Mathematical Modeling of Heterogeneity and Drug Response in Lung Cancer

By

David J. Wooten

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Cancer Biology

September 30, 2018

Nashville, Tennessee

Approved:

Jonathan Irish, Ph.D.

Mark Ellingham, Ph.D.

Ken Lau, Ph.D.

Carlos Lopez, Ph.D.

Vito Quaranta, M.D.

ACKNOWLEDGMENTS

I offer my sincere gratitude to the colleagues and friends that I have gained during my studies these past years. First, I offer thanks to Akshata Udyavar, who developed the foundation of my work in small-cell lung cancer, and was open enough to gave me a chance and a tangible project to start on. Thank you for all you taught me, and the good work we did together.

I am also extremely grateful for the collaborative work I was able to do with Christian Meyer, Sarah Maddox, Leonard Harris, and Peter Frick. This dissertation would not have been possible without all of my great discussions shared work with you, and your collective willingness to explore undeveloped ideas with me. Including ALL the arguments!

Much thanks also to Shawn Garbett. Whenever I had had enough with biology, I could always count on you to discuss the craziest and most esoteric corners of math, physics, computer science, and yes, even cheese.

Finally I would like to thank the Vanderbilt Cancer Biology program, Quantitative & Chemical Biology program, all the members of my committee, and especially my adviser, Vito Quaranta. Thank you all for your guidance, your patience, and your willingness to give a chance to a student who was way out of his depths.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Despite President Richard Nixon signing the National Cancer Act of 1971, publicly declaring 'war' on cancer, in 2018 cancer remains the 2nd leading cause of death in the world. Nevertheless, substantial progress has been made over the long history of this disease. Surgical resection, which is still often the most effective treatment for noninvasive solid tumors, could do little for metastatic disease. Chemotherapy, which gained prominence through the 1940's and 50's offered a systemic approach that targeted and killed cancer cells all throughout the body (DeVita and Chu, 2008). Nevertheless, treatment with a single chemotherapeutic agent was generally insufficient to eradicate cancer from the body, leading to combination chemotherapies, and eventually a surge in targeted therapies (Joo et al., 2013).

The idea of targeted therapies, which gained considerable momentum in the 1990's due to the success of imatinib in the treatment of chronic myeloid leukemia, was to identify a specific genetic or molecular aberration that distinguishes a patient's cancer from their healthy cells. A successful targeted therapy will potently kill cells harboring that aberration, while having minimal or no impact on healthy cells. Nevertheless, despite several successes, even targeted therapies often fail eventually (Huang et al., 2014; Bozic et al., 2012).

1.1 Heterogeneity and relapse in cancer

A common wisdom has emerged in oncology that no two cancers are the same. This is further exacerbated within a patient who might have several disconnected tumors with different types, or even different cells within the same tumor that are genetically and molecularly distinct. In 1976, Peter Nowell proposed a model of clonal evolution within a pop-

ulation of tumor cells (Nowell, 1976), showing that treatment with a drug that kills most of the tumor may simply be putting evolutionary pressure, selecting for the eventual emergence of a population of tumor cells resistant to treatment. Nevertheless, in many cancers, development of targeted therapies has had major a impact on prolonging and improving patients' lives.

1.2 Lung cancer

Lung cancer is a highly prevalent disease, and comes in second only behind prostate (in men) and breast (in women) cancer for new diagnoses each year. However, lung cancer is still the number one cause of cancer related deaths in both men and women (American Cancer Society, 2018). Lung cancer is broadly divided into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), the latter type encompassing lung adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. Both types carry poor prognoses, with 5-year survival rates of about 30% for NSCLC, and 10% for SCLC (American Cancer Society, 2018). Targeted therapies, especially for EGFR and ALK, have in recent years produced a major impact on the treatment of patients with advanced NSCLC (Chan and Hughes, 2015), however, patients with SCLC are still prescribed a combination of chemo- and radiation-therapy as standard of care (Rudin et al., 2015).

1.2.1 Small-cell lung cancer

Small Cell Lung Cancer (SCLC) is a neuroendocrine tumor representing 15% of lung cancer. Due to its fast growth, early metastatic spread, and rapid relapse following dramatic chemotherapy responses, it is the most lethal form of lung cancer. Current standard of initial care is confined to combination chemo- or radiation-therapy (Table 1.1), regardless of stage, alongside prophylactic cranial irradiation to delay onset of tumor metastasis to the brain (Rudin et al., 2015). Proposals to stratify patients based on tumor phenotype have been met with resistance due to unclear clinical relevance and lack of actionable informa-

Stage	Standard Treatment Options
Limited Stage	Chemotherapy and radiation therapy
	Combination chemotherapy alone
	Surgery + chemotherapy or chemoradiation therapy
	Prophylactic cranial irradiation
Extensive Stage	Combination chemotherapy
	Radiation therapy
	Prophylactic cranial irradiation
Recurrent Disease	Chemotherapy
	Palliative therapy

Table 1.1: Standard of care options for patients with limited, extensive, or recurrent stage disease.

tion (Hiltermann et al., 2012; Jones et al., 2004; Subramanian and Simon, 2010; Travis, 2012). Indeed, SCLC cells are strikingly uniform by histopathology, displaying a classic "small blue round cell" morphology. Nevertheless, recent years have yielded an explosion of reports of SCLC phenotypic heterogeneity (George et al., 2015; Borromeo et al., 2016; Mollaoglu et al., 2017; Lim et al., 2017; Jahchan et al., 2016; Calbo et al., 2011; Udyavar et al., 2017), and it is widely anticipated that improved clinical outcomes will emerge from phenotypic SCLC stratification (George et al., 2015; Mollaoglu et al., 2017; Lim et al., 2017; Jahchan et al., 2016). Heterogeneity was implicated in the aggressiveness of the disease when Calbo *et al.* (Calbo et al., 2011) demonstrated that interactions between two distinct SCLC phenotypes, one with more neuroendocrine character and one with more mesenchymal character, were necessary to promote metastasis. Others have identified distinct SCLC phenotypes characterized by MYC (Mollaoglu et al., 2017), NEUROD1 (Borromeo et al., 2016; Mollaoglu et al., 2017), NOTCH (George et al., 2015), REST (Lim et al., 2017), and others. Indeed, a combination of older (Carney et al., 1985) and newer (George et al., 2015; Mollaoglu et al., 2017; Lim et al., 2017; Jahchan et al., 2016; Calbo et al., 2011; Udyavar et al., 2017) observations indicate that the uniform morphology may

hide a variety of functionally distinct phenotypes with the ability to form an SCLC ecosystem that is robust to perturbations and treatment.

One promising avenue for treatment strategies in SCLC is based on an understanding of how SCLC cells, within a landscape of heterogeneous phenotypes, are able to evade treatment and acquire drug resistance. Phenotype functional diversification, within the constraint of "small blue round cells", may provide escape routes so that cells transition from sensitive to resistant phenotypes either prior to or soon after a drug challenge (Udyavar et al., 2017). Nevertheless, there is no standardized atlas of SCLC phenotypic heterogeneity, and similarities between reported phenotypes are not clear, especially across the diverse range of model systems. Defining phenotypes at the molecular and genetic levels, across model systems, and quantifying their routes and rates of plasticity is thus critical for connecting reported heterogeneity with patients' tumor clonal dynamics. The ability to predict drug-response trajectories based on initial tumor phenotypic composition and detected changes over time has major implications for drug discovery and treatment optimizations in SCLC. Finally, strategies to "reprogram" SCLC cells to a sensitive state may provide an alternative or adjuvant avenue to chemotherapy. Thus, we are interested in understanding the relationship between SCLC cell identity and drug response.

1.3 Modeling cell differentiation and reprogramming in cancer

1.3.1 Acknowledgements

This section is derived from work previously published. Reprinted by permission from Elsevier B.V.: Wooten DJ and Quaranta V, Mathematical models of cell phenotype regulation and reprogramming: Make cancer cells sensitive again!, *Biochimica et Biophysica Acta - Reviews on Cancer*, 2017, v1867:167-175 (Wooten and Quaranta, 2017).

1.3.2 Introduction

Cancer is traditionally viewed as a genetic disease caused by the random accumulation of mutations in critical genes or pathways that control proliferation and other hallmark traits (Hanahan and Weinberg, 2011). Heterogeneity within a tumor would then arise through classic Darwinian evolutionary processes of mutation and clonal selection (Nowell, 1976). Expansion of heterogeneous phenotypes can then limit the effectiveness of treatment which is inevitably directed to the majority (average) clones, as insensitive phenotypic variants emerge.

However, it is becoming increasingly clear that the phenotype of a cancer cell is not just determined by its genotype. Epigenetic (Ballestar and Esteller, 2008) and microenvironmental (Mbeunkui and Johann Jr, 2009) factors provide additional significant contributions, such that two cancer cells with identical genotype may actually exhibit distinct phenotypes (Figure 1.1).

This establishes provocative parallels with embryonic development in which a single genome can give rise to widely diverging differentiated phenotypes. This review discusses key advances toward a systems-level understanding of cell identity and reprogramming, first in the context of normal development, then connecting it to cancer heterogeneity and evolution. We review the foundational theory of cellular reprogramming, and discuss quantitative methods to predict or improve reprogramming efficiency and outcomes. In the last section, we position these exciting recent cell biological breakthroughs in the context of cancer heterogeneity and tumor evolution.

1.3.3 Cellular Reprogramming

Within developmental biology, the traditional dogma of cellular differentiation has been that an organism begins as a zygote which gives rise to pluripotent stem cells (Gilbert, 2016). Upon division, environmental cues (Derynck and Akhurst, 2007) or stochastic ef-

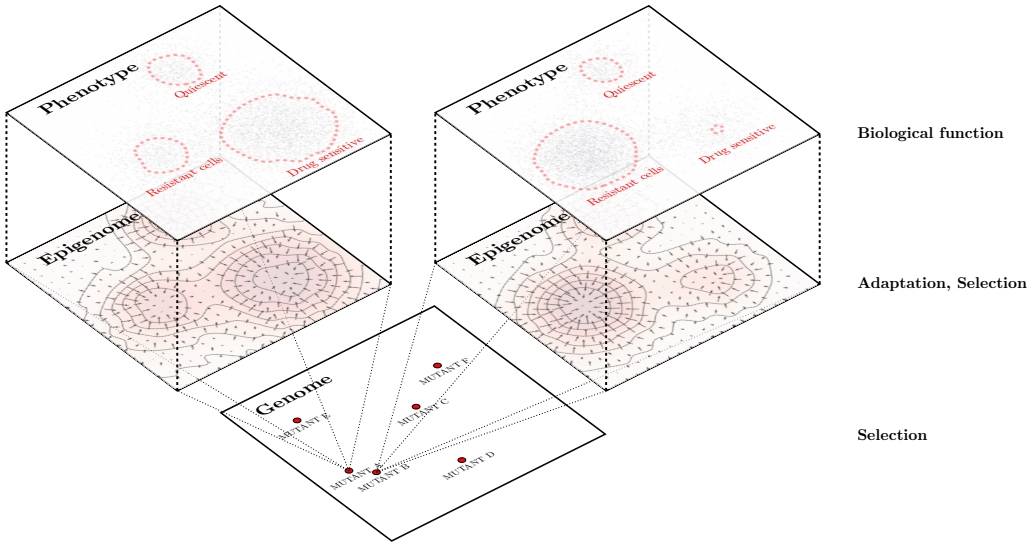


Figure 1.1: Levels of biological regulation. Due to genetic instability, cancer cells in a tumor may have several distinct genomes (bottom). Each distinct genome underlies a unique epigenetic landscape (middle), which defines what cell phenotypes are possible (top). This allows both for cells with identical genomes to adopt distinct phenotypes, and also for cells with distinct genomes to identify identical phenotypes (left VS right). This opens up powerful possibilities for reprogramming cancer cells with diverse genetic backgrounds into more treatable or less malignant phenotypes.

fects (McHale and Lander, 2014) can give rise to a hierarchy of cells with increasingly differentiated states. Differentiation was considered an irreversible process, in which histone modifications and DNA methylation controlled the accessibility of key DNA regions (Gilbert, 2016) through opening or closing of the chromatin structure. However, differentiated cells still maintain all the DNA required for pluripotency. Indeed, it was shown that implanting the nucleus of a somatic cell into a denucleated oocyte could produce a stem cell (Hwang et al., 2004), demonstrating the existence of unknown regulatory mechanisms in the oocyte cytoplasm which were able to re-activate the locked pluripotency state.

In 2006, Takahashi and Yamanaka found a set of 4 transcription factors (TFs): Oct3/4, Sox2, Klf4, and c-Myc (collectively "OSKM"), that could cause dedifferentiation of mouse embryonic fibroblasts to induced pluripotent stem cells (iPSCs) (Takahashi and Yamanaka, 2006). On average, though, only about 0.05% of transduced cells underwent transformation, and upon relaxation of the TF cocktail, the cells fell back into their previous, differen-

tiated state (Okita et al., 2007). Other studies have extended the gamut of reprogrammable cell types (Buganim et al., 2013), and while in many cases efficiency has been greatly improved (Hanna et al., 2009; Rais et al., 2013), deciphering the regulatory programs controlling cell identity promises to enable many biomedical applications (Khalil and Collins, 2010; Cherry and Daley, 2012), and may have future impacts on cancer therapy.

The OSKM TFs have been classified as "pioneer transcription factors," able to bind enhancers in a closed chromatin state (Soufi et al., 2012). These factors were also found to promiscuously activate multiple off-target genes, such that final establishment of the pluripotent state very likely requires a system-wide rebalancing of the gene regulatory network (GRN) (Soufi et al., 2012).

Mathematical modeling is appropriate for understanding the GRN dynamics underlying this rebalancing, and could accelerate discovery of key TFs to reprogram cells to a target phenotype. This may be especially true in cancer cells, in which phenotypes are often not clearly classifiable, especially with respect to treatment sensitivity. In the next sections, we will discuss theoretical frameworks which aim to clarify the topology of epigenetic landscapes in mathematical terms, and could help resolve the nature of cancer cell phenotypes and their drivers.

1.3.4 The epigenetic landscape and theory of attractors

In 1957, CH Waddington proposed the concept of an epigenetic landscape (Waddington, 1957) (Figure 1.2), in which cells roll downhill through bifurcating channels representing differentiation pathways. As cells progress down these metaphorical slopes, they become increasingly committed to a terminal phenotype at the bottom. Distinct pathways are separated by ridges, confining cells to their differentiated identity. While this framework was intended purely as a conceptual tool to obtain a "rough and ready picture" that "cannot be interpreted rigorously" (Waddington, 1957), it was nonetheless developed within the mathematical context of dynamical systems theory.

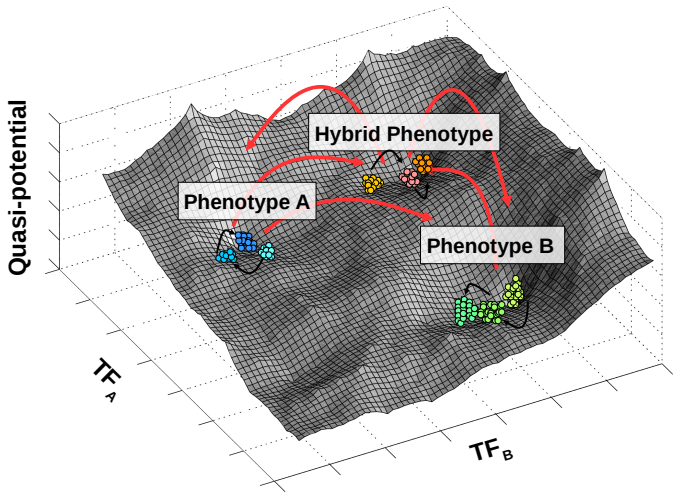


Figure 1.2: Epigenetic landscape. Waddington envisioned cell differentiation as cells rolling downhill in an "epigenetic landscape", becoming increasingly more committed to a terminal cell fate. Here we show an artistic rendering of a bumpy epigenetic landscape characterized by a quasi-potential. Cells move downhill and settle into one of the minima, termed "basins of attraction". Some cell states may be more or less stable, depending on the height of the energy barrier between them. Small perturbations will not push cells out of their basin of attraction, however large and properly directed perturbations can reprogram cells by moving them into a neighboring basin (shown as arrows depicting transitions).

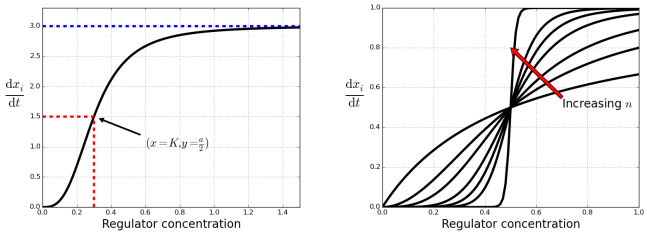
This was reasonable since, within this theory, stable states (named attractors) commonly arise from dissipative systems which must exchange energy and matter with their environment to sustain function (Gros, 2008), a seemingly realistic and necessary behavior for cells. Thus, biologically, an attractor describes a state in which a cell identity can stably persist.

Over the past 50 years, several researchers have taken on the task of formalizing this attractor framework in the context of biology, in order to understand how signaling pathways and GRNs may robustly coordinate cell behavior (Ferrell, 2012; Kauffman, 1969b,a; Huang et al., 2005; Huang, 2012; Wang et al., 2010; Choi et al., 2012; Tyson et al., 2003; Woodcock, 1978; Huang, 2002; Zhou et al., 2012; Wang et al., 2008; Li and Wang, 2014, 2013a). The next section highlights these efforts and their potential relevance to cell reprogramming.

1.3.5 Gene regulatory dynamics and attractors

Stuart Kauffman proposed the idea of Boolean network models, in which genes can either be ON or OFF (Kauffman, 1969b), in order to simulate the dynamics of GRNs. His models revealed that networks with certain structural properties did indeed settle into a small number of stable attractors, providing the first evidence that cell types may correlate with GRN attractors (Kauffman, 1969b,a). A few decades later, Huang and coworkers (Huang et al., 2005) provided an experimental justification for this intriguing idea. In mathematics, attractors by definition have an associated region called the "basin of attraction" (Figure 1.2) corresponding to all states that will eventually approach the attractor (Strogatz, 2014). Huang and coworkers exploited the fact that HL60 human promyelocytic leukemia cells can be induced to differentiate into neutrophils via treatment with either DMSO or ATRA. Tracking the trajectory of a 2773-gene expression panel, they showed that HL60 cells respond divergently to treatment with these agents. However, the trajectories eventually converged to an identical neutrophil state. They reasoned that the divergent trajectories must both have been within the basin of attraction of the neutrophil state, which must then be an attractor (Huang et al., 2005).

This basin of attraction is ultimately responsible for the stability of an attractor, as small deviations from the attractor will remain confined within the same basin (Figure 1.2)(Kaneko, 2006; Strogatz, 2014). Within this framework, it has been suggested that reprogramming can be achieved by forcing a cell out of its basin of attraction either through an external push, or an internal rewiring which causes the landscape to shift, such that the cell will be ultimately drawn toward an alternative attractor (Figure 1.2) (Wang et al., 2011; Kaneko, 2006; Zhou et al., 2011; Zhou and Huang, 2011; Li and Wang, 2013b; Huang, 2009; Zhou and Huang, 2013; MacArthur et al., 2008; Huang, 2002). These efforts provided a motivation for developing quantitative models to connect the dynamics of GRNs to cellular phenotype and reprogramming, as described in the next section.



(a) (b)

Figure 1.3: Hill kinetic gene regulatory functions. Hill kinetics are commonly used to quantify the dynamics of GRNs. (a) A single activating interaction showing the characteristic sigmoidal shape of the Hill curve. Low regulator concentrations do little to induce expression of the target gene, however as the concentration approaches a threshold value (red dotted line) the regulatory strength increases rapidly. At the red dotted line, the regulator has half of its maximum possible influence on its target. Further increases in concentration gradually cause the influence to asymptotically approach the maximum regulation (blue dotted line). Here, $a = b = 3$, $K = 0.3$, and $n = 3$. (b) The Hill curve becomes more switchlike as n becomes larger. Here, $a = b = 1$, $K = 0.5$, and $n \in \{1, 2, 3, 5, 7, 10, 50\}$ as the switch becomes more instantaneous.

1.3.6 Quantifying, drawing, and analyzing the epigenetic landscape

For biological systems, GRNs are comprised of hundreds to thousands of interacting genes. Mathematically, each gene represents a single dimension, and it is possible to routinely compute with all these dimensions. However, most humans can only visualize a maximum of 3 dimensions at once (2 genes, plus the landscape height, Figure 1.2), so that it is in general impossible to visually represent the entire high-dimensional epigenetic landscape for real biological systems. Nevertheless, quantifying this landscape can provide information about the barriers between separate basins, and visualizing 2D slices can promote an intuitive understanding of cells' behaviors.

A common method to quantify the dynamics of a GRN is based on Hill kinetics (de Jong, 2002; Wang et al., 2008, 2011; Li and Wang, 2014; Zhou et al., 2011; Foster et al., 2009; Li and Wang, 2013a; Savageau, 1995) (Figure 2). When each regulator of a given TF (say, x_i) acts independently, the dynamics take the form

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \sum_{j \in activators} \frac{a_{j,i} x_j^{n_{j,i}}}{K_{j,i}^{n_{j,i}} + x_j^{n_{j,i}}} + \sum_{j \in repressors} \frac{b_{j,i} K_{j,i}^{n_{j,i}}}{K_{j,i}^{n_{j,i}} + x_j^{n_{j,i}}} - k_i x_i \tag{1.1}$$

where $a_{j,i}$ and $b_{j,i}$ represent the maximum contribution or inhibition of the j^{th} TF on expression of x_i (Figure 1.3), $K_{j,i}$ reflects the threshold for the j^{th} TF to influence x_i, $n_{j,i}$ controls how switchlike or gradual the regulation is (Figure 1.3), and k_i represents the natural decay rate of x_i. More complicated forms are able to account for interactions between TFs (Buchler et al., 2003).

The epigenetic landscape is often thought of as analogous to potential energy from physics, and is ideally computed as a function, $U(\vec{x})$, such that cells roll down the gradient of $U(\vec{x})$ under the dynamics of the GRN

$$\frac{\mathrm{d}\vec{x}}{\mathrm{d}t} = -\nabla U(\vec{x}) \tag{1.2}$$

In practice, however, even simple GRNs are not gradient systems (Ferrell, 2012; Wang

et al., 2008; Zhou et al., 2012; Bhattacharya et al., 2011), and cannot be fully described by Equation (1.2). One proposed method to overcome this is to split Equation (1.1) into a portion which is describable by a landscape and some remainder (Wang et al., 2008; Zhou et al., 2012)

$$\frac{\mathrm{d}\vec{x}}{\mathrm{d}t} = -\nabla\widetilde{U}(\vec{x}) + \vec{F}_{remainder}(\vec{x}) \tag{1.3}$$

where $\widetilde{U}(\vec{x})$ is commonly referred to as a "quasi-potential" (Figure 1.2). Equation (1.3) essentially separates the dynamics of relaxing *toward* the attractor ($-\nabla\widetilde{U}(\vec{x})$) from the dynamics *along* the attractor ($\vec{F}_{remainder}(\vec{x})$) (Wang et al., 2008; Li and Wang, 2014). However, there are infinitely many ways to construct such a decomposition (Equation (1.3)) and the precise method chosen will influence the resulting landscape and difficulty of computation (Zhou et al., 2012). Nevertheless, given a landscape it is possible to identify pathways of least resistance (Li and Wang, 2013a; Wang et al., 2010, 2011; Zhou et al., 2012; Zhou and Huang, 2013) for cells to transition between attractors, and it has been proposed (Li and Wang, 2013a; Zhou and Huang, 2013) that such paths may provide a roadmap for efficient reprogramming strategies. Indeed, by varying parameters, from Equation (1.1) it is possible to quantify how much each individual interaction contributes to the barrier height between attractors. In a model of stem cell differentiation, Li found that previously established reprogramming factors had a large impact on the barrier height along the path of least resistance between attractors in a model of stem cell differentiation (Li and Wang, 2013a).

Alternative methods have been proposed to quantify the energy landscape, including network entropy (Banerji et al., 2013) and a stitched-together landscape of potential energy changes (Bhattacharya et al., 2011), and although these methods are more phenomenological, they have been shown to qualitatively agree with intuitive expectations, and may be appropriate for very high-dimensional systems.

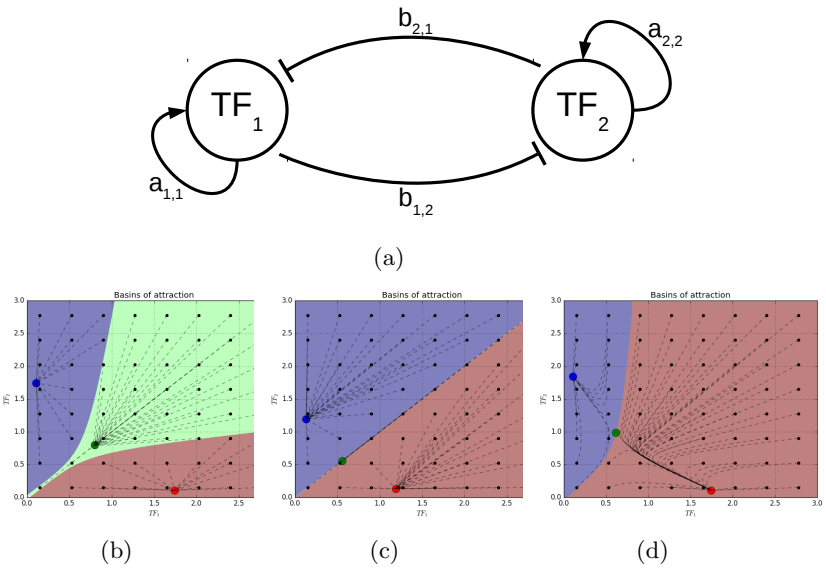


Figure 1.4: Regulatory networks and basins of attraction. (a) A common regulatory motif which consists of two TFs which repress one-another's expression, while activating their own. (b-d) Possible attractors and basins of attraction for different parameter sets from Equation (1.1). While there is a stable "hybrid" state in (b), it can be destabilized as in (c) and (d) by reducing the strength of self-activation. Similarly, by adjusting the activating and repressing parameters, one phenotype may become more epigenetically favorable (d) than the other (c).

1.3.7 Boolean network models can identify reprogramming strategies

Equation (1.1) models regulatory interactions as sigmoidal functions (Figure 1.3), so that low concentrations have no regulatory impact, and high concentrations saturate toward a maximal impact. By letting the Hill coefficient n become large, the transition from zero to maximal impact becomes instantaneous (Figure 1.3)(Alon, 2006), mimicking the Boolean structure proposed by Kauffman (Kauffman, 1969b). Boolean models therefore represent a coarse-grained approach which is useful to derive qualitative results.

Using Boolean models to describe cellular reprogramming, Crespo and colleagues (Crespo et al., 2013) developed network models of several biological systems. Simulations of activations and knockdowns which drove the systems between attractors revealed

experimentally validated cocktails of reprogramming factors.

Alternatively, Lang and colleagues (Lang et al., 2014) built a Hopfield neural network (a type of Boolean model) to simulate dynamics leading toward 63 distinct cell states. This approach revealed multiple attractors between cell types, which they suggested may be responsible for the partial reprogramming observed experimentally. Furthermore, they computed the contribution of each TF to the stability of a state and were able to identify many known reprogramming TFs.

1.3.8 Simple models reveal topological properties of reprogramming

It has been proposed that GRNs maintain separated basins through clusters of TFs which mutually activate one another's expression via positive feedback, and reciprocally inhibit expression of alternative clusters through negative feedback (Angeli et al., 2004; Crespo and Del Sol, 2013; Kim and Wang, 2007; MacArthur et al., 2008) giving rise to multistability (Thomas, 1978; Ferrell, 2012; Huang et al., 2007; Foster et al., 2009) (Figure 1.4). Intuitively, once one module becomes dominant, it simultaneously reinforces its own activity, while silencing its repressors. Depending on the parameters of such systems, which are potentially influenced by multiple external factors (Ferrell, 2012; Huang et al., 2007), stable states may become destabilized (Figure 1.4). This is one way to force cells into the basin of an alternative attractor (Figure 1.2). Thus, positive and negative feedback loops are a key driver of multistability, and represent a promising motif for cellular reprogramming.

Inspired by this observation, Crespo and Del Sol (Crespo and Del Sol, 2013) identified reprogramming factors by simply searching for positive feedback loops whose elements are differentially expressed between distinct cell states. Within these feedback loops, genes with maximum out-degree were predicted to be master regulators. This technique was able to identify known reprogramming recipes for the several cell types they tested (Crespo and Del Sol, 2013).

Other researchers have studied structural properties of simple models in which entire clusters of TFs are simulated as single, auto-activating and cross-inhibiting nodes (Figure 1.5) (MacArthur et al., 2008; Artyomov et al., 2010; Foster et al., 2009). For instance, MacArthur and colleagues (MacArthur et al., 2008) built a simple model for stem cell differentiation by coupling 3 auto-activating lineage-specific modules and a pluripotency module with negative feedback loops. They found that increasing differentiation factors led to an irreversible progression through pluripotent, tripotent, bipotent, and finally terminally differentiated states. Noise added to the pluripotency module enabled differentiated cells to stochastically dedifferentiate, while noise in other modules had no effect.

This TF clustering approach was adopted by Cahan, Morris, and coworkers (Cahan et al., 2014; Morris et al., 2014) who developed CellNet to score performance of reprogramming experiments between a predefined set of 21 cell types (though it has now expanded to include 36: 16 human cell types and 20 mouse), and suggest TFs to push partially reprogrammed cells the rest of the way. To build CellNet, the authors identified network-based clusters of TFs with specific enrichment for a distinct cell type. TFs are prioritized as reprogramming targets based on their differential expression levels and centrality within these clusters. Applying their method to established protocols of cellular reprogramming, the authors consistently found evidence of partial reprogramming. In many cases, further manipulation of prioritized TFs improved target cell function and stability (Morris et al., 2014).

Others have considered GRNs arranged as a hierarchy of positive and negative feedback loops, which lead to hierarchies of differentiation (Foster et al., 2009; Zhou et al., 2011). Artyomov and coworkers studied topological reprogramming properties of such networks (Artyomov et al., 2010), and observed that successful reprogramming was only achievable by targeting TF modules of the immediate progenitor or the immediate neighbor of the current state. (Artyomov et al., 2010).

Rackham and colleagues developed Mogrify (Rackham et al., 2016) based on a simi-

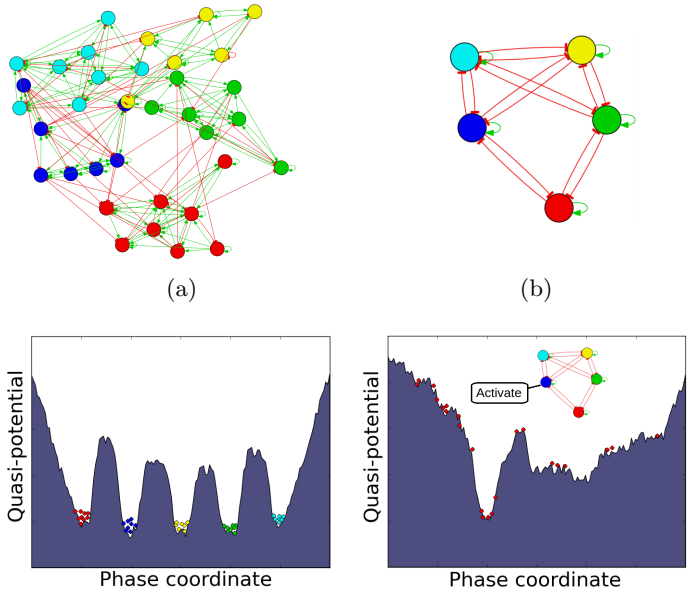


Figure 1.5: Multistable regulatory network and reprogramming. (a) Example topology of a GRN extending the motif in Figure 1.4 in which TFs specific to a certain cell type (given as colors) predominately promote one another's expression through positive feedback loops, and inhibit TFs specific to alternative cell types. (b) This topology is often ammenable to reducing to interactions between clusters of TFs. (c) Such topologies commonly give rise to multistable systems where each cluster of TFs are activated in particular phenotypes. (d) Reprogramming by activating TFs from a target phenotype may destabilize several of the phenotypic attractors, but may still maintain a rugged landscape with partially reprogrammed cells and off target effects. Further activations or knockdowns can alleviate this and result in more complete and stable reprogramming.

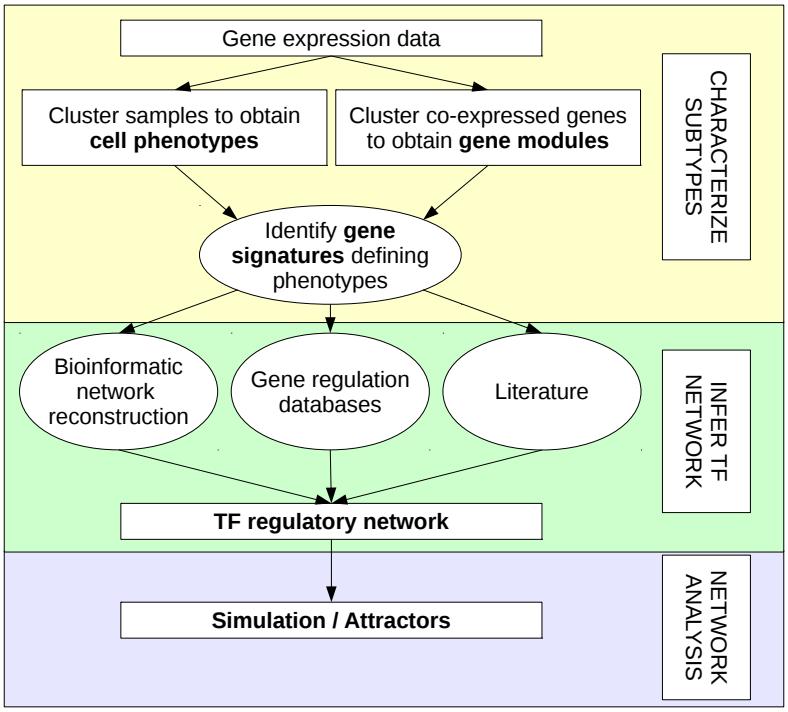


Figure 1.6: Network construction and analysis workflow. A workflow we used in our recent work (Udyavar et al., 2017) in which we identify and model a TF network controlling phenotypic heterogeneity in small-cell lung cancer. It consists of three phases. First, analyzing gene expression data to identify distinct cellular phenotypes and co-expressed clusters of genes which distinguish those phenotypes (as in Figure 1.5). Second, identifying TFs which regulate those co-expressed gene clusters and assembling them into a TF network. Finally, simulating the dynamics of the TF network to identify stable attractors. We used a Boolean modeling approach to find SCLC phenotype attractors.

lar principle by prioritizing TFs based on their proximity to differentially expressed genes. They integrated topological information from gene regulatory networks from the STRING database (Szklarczyk et al., 2011) and FANTOM consortium (Kawaji et al., 2009), and assign a score to TFs based on a combined score of their own differential expression between the two tissue types, and also by weighting the differential expression of downstream target genes based on their regulatory distance and specificity. They have applied this approach to prioritize TFs to reprogram between any of 173 cell types and 134 tissues in humans, identifying and validating several novel reprogramming experiments.

1.3.9 Reprogramming in cancer

While often thought of as a purely mutational disease, cancer is recognized as having a significant epigenetic component (Ballestar and Esteller, 2008), supporting the idea that the methods of cellular reprogramming may find new applications in treatment.

Tumors are known to be heterogeneous cell populations including multiple types of cancer cells as well as various non-malignant supporting cells. The behaviors and interactions between all these cell types leads to the overall phenotype of the tumor. For instance, clonal evolution in cancer can lead to selection of resistant subpopulations of tumor cells during treatment (Greaves and Maley, 2012). The cancer stem cell hypothesis suggests that tumors have a small population of cells which persist through treatment, and which are able to give rise to and replenish the entire pool of tumor phenotypic diversity (Jordan et al., 2006).

Different groups have shown that tumor cells can be reprogrammed to iPSCs using the same factors as in normal cells (Miyoshi et al., 2010; Zhang et al., 2013; Suvà et al., 2014). For instance, Miyoshi and colleagues (Miyoshi et al., 2010) found that the Yamanaka OSKM factors were able to induce an embryonic stem-cell-like phenotype in gastrointestinal cancer cell lines, and found that these transformed cells showed reactivation of tumor suppressor genes, increased chemosensitivity, and decreased invasiveness.

Likewise, Zhang and coworkers (Zhang et al., 2013) showed that overexpression of Oct4, Nanog, Sox2, Lin28, Klf4, and c-Myc was able to reprogram human sarcomas into iPSCs. Furthermore they showed that these tumorous iPSCs could be redifferentiated into connective tissue and red blood cells, and that this process resulted in a loss of tumorigenicity.

Glioblastomas (GBMs) contain rare populations of tumor propagating cells with stem-like properties which are known to drive tumor progression and therpeutic resistance. Suva and colleagues (Suvà et al., 2014) showed that activation of Pou3f2, Sox2, Sall2, and Olig2 is able to transform GBM cells to behave like these tumor propagating cells.

Taking another direction, Vêncio and colleagues (Vêncio et al., 2012) showed that reprogramming the tumor microenvironment could also have a therapeutic effect. They found that CD90+ prostate cancer-associated stromal cells could be reprogrammed to iPSCs by overexpressing Pou5f1, Lin28, Nanog, and Sox2, causing the reprogrammed stromal cells to lose their tumor supporting phenotype.

These studies have relied on the vast amount of knowledge and expertise developed in the stem-cell reprogramming field, and serve as proof-of-principles to establish reprogramming as a viable avenue for treatment. However, the extreme diversity of cancer heterogeneity provides challenges and opportunities for additional reprogramming strategies between different cancer-specific phenotypes, or even reverting the cancer phenotype to a non-malignant one. Prioritizing targets to drive conversions between these different phenotypes will benefit from the application of the computational and mathematical frameworks reviewed in this paper.

Indeed, it has been argued that the view of cell types as attractors of some epigenetic landscape is applicable not only to understand epigenetic regulation in normal tissue, but also cancer and cancer heterogeneity. For instance, several groups have advocated that the malignant state itself represents an attractor of some epigenetic landscape, and that healthy, non-malignant states may still be reachable via reprogramming (Kauffman, 1971; Huang et al., 2009; Li et al., 2015; Szedlak et al., 2014).

Additionally, other groups including ourselves have found that distinct TF network attractors specify heterogeneous cancer phenotypes. For instance, we (Udyavar et al., 2017) (Chapter 2 here) found that neuroendocrine-epithelial and mesenchymal-like heterogeneity in small-cell lung cancer could be explained as attractors of a TF network derived following a mixed bioinformatics, modeling, and experimental workflow in Figure 1.6. Significantly we found that treatment with chemotherapeutics caused shifts toward a hybrid phenotype, suggesting that epigenetically modulated cancer heterogeneity contributes to treatment resistance. Reprogramming cells out of resistant states and into sensitive ones is likely to

improve treatment effectiveness in such cases.

Many of these results have also been demonstrated by other groups using microRNA perturbations to reprogram cancer cells, instead of TFs. This builds off of reports that microRNAs can efficiently reprogram cells (Anokye-Danso et al., 2011; Hu et al., 2013). Tsuno and colleagues (Tsuno et al., 2014) showed that lentiviral induction of miR-520d was able to reprogram hepatoma cells to a stem-like state which, when injected into mice, showed no malignancy. Importantly, Ogawa and colleagues (Ogawa et al., 2015) administered miR-302 and miR-369 *in vivo* to reprogram colon tumors to less malignant states, showing that the reprogrammed tumor cells had higher activation of apoptosis. Other groups have also shown the potential of miR-302 to reprogram cancer cells toward a stem-like state (Lin et al., 2008; Koga et al., 2014), and may improve drug sensitivity (Koga et al., 2014).

1.3.10 Moving to the clinic: Opportunities and Challenges

One of the primary challenges facing cancer researchers and clinicians today is that while treatment may be initially efficacious, the tumor eventually relapses, and has become refractory to further treatment (Kottke et al., 2013). It would therefore be powerful to find ways to reprogram tumor cells back to a sensitive phenotype, or adjuvant treatments which prevent the emergence of resistant phenotypes altogether.

Early efforts to develop such treatments will benefit by focusing on cancer types with well-established signatures of heterogeneity, such as small-cell lung cancer (Udyavar et al., 2017), glioblastoma (Parker et al., 2016), non-small-cell lung cancer (Boutros et al., 2009), or breast cancer (Almendro et al., 2014b; Dai et al., 2015), and phenotype-specific drug sensitivity patterns. Based on these gene signatures, a workflow like that shown in Figure 1.6 can be used to to identify a GRN underlying that cancer's epigenetic landscape, and transitions between sensitive and resistant attractors.

Clinically, transcriptomic profiling of multiregion biopsies (Gerlinger et al., 2012) or of

circulating tumor cells (liquid biopsies) (Hayes and Paoletti, 2013) could be used to determine the specific phenotypic composition of a patient's primary tumor and/or metastases. Optimal combinations of TF activations or knockdowns for this specific patient can then be identified to stabilize the sensitive cells, and destabilize the resistant ones.

Historically it has been difficult to develop chemical therapeutics targeting transcription factor function, but recent advances have shown that this is a promising area of future research (Bhagwat and Vakoc, 2015). Nevertheless, there are several, alternate approaches which may result in indirect perturbations of a patient's TF network, including combinations of drugs, signaling pathways, microRNAs, and epigenetic agents.

Bioinformatically, the Connectivity Map (Lamb et al., 2006) and LINCS L1000 (Duan et al., 2014) databases provide information about gene expression changes resulting from pharmacological or genetic perturbations. Interrogating these online resources with appropriate datasets from clinical tumors may help identify existing approved compounds which are statistically linked to activation or inhibition of target TFs. Nevertheless, the mechanism of these changes is generally not well understood, and off-target effects may limit the effectiveness of this approach.

Mechanistically, signal transduction pathways carry information through a cell and often end in the activation, deactivation, and/or degradation of one or many TFs. Importantly, the phosphorylation events that carry information through signaling pathways are well suited to pharmacological inhibition or activation.

Another promising approach is through clinical overexpression or inhibition of microRNAs. Currently no microRNA therapeutics are FDA approved, however there are several under preclinical investigation, and a few in clinical trials. Significantly, as discussed above, in some cases microRNAs have been shown to be effective reprogramming agents, including in cancer, and they may therefore become important components of reprogramming therapies.

Gene expression is intrinsically limited by the accessibility of a region of DNA, and

chromatin structure can help lock in a cell's identity. Chemical agents which act as histone deacetylase inhibitors or DNA methyltransferases can cause extensive changes in gene expression and plasticity, and have been shown as effective reprogramming agents. Furthermore, there are several FDA approved epigenetic therapeutics which may be used in combination with other strategies to improve reprogramming efficacy.

Biologically, the actions of signaling networks, microRNAs, chromatin structure, and TF regulation occur over several timescales, from seconds or minutes (within a single cell's lifetime), to hours and days (spanning cell generations) (Voss and Hager, 2014). We foresee that a key theoretical challenge will be to reconcile the dynamics of these different timescales, and across cell divisions.

1.3.11 Discussion

60 years after its introduction, Waddington's epigenetic landscape remains a powerful metaphor to understand differentiation and reprogramming. Recent advances in cancer biology have painted a clearer picture of the importance of epigenetic regulation in maintaining malignancy and heterogeneity, and indeed several studies have already shown the feasibility of using TFs to reprogram cancer cells, particularly in the context of differentiation.

Analysis of stability, perturbations, and topology have all been applied to identify possible reprogramming targets with great success in normal developmental biology, and provided insights into how cells maintain barriers between differentiated states through feedback loops.

However, it is not always clear how dysregulation in cancer GRNs may give rise to unique phenotypes. Nevertheless, recent results have suggested that these same tools may be useful to decode cancer epigenetic regulation.

As our understanding of epigenetic regulation improves, we should be able to create detailed models which provide a theoretical, high-resolution depiction of the control of

cancer and cancer heterogeneity which will be indispensable in the search for therapeutic options. The outcome of such research could fundamentally reverse the standard approach to personalized therapy, in which new drugs are developed to match specific subtype of cancer. Instead, we may find ways to develop therapies which reprogram resistant cells into a sensitive state, matching cells to the drug.

1.4 Outline of dissertation

In this dissertation, I will report the methods and results of my studies into the relationship between phenotypic heterogeneity, cell identity, and drug response in cancer, as well as methods to identify systems-based therapies, and quantify their effectiveness. The first two chapters report the results of my investigation into phenotypic heterogeneity in SCLC. I identify several transcriptionally and functionally distinct SCLC phenotypes, including the characterization of a multi-drug resistant, and develop mathematical models to understand the key regulatory genes that stabilize these phenotypes. The next two chapters report the results of my investigation into modeling the drug response of heterogeneous populations of cells, as well as a new method to quantify the synergystic behaviors of combinations of drugs. In total, this work represents progress towards a systems-level understanding of the relationship between cancer, heterogeneity, cell identity, and drug response.

Chapter 2

Transcription factor network regulates NE and ML differentiation in SCLC, and reveals distinct, variant SCLC phenotypes

2.1 Acknowledgement

This chapter is derived from work previously published. Reprinted by permission from the American Association for Cancer Research: Udyavar AR *et. al*, Novel Hybrid Phenotype Revelaed in Small Cell Lung Cacner by a Transcription Factor Network Model That Can Explain Tumor Heterogeneity, *Cancer Research*, 2017, 77(5):1063-1074 (Udyavar et al., 2017).

2.2 Introduction

Small cell lung cancer (SCLC), accounting for 13% of lung cancers (American Cancer Society, 2016), is exceptionally aggressive. Patients with extensive disease die 1 year from diagnosis, and patients with limited disease experience a dismal 20% cure rate (Rudin et al., 2015; Fischer et al., 2007; Hann and Rudin, 2007). Standard of care (Rudin et al., 2015), confined to chemo and radiotherapy for half a century, is largely ineffective as SCLC patients exhibit high initial response rates rapidly followed by treatment-refractory relapse.

Expression-based subtyping, impactful in other cancers (Marusyk et al., 2012), may be effective in SCLC because of phenotypic variability (Rudin et al., 2015; Carney et al., 1985) with respect to neuroendocrine features of its cell of origin (Sutherland et al., 2011; Calbo et al., 2011). A recent study (George et al., 2015) identified two transcriptional SCLC subtypes distinguishable by Notch pathway activity and aggressiveness, but without mutational differences. In genetic mouse models of SCLC, Calbo et al. showed that spontaneously occurring neuroendocrine and non-neuroendocrine cell phenotypes coexist and cooperate to promote metastasis (Calbo et al., 2011).

These reports indicate that a deeper understanding of cellular phenotypes could produce insights into biology and evolution of SCLC. A limitation of previous studies (Calbo et al., 2011; George et al., 2015) is that analyses were based on population averages, whereas variability in tumors should be considered at the single-cell level (Marusyk et al., 2012; Almendro et al., 2014a; Meacham and Morrison, 2013). It also remains unclear why this heterogeneity emerges.

To fill these knowledge gaps, we investigate SCLC phenotypic heterogeneity at the single-cell level using an integrative computational and experimental approach. Consistent with previous reports, we found two transcriptional subtypes at the population level in cell lines and patient specimens, characterized by gene co-expression modules enriched in neuroendocrine/epithelial (NE) and mesenchymal-like (ML) features. To understand how these phenotypes may arise in the absence of driving mutations (George et al., 2015), we hypothesized that they are attractors of a regulatory TF network. This approach is grounded in the mathematical interpretation of Waddington's epigenetic landscape (Waddington, 1957; Huang, 2012), whereby attractors correspond to biological differentiation states or stable phenotypes. Based on this view, it has previously been proposed that malignant phenotypes in cancer correspond to attractors (Kauffman, 1971; Huang et al., 2009), and some have suggested "differentiation therapy" from malignant to benign attractors as a possible treatment strategy (Kauffman, 1971; Huang et al., 2009; Li et al., 2015; Szedlak et al., 2014).

To this end we construct an SCLC master regulatory network of transcription factors (TFs) from NE and ML gene-expression signatures. We then adopt a discrete Boolean modeling approach to simulate the behavior of this TF network and evaluate its ability to dynamically control NE and ML phenotypes. Discrete models are well suited to provide insight into complex TF networks by identifying steady state TF patterns of expression, termed attractors. Mathematically, attractors represent the stable configurations of the dynamic TF network. Biologically, attractors correspond to transcriptional steady states of

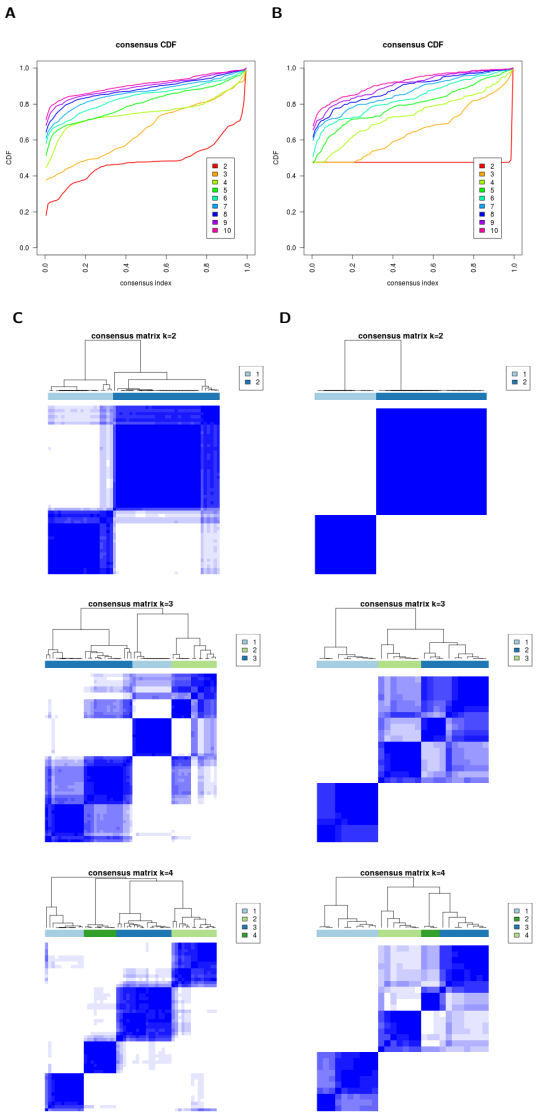


Figure 2.1: Identification of two phenotypes in SCLC cell lines and patients. Consensus clustering was used to identify robust clusters of CCLE cell lines (A and C) and CLCGP tumor specimens (B and D). The consensus score is the frequency that a given pair of samples was placed in the same cluster over 1,000 iterations. The cumulative distribution of SCLC cell line (A) and patient (B) consensus scores is shown for values of k from 2 to 10. These results most strongly support 2 transcriptional subtypes in SCLC cell lines and patients, though in cell lines 4 subtypes may be a good fit. Corresponding consensus score heatmaps for SCLC cell lines (C) and patients (D) are shown for k=2 to k=4. Each point represents a pair of samples, colored by consensus score from white (0, never co-cluster) to blue (1, always co-cluster).

an epigenetic landscape formed by active (ON) and silent (OFF) TFs regulating each other. While discrete models are coarse approximations, they are nevertheless informative and circumvent the obstacle of unfeasible parameter acquisition (Albert et al., 2008; Wynn et al., 2012).

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

2.3 Characterization of SCLC NE and ML phenotypes

We applied consensus clustering (Wilkerson and Hayes, 2010) to the 53 SCLC cell lines in the Cancer Cell Line Encyclopedia (CCLE) database (Barretina et al., 2012), and to 28 tumor samples from the Clinical Lung Cancer Genome Project (CLCGP) (, CLCGP), with the expectation to detect previously observed neuroendocrine and non-neuroendocrine phenotypes (Calbo et al., 2011; George et al., 2015). Consensus clustering analysis, which provides a rationale for determining the number of robustly separated subtypes, indicated that the CCLE cell lines and the CLCGP patient specimens are most consistently separated into 2 distinct clusters (termed Clusters 1 and 2, Figure 2.1).

To characterize these clusters, we applied weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008) to the 53 CCLE cell lines. WGCNA identified 13 gene co-expression modules (Figure 2.2). To summarize the overall expression of each module, eigengenes (first principal component of all genes within a module) were

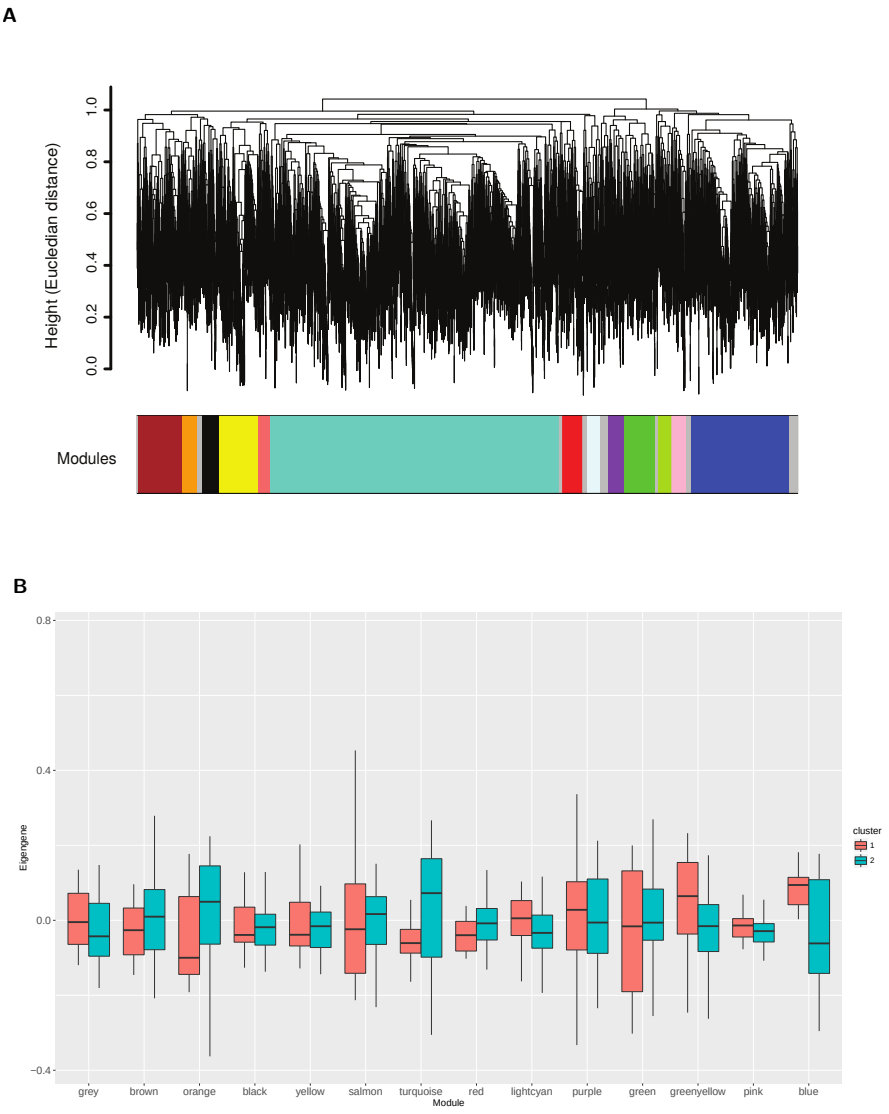


Figure 2.2: WGCNA co-expression modules. (A) Dendrogram generated by WGCNA unsupervised hierarchical clustering of genes from SCLC dataset. Adopting a cutoff height of 0.95 resulted in 14 gene modules. 13 modules contain co-expressed genes and are assigned colors, as is standard for the WGCNA package. The grey module contains all genes which could not be clustered using this cutoff. (B) Boxplots comparing distribution of module eigengenes for the consensus clusters assigned in Figure 2.1. Only the Blue and Turquoise module were statistically significantly different.

computed (Figure 2.3). The Blue module eigengene was upregulated (Bonferroni adjusted p ¡ 0.001) within Cluster 2, while the Turquoise module eigengene was upregulated (Bonferroni adjusted p ¡ 0.05) within Cluster 1 (Figure 2.3 and Figure 2.2B). No other module eigengenes showed statistically significant differences between the consensus clusters (Figure 2.2B).

WGCNA modules are derived agnostically, and reveal co-expressed genes participating in similar biology. Blue and Turquoise module genes therefore provide information about the overall character of the consensus clusters. Visual inspection reveals that hubs (highly connected genes) in the Blue and Turquoise modules are well-known neuroendocrine/epithelial or mesenchymal biomarkers, respectively (Figures 2.4 and 2.5). Statistical analyses by Gene Ontology and EnrichmentMap (Merico et al., 2010) confirm enrichment of the Blue module for neuroendocrine and epithelial differentiation processes, while the Turquoise module is enriched for pathways involved in mesenchymal phenotype and epithelial-to-mesenchymal transition (EMT) (Figure 2.6 and 2.7). Additionally, we performed gene-set enrichment analysis (GSEA) with published signatures of pro-neural, mesenchymal and proliferative glioblastoma subtypes (Carro et al., 2010) on the 12 cell lines with highest Blue module eigengene expression, and 9 cell lines with highest Turquoise module eigengene expression. The pro-neural (Figure 2.7B, lower panel) and mesenchymal (Figure 2.7B, upper panel) signatures were enriched in the cell lines with high Blue module and Turquoise module expression, respectively. As a control, the proliferative signature was significantly enriched in neither (p-value > 0.1, data not shown).

Thus the Blue module genes are a signature of the canonical neuroendocrine state of SCLC, while the Turquoise module genes are a signature of non-neuroendocrine / mesenchymal cells. We therefore refer to Cluster 2 (in which Blue module genes are high) as the SCLC neuroendocrine/epithelial (NE) subtype, and to the Cluster 1 (in which Turquoise module genes are high) as the SCLC mesenchymal-like (ML) subtype (Figure 2.1B, C).

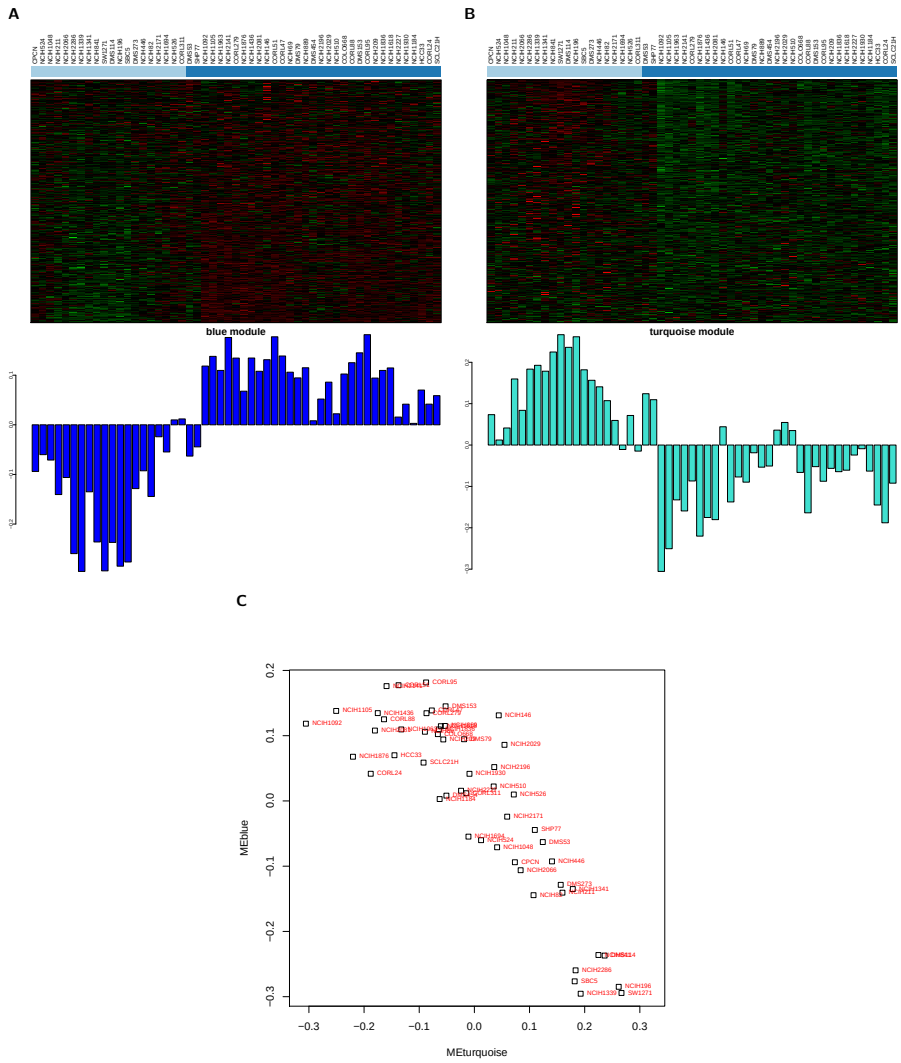


Figure 2.3: Two anti-correlated gene co-expression networks distinguish phenotypic clusters. (A and B) Upper panels: Heatmap view of the Blue (A) and Turquoise (B) module genes (rows) across 53 SCLC cell lines (columns). Cell lines are ordered and marked as in Figure 2.1C consensus clustering. Lower panels: The gene expression profile for each cell line is summarized by the eigengene. The Blue and Turquoise modules had significantly different eigengene expression between the two consensus clusters. (C) The Blue and Turquoise module eigengenes plotted for each cell line reveals anti-correlated expression of Blue vs. Turquoise module (Pearson's correlation: -0.86, p-value: 1.6e-16).

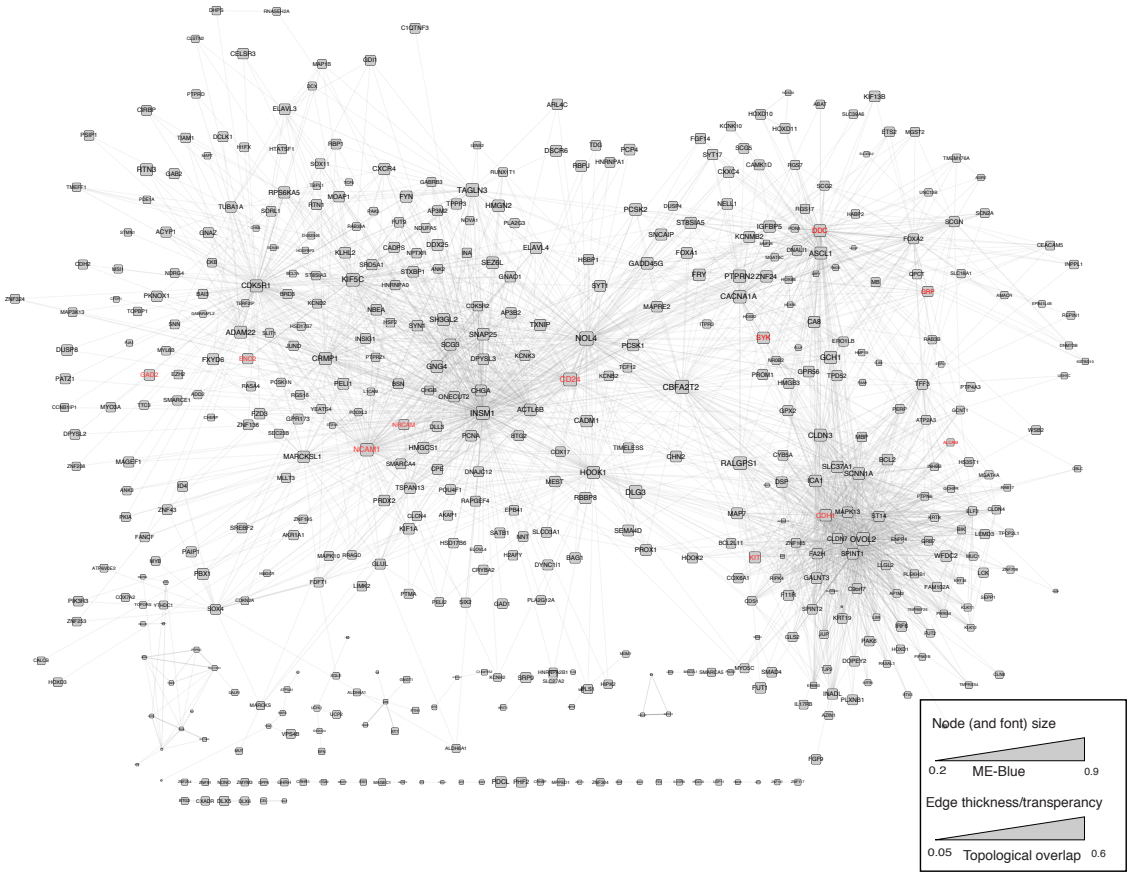


Figure 2.4: Blue module network topology given by WGCNA. This figure shows nodes within the Blue module defined via WGCNA while edges denote a topological overlap measure (TOM). TOM is a metric for the degree of co-expression/correlation between a pair of genes (Langfelder and Horvath, 2008). If the TOM is significant, an edge is drawn between a pair of genes. The thickness of the edges denotes the magnitude of TOM. The size of the node and its font denotes intramodular connectivity ('hubness') (Langfelder and Horvath, 2008) within the Blue module, higher the value - larger the size of the node and its font. The nodes in red denote well-known biomarkers of neuroendocrine and epithelial differentiation.

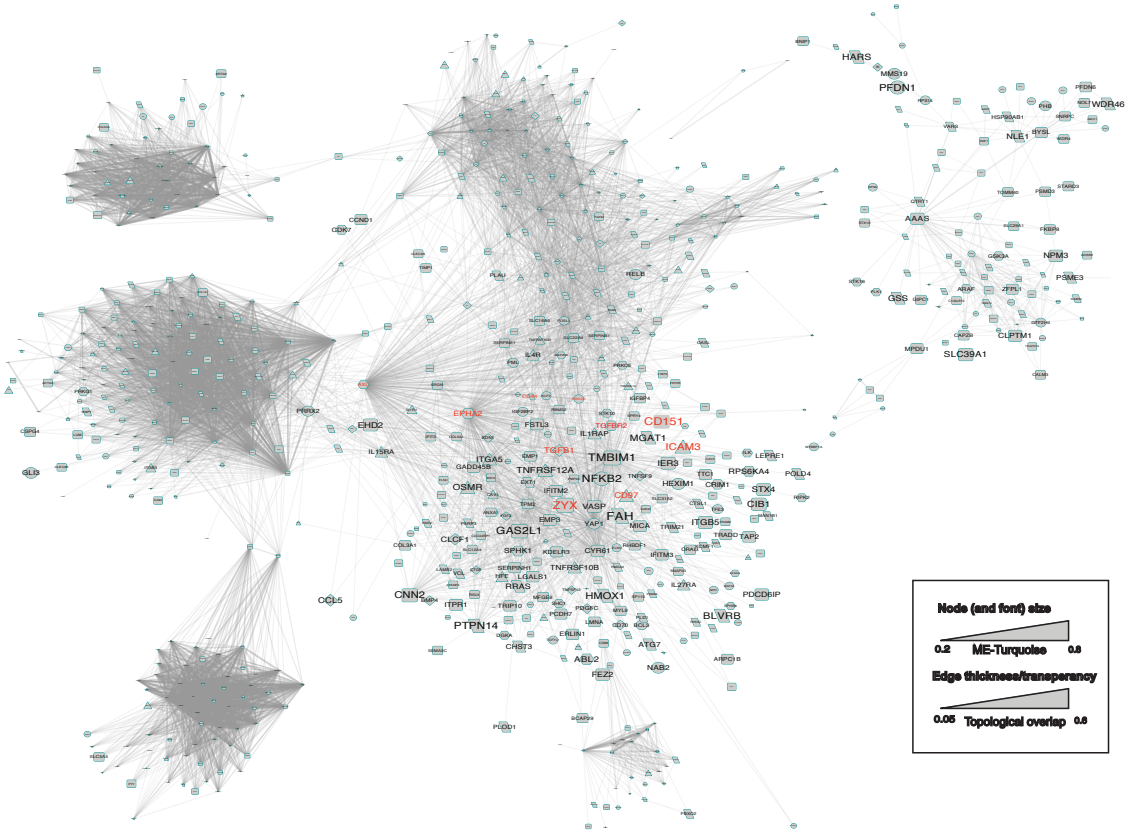


Figure 2.5: Turquoise module network topology given by WGCNA. Same as Figure 2.4, except that the nodes in red denote well-known biomarkers of Epithelial-mesenchymal transition (EMT) or mesenchymal differentiation.

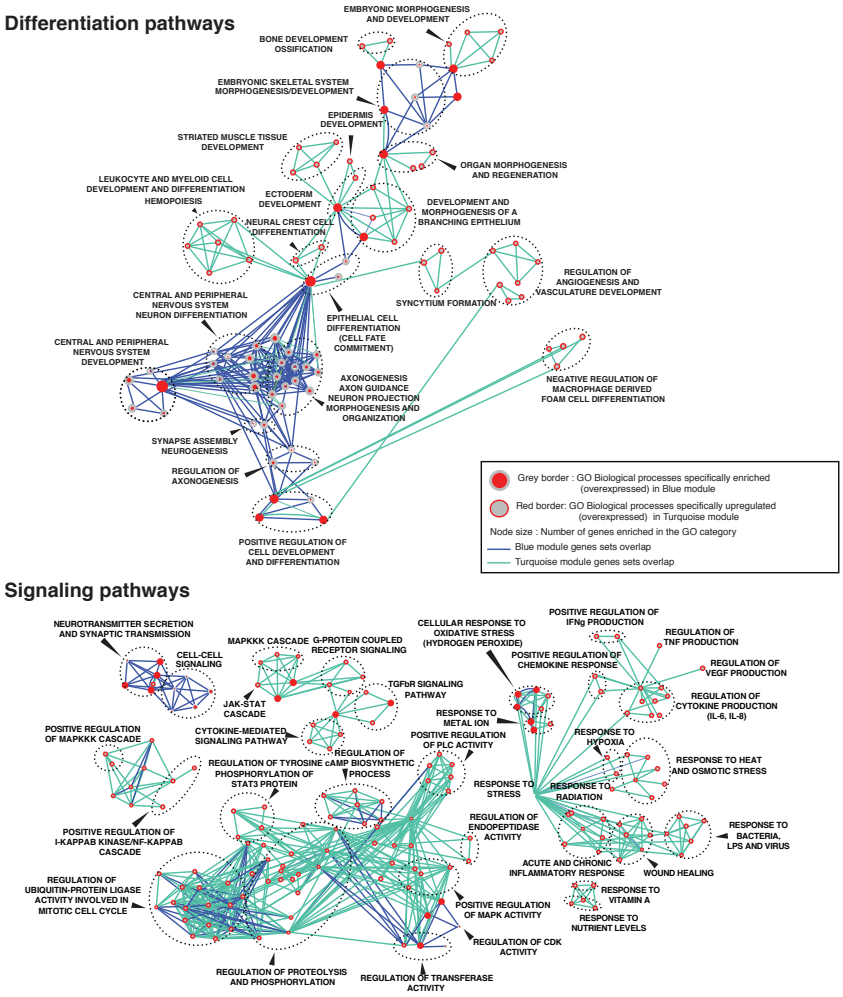


Figure 2.6: Biological processes associated with Blue and Turquoise module genesets. The Blue and Turquoise modules exhibit statistically significant differences (FDR¡0.05) in differentiation (upper part) and signaling (lower part) pathways, by comparative enrichment analysis of Gene Ontology (GO) pathways using BINGO and EnrichmentMap (Merico et al., 2010) in Cytoscape (www.cytoscape.org). These differences are presented as a network, where nodes denote the GO categories and edges denote GO connections between the pathways. Solid red dots are umbrella nodes that connect distinct but related biological processes (manually encircled with dotted lines). The characteristics of the other dots are indicated in the box. 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, MAPK, JAK-STAT, NF-kB, TGFbeta, and cytokine signaling cascades (TNF, VEGF, IL-6, IL-8) that are known to be associated with a mesenchymal/EMT phenotype. Note that statistically significant differences between Blue and Turquoise modules were also found for the following pathways: metabolism, adhesion, transcription, proliferation and apoptosis (data not shown).

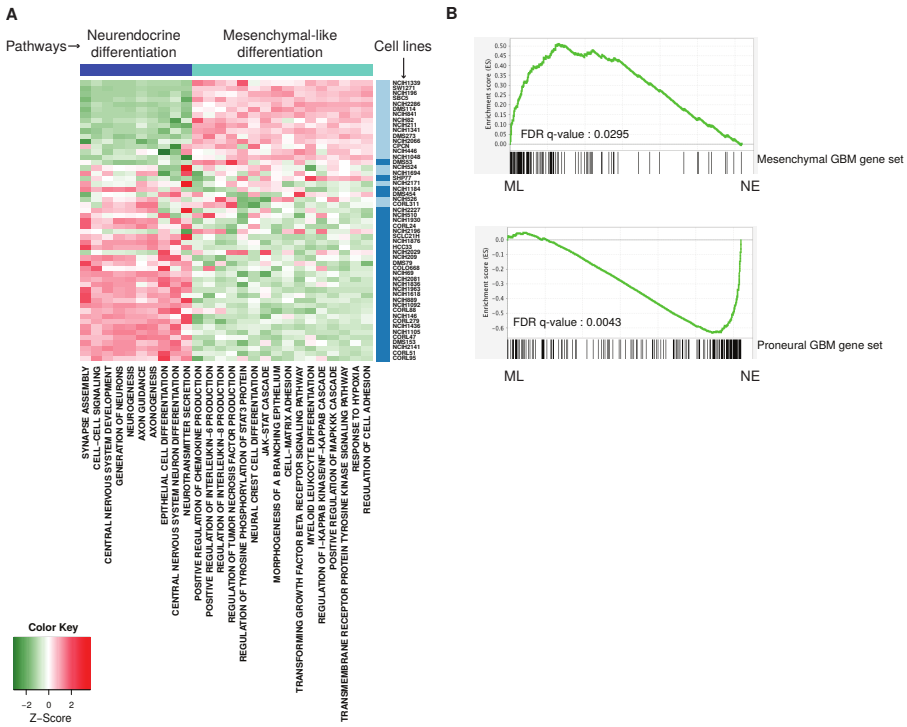


Figure 2.7: Pathway expression of the Blue and Turquoise modules given by comparative pathway enrichment analysis. (A) In columns, expression of genes comprised in differentiation and signaling pathways that distinguish the Blue and Turquoise module (Figure 2.6), averaged within each of the 53 SCLC cell lines (rows, ordered by Blue eigengene as in Figure 2.3). Blue module shows enrichment for neuronal signaling, axon guidance, neurotransmitter secretion and cell-cell signaling. Turquoise module shows enrichment for MAPK, 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) Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) of pro-neural and mesenchymal glioblastoma subtype signatures (Carro et al., 2010) in SCLC cell lines. In this analysis, the 15477 CCLE genes are rank-ordered (left to right) based on correlation between their expression to the top 9 (ML phenotype) and bottom 12 (NE phenotype) SCLC cell lines as ordered in (A). These rank-ordered genes are then assessed for enrichment with gene signatures of pro-neural (bar code in upper panel, as indicated) and mesenchymal (bar code in lower panel, as indicated) glioblastoma subtypes (Carro et al., 2010). Upper panel: the Enrichment Score (ES) for the mesenchymal glioblastoma signature quickly rises in the part of the gene ranking list correlated with the SCLC ML cell lines, and decreases thereafter. Lower panel: the ES for the pro-neural glioblastoma signature rises alongside genes correlated with the SCLC NE cell lines.

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

Since NE and ML subtypes are not associated with driver mutations (George et al., 2015), we hypothesized they may be regulated epigenetically, as during normal cell differentiation. Cell identity in differentiation is largely controlled by regulatory networks of transcription factors (TFs) that coordinate expression of each other and of target genes (Dowen et al., 2014; Gilbert and Barresi, 2017). To understand specification of NE or ML cell identity, we derived a network of TFs which regulate expression of genes within the Blue and Turquoise modules (see Methods; Figure 2.8). By pruning nodes which had no outgoing edges, we reduced this network to a core set of 33 TFs and 361 interactions (Figure 2.9). We used random asynchronous order Boolean simulation with both threshold and inhibitory dominant update rules to identify attractors. Using the threshold update rule, we found 57 stable fixed-point attractors (Figure 2.9A) and no oscillating attractors. With the inhibitory dominant rule we found only 6 fixed point attractors, but 5 two-state oscillating attractors as well. Both the threshold and inhibitory dominant attractors revealed similar features in all subsequent analyses, suggesting that these results are robust to the precise nature of the regulatory interactions. We also inferred network robustness to perturbations using Derrida curves (Derrida and Pomeau, 1986) (Figure 2.10). These results indicate that the basins of attraction are more robust than those of a random network, and that the network dynamics are ordered, rather than chaotic. Together, these observations suggest that the internal structure of interactions imposed to the network (Figure 2.9A) is non-random and naturally leads to well-regulated stationary states, and that our coarse modeling approach is acceptable for this system.

Each attractor is a 33-dimensional vector of TFs, differing by the overall TF ON-OFF expression pattern (columns in Figures 2.9B). Many of the attractors differ only by one or a few TFs. Hierarchical clustering segregated the threshold attractors into four distinct clusters (Figure 2.9B). Results from the inhibitory dominant network are qualitatively sim-

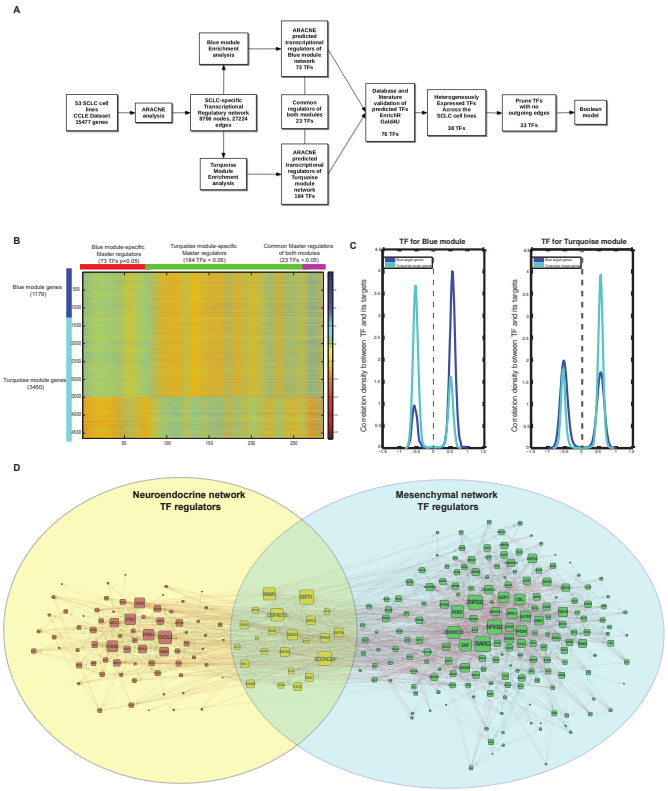


Figure 2.8: Identification of transcription factors that regulate SCLC phenotypic states. (A) Overview of Boolean model network generation analysis. To identify a global SCLC transcriptional regulatory network, ARACNE analysis (based on mutual information between genes) was performed on 53 SCLC cell lines and 15477 genes in the CCLE dataset. The analysis yielded a network of 8706 nodes (genes) and 27224 edges (see Methods). This core SCLC network was analyzed using Fisher's exact test to identify top transcription factors (TF) that act as master regulators of either the NE or ML networks (identified via WGCNA). These TFs were independently validated using literature and transcription factor ChIP-Seq and TF-binding site prediction databases via EnrichR (Chen et al., 2013), leading to a 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.

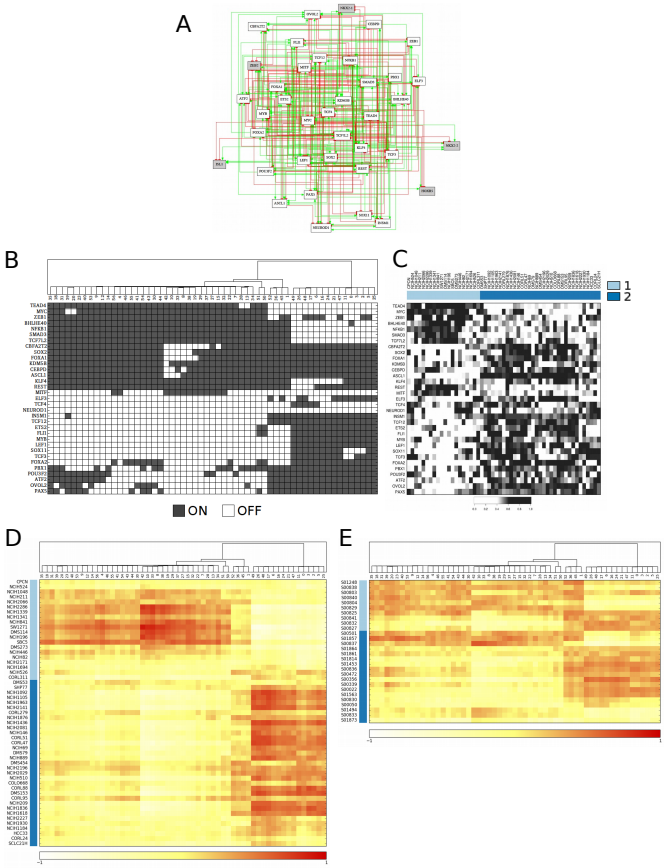


Figure 2.9: Transcription factor network predicts NE and ML attractors. (A) The 33 TF SCLC regulatory network. Green edges indicate activation, while red indicate inhibition. (B) Boolean simulations identified 57 stable attractors (columns). TF expression patterns for all attractors is displayed in rows: shaded cells represent TFs that are ON, white represents OFF. (C) TF expression in the CCLE dataset reveals similar expression patterns to in silico attractors. (D) Correlation score between attractors (columns) and cell lines (rows). Positive correlation is indicated in red, no correlation in yellow, and negative correlation in white. Attractors 49-25 show high correlation with the NE cell lines, while attractors 35-34 show high correlation with the ML cell lines. Nevertheless several cell lines are uncorrelated with NE or ML attractors, possibly revealing a hybrid phenotype. (E) Pearson correlation scores were computed as in (D) between the 57 attractors (columns) and 28 SCLC patient tumor specimens from the CLCGP (rows).

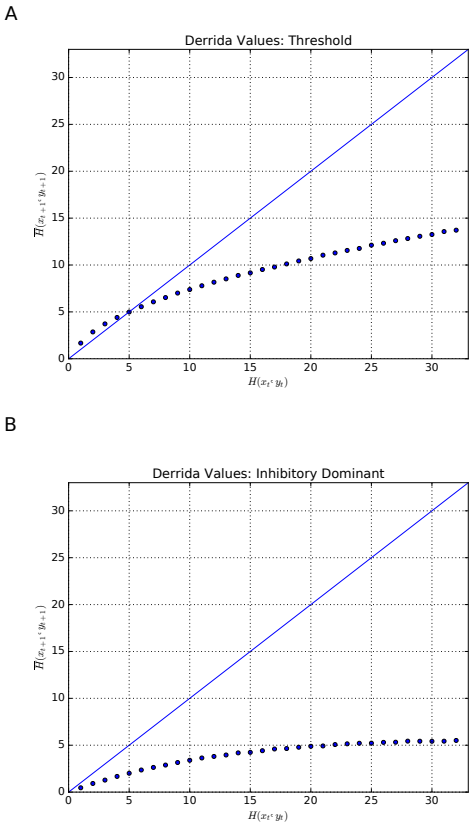


Figure 2.10: Derrida analyses show that network dynamics are ordered. Derrida plots showing the average Hamming distance between two states after a single update of the Boolean network, as a function of the Hamming distance between the two initial states for (A) threshold update, and (B) inhibitory dominant updates. The solid blue line distinguishes whether states get farther apart (above the line) or closer together (below the line) after a single update. Chaotic trajectories tend to get farther apart, and therefore lie above the line, while ordered trajectories converge, and lie below. These results suggest that the SCLC TF network dynamics are within an ordered regime.

ilar, and are shown in Figure 2.11. Based on the TF ON-OFF patterning in each attractor within the clusters, attractors 49 to 25 (along the dendrogram, Figure 2.9B) have active TFs known to be either neuroendocrine (INSM1, POU3F2, SOX2, SOX11) or epithelial (FOXA2, OVOL2). In contrast, attractors 35 to 34 (along the dendrogram, Figure 2.9B) contain active TFs (MYC, NFKB1, SMAD3) involved in mesenchymal differentiation or EMT.

We computed a correlation score between attractors and both cell lines and patients by scaling the CCLE and CLCGP expression data from 0 to 1 (Figures 2.9C) and calculating Pearson's correlation coefficient pairwise between each attractor and sample. Several attractors exhibited high correlation with cell lines within the NE consensus cluster, while others exhibited high correlation with cell lines within the ML consensus cluster (Figure 2.9D). Similar results were observed with patients (Figure 2.9E), and in both cases the model's attractors were found to be significantly more correlated to the samples than random (Figure 2.12).

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

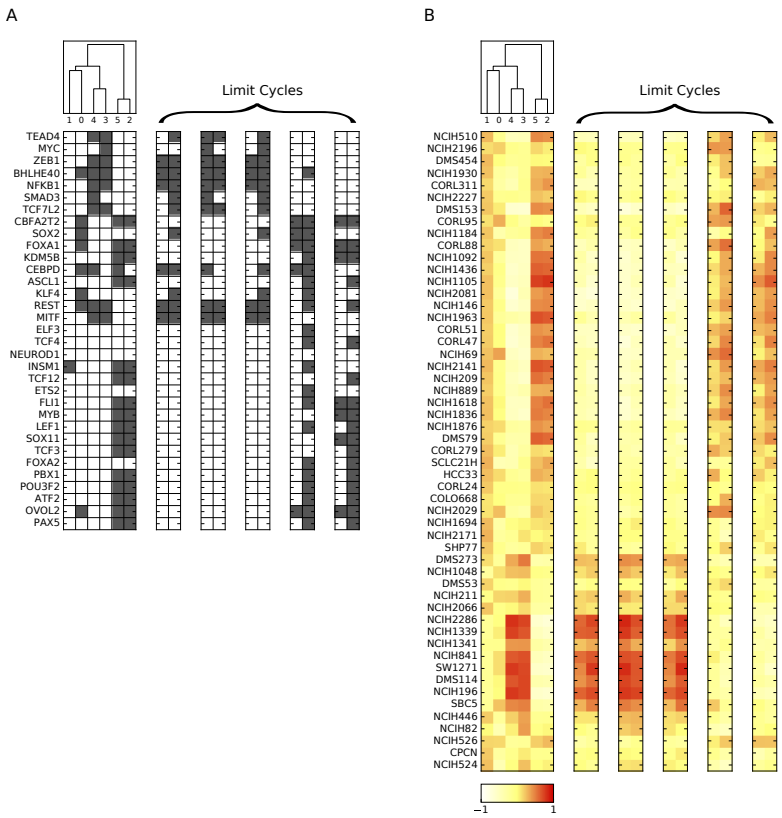


Figure 2.11: Inhibitory dominant dynamics show qualitatively similar attractors to threshold updates. (A) Attractors identified using the inhibitory dominant update method (see Methods). We found 6 steady state attractors, as well as 5 distinct 2-state oscillating attractors, or limit cycles. (B) Correlation score of the attractors with the CCLE cell lines, as in Figure 2.4. The inhibitory dominant update method still identifies several attractors which are correlated to the NE and ML cell lines, however several cell lines still show no significant correlation with any attractor. The limit cycle attractors transition between similar states, suggesting that each limit cycle may still be thought of as only either NE or ML, not transitions between them.

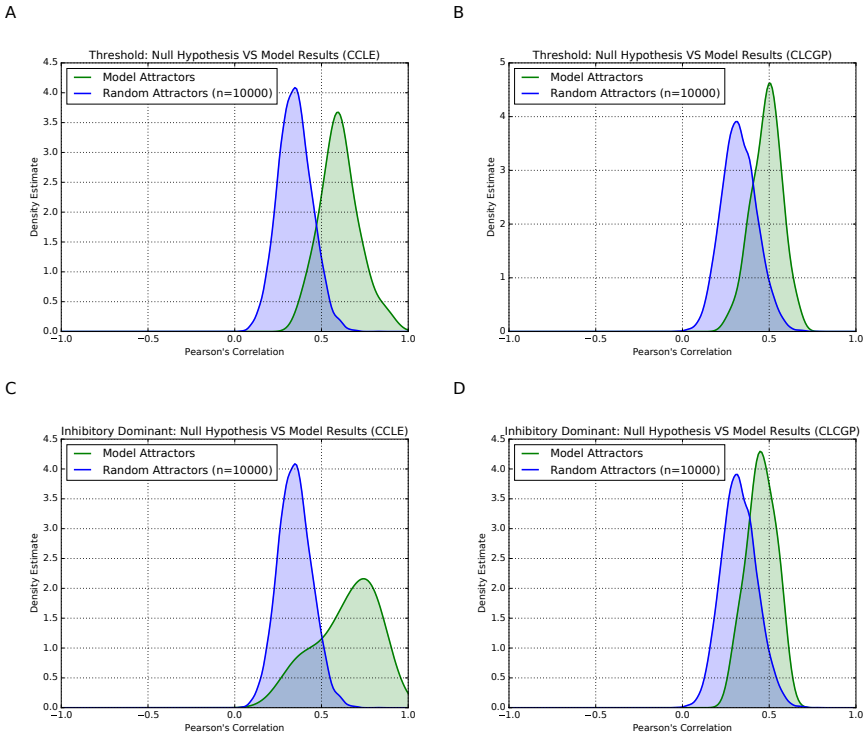


Figure 2.12: Statistical significance of TF network attractors. 10,000 random attractors were generated to derive a correlation distribution for the null hypothesis that the model finds random attractors. (A and C) The distribution of attractor correlations with the CCLE dataset for threshold updates (A) and inhibitory dominant updates (C). For each attractor, only the highest correlation was considered to avoid saturating the distribution with poor correlations (i.e. if an attractor correlates with a NE cell line, it will almost certainly be anti-correlated or poorly correlated with the ML cell lines). The blue distribution shows the best correlation of random attractors, while the green distribution shows the best correlation of the model's attractors. By the Mann-Whitney U test, the model's correlations are significantly higher than random (threshold: p-value = 9.5e-34, inhibitory-dominant: p-value = 7.6e-9). (B and D) The distribution of attractor correlations with the CLCGP dataset for threshold updates (B) and inhibitory dominant updates (D). As in (A) only the highest correlations are considered. By the Mann-Whitney U test, the model's correlations are significantly higher than random (threshold: p-value = 1.7e-23, inhibitory-dominant: p-value = 1.7e-6).

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

We probed the expression of cell lines for 10 of the 33 TFs. In general, NE TFs were expressed at a higher level in NE than ML cell lines, and vice versa (Figure 2.13A). Consistent with the finding that some cell lines did not correlate with either the NE or ML attractors (Figures 2.9D), we observed cell lines which simultaneously expressed both NE and ML TFs. Similarly, CD56 (NE marker) and CD44 (ML marker) were found co-expressed in 3 out of 10 patient samples (pt1112-1, pt216-1, and pt460-1, Figure 2.13B). The other tumor samples had mutually exclusive expression of either CD56 or CD44.

These double-positive cell lines may be either composed of mixed populations of NE and ML cells, or hybrid single-cells simultaneously co-expressing features of both phenotypes. We investigated this using single-cell flow cytometry with well-established neuroendocrine/epithelial and mesenchymal differentiation biomarkers (Figure 2.14 legend). To aid visualization, we defined overall NE and ML scores as the unweighted sum of NE and ML biomarkers, respectively.

In biaxial plots of these scores, NE cell lines consist primarily of NE+ML- single cells, while ML cell line single cells were NE-ML+ (Figure 2.14A, C). However, in several cell lines, single cells simultaneously expressed similar levels of both NE and ML biomarkers (Figure 2.14B). These results confirm the existence of a novel SCLC NE+ML+ hybrid phenotype comprised of both neuroendocrine/epithelial and mesenchymal-like characteristics.

Other reports of hybrid phenotypes in cancer have associated the hybrid cells with a more stem-like phenotype (Jolly et al., 2015; Xue et al., 2015). We measured single-cell expression of CD133, a cancer stem-cell marker, and found no significant difference between the NE, ML, and hybrid cell lines (Figure 2.15).

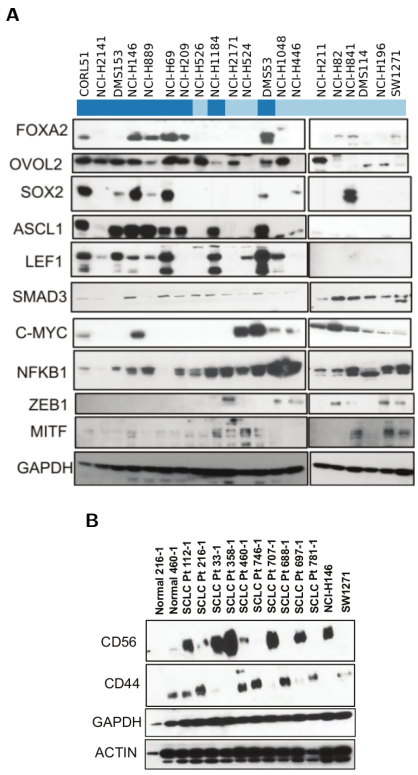


Figure 2.13: Experimental validation of TF network states in human SCLC. (A) 5 NE TFs (FOXA2, OVOL2, SOX2, ASCL1, and LEF1) and 5 ML TFs (SMAD3, MYC, NFKB1, ZEB1, MITF) validated by western blot. The NE TFs show higher expression in the NE cell lines (dark blue, as in consensus clustering) than the ML cell lines (light blue), and vice versa. Nevertheless, several cell lines showed similar levels of expression of both NE and ML TFs, including NCI-H146, NCI-H209, NCI-H1184, DMS53, NCI-H1048, NCI-H446. Many of these cell lines, such as DMS53, also showed poor correlation with any attractor in Figures 2.9 and 2.11, and were near the center in Figure 2.3C, suggesting a non-NE and non-ML phenotype. These cell lines are tentatively denoted as "hybrid" phenotypes. (B) CD56 (a NE marker) and CD44 (a ML marker) were probed by western blots in 10 SCLC patient samples. Most patients show expression of only one or the other, however Pt112-1, Pt216-1, and Pt460-1 show double positive expression of both markers, suggesting the hybrid phenotype may be important in patients as well.

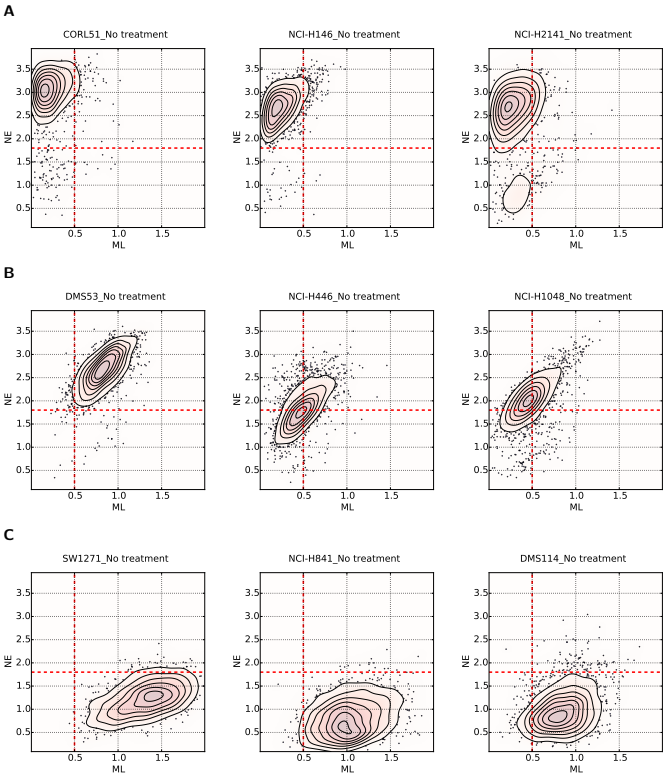


Figure 2.14: Single-cell level expression of phenotypic biomarkers in SCLC cell lines reveals hybrid cells. Flow cytometry was performed using 4 NE surface markers (CD56, CD24, CADM1, and ALCAM) and 2 ML surface markers (EPHA2 and CD151). Axes represent unweighted sums of normalized NE and ML fluorescence. (A) CORL51, NCI-H146, and NCI-H2141 were classified as NE by consensus clustering, and single cells show NE+ML- phenotype. (B) DMS53, NCI-H446, and NCI-H1048 showed comparable levels of expression of NE and ML TFs (Figure 2.13), Blue- and Turquoise-module eigengenes (Figure 2.3), and poor correlation with either NE or ML attractors (Figures 2.4, 2.11). At the single cell level, these cells also show similar levels of expression of NE and ML markers, suggesting these cells are in a hybrid NE+ML+ or NE-ML- phenotype, not well described by a NE vs ML dichotomy. The hybrid state is characterized by cells along the diagonal of these plots. (C) SW1271, NCI-H841, and DMS114 were consistently classified as ML, and single-cells exhibit NE-ML+ phenotype.

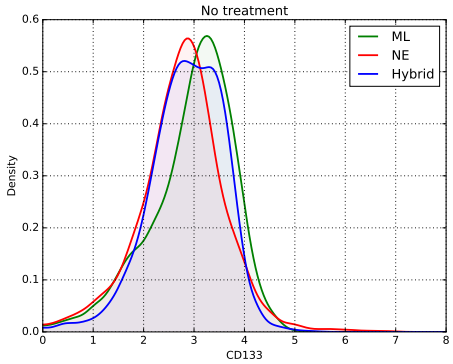


Figure 2.15: Hybrid cells are not enriched for stem cell marker. Distribution of fluorescence intensities of CD133, a cancer stem-cell marker, for NE, ML, and hybrid cell lines. Density estimates show that there is no significant difference in the expression levels of this marker between the three subsets.

2.6 Modulation of SCLC phenotypes with chemotherapy or epigenetic drugs

Next, SCLC cell lines were treated with etoposide and cisplatin, and with epigenetic modulators valproic acid (HDAC inhibitor) and 5-azacytidine (DNA methylation inhibitor). Flow cytometry measurement of NE and ML markers (see Figure 2.14 legend) was used to characterize how treatment shifted the phenotypic identity of SCLC cells. Biaxial plots of NE vs ML scores showed phenotypic shifts at the single-cell level for all perturbations (Figure 2.16), converging toward hybrid populations. The hybrid phenotype was thus an end-state for SCLC cells subjected to stress.

2.7 Discussion

In this work, we identified signatures of phenotypic heterogeneity in SCLC, and a set of TFs regulating expression of these genes. Through discrete Boolean model simulations, we showed that a master SCLC TF network naturally settles into states which were identified as NE or ML. This suggests that the NE and ML cell identities can naturally emerge from regulatory dynamics of TF networks, rather than being driven by genomic mutations. Most significantly, by distilling the NE and ML phenotypes to their essential TF drivers, the

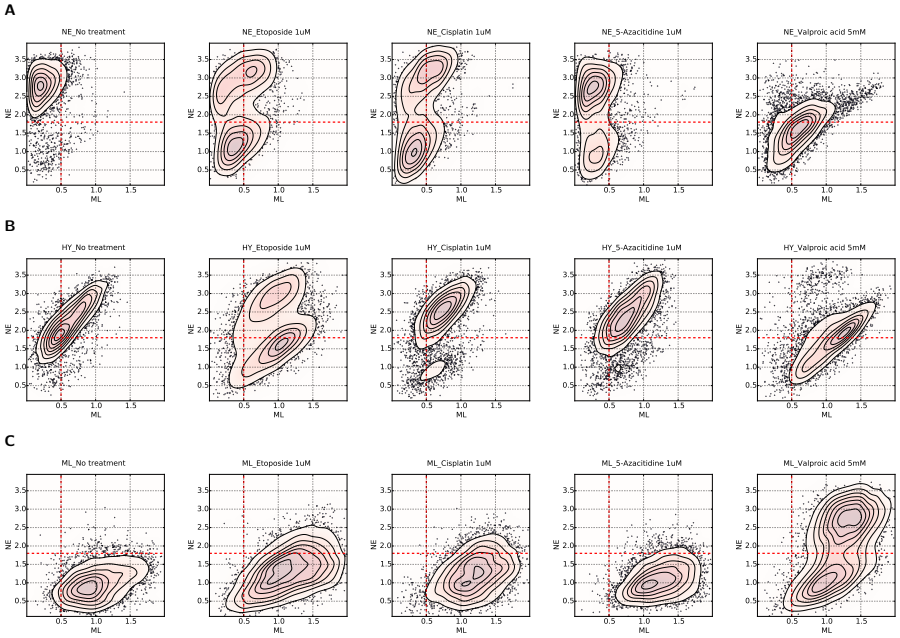


Figure 2.16: Phenotypic modulation by treatment pushes cells toward hybrid state. NE, ML, and Hybrid (HY) cell lines (from Figure 2.14) are combined to show overall phenotypic responses. NE and ML axes are the same as defined in Figure 2.14. Each column represents a unique treatment condition: Untreated (column 1); 1uM Etoposide (column 2); 1uM Cisplatin (column 3); 1uM Azacitidine (column 4); 5mM Valproic Acid (column 5). Cells were treated for 2 days. (A) The NE cell lines show a shift toward the NE+ML+ or NE-ML- quadrants. (B) The HY cell lines show no significant shift. (C) The ML cell lines show a shift toward the NE+ML+ quadrant.

model also exposed some cell lines which had been improperly classified as NE or ML through transcriptome based clustering. In vitro validation confirmed existence of SCLC cell lines and patient samples that simultaneously express both NE and ML markers and TFs. We verified this hybrid phenotype at the single-cell level in cell lines, and showed that cell lines transition toward this hybrid state for survival upon drug treatment.

As resolving cancer heterogeneity can have profound impact on patient care and outcomes (Fisher et al., 2013), a deeper understanding of SCLC heterogeneity, i.e., NE, ML, hybrid and beyond, may translate to benefits in the clinic. Historically, attempts to separate SCLC into subtypes in the clinic were abandoned due to poor reproducibility among pathologists and unclear clinical relevance (Travis, 2012). However, our findings indicate that SCLC heterogeneity is dynamic, since a core TF network specifies both the NE and ML phenotypes, which may confound static associations. The hybrid phenotype may be an additional confounder, exacerbated by the tendency of both NE and ML cells to move toward the hybrid phenotype under treatment. We argue that SCLC heterogeneity is best interpreted as states within a phenotypic landscape, and understanding phenotypic mobility within this landscape could provide ties to clinical relevance.

While SCLC has a high mutation rate, no correlations between mutations and distinct subtypes have been observed (George et al., 2015). In this work we show a non-mutational mechanism for regulation of distinct SCLC cell identities, as attractors of a TF regulatory network. This framework of equating cell types to attractors was first advanced by Waddington in his eponymous epigenetic landscape (Waddington, 1957), and has since been expounded upon by many mathematical biologists (Huang, 2012). It has further been argued that malignant cancer states and cancer heterogeneity are best described by the concept of attractors (Kauffman, 1971; Huang et al., 2009; Li et al., 2015). Significantly, this framework has been shown to provide possible therapeutic strategies by identifying targets *in silico* which most significantly perturb the attractor states (Szedlak et al., 2014). We do not exclude the possibility that mutations can drive SCLC heterogeneous phenotypes, but

our results indicate that epigenetic causes should also be considered.

The Boolean modeling approach used here is an established way of obtaining a coarse-grained picture of large network behavior, and is well suited to identifying the stable states of TF networks (Albert et al., 2008; Wynn et al., 2012). Our approach was novel in constructing the TF network in blind fashion from SCLC gene expression datasets, TF databases and literature, and was validated by its ability to reproduce correlations with known phenotypes. Significantly our model revealed the existence of a previously unrecognized hybrid phenotype, by showing that some cells do not correlate to either the NE or ML attractors. However, we did not identify attractors correlated with the hybrid phenotype. Hybrid EMT phenotypes have been previously reported in NSCLC (Lecharpentier et al., 2011) and lung adenocarcinoma (Schliekelman et al., 2015), and other groups have recently reported computational modeling of hybrid EMT phenotypes by driving EMT networks with external stimuli (Jolly et al., 2015; Steinway et al., 2015). Additionally, the Boolean modeling approach cannot capture intermediate levels of expression, and therefore attractors corresponding to the hybrid state may not be identifiable using this method. Continuum modeling approaches may be needed to better understand the hybrid state. Identification of gene co-expression modules enriched in the hybrid phenotype may also reveal additional relevant TFs. Considering all of these possibilities, current ongoing work in our laboratory is directed at identifying hybrid attractors.

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

2.8 Methods

2.8.1 Data normalization

CCLE dataset (Barretina et al., 2012) was downloaded from Broad Institute as CEL files. Data were normalized and median centered using quantile RMA normalization using Affy Bioconductor package (Gautier et al., 2004) in R. Probe-level data for all the datasets was converted to gene-level data by probe merging using collapseRows (Miller et al., 2011). Probes with no known gene symbols were removed.

2.8.2 Consensus clustering

Consensus clustering was performed using ConsensusClusterPlus v1.24.0 package in R v3.2.3 (Wilkerson and Hayes, 2010) on both the 53 SCLC cell line dataset from CCLE and 28 SCLC patients from CLCGP, with 80% sub-sampling of both genes and samples, 1000 repetitions, 1 - Pearson correlation, and k-means. Both the CCLE and GLCGP datasets were subsetted to only include genes measured in the GSE6044 (Rohrbeck et al., 2008) dataset, in order to maximize overlap with our previous work (Udyavar et al., 2013).

2.8.3 WGCNA and Network analysis

Co-expression network analysis was performed in R v3.2.3 using the WGCNA package v1.49 (Langfelder and Horvath, 2008). As with the consensus clustering analysis, the CCLE SCLC dataset was subsetted to only include genes measured in GSE6044 (Rohrbeck et al., 2008). We used 1 - Pearson's correlation to build a co-expression based dissimilarity matrix. Modules were generated using unsupervised average-linked hierarchical clustering with a static height of 0.95. We required that each module contain at least 100 genes.

Comparative Pathway enrichment analysis of the Blue and Turquoise modules was performed using BINGO and Enrichment map (Merico et al., 2010) and visualized Cytoscape. Expression values were transformed to Z-scores, and 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.

2.8.4 Transcriptional regulatory network construction

To generate an SCLC-specific transcriptional network, we applied the bootstrap version of ARACNE (Margolin et al., 2006) on gene expression profiles from the 53 CCLE SCLC cell lines using the following parameters: p-value = 10-7, dpi = 0 and 100 bootstraps, resulting in 27,224 interactions among 8,706 nodes. To evaluate if genes in Blue and Turquoise module are enriched for targets of any specific TF, we used Fisher's exact test (FET) (Carro et al., 2010) to compute TF enrichment with the Blue and Turquoise module genes. 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 2.8).

Modeling gene regulatory networks which were inferred entirely from the available data can suffer from circular reasoning: a dataset generates a network, which then predicts the dataset. To avoid this fallacy, we built the network topology strictly using sources external to the datasets of interest. Thus, we first filtered ARACNE TF predictions based on gold-standard TF-target binding site databases CHEA, ENCODE, TRANSFAC, JASPAR using EnrichR (Lachmann et al., 2010; Consortium, 2004; Matys et al., 2003; Mathelier et al., 2014; Chen et al., 2013), and literature databases such as Pubmed and Glad4U (Jourquin et al., 2012)A. These filtration steps produced a list of 76 likely TF regulators of NE and/or ML differentiation. We took only heterogeneously expressed TFs (median absolute deviation above the 50th percentile) yielding a list of 38 TFs that we used to build a Boolean network for SCLC. Next we extracted directed interactions between these TFs using only information from the above gold-standard references and literature. Where possible, interactions were classified as activating or inhibiting by manually searching the literature. Interactions which we could not find in literature were classified as activating if the TFs are

positively correlated across the CCLE dataset, or inhibitory given negative correlation. 5 TFs did not have outgoing edges and were eliminated, leaving 33 TFs.

2.8.5 Boolean network simulation and analysis

The TF network was simulated as a Boolean network where each node was either ON (active) or OFF (silent). Nodes were updated using the random order asynchronous method (Albert et al., 2008). We considered two distinct approaches updating TFs: 1) threshold updating, in which the total number of ON inputs are compared such that the target node is switched ON when there are more activators, and OFF when there are more inhibitors, and 2) inhibitory dominant, in which having any inhibitor ON is sufficient to turn the target node OFF. The threshold update rule is referred to as a Hopfield neural network in some literature. Because the network's state space is so large, we only simulated a random subsample of the states. For both update rules, a state transition network was seeded with 8,000 randomly generated initial states, in which each TF had a 50% chance of being ON or OFF. When TFs are updated in asynchronous random order, it is possible that the state at time t may have several possible outgoing trajectories to time t+1. To account for this, each state that we observed was initialized and updated 30 times to sample distinct trajectories to the next state that may be influenced by the update order. Each newly observed state was queued to be updated in this fashion, until there were no new states identified. Attractors were identified by applying the attracting components algorithm from NetworkX to the state transition graph for each update rule. Using threshold updates, we found 57 fixed point attractors, and no oscillating ones. We observed the same set of attractors using synchronous updates with 217 and 218 initial states. Using the inhibitory dominant update rule, we found 6 fixed point attractors, and 5 two-state oscillating attractors, finding the same set of attractors using synchronous updates from 212 to 218 initial states.

To score the correlation between samples and attractors, CCLE and CLCGP expression data were independently scaled between 0 and 1, and Pearson's r was calculated pairwise

between attractors and cell lines/patients. Statistical significance was determined by considering the highest correlation between each attractor and cell line, and comparing against a null distribution, obtained by generating 10,000 random TF vectors. The Mann-Whitney U test was used to compare the score distributions. The model attractors had higher correlation with cell lines than the random attractors (threshold: p-value = 9.5e-34, inhibitory-dominant: p-value = 7.6e-9), and higher correlation with patients than the random attractors with (threshold: p-value = 1.7e-23, inhibitory-dominant: p-value = 1.7e-6). Individual attractors were assigned a p-value for their highest correlation with each sample by direct comparison with the null distribution.

2.8.6 Robustness of NE and ML attractor states

Cells need to be able to robustly guide their differentiation choices depending on driving signals, and therefore we would expect trajectories toward cell attractors to be robust. Structural coherence (Willadsen and Wiles, 2007) is a topological metric which measures how reliably an initial condition evolves toward its appropriate attractor given a perturbation. This metric requires an estimate of the total size of the basin of attraction, so we were only able to apply it to attractors for which we could reliably estimate this size (a few basins were too small to reliably estimate their size). Values of random, maximum, observed, and structural coherence are reported in Table S4. We also calculated Derrida curves (Derrida and Pomeau, 1986) (Figure 2.10), as the average increase growth or decay of an initial perturbation after one step, for perturbation sizes ranging from 1 to 32 TFs.

2.8.7 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).

2.8.8 Cell culture

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

2.8.9 Flow cytometry data generation and analysis

1-2 million cells were plated in T75 or T150 flasks the previous day and collected for flow experiment the next day as described below. For drug treatment experiments, cells were plated same as above the previous day, followed by drug addition the next day. Cells were incubated with drugs at 37 degrees 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 or permeabilization with ice-cold 100% methanol at -20 for 30 minutes. Cells were then stained with fluorescent conjugated antibodies for 30 minutes in dark at room temperature. Samples were washed with PBS and run on BD 5-laser instrument at the Vanderbilt Flow cytometry core. Fluorescent channels were compensated using anti-mouse IgK beads (BD Biosciences) that were tagged with fluorescent antibodies.

First intact cells were gated on the Forward Scatter Area (FSC-A) and Side-Scatter Area (SSC-A) plot, where debris has a low FSC-SSC ratio. This was followed by gating for Alexa700 negative viable cells as described previously (Box 1 in (Irish and Doxie, 2014)), where A700 positive populations are dead/dying cells. Finally, A700 negative viable cells are further gated to include only singlet populations. 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. The raw cytometer intensity readouts for fluorescent channels were first converted to log scale by using the asinh() function with a co-factor of 150. Gating was conducted in Cytobank (Kotecha et al., 2010).

2.8.10 Western blotting

Cell lines were plated for 2 days in complete medium to equilibrate. Lysates were prepared by spinning cells at 4C, aspirating the media, and adding M-PER lysis buffer (Pierce) containing 1X phosphatase inhibitors 2 and 3 and protease inhibitor (Sigma-Aldrich).

Lysates were incubated for five minutes at room temperature, vortexed for 30 seconds and centrifuged at 15000 rpm for 15 minutes (at 4C). The protein concentration was quantified using BCA assay (Pierce). Lysates were boiled for 10 minutes at 100 degrees with 1X NuPage sample buffer (Molecular Probes) and run on 8% or 4-12% Tris-glycine gels (Molecular Probes). Semi-dry transfer was followed by blocking with 1X Casein-TBS. Blots were imaged by chemiluminescence.

Chapter 3

Variant-NE SCLC phenotypes show distinct gene expression, drug sensitivity, and transcriptional regulatory programs

3.1 Gene co-expression modules differentiate neuroendocrine, non-neuroendocrine, and distinct variant phenotypes

Chapter 2 reported variant SCLC phenotypes (termed "hybrid") that are distinct from the canonical NE and non-NE phenotypes. This chapter provides a more comprehensive characterization of the transcriptional and phenotypic nature of these 2 additional variant SCLC subtypes. Additionally, analyses in this chapter are based on an updated RNA-seq expression from the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012), including a total of 50 SCLC cell lines.

To identify robustly distinguishable phenotypes in SCLC, we applied Consensus Clustering (Wilkerson and Hayes, 2010) to this new RNA-seq dataset. The consensus clusters suggest that the SCLC cell lines are best described as either two or four distinct phenotypes (Figure 3.1. Because of our previous results, and growing literature on SCLC heterogeneity supporting more than two distinct subtypes, here we further investigate the four-cluster

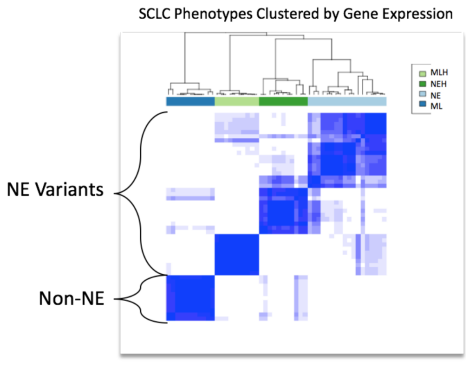


Figure 3.1: Consensus clustering robustly separates SCLC cell lines by into four groups, including a non-NE cluster, and 3 distinct NE clusters.

result.

We compared expression of 19 well-studied markers previously associated with heterogeneity in SCLC (Jahchan et al., 2016; Lim et al., 2017; Calbo et al., 2011; Borromeo et al., 2016; Mollaoglu et al., 2017; Williamson et al., 2016; Huang et al., 2018) (Figure 3.2). Three of the four consensus clusters overlap well with previously reported phenotypes: the canonical neuroendocrine (NE) ASCL1$^+$ NEUROD1$^-$ SOX2$^+$ (Borromeo et al., 2016) phenotype, an NE-variant NEUROD1$^+$ MYC$^+$ ASCL1$^-$ phenotype (Borromeo et al., 2016; Mollaoglu et al., 2017), and a non-NE variant REST$^+$ NOTCH$^+$ HES1$^+$ phenotype (Lim et al., 2017). The final phenotype, which we here refer to as NE Variant-2, expressed canonical NE markers such as ASCL1, DDC, and GRP, but also HES1, which has been reported as an indicator of Notch pathway activity in SCLC (Lim et al., 2017). Interestingly, in the NE Variant-2 phenotype, HES1 is expressed despite comparatively low NOTCH expression (Figure 3.2).

To identify gene programs driving the distinction between the four SCLC phenotypic clusters, we performed weighted gene co-expression network analysis (WGCNA) on the RNAseq data. Using a scaling factor of 12, WGCNA identified 17 co-expressed gene modules. We computed module eigengenes, which summarize the overall expression of all genes in a given module, for each cell line and found 11 out of 17 modules were able to statistically distinguish between the four cell line clusters (Figure 3.3, Kruskal-Wallis fdr-adjusted p-value < 0.05).

3.2 Expression of four cell-surface markers is able to robustly distinguish SCLC phenotypes.

The identification of these four SCLC phenotypes prompted us to search for biomarkers that could be used as a proxy for presence of phenotypes in heterogeneous tumors. Our computational approach uses linear discriminant analysis (LDA) to identify a set of genes that are able to distinguish between the four phenotypes. Our search was limited to cell

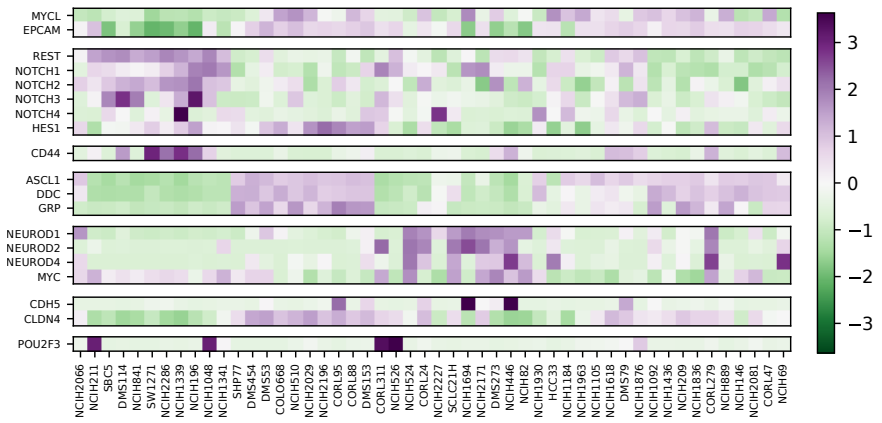


Figure 3.2: Consensus phenotypes are well represented by literature reports. Several recent reports have identified markers and characteristics of heterogeneous SCLC subpopulations. Here, we compare these markers with the consensus clusters identified in Figure 3.1. The top group includes markers for the "tumor propagating cell" (TPC) subtype characterized in mice by Jahchan *et. al.* (Jahchan et al., 2016), which can distinguish the NE and NE Variant-2 clusters from the non-NE and NE Variant-1. The second row contains markers of the HES1+, NOTCH+ phenotype reported in mice by Lim *et. al.* (Lim et al., 2017), which are enriched in the non-NE cluster. The third row shows expression of an alternative non-NE marker, CD44 which was reported in mice and humans by Calbo *et. al.* (Calbo et al., 2011). Interestingly, CD44 is enriched specifically within a subset of the non-NE cluster, suggesting the possibility of additional heterogeneity. The fourth row contains markers of the canonical NE SCLC phenotype (Borromeo et al., 2016), which are highly expressed in the NE and NE Variant-2 subtypes. The fifth row shows markers of an alternative NEUROD+ ASCL1- phenotype which has been reported in humans (Borromeo et al., 2016) and mice (Mollaoglu et al., 2017). The sixth row shows markers of a rare population found in circulating tumor cells that show vasculogenic mimicry (Williamson et al., 2016). Here, CDH5 is poorly distinguishing across our subtypes, though CLDN4A is specific to the NE and NE Variant-2 types. The last row shows expression of POU2F3, which has recently been proposed as a master regulator in SCLC cells with a potentially distinct cell of origin (Huang et al., 2018). However, this gene has no distinguishing power across the subtypes defined here.

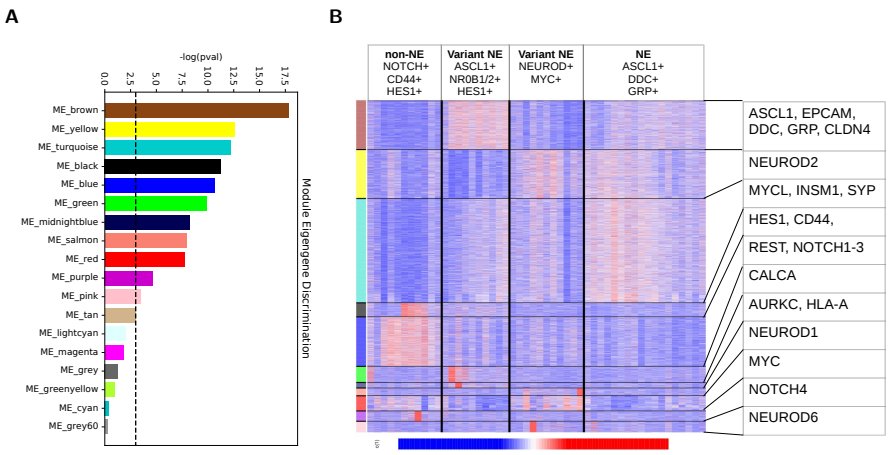


Figure 3.3: WGCNA analysis of four phenotypes. (A) WGCNA analysis identifies 11 statistically distinguishing modules of co-expressed genes across this panel of 50 cell lines. (N) Expression of all genes within each of the 11 distinct modules reveals a mosaic pattern of expression distinguishing SCLC subtypes, such that most modules are upregulated in more than one phenotype. A few key markers are highlighted for each module and phenotype.

surface markers, based on the Jensen Compartments database (Binder et al., 2014), so that live cells could be used for future analyses. Because an exhaustive search of marker sets would be computationally expensive, we prioritized markers based on their ability to distinguish between at least two phenotypes (pairwise fdr-corrected Mann-Whitney U test between each phenotype). We then selected four of the markers at a time to make a candidate marker set, which were used as "features" in the gene expression data (RNA-seq data from the CCLE) on 50 cell lines. LDA attempts to find the best model for separating two or more defined populations of data; in our analysis, we sought the LDA model that best separates the four phenotypes based on the candidate set of features in phenotypic space, where each feature represents a different dimension of the space.

Using a random 80% of the data, we performed cross-validation on our model by training a model on each possible combination from the candidate list, and scoring the LDA's predictive accuracy on the remaining 20% of the data, and repeating this process 100 times. The accuracy score of each model was then computed as the average accuracy score over

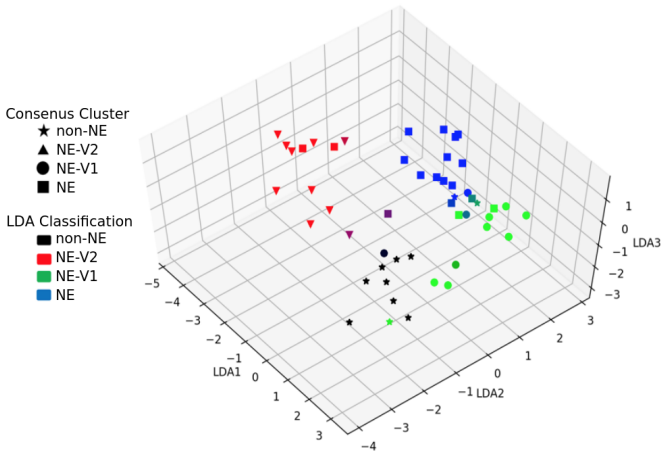


Figure 3.4: Linear discriminant analysis projection of RET, TYRO3, EPHA2, and CEACAM5. Each dot represents one cell line. Over 1000 iterations, 80% of cell lines were used to train the LDA model, while 20% were used for validation. The shape of each point reflects its phenotype based on the consensus clustering results, while the color indicates how often each cell line was assigned to each type by the LDA model over the 1000 iterations.

100 iterations. Because of the inherent randomness associated with selecting 80% of the data, we repeated this process 10 times to determine the best marker sets for discriminating between phenotypes. Based on this cross-validation, as well as availability of high quality antibodies, we identified the marker set RET, TYRO3, EPHA2, and CEACAM5 (Figure 3.4) as the best marker set. RET and CEACAM5 are both members of the brown module, and TYRO3 and EPHA2 are both members of the blue module. Figure 3.4 shows clear separation between the four phenotypes based on expression of these four markers. Additional validation at the single-cell level with flow cytometry is currently ongoing work in the lab, and alternative marker sets may be used if protein expression of these markers does not match our prediction based on RNA.

3.3 SCLC phenotypes are differentially enriched in diverse biological processes, including drug metabolism and catabolism

To assess the phenotypic character of these gene modules, we performed gene ontology (GO) enrichment analysis using the Consensus Path Database (Kamburov et al., 2013).

GO enrichment analysis of each of the 11 significant gene modules (Figure 3.3) resulted in a combined total of 1,763 statistically enriched biological processes. To visualize this GO space, we used the GoSemSim package (Yu et al., 2010) in R to compute a pairwise dissimilarity score between all GO terms that were statistically enriched ($p < 0.05$) in at least one of the 11 significant modules. We projected the GO terms into a 2D space using t-distributed stochastic neighbor embedding (tSNE) (Van Der Maaten and Hinton, 2008). tSNE is popular method which computes a low-dimensional embedding of datapoints that seeks to preserve the high-dimensional distance between points in the low-dimensional space. Therefore, GO terms that describe semantically similar biological functions are placed close to one-another in the tSNE projected GO space (Figure 3.5).

This analysis revealed distinct phenotypic identities for the gene modules, spanning tissue morphogenesis, cell migration, and cell-type specification, immune response regulation and response to environmental signals, secretion, and metabolism (Figure 3.5). In particular, the yellow, salmon, and pink modules are enriched for neuroendocrine differentiation and neurotransmitter secretion and are highly upregulated in the canonical NE and NEUROD+ NE variant phenotypes (Figure 3.6). Genes in these modules are specifically upregulated in the NE phenotypes (salmon and pink modules are specific to the variant phenotype) (Figure 3.3). In contrast, the blue, black, and purple modules are enriched for cell adhesion and migration ontologies (Figure 3.6, and are highly upregulated in the non-NE variant phenotype (Figure 3.3).

Genes within the green, midnight blue, and brown modules tended to be more highly expressed in the NE variant-2 phenotype, though brown-module genes are also expressed in the canonical NE subtype, suggesting that these modules may describe the unique character of this previously unreported variant phenotype.

The brown module was enriched for canonical phenotypic features of SCLC, particularly cellular secretion and epithelial differentiation. Interestingly, the green module was enriched for immune response and drug/xenobiotic metabolism, as well as cellular response

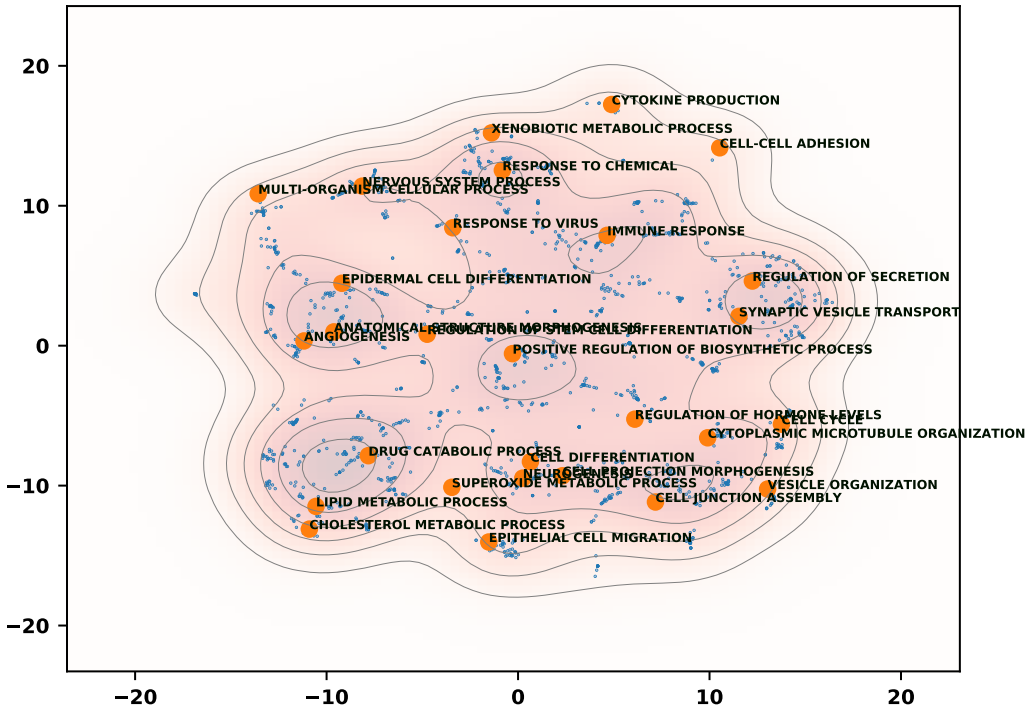


Figure 3.5: SCLC heterogeneity biological process phenospace. A dissimilarity score between pairs of SCLC enriched GO terms was calcuated using GoSemSim (Yu et al., 2010), and used to create a tSNE projection grouping similar biological processes together. Each blue dot is a GO term, with selected terms highlighted. Several distinct clusters of related processes can be seen.

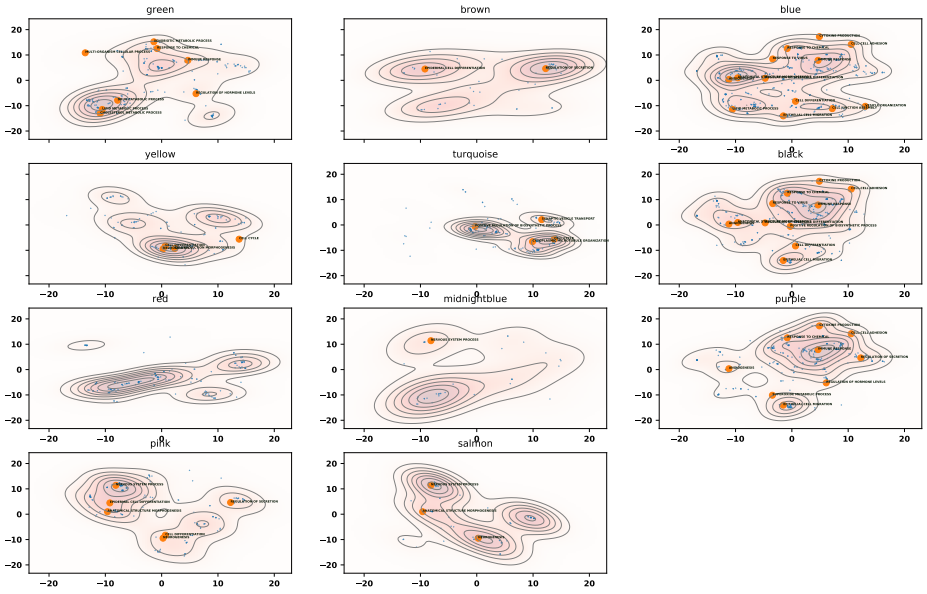


Figure 3.6: Module specific phenospace. A breakout of where each of the 11 statistically significant WGCNA modules individual falls in the GO space from Figure 3.5. Of particular interest, the green module, which is highly upregulated in the NE Variant-2 phenotype, is highly enriched in metabolic ontologies, including drug catabolism and metabolism and xenobotic metabolism.

to environmental signals. The high enrichment for drug metabolism suggests that NE variant-2 cells may represent a drug-resistant SCLC phenotype. This could further explain the drug-induced transitions toward the hybrid phenotypes reported in Chapter 2 and (Udyavar et al., 2017), as drug treated cells move toward the NE variant-2 phenotype and become more resistant.

3.4 NE variant-2 phenotype shows decreased sensitivity to a broad range of drugs

The differential gene expression profiles and GO enrichment of the 4 SCLC subtypes prompted us to investigate the response of each subtype to therapeutic agents. Because the ontology enrichment analysis suggested that genes related to drug metabolism may be overexpressed in Variant 2, compared to other phenotypes. Therefore, we sought to test the hypothesis that the Variant 2 phenotype acquires resistance by increased ability to metabolize therapeutic agents. We computed activity area of dose response curves, which accounts for both potency and efficacy, as a metric of drug sensitivity.

$$AA = \sum_{d \in doses} [1 - V(d)] \tag{3.1}$$

where $V(d)$ is the percent-viability of cells treated at dose d.

We analyzed dose-response curves from a previously published (Polley et al., 2016) drug panel containing 103 FDA-approved oncology agents and 423 investigational agents. This drug response panel included some SCLC cell lines not present in the CCLE, and therefore not assigned to the consensus clusters (Figure 3.1). For these additional cell lines, we predicted their phenotypic classification using the microarray gene expression of RET, TYRO3, CEACAM5, and EPHA2 fom (Polley et al., 2016), as inputs to the LDA model in Figure 3.4. In more than half of drugs tested, the NE Variant-2 phenotype was most resistant, suggesting that this may be a broadly drug resistant SCLC phenotype. Across the entire panel, 43 drugs had statistically significant different response based on the activity

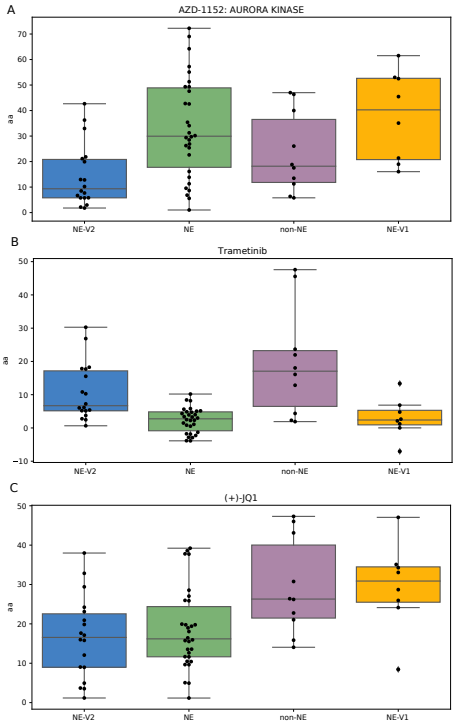


Figure 3.7: Drugs show differential response accross SCLC phenotypes. The y-axis shows the activity area (3.1), and is correlated with drug sensitivity. (A) The NE-V1 phenotype shows the most sensitivity to AZD-1152, an inhibitor of AURKB, in line with previous reports (Mollaoglu et al., 2017). Conversely, the NE-V2 phenotype is the most resistant. (B) Most SCLC cell lines are resistant to trametinib, however the NE-V2 and non-NE phenotypes are significantly more sensitive than the NE and NE-V1. (C) JQ1, a BET bromodomain inhibitor that has been used as a MYC inhibitor in SCLC (Kaur et al., 2016) has the most activity in NE-V1 and non-NE subtypes, which have the highest expression of MYC (Figure 3.2.

area (ANOVA, fdr corrected $p < 0.05$) between the four SCLC phenotypes. Of these, Variant 2 was the most resistant to 22 drugs.

To investigate patterns across the drug panel, we grouped drugs together into classes representing broad biological targets, using information from the Probes & Drugs database. The NE Variant-1 cell lines were more sensitive to aurora kinase (AURK) inhibitors than other subtypes as was previously reported. This phenotype is associated with high expression of MYC and NEUROD (Figure 3.2), reflecting similar results recently published in a NEUROD1 mouse model of SCLC with increased sensitivity to AURK inhibitors (Mollaoglu et al., 2017). Nevertheless, while we find the NE Variant-1 phenotype to be the most

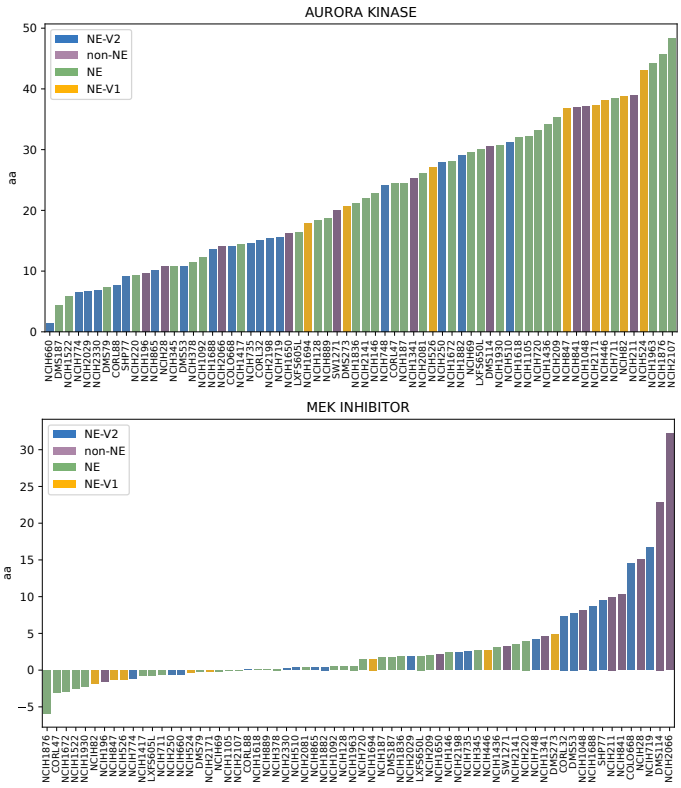


Figure 3.8: Average response of 15 AURK inhibitors and 13 MEK pathway inhibitors show differential activity across the phenotypes. Broadly, AURK inhibitors have high activity in NE-V1 and low activity in NE-V2 cells, while MEK inhibitors have the highest activity in non-NE and NE-V2 cells.

sensitive to AURK inhibition, the more notable trend is the increased resistance of the NE Variant-2 subtype to AURK inhibition (Figure 3.8).

Conversely, the MEK inhibitor class was the only category analyzed that showed higher sensitivity of Variant 2 cells that the other subtypes, though most cell lines were not affected by these inhibitors. This suggests a possible area for further study.

3.5 Transcription factor network defines SCLC phenotypic heterogeneity and reveals master regulators

We previously (Udyavar et al., 2017) identified a TF network that was able to explain the stability of NE and non-NE SCLC phenotypes (see Chapter 2), however that analy-

sis did not find attractors that distinguish the NE phenotype from the NE Variant-1 and Variant-2 phenotypes reported here. To understand how these NE-variant phenotypes are stabilized, we identified putative master transcription factor (TF) regulators from each of the 11 significant WGCNA modules. From each module, we selected master TFs based on their differential expression across the four SCLC phenotypes. Regulatory relationships between these TFs were extracted from public TF regulatory databases, including ChEA, TRANSFAC, JASPAR, and ENCODE, based on evidence of TF $\rightarrow$ DNA binding in the promoter region of a target TF, as well as several sources from the literature. The network is shown in Figure 3.9.

In Chapter 2, we simulated the network using coarse-grained Boolean modeling approaches, specifically threshold updates and inhibitor dominant rules. Nevertheless, more complex Boolean regulatory patterns are possible, and likely to be common in biology. Here we describe a method to infer logical relationships between regulators based on steady state gene expression data, and show that the inferred model is able to describe the stabilization of all four SCLC phenotype.

3.5.1 Inference of logical relationships in the TF network

A Boolean function of N input variables is a function $F : \{0,1\}^N \mapsto \{0,1\}$. The domain of F is a finite set with 2^N elements, and therefore F is completely specified by a 2^N dimensional vector in the space $\{0,1\}^{2^N}$ in which each component of the vector corresponds to the output of F for one possible input. For instance, if $F = F(A,B)$ ($N = 2$ input regulators), one possible version of F is given by the right column of Table 3.1, corresponding to $F = A \vee B$. In general, knowledge of the steady states of F is unlikely to be sufficient to fully constrain all 2^N components of the vector describing F. Therefore we adopt a probabilistic framework, in which each component of the vector is a continuous real-value $v_i \in [0,1]$ reflecting our confidence in the output of F, based on the available constraints. Components of F that are near 0.5 will indicate uncertainty about whether the

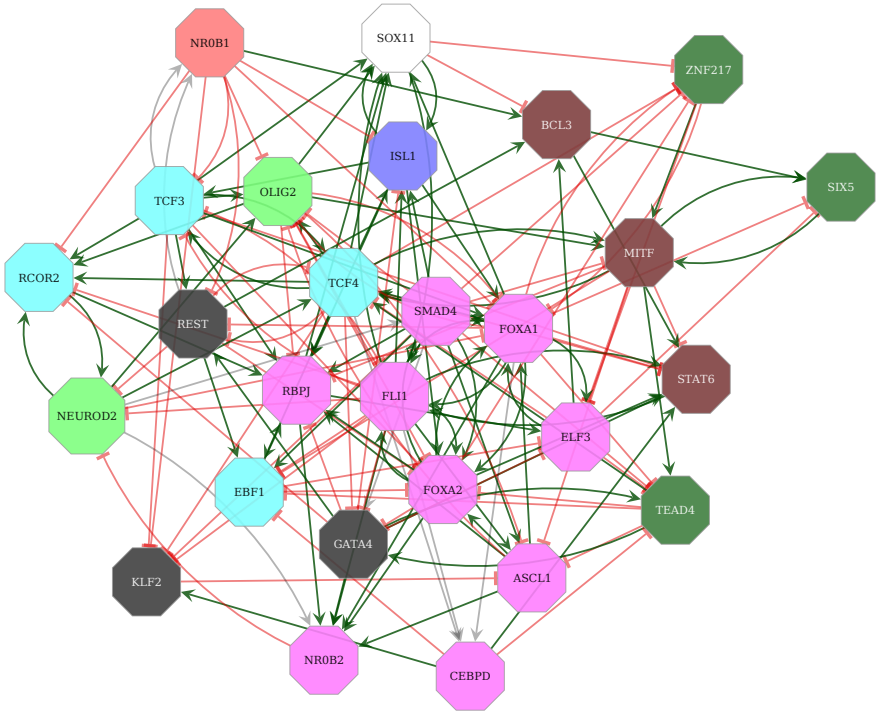


Figure 3.9: Regulatory network of differentially expressed TFs from each of the 11 coexpressed gene modules (Figure 3.3. Colors indicate which phenotype each TF is upregulated in. Black is upregulated in non-NE, light green NE-V1, blue in NE, red in NE-V2, cyan in NE-V1 and NE, magenta in NE and NE-V2, brown in NE-V2 and non-NE, dark green in NE-V1 and non-NE, and white in NE, NE-V1, and NE-V2. Red edges indicate inhibition (on average), and green activation (on average).

A	B	F(A,B)
0	0	0
0	1	1
1	0	1
1	1	1

Table 3.1: An example Boolean function, $F(A,B) = A \vee B$.

output should be 0 or 1, given the available constraining data.

Given M observations (in our case, each observations is a measurement of gene expression of the N regulator TFs and the target TF in $M = 50$ cell lines), we want to compute this probabilistic vector $\vec{V}$ describing a probabilistic Boolean function F of N variables. First, we organize the input-output relationship as a binary decision tree with N layers leading to the 2^N leaves, each of which corresponds to a component of vector $\vec{V}$. For instance, given two regulators A and B ($N = 2$), the leaves of the binary decision tree correspond to the probabilities that $(\overline{A} \wedge \overline{B})$, $(\overline{A} \wedge B)$, $(A \wedge \overline{B})$, and $(A \wedge B)$. Collectively, the observations define an $M \times N$ matrix $\mathbf{R} = \left[\vec{R}_1, \vec{R}_2, \ldots, \vec{R}_N\right]$ quantifying the input regulator variables (columns) for each observation (rows), as well as an M dimensional vector $\vec{T} = [t_1, t_2, \ldots, t_M]$ quantifying the target, output variable. A Gaussian mixed model is then used to transform the columns of $\mathbf{R}$ (regulator variables) and the vector $\vec{T}$ into probabilities $\mathbf{R}'$ and $\vec{T}'$ of the variables being OFF or ON in each observation (row).

Let $P_j(\vec{R'}_i)$ be a function that quantifies the probability that the input variables of the i^{th} observation belong to the j^{th} leaf of the binary decision tree. For instance using the example above, the second leaf of the binary tree is $(\overline{A} \wedge B)$. Therefore, $P_{j=2}(A,B) = (1-A) \cdot B$. Note that by this definition, $\sum_{j=1}^{2^N} P_j(\vec{R'}_i) = 1$. Using this, we define an $M \times 2^N$ weight matrix $\mathbf{W} = w_{i,j}$ as

$$w_{i,j} = P_j(\vec{R'}_i) \tag{3.2}$$

that describes how much the i^{th} observation constrains the j^{th} component of $\vec{V}$. Additionally, to avoid overfitting underdetermined leaves, we define the uncertainty $\vec{U} =$

$[u_1, u_2, \ldots, u_{2^N}]$ of each leaf

$$u_j = 1 - \max_{i \in \{1,\ldots,M\}} (w_{i,j}) \tag{3.3}$$

From these, we then define the vector $\vec{V}$ describing function F as

$$v_j = \frac{\sum_{i=1}^{M} t'_i \cdot w_{i,j} + 0.5 \cdot u_j}{\sum_{i=1}^{M} w_{i,j} + u_j} \tag{3.4}$$

Thus, each component of $\vec{V}$ is the average of the output target variable $\vec{T}$ weighted by $\mathbf{W}$, with an additional uncertainty term $\vec{U}$ to avoid overfitting. For leaves j of the binary tree that are poorly constrained by any of the observables, $v_j \approx 0.5$, indicating maximal uncertainty in the output of F at those leaves. Uncertainty of a leaf j also arises when observations i with large weight $w_{i,j}$ have inconsistent values for t'_i, such as if $t'_1 = 0$ and $t'_2 = 1$.

3.5.2 Probabilistic simulation of SCLC GRN identifies stable states corresponding to the four SCLC subtypes

Probabilistic logical relationships were derived for all nodes of the SCLC TF network in Figure 3.9. Figure 3.10 shows the rule fitting results for ASCL1, NEUROD2, REST, and ELF3. The dynamics of Boolean networks can be simulated as discrete state transitions forming a state transition graph (STG), where each state is defined by the TFs that are ON and OFF. Here we simulate the dynamics of the SCLC network using the inferred probabilistic rules, and a general-asynchronous (Albert et al., 2008) update scheme. In a general-asynchronous update, one TF is selected at random to be updated. Therefore, in this 27-node TF network, each state in the STG has 27 possible neighbors, corresponding to switching each one of the TFs, and the probability of traversing that edge is given by the probability in the inferred rule.

The large size of the STG, which has $2^{27} = 134,217,728$ possible states, precludes an

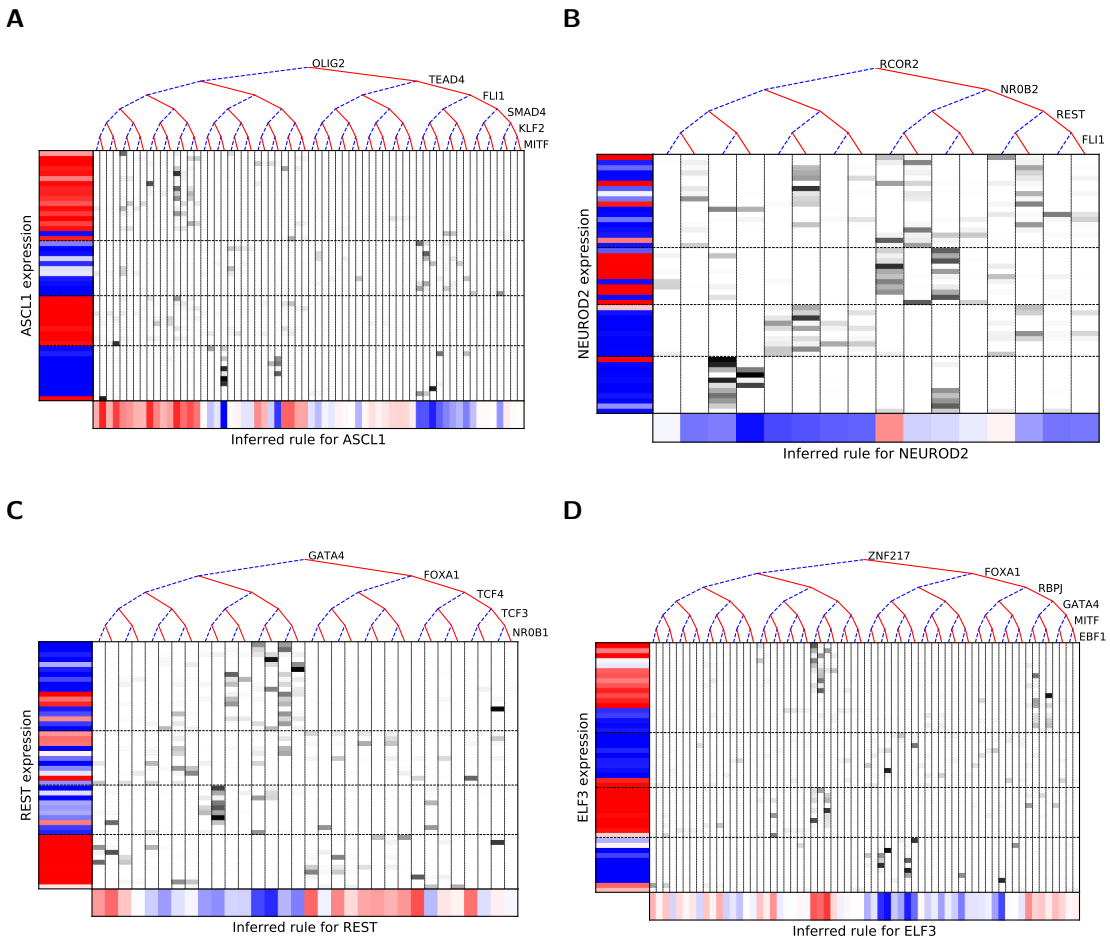


Figure 3.10: Probabilistic Boolean rule fits for (A) ASCL1, (B) NEUROD2, (C) REST, and (D) ELF3. In each case, the target gene is a function of all the genes along the binary tree at the top, while expression of the target is shown on the left. Each row represents one cell line, each column represents on possible input state, and the bottom shows the inferred function F for every possible input state. Color ranges from 0=blue (highly confident the TF is off), to 0.5=white (no information about whether the TF should be on or off), to 1=red (highly confident the TF is on). The greyscale indicates how much each cell line (row) constrains each state (column). Rows are organized by subtype (top to bottom: NE, NE-V1, NE-V2, non-NE). For instance in (A), many NE cell lines constrain the state OLIG2=0, TEAD4=0, FLI1=1, SMAD4=1, KLF2=0, MITF=0, and in all of those cell lines, ASCL1 is on. Therefore, that leaf of the rule at the bottom is highly confident that ASCL1 should be ON.

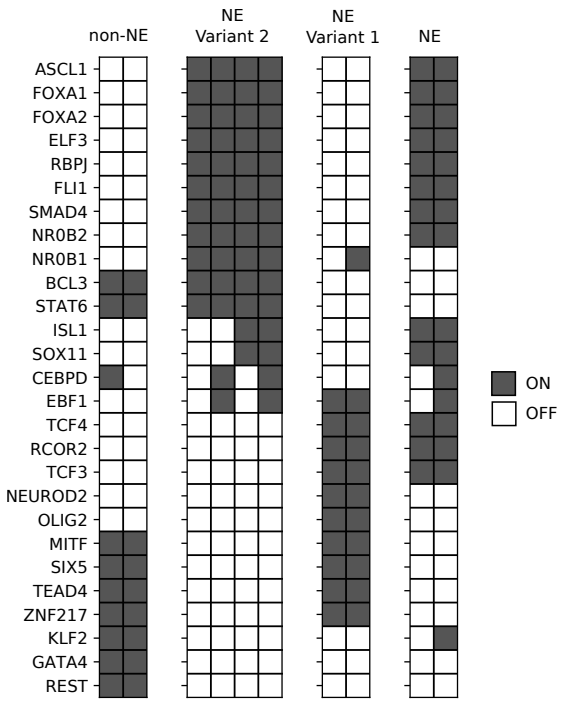


Figure 3.11: Stable states of the inferred probabilistic rules. Each state has a probability $<$ 0.5 of transitioning away from itself for all TFs, under the probabilistic Boolean dynamics. Based on gene expression, these attractors correspond to the four SCLC phenotypes, and reveal the overlaps between them.

exhaustive search through the state space. Nevertheless, regions of interest can be probed, both by exhaustive search of neighborhood of states we are interested in, and by random walks. To focus on areas of interest, we calculated an "average" state for each phenotype by averaging the expression of TFs across all cell lines belonging to the phenotype, and rounding values to 0 or 1. Next, we exhaustively searched the neighborhood of each of these starting states out to a distance of 6 TF changes.

Within this subset of the overall state space, we found 10 states for which all 27 TFs had a probability less than 50% of switching. Transitions into these states are therefore more likely than transitions out, such that they are stable states of the network dynamics (Figure 3.11).

3.5.3 *In silico* activation and knockdown of TFs identifies master regulators that increase or decrease stability of SCLC phenotypes

To quantify the baseline stability of the steady states in Figure 3.11, we performed random walks starting from each stable state, and counted how many steps were required to reach a state with outside of a 4-TF neighborhood around the starting state. The random walk was executed following Algorithm 1.

Algorithm 1 Probabilistic Boolean Random Walk

procedure RANDOM WALK
Input:
state_init: Starting Boolean state
rules: Probabilistic update rules that map *state* to a probability of turning each TF ON or OFF
R: Maximum walk radius
Output:
Number of steps taken before the random walk is a distance greater than *R* from *state_init*
    $state \leftarrow state\_init$
    $steps \leftarrow 0$
    while $d(state, state\_init) \leq R$ **do**
        $steps \leftarrow steps + 1$
        $TF \leftarrow$ a randomly chosen TF
        $probability\_update \leftarrow rules(state, TF)$
        $r = random\_uniform(0,1)$
        if $r < probability\_update$ **then**
            $state[TF] \leftarrow 1 - state[TF]$
    Return: $steps$

1000 random walks were executed from each stable state, and the average number of steps before the walk left a 4-TF neighborhood of the starting state was used as a measure of the stability of the state. *In silico* TF activation and knockdowns were carried out by fixing the target TF as ON or OFF, and setting the probability of it changing to 0. Each TF in the network was individually activated and knocked down in each of the stable states, and 1000 random walks were executed for each condition, keeping the perturbed TFs constant. The percent increase or decrease of the stability relative to the baseline stability was calculated,

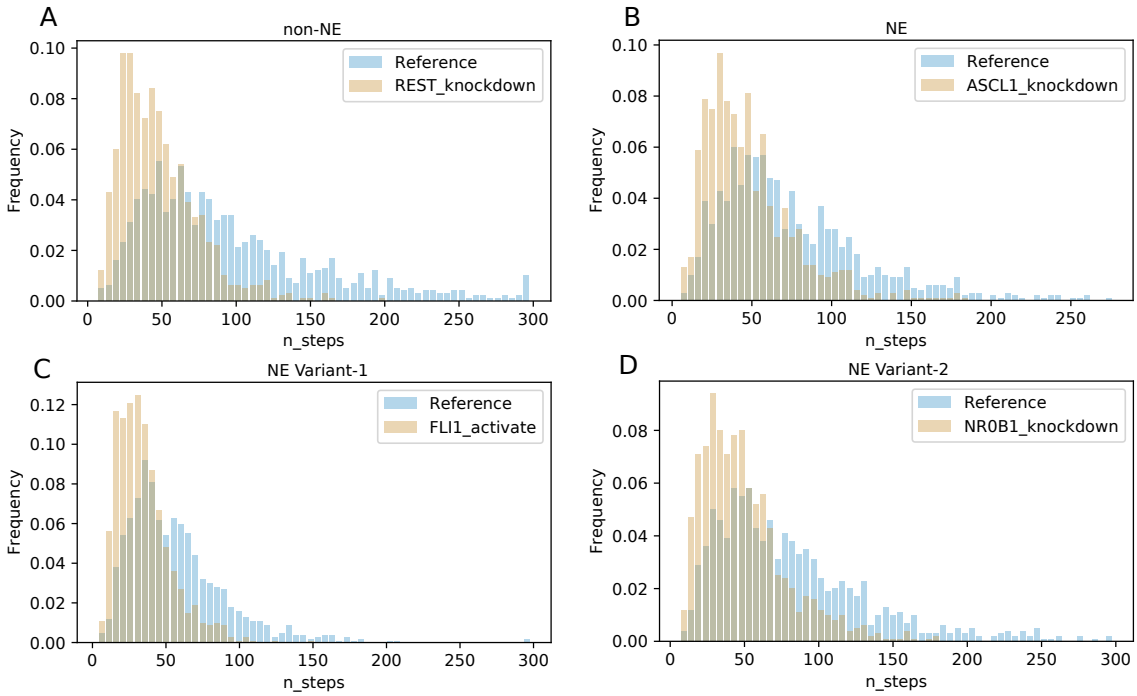


Figure 3.12: Random walks starting from the attractors in Figure 3.11 will eventually get away from the start state due to uncertainty in the Boolean rules. "Reference" counts how many random steps are required to get a distance greater than 4 TFs away from the start state under the network's natural dynamics. The knockdowns and activations shown here hold expression of the perturbed gene OFF or ON. In these four examples, the perturbation destabilized the start state, such that the random walk left the neighborhood sooner.

resulting in a score of how much activating or knocking out each TF stabilizes the starting state phenotype (Figure 3.13).

Figure 3.13A shows thatt the model predicts that activation of many TFs, including FOXA1/2, SMAD4, OLIG2, ELF3, NEUROD2, TCF4, and ASCL1, destabilizes the non-NE phenotype, suggesting that these are collectively master NE or NE-Variant regulators. Conversely, activation of TCF3 stabilizes the non-NE phenotype, as does knockout of TEAD4 and REST. REST has previously been implicated as a master regulator of non-NE differentiation in SCLC (Lim et al., 2017), as has TEAD4 through its role as an activator of YAP and TAZ through the hippo pathway (Horie et al., 2016).

Similarly key regulators are predicted for the other subtypes. Of note, ASCL1 and FOXA1/2 are predicted as regulators stabilizing the NE and NE Variant-2 phenotypes, while NEUROD2 and OLIG2 stabilize the NE Variant-1. SMAD4 and TCF4 are both found to stabilize the NE and NE Variant-1 phenotypes, while SOX11 stabilizes both NE variants.

While many of the predicted master regulators, including REST, TEAD4, ASCL1, and NEUROD2, match previously reported regulators of SCLC subtypes, further work modulating expression of the additional TFs will be required to understand and validate their roles. Particularly of interest, knockdown of ELF3 and NR0B1 is predicted to destabilize only the NE Variant-2 phenotype, and therefore modulation of these genes may provide key therapeutic advantages by forcing SCLC cells to move away from the more drug resistant phenotype.

3.6 Discussion

These findings further clarify the role that SCLC phenotypic heterogeneity plays in the context of differential drug sensitivity, and potentially therapeutic response. We have positioned our four subtypes (non-NE, NE, NE Variant-1, and NE Variant-2) within the broader literature of SCLC heterogeneity. One advantage of this analysis is that instead of

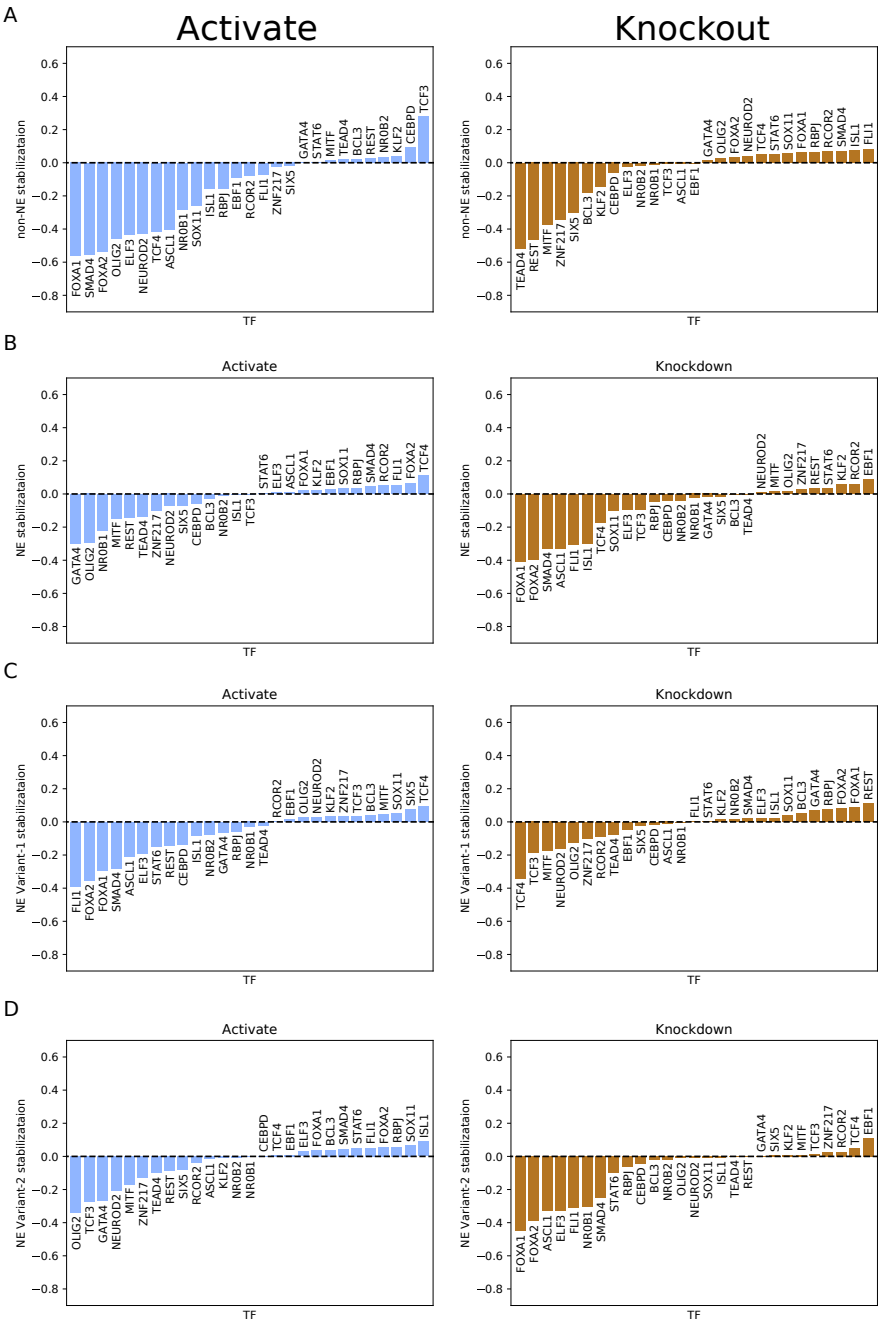


Figure 3.13: Stablization of SCLC phenotypes by TF knockdown and activation. The percent change of stability measures the percent change in the average number of steps needed to leave the neighborhood of the stable states. Negative indicates destabilizing, while positive indicates increasing stability. Results are shown for 1000 iterations starting from (A) non-NE, (B) NE, (C) NE-V1, and (D) NE-V2.

defining each subtype by expression of one or a few markers, we have identified distinct co-expressed gene programs that participate in unique biological processes. A key finding of this was the enrichment of the NE Variant-2 subtype with drug/xenobiotic metabolism/catabolism genes. A drug screen across a broad range of compounds indicated that this phenotype is generally more resistant than the others, though the trend may be reversed in MEK pathway genes, which may be a promising target for future combination therapies.

The distinction between the NE and NE Variant-2 phenotypes is significant, because previous markers of SCLC heterogeneity were unable to differentiate these two subtypes. Nevertheless, validation in mice or in humans of the existence of the NE Variant-2 subtype, and its drug resistant properties will be of significant importance. To this end, we developed a panel of four cell-surface biomarkers that, *in vitro*, are able to clearly distinguish all four subtypes.

Finally, we have introduced a novel method for identifying key master regulators of phenotypic heterogeneity. A key benefit of this method is that it does not overfit the data - each TF is assumed to diffuse randomly in states for which there has been no data to constrain its dynamics. With this method we were able to recapitulate known master regulators of SCLC heterogeneity, as well as making novel predictions. One future direction currently being pursued in the lab is reprogramming cells via induced expression or knockout of target TFs, and future studies will be directed at tracking cell phenotype identities, as well as drug sensitivities, after such perturbations.

In summary, we have provided a framework to more comprehensively understand diverse biological processes, regulation, and drug response associated with SCLC heterogeneity.

Chapter 4

Modeling tumor relapse in heterogeneous populations of cancer cells

4.1 Introduction

Heterogeneity has been heavily implicated in the emergence of resistance, and eventual relapse, in almost all forms of cancer. A recent report introduced the clonal fractional proliferation method (Frick et al., 2015) to quantify the heterogeneous proliferative dynamics of clonal lineages within cancer cell populations. This method quantifies the drug induced proliferation (DIP) rate for single-cell clonal lineages within a cell population Here, we develop a mathematical framework, named the "heterogeneous growth model" (HGM) for predictive modeling of population level drug response based on the underlying distribution of clonal DIP rates. The HGM explicitly quantifies the role of clonal fitness variation in driving the rate of relapse, and is flexible enough to be applied to several possible DIP rate distributions. We show that the HGM can be used to make quantitative predictions of time-to-relapse (TTR), defined as the time it takes for a drug treated population of cells to regress, and return back to the original population level. Sensitivity analysis is performed to identify cases for which combination treatments targeting the DIP rate distribution mean, variance, or skewness are likely to have the most significant impact on TTR.

4.2 Notation

Throughout this chapter we will use the following notation: P = Total population

r = Proliferation rate of a given set of cells

$x(r)$ = Number of cells with proliferation rate r

$z(r)$ = Normalized proliferation rate distribution (i.e. $z(r) = \frac{x(r)}{P}$)

μ = Average proliferation rate of entire population

σ = Standard deviation of proliferation rates

μ_n = The n^{th} central moment of the proliferation rate distribution. Note $\mu_2 = \sigma^2$

μ'_n = The n^{th} raw moment of the proliferation rate distribution. Note $\mu'_1 = \mu$

t = Time

Unless it is otherwise made explicitly clear, each variable (except for r) is considered to potentially be a function of time (*e.g.* $P = P(t)$, $x(r) = x(r,t)$). Occasionally for clarity this dependence will be stated explicitly, but for the sake of brevity it will generally be dropped.

4.3 Population level proliferative dynamics can be quantified by the moment generating function of the underlying DIP rate distribution

4.3.1 HGM Assumptions

To quantitatively describe the rebounding growth dynamics of a population of cells whose proliferation rates are distributed with non-zero variance, we will here make two biological assumptions, and one technical assumption about the admissible DIP rate distributions. Using data from PC9 cells treated with erlotinib, we later demonstrate the validity of the biological assumptions for at least some systems, though we also discuss situations in which they may not hold.

1. Proliferation rate is completely heritable. That is, all cells with a given proliferation rate, r, produce daughter cells which have the same proliferation rate.

2. A population of cells with a given proliferation rate r grows or shrinks according to exponential population growth, independent of all other populations.

3. The initial distribution of proliferation rates has a well defined and finite moment generating function (MGF) $\forall t \geq 0$ (where t is the conventional MGF parameter).

It should be noted that for simplicity the following derivations treat the proliferation rate as a continuous distribution, but the results may be equally formulated as a discrete distribution by replacing integrals with sums where relevant.

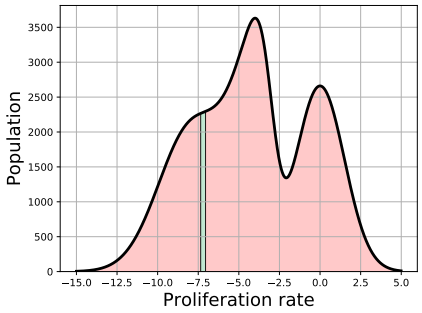


Figure 4.1: Cell population with heterogeneous proliferation rates. Cells are divided into continuous subpopulations with a fixed proliferation rate r. Here we highlight in green a subpopulation with negative proliferation rate, which will thus decay exponentially over time. The total population is equivalent to the total area under the curve.

4.3.2 Population dynamics follow the moment generating function of the initial DIP rate distribution

Assumptions 1 and 2 can be written as a differential equation

$$\frac{\mathrm{d}x(r,t)}{\mathrm{d}t} = r \cdot x(r,t) \tag{4.1}$$

This has the well-known solution

$$x(r,t) = x(r,0)e^{rt} \tag{4.2}$$

Here, $x(r,0)$ is the initial measurement of cell counts with proliferation rate r. To obtain the full population, we integrate over all r

$$P(t) = \int_{-\infty}^{\infty} x(r,t)\,\mathrm{d}r \tag{4.3}$$

See Figure 4.1 for an example of the model setup.

Substituting (4.2) into (4.3), and recalling that $z(r) = \frac{x(r)}{P}$ we find

$$P(t) = P(0) \int_{-\infty}^{\infty} z(r,0) e^{rt} \, \mathrm{d}r \tag{4.4}$$

which is the definition of the MGF (Kenney, 1939) of the initial proliferation rate distribution, $z(r,0)$, multiplied by the initial population. All further analyses thus rely on the convergent behavior of the integral in equation (4.4), as stated in assumption 3. Since the MGF of many common distributions is known, it is possible to compute the total population at any time t by looking up the appropriate MGF. Nevertheless, a more intuitive understanding of the model may be achieved by further analysis.

4.4 The variance of the DIP rate distribution drives the population rebound

To put (4.3) into a more meaningful form, we compute $\frac{\mathrm{d}P}{\mathrm{d}t}$ as

$$\begin{aligned} \frac{\mathrm{d}P(t)}{\mathrm{d}t} &= \int_{-\infty}^{\infty} \frac{\mathrm{d}}{\mathrm{d}t} x(r,t) \, \mathrm{d}r \\ &= \int_{-\infty}^{\infty} r \cdot x(r,t) \, \mathrm{d}r \\ &= \frac{\int_{-\infty}^{\infty} r \cdot x(r,t) \, \mathrm{d}r}{\int_{-\infty}^{\infty} x(r,t) \, \mathrm{d}r} \int_{-\infty}^{\infty} x(r,t) \, \mathrm{d}r \end{aligned}$$

Because $P(t) = \int_{-\infty}^{\infty} x(r,t) \, \mathrm{d}r$, we find

$$\frac{\int_{-\infty}^{\infty} r \cdot x(r,t) \, \mathrm{d}r}{\int_{-\infty}^{\infty} x(r,t) \, \mathrm{d}r} = \int_{-\infty}^{\infty} r \cdot z(r,t) \, \mathrm{d}r \equiv \mu(t) \tag{4.5}$$

Thus we find that

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = \mu(t) P(t) \tag{4.6}$$

Thus, at any time t, the total population proliferates at a rate equal to the average proliferation of the underlying clonal lineages. Furthermore, the general solution of equations of the form (4.6) is given by

$$P(t) = P_0 e^{\int_0^t \mu(t') \, \mathrm{d}t'} \tag{4.7}$$

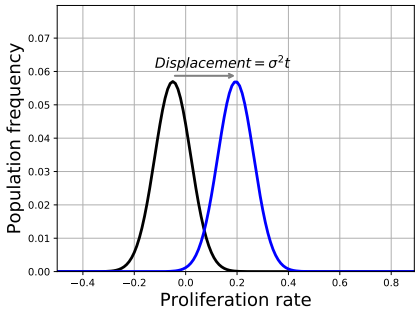


Figure 4.2: A normal distribution of proliferation rates moves to the right under (4.1) with constant velocity given by its variance. The shape of the distribution stays the same as the mean increases.

From (4.6), it is clear that if $\mu(t)$ is constant, the population dynamics reduce to exponential growth or decay given by the average proliferation rate. However, if $\mu(t)$ is increasing, the population growth is predicted to be super-exponential, whereas if $\mu(t)$ is decreasing, the population growth is sub-exponential. To understand how $\mu(t)$ changes with time, we compute

$$\begin{aligned}
\frac{\mathrm{d}\mu(t)}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\frac{\int_{-\infty}^{\infty} r\cdot x(r,t)\,\mathrm{d}r}{\int_{-\infty}^{\infty} x(r,t)\,\mathrm{d}r} \\
&= \frac{\int_{-\infty}^{\infty} x\,\mathrm{d}r\int_{-\infty}^{\infty} r\frac{\mathrm{d}x}{\mathrm{d}t}\,\mathrm{d}r - \int_{-\infty}^{\infty} rx\,\mathrm{d}r\int_{-\infty}^{\infty}\frac{\mathrm{d}x}{\mathrm{d}t}\,\mathrm{d}r}{(\int_{-\infty}^{\infty} x\,\mathrm{d}r)^2} \\
&= \frac{\int_{-\infty}^{\infty} x\,\mathrm{d}r\int_{-\infty}^{\infty} r^2 x\,\mathrm{d}r}{(\int_{-\infty}^{\infty} x\,\mathrm{d}r)^2} - \frac{\int_{-\infty}^{\infty} rx\,\mathrm{d}r\int_{-\infty}^{\infty} rx\,\mathrm{d}r}{(\int_{-\infty}^{\infty} x\,\mathrm{d}r)^2}
\end{aligned}$$

In the first term we may cancel $\int_{-\infty}^{\infty} x\,\mathrm{d}r$ from the numerator and denominator. In the second, recall from (4.5) that $\frac{\int_{-\infty}^{\infty} rx\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} = \mu$

$$\begin{aligned}
&= \frac{\int_{-\infty}^{\infty} r^2 x\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} - \mu^2 \\
&= \mathrm{E}\left[r^2\right] - (\mathrm{E}[r])^2 \\
&= \sigma^2(t) \qquad (4.8)
\end{aligned}$$

Thus if a population of cells has a distribution of proliferation rates with non-zero variance, it will grow super-exponentially as long as assumptions 1 and 2 are satisfied (if there is no density of positive DIP clones it will not "grow", but the exponential decay will slow down). This finding is a form of Fisher's fundamental theorem of natural selection which states that "the rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time." (Qian, 2013; Price, 1972; Fisher, 1930). In cancer, this is particularly significant when drug treatment initially induces an on-average negative DIP rate for a population of cells, which is then driven to be on-average positive DIP rate by variance in the underlying DIP rate, resulting in a rebounding population.

It should be noted that while this model in general may allow for limitless increase of μ, it is expected that for large times either assumption 1 or 2 may break down. Specifically, there are physical constraints on the maximum possible proliferation rate, and we expect the variance to decrease as the cell population becomes dominated by the fastest proliferating subpopulation. Furthermore, other recent reports suggest that in other cancer types, phenotypic transitions may cause a violation of assumption 1 (Paudel et al., 2018), leading to deviations from the HGM predictions.

In the next sections, we explore the dynamics of (4.6) and (4.8) in the specific cases of normally distributed and skewed DIP rate distributions.

4.4.1 Normal distribution

If proliferation rates are normally distributed, the total population (4.4) is the well-known MGF for the normal distribution:

$$P(t) = P(0)e^{\mu(0)t+\frac{\sigma(0)^2t^2}{2}} \tag{4.9}$$

Here we can see (as we later prove directly) that the variance must be constant (Figure

4.2), allowing us to directly integrate equation (4.8) as

$$\frac{\mathrm{d}\mu(t)}{\mathrm{d}t} = \sigma^2$$
$$\int_0^t \mathrm{d}\mu(t') = \int_0^t \sigma^2 \,\mathrm{d}t'$$
$$\mu(t) = \mu(0) + \sigma^2 t$$

and then directly integrating the exponent of (4.7). To understand (and later prove) why the variance is constant, consider

$$\sigma^2 = \frac{\int_{-\infty}^{\infty} x(r-\mu)^2 \,\mathrm{d}r}{\int_{-\infty}^{\infty} x \,\mathrm{d}r}$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma^2 = \frac{\int_{-\infty}^{\infty} x \,\mathrm{d}r \int_{-\infty}^{\infty} \left[xr(r-\mu)^2 - 2x\frac{\mathrm{d}\mu}{\mathrm{d}t}(r-\mu)\right] \mathrm{d}r - \int_{-\infty}^{\infty} x(r-\mu)^2 \,\mathrm{d}r \int_{-\infty}^{\infty} xr \,\mathrm{d}r}{\left(\int_{-\infty}^{\infty} x \,\mathrm{d}r\right)^2}$$

Splitting up the terms, and noting that $\frac{\mathrm{d}\mu}{\mathrm{d}t} = \sigma^2$ we find

$$= \frac{\int_{-\infty}^{\infty} xr(r-\mu)^2 \,\mathrm{d}r}{\int_{-\infty}^{\infty} x \,\mathrm{d}r} - 2\sigma^2 \frac{\int_{-\infty}^{\infty} xr \,\mathrm{d}r}{\int_{-\infty}^{\infty} x \,\mathrm{d}r} + 2\sigma^2\mu - \frac{\int_{-\infty}^{\infty} x(r-\mu)^2 \,\mathrm{d}r}{\int_{-\infty}^{\infty} x \,\mathrm{d}r} \cdot \frac{\int_{-\infty}^{\infty} xr \,\mathrm{d}r}{\int_{-\infty}^{\infty} x \,\mathrm{d}r}$$

Notice the occurance of the average rate (4.5) in the second and last term, and the variance (4.8) also in the last term.

$$= \frac{\int_{-\infty}^{\infty} xr(r-\mu)^2 \,\mathrm{d}r}{\int_{-\infty}^{\infty} x \,\mathrm{d}r} - 2\sigma^2\mu + 2\sigma^2\mu - \sigma^2\mu$$
$$= \frac{\int_{-\infty}^{\infty} xr(r-\mu)^2 \,\mathrm{d}r}{\int_{-\infty}^{\infty} x \,\mathrm{d}r} - \sigma^2\mu$$

Expanding the numerator of the first term gives

$$= \frac{\int_{-\infty}^{\infty} xr^3 \,\mathrm{d}r}{\int_{-\infty}^{\infty} x \,\mathrm{d}r} - 2\mu \frac{\int_{-\infty}^{\infty} xr^2 \,\mathrm{d}r}{\int_{-\infty}^{\infty} x \,\mathrm{d}r} + \mu^2 \frac{\int_{-\infty}^{\infty} xr \,\mathrm{d}r}{\int_{-\infty}^{\infty} x \,\mathrm{d}r} - \sigma^2\mu \tag{4.10}$$

Notice here that the integrals in the third term are the average proliferation rate (4.5). To

deal with the second term recall that the variance is defined in terms of the *expectation*, $E[A(r)] = \int_{-\infty}^{\infty} A(r) \cdot z(r)\,\mathrm{d}r$ as

$$\begin{aligned}\sigma^2 &= \mathrm{E}\left[r^2\right] - (\mathrm{E}[r])^2 \\ &= \frac{\int_{-\infty}^{\infty} x r^2\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} - \mu^2\end{aligned}$$

Thus we can see that

$$\frac{\int_{-\infty}^{\infty} x r^2\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} = \sigma^2 + \mu^2 \tag{4.11}$$

Plugging this back into (4.10) we get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\sigma^2 &= \frac{\int_{-\infty}^{\infty} x r^3\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} - 2\mu(\sigma^2 + \mu^2) + \mu^3 - \sigma^2\mu \\ &= \frac{\int_{-\infty}^{\infty} x r^3\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} - \mu^3 - 3\sigma^2\mu \\ &= \frac{\int_{-\infty}^{\infty} x(r^3 - \mu^3)\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} - 3\sigma^2\mu \qquad (4.12)\end{aligned}$$

To continue we show that (4.12) is in fact equivalent to the third central moment, μ_3. To do this, consider the definition of μ_3

$$\begin{aligned}\mu_3 &= \frac{\int_{-\infty}^{\infty} x(r-\mu)^3\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} \qquad (4.13) \\ &= \frac{\int_{-\infty}^{\infty} x(r^3 - 3r^2\mu + 3r\mu^2 - \mu^3)\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} \\ &= \frac{\int_{-\infty}^{\infty} x(r^3 - \mu^3)\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} - 3\mu\frac{\int_{-\infty}^{\infty} x r^2\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} + 3\mu^2\frac{\int_{-\infty}^{\infty} x r\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r}\end{aligned}$$

The second term is the second raw moment, which we calculated in (4.11), and the third term is again the definition of average growth rate from (4.5). Substituting these yields

$$\begin{aligned} &= \frac{\int_{-\infty}^{\infty} x(r^3 - \mu^3)\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} - 3\mu(\sigma^2 + \mu^2) + 3\mu^3 \\ &= \frac{\int_{-\infty}^{\infty} x(r^3 - \mu^3)\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} - 3\sigma^2\mu \end{aligned} \tag{4.14}$$

Thus in (4.14) we show that (4.12) is equivalent to the third central moment μ_3 as defined in (4.13), or

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma^2 = \mu_3 \tag{4.15}$$

The third central moment is closely related to the skewness ($\gamma = \frac{\mu_3}{\sigma^3}$), and thus because the normal distribution has no skewness, its variance is constant. This is of course not a proof, as we have not yet addressed whether a distribution which is initially normal may become skewed over time. As we show later, however, the shape of the normal distribution is invariant under the HG model, and the variance is indeed constant.

4.4.2 Skewed distributions

Equation (4.15) immediately leads to consideration of skewed distributions which have in general $\mu_3 \neq 0$. Indeed, if $\mu_3 < 0$, that is, the distribution is negatively skewed, the variance decreases over time leading to slower rebound (Figure 4.3). If instead $\mu_3 > 0$, that is, the distribution is positively skewed, the variance increases leading to faster rebound (Figure 4.4).

Interestingly, through numerical simulation of (4.4), it is further observed that over time, a positively or negatively skewed skew-normal(Azzalini, 2005) distribution will approach a normal distribution, which then proceeds with a constant variance (Figures 4.3 - 4.5).

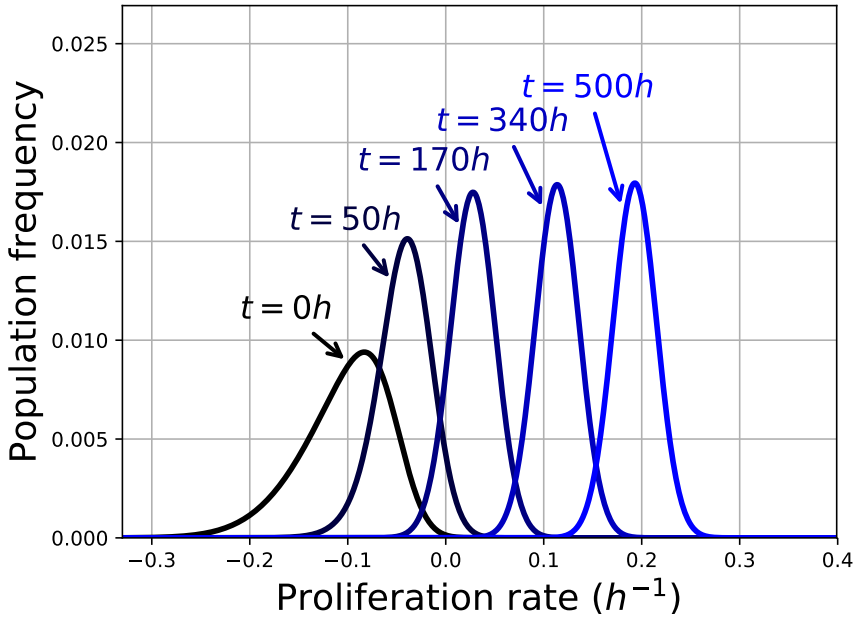


Figure 4.3: Negative skew causes decreased variance. A negative skewnormal ($\alpha = -3$) distribution has a decreasing variance, until it reaches steady state. The decreased variance leads to significantly slower rebound dynamics.

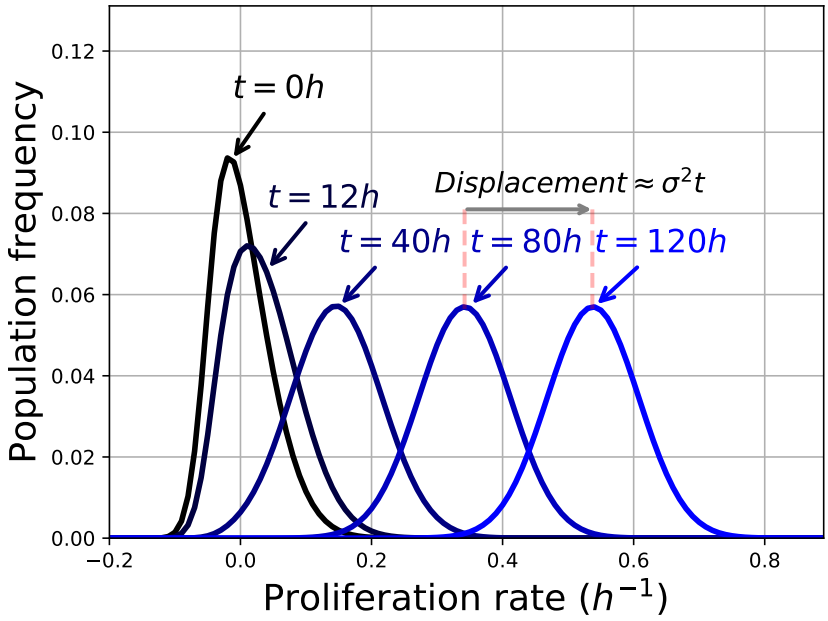


Figure 4.4: Positive skew causes decreased variance. A positive skewnormal ($\alpha = 3$) distribution has an increasing variance, until it reaches steady state. The increased variance leads to accelerated rebound dynamics.

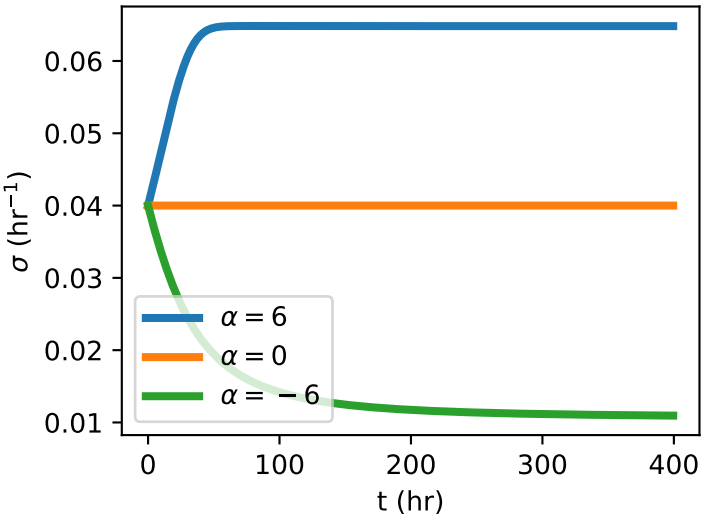


Figure 4.5: Over time, a skewnormal distribution becomes more normal until it reaches a steady state. Negative skew cases a decrease in variance, while positive skew causes an increase.

4.4.3 Higher order moments

While the increase in average (and thus observed, (4.6)) proliferation rate is driven by the variance, as described by equation (4.8), the variance may in general be modulated through over time by the general shape of the distribution. In (4.15) we showed that the derivative of variance is exactly the third central moment, but of course now we must consider how the third central moment changes over time. More generally, consider the n^{th} central moment μ_n given by

$$\mu_n = \frac{\int_{-\infty}^{\infty} x(r) \cdot (r-\mu)^n \, \mathrm{d}r}{\int_{-\infty}^{\infty} x(r) \, \mathrm{d}r} \tag{4.16}$$

We seek to find a general form for its derivative, as

$$\frac{\mathrm{d}\mu_n}{\mathrm{d}t} = \frac{\int_{-\infty}^{\infty} x\,\mathrm{d}r \int_{-\infty}^{\infty} \left[xr(r-\mu)^n - nx\sigma^2(r-\mu)^{n-1}\right]\mathrm{d}r - \int_{-\infty}^{\infty} x(r-\mu)^n\,\mathrm{d}r \int_{-\infty}^{\infty} xr\,\mathrm{d}r}{(\int_{-\infty}^{\infty} x\,\mathrm{d}r)^2}$$
$$= \frac{\int_{-\infty}^{\infty} xr(r-\mu)^n\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} - n\sigma^2 \frac{\int_{-\infty}^{\infty} x(r-\mu)^{n-1}\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} - \frac{\int_{-\infty}^{\infty} x(r-\mu)^n\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} \cdot \frac{\int_{-\infty}^{\infty} xr\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r}$$
$$= \frac{\int_{-\infty}^{\infty} xr(r-\mu)^n\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} - n\sigma^2\mu_{n-1} - \mu_n\mu$$

Because n must be a positive integer, the first term may be expanded using the binomial theorem

$$= \sum_{k=0}^{n} \binom{n}{k} (-\mu)^k \frac{\int_{-\infty}^{\infty} xr^{n-k+1}\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} - n\sigma^2\mu_{n-1} - \mu_n\mu$$

By separating the first and last terms from the sum, we may find the terms associated with the $(n+1)^{th}$ moment:

$$= \frac{\int_{-\infty}^{\infty} xr^{n+1}\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} + (-1)^n(\mu)^{n+1} + \sum_{k=1}^{n-1} \binom{n}{k} (-\mu)^k \frac{\int_{-\infty}^{\infty} xr^{n-k+1}\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} - n\sigma^2\mu_{n-1} - \mu_n\mu$$

Here we must be careful, as the sign of $(-1)^n$ becomes dependent on n. If n is odd:

$$\frac{\mathrm{d}\mu_n}{\mathrm{d}t} = \frac{\int_{-\infty}^{\infty} x(r^{n+1} - \mu^{n+1})\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} + \sum_{k=1}^{n-1} \binom{n}{k} (-\mu)^k \frac{\int_{-\infty}^{\infty} xr^{n-k+1}\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} - n\sigma^2\mu_{n-1} - \mu_n\mu \quad (4.17)$$

If n is even:

$$\frac{\mathrm{d}\mu_n}{\mathrm{d}t} = \frac{\int_{-\infty}^{\infty} x(r^{n+1} - \mu^{n+1})\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} + 2\mu^{n+1} + \sum_{k=1}^{n-1} \binom{n}{k} (-\mu)^k \frac{\int_{-\infty}^{\infty} xr^{n-k+1}\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r}$$
$$- n\sigma^2\mu_{n-1} - \mu_n\mu \quad (4.18)$$

In (4.18) we have included the term $+2\mu^{n+1}$ to balance out the $-\mu^{n+1}$ in the integrand in the numerator of the first term, leaving a net $+1\mu^{n+1}$. The first term in both (4.17) and (4.18) appears to be related to the $(n+1)^{th}$ central moment. Working from this direction,

we expand the $(n+1)^{th}$ central moment from (4.16) using the binomial theorem:

$$\begin{aligned}\mu_{n+1} &= \frac{\int_{-\infty}^{\infty} x(r-\mu)^{n+1}\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} \\ &= \sum_{k=0}^{n+1}\binom{n+1}{k}(-\mu)^k\,\frac{\int_{-\infty}^{\infty} x r^{n-k+1}\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r}\end{aligned}$$

Separating the first and last terms, and considering first the case where n is odd ($n+1$ is even):

$$= \frac{\int_{-\infty}^{\infty} x(r^{n+1}-\mu^{n+1})\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} + 2\mu^{n+1} + \sum_{k=1}^{n}\binom{n+1}{k}(-\mu)^k\,\frac{\int_{-\infty}^{\infty} x r^{n-k+1}\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r}$$

Here the $+2\mu^{n+1}$ is again needed because the $(-1)^{n+1}$ would have been positive. However, by separating the $k=n$ term from the summation, we find another source of μ^{n+1}

$$\begin{aligned}&= \frac{\int_{-\infty}^{\infty} x(r^{n+1}-\mu^{n+1})\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} + 2\mu^{n+1} - (n+1)\mu^{n+1} \\ &+ \sum_{k=1}^{n-1}\binom{n+1}{k}(-\mu)^k\,\frac{\int_{-\infty}^{\infty} x r^{n-k+1}\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} \\ &= \frac{\int_{-\infty}^{\infty} x(r^{n+1}-\mu^{n+1})\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} - (n-1)\mu^{n+1} + \sum_{k=1}^{n-1}\binom{n+1}{k}(-\mu)^k\,\frac{\int_{-\infty}^{\infty} x r^{n-k+1}\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r}\end{aligned}$$

This may be solved to find

$$\frac{\int_{-\infty}^{\infty} x(r^{n+1}-\mu^{n+1})\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} = \mu_{n+1} + (n-1)\mu^{n+1} - \sum_{k=1}^{n-1}\binom{n+1}{k}(-\mu)^k\frac{\int_{-\infty}^{\infty} x r^{n-k+1}\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} \quad (4.19)$$

if n is odd. Repeating this analysis for n even, we find after separating the first and last terms from the sum

$$\mu_{n+1} = \frac{\int_{-\infty}^{\infty} x(r^{n+1}-\mu^{n+1})\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} + \sum_{k=1}^{n}\binom{n+1}{k}(-\mu)^k\,\frac{\int_{-\infty}^{\infty} x r^{n-k+1}\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r}$$

And again separating $k = n$ we get

$$= \frac{\int_{-\infty}^{\infty} x(r^{n+1} - \mu^{n+1})\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} + (n+1)\mu^{n+1} + \sum_{k=1}^{n-1} \binom{n+1}{k} (-\mu)^k \frac{\int_{-\infty}^{\infty} x r^{n-k+1}\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r}$$

This may be solved to find

$$\frac{\int_{-\infty}^{\infty} x(r^{n+1} - \mu^{n+1})\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} = \mu_{n+1} - (n+1)\mu^{n+1} - \sum_{k=1}^{n-1} \binom{n+1}{k} (-\mu)^k \frac{\int_{-\infty}^{\infty} x r^{n-k+1}\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r} \quad (4.20)$$

To simplify the notation, recall that $\frac{\int_{-\infty}^{\infty} x r^n\,\mathrm{d}r}{\int_{-\infty}^{\infty} x\,\mathrm{d}r}$ is the n^{th} *raw* moment, which we will denote by μ'_n. Using this notation, and plugging (4.19) and (4.20) into (4.17) and (4.18) respectively, we find:

$$\frac{\mathrm{d}\mu_n}{\mathrm{d}t} = \mu_{n+1} + (n-1)\mu^{n+1} - n\sigma^2\mu_{n-1} - \mu_n\mu - \sum_{k=1}^{n-1} \left[\binom{n+1}{k} - \binom{n}{k} \right] (-\mu)^k \mu'_{n-k+1}$$

Reordering some terms and applying Pascal's rule

$$= \mu_{n+1} - \mu_n\mu - n\sigma^2\mu_{n-1} + (n-1)\mu^{n+1} - \sum_{k=1}^{n} \binom{n}{k-1} (-\mu)^k \mu'_{n-k+1}$$

for n odd, and

$$\begin{aligned}\frac{\mathrm{d}\mu_n}{\mathrm{d}t} &= \mu_{n+1} - (n+1)\mu^{n+1} + 2\mu^{n+1} - n\sigma^2\mu_{n-1} - \mu_n\mu \\ &\quad - \sum_{k=1}^{n-1} \left[\binom{n+1}{k} - \binom{n}{k} \right] (-\mu)^k \mu'_{n-k+1} \\ &= \mu_{n+1} - \mu_n\mu - n\sigma^2\mu_{n-1} - (n-1)\mu^{n+1} - \sum_{k=1}^{n-1} \binom{n}{k-1} (-\mu)^k \mu'_{n-k+1}\end{aligned}$$

for n even. Notice that the only difference between the even and odd cases is the sign of the $(n-1)\mu^{n+1}$ term, leading to the conclusion that for all $n \in \mathbb{N}$ (as long as typical

conventions are followed regarding empty sums):

$$\frac{\mathrm{d}\mu_n}{\mathrm{d}t} = \mu_{n+1} - \mu_n\mu - n\sigma^2\mu_{n-1} - (-1)^n(n-1)\mu^{n+1} - \sum_{k=1}^{n-1}\binom{n}{k-1}(-\mu)^k\mu'_{n-k+1} \quad (4.21)$$

To further simplify (4.21) we must first examine the sum

$$-SUM = -\sum_{k=1}^{n-1}\binom{n}{k-1}(-\mu)^k\mu'_{n-k+1}$$

The raw moment μ'_{n-k+1} may be expanded as a sum of central moments(Papoulis, 1984)

$$= -\sum_{k=1}^{n-1}\binom{n}{k-1}(-\mu)^k\sum_{k'=0}^{n-k+1}\binom{n-k+1}{k'}\mu_{k'}\mu^{n-k+1-k'}$$

The μ^k may be distributed to the inner sum

$$= \sum_{k=1}^{n-1}(-1)^{k+1}\binom{n}{k-1}\sum_{k'=0}^{n-k+1}\binom{n-k+1}{k'}\mu_{k'}\mu^{n-k'+1}$$

Consider that this equation may be recast as

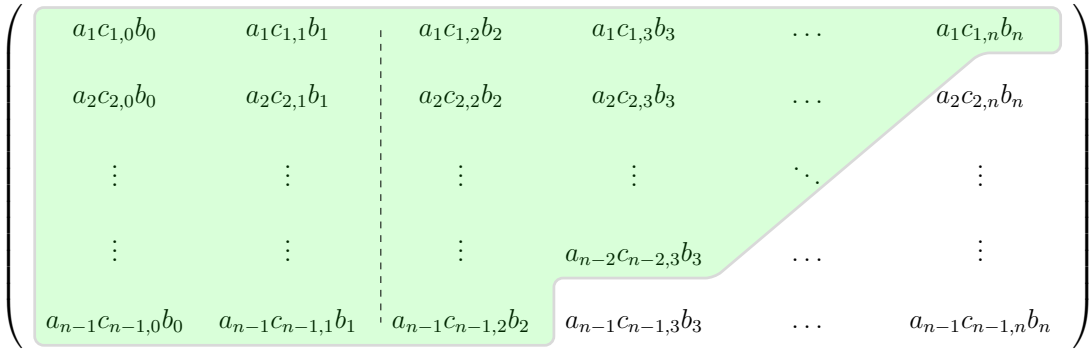
$$-SUM = \sum_{k=1}^{n-1}a_k\sum_{k'=0}^{n-k+1}c_{k,k'}b_{k'} \quad (4.22)$$

where

$$a_k = (-1)^{k+1}\binom{n}{k-1}$$
$$b_{k'} = \mu_{k'}\mu^{n-k'+1}$$
$$c_{k,k'} = \binom{n-k+1}{k'}$$

Let $\mathbf{A}$ be an $(n-1)\times(n+1)$ matrix with each element $A_{i,j} = a_{i+1}c_{i+1,j}b_j$ (the $+1$ in

the index for a and c is arbitrary to keep our indices more relevant to the summation).



The sum in (4.22) is equivalent to summing over the shaded region of the matrix, which we separate into two sections, a rectangular region spanning the first 2 columns, and the remaining triangular region. The value of this is that the rectangular region may be reduced to a non-nested double sum, and the order of the summation in the triangular region may be reversed.

$$-SUM = \sum_{k=1}^{n-1}\sum_{k'=0}^{1} a_k c_{k,k'} b_{k'} + \sum_{k'=2}^{n} b_{k'} \sum_{k=1}^{n-k'+1} c_{k,k'} a_k$$

Substituting in values for a_k, b_k, and $c_{k,k'}$

$$= \sum_{k=1}^{n-1}\sum_{k'=0}^{1} (-1)^{k+1} \binom{n}{k-1} \binom{n-k+1}{k'} \mu_{k'} \mu^{n-k'+1}$$
$$+ \sum_{k'=2}^{n} \mu_{k'} \mu^{n-k'+1} \sum_{k=1}^{n-k'+1} (-1)^{k+1} \binom{n}{k-1} \binom{n-k+1}{k'}$$

The first set of sums may be explicity expanded over k', noting that $\mu_0 = 1$ and $\mu_1 = 0$, while the $k' = n$ term may be separated from the second, nested sum

$$= \mu^{n+1} \sum_{k=1}^{n-1} (-1)^{k+1} \binom{n}{k-1} + \mu_n \mu + \sum_{k'=2}^{n-1} \mu_{k'} \mu^{n-k'+1} \sum_{k=1}^{n-k'+1} (-1)^{k+1} \binom{n}{k-1} \binom{n-k+1}{k'}$$

It turns out the inner sum of the nested sum is 0; to see this we first shift the index from $k \to k-1$

$$= \mu^{n+1} \sum_{k=1}^{n-1} (-1)^{k+1} \binom{n}{k-1} + \mu_n \mu + \sum_{k'=2}^{n-1} \mu_{k'} \mu^{n-k'+1} \sum_{k=0}^{n-k'} (-1)^{k} \binom{n}{k} \binom{n-k}{k'} \tag{4.23}$$

Consider the inner sum from $k = 0$ to $n - k'$:

$$\begin{aligned}
\sum_{k=0}^{n-k'} (-1)^k \binom{n}{k} \binom{n-k}{k'} &= \sum_{k=0}^{n-k'} (-1)^k \frac{n!}{k!(n-k)!} \cdot \frac{(n-k)!}{k'!(n-k-k')!} \\
&= \frac{n!}{k'!} \sum_{k=0}^{n-k'} (-1)^k \frac{1}{k!(n-k-k')!} \\
&= \frac{n!}{k'!(n-k')!} \sum_{k=0}^{n-k'} (-1)^k \frac{(n-k')!}{k!(n-k-k')!} \\
&= \binom{n}{k'} \sum_{k=0}^{n-k'} (-1)^k \binom{n-k'}{k} \\
&= \binom{n}{k'} \sum_{k=0}^{n'} (-1)^k \binom{n'}{k}
\end{aligned}$$

From the properties of Pascal's triangle this last sum is 0 for $k' < n$, thus the nested sum in the last term of (4.23) is 0. Thus the sum from (4.21) is

$$-SUM = \mu^{n+1} \sum_{k=1}^{n-1} (-1)^{k+1} \binom{n}{k-1} + \mu_n \mu$$

Plugging this back into (4.21) we get

$$\frac{\mathrm{d}\mu_n}{\mathrm{d}t} = \mu_{n+1} - \mu_n\mu - n\sigma^2\mu_{n-1} - (-1)^n(n-1)\mu^{n+1} + \mu^{n+1}\sum_{k=1}^{n-1}(-1)^{k+1}\binom{n}{k-1} + \mu_n\mu$$

Here we may cancel out the $\mu_n\mu$ terms, and group the μ^{n+1} terms together

$$= \mu_{n+1} - n\sigma^2\mu_{n-1} + \mu^{n+1}\left[-(-1)^n(n-1) + \sum_{k=1}^{n-1}(-1)^{k+1}\binom{n}{k-1}\right]$$

We may shift the summation indices from $k \to k-1$ as before

$$= \mu_{n+1} - n\sigma^2\mu_{n-1} + \mu^{n+1}\left[-(-1)^n(n-1) + \sum_{k=0}^{n-2}(-1)^{k}\binom{n}{k}\right]$$

The summation may be completed all the way to $k = n$ if we then explicity subtract both missing terms

$$= \mu_{n+1} - n\sigma^2\mu_{n-1} + \mu^{n+1}\left[-(-1)^n(n-1) + \sum_{k=0}^{n}(-1)^{k}\binom{n}{k} - (-1)^{n-1}n - (-1)^n\right]$$

As before, this summation now vanishes, leaving

$$= \mu_{n+1} - n\sigma^2\mu_{n-1} + \mu^{n+1}\left[-(-1)^n(n-1) + (-1)^n n - (-1)^n\right]$$

Factoring out $(-1)^n$

$$= \mu_{n+1} - n\sigma^2\mu_{n-1} + \mu^{n+1}\left[(-1)^n(-n+1+n-1)\right]$$

Thus the entire last term vanishes. We have now reduced (4.21) to its simplest form

$$\frac{\mathrm{d}\mu_n}{\mathrm{d}t} = \mu_{n+1} - n\sigma^2\mu_{n-1} \tag{4.24}$$

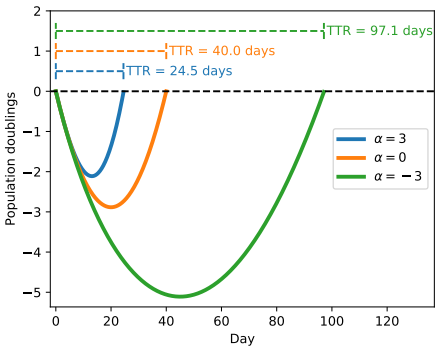


Figure 4.6: Predictions of the TTR for 3 distributions, spanning negative skew ($\alpha = -3$), normal ($\alpha = 0$), and positive skew ($\alpha = 3$). As expected, the negative skew distribution has significantly slower relapse due to the skew decreasing the variance, while the positive skew distribution is accelerated.

This formula describes how higher order central moments ultimately influence the variance, and equations (4.7) and (4.8) describe how the variance then contributes to the growth of the population.

4.4.4 Steady state solution

Together with equation (4.8), (4.24) defines how the distribution of proliferation rates changes over time. Due to (4.8), the only stationary distributions of (4.1) are Dirac delta distributions, with the location as a free parameter. However, if we relax the criteria that the mean be constant (i.e. we permit distributions with positive variance, corresponding to an increasing mean) we find one more steady state solution

$$\mu_{n+1} = \begin{cases} n\sigma^2\mu_{n-1} & : n \geq 1 \text{ is odd} \\ 0 & : n \geq 2 \text{ is even} \end{cases}$$

The moments of a normal distribution satisfy this constraint(Siegrist, 2014), and thus we have shown (as hinted above) that the shape of a normal distribution remains constant, while the distribution itself moves in the positive direction with velocity equal to the variance.

4.5 DIP rate distribution parameters can predict time-to-relapse (TTR)

In cancer, we are predominately interested in DIP rate distributions with a negative mean, but enough variance to drive the mean to be positive. In this case, from (4.7), we can define the time-to-relapse, T, which must satisfy

$$\int_0^T \mu(t)\,\mathrm{d}t = 0 \tag{4.25}$$

While they dynamics of $\mu(t)$ are not known for arbitrary DIP rate distributions, we may analyze the behavior of T for the normal and skew-normal cases. For a normal distribution, $\mu(t) = \mu(0) + \sigma^2 t$, such that the TTR satisfies

$$\mu(0)T + \frac{\sigma^2 T^2}{2} = 0$$

Ignoring the trivial solution of $T = 0$, we find for a normal distribution

$$T = -\frac{2\mu(0)}{\sigma^2} \tag{4.26}$$

For a skew-normal distribution, we recall that the total population is equivalent to the MGF scaled by the initial population (4.4). The MGF of the skew-normal distribution is given by

$$MGF(t) = e^{\xi t + \frac{\omega^2 t^2}{2}} \left(1 + erf\left(\frac{\omega\delta t}{\sqrt{2}}\right)\right) \tag{4.27}$$

where $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$, the scale parameter $\omega = \frac{\sigma\pi}{\sqrt{\pi^2 - 2\delta^2\pi}}$, the location parameter $\xi = \mu - \omega\delta\sqrt{\frac{2}{\pi}}$, and α is the skew parameter(Azzalini, 2005).

Thus we find that the TTR for a skew-normal distribution must satisfy

$$1 + erf\left(\frac{\omega\delta T}{\sqrt{2}}\right) = e^{-\left(\xi T + \frac{\omega^2 T^2}{2}\right)} \tag{4.28}$$

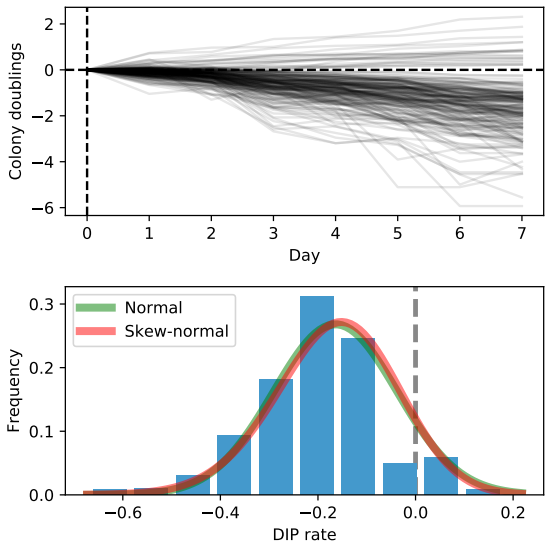


Figure 4.7: Measurement of DIP rate distribution in PC9 treated with erlotinib. Top: PC9 cells were treated with $3\mu M$ erlotinib and displayed exponential growth over a 7 day period. Bottom: DIP rates (day^{-1}), showing fits for normal ($\mu = -0.164$, $\sigma = 0.126$ and skew-normal ($\alpha = -1.16$, $\xi = -0.07$, $\omega = 0.158$) distributions.

As expected, a negatively skewed DIP rate distribution has a longer TTR because the skew decreases the variance, while a positively skewed distribution has a shorter TTR because the skew increases the variance (Figure 4.6).

4.5.1 HGM predicts relapse of PC9 cells treated with erlotinib

A DIP rate distribution for PC9 cells in $3\mu M$ erlotinib was measured using the clonal fractional proliferation method from Frick *et. al.* (Frick et al., 2015). Cells were treated with erlotinib for 72 hours prior to time t=0, and imaged once per day starting at 72 hours. Over a 1 week period, each colony of cells displayed exponential growth, characterized by straight lines in Figure 4.7. A linear model was fit to each colony's trajectory, where the slope represents that colony's DIP rate. The resulting DIP rate distribution, as well as normal and skew-normal fits is shown in Figure 4.7.

From these estimates of the normal distribution fit, the HGM would predict (4.26) a TTR of about 21 days, while the skew-normal prediction (4.28) is about 24 days. Further-

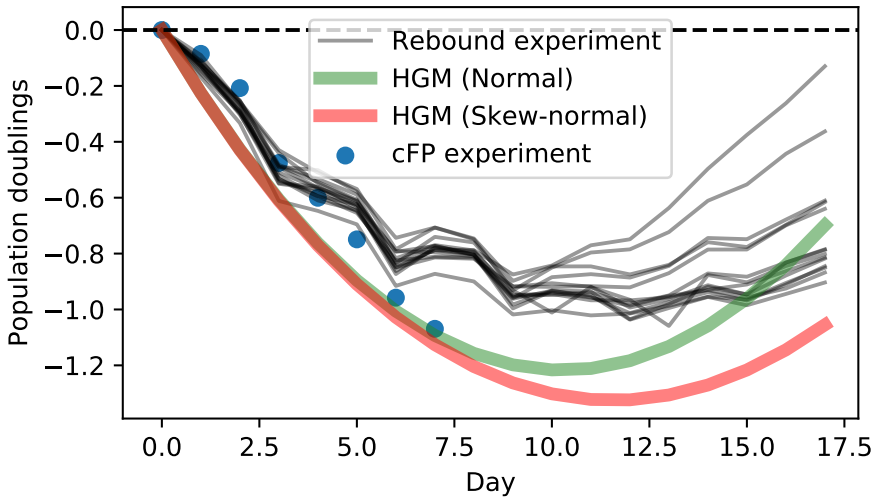


Figure 4.8: PC9 rebound matches HGM. Heterogeneous PC9 population level response to treatment with $3\mu M$ erlotinib shows response dynamics that are closely predicted by the cFP derived HGM. There is some slight deviation between the cFP and rebound response on days 6 and 7, that may be responsible for the difference in depth of response.

more, the normal distribution prediction suggests that the PC9 population response should reach its minimum point at about 10 days. PC9 population rebound was tracked for 17 days over 14 replicates (Figure 4.8, with a rebound closely following the HGM prediction. This is true even though the data collected through the clonal fractional proliferation assay (Figure 4.7) only went out 7 days, and had not shown signs of rebound (Figure 4.8).

4.5.2 Parameter Sensitivity

The goal of cancer treatment is to push T back as far as possible, ideally beyond the expected lifetime of a patient. Here we investigate sensitivity of TTR to different parameters of normal and skew-normal DIP rate distributions.

4.5.2.1 Normal distribution

For a normal distribution, the relative sensitivity of T (equation (4.26) to the initial mean, μ (for brevity we have dropped the argument of $\mu(0)$, but it should be noted that this

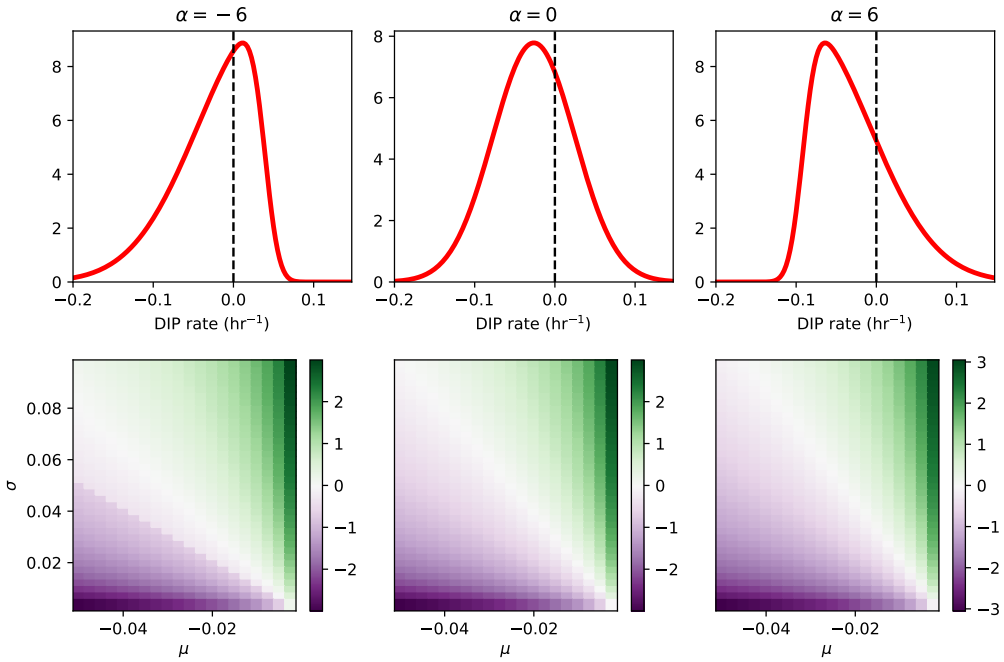


Figure 4.9: Sensitivity analysis of TTR. The relative sensitivity of TTR to μ and σ is shown for a negative skew (left), normal (middle), and positive skew (right) distribution. Top: Example skewnormal distributions, showing the effect of the shape parameter α. Bottom: Relative sensitivity of TTR to varying μ VS varying σ. Color is $ln\left[\frac{sens_\mu}{sens_\sigma}\right]$. Green regions reflect greater sensitivity to μ, while purple regions reflect greater sensitivity to σ. The influence of α on the relative sensitivity reveals that negative skew increases the relative importance of μ, while positive skew increases the relative sensitivity of σ.

is the initial proliferation rate mean) and the standard deviation σ, may be determined by comparing the partial derivatives with respect to each variable.

$$\frac{\partial T}{\partial \mu} = -\frac{2}{\sigma^2}$$
$$\frac{\partial T}{\partial \sigma} = \frac{4\mu}{\sigma^3}$$

The ratio of the magnitude of the effects between them is given by

$$R(\mu,\sigma) = \left|\frac{\partial T}{\partial \sigma}\right| : \left|\frac{\partial T}{\partial \mu}\right| = \frac{2|\mu|}{\sigma} \tag{4.29}$$

Thus as long as $\sigma < 2|\mu|$, TTR is predicted to be more sensitive to changes in σ. On the other hand, if $\sigma > 2|\mu|$, TTR is predicted to be more sensitive to changes in μ.

4.5.3 Skew-normal distribution

It is impossible to analytically solve (4.28) for for T, but it may be numerically explored. Simulation over a grid with $\mu \in [-1, -0.1]$, $\sigma \in [0.1, 2)$, and $\alpha \in [-3, 3]$ with grid points sampled with regular step lengths of 0.1 revealed no points with maximum sensitivity to α. Figure 4.9 shows the results of simulation at three values of α, with $0.01s^{-1}$ resolution in μ and σ, and reveals that skew has very little impact on where TTR is more sensitive to either μ or σ.

4.6 Discussion

While heterogeneity is ubiquitously understood to be a major driver of relapse in drug-treated tumors, we have here introduced a model, HGM, that explicitly quantifies the impact of the variance of clonal subpopulation fitness. Due to the stability of the HGM solution for DIP rates following a normal distribution, we conclude that even for approximately normal DIP rate distributions, the predictions of the HGM should be close to the observed population level response. Nevertheless, we were able to show that positive (negative) skewness can accelerate (decelerate) the rate of relapse by driving an increase (decreases) in the DIP rate variance.

We propose that treatments with a second agent that lead to decreased mean or decreased variance of the DIP rate distribution can lead to an extension of the TTR. Furthermore, depending on the exact shape of the DIP rate distribution, we are able to identify which parameter is likely to have a greater impact on the TTR. Nevertheless, further work remains to understand cases in which assumption 1 may be invalid, such as in the recent work by Paudel *et. al.* (Paudel et al., 2018).

Chapter 5

MuSyC: A new framework for quantifying synergy of drug combination response surfaces unifies the landscape of synergy metrics by decoupling synergy of potency and efficacy

5.1 Introduction

Recent decades have witnessed an exponential expansion of available drugs for the treatment of disease (Inglese et al., 2007; Janzen, 2014; Mott et al., 2015; Gong et al., 2017). This expansion has been concomitant with an evolving understanding of disease complexity (Hanahan and Weinberg, 2011; Mitchell, 2012); complexity commonly necessitating combination therapy (He et al., 2016). However, the clinical use of combination therapy is often limited by tolerable dose ranges, and therefore, it is desirable to identify combinations that enable dose reduction (Tallarida, 2011), i.e., synergistic potency. Additionally, combining drugs does not guarantee *a priori* an increase in efficacy over the single agents, and therefore it is desirable to identify combinations with effects greater than achievable with either drug alone (Foucquier and Guedj, 2015), i.e., synergistic efficacy. To assess a combination's performance, several drug synergy metrics have been proposed (Foucquier and Guedj, 2015). The roots of current synergy metrics can be traced back to either Loewe, who advanced the Dose Additivity Principle (Loewe, 1926) or Bliss who first described the Multiplicative Survival Principle (BLISS, 1939). Nearly a century later, methods to quantify drug synergy continue to appear (Chou and Talalay, 1984; Greco et al., 1992; Yadav et al., 2015; Twarog et al., 2016; Zimmer et al., 2016; Schindler, 2017). However, none of these methods distinguish between synergistic potency and synergistic efficacy. Instead, they either make no distinction (Loewe, 1926; BLISS, 1939; Chou and Talalay, 1984; Twarog et al., 2016) or assume the only form of synergism is through potency (Yadav et al., 2015; Zimmer et al., 2016). Nevertheless, this distinction is essential to arrive at an unambiguous definition of synergy and properly rationalize the deployment

of drug combinations, e.g. in personalized medicine. Indeed, conflating them may mislead drug combination discovery efforts. For instance, a search for improved efficacy based on traditional synergy frameworks may be confounded by their inability to sort out synergistically potent combinations.

To address this critical shortcoming and resolve these two independent types of synergy, herein we propose a new synergy framework termed Multi-dimensional Synergy of Combinations (MuSyC), which is based on a two-dimensional (2D) extension of the Hill equation. MuSyC distinguishes between synergistic potency and synergistic efficacy based on parameters arising from the 2D Hill equation. These parameters can be directly related to the shape of sigmoidal dose-response curves, and therefore, to standard pharmacologic measures of potency and efficacy. We visualize synergy of potency and efficacy on Drug Synergy Diagrams (DSDs) which stratify drug combinations along orthogonal axes of synergy facilitating rapid comparisons between different combinations.

In direct comparisons, we found that traditional major synergy frameworks emerge as special cases of MuSyC, and that when certain conditions are not met, these methods introduce biases and/or ambiguous conclusions.

To demonstrate the value of MuSyC, we investigate a panel of anti-cancer compounds in combination with osimertinib, a third-generation mutant-EGFR inhibitor, in EGFR-mutant non-small cell lung cancer (NSCLC). We find that drugs targeting epigenetic regulators and microtubule polymerization are synergistically efficacious with osimertinib. In contrast, drugs co-targeting kinases in the MAPK pathway affected potency, not the efficacy of osimertinib. These conclusions have implications for drug combination deployment in NSCLC where increasing the efficacy of EGFR-inhibitors has historically relied on trial and error with no overarching principles to guide development (Tang et al., 2013; Schiffmann et al., 2016).

We also apply MuSyC to study the clinically-relevant combination targeting RAF and MEK in BRAF-mutant melanoma (Long et al., 2014). Overwhelmingly, we find this com-

bination to be synergistically efficacious, though in several cases at the cost of antagonistic potency. We then identify NOX5 as an unsuspected molecular co-target for mutant-BRAF melanomas and show the synergistic efficacy of BRAF and NOX5 inhibitors is proportional to NOX5 expression.

5.2 Generalized multi-dimensional Hill equation describes orthogonal synergies of potency, efficacy, and cooperativity

5.2.1 One-dimensional Hill equation

In pharmacology, the effect of a drug is usually described by the Hill equation, which arises from the equilibrium of a reversible transformation between an unaffected population fraction (U) and an affected population fraction (A)

$$U + h \cdot d \underset{r_{-1}}{\overset{r_1}{\rightleftharpoons}} A \tag{5.1}$$

Here, d is the concentration of the drug, h is the Hill slope, often called cooperativity, and r_1 and r_{-1} are constants corresponding to its rate of action. This reaction is represented in Figure 5.1A as transitions (given by edges) between states (given by nodes). It has been shown that it is reasonable to assume that populations U and A reach equilibrium much faster than they proliferate on their own (Harris et al., 2016), such that we may solve for the equilibrium as

$$\frac{\mathrm{d}U}{\mathrm{d}t} = A \cdot r_{-1} - U \cdot r_1 d^h \equiv 0$$

$$\frac{A}{U} = \frac{r_1 d^h}{r_{-1}}$$

When $d^h = \frac{r_{-1}}{r_1}$, then half the population is affected, and half is unaffected ($A = U$). This dose is called the EC50, denoted as C, such that $C^h = \frac{r_{-1}}{r_1}$. Because 100% of the population

is either unaffected or affected, we also have

$$U + A = 1$$

which leads to the 1D Hill equation

$$U = \frac{C^h}{C^h + d^h} \tag{5.2}$$

If the unaffected and affected populations differ phenotypically by some effect (such as proliferation rate), the average observed effect over the whole population at dose d of some drug will be a weighted average of the two effects by the percent affected and unaffected. Namely,

$$E_d = U \cdot E_0 + A \cdot E_m,$$

where E_0 is the effect characteristic of the unaffected population, and E_m is the effect characteristic of the affected population. From this we find the final form of a 4-parameter sigmoidal equation describing dose-response due to Hill-kinetics:

$$\frac{E - E_m}{E_0 - E_m} = \frac{C^h}{C^h + d^h} \tag{5.3}$$

where E is the observed effect at dose d. The parameters of equation 5.3 correspond to efficacy ($E_0 - E_m$), potency C, and the Hill coefficient, which is often called cooperativity h.

5.2.2 Two-dimensional Hill equation

To extend the 1D Hill equation to two drugs, we consider four possible states, U, A_1, A_2, and $A_{1,2}$ reflecting cell populations that are unaffected, affected by drug 1 alone, affected by drug 2 alone, or affected by both drugs 1 and 2, respectively. We model this system

using four distinct reactions:

$$
\begin{aligned}
U + h_1 \cdot d_1 &\underset{r_{-1}}{\overset{r_1}{\rightleftharpoons}} A_1 \\
U + h_2 \cdot d_2 &\underset{r_{-2}}{\overset{r_2}{\rightleftharpoons}} A_2 \\
A_1 + (\gamma_1 h_2) \cdot d_2 &\underset{r_{-2}}{\overset{\alpha_1^{h_2} r_2}{\rightleftharpoons}} A_{1,2} \\
A_2 + (\gamma_2 h_1) \cdot d_1 &\underset{r_{-1}}{\overset{\alpha_2^{h_1} r_1}{\rightleftharpoons}} A_{1,2}
\end{aligned}
\tag{5.4}
$$

Here, α_1 and α_2 reflect the modulation of drug 1 and drug 2 on cells' affinity for drug 2 and drug 1, respectively.

This leads to the following underdetermined system of differential equations

$$
\begin{aligned}
\frac{\mathrm{d}U}{\mathrm{d}t} &= -U \cdot \left(r_1 d_1^{h_1} + r_2 d_2^{h_2} \right) + A_1 \cdot r_{-1} + A_2 \cdot r_{-2} \\
\frac{\mathrm{d}A_1}{\mathrm{d}t} &= -A_1 \cdot \left(r_{-1} + r_2 (\alpha_1 d_2)^{\gamma_1 h_2} \right) + U \cdot r_1 d_1^{h_1} + A_{1,2} \cdot r_{-2} \\
\frac{\mathrm{d}A_2}{\mathrm{d}t} &= -A_2 \cdot \left(r_1 (\alpha_2 d_1)^{\gamma_2 h_1} + r_{-2} \right) + U \cdot r_2 d_2^{h_2} + A_{1,2} \cdot r_{-1} \\
\frac{\mathrm{d}A_{1,2}}{\mathrm{d}t} &= -A_{1,2} \cdot (r_{-1} + r_{-2}) + A_1 \cdot \alpha_1 r_2 d_2^{h_2} + A_2 \cdot r_1 (\alpha_2 d_1)^{h_1}
\end{aligned}
\tag{5.5}
$$

Finally, as in the 1D case we have

$$
U + A_1 + A_2 + A_{1,2} = 1. \tag{5.6}
$$

At equilibrium, the equations 5.5 must be equal to zero; however, the system only defines a rank 3 matrix. Taking the first three equations from 5.5 with 5.6, we define

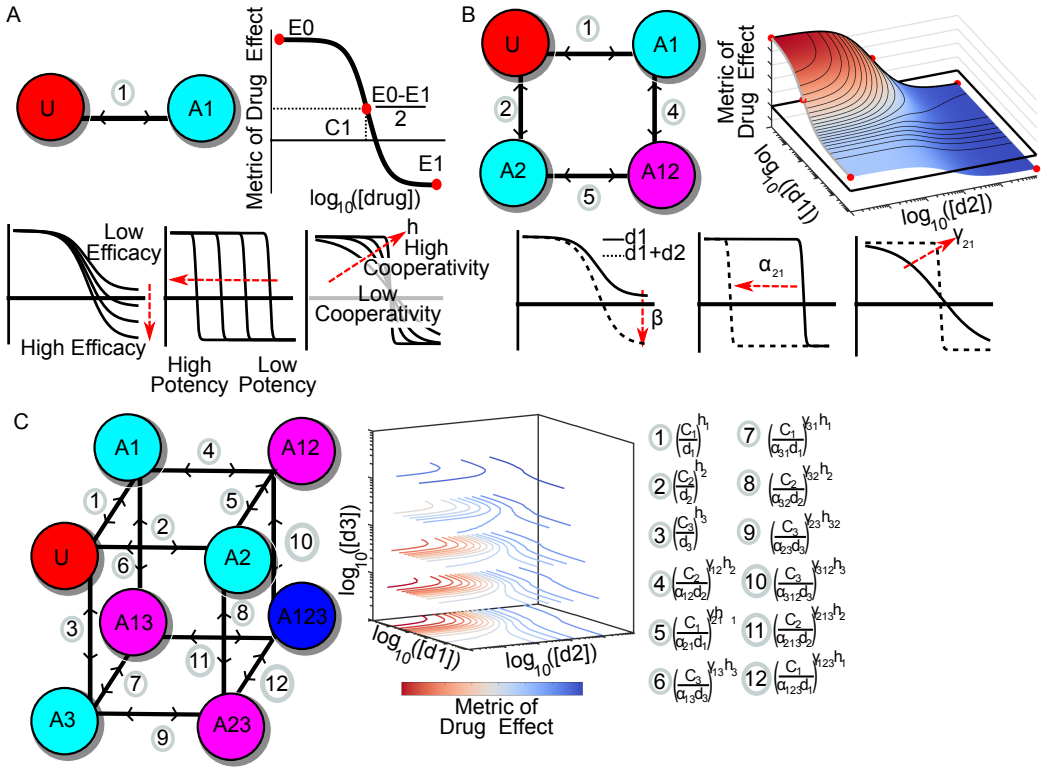


Figure 5.1: Mass action kinetics based, multi-dimensional model of drug effect. A) The commonly observed sigmoidal dose dependent response of a system to a single drug is classically fit to a generalized Hill equation which stems from a two state model of drug action. Three parameters of the Hill equation denote the compound's efficacy (E_1), potency (C_1), and cooperativity (h). B) Two dimensional extension for two drug case to four-state model. Correspondingly, the fit is now a Hill surface with parameters for synergistic efficacy (β), synergistic potency (α)m and synergistic cooperativity (γ). β is the percent increase in effect of the combination beyond most efficacious single agent. α is the multiplicative change in the effective dose of one drug given saturating doses of the other. γ is the multiplicative change in the cooperativity (i.e. switch-like behavior) of one drug with saturating doses of the other. C) Three drug extension to hyper-cube and corresponding hyper-dose-response surface. Representations depicted have metrics of effect for which $E_0 > E_3$; however, MuSyC can also be applied if the drug increases the effect $E_0 < E_3$ by changing the sign on β.

$$\mathbf{M} := \begin{bmatrix} -\left(r_1 d_1^{h_1} + r_2 d_2^{h_2}\right) & r_{-1} & r_{-2} & 0 \\ r_1 d_1^{h_1} & -\left(r_{-1} + r_2(\alpha_1 d_2)^{h_2}\right) & 0 & r_{-2} \\ r_2 d_2^{h_2} & 0 & -\left(r_1(\alpha_2 d_1)^{h_1} + r_{-2}\right) & r_{-1} \\ 1 & 1 & 1 & 1 \end{bmatrix} \tag{5.7}$$

such that

$$\mathbf{M} \cdot \begin{bmatrix} U & A_1 & A_2 & A_{1,2} \end{bmatrix}^\intercal = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^\intercal$$

or

$$\begin{bmatrix} U & A_1 & A_2 & A_{1,2} \end{bmatrix}^\intercal = \mathbf{M}^{-1} \cdot \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^\intercal \tag{5.8}$$

If we once again consider distinct effects E_0, E_1, E_2, and E_3 distinguishing populations U, A_1, A_2, and $A_{1,2}$, we find

$$E = \begin{bmatrix} E_0 & E_1 & E_2 & E_3 \end{bmatrix} \cdot \mathbf{M}^{-1} \cdot \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^\intercal \tag{5.9}$$

Equation 5.9 has the following twelve explicit parameters: r_1, r_{-1}, r_2, r_{-2}, E_0, E_1, E_2, E_3, h_1, h_2, α_1 and α_2. As before we define EC50, C, such that $C_1^{h_1} = \frac{r_{-1}}{r_1}$, and similarly for C_2. C_1 and C_2 describe the EC50s of the first two reactions in 5.4. However, the EC50s for the third and fourth reactions in 5.4 are $C_1^{effective} = \frac{C_1}{\alpha_2}$, and likewise for $C_2^{effective}$. Thus, the α parameters can be thought of as multiplicative factors for the EC50s. Likewise, the Hill slopes of the third and fourth reactions are multiplied by the γ parameters.

We therefore interpret α_1 and α_2 as parameters for potency synergy, and γ_1 and γ_2 as parameters for cooperativity synergy. Likewise, we define

$$\beta := \frac{min(E_1, E_2) - E_3}{E_0 - min(E_1, E_2)} \tag{5.10}$$

(Note: This form of β is used when the drugs cause a *decrease* in E. If drugs *increase* E,

$max(E_1, E_2)$ is used instead. Thus, β quantifies the percent increase or decrease in efficacy of the combination of drugs, relative to the more efficacious single agent.

Synergistic potency (α) quantifies the extent to which one drug modulates the "effective" dose of the other. In general, each drug can independently modify the effective dose of the other therefore, for two drugs, there are two αs. Finally, synergistic cooperativity (γ) corresponds to the amount the hill coefficient of one compound is altered by the other. As with α, γ is asymmetric. These three types of modifications are depicted in Figure 5.1B. From equation 5.9 a 'null' dose-response surface can be generated where there is no synergy ($\beta = 0, \alpha_1 = \alpha_2 = 1, \gamma_1 = \gamma_2 = 1$) (Figure 5.1B).

This framework is easily extended to multiple drugs following a cubic geometry as depicted in Figure 5.1C. This geometry results in a hyper-dose response surface depicted in Figure 5.1C and can be conceptualized as a stacking of dose response surfaces along the third dimension. Due to MuSyC's geometric foundation, it is possible to continue to extend MuSyC to increasing number of drugs, at the cost of an increasing number of synergy parameters.

There is a corresponding increase in the number of parameters with α and γ scaling by $n * 2^{n-1} - n$ (the number of edges for a hyper-cube minus the edges connecting unaffected and singly affected states). This and the commensurate data necessary to fully constrain these hyper-surfaces invokes a parameter identifiability problem. Because efficacy and potency are the most important for cancer therapeutics, we generally impose that $\gamma_1 = \gamma_2 = 1$, and in our own data this approximation does not lead to poor dose-response fits. Nevertheless, in other systems cooperativity synergy may be of greater importance and interest.

5.3 MuSyC unifies dominant synergy frameworks

Several other methods for calculating synergy exist, including long-standing traditional methods Loewe (Loewe, 1927), Bliss (BLISS, 1939), and CI (Chou and Talalay, 1984),

as well as more recent methods such as ZIP (Yadav et al., 2015), BRAID (Twarog et al., 2016), and the effective dose model (Zimmer et al., 2016). All of these methods, as well as MuSyC, define a null surface that quantifies the expected effect of a combination. Combinations with effects greater than or less than expected surface are deemed synergistic or antagonistic respectively. These methods broadly use one of two approaches to quantify synergy. Loewe, Bliss, CI, and ZIP quantify synergy at every concentration based on how the experimentally measured response relates to the null surface. Conversely, BRAID, the effective dose model, and MuSyC provide equations describing the entire surface, containing synergy parameters which are fit to experimental data using non-linear optimization techniques.

Here, we briefly compare our model to each of these others and show that our model (1) is able to describe combination surfaces that other methods cannot, (2) results in synergy parameters which are straightforward to interpret, (3) is not restricted to a special class of effects with bounded scales, and (4) reduces to the most common approaches (Loewe, CI, and Bliss) in special cases.

5.3.1 The Dose Equivalence Principle: Loewe and CI

The first prevalent foundational principle, established by Loewe (Loewe, 1926), and subsequently expanded on by CI(Chou and Talalay, 1984), is the dose equivalence principle. This states that for a given effect magnitude E, such that dose d_1 of drug 1 alone, or dose d_2 of drug 2 alone achieves that effect, then there is a constant ratio $R = \frac{d_1}{d_2}$ such that using b less of drug 2 can always be compensated for by using $a = Rb$ more of drug 1. One limitation of this definition is that the null surface is only defined for combinations whose magnitude of effect is less than the *weaker* drug's maximal effect. This is because beyond such concentrations, no amount a of the weaker drug can compensate for reducing the dose of the stronger drug by b.

An "isobole" is a curve comprised of concentrations d_1 and d_2 of drugs 1 and 2 that all

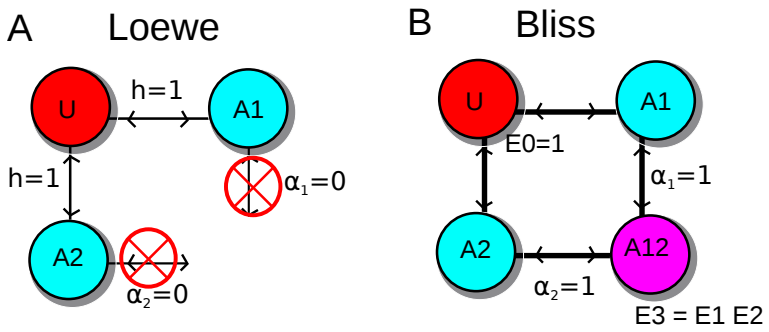


Figure 5.2: Loewe additivity and Bliss independence emerge as special cases of MuSyC. (A) As has previously been reported, Loewe additivity corresponds to mutually exclusive drugs (Chou and Talalay, 1984), such that state $A_{1,2}$ does not exist. This is achieved in MuSyC by setting $\alpha_1 = \alpha_2 = 0$, and also require $h_1 = h_2 = 0$. (B) The Bliss independence null model can be matched by MuSyC by setting $\alpha_1 = \alpha_2 = 0$, and $E_3 = E_1 E_2$. The first constraint is equivalent to the MuSyC null model, while the second simply describes a shift in the expected effect E_3. Exempting that, the Bliss and MuSyC null models are equivalent.

achieve the same effect (Greco et al., 1995). The null surface defined by Loewe additivity is characterized by linear isoboles. MuSyC's null surface shares this property when the two drugs are maximally antagonistic with respect to potency and $h_1 = h_2 = 1$ (Figure 5.2). This can be seen by setting $\alpha_1 = \alpha_2 = 0$, in which case equation 5.9 reduces to

$$(E - E_0) + (E - E_1)\left(\frac{d_1}{C_1}\right)^{h_1} + (E - E_2)\left(\frac{d_2}{C_2}\right)^{h_2} = 0 \tag{5.11}$$

By this it is easy to see when $h_1 = h_2 = 1$, iso-effect lines are represented by the linear isoboles characteristic of Loewe Additivity and the CI null models. However, even in this case MuSyC is not limited by the weaker drug, and can therefore extend Loewe's isoboles to any combination doses.

5.3.2 The Multiplicative Survival Principle: Bliss and Effective Dose

The other prevalent foundational synergy principle is multiplicative survival, described by Bliss (BLISS, 1939). Bliss defines a null model by assuming the probability of a cell being unaffected by drug 1 (U_1) is independent of the probability of a cell being unaffected

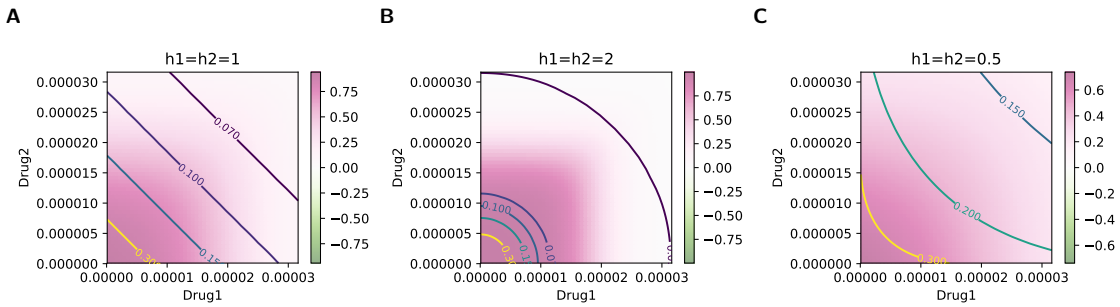


Figure 5.3: Isoboles show linear or nonlinear shapes depending on h. (A) As in Figure 5.2, here we show that when $h = 1$ and $\alpha = 0$, MuSyC results in linear isoboles, which are defined as contours of equivalent effect. However in (B) and (C), we show that MuSyC predicts nonlinear isoboles when $h \neq 1$.

by drug 2 (U_2). From this, the null surface states the probability of being unaffected by both drug 1 and drug 2 in combination is $U = U_1 \cdot U_2$ (BLISS, 1939). The MuSyC model leads Bliss null surfaces when there is no potency synergy ($\alpha_1 = \alpha_2 = 1$), and $E_3 = E_1 E_2$ (Figure 5.2).

To show this, we first assume that each drug in isolation has a 1D hill response

$$\overline{U_i} = \frac{1}{1 + \left(\frac{d_i}{C_i}\right)^{h_i}} \tag{5.12}$$

where $\overline{U_i}$ reflects the portion of cells unaffected by drug i alone. For the 2D case, when $\alpha_1 = \alpha_2 = 1$, each edge in Figure 5.1B satisfies detailed balance

$$A_1 = \left(\frac{d_1}{C_1}\right)^{h_1} U$$
$$A_2 = \left(\frac{d_2}{C_2}\right)^{h_2} U$$
$$A_{1,2} = \left(\frac{d_2}{C_2}\right)^{h_2} A_1$$

Including equation 5.6, we find that for a combination of two drugs

$$
\begin{aligned}
U &= \frac{1}{1+\left(\frac{d_1}{C_1}\right)^{h_1}+\left(\frac{d_2}{C_2}\right)^{h_2}+\left(\frac{d_1}{C_1}\right)^{h_1}\left(\frac{d_2}{C_2}\right)^{h_2}} \\
A_1 &= \frac{\left(\frac{d_1}{C_1}\right)^{h_1}}{1+\left(\frac{d_1}{C_1}\right)^{h_1}+\left(\frac{d_2}{C_2}\right)^{h_2}+\left(\frac{d_1}{C_1}\right)^{h_1}\left(\frac{d_2}{C_2}\right)^{h_2}} \\
A_2 &= \frac{\left(\frac{d_2}{C_2}\right)^{h_2}}{1+\left(\frac{d_1}{C_1}\right)^{h_1}+\left(\frac{d_2}{C_2}\right)^{h_2}+\left(\frac{d_1}{C_1}\right)^{h_1}\left(\frac{d_2}{C_2}\right)^{h_2}} \\
A_{1,2} &= \frac{\left(\frac{d_1}{C_1}\right)^{h_1}\left(\frac{d_2}{C_2}\right)^{h_2}}{1+\left(\frac{d_1}{C_1}\right)^{h_1}+\left(\frac{d_2}{C_2}\right)^{h_2}+\left(\frac{d_1}{C_1}\right)^{h_1}\left(\frac{d_2}{C_2}\right)^{h_2}}
\end{aligned}
\tag{5.13}
$$

From this, it is easy to verify that $U = U_1 \cdot U_2$, which is equivalent to the Bliss Independence null model. Furthermore, given $E_0 = 1$

$$E = U + A_1E_1 + A_2E_2 + A_{1,2}E_3$$

We define $\overline{U_i}$, $\overline{A_i}$, and $\overline{E_i} = \overline{U_i} + \overline{A_i}E_i$ to be the fraction of unaffected cells, fraction of affected cells, and observed effect for treatment with the single drug i, as described by equation 5.12. Then

$$
\begin{aligned}
\overline{E_1} \cdot \overline{E_2} &= \left[\overline{U_1} + (1-\overline{U_1}E_1\right]\left[\overline{U_2} + (1-\overline{U_2}E_2)\right] \\
&= \overline{U_1U_2} + E_1(\overline{U_2} - \overline{U_1U_2}) + E_2(\overline{U_1} - \overline{U_1U_2}) + E_1E_2(1-\overline{U_1})(1-\overline{U_2})
\end{aligned}
$$

From 5.13, we know $U = \overline{U_1} \cdot \overline{U_2}$, and $A_{1,2} = \overline{A_1} \cdot \overline{A_2}$, leading to

$$= U + E_1(\overline{U_2} - U) + E_2(\overline{U_1} - U) + E_1E_2A_{1,2}$$

Similarly, it is simple to show $A_1 = \overline{U_2} - U$, and similarly for A_2

$$= U + E_1A_1 + E_2A_2 + E_1E_2A_{1,2}$$

Therefore, given $\alpha_1 = \alpha_2 = 1$, $E_0 = 1$, and $E_3 = E_1E_2$, then $\overline{E_1} \cdot \overline{E_2} = E$. Thus, while Bliss was derived purely within the scope of "percent affected", MuSyC shows that the Bliss model may be appropriately extended to any measure of effect for which $E_0 = 1$ and effects are expected to be multiplicative. Nevertheless, for effects which do not satisfy these criteria, the Bliss model is not appropriate, while MuSyC may still be used for arbitrary effects.

5.3.3 Bliss and Loewe cannot distinguish synergistic potency from synergistic efficacy

While the null models of Bliss and Loewe emerge as special cases of MuSyC, and indeed the Bliss null model is equivalent regardless of h, neither Bliss nor Loewe are designed to distinguish synergistic potency and efficacy. Figure 5.4 shows this by plotting lines for which Loewe (A) or Bliss (B) are constant, but which span synergistic and antagonistic potency and efficacy in MuSyC.

5.4 Comparison of MuSyC to other metrics

Several newer methods of calculating synergy have been proposed in recent years. Here we compare several newer models with MuSyC. We find many similarities, and show that many alternative metrics quantify either potency or efficacy synergy, but not both. In addition, most of the alternative frameworks are limited to specific types of effect metrics, or limited by the magnitude of the effects. Collectively, key limitations and differences of several synergy models in comparison with MuSyC are summarized in Table 5.1.

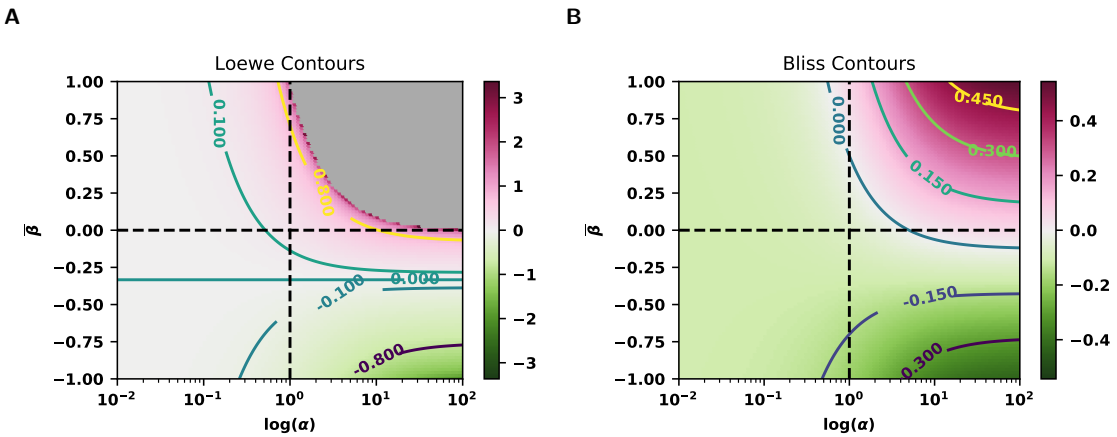


Figure 5.4: Contour lines showing constant values of Loewe (A) and Bliss (B) for various values of $\alpha_1 = \alpha_2$ and β. Loewe and Bliss are calculated for each plot at the EC50 of each drug. Parameters to generate: h=1, E0=1, E1=E2=0.5, and beta ranging from -1 (E3=1) to 1 (E3=0).

	Loewe	Bliss	CI	HSA	ZIP	Effective Dose	BRAID	MuSyC
Usable for arbitrary effect (not just percent affected)	X			X	X		X	X
Effect which exceeds weaker		X		X	X		X	X
Effect which exceeds stronger		X		X				X
Concentration independent						X	X	X
Satisfies sham	X		X				X	
Needs marginal data only		X		X				
Distinguishes potency and efficacy								X
Directly related to traditional dose response parameters				X		X		X

Table 5.1: Comparison of traditional and modern frameworks for calculating synergy.

5.4.1 Effective dose model

Zimmer *et. al.* (Zimmer et al., 2016) introduced the effective dose model as a parameterized extension of Bliss. In Bliss, the combination's expected "unaffected" cell fraction is the product of unaffected fractions from each individual drug. The effective dose model is constructed by fitting the dose response of each single drug to a 1D Hill equation as in equation 5.12, which they call $g_i(D_i)$, which is a function of the dose D_i. They then transform the doses D_i to "effective doses" via a Michaelis-Menten-like equation: $D_{i,eff} = D_i \prod_{j \neq i} \left(1 + a_{i,j} \frac{D_{j,eff}/D_{j,0}}{1+D_{j,eff}/D_{j,0}}\right)^{-1}$. ($D_{j,0}$ represents the IC50 of the j^{th} drug - each drug is fit explicitly assuming that the effect ranges from 0 to 1). The parameter $a_{i,j}$ represents how drug j modifies the effective dose synergistically ($a_{i,j} < 0$) or antagonistically ($a_{i,j} > 0$) drug i. Note that as $a_{i,j} \to -\left(\frac{1+D_{j,eff}/D_{j,0}}{D_{j,eff}/D_{j,0}}\right)$, $D_{i,eff} \to +\infty$, and as $a_{i,j} \to +\infty$, $D_{i,eff} \to 0$, which defines the bounds over which they can define $a_{i,j}$.

From this, they arrive at a surface fitting model by using their effective doses in a Bliss-like equation

$$U_{1,2} = g_1(D_{1,eff}) \cdot g_2(D_{2,eff})$$

There are obvious similarities between their parameter $a_{i,j}$ and MuSyC's parameters α_i, as both reflect a transformation of dose, however the exact method is slightly different. Their formulation assumes that each drug has a Michaelis-Menten like effect on the potency of the other drugs, whereas our model can account for non-Michaelis-Menten effects, specifically when $h! = 1$. Furthermore, by using equation 5.12, they explicitly assume that the measured drug effect ranges from 100% to 0%, and fit the data with this constraint. Therefore their model is unable to describe combinations where the two drugs either have unequal maximum effects, or the combination has a greater effect than the drugs can achieve alone, features which are commonly observed (Fallahi-Sichani et al., 2013). In contrast, MuSyC is able to fit dose response surfaces with arbitrary effect magnitudes. Nevertheless, the differences between the effective dose model and MuSyC mean that the

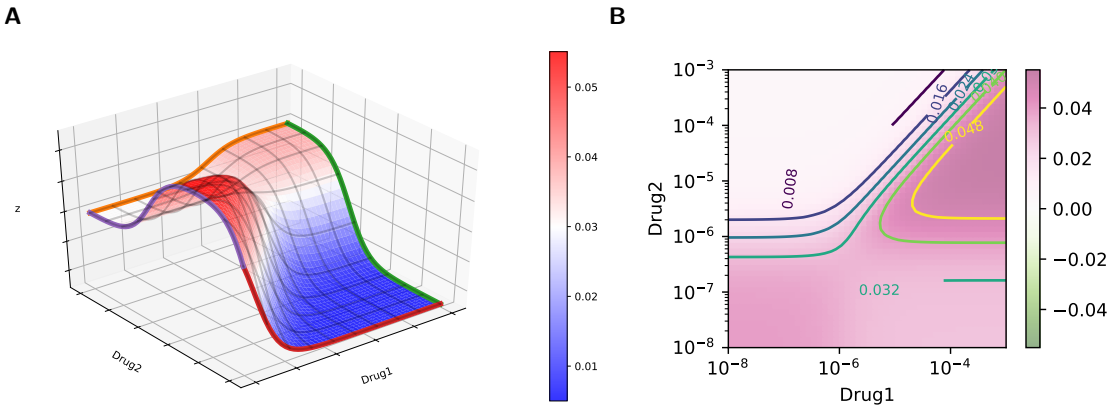


Figure 5.5: MuSyC is capable of describing highly complex dose response surfaces. At high concentrations of Drug1, Drug2's dose response becomes non-monotonic in this example. In this case, Drug2's antagonism of Drug1's potency causes A_2 to dominate at high concentrations of Drug2, and E_2 is a better response than both E_1 and E_3. Parameters to generate: $E_0 = 0.04$, $E_1 = 0.03$, $E_2 = 0.005$, $E_3 = 0.07$, $h_1 = 1.4$, $h_2 = 1.3$, $\alpha_1 = 1$, $\alpha_2 = 0$

effective dose model is able to generate surfaces that MuSyC cannot.

5.4.2 ZIP

Like the effective dose model (Zimmer et al., 2016), as well as our potency synergy (α), ZIP works by quantifying how one drug shifts the potency of the other (Yadav et al., 2015). ZIP is formulated for arbitrary E_0 and E_{max}; however, it assumes E_{max} is the same for both drugs, as well as the combination ($E_1 = E_2 = E_3$). To identify potency shifts, the ZIP method fixes the concentration of one drug, then fits a Hill-equation dose response for the other drug. However, for combinations with efficacy synergy or antagonism, slices of the dose-response surface can have non-Hill, and even non-monotonic shapes (Figure 5.5). Because MuSyC accounts explicitly for efficacy synergy, its surfaces are able to describe such complex drug combination surfaces where ZIP cannot.

Furthermore, ZIP calculates synergy at every concentration. This is similar to the approach taken by Bliss, Loewe, and CI, and can be used to find doses which maximize the observed synergy. Nevertheless, clinically it is less relevant to find the dose leading to

maximum synergy, because high synergy still does not mean high efficacy and quantifying synergy on a dose by dose basis can misleading. Additionally, this dose dependence often leads to ambiguous results about whether a given combination is synergistic or not, as it synergizes at some concentrations, and antagonizes at others.

5.4.3 BRAID

BRAID assumes that each drug alone has a Hill-like response, and constructs an *ad hoc* Hill-like equation for the combination (Twarog et al., 2016). This equation uses a single dose parameter which combines the doses of both individual drugs using a parameter κ. To uniquely solve for κ, this formalism, like Loewe additivity, adds the constraint that a drug in combination with itself must be neither synergistic nor antagonistic. By adjusting κ, BRAID is able to fit complex drug combination surfaces, including non-monotonic responses, unlike ZIP. Additionally, because BRAID fits the whole combination surface using a single parameter, it can be used to make unambiguous statements about whether the combination is synergistic or antagonistic. Nevertheless, BRAID does not account for differences in synergy due to efficacy vs. potency, whereas we find many combinations that are synergistic with respect to one, but antagonistic with respect to the other. Indeed, the biochemical interpretation of κ is not straightforward. Furthermore, the BRAID model is unable to fit combination surfaces with synergistic efficacy, as it assumes that the maximum effect of the combination is equal to the maximum effect of the stronger single-drug.

5.4.4 Highest Single Agent

Highest single agent (HSA) (Foucquier and Guedj, 2015) is a simple model that defines synergy dose by dose, as the net difference between the combination response, and the stronger single-drug response $HSA = min(E(d_1,0),E(0,d_2)) - E(d_1,d_2)$ (this form assumes that drug decreases E, though it can also be defined for drugs that increase E). At high concentrations of d_1 and d_2, HSA becomes similar to our definition of efficacy syn-

ergy (β). Nevertheless, at intermediate doses, HSA will conflate synergy of potency and efficacy.

5.5 MuSyC resolves Hill-dependent biases introduced by "sham-compliant" synergy frameworks

Several combination synergy frameworks are founded, at least partly, in the "sham combination principle". This principle is based on a thought experiment in which a single drug is divided into two vials, and a drug combination experiment is conducted treating those vials as though they are distinct drugs. The sham combination principle states that a synergy framework should not identify such a sham combination as either synergistic or antagonistic. The sham experiment has become a crucial test for synergy frameworks. Famously, Bliss is known to be non sham-compliant (Foucquier and Guedj, 2015), and this extends also to the Zimmer's effective dose model. Some synergy models, including BRAID, are derived by asserting sham compliance as an assumption of their model (Twarog et al., 2016; Schindler, 2017), while others, including Loewe, CI, and ZIP demonstrate sham compliance as a consequence of their models (Loewe, 1926; Chou and Talalay, 1984; Yadav et al., 2015).

Here we investigate the sham experiment in the context of MuSyC, and show that the biochemistry of the sham experiment is sufficiently different than the biochemistry of non-sham combinations, leading to a Hill-coefficient dependent bias in sham-compliant synergy frameworks.

5.5.1 Sham compliance of MuSyC

Traditionally, to be sham-compliant, MuSyC should predict neither synergy nor antagonism for sham combinations. Nevertheless, the mass-action framework of MuSyC requires that a sham combination should result in $\alpha_1 = \alpha_2 = 0$, because $A_{1,2}$, which reflects a cell state that is affected by both drugs, does not exist if there is only one drug. Furthermore,

if the $A_{1,2}$ state does not exist ($A_{1,2} = 0$), then E_3 does not exist, and therefore β is completely unconstrained. Therefore, to test the sham compliance of MuSyC, we only consider the case when $\alpha_1 = \alpha_2 = 0$.

Because both drugs are actually the same drug, we can further assert that $h_1 = h_2 = h$, $C_1 = C_2 = C$, and $E_1 = E_2$. Using this with equation 5.11, we find

$$E_{sham}^{MuSyC}(d_1, d_2) = \frac{E_0 + E_1 \frac{d_1^h + d_2^h}{C^h}}{1 + \frac{d_1^h + d_2^h}{C^h}} \tag{5.14}$$

Conversely, the null surface of a sham experiment can defined directly using its 1D dose response from equation 5.3 as

$$\begin{aligned} E_{sham}^{True}(d_1, d_2) &= E_{single}(d_1 + d_2) \\ &= \frac{E_0 + E_1 \left(\frac{d_1 + d_2}{C}\right)^h}{1 + \left(\frac{d_1 + d_2}{C}\right)^h} \end{aligned} \tag{5.15}$$

Equations 5.14 and 5.15 are only equivalent when $h = 1$, suggesting that MuSyC is not, in general, sham-compliant. To understand this discrepancy, consider a sham experiment where the drug has 3 binding sites ($h = 3$) and follows Hill-type kinetics, and is split into two vials used for a sham 2-drug combination experiment. Following the reasoning above, and referencing equation 5.2, we find

$$\begin{aligned} \frac{A}{U} &= \left(\frac{d_1 + d_2}{C}\right)^3 \\ &= \frac{d_1^3}{C^3} + 3\frac{d_1^2 d_2}{C^3} + 3\frac{d_1 d_2^2}{C^3} + \frac{d_2^3}{C^3} \end{aligned}$$

The two cross-terms in the middle account for the difference between the true sham experiment result and MuSyC. Hill-kinetics formally result from equilibria of chemical reactions of the form $U + hDrug \rightleftharpoons A$, such that all h drug molecules bind simultaneously.

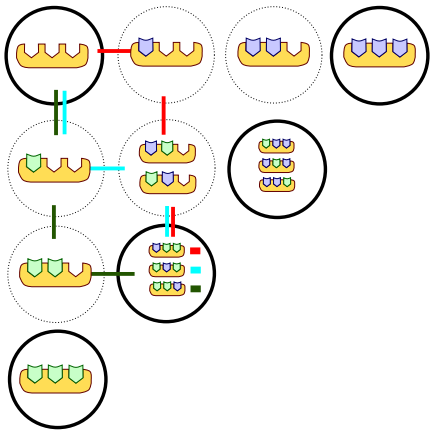


Figure 5.6: An illustration of the unique biochemistry of the sham experiment. The top left state represents an undrugged cell, with three binding sites. Hill kinetics are derived from the assumption that all three drugs bind instantaneously, however intermediate transition steps are shown here. In a sham experiment, a drug is treated as though it it two separate drugs (green and blue). As shown in this figure, mixed states of the binding sites being bound by both green and blue drugs exist. Because green and blue are the same, these all represent the same fully drugged state. We highlight the three paths that can be followed to reach a mixed-drugged state.

The first cross-term above represents a mixed case, in which only 2 molecules of drug_1 are bound and 1 molecule of drug_2 is bound. This is fine in a sham experiment, because drugs 1 and 2 are the same, so the Hill assumption is satisfied when combinations of drug 1 and 2 add up to account for a total of h (3, in our example) molecules being bound, as shown in Figure 5.6. (the coefficients of the cross terms correspond to the 3 ways this can be achieved: '112', '121', '211').

However, in combinations of two different drugs, these cross-terms represent partially drugged states that violate the Hill-kinetic assumption that all h molecules of each drug must bind simultaneously. Furthermore, if the two drugs have different Hill-coefficients, there is no evident rationale to combine them into a single, meaningful Hill-coefficient as in 5.15, and there is no clearly defined diagonal in Figure 5.6.

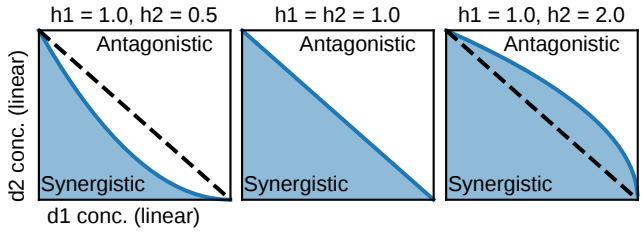


Figure 5.7: The hill coefficient induces nonlinearities that bias Loewe. Loewe and CI assume the dose-response surface contours (i.e., the DIP rate axis comes out of the page) are linear (middle panel). The blue area indicates regions synergistic by MuSyC, while white regions are antagonistic. When h2<1 (left panel), Loewe and CI misclassify the hatched region as synergistic, while when h2>1 (right panel), they misclassify the hatched region as antagonistic.

5.5.2 Sham compliance introduces a Hill-coefficient dependent bias on synergy

The centrality of the sham principle within the field of drug synergy cannot be overstated, as new frameworks are praised or critiqued based on compliance, and several methods are even derived with the sham principle as an explicit assumption (Cokol et al., 2011; O'Neil et al., 2016). Because sham compliance is characterized by linear isoboles, we can use equation 5.11 to explore the consequence of h on sham compliant models.

Specifically, for drugs with $h < 1$, we predict that synergy models with linear isoboles will tend to overestimate synergy, while drugs with $h > 1$ will lead to underestimated synergy (Figure 5.7). Using two previously published drug combination datasets (Cokol et al., 2011; O'Neil et al., 2016), we explored the impact of h on synergy, specifically using Loewe additivity. To this end, we calculated an expected baseline synergy value for these methods by constructing a MuSyC null surface ($\alpha_1 = \alpha_2 = 1$ and $\beta = 0$) for each combination, and using the other frameworks to calculate the synergy of the MuSyC null surface.

Figures 5.8 reveal that these methods are biased toward synergism at low values of h. Interestingly, the underestimation of synergy when $h > 1$ is largely counterbalanced by the overestimation of synergy that these models make by assuming $\alpha_1 = \alpha_2 = 0$ (that is, if $\alpha_1 = 0.5$, MuSyC shows that drug 1 antagonizes the potency of drug 2, but because $\alpha_1 > 0$,

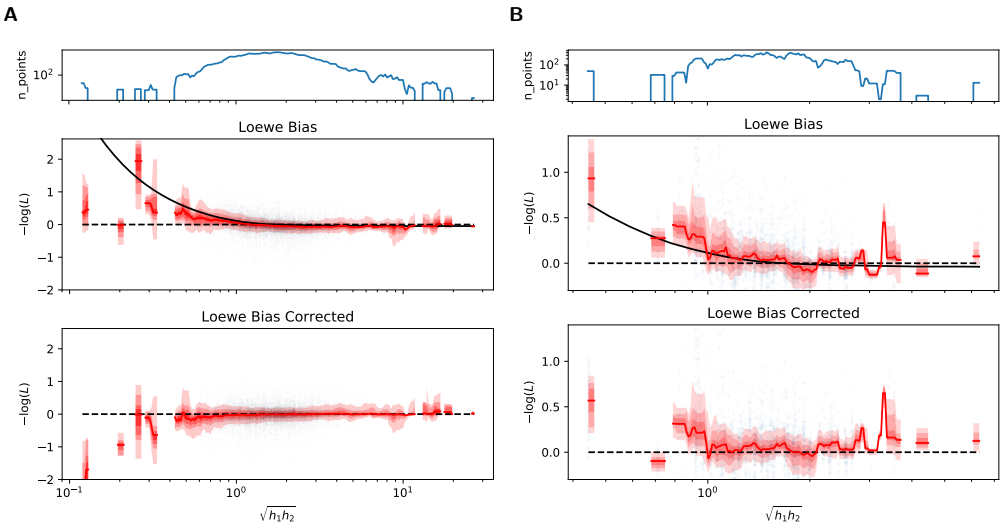


Figure 5.8: Hill dependent bias observed in large published drug screens. Loewe synergy calculated for two large drug screens, from (Cokol et al., 2011; O'Neil et al., 2016). The red shaded regions show moving window percentiles (10th through 90th percentiles, in steps of 10). The top plots show how many are present in the window. Black curve in middle plot shows the median predicted Loewe baseline with $\alpha = 1$ as a function of Hill slope with $h_1 = h_2$. This curve is just an average, as each individual datapoint has its own Loewe bias baseline based on its concentration. Subtracting each datapoint's individual bias yields the bottom plot, which show the elimination of the bias.

this is still more synergistic than the sham null model of infinite potency antagonism). By subtracting the MuSyC predicted baseline synergy for each tested combination, the Hill-dependent bias is significantly reduced.

5.6 2D Hill equation distinguishes synergy of potency from synergy of efficacy for drug combinations

As we have shown, the dose-effect relationship of a single drug is traditionally quantified by the Hill equation, which contains parameters describing efficacy (Emax) and potency (EC50) of a dose-response curve (Figure 5.9A). To characterize the dose-effect relationship for drug combinations, we derived a 2D generalization of the Hill equation from a drug-combination model (Figure 5.1), using principles of mass action kinetics. The 2D Hill equation parameterizes a dose-response surface (Figure 5.9B) (Greco et al., 1992; Yadav et al., 2015), a 2D generalization of 1D dose-response curves (Figure 5.9A). In this equation, the alteration in the potency and changes in the efficacy are quantified by parameters for synergistic potency, denoted by α, and synergistic efficacy, denoted by β. These parameters govern the shape of the dose-response surface and can capture complex patterns in the experimental data.

The parameter β is defined as the percent increase in a drug combination's effect beyond the most efficacious single drug. For instance, in the case of synergistic efficacy ($\beta > 0$), the effect at the maximum concentration of both drugs (E3) exceeds the maximum effect of either drug alone (E1, E2) (Figure 5.9C Quadrants I, II). For antagonistic efficacy ($\beta < 0$) (Figure 5.9C Quadrants III, IV), at least one or both drugs are more efficacious as single agents than in combination.

The parameter α quantifies how the effective dose of one drug is altered by the presence of the other. In the case of synergistic potency ($\alpha > 1$), the half-maximal effective concentration or EC50 (denoted C in Figure 5.9C) decreases due to the addition of the other drug (Figure 5.9C Quadrants I, IV). In the case of antagonistic potency ($0 \leq \alpha < 1$), the EC50

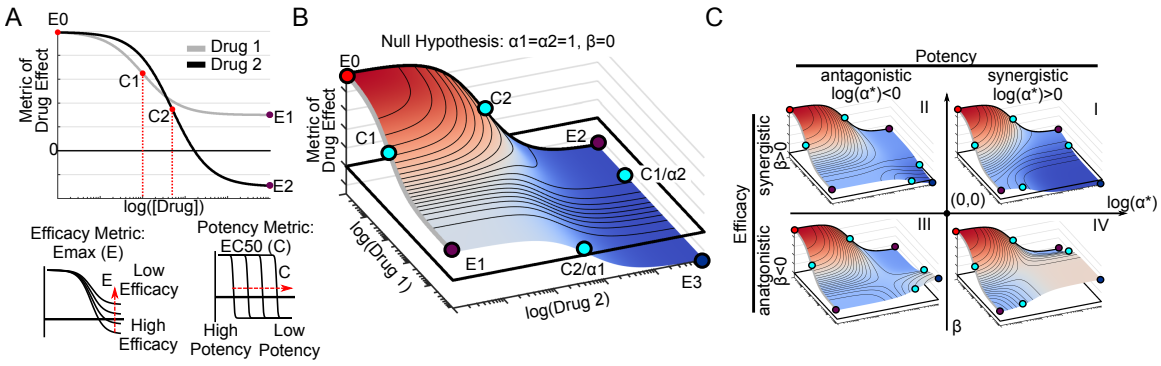


Figure 5.9: 2D Hill equation for parameterizing dose-response surfaces distinguishes between synergistic efficacy and synergistic potency of a drug combination. A) Sigmoidal dose-response curves relating drug concentration to a measured effect, fit to the 1D Hill equation. The equation contains parameters for estimating a drug's potency, (C) the concentration required for half-maximal effect, and efficacy (Emax), the maximal effect. Here, Drug 1 is more potent than Drug 2 (C1<C2) while Drug 2 is more efficacious (E2<E1). For simplicity, in this diagram we only depict a metric for which increasing efficacy results in lower values (e.g., anti-proliferative drugs). However, the same considerations are valid for metrics which increase in value with increased efficacy. B) A dose-response surface for Drugs 1 and 2, based on the 2D Hill equation, under the null hypothesis of no synergy of efficacy and potency ($\alpha_1 = \alpha_2 = 1$), and ($\beta = 0$). C) Representative dose-response surfaces for each quadrant on a Drug Synergy Diagram (DSD). The vertical axis is divided into antagonistic ($\beta < 0$) and synergistic ($\beta > 0$) efficacy. The horizontal axis is divided into antagonistic ($log(\alpha) < 0$) and synergistic ($\log(\alpha) > 0$) potency where α can be either α_1 or α_2. Quadrant I corresponds to synergistic potency and efficacy. In contrast, combinations in quadrant IV have synergistic potency, but antagonistic efficacy corresponding to a blunting in efficacy at lower doses.

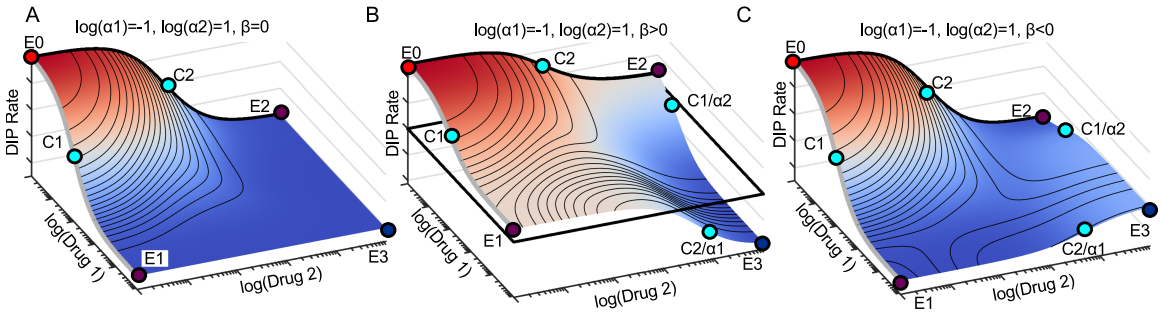


Figure 5.10: Combination surfaces with asymmetric synergistic potency. A) Combination surface with asymmetric potency and no synergistic efficacy ($E_1 = E_2 = E_3$). At saturating concentrations of one drug there is no difference between the maximal effect and minimal effect of the other drug. B) Combination surface with asymmetric synergistic potency with synergistic efficacy. C) Combination surface with asymmetric synergistic potency with antagonistic efficacy.

increases (Figure 5.9C Quadrants II, III). Since each drug can modulate the effective dose of the other independently (Zimmer et al., 2016), we define two α values (α_1 and α_2). This separation of α values in the 2D Hill equation makes it possible for a given drug combination to have synergism of potency in one direction ($\alpha_1 > 1$), and antagonism of potency in the other ($\alpha_2 < 1$), or vice versa (see Figure 5.10 for example surfaces).

In summary, the 2D Hill equation enables a formalism, termed Multi-dimensional Synergy of Combinations (MuSyC), in which synergistic efficacy and synergistic potency are orthogonal and quantified by the parameters β, related to Emax, and α related to EC50.

5.7 MuSyC quantifies synergy of potency and efficacy in a drug combination screen

We applied MuSyC to evaluate the synergistic potency and efficacy of 64 drugs in combination with osimertinib, a mutant EGFR-tyrosine kinase inhibitor recently approved for first-line treatment of EGFR-mutant NSCLC (Soria et al., 2018). The selected drugs span a diverse array of cellular targets that can be broadly grouped into four categories: kinases, receptors and channels, epigenetic regulators, and mitotic check-points (Figure 5.12C), each with several sub-categories. The combinations were tested in PC9 cells, a canonical model of EGFR-mutant NSCLC (Sharma et al., 2010) using a high-throughput,

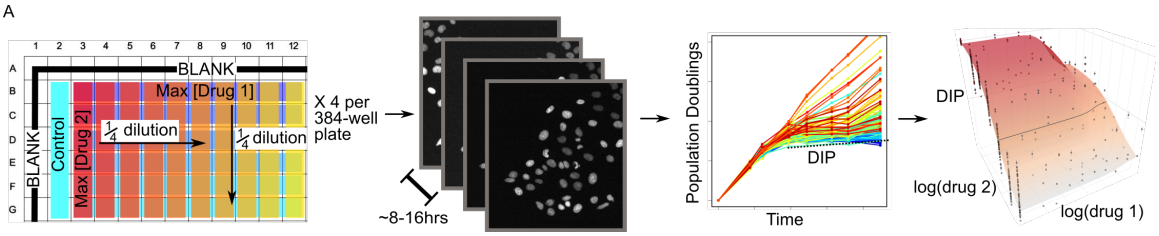


Figure 5.11: High throughput experimental pipeline for generating dose response surfaces. A) Initial drug matrix is prepared on 384-well plate and transferred to cells seeded at sub-confluent densities. Cells have are engineered to express a fluorescently tagged histone (H2B-RFP) allowing for cell counts using automated segmentation software. Each condition is imaged every 6-8 hours resulting in growth curves. The growth curves are fit for the DIP rate (Harris et al., 2016) to quantify drug effect. This matrix of effects is then fit to estimate the dose response surfaces.

in vitro, drug-screening assay (Figure 5.11). We quantified drug effect using the drug-induced proliferation (DIP) rate metric (Harris et al., 2016), a recently developed drug effect metric which avoids temporal biases characteristic of traditional endpoint assays (see section 5.11.1).

We extracted MuSyC synergy parameters from fitted combination surfaces for all combinations (Figure 5.12). $\beta(obs)$ is the observed synergistic efficacy at the maximum tested dose range). As single agents, the drugs display wide ranges of efficacy (E2) (Figure 5.13) and potency (C). The efficacy and potency of the single agents have no relationship with the synergistic efficacy and synergistic potency when combined with osimertinib (Pearson correlation$<$0.2) (Figure 5.13). This confirms MuSyC's synergy parameters are independent of single-agents' dose-response curve and therefore cannot be predicted from single agent pharmacologic profiles.

As specific examples highlighting synergistic efficacy and potency, the dose-response surface for M344 (a histone deacetylase (HDAC) inhibitor) in combination with osimertinib exhibits extreme synergistic efficacy ($\beta(obs) = 1.25 \pm 0.03$), reflecting a 125% increase in efficacy over osimertinib alone (Figure 5.12A,D). However, this combination is antagonistically potent ($log(\alpha_2) = -0.90$) as observed in the shift in the EC50 of osimertinib in the presence of 1uM M344 (Figure 5.12A red to purple dotted line). In contrast, ceritinib, an

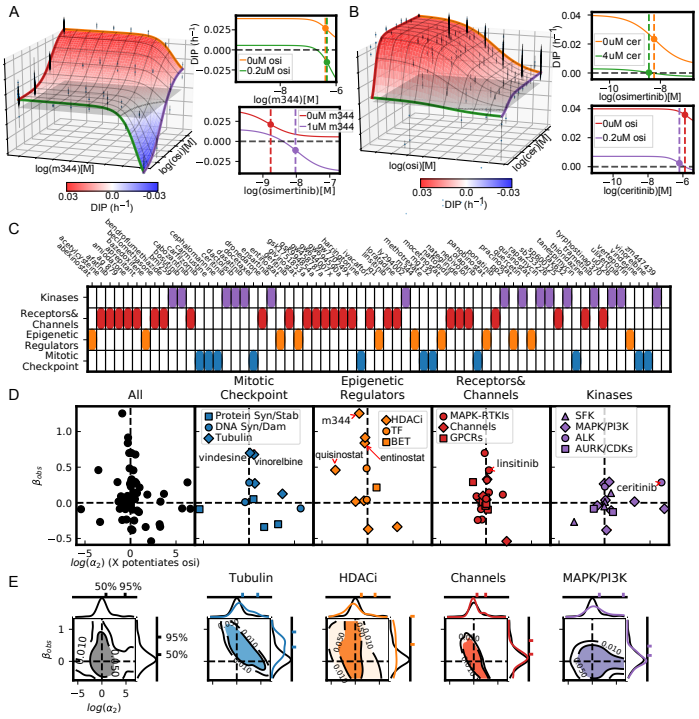


Figure 5.12: High throughput screen of 64 drugs combined with osimertinib demonstrates the drug class dependence of synergistic potency and efficacy in NSCLC. A) Combination surface of M344, an HDACi, and osimertinib (osi). Grey plane indicates a cytostatic growth rate (ie DIP rate=0 h-1). Left are the dose-response curves for each drug alone (orange and red curves) and each drug with the maximum tested concentration of the other (green and purple). Colors correspond to the colored lines on the combination surface. The dotted lines demarcate the EC50 for each curve. B) Combination surface for ceritinib (cer), an ALK inhibitor with off-target effects on IGF1R, in combination with osimertinib. Ceritinib increases the potency of osimertinib at maximum tested concentration, as observed in the shift of the EC50 between orange and green curves in the top left panel. The shift is proportional to the concentration used and would, therefore, increase at higher concentrations; however, such concentrations are not physiologically realizable due to the low potency of ceritinib in this system (EC50=2.02 uM). C) Drug panel used in combination with osimertinib (mutant EGFR-TKI) matched by target's biological function and grouped in the 4 indicated categories. D) DSDs for drug combinations. In practice, it is best to quantify the synergistic efficacy at the maximum tested concentrations (denoted $\beta(obs)$) instead of the theoretical maximal effect (β). The vertical axis quantifies synergistic efficacy, ($\beta(obs)$). The horizontal axis ($log(\alpha_2)$) quantifies how osimertinib's potency is modulated by each drug (see Figure 5.14 for α_1 plot). Error bars represent the parameter's fit uncertainty based on the MCMC optimization (See methods 5.11 for fitting algorithm). E) 2D density plots and associated marginal distributions for $\beta(obs)$ (vertical axis) and α_2 (horizontal axis) for all drugs (black) and selected category subclasses. Colored tick marks indicate the 50% and 95% probability density intervals for each class.

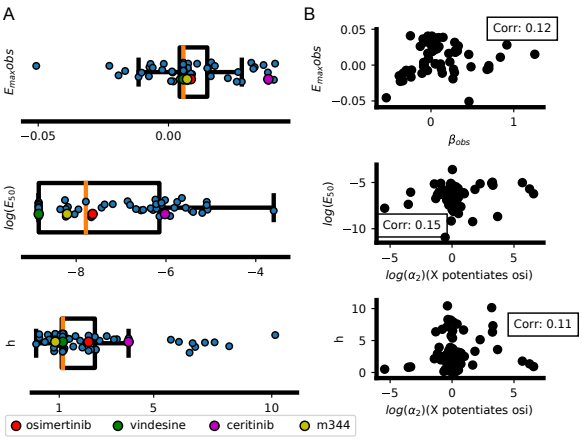


Figure 5.13: Correlation of efficacy ($E_{max}(obs)$) and potency (C) with the synergistic potency and efficacy in combination with osimertinib. A) Jitter plot of the single drug's ($E_{max}(obs)$), $C[uM]$, and hill slope (h). B) Synergy parameters do not correlate with a single drug's potency and efficacy in isolation. (α_2=X potentiates osimertinib)

ALK inhibitor with off-target effects on IGF1R (Shaw et al., 2014), increases osimertinib's potency ($log(\alpha_2) = 6.25$) (Figure 5.12B green to orange dotted line) at 4uM (maximal tested concentration). However, the concentration of ceritinib required to observe a maximal reduction in the EC50 of osimertinib is greater than the maximum tested concentration due to the low potency of ceritinib alone (EC50=2.02uM) highlighting the importance of interpreting synergistic potency in the context of the absolute potency (C) fit in the 2D Hill equation.

To visualize synergistic/antagonistic efficacy and potency across the entire drug screen at once, we plot each drug combination on DSDs constructed on axes representing the synergistic parameters extracted from the dose-response surfaces (Figure 5.12A,B). Specifically, synergistic efficacy ($\beta(obs)$) is plotted on the vertical axis, and synergistic potency on the horizontal axis (Figure 5.12D). Because $\alpha_1 \neq \alpha_2$, here we plot only the synergistic effect on osimertinib's potency on the horizontal axis (osimertinib's alteration of the other drug's potency is shown in Figure 5.14). These DSDs reveal distinguishing trends between the drug categories. The marginal distributions of tubulin destabilizers (which affect microtubule polymerization) or epigenetic regulators (which broadly affect gene expression) in

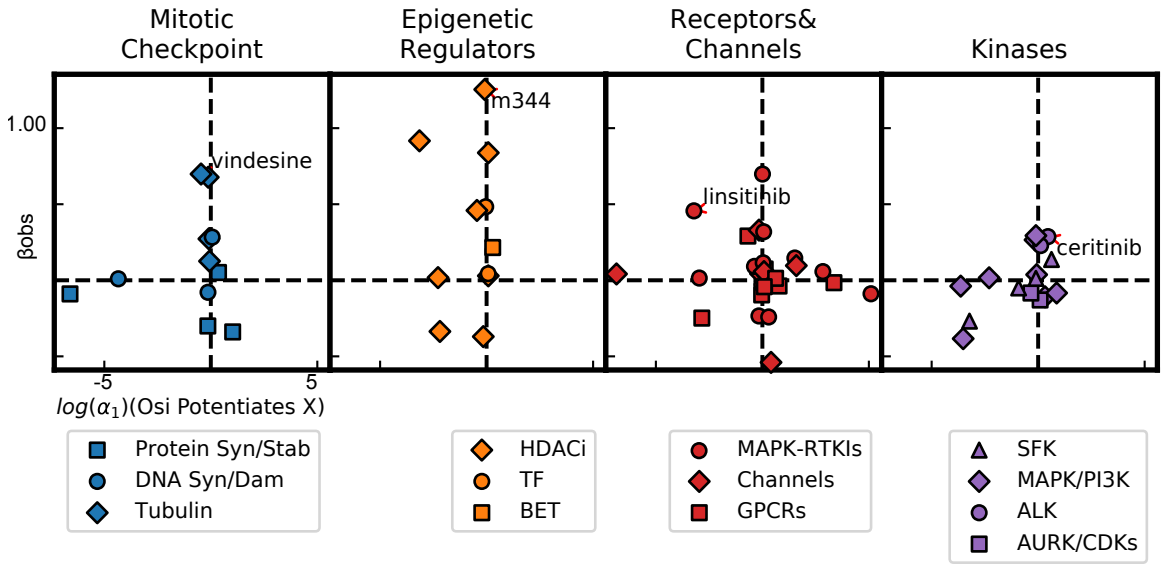


Figure 5.14: DSDs for by drug class for examining the potentiation of drug X by osimertinib. Notably, ceritinib is not potentiated by osimertinib (last panel) and linsitinib's potency is antagonized by osimertinib.

combination with osimertinib shows an upward shift along the axis of synergistic efficacy (Figure 5.12E comparing blue and orange to black vertical distributions). In contrast, we observe limited synergistic/antagonistic efficacy in drugs targeting receptors and channels with the notable exception of linsitinib (an IGF1R inhibitor) (Figure 5.12D). Cross-talk has been demonstrated between the IGF1R pathway and the EGFR pathway (van der Veeken et al., 2009).

We observed limited synergistic efficacy in directly co-targeting kinases in the MAPK pathway suggesting this may be an unproductive avenue in EGFR-mutant NSCLC (Figure 5.12E). As expected, no synergistic benefit was observed for combinations within the channels drug subclass (Figure 5.12E comparing blue and orange distributions to black distribution). Epigenetic regulators have previously been suggested to prime NSCLC for sensitivity to EGFRi (Sharma et al., 2010; Greve et al., 2015; Schiffmann et al., 2016) and the HDAC inhibitor entinostat in combination with erlotinib (first generation EGFR-TKI) has been shown to increase overall survival in EGFR-mutant NSCLC cases with high expression of E-cadherin (Witta et al., 2006, 2012). Consistent with this, we also observe

entinostat was synergistically efficacious with osimertinib ($\beta(obs) = 0.84 \pm 0.23$) in PC9 cells, an E-cadherin high expressing cell line (Shimoyama et al., 1992).

In summary, by quantifying synergy of potency separate from synergy of efficacy, MuSyC reveals drug-class trends which can be used to guide subsequent screens and translational applications in NSCLC.

5.8 MuSyC validates co-targeting RAF and MEK in BRAF-mutant melanoma

The NSCLC osimertinib drug screen (Figure 5.12) suggests combinations targeting molecules within the same signaling pathway may not be productive avenues for increasing efficacy. However, a combination used clinically in BRAF-mutant melanoma (Long et al., 2014; Eroglu and Ribas, 2016) co-targets BRAF and MEK in the MAPK pathway. To investigate this combination in more detail, we screened a panel of four BRAF inhibitors (BRAFi) in pairwise combination with four MEK inhibitors (MEKi), with a panel of eight BRAFV600 mutant melanoma cell lines for a total of 128 experiments (see (Paudel et al., 2018) for cell-line information).

Based on the mean $\beta(obs)$ across cell-lines, all 16 combinations were synergistically efficacious (Figure 5.15A, Figure 5.16) indicating MuSyC would have identified this combination prospectively. MuSyC is sensitive enough to detect variations in synergistic efficacy between cell lines (Figure 5.15A, 5.16) pointing to cell-intrinsic mechanisms modulating synergistic efficacy of the particular combination. For instance, A2058, a canonically insensitive cell-line (Hardeman et al., 2017), displayed low average synergistic efficacy to combined MAPK inhibition suggesting this cell-line does not depend on MEK reactivation in the presence of BRAFi as a resistance mechanism, which may be mediated instead by an altered metabolic phenotype (Parmenter et al., 2014; Hardeman et al., 2017).

These data indicate the magnitude of a drug combination's synergistic efficacy depends upon the molecular and cellular context, i.e., co-targeting within the MAPK pathway may work for mutant-BRAF melanoma, not for mutant-EGFR NSCLC.

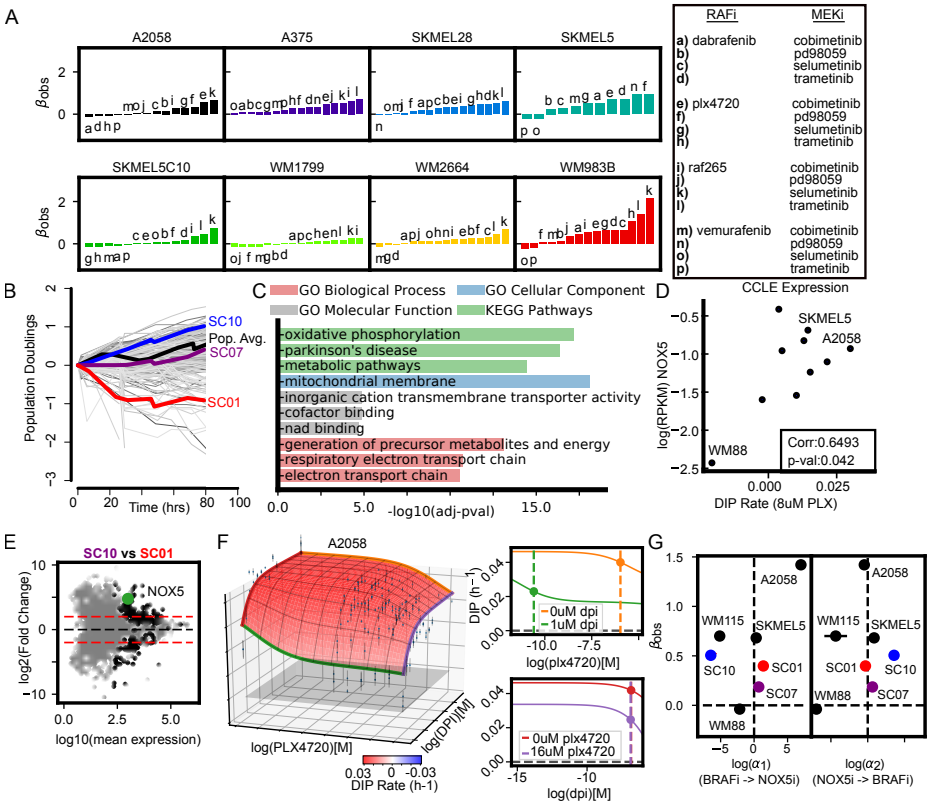


Figure 5.15: Synergistic efficacy and/or potency of drug combinations in BRAF-mutant melanoma. A) Eight BRAF-mutant melanoma cell lines were treated with all possible pairwise combinations of 4 RAF and 4 MEK inhibitors for a total of 128 unique combinations. Waterfall plots of $\beta(obs)$ for each cell line with all combinations which converged in fitting. Drug combinations noted by letter in the legend to right. B) Growth curves of single-cell-derived subclones from SKMEL5 treated with 8uM PLX4720. Grey curves represent colony growth according to the clonal fractional proliferation assay (Tyson et al., 2012). The parental response indicated in black curve. Subclones with varying sensitivity to BRAFi (SC01, SC07, SC10) were subsequently used to identify 200 DEGs. C) Top gene set enrichment terms for 200 DEGs. D) NOX5 expression correlates with BRAFi sensitivity in 10 BRAF-mutant melanoma cells. Drug sensitivity was quantified as DIP rate measured in 8M PLX4720. E) Pairwise comparison between SC01 and SC10 of DEGs (FDR<0.001) identified using DESeq2 (Love et al., 2014). The 200 identified DEGs from ANOVA between all 3 subclones are in black. Dotted red lines denote plus/minus 4X fold change. F) Dose-response surface for PLX4270+DPI (NOX5 inhibitor) in A2058. G) DSD for NOX5i+BRAFi reveals NOX5 expression correlates with the synergistic efficacy of NOX5i+BRAFi.

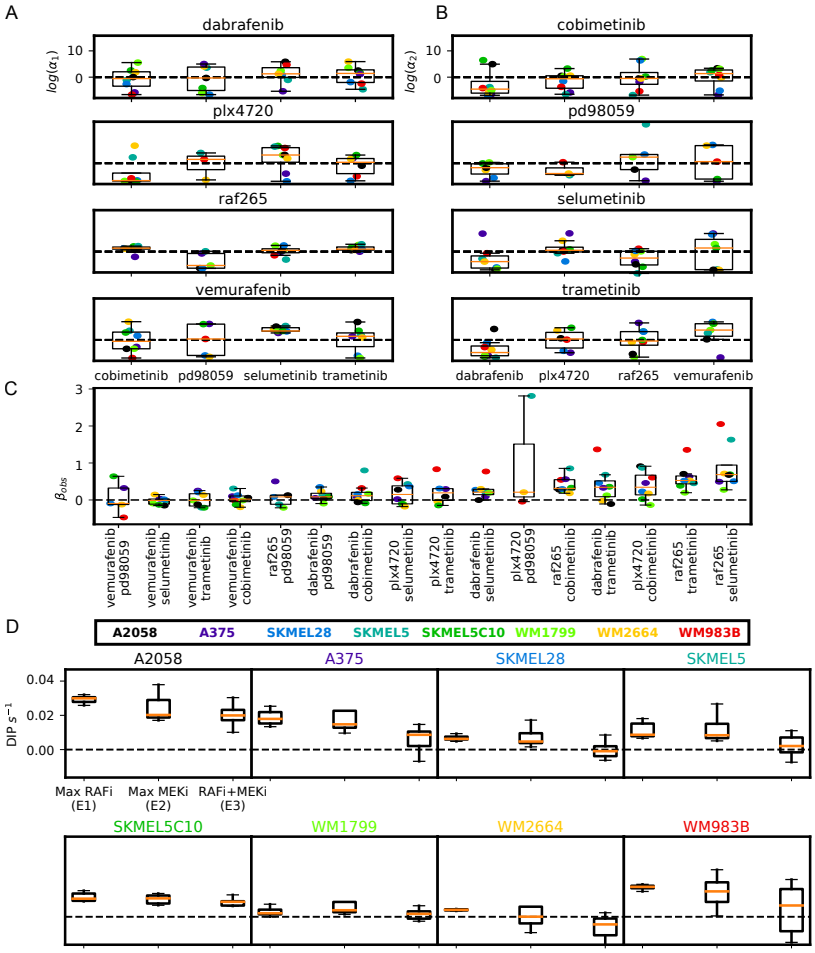


Figure 5.16: Synergistic Potency, synergistic efficacy, and maximal effect of combined RAF and MEK inhibition. A) Jitter plots of $log(\alpha_1)$ for each RAF inhibitor for the four MEK inhibitors tested. α_1 corresponds to the alteration in MEK inhibitor's effective dose due to the presence of a RAF inhibitor. Dashed line denotes zero separating synergistic and antagonistic potency. The color of plotted points is corresponds to the cell line as annotated at the bottom of the figure. B) Jitter plots of $log(\alpha_2)$ for each MEK inhibitor for the four RAF inhibitors tested. α_2 corresponds to the alteration in RAF inhibitor's effective dose due to the presence of a MEK inhibitor. C) Rank ordered jitter plots of the median β_{obs} for each drug combination across all cell lines. D) Distribution of maximal effects for RAFi alone (E1, 4 drugs), MEKi alone (E2, 4 drugs), and the combination (E3, 16 combinations) for each cell line. Orange bar denotes mean.

MuSyC also provides information on synergistic potency for these combinations. Specifically, it determines whether a combination is antagonistically or synergistically potent in a particular cell line. Notably, a popular combination (Dabrafenib and Trametinib) is synergistically efficacious but antagonistically potent in all cell lines except one (Figure 5.16), a trade-off that may be relevant for translation.

5.9 NOX5 is a synergistically efficacious BRAF co-target in BRAF-mutant melanoma

While drug combinations are commonly initially identified from top-down approaches, e.g. high throughput drug screens, many others, including BRAFi/MEKi, were discovered from a bottom-up approach via molecular correlates of insensitivity (Paraiso et al., 2010; Sturm et al., 2010). To demonstrate how MuSyC can also be applied to these studies, we looked for molecular correlates of BRAFi insensitivity between subclones of a BRAF-mutant melanoma cell line (SKMEL5). We isolated three SKMEL5 with differential sensitivity to BRAFi (Figure 5.15B), quantified gene expression using RNAseq, and identified 200 differentially expressed genes (DEGs) (FDR$<$0.001). This gene set was significantly enriched in processes, cellular components, and molecular functions relating to metabolism (Figure 5.15C), aligning with previous reports on the relationship between altered metabolism and resistance to BRAFi (Haq et al., 2013; Smith et al., 2016; Hardeman et al., 2017).

We computed the correlation of the 200 DEGs' expression to BRAF sensitivity across a 10 cell line panel (see Methods 5.11) using data from (Subramanian et al., 2017). NADPH oxidase 5 (NOX5) stood out as one of five genes with a significant, positive correlation with BRAFi insensitivity (Pearson r=0.65, p-val=0.042) (Figure 5.15D) and was significantly up-regulated in the BRAFi insensitive subclone (SC10) compared with the sensitive subclone (SC01) (Figure 5.15E). Previously unconsidered, NOX5 is an interesting target due to its convergent regulation on metabolic and redox signaling at mitochondria, (Lu et al., 2012; LeBleu et al., 2014), processes significantly enriched in the DEGs (Figure

5.15C).

We tested PLX4720 in combination with DPI, a NOX5 inhibitor (Jaquet et al., 2011), in a panel of 7 melanoma cell lines selected based on differential NOX5 expression. The synergistic efficacy of the combination correlated with NOX5 expression (Pearson r=0.77, p-value=0.043) (Figure 5.15G). Of note, A2058, well-known for its resistance to BRAFi exhibited the greatest synergistic efficacy and NOX5 expression, among the cell lines (Figure 5.15F). Indeed, in A2058, the NOX5i/BRAFi combination was more synergistically efficacious than all tested MEKi/BRAFi combinations (Figure 5.15A).

Taken together, these results suggest co-targeting NOX5 in BRAF-mutant melanoma could lead to improved outcomes for BRAF-mutant melanoma patients with a unique metabolic program for which NOX5 is a biomarker. Furthermore, this study demonstrates the utility of MuSyC in translating putative combinations based on molecular data to the clinically-relevant scale of the cancer cell population. These results demonstrate the broad applicability of MuSyC to discover synergistically potent and/or efficacious combinations from both high throughput drug screens and molecularly directed studies.

5.10 Discussion

The goal of using synergistic drugs is to achieve more with less. It is therefore intuitive that two types of synergy exist: one corresponding to how much more is achievable (synergistic efficacy), the other to how much less is required (synergistic potency). Finding such combinations is vital for optimizing therapeutic windows, as there exists a fundamental trade-off between clinical efficacy and tolerable doses. Synergistic combinations widen the therapeutic window thereby increasing the odds of success. By stratifying synergy along distinct axes of potency and efficacy using MuSyC, informed choices can be made about this trade-off. Furthermore, the distinction facilitates identifying drug-classes that can be iteratively expanded in future drug combination screens to optimize synergistic efficacy or synergistic potency, whichever is desirable for a particular system.

Conditions for which single-drug efficacy is deemed sufficient would benefit from synergistically potent combinations to drive down toxicity. Alternatively, conditions with low efficacy treatments are in pressing need of synergistically efficacious combinations in order to improve depth and durability of response. In this respect, MuSyC provides a global view of the synergistic behavior of whole classes of drugs, e.g. from a high-throughput drug screen, via DSDs. It also facilitates detailed investigations of a given combination via dose-response surfaces. In the case of osimertinib, which targets a signaling pathway, combinations targeting independent biological processes (chromatin organization and microtubule polymerization) tend to produce a gain in efficacy, whereas combinations targeting co-dependent processes (such as other signaling pathways) tend to vary along the axis of potency. These insights should guide future screens for translational applications of combinations with desired synergistic properties.

Loewe additivity and Bliss independence have maintained dominance in the field, along with the related work of Chou and Talalay. Yet there is broad uncertainty regarding the appropriate use of each of these methods because they are based on differing foundational principles, often leading to incompatible results (Tang et al., 2015). MuSyC removes this uncertainty by unifying these methods into a single framework, within which Loewe and Bliss emerge as special cases.

There has been much critical analysis over the past twenty-five years on the term 'synergy' (Greco et al., 1992, 1995), arguably rooted in the practice of defining synergy with respect to arbitrary expectations of drug additivity implicitly codified in their foundational principles. In contrast, ambiguity about the meaning of 'synergy' disappears in MuSyC, because its parameters directly relate to standard pharmacology dose-response parameters: efficacy and potency. Indeed, a major advance of MuSyC is the decisive shift toward parameters directly relating to a quantifiable change in efficacy and/or potency. Thus, ambiguous questions, such as "Is there synergy?", can be recast into more precise questions, such as "How much does efficacy/potency change when these drugs are used in combination?".

Such precise language may promote a move away from arbitrary cut-offs for "significant synergy" (Meletiadis et al., 2010).

Traditional frameworks define synergy on a dose-by-dose basis, commonly leading to ambiguous results across the dose matrix. This ambiguity precludes connecting gene expression with the synergy of a given combination, hampering its clinical translatability. MuSyC instead fits synergy parameters for the entire dose-response surface, and when applied across diverse drug panels, can help elucidate the connections between gene expression and synergy of potency or efficacy for a combination.

While we focused on the DIP rate as our metric of effect, MuSyC may be applied to any quantifiable phenotype whose dose-response is suitable to be fit by a Hill equation. In contrast, all other synergy models we surveyed impose strict constraints on the type and/or magnitude of the drug effect metric. Thus, MuSyC opens up the potential to study synergy of drug effects which were previously impossible to address by existing methods.

In conclusion, we have presented MuSyC, a drug synergy framework that maintains a distinction between two intuitive types of synergy and that may be applied to any drug effect metric. We showed this framework allows for a richer understanding of combination drug interactions, with practical translational consequences. We foresee this approach will streamline drug discovery pipelines and facilitate the deployment of precision approaches for therapeutic combinations.

5.11 Methods

5.11.1 Calculating the DIP Rate

Traditionally, the efficacy of an anti-proliferative compound was measured as the percent of viable cells (relative to control) after a treatment interval (Fallahi-Sichani et al., 2013; Yang et al., 2013); however, it has been recently shown this metric is subject to temporal biases (Hafner et al., 2016; Harris et al., 2016). To address these biases, we pre-

viously developed an unbiased metric of drug effect termed the drug-induced proliferation (DIP) rate(Harris et al., 2016). The DIP rate is defined as the steady state proliferation rate achieved after drug equilibration. A positive DIP rate indicates an exponentially growing population, while a negative DIP rate indicates a regressing one. A rate of zero indicates cytostasis, which may result from cells entering a non-dividing state or from balanced death and division (Paudel et al., 2018). A single compound is often insufficient to achieve a negative DIP rate, prompting the search for combinations of drugs to improve increase the efficacy. This mirrors treatment in the clinic which commonly involves a cocktail of cytotoxic therapies and/or targeted therapies (Yap et al., 2013; Mokhtari et al., 2015; Ettinger et al., 2017). The method used to calculate the DIP rate is extensively described in (Harris et al., 2016).

5.11.2 Bioinformatic identification of gene expression correlating to BRAFi insensitivity

RNA-seq of melanoma cell lines

Total RNA was isolated from untreated SKMEL5 single-cell derived sublines, each in triplicate, using Trizol isolation method (Invitrogen) according to the manufacturers instructions. RNA samples were submitted to Vanderbilt VANTAGE Core services for quality check, where mRNA enrichment and cDNA library preparation were done with Illumina Tru-Seq stranded mRNA sample prep kit. Sequencing was done at Paired-End 75 bp on the Illumina HiSeq 3000. Reads were aligned to the GRCh38 human reference genome using HISAT2 (Siren et al., 2014; Kim et al., 2015) and gene counts were obtained using featureCounts (Liao et al., 2014). All downstream analyses were performed in R.

Differential Expression Gene Analysis and Functional Enrichment Analysis

Differentially Expressed Genes (DEGs) were selected by ANOVA on baseline gene expression data on three clones based on a statistical cutoff of Likelihood Ratio Test (LRT) (p-values < 0.001). Functional enrichment analyses, including GO Term Enrichment and Pathway Enrichment Analysis were done using CRAN Package Enrichr

(https://cran.r-project.org/web/packages/enrichR/index.html), based on a web-based tool for analyzing gene sets and enrichment of common annotated biological functions (http://amp.pharm.mssm.edu/Enrichr/). The enriched GO terms and enriched KEGG pathways were restricted to those with p-values corrected for multiple testing less than adjusted.p.values < 0.001. The top GO Biological Processes included generation of precursor metabolites and energy, electron transport chain, inorganic cation transmembrane transport, and metabolic process. The top GO Molecular Function terms included inorganic cation transmembrane transporter activity, cofactor binding, NAD binding, and ATPase activity. The top GO Cellular Component terms included mitochondria. Top KEGG pathways enriched for DEGs include metabolic pathways, oxidative phosphorylation, carbon metabolism and TCA cycle (Figure 5.15). Overall, these enriched GO terms and pathways point toward differences in the regulators of metabolic function in the three subclones. This is consistent with previous reports that suggest altered metabolism is implicated in drug sensitivity and melanoma resistance to BRAFi (Haq et al., 2013; Hardeman et al., 2017; Vazquez et al., 2013). Correlation of BRAFi insensitivity was computed for each identified DEG according to 10 cell line panel. Pair-wise comparisons of DEGs was was performed on genes (after low count genes were removed) using DESeq2 pipeline (Love et al., 2014).

RT-qPCR quantification of NOX5 expression

Total RNA was extracted using Trizol isolation method (Invitrogen) according to the manufacturers instructions [1]. cDNA synthesis was performed with QuantiTect Reverse Transcription Kit (Cat# 205311) from Qiagen. RT-qPCR was performed using the IQTM SYBR Green Supermix from BioRad (Cat# 1708880). Amplifications were performed in BioRad CFX96 TouchTM Real-Time PCR Detection System. All experiments were done at least in 3+ technical replicates. Log2 of the transcript expressions were normalized to their expressions to SKMEL5 subline SC01. HPRT or 36B4 were used as housekeeping gene in all the experiments.

Chapter 6

Conclusion

6.1 Discussion

A major theme throughout this work has been the idea of stability. Living systems all exist in a shared environment with other living systems, and continuous competition and evolution ensure an ever changing landscape that they all must contend with. Indeed, given the enormous combinatorial complexity of behaviors that are possible, one of the most paradoxical aspects of biology is that it can be understood at all; that it finds robust, stable patterns of behavior.

Stability likewise lies at the heart of dynamical systems theory in mathematics. Systems which may have arbitrarily complex behavior throughout the entire phase space often settle into subregions, where the behavior becomes more understandable, and more predictable.

Nevertheless, the second major theme throughout this work must be instability. Cancer has taught us that it won't be confined into a box easily. Cancer therapies, especially in small cell lung cancer, often have great effect at the start of treatment. And yet tumors are able to evolve, both genetically, and with growing appreciation throughout the field, epigenetically. In many ways, cancer exemplifies Bruce Lee's famous adage to be "formless, shapeless, like water." Cancer adapts to our therapies, whether passively or actively, and evades our attacks. The longterm goal of this work is to understand how to better control the way cancer adapts, and as much as possible, to make it adapt in ways that we choose.

In this effort, Waddington's original idea of a "creode" (Waddington, 1957) is exceptionally important. It suggests that not only are biological identities stable, but that transition pathways between distinct identities are themselves stable - some paths are easier to traverse than others, and once a cell starts down a path, it becomes (loosely) locked in. Understanding the paths that cancer takes then, and how those paths are regulated, could

have a dramatic impact on our ability to disrupt cancer's adaptability.

In chapters 2 and 3, I studied the stability of small cell lung cancer's phenotypic identity. In this work, I illustrated how existing SCLC heterogeneity may emerge spontaneously through pre-wired genetic regulatory programs, driven by master regulator transcription factors. I identified and characterized a novel SCLC subtype, NE Variant-2, which is distinguished phenotypically from the other subtypes as being exceptionally resistant to treatment. In chapter 2 I showed that under drug treatment, NE and ML (non-NE) cells shift toward the variant subtypes, however a full characterization of drug treated cells remains to be completed. In chapter 3, I identified possible master regulators of the NE Variant-2 subtype, and ongoing work in the lab is directed at reprogramming those cells into other states, to possibly increase their sensitivity. Nevertheless, to achieve clinical utility, these master regulators must be targeted through signaling pathways, and connecting the complexity of fast-acting signaling pathways with slow transcriptional cell identity programs remains a large challenge.

In chapters 4 and 5, I developed tools that can be used to understand and model a cellular population's response to treatment. In chapter 4, I extended the traditional model of exponential population growth to the case where the growth rate is distributed over subpopulations with stable identity. In doing so, I showed the significance of different parameters of the proliferation rate distribution, with respect to how they contribute to the overall relapse. Furthermore, I demonstrated the stability of the Gaussian distribution in drug treated populations. Because of this stability, the model can make reliable predictions, even if the DIP rate distribution is slightly non-normal: it will become more normal over time.

This derivation assumed that daughter cells will inherit their parent cell's proliferation rate, and this assumptions was shown to be reasonable in erlotinib treated PC9 cells. Nevertheless, over longer terms, or in other cancers, this assumption may be less likely to hold. Understanding the interplay between drug-induced phenotype transitions, and drug-

induced proliferation rates is an ongoing challenge in the lab, especially in melanoma, for which the DIP rates appear much less stable (Paudel et al., 2018).

Finally, I introduced a novel and unifying framework for modeling drug combination synergy: MuSyC. This too has its roots in stability, as the population is modeled as an equilibrium between multiple, distinct drugged states. MuSyC is powerful in its flexibility, but most importantly in its interpretability. Synergy parameters in MuSyC reflect real, geometrical features of the dose-response surface, and have unambiguous meaning. We demonstrated MuSyC's usefulness in drug combination screens in non-small-cell lung cancer, and in melanoma. Interestingly we were able to identify trends between drug classes and synergistic efficacy vs synergistic potency. A systematic and predictive connection between these tradeoffs would be a significant advance in targeted therapy.

Overall, each of the chapters in this work represent pillars, but the goal is to connect the pillars to arrive at a comprehensive and predictive framework for identifying holistic combination treatments that confine the tumor's response in a phenotypic region that it can be killed. Of course, the ultimate goal is the improvement of patients lives, and the eradication of this terrible disease, and it is my hope that this work represents a small step in that direction.

6.2 Future Directions

This work leads to several possible directions of future research.

6.2.1 SCLC

Our work in SCLC revealed distinct drug sensitivity patterns associated with distinct transcriptional patterns. Jahchan *et. al.* (Jahchan et al., 2016) identified a tumor propagating cell phenotype in a mouse model of SCLC that is capable of giving rise to several different cell types. Nevertheless it is not known how plastic these phenotypes are, whether all types may spontaneously transition or transdifferentiate into the others, nor how different

selective pressures within the tumor microenvironment may promote or inhibit transitions. Many standard SCLC cell lines have been established for several decades (Carney et al., 1985), and their phenotyptic traits are well conserved. Nevertheless, future work tracking single-cell fates over time of SCLC cells alone, in mixed populations, in response to perturbations such as drug treatment, or in *in vitro* xenograft models could quantify the relative baseline stability of these phenotypes in different contexts.

Furthermore, given recent proliferation of literature on SCLC heterogeneity (Calbo et al., 2011; George et al., 2015; Borromeo et al., 2016; Jahchan et al., 2016; Udyavar et al., 2017; Mollaoglu et al., 2017; Lim et al., 2017; Williamson et al., 2016; Huang et al., 2018), there is uncertainty within the field about what are the significant differences or similarities between these distinct phenotypes, a problem that is compounded by the use of diverse model systems spanning cell lines, mouse xenografts and genetically engineered mouse models, and human patients. An atlas showing transcriptional and ontological similarities and differences, spanning these phenotypes and model systems, would clarify to what extent results obtained in one system may be applicable to another.

We identified key TF drivers of the non-NE and 3 NE phenotypes. Future work is ongoing to explore the effects of targeted perturbation of these TFs. Many cellular reprogramming efforts are hindered by partially reprogrammed cells, and it is not known to what extent semi-stable states may exist in-between the phenotypes we have identified. Furthermore, it remains unclear whether TF perturbation will induce changes in cell surface markers, gene co-expression modules, and/or drug sensitivity. There may be SCLC-specific regulatory interactions that are not well reflected in the literature or databases. Experimental GRN reconstruction methods, such as ATAC-seq and DNA footprinting may be used to identify SCLC-specific regulation. Even if TF perturbations do induce transitions, it is often challenging to target TFs in clinical settings, while molecules in signaling pathways are generally more amenable targets. Identification of differentially active signaling pathways across the phenotypes may reveal novel, actionable targets. Coupling signaling pathways

to downstream activation or silencing of TF activity could provide further specificity for identifying targets. Mechanistic models coupling the dynamics and cross-talk of multiple signaling pathways with the TF GRN in figure 3.9 would require modeling approaches bridging the timescale differences between signaling networks and TF regulatory networks, but could result in systematic prediction of targets that could drive SCLC cells toward death states.

Furthermore, modern single-cell technologies are opening up the possibility to quantitatively sample the "position" of cells along their creodes, in a way that will greatly improve our ability to understand the regulation, and the stability, of these creodes (Herring et al., 2018). By treating them as timeseries data, pseudotime trajectories inferred from single-cell derived creodes may be used to better fit the rules governing transitions .

6.2.2 TF network modeling

The probabilistic, uncertainty Boolean network model approach developed in section 1.3.5 can also be expand to account for more complex situations. Because of the expected robustness in biological regulation, it may be possible to estimate regulatory output on leaves for which there are no constraining biological observations. Each leaf has N neighbors, corresponding to the N regulatory inputs to the target TF, and in cases for which all regulatory neighbors have similar output, it may be biologically reasonable to assume that the unconstrained leaf ought to behave similarly. This could be explored by comparing leaves' regulation to their neighbors across many different networks and datasets.

Additionally, knowledge about the family, structure, and molecular function of TFs, or knowledge of their protein-protein interactions, may be added to constrain their regulation. For instance, a regulatory function like $A = B \wedge C$ may reflect an underlying protein interaction, or coactivator activity of B and C. Such pre-defined knowledge may be useful to better constrain uncertain leaves.

This uncertainty network approach can also be used to identify functionally irrelevant

regulators, as well as detecting the absence of necessary regulators. The first may be observed when all pairs of leaves following a given input TF being ON or OFF are nearly identical, that is, whether the TF is ON or OFF makes no effective difference to the output. The latter case may be observed when two samples that constrain the same leaf differ as to their predicted output. For instance, if $A = f(B,C)$, and one sample is $(A_1,B_1,C_1) = (1,1,0)$, and another is $(A_2,B_2,C_2) = (0,1,0)$, then this suggests that A must be a function of at least one other TF.

Finally, the model presented here is specifically designed for non-perturbed, steady-state data. When validating model predictions, it may be useful to collect timeseries data or data corresponding to perturbed cells, and this data should be useful at further constraining the rules, especially in regions of expression space between attractors. Timeseries data may be included by adding a time dependent component, and allowing that $A = f(B,C)$ be replaced with $A(t_{i+1}) = f(B_i,C_i)$, such that the distribution of how each sample at each timepoint constrains the output TF at the next timepoint. Care should be taken in how timesteps are chosen experimentally. Data corresponding to cells with one or more perturbed TFs may be used, but can only constrain rules for those TFs which are not perturbed. While cells may get stuck in previously unaccounted attractors, adding these data to the model may therefore be used to troubleshoot and debug reprogramming experiments.

6.2.3 Heterogeneous proliferation model

The heterogeneous proliferation model derived in chapter 4 is applicable for systems with low phenotypic plasticity, however transitions between different phenotypes are likely to play a significant role in the relapse of more phenotypically plastic cancers. Phenotypic stability may be less in other cancer types, such as in (Paudel et al., 2018). Furthermore, spatial constraints and microenvironmental interactions may additionally induce phenotypic plasticity in *in vivo* systems. The stability of proliferation rates observed in figure 4.7 should be assessed in 3D culture and *in vivo*.

Plasticity may be reasonably modeled using a transition-proliferation model

$$\frac{\mathrm{d}x_i(t)}{\mathrm{d}t} = (r_i - \sum_j T_{i,j}) \cdot x_i(t) + \sum_j T_{j,i} x_j(t) \tag{6.1}$$

where $T_{i,j}$ quantifies the transition rate from phenotype i to phenotype j.

Additionally, it remains unclear how the proliferation rate connects to underlying molecular states. Understanding this relationship may reveal molecular correlates of DIP rate variance or skewness, and targeting these correlates may then be able to modulate the underlying distribution to delay the time-to-relapse.

6.2.4 Drug combination synergy

In chapter 5, we demonstrated the application of MuSyC to combinations of 2 drugs. Nevertheless, in many cases drug treatment may require administration of 3 or more compounds. In general for a combination of N drugs, MuSyC must fit $2^N - N - 1$ distinct efficacy synergies and $2^{N-1} \cdot N - N$ distinct potency and cooperativity synergies each. As discussed in chapter 5, for $N = 2$, this results in 1 efficacy synergy and 2 each of potency and cooperativity synergies. Nevertheless, for combinations of $N = 3$ drugs, this requires 4 distinct efficacy synergies, corresponding to the efficacy gains of combination of drugs (1,2), (2,3), (1,3), and (1,2,3), and 9 distinct potency and cooperativity synergies. Further work remains to understand to what extent all of these higher order synergies are important, as well as the best strategies for sampling drug-responses to optimally constrain the greatest amount of these parameters.

Additionally, clinically it is common to treat with one drug, then after a pause, treat with a second drug. The MuSyC model was designed under the assumption that equilibrium is reached for both drugs simultaneously, but in the time-staggered case, this assumption is not valid. While MuSyC's synergy parameters have clear, geometrical interpretations for the shape of the edges of the dose response surface, there may be significant differences in

the interior of the surface depending on the order of treatment.

The identification of NOX5 as anti-correlated with DIP rate following BRAF inhibition suggested combination of BRAFi + NOX5i, which we showed to be synergistic in Figure 5.15. This suggests that systematic connection of gene expression, molecular networks, and single-drug sensitivity may enable accurate prediction of synergistic combinations. To this end, connecting the gene co-expression modules from Figure 3.3 to the drug sensitivity in Figure 3.7 may be useful to identify synergistic combinations.

Prediction of synergistic targets in this way may further be improved by understanding which co-targets will result in synergistic efficacy VS other types. For instance, it may be that co-targeting two molecules in the same pathway tends to result in one type of synergy, while co-targeting disparate pathways, or biological processes, tends to result in another type. Mechanistic models of signaling pathways and crosstalk may be systematically explored with *in silico* perturbations to identify patterns that may result in one or another type of synergy.

BIBLIOGRAPHY

I. Albert, J. Thakar, S. Li, R. Zhang, and R. Albert. Boolean network simulations for life scientists. *Source Code for Biology and Medicine*, 3:16, 2008. ISSN 1751-0473. doi: 10.1186/1751-0473-3-16.

V. Almendro, Y.-K. Cheng, A. Randles, S. Itzkovitz, A. Marusyk, E. Ametller, X. Gonzalez-Farre, M. Muñoz, H. G. Russnes, A. Helland, I. H. Rye, A.-L. Borresen-Dale, R. Maruyama, A. van Oudenaarden, M. Dowsett, R. L. Jones, J. Reis-Filho, P. Gascon, M. Gönen, F. Michor, and K. Polyak. Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. *Cell Reports*, 6(3):514–527, 2014a. ISSN 2211-1247. doi: 10.1016/j.celrep.2013.12.041.

V. Almendro, H. J. Kim, Y.-K. Cheng, M. Gönen, S. Itzkovitz, P. Argani, A. van Oudenaarden, S. Sukumar, F. Michor, and K. Polyak. Genetic and phenotypic diversity in breast tumor metastases. *Cancer Res.*, 74(5):1338–1348, 3 2014b. ISSN 1538-7445. doi: 10.1158/0008-5472.CAN-13-2357-T.

U. Alon. *An introduction to systems biology: design principles of biological circuits*. CRC press, 2006.

American Cancer Society. Cancer Facts & Figures 2016, 8 2016.

American Cancer Society. Cancer Facts & Figures 2018. *American Cancer Society*, 2018. ISSN 15280020. doi: 10.1182/blood-2015-12-687814.

D. Angeli, J. E. Ferrell, and E. D. Sontag. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proceedings of the National Academy of Sciences of the United States of America*, 101(7):1822–1827, 2004.

F. Anokye-Danso, C. M. Trivedi, D. Juhr, M. Gupta, Z. Cui, Y. Tian, Y. Zhang, W. Yang, P. J. Gruber, J. A. Epstein, and E. E. Morrisey. Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell*, 8(4): 376–388, 4 2011. ISSN 1934-5909. doi: 10.1016/j.stem.2011.03.001.

M. N. Artyomov, A. Meissner, and A. K. Chakraborty. A model for genetic and epigenetic regulatory networks identifies rare pathways for transcription factor induced pluripotency. *PLoS Comput Biol*, 6(5):e1000785–e1000785, 2010.

A. Azzalini. The Skew-normal Distribution and Related Multivariate Families. *Scandinavian Journal of Statistics*, 32:159–188, 2005.

E. Ballestar and M. Esteller. Epigenetic gene regulation in cancer. *Advances in genetics*, 61:247–267, 2008.

C. R. S. Banerji, D. Miranda-Saavedra, S. Severini, M. Widschwendter, T. Enver, J. X. Zhou, and A. E. Teschendorff. Cellular network entropy as the energy potential in Waddington's differentiation landscape. *Scientific reports*, 3:3039, 2013. ISSN 2045-2322. doi: 10.1038/srep03039.

J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Kim, C. J. Wilson, J. Lehár, G. V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M. F. Berger, J. E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jané-Valbuena, F. A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I. H. Engels, J. Cheng, G. K. Yu, J. Yu, P. Aspesi, M. de Silva, K. Jagtap, M. D. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, R. C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J. P. Mesirov, S. B. Gabriel, G. Getz, K. Ardlie, V. Chan, V. E. Myer, B. L. Weber, J. Porter, M. Warmuth, P. Finan, J. L. Harris, M. Meyerson, T. R. Golub, M. P. Morrissey, W. R. Sellers, R. Schlegel, and L. A. Garraway. The Cancer Cell Line Encyclopedia

enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–607, 3 2012. ISSN 1476-4687. doi: 10.1038/nature11003.

A. S. Bhagwat and C. R. Vakoc. Targeting Transcription Factors in Cancer. *Trends in Cancer*, 1(1):53–65, 9 2015. ISSN 2405-8033. doi: 10.1016/j.trecan.2015.07.001.

S. Bhattacharya, Q. Zhang, and M. E. Andersen. A deterministic map of Waddington's epigenetic landscape for cell fate specification. *BMC systems biology*, 5(1):85, 1 2011. ISSN 1752-0509. doi: 10.1186/1752-0509-5-85.

J. X. Binder, S. Pletscher-Frankild, K. Tsafou, C. Stolte, S. I. O'Donoghue, R. Schneider, and L. J. Jensen. COMPARTMENTS: Unification and visualization of protein subcellular localization evidence. *Database*, 2014. ISSN 17580463. doi: 10.1093/database/bau012.

C. I. BLISS. THE TOXICITY OF POISONS APPLIED JOINTLY1. *Annals of Applied Biology*, 26(3):585–615, 8 1939. ISSN 00034746. doi: 10.1111/j.1744-7348.1939.tb06990.x.

M. D. Borromeo, T. K. Savage, R. K. Kollipara, M. He, A. Augustyn, J. K. Osborne, L. Girard, J. D. Minna, A. F. Gazdar, M. H. Cobb, and J. E. Johnson. ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs. *Cell Reports*, 2016. ISSN 22111247. doi: 10.1016/j.celrep.2016.06.081.

P. C. Boutros, S. K. Lau, M. Pintilie, N. Liu, F. A. Shepherd, S. D. Der, M.-S. Tsao, L. Z. Penn, and I. Jurisica. Prognostic gene signatures for non-small-cell lung cancer. *PNAS*, 106(8):2824–2828, 2 2009. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0809444106.

I. Bozic, B. Allen, and M. A. Nowak. Dynamics of targeted cancer therapy, 2012. ISSN 14714914.

N. E. Buchler, U. Gerland, and T. Hwa. On schemes of combinatorial transcription logic. *Proceedings of the National Academy of Sciences*, 100(9):5136–5141, 2003.

Y. Buganim, D. A. Faddah, and R. Jaenisch. Mechanisms and models of somatic cell reprogramming. *Nature Reviews Genetics*, 14(6):427–439, 2013.

P. Cahan, H. Li, S. A. Morris, E. L. da Rocha, G. Q. Daley, and J. J. Collins. CellNet: network biology applied to stem cell engineering. *Cell*, 158(4):903–915, 2014.

J. Calbo, E. van Montfort, N. Proost, E. van Drunen, H. B. Beverloo, R. Meuwissen, and A. Berns. A functional role for tumor cell heterogeneity in a mouse model of small cell lung cancer. *Cancer Cell*, 19(2):244–256, 2011. ISSN 1878-3686. doi: 10.1016/j.ccr.2010.12.021.

D. N. Carney, A. F. Gazdar, G. Bepler, J. G. Guccion, P. J. Marangos, T. W. Moody, M. H. Zweig, and J. D. Minna. Establishment and identification of small cell lung cancer cell lines having classic and variant features. *Cancer Research*, 45(6):2913–2923, 1985. ISSN 0008-5472.

M. S. Carro, W. K. Lim, M. J. Alvarez, R. J. Bollo, X. Zhao, E. Y. Snyder, E. P. Sulman, S. L. Anne, F. Doetsch, H. Colman, A. Lasorella, K. Aldape, A. Califano, and A. Iavarone. The transcriptional network for mesenchymal transformation of brain tumours. *Nature*, 463(7279):318–325, 2010. ISSN 1476-4687. doi: 10.1038/nature08712.

B. A. Chan and B. G. M. Hughes. Targeted therapy for non-small cell lung cancer: current standards and the promise of the future. *Translational lung cancer research*, 2015. ISSN 2218-6751. doi: 10.3978/j.issn.2218-6751.2014.05.01.

E. Y. Chen, C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark, and A. Ma'ayan. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*, 14:128, 2013. ISSN 1471-2105. doi: 10.1186/1471-2105-14-128.

A. B. C. Cherry and G. Q. Daley. Reprogramming cellular identity for regenerative medicine. *Cell*, 148(6):1110–1122, 2012.

M. Choi, J. Shi, S. H. Jung, X. Chen, and K.-H. Cho. Attractor landscape analysis reveals feedback loops in the p53 network that control the cellular response to DNA damage. *Science signaling*, 5:ra83, 2012. ISSN 1937-9145. doi: 10.1126/scisignal.2003363.

T. C. Chou and P. Talalay. Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. *Advances in enzyme regulation*, 22:27–55, 1984. ISSN 0065-2571.

C. L. C. G. P. (CLCGP) and N. G. M. (NGM). A genomics-based classification of human lung tumors. *Science Translational Medicine*, 5(209):209ra153, 2013. ISSN 1946-6242. doi: 10.1126/scitranslmed.3006802.

M. Cokol, H. N. Chua, M. Tasan, B. Mutlu, Z. B. Weinstein, Y. Suzuki, M. E. Nergiz, M. Costanzo, A. Baryshnikova, G. Giaever, C. Nislow, C. L. Myers, B. J. Andrews, C. Boone, and F. P. Roth. Systematic exploration of synergistic drug pairs. *Molecular Systems Biology*, 7(1):544, 11 2011. ISSN 1744-4292. doi: 10.1038/msb.2011.71.

E. P. Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science (New York, N.Y.)*, 306(5696):636–640, 2004. ISSN 1095-9203. doi: 10.1126/science.1105136.

I. Crespo and A. Del Sol. A general strategy for cellular reprogramming: The importance of transcription factor cross-repression. *Stem cells*, 31(10):2127–2135, 2013.

I. Crespo, T. M. Perumal, W. Jurkowski, and A. Del Sol. Detecting cellular reprogramming determinants by differential stability analysis of gene regulatory networks. *BMC systems biology*, 7(1):140, 2013.

X. Dai, Y. Li, Z. Bai, and X.-Q. Tang. Molecular portraits revealing the heterogeneity of

breast tumor subtypes defined using immunohistochemistry markers. *Scientific Reports*, 5:14499, 9 2015. ISSN 2045-2322. doi: 10.1038/srep14499.

H. de Jong. Modeling and simulation of genetic regulatory systems: a literature review. *Journal of computational biology : a journal of computational molecular cell biology*, 9 (1):67–103, 1 2002. ISSN 1066-5277. doi: 10.1089/10665270252833208.

B. Derrida and Y. Pomeau. Random networks of automata: A simple annealed approximation. *Europhys Lett*, pages 45–49, 1986.

R. Derynck and R. J. Akhurst. Differentiation plasticity regulated by TGF-beta family proteins in development and disease. *Nature Cell Biology*, 9(9):1000–1004, 2007.

V. T. DeVita and E. Chu. A history of cancer chemotherapy, 2008. ISSN 00085472.

J. M. Dowen, Z. P. Fan, D. Hnisz, G. Ren, B. J. Abraham, L. N. Zhang, A. S. Weintraub, J. Schujiers, T. I. Lee, K. Zhao, and R. A. Young. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*, 159(2):374–387, 2014. ISSN 1097-4172. doi: 10.1016/j.cell.2014.09.030.

Q. Duan, C. Flynn, M. Niepel, M. Hafner, J. L. Muhlich, N. F. Fernandez, A. D. Rouillard, C. M. Tan, E. Y. Chen, T. R. Golub, P. K. Sorger, A. Subramanian, and A. Ma'ayan. LINCS Canvas Browser: interactive web app to query, browse and interrogate LINCS L1000 gene expression signatures. *Nucleic Acids Res.*, 42(Web Server issue):W449–460, 7 2014. ISSN 1362-4962. doi: 10.1093/nar/gku476.

Z. Eroglu and A. Ribas. Combination therapy with BRAF and MEK inhibitors for melanoma: latest evidence and place in therapy. *Therapeutic advances in medical oncology*, 8(1):48–56, 1 2016. ISSN 1758-8340. doi: 10.1177/1758834015616934.

D. S. Ettinger, D. E. Wood, D. L. Aisner, W. Akerley, J. Bauman, L. R. Chirieac, T. A. D'Amico, M. M. DeCamp, T. J. Dilling, M. Dobelbower, R. C. Doebele, R. Govindan,

M. A. Gubens, M. Hennon, L. Horn, R. Komaki, R. P. Lackner, M. Lanuti, T. A. Leal, L. J. Leisch, R. Lilenbaum, J. Lin, B. W. Loo, R. Martins, G. A. Otterson, K. Reckamp, G. J. Riely, S. E. Schild, T. A. Shapiro, J. Stevenson, S. J. Swanson, K. Tauer, S. C. Yang, K. Gregory, and M. Hughes. Non-Small Cell Lung Cancer, Version 5.2017, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network : JNCCN*, 15(4):504–535, 4 2017. ISSN 1540-1413. doi: 10.6004/JNCCN.2017.0050.

M. Fallahi-Sichani, S. Honarnejad, L. M. Heiser, J. W. Gray, and P. K. Sorger. Metrics other than potency reveal systematic variation in responses to cancer drugs. *Nature Chemical Biology*, 9(11):708–714, 9 2013. ISSN 1552-4450. doi: 10.1038/nchembio.1337.

J. E. Ferrell. Bistability, bifurcations, and Waddington's epigenetic landscape, 2012. ISSN 09609822.

B. Fischer, M. Marinov, and A. Arcaro. Targeting receptor tyrosine kinase signalling in small cell lung cancer (SCLC): what have we learned so far? *Cancer Treatment Reviews*, 33(4):391–406, 2007. ISSN 0305-7372. doi: 10.1016/j.ctrv.2007.01.006.

R. Fisher, L. Pusztai, and C. Swanton. Cancer heterogeneity: implications for targeted therapeutics. *British Journal of Cancer*, 108(3):479–485, 2013. ISSN 1532-1827. doi: 10.1038/bjc.2012.581.

R. A. Fisher. *The Genetical Theory Of Natural Selection*. The Clarendon Press, 1930. ISBN 9781406707830.

D. V. Foster, S. Huang, S. A. Kauffman, and J. G. Foster. Sequential Branch Points in Hierarchical Cell Fate Determination. *Journal of Theoretical Biology*, 260:589–597, 2009. ISSN 00225193. doi: 10.1016/j.jtbi.2009.07.005.

J. Foucquier and M. Guedj. Analysis of drug combinations: current methodological land-

scape. *Pharmacology research & perspectives*, 3(3):e00149, 6 2015. ISSN 2052-1707. doi: 10.1002/prp2.149.

P. Frick, B. Paudel, D. Tyson, and V. Quaranta. Quantifying heterogeneity and dynamics of clonal fitness in response to perturbation. *Journal of cellular physiology*, 230(7): 1403–1412, 8 2015. ISSN 0021-9541. doi: 10.1002/jcp.24888.

L. Gautier, L. Cope, B. M. Bolstad, and R. A. Irizarry. affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics (Oxford, England)*, 20(3):307–315, 2004. ISSN 1367-4803. doi: 10.1093/bioinformatics/btg405.

J. George, J. S. Lim, S. J. Jang, Y. Cun, L. Ozretić, G. Kong, F. Leenders, X. Lu, L. Fernández-Cuesta, G. Bosco, C. Müller, I. Dahmen, N. S. Jahchan, K.-S. Park, D. Yang, A. N. Karnezis, D. Vaka, A. Torres, M. S. Wang, J. O. Korbel, R. Menon, S.-M. Chun, D. Kim, M. Wilkerson, N. Hayes, D. Engelmann, B. Pützer, M. Bos, S. Michels, I. Vlasic, D. Seidel, B. Pinther, P. Schaub, C. Becker, J. Altmüller, J. Yokota, T. Kohno, R. Iwakawa, K. Tsuta, M. Noguchi, T. Muley, H. Hoffmann, P. A. Schnabel, I. Petersen, Y. Chen, A. Soltermann, V. Tischler, C.-m. Choi, Y.-H. Kim, P. P. Massion, Y. Zou, D. Jovanovic, M. Kontic, G. M. Wright, P. A. Russell, B. Solomon, I. Koch, M. Lindner, L. A. Muscarella, A. la Torre, J. K. Field, M. Jakopovic, J. Knezevic, E. Castaños-Vélez, L. Roz, U. Pastorino, O.-T. Brustugun, M. Lund-Iversen, E. Thunnissen, J. Köhler, M. Schuler, J. Botling, M. Sandelin, M. Sanchez-Cespedes, H. B. Salvesen, V. Achter, U. Lang, M. Bogus, P. M. Schneider, T. Zander, S. Ansén, M. Hallek, J. Wolf, M. Vingron, Y. Yatabe, W. D. Travis, P. Nürnberg, C. Reinhardt, S. Perner, L. Heukamp, R. Büttner, S. A. Haas, E. Brambilla, M. Peifer, J. Sage, and R. K. Thomas. Comprehensive genomic profiles of small cell lung cancer. *Nature*, 524(7563):47–53, 2015. ISSN 1476-4687. doi: 10.1038/nature14664.

M. Gerlinger, A. J. Rowan, S. Horswell, J. Larkin, D. Endesfelder, E. Gronroos, P. Martinez, N. Matthews, A. Stewart, P. Tarpey, I. Varela, B. Phillimore, S. Begum, N. Q.

McDonald, A. Butler, D. Jones, K. Raine, C. Latimer, C. R. Santos, M. Nohadani, A. C. Eklund, B. Spencer-Dene, G. Clark, L. Pickering, G. Stamp, M. Gore, Z. Szallasi, J. Downward, P. A. Futreal, and C. Swanton. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *New England Journal of Medicine*, 366 (10):883–892, 3 2012. ISSN 0028-4793. doi: 10.1056/NEJMoa1113205.

S. F. Gilbert. Developmental plasticity and developmental symbiosis: the return of eco-devo. *Current topics in developmental biology*, 116:415–433, 2016.

S. F. Gilbert and M. J. F. Barresi. *Developmental Biology*. Oxford University Press, 2017. ISBN 978-1-60535-738-6.

Z. Gong, G. Hu, Q. Li, Z. Liu, F. Wang, X. Zhang, J. Xiong, P. Li, Y. Xu, R. Ma, S. Chen, and J. Li. Compound Libraries: Recent Advances and Their Applications in Drug Discovery. *Current Drug Discovery Technologies*, 14(4):216–228, 10 2017. ISSN 15701638. doi: 10.2174/1570163814666170425155154.

M. Greaves and C. C. Maley. Clonal evolution in cancer. *Nature*, 481(7381):306–313, 2012.

W. Greco, H. Unkelbach, G. Poch, J. Suhnel, M. Kundi, and W. Bodeker. Consensus on Concepts and Terminology for Combined-action Assessment: The Saariselka Agreement. *ACES*, 4(3):65–69, 1992.

W. R. Greco, G. Bravo, and J. C. Parsons. The search for synergy: a critical review from a response surface perspective. *Pharmacological reviews*, 47(2):331–85, 6 1995. ISSN 0031-6997.

G. Greve, I. Schiffmann, D. Pfeifer, M. Pantic, J. Schüler, and M. Lübbert. The pan-HDAC inhibitor panobinostat acts as a sensitizer for erlotinib activity in EGFR-mutated and -wildtype non-small cell lung cancer cells. *BMC cancer*, 15:947, 12 2015. ISSN 1471-2407. doi: 10.1186/s12885-015-1967-5.

C. Gros. *Complex and Adaptive Dynamical Systems*. Springer, 2008.

M. Hafner, M. Niepel, M. Chung, and P. K. Sorger. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nature Methods*, 13(6):521–527, 5 2016. ISSN 1548-7091. doi: 10.1038/nmeth.3853.

D. Hanahan and R. A. Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5): 646–74, 3 2011. ISSN 1097-4172. doi: 10.1016/j.cell.2011.02.013.

C. L. Hann and C. M. Rudin. Fast, hungry and unstable: finding the Achilles' heel of small-cell lung cancer. *Trends in Molecular Medicine*, 13(4):150–157, 2007. ISSN 1471-4914. doi: 10.1016/j.molmed.2007.02.003.

J. Hanna, K. Saha, B. Pando, J. Van Zon, C. J. Lengner, M. P. Creyghton, A. van Oudenaarden, and R. Jaenisch. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature*, 462(7273):595–601, 2009.

R. Haq, J. Shoag, P. Andreu-Perez, S. Yokoyama, H. Edelman, G. C. Rowe, D. T. Frederick, A. D. Hurley, A. Nellore, A. L. Kung, J. A. Wargo, J. S. Song, D. E. Fisher, Z. Arany, and H. R. Widlund. Oncogenic BRAF regulates oxidative metabolism via PGC1α and MITF. *Cancer cell*, 23(3):302–15, 3 2013. ISSN 1878-3686. doi: 10.1016/j.ccr.2013.02.003.

K. N. Hardeman, C. Peng, B. B. Paudel, C. T. Meyer, T. Luong, D. R. Tyson, J. D. Young, V. Quaranta, and J. P. Fessel. Dependence On Glycolysis Sensitizes BRAF-mutated Melanomas For Increased Response To Targeted BRAF Inhibition. *Scientific reports*, 7: 42604, 2 2017. ISSN 2045-2322. doi: 10.1038/srep42604.

L. A. Harris, P. L. Frick, S. P. Garbett, K. N. Hardeman, B. B. Paudel, C. F. Lopez, V. Quaranta, and D. R. Tyson. An unbiased metric of antiproliferative drug effect in vitro. *Nature Methods*, 13(6):497–500, 5 2016. ISSN 1548-7091. doi: 10.1038/nmeth.3852.

D. F. Hayes and C. Paoletti. Circulating tumour cells: insights into tumour heterogeneity. *J. Intern. Med.*, 274(2):137–143, 8 2013. ISSN 1365-2796. doi: 10.1111/joim.12047.

B. He, C. Lu, G. Zheng, X. He, M. Wang, G. Chen, G. Zhang, and A. Lu. Combination therapeutics in complex diseases. *Journal of cellular and molecular medicine*, 20(12): 2231–2240, 2016. ISSN 1582-4934. doi: 10.1111/jcmm.12930.

C. A. Herring, A. Banerjee, E. T. McKinley, A. J. Simmons, J. Ping, J. T. Roland, J. L. Franklin, Q. Liu, M. J. Gerdes, R. J. Coffey, and K. S. Lau. Unsupervised Trajectory Analysis of Single-Cell RNA-Seq and Imaging Data Reveals Alternative Tuft Cell Origins in the Gut. *Cell Systems*, 2018. ISSN 24054720. doi: 10.1016/j.cels.2017.10.012.

T. J. N. Hiltermann, M. M. Pore, A. van den Berg, W. Timens, H. M. Boezen, J. J. W. Liesker, J. H. Schouwink, W. J. A. Wijnands, G. S. M. A. Kerner, F. A. E. Kruyt, H. Tissing, A. G. J. Tibbe, L. W. M. M. Terstappen, and H. J. M. Groen. Circulating tumor cells in small-cell lung cancer: a predictive and prognostic factor. *Annals of Oncology*, 2012. ISSN 0923-7534. doi: 10.1093/annonc/mds138.

M. Horie, A. Saito, M. Ohshima, H. I. Suzuki, and T. Nagase. YAP and TAZ modulate cell phenotype in a subset of small cell lung cancer. *Cancer Science*, 107(12):1755–1766, 12 2016. doi: 10.1111/cas.13078.

S. Hu, K. D. Wilson, Z. Ghosh, L. Han, Y. Wang, F. Lan, K. J. Ransohoff, and J. C. Wu. MicroRNA-302 Increases Reprogramming Efficiency via Repression of NR2F2. *Stem Cells*, 31(2):259–268, 2 2013. ISSN 1066-5099. doi: 10.1002/stem.1278.

M. Huang, A. Shen, J. Ding, and M. Geng. Molecularly targeted cancer therapy: Some lessons from the past decade, 2014. ISSN 01656147.

S. Huang. Regulation of cellular states in mammalian cells from a genomewide view. *Gene Regulation and Metabolism: Postgenomic Computational Approaches*, pages 181–220, 2002.

S. Huang. Reprogramming cell fates: Reconciling rarity with robustness. *BioEssays*, 31: 546–560, 2009. ISSN 02659247. doi: 10.1002/bies.200800189.

S. Huang. The molecular and mathematical basis of Waddington's epigenetic landscape: a framework for post-Darwinian biology? *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 34(2):149–157, 2012. ISSN 1521-1878. doi: 10.1002/bies.201100031.

S. Huang, G. Eichler, Y. Bar-Yam, and D. E. Ingber. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Physical review letters*, 94(12):128701, 2005.

S. Huang, Y.-P. Guo, G. May, and T. Enver. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental biology*, 305(2):695–713, 2007.

S. Huang, I. Ernberg, and S. Kauffman. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Seminars in Cell & Developmental Biology*, 20(7):869–876, 2009. ISSN 1096-3634. doi: 10.1016/j.semcdb.2009.07.003.

Y. H. Huang, O. Klingbeil, X. Y. He, X. S. Wu, G. Arun, B. Lu, T. D. Somerville, J. P. Milazzo, J. E. Wilkinson, O. E. Demerdash, D. L. Spector, M. Egeblad, J. Shi, and C. R. Vakoc. POU2F3 is a master regulator of a tuft cell-like variant of small cell lung cancer. *Genes and Development*, 2018. ISSN 15495477. doi: 10.1101/gad.314815.118.

W. S. Hwang, Y. J. Ryu, J. H. Park, E. S. Park, E. G. Lee, J. M. Koo, H. Y. Jeon, B. C. Lee, S. K. Kang, S. J. Kim, and others. Evidence of a pluripotent human embryonic stem cell line derived from a cloned blastocyst. *Science*, 303(5664):1669–1674, 2004.

J. Inglese, R. L. Johnson, A. Simeonov, M. Xia, W. Zheng, C. P. Austin, and D. S. Auld. High-throughput screening assays for the identification of chemical probes. *Nature*

Chemical Biology, 3(8):466–479, 8 2007. ISSN 1552-4450. doi: 10.1038/nchembio.2007.17.

J. M. Irish and D. B. Doxie. High-dimensional single-cell cancer biology. *Current Topics in Microbiology and Immunology*, 377:1–21, 2014. ISSN 0070-217X. doi: 10.1007/82{_}2014{_}367.

N. S. Jahchan, J. S. Lim, B. Bola, K. Morris, G. Seitz, K. Q. Tran, L. Xu, F. Trapani, C. J. Morrow, S. Cristea, G. L. Coles, D. Yang, D. Vaka, M. S. Kareta, J. George, P. K. Mazur, T. Nguyen, W. C. Anderson, S. J. Dylla, F. Blackhall, M. Peifer, C. Dive, and J. Sage. Identification and Targeting of Long-Term Tumor-Propagating Cells in Small Cell Lung Cancer. *Cell Reports*, 16(3):644–656, 2016. ISSN 2211-1247. doi: 10.1016/j.celrep.2016.06.021.

W. Janzen. Screening Technologies for Small Molecule Discovery: The State of the Art. *Chemistry & Biology*, 21(9):1162–1170, 9 2014. ISSN 1074-5521. doi: 10.1016/J.CHEMBIOL.2014.07.015.

V. Jaquet, J. Marcoux, E. Forest, K. G. Leidal, S. McCormick, Y. Westermaier, R. Perozzo, O. Plastre, L. Fioraso-Cartier, B. Diebold, L. Scapozza, W. M. Nauseef, F. Fieschi, K.-H. Krause, and K. Bedard. NADPH oxidase (NOX) isoforms are inhibited by celastrol with a dual mode of action. *British journal of pharmacology*, 164(2b):507–20, 9 2011. ISSN 1476-5381. doi: 10.1111/j.1476-5381.2011.01439.x.

M. K. Jolly, M. Boareto, B. Huang, D. Jia, M. Lu, E. Ben-Jacob, J. N. Onuchic, and H. Levine. Implications of the Hybrid Epithelial/Mesenchymal Phenotype in Metastasis. *Frontiers in Oncology*, 5:155, 2015. ISSN 2234-943X. doi: 10.3389/fonc.2015.00155.

M. H. Jones, C. Virtanen, D. Honjoh, T. Miyoshi, Y. Satoh, S. Okumura, K. Nakagawa, H. Nomura, and Y. Ishikawa. Two prognostically significant subtypes of high-grade lung neuroendocrine tumours independent of small-cell and large-cell neuroendocrine

carcinomas identified by gene expression profiles. *Lancet*, 2004. ISSN 01406736. doi: 10.1016/S0140-6736(04)15693-6.

W. D. Joo, I. Visintin, and G. Mor. Targeted cancer therapy - Are the days of systemic chemotherapy numbered?, 2013. ISSN 03785122.

C. T. Jordan, M. L. Guzman, and M. Noble. Cancer stem cells. *New England Journal of Medicine*, 355(12):1253–1261, 2006.

J. Jourquin, D. Duncan, Z. Shi, and B. Zhang. GLAD4U: deriving and prioritizing gene lists from PubMed literature. *BMC genomics*, 13 Suppl 8:S20, 2012. ISSN 1471-2164. doi: 10.1186/1471-2164-13-S8-S20.

A. Kamburov, U. Stelzl, H. Lehrach, and R. Herwig. The ConsensusPathDB interaction database: 2013 Update. *Nucleic Acids Research*, 2013. ISSN 03051048. doi: 10.1093/nar/gks1055.

K. Kaneko. *Life: an introduction to complex systems biology*. Springer, 2006.

S. Kauffman. Homeostasis and differentiation in random genetic control networks. *Nature*, 224:177–178, 1969a.

S. Kauffman. Differentiation of malignant to benign cells. *Journal of Theoretical Biology*, 31(3):429–451, 1971. ISSN 0022-5193.

S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of theoretical biology*, 22(3):437–467, 1969b.

G. Kaur, R. A. Reinhart, A. Monks, D. Evans, J. Morris, E. Polley, and B. A. Teicher. Bromodomain and hedgehog pathway targets in small cell lung cancer. *Cancer Letters*, 2016. ISSN 18727980. doi: 10.1016/j.canlet.2015.12.001.

H. Kawaji, J. Severin, M. Lizio, A. Waterhouse, S. Katayama, K. M. Irvine, D. A. Hume, A. R. R. Forrest, H. Suzuki, P. Carninci, and others. The FANTOM web resource: from

mammalian transcriptional landscape to its dynamic regulation. *Genome biology*, 10(4): R40, 2009.

J. F. Kenney. *Mathematics of Statistics*. Number pt. 2 in Mathematics of Statistics. D. Van Nostrand Company, Inc., 1939.

A. S. Khalil and J. J. Collins. Synthetic biology: applications come of age. *Nature Reviews Genetics*, 11(5):367–379, 2010.

D. Kim, B. Langmead, and S. L. Salzberg. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12(4):357–360, 4 2015. ISSN 1548-7091. doi: 10.1038/nmeth.3317.

K. Y. Kim and J. Wang. Potential energy landscape and robustness of a gene regulatory network: Toggle switch. *PLoS Computational Biology*, 3:565–577, 2007. ISSN 1553734X. doi: 10.1371/journal.pcbi.0030060.

C. Koga, S. Kobayashi, H. Nagano, Y. Tomimaru, N. Hama, H. Wada, K. Kawamoto, H. Eguchi, M. Konno, H. Ishii, K. Umeshita, Y. Doki, and M. Mori. Reprogramming Using microRNA-302 Improves Drug Sensitivity in Hepatocellular Carcinoma Cells. *Ann Surg Oncol*, 21(4):591–600, 12 2014. ISSN 1068-9265, 1534-4681. doi: 10.1245/s10434-014-3705-7.

N. Kotecha, P. O. Krutzik, and J. M. Irish. Web-based analysis and publication of flow cytometry experiments. *Current Protocols in Cytometry*, Chapter 10:Unit10.17, 2010. ISSN 1934-9300. doi: 10.1002/0471142956.cy1017s53.

T. Kottke, N. Boisgerault, R. M. Diaz, O. Donnelly, D. Rommelfanger-Konkol, J. Pulido, J. Thompson, D. Mukhopadhyay, R. Kaspar, M. Coffey, H. Pandha, A. Melcher, K. Harrington, P. Selby, and R. Vile. Detecting and targeting tumor relapse by its resistance to innate effectors at early recurrence. *Nat Med*, 19(12):1625–1631, 12 2013. ISSN 1078-8956. doi: 10.1038/nm.3397.

A. Lachmann, H. Xu, J. Krishnan, S. I. Berger, A. R. Mazloom, and A. Ma'ayan. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics (Oxford, England)*, 26(19):2438–2444, 2010. ISSN 1367-4811. doi: 10.1093/bioinformatics/btq466.

J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935, 9 2006. ISSN 1095-9203. doi: 10.1126/science.1132939.

A. H. Lang, H. Li, J. J. Collins, and P. Mehta. Epigenetic Landscapes Explain Partially Reprogrammed Cells and Identify Key Reprogramming Genes. *PLoS computational biology*, 10:e1003734, 2014. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1003734.

P. Langfelder and S. Horvath. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, 9:559, 2008. ISSN 1471-2105. doi: 10.1186/1471-2105-9-559.

V. S. LeBleu, J. T. OConnell, K. N. Gonzalez Herrera, H. Wikman, K. Pantel, M. Haigis, F. M. de Carvalho, A. Damascena, L. T. Domingos Chinen, R. M. Rocha, J. M. Asara, and R. Kalluri. PGC-1α mediates mitochondrial biogenesis and oxidative phosphorylation in cancer cells to promotemetastasis. *Nature Cell Biology*, 16(10):992–1003, 10 2014. ISSN 1465-7392. doi: 10.1038/ncb3039.

A. Lecharpentier, P. Vielh, P. Perez-Moreno, D. Planchard, J. C. Soria, and F. Farace. Detection of circulating tumour cells with a hybrid (epithelial/mesenchymal) phenotype in patients with metastatic non-small cell lung cancer. *British Journal of Cancer*, 105(9):1338–1341, 2011. ISSN 1532-1827. doi: 10.1038/bjc.2011.405.

C. Li and J. Wang. Quantifying Cell Fate Decisions for Differentiation and Reprogramming of a Human Stem Cell Network: Landscape and Biological Paths. *PLoS Computational Biology*, 9, 2013a. ISSN 1553734X. doi: 10.1371/journal.pcbi.1003165.

C. Li and J. Wang. Quantifying Waddington landscapes and paths of non-adiabatic cell fate decisions for differentiation, reprogramming and transdifferentiation. *Journal of the Royal Society, Interface / the Royal Society*, 10:20130787, 2013b. ISSN 1742-5662. doi: 10.1098/rsif.2013.0787.

C. Li and J. Wang. Landscape and flux reveal a new global view and physical quantification of mammalian cell cycle. *Proceedings of the National Academy of Sciences*, 111(39): 14130–14135, 2014.

S. Li, X. Zhu, B. Liu, G. Wang, and P. Ao. Endogenous molecular network reveals two mechanisms of heterogeneity within gastric cancer. *Oncotarget*, 6(15):13607–13627, 2015. ISSN 1949-2553. doi: 10.18632/oncotarget.3633.

Y. Liao, G. K. Smyth, and W. Shi. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–930, 4 2014. ISSN 1367-4803. doi: 10.1093/bioinformatics/btt656.

J. S. Lim, A. Ibaseta, M. M. Fischer, B. Cancilla, G. OYoung, S. Cristea, V. C. Luca, D. Yang, N. S. Jahchan, C. Hamard, M. Antoine, M. Wislez, C. Kong, J. Cain, Y.-W. Liu, A. M. Kapoun, K. C. Garcia, T. Hoey, C. L. Murriel, and J. Sage. Intratumoral heterogeneity generated by Notch signaling promotes small cell lung cancer. *Nature*, 545(7654):360–364, 8 2017. ISSN 0028-0836. doi: 10.1038/nature22323.

S.-L. Lin, D. C. Chang, S. Chang-Lin, C.-H. Lin, D. T. S. Wu, D. T. Chen, and S.-Y. Ying. Mir-302 reprograms human skin cancer cells into a pluripotent ES-cell-like state. *RNA*, 14(10):2115–2124, 10 2008. ISSN 1355-8382, 1469-9001. doi: 10.1261/rna.1162708.

S. Loewe. über Kombination swirkungen. *Arch fur Exp Pathology*, 114:313–326, 1926.

S. Loewe. Versuch einer allgemeinen Pharmakologie der Arznei- kombinationen. *Klinische Wochenschrift*, 6(23):1078–1085, 1927.

G. V. Long, D. Stroyakovskiy, H. Gogas, E. Levchenko, F. de Braud, J. Larkin, C. Garbe, T. Jouary, A. Hauschild, J. J. Grob, V. Chiarion Sileni, C. Lebbe, M. Mandalà, M. Millward, A. Arance, I. Bondarenko, J. B. Haanen, J. Hansson, J. Utikal, V. Ferraresi, N. Kovalenko, P. Mohr, V. Probachai, D. Schadendorf, P. Nathan, C. Robert, A. Ribas, D. J. DeMarini, J. G. Irani, M. Casey, D. Ouellet, A.-M. Martin, N. Le, K. Patel, and K. Flaherty. Combined BRAF and MEK Inhibition versus BRAF Inhibition Alone in Melanoma. *New England Journal of Medicine*, 371(20):1877–1888, 11 2014. ISSN 0028-4793. doi: 10.1056/NEJMoa1406037.

M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12):550, 2014. ISSN 1474-760X. doi: 10.1186/s13059-014-0550-8.

W. Lu, Y. Hu, G. Chen, Z. Chen, H. Zhang, F. Wang, L. Feng, H. Pelicano, H. Wang, M. J. Keating, J. Liu, W. McKeehan, H. Wang, Y. Luo, and P. Huang. Novel Role of NOX in Supporting Aerobic Glycolysis in Cancer Cells with Mitochondrial Dysfunction and as a Potential Target for Cancer Therapy. *PLoS Biology*, 10(5):e1001326, 5 2012. ISSN 1545-7885. doi: 10.1371/journal.pbio.1001326.

B. D. MacArthur, C. P. Please, and R. O. Oreffo. Stochasticity and the molecular mechanisms of induced pluripotency. *PloS one*, 3(8):e3086, 2008.

A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. Dalla Favera, and A. Califano. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC bioinformatics*, 7 Suppl 1:S7, 2006. ISSN 1471-2105. doi: 10.1186/1471-2105-7-S1-S7.

A. Marusyk, V. Almendro, and K. Polyak. Intra-tumour heterogeneity: a looking glass

for cancer? *Nature Reviews. Cancer*, 12(5):323–334, 2012. ISSN 1474-1768. doi: 10.1038/nrc3261.

A. Mathelier, X. Zhao, A. W. Zhang, F. Parcy, R. Worsley-Hunt, D. J. Arenillas, S. Buchman, C.-y. Chen, A. Chou, H. Ienasescu, J. Lim, C. Shyr, G. Tan, M. Zhou, B. Lenhard, A. Sandelin, and W. W. Wasserman. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Research*, 42(Database issue):142–147, 2014. ISSN 1362-4962. doi: 10.1093/nar/gkt997.

V. Matys, E. Fricke, R. Geffers, E. Gössling, M. Haubrock, R. Hehl, K. Hornischer, D. Karas, A. E. Kel, O. V. Kel-Margoulis, D.-U. Kloos, S. Land, B. Lewicki-Potapov, H. Michael, R. Münch, I. Reuter, S. Rotert, H. Saxel, M. Scheer, S. Thiele, and E. Wingender. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Research*, 31(1):374–378, 2003. ISSN 1362-4962.

F. Mbeunkui and D. J. Johann Jr. Cancer and the tumor microenvironment: a review of an essential relationship. *Cancer chemotherapy and pharmacology*, 63(4):571–582, 2009.

P. T. McHale and A. D. Lander. The protective role of symmetric stem cell division on the accumulation of heritable damage. *PLoS Comput Biol*, 10(8):e1003802, 2014.

C. E. Meacham and S. J. Morrison. Tumour heterogeneity and cancer cell plasticity. *Nature*, 501(7467):328–37, 9 2013. ISSN 1476-4687. doi: 10.1038/nature12624.

J. Meletiadis, S. Pournaras, E. Roilides, and T. J. Walsh. Defining fractional inhibitory concentration index cutoffs for additive interactions based on self-drug additive combinations, Monte Carlo simulation analysis, and in vitro-in vivo correlation data for antifungal drug combinations against Aspergillus fumi. *Antimicrobial agents and chemotherapy*, 54(2):602–9, 2 2010. ISSN 1098-6596. doi: 10.1128/AAC.00999-09.

D. Merico, R. Isserlin, O. Stueker, A. Emili, and G. D. Bader. Enrichment map: a network-

based method for gene-set enrichment visualization and interpretation. *PloS One*, 5(11): e13984, 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0013984.

J. A. Miller, C. Cai, P. Langfelder, D. H. Geschwind, S. M. Kurian, D. R. Salomon, and S. Horvath. Strategies for aggregating gene expression data: the collapseRows R function. *BMC bioinformatics*, 12:322, 2011. ISSN 1471-2105. doi: 10.1186/1471-2105-12-322.

K. J. Mitchell. What is complex about complex disorders? *Genome biology*, 13(1):237, 1 2012. ISSN 1474-760X. doi: 10.1186/gb-2012-13-1-237.

N. Miyoshi, H. Ishii, K.-i. Nagai, H. Hoshino, K. Mimori, F. Tanaka, H. Nagano, M. Sekimoto, Y. Doki, and M. Mori. Defined factors induce reprogramming of gastrointestinal cancer cells. *Proceedings of the National Academy of Sciences*, 107(1):40–45, 2010.

R. B. Mokhtari, T. S. Homayouni, N. Baluch, E. Morgatskaya, S. Kumar, B. Das, and H. Yeger. Combination therapy in combating cancer. *Oncotarget*, 8(23):38022–38043, 7 2015. ISSN 1949-2553. doi: 10.18632/oncotarget.16723.

G. Mollaoglu, M. R. Guthrie, S. Böhm, J. Brägelmann, I. Can, P. M. Ballieu, A. Marx, J. George, C. Heinen, M. D. Chalishazar, H. Cheng, A. S. Ireland, K. E. Denning, A. Mukhopadhyay, J. M. Vahrenkamp, K. C. Berrett, T. L. Mosbruger, J. Wang, J. L. Kohan, M. E. Salama, B. L. Witt, M. Peifer, R. K. Thomas, J. Gertz, J. E. Johnson, A. F. Gazdar, R. J. Wechsler-Reya, M. L. Sos, and T. G. Oliver. MYC Drives Progression of Small Cell Lung Cancer to a Variant Neuroendocrine Subtype with Vulnerability to Aurora Kinase Inhibition. *Cancer Cell*, 2017. ISSN 18783686. doi: 10.1016/j.ccell.2016.12.005.

S. A. Morris, P. Cahan, H. Li, A. M. Zhao, A. K. San Roman, R. A. Shivdasani, J. J. Collins, and G. Q. Daley. Dissecting engineered cell types and enhancing cell fate conversion via CellNet. *Cell*, 158(4):889–902, 2014.

B. T. Mott, R. T. Eastman, R. Guha, K. S. Sherlach, A. Siriwardana, P. Shinn, C. McKnight, S. Michael, N. Lacerda-Queiroz, P. R. Patel, P. Khine, H. Sun, M. Kasbekar, N. Aghdam, S. D. Fontaine, D. Liu, T. Mierzwa, L. A. Mathews-Griner, M. Ferrer, A. R. Renslo, J. Inglese, J. Yuan, P. D. Roepe, X.-z. Su, and C. J. Thomas. High-throughput matrix screening identifies synergistic and antagonistic antimalarial drug combinations. *Scientific Reports*, 5(1):13891, 11 2015. ISSN 2045-2322. doi: 10.1038/srep13891.

P. C. Nowell. The clonal evolution of tumor cell populations. *Science*, 194(4260):23–28, 1976.

H. Ogawa, X. Wu, K. Kawamoto, N. Nishida, M. Konno, J. Koseki, H. Matsui, K. Noguchi, N. Gotoh, T. Yamamoto, K. Miyata, N. Nishiyama, H. Nagano, H. Yamamoto, S. Obika, K. Kataoka, Y. Doki, M. Mori, and H. Ishii. MicroRNAs Induce Epigenetic Reprogramming and Suppress Malignant Phenotypes of Human Colon Cancer Cells. *PLOS ONE*, 10(5):e0127119, 5 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0127119.

K. Okita, T. Ichisaka, and S. Yamanaka. Generation of germline-competent induced pluripotent stem cells. *nature*, 448(7151):313–317, 2007.

J. O'Neil, Y. Benita, I. Feldman, M. Chenard, B. Roberts, Y. Liu, J. Li, A. Kral, S. Lejnine, A. Loboda, W. Arthur, R. Cristescu, B. B. Haines, C. Winter, T. Zhang, A. Bloecher, and S. D. Shumway. An Unbiased Oncology Compound Screen to Identify Novel Combination Strategies. *Molecular Cancer Therapeutics*, 2016. ISSN 1535-7163. doi: 10.1158/1535-7163.MCT-15-0843.

A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill Series in Electrical Engineering. McGraw-Hill, 1984. ISBN 9780070484689.

K. H. T. Paraiso, I. V. Fedorenko, L. P. Cantini, A. C. Munko, M. Hall, V. K. Sondak, J. L. Messina, K. T. Flaherty, and K. S. M. Smalley. Recovery of phospho-ERK activity allows

melanoma cells to escape from BRAF inhibitor therapy. *British Journal of Cancer*, 102 (12):1724–1730, 6 2010. ISSN 0007-0920. doi: 10.1038/sj.bjc.6605714.

N. R. Parker, A. L. Hudson, P. Khong, J. F. Parkinson, T. Dwight, R. J. Ikin, Y. Zhu, Z. J. Cheng, F. Vafaee, J. Chen, H. R. Wheeler, and V. M. Howell. Intratumoral heterogeneity identified at the epigenetic, genetic and transcriptional level in glioblastoma. *Scientific Reports*, 6:22477, 3 2016. ISSN 2045-2322. doi: 10.1038/srep22477.

T. J. Parmenter, M. Kleinschmidt, K. M. Kinross, S. T. Bond, J. Li, M. R. Kaadige, A. Rao, K. E. Sheppard, W. Hugo, G. M. Pupo, R. B. Pearson, S. L. McGee, G. V. Long, R. A. Scolyer, H. Rizos, R. S. Lo, C. Cullinane, D. E. Ayer, A. Ribas, R. W. Johnstone, R. J. Hicks, and G. A. McArthur. Response of BRAF-mutant melanoma to BRAF inhibition is mediated by a network of transcriptional regulators of glycolysis. *Cancer Discovery*, 2014. ISSN 21598290. doi: 10.1158/2159-8290.CD-13-0440.

B. B. Paudel, L. A. Harris, K. N. Hardeman, A. A. Abugable, C. E. Hayford, D. R. Tyson, and V. Quaranta. A Nonquiescent "Idling" Population State in Drug-Treated, BRAF-Mutated Melanoma. *Biophysical journal*, 114(6):1499–1511, 3 2018. ISSN 1542-0086. doi: 10.1016/j.bpj.2018.01.016.

E. Polley, M. Kunkel, D. Evans, T. Silvers, R. Delosh, J. Laudeman, C. Ogle, R. Reinhart, M. Selby, J. Connelly, E. Harris, N. Fer, D. Sonkin, G. Kaur, A. Monks, S. Malik, J. Morris, and B. A. Teicher. Small Cell Lung Cancer Screen of Oncology Drugs, Investigational Agents, and Gene and microRNA Expression. *Journal of the National Cancer Institute*, 2016. ISSN 14602105. doi: 10.1093/jnci/djw122.

G. R. Price. Fisher's 'fundamental theorem' made clear. *Annals of Human Genetics*, 36: 129–140, 1972.

H. Qian. Evolution Meets Thermodynamics: Fitness and Entropy Production in a Cell

Population Dynamics with Epigenetic Phenotype Switching. *arXiv:1309.1898 q-bio*, (arXiv:1309.1898), 2013.

O. J. L. Rackham, J. Firas, H. Fang, M. E. Oates, M. L. Holmes, A. S. Knaupp, H. Suzuki, C. M. Nefzger, C. O. Daub, J. W. Shin, and others. A predictive computational framework for direct reprogramming between human cell types. *Nature genetics*, 2016.

Y. Rais, A. Zviran, S. Geula, O. Gafni, E. Chomsky, S. Viukov, A. A. Mansour, I. Caspi, V. Krupalnik, M. Zerbib, and others. Deterministic direct reprogramming of somatic cells to pluripotency. *Nature*, 502(7469):65–70, 2013.

A. Rohrbeck, J. Neukirchen, M. Rosskopf, G. G. Pardillos, H. Geddert, A. Schwalen, H. E. Gabbert, A. von Haeseler, G. Pitschke, M. Schott, R. Kronenwett, R. Haas, and U.-P. Rohr. Gene expression profiling for molecular distinction and characterization of laser captured primary lung cancers. *Journal of Translational Medicine*, 6:69, 2008. ISSN 1479-5876. doi: 10.1186/1479-5876-6-69.

C. M. Rudin, N. Ismaila, C. L. Hann, N. Malhotra, B. Movsas, K. Norris, M. C. Pietanza, S. S. Ramalingam, A. T. Turrisi, and G. Giaccone. Treatment of Small-Cell Lung Cancer: American Society of Clinical Oncology Endorsement of the American College of Chest Physicians Guideline. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 33(34):4106–4111, 2015. ISSN 1527-7755. doi: 10.1200/JCO.2015.63.7918.

M. A. Savageau. Michaelis-Menten mechanism reconsidered: Implications of fractal kinetics. *Journal of Theoretical Biology*, 1995. ISSN 10958541. doi: 10.1006/jtbi.1995.0181.

I. Schiffmann, G. Greve, M. Jung, and M. Lübbert. Epigenetic therapy approaches in non-small cell lung cancer: Update and perspectives. *Epigenetics*, 11(12):858–870, 12 2016. ISSN 1559-2308. doi: 10.1080/15592294.2016.1237345.

M. Schindler. Theory of synergistic effects: Hill-type response surfaces as null-interaction models for mixtures. *Theoretical Biology and Medical Modelling*, 14(1):15, 12 2017. ISSN 1742-4682. doi: 10.1186/s12976-017-0060-y.

M. J. Schliekelman, A. Taguchi, J. Zhu, X. Dai, J. Rodriguez, M. Celiktas, Q. Zhang, A. Chin, C.-H. Wong, H. Wang, L. McFerrin, S. A. Selamat, C. Yang, E. M. Kroh, K. S. Garg, C. Behrens, A. F. Gazdar, I. A. Laird-Offringa, M. Tewari, I. I. Wistuba, J. P. Thiery, and S. M. Hanash. Molecular portraits of epithelial, mesenchymal, and hybrid States in lung adenocarcinoma and their relevance to survival. *Cancer Research*, 75(9): 1789–1800, 2015. ISSN 1538-7445. doi: 10.1158/0008-5472.CAN-14-2535.

S. V. Sharma, D. Y. Lee, B. Li, M. P. Quinlan, F. Takahashi, S. Maheswaran, U. McDermott, N. Azizian, L. Zou, M. A. Fischbach, K.-K. K. Wong, K. Brandstetter, B. Wittner, S. Ramaswamy, M. Classon, and J. Settleman. A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations. *Cell*, 141(1):69–80, 4 2010. ISSN 00928674. doi: 10.1016/j.cell.2010.02.027.

Y. Shimoyama, A. Nagafuchi, S. Fujita, M. Gotoh, M. Takeichi, S. Tsukita, and S. Hirohashi. Cadherin dysfunction in a human cancer cell line: possible involvement of loss of alpha-catenin expression in reduced cell-cell adhesiveness. *Cancer research*, 52(20): 5770–4, 10 1992. ISSN 0008-5472.

K. Siegrist. Virtual Laboratories in Probability and Statistics. University of Alabama Huntsville, 2014. URL http://www.math.uah.edu/stat/.

J. Siren, N. Valimaki, and V. Makinen. Indexing Graphs for Path Queries with Applications in Genome Research. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 11(2):375–388, 3 2014. ISSN 1545-5963. doi: 10.1109/TCBB.2013.2297101.

M. P. Smith, H. Brunton, E. J. Rowling, Z. A. Cooper, J. A. Wargo, C. W. Correspondence, J. Ferguson, I. Arozarena, Z. Miskolczi, J. L. Lee, M. R. Girotti, R. Marais, M. P.

Levesque, R. Dummer, D. T. Frederick, K. T. Flaherty, and C. Wellbrock. Inhibiting Drivers of Non-mutational Drug Tolerance Is a Salvage Strategy for Targeted Melanoma Therapy. *Cancer Cell*, 29:270–284, 2016. doi: 10.1016/j.ccell.2016.02.003.

J.-C. Soria, Y. Ohe, J. Vansteenkiste, T. Reungwetwattana, B. Chewaskulyong, K. H. Lee, A. Dechaphunkul, F. Imamura, N. Nogami, T. Kurata, I. Okamoto, C. Zhou, B. C. Cho, Y. Cheng, E. K. Cho, P. J. Voon, D. Planchard, W.-C. Su, J. E. Gray, S.-M. Lee, R. Hodge, M. Marotti, Y. Rukazenkov, S. S. Ramalingam, and FLAURA Investigators. Osimertinib in Untreated ¡i¿EGFR¡/i¿ -Mutated Advanced NonSmall-Cell Lung Cancer. *New England Journal of Medicine*, 378(2):113–125, 1 2018. ISSN 0028-4793. doi: 10.1056/NEJMoa1713137.

A. Soufi, G. Donahue, and K. S. Zaret. Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome. *Cell*, 151(5):994–1004, 2012. ISSN 0092-8674. doi: http://dx.doi.org/10.1016/j.cell.2012.09.045.

S. N. Steinway, J. G. T. Zañudo, P. J. Michel, D. J. Feith, T. P. Loughran, and R. Albert. Combinatorial interventions inhibit TGFβ-driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. *NPJ systems biology and applications*, 1:15014, 2015. ISSN 2056-7189. doi: 10.1038/npjsba.2015.14.

S. H. Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview press, 2014.

O. E. Sturm, R. Orton, J. Grindlay, M. Birtwistle, V. Vyshemirsky, D. Gilbert, M. Calder, A. Pitt, B. Kholodenko, and W. Kolch. The Mammalian MAPK/ERK Pathway Exhibits Properties of a Negative Feedback Amplifier. *Science Signaling*, 3(153):ra90–ra90, 12 2010. ISSN 1937-9145. doi: 10.1126/scisignal.2001212.

A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene

set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43):15545–15550, 2005. ISSN 0027-8424. doi: 10.1073/pnas.0506580102.

A. Subramanian, R. Narayan, S. M. Corsello, D. D. Peck, T. E. Natoli, X. Lu, J. Gould, J. F. Davis, A. A. Tubelli, J. K. Asiedu, D. L. Lahr, J. E. Hirschman, Z. Liu, M. Donahue, B. Julian, M. Khan, D. Wadden, I. C. Smith, D. Lam, A. Liberzon, C. Toder, M. Bagul, M. Orzechowski, O. M. Enache, F. Piccioni, S. A. Johnson, N. J. Lyons, A. H. Berger, A. F. Shamji, A. N. Brooks, A. Vrcic, C. Flynn, J. Rosains, D. Y. Takeda, R. Hu, D. Davison, J. Lamb, K. Ardlie, L. Hogstrom, P. Greenside, N. S. Gray, P. A. Clemons, S. Silver, X. Wu, W.-N. Zhao, W. Read-Button, X. Wu, S. J. Haggarty, L. V. Ronco, J. S. Boehm, S. L. Schreiber, J. G. Doench, J. A. Bittker, D. E. Root, B. Wong, and T. R. Golub. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*, 171(6):1437–1452, 11 2017. ISSN 1097-4172. doi: 10.1016/j.cell.2017.10.049.

J. Subramanian and R. Simon. Gene expression-based prognostic signatures in lung cancer: Ready for clinical use?, 2010. ISSN 00278874.

K. D. Sutherland, N. Proost, I. Brouns, D. Adriaensen, J.-Y. Song, and A. Berns. Cell of origin of small cell lung cancer: inactivation of Trp53 and Rb1 in distinct cell types of adult mouse lung. *Cancer Cell*, 19(6):754–764, 2011. ISSN 1878-3686. doi: 10.1016/j.ccr.2011.04.019.

M. L. Suvà, E. Rheinbay, S. M. Gillespie, A. P. Patel, H. Wakimoto, S. D. Rabkin, N. Riggi, A. S. Chi, D. P. Cahill, B. V. Nahed, and others. Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell*, 157(3):580–594, 2014.

A. Szedlak, G. Paternostro, and C. Piermarocchi. Control of asymmetric Hopfield networks

and application to cancer attractors. *PloS One*, 9(8):e105842, 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0105842.

D. Szklarczyk, A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguez, T. Doerks, M. Stark, J. Muller, P. Bork, and others. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic acids research*, 39(suppl 1):D561–D568, 2011.

K. Takahashi and S. Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4):663–676, 2006.

R. J. Tallarida. Quantitative methods for assessing drug synergism. *Genes & cancer*, 2(11): 1003–8, 11 2011. ISSN 1947-6027. doi: 10.1177/1947601912440575.

J. Tang, R. Salama, S. M. Gadgeel, F. H. Sarkar, and A. Ahmad. Erlotinib Resistance in Lung Cancer: Current Progress and Future Perspectives. *Frontiers in Pharmacology*, 4: 15, 2 2013. ISSN 1663-9812. doi: 10.3389/fphar.2013.00015.

J. Tang, K. Wennerberg, and T. Aittokallio. What is synergy? The Saariselkä agreement revisited. *Frontiers in pharmacology*, 6:181, 2015. ISSN 1663-9812. doi: 10.3389/fphar.2015.00181.

R. Thomas. Logical analysis of systems comprising feedback loops. *Journal of Theoretical Biology*, 1978. ISSN 10958541. doi: 10.1016/0022-5193(78)90127-3.

W. D. Travis. Update on small cell carcinoma and its differentiation from squamous cell carcinoma and other non-small cell carcinomas. *Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc*, 25 Suppl 1:18–30, 2012. ISSN 1530-0285. doi: 10.1038/modpathol.2011.150.

S. Tsuno, X. Wang, K. Shomori, J. Hasegawa, and N. Miura. Hsa-miR-520d induces

hepatoma cells to form normal liver tissues via a stemness-mediated process. *Scientific Reports*, 4:3852, 1 2014. ISSN 2045-2322. doi: 10.1038/srep03852.

N. R. Twarog, E. Stewart, C. V. Hammill, and A. A. Shelat. BRAID: A Unifying Paradigm for the Analysis of Combined Drug Action. *Scientific Reports*, 6(1):25523, 7 2016. ISSN 2045-2322. doi: 10.1038/srep25523.

D. R. Tyson, S. P. Garbett, P. L. Frick, and V. Quaranta. Fractional proliferation: A method to deconvolve cell population dynamics from single-cell data. *Nature Methods*, 2012. ISSN 15487091. doi: 10.1038/nmeth.2138.

J. J. Tyson, K. C. Chen, and B. Novak. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Current Opinion in Cell Biology*, 15(2): 221–231, 4 2003. ISSN 09550674. doi: 10.1016/S0955-0674(03)00017-6.

A. R. Udyavar, M. D. Hoeksema, J. E. Clark, Y. Zou, Z. Tang, Z. Li, M. Li, H. Chen, A. Statnikov, Y. Shyr, D. C. Liebler, J. Field, R. Eisenberg, L. Estrada, P. P. Massion, and V. Quaranta. Co-expression network analysis identifies Spleen Tyrosine Kinase (SYK) as a candidate oncogenic driver in a subset of small-cell lung cancer. *BMC systems biology*, 7 Suppl 5:S1, 2013. ISSN 1752-0509. doi: 10.1186/1752-0509-7-S5-S1.

A. R. Udyavar, D. J. Wooten, M. Hoeksema, M. Bansal, A. Califano, L. Estrada, S. Schnell, J. M. Irish, P. P. Massion, and V. Quaranta. Novel Hybrid Phenotype Revealed in Small Cell Lung Cancer by a Transcription Factor Network Model That Can Explain Tumor Heterogeneity. *Cancer research*, 77(5):1063–1074, 12 2017. ISSN 1538-7445. doi: 10.1158/0008-5472.CAN-16-1467.

L. J. P. Van Der Maaten and G. E. Hinton. Visualizing high-dimensional data using t-sne. *Journal of Machine Learning Research*, 2008. ISSN 1532-4435. doi: 10.1007/s10479-011-0841-3.

J. van der Veeken, S. Oliveira, R. M. Schiffelers, G. Storm, P. M. P. van Bergen En Henegouwen, and R. C. Roovers. Crosstalk between epidermal growth factor receptor- and insulin-like growth factor-1 receptor signaling: implications for cancer therapy. *Current cancer drug targets*, 9(6):748–60, 9 2009. ISSN 1873-5576.

F. Vazquez, J.-H. Lim, H. Chim, K. Bhalla, G. Girnun, K. Pierce, C. Clish, S. Granter, H. Widlund, B. Spiegelman, and P. Puigserver. PGC1α Expression Defines a Subset of Human Melanoma Tumors with Increased Mitochondrial Capacity and Resistance to Oxidative Stress. *Cancer Cell*, 23(3):287–301, 3 2013. ISSN 15356108. doi: 10.1016/j.ccr.2012.11.020.

E. F. Vêncio, A. M. Nelson, C. Cavanaugh, C. B. Ware, D. G. Milller, J. C. O. Garcia, R. Z. N. Vêncio, M. A. Loprieno, and A. Y. Liu. Reprogramming of prostate cancer-associated stromal cells to embryonic stem-like. *The Prostate*, 72(13):1453–1463, 2012.

T. C. Voss and G. L. Hager. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat Rev Genet*, 15(2):69–81, 2 2014. ISSN 1471-0056. doi: 10.1038/nrg3623.

C. H. Waddington. *The Strategy of the Genes*. Routledge, 1957. ISBN 978-1-317-65755-2.

J. Wang, L. Xu, and E. Wang. Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. *Proceedings of the National Academy of Sciences of the United States of America*, 105:12271–12276, 2008. ISSN 0027-8424. doi: 10.1073/pnas.0800579105.

J. Wang, L. Xu, E. Wang, and S. Huang. The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophysical Journal*, 99:29–39, 2010. ISSN 00063495. doi: 10.1016/j.bpj.2010.03.058.

J. Wang, K. Zhang, L. Xu, and E. Wang. Quantifying the Waddington landscape and biological paths for development and differentiation. *Proceedings of the National Academy*

of Sciences of the United States of America, 108:8257–8262, 2011. ISSN 0027-8424. doi: 10.1073/pnas.1017017108.

M. D. Wilkerson and D. N. Hayes. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics (Oxford, England)*, 26(12): 1572–1573, 2010. ISSN 1367-4811. doi: 10.1093/bioinformatics/btq170.

K. Willadsen and J. Wiles. Robustness and state-space structure of Boolean gene regulatory models. *Journal of Theoretical Biology*, 249(4):749–765, 2007. ISSN 0022-5193. doi: 10.1016/j.jtbi.2007.09.004.

S. C. Williamson, R. L. Metcalf, F. Trapani, S. Mohan, J. Antonello, B. Abbott, H. S. Leong, C. P. Chester, N. Simms, R. Polanski, D. Nonaka, L. Priest, A. Fusi, F. Carlsson, A. Carlsson, M. J. Hendrix, R. E. Seftor, E. A. Seftor, D. G. Rothwell, A. Hughes, J. Hicks, C. Miller, P. Kuhn, G. Brady, K. L. Simpson, F. H. Blackhall, and C. Dive. Vasculogenic mimicry in small cell lung cancer. *Nature Communications*, 2016. ISSN 20411723. doi: 10.1038/ncomms13322.

S. E. Witta, R. M. Gemmill, F. R. Hirsch, C. D. Coldren, K. Hedman, L. Ravdel, B. Helfrich, R. Dziadziuszko, D. C. Chan, M. Sugita, Z. Chan, A. Baron, W. Franklin, H. A. Drabkin, L. Girard, A. F. Gazdar, J. D. Minna, and P. A. Bunn. Restoring E-Cadherin Expression Increases Sensitivity to Epidermal Growth Factor Receptor Inhibitors in Lung Cancer Cell Lines. *Cancer Research*, 66(2):944–950, 1 2006. ISSN 0008-5472. doi: 10.1158/0008-5472.CAN-05-1988.

S. E. Witta, R. M. Jotte, K. Konduri, M. A. Neubauer, A. I. Spira, R. L. Ruxer, M. Varella-Garcia, P. A. Bunn, and F. R. Hirsch. Randomized phase II trial of erlotinib with and without entinostat in patients with advanced non-small-cell lung cancer who progressed on prior chemotherapy. *Journal of clinical oncology : official journal of the American*

Society of Clinical Oncology, 30(18):2248–55, 6 2012. ISSN 1527-7755. doi: 10.1200/JCO.2011.38.9411.

A. E. R. Woodcock. Landscapes of change: Catastrophe theory and biological processes. *Behavioral Science*, 23(4):390–401, 1978.

D. J. Wooten and V. Quaranta. Mathematical models of cell phenotype regulation and reprogramming: Make cancer cells sensitive again! *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1867(2):167–175, 4 2017. ISSN 0304419X. doi: 10.1016/j.bbcan.2017.04.001.

M. L. Wynn, N. Consul, S. D. Merajver, and S. Schnell. Logic-based models in systems biology: a predictive and parameter-free network analysis method. *Integrative Biology: Quantitative Biosciences from Nano to Macro*, 4(11):1323–1337, 2012. ISSN 1757-9708. doi: 10.1039/c2ib20193c.

J. Xue, Y. Zhu, Z. Sun, R. Ji, X. Zhang, W. Xu, X. Yuan, B. Zhang, Y. Yan, L. Yin, H. Xu, L. Zhang, W. Zhu, and H. Qian. Tumorigenic hybrids between mesenchymal stem cells and gastric cancer cells enhanced cancer proliferation, migration and stemness. *BMC cancer*, 15:793, 2015. ISSN 1471-2407. doi: 10.1186/s12885-015-1780-1.

B. Yadav, K. Wennerberg, T. Aittokallio, and J. Tang. Searching for Drug Synergy in Complex DoseResponse Landscapes Using an Interaction Potency Model. *Computational and Structural Biotechnology Journal*, 13:504–513, 1 2015. ISSN 2001-0370. doi: 10.1016/J.CSBJ.2015.09.001.

W. Yang, J. Soares, P. Greninger, E. J. Edelman, H. Lightfoot, S. Forbes, N. Bindal, D. Beare, J. A. Smith, I. R. Thompson, S. Ramaswamy, P. A. Futreal, D. A. Haber, M. R. Stratton, C. Benes, U. McDermott, and M. J. Garnett. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells.

Nucleic acids research, 41(Database issue):955–61, 1 2013. ISSN 1362-4962. doi: 10.1093/nar/gks1111.

T. A. Yap, A. Omlin, and J. S. de Bono. Development of Therapeutic Combinations Targeting Major Cancer Signaling Pathways. *Journal of Clinical Oncology*, 31(12):1592–1605, 4 2013. ISSN 0732-183X. doi: 10.1200/JCO.2011.37.6418.

G. Yu, F. Li, Y. Qin, X. Bo, Y. Wu, and S. Wang. GOSemSim: An R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*, 2010. ISSN 13674803. doi: 10.1093/bioinformatics/btq064.

X. Zhang, F. D. Cruz, M. Terry, F. Remotti, and I. Matushansky. Terminal differentiation and loss of tumorigenicity of human cancers via pluripotency-based reprogramming. *Oncogene*, 32(18):2249–2260, 2013.

J. X. Zhou and S. Huang. Understanding gene circuits at cell-fate branch points for rational cell reprogramming, 2011. ISSN 01689525.

J. X. Zhou and S. Huang. Theoretical Considerations for Reprogramming Multicellular Systems. In *Synthetic Biology*, pages 81–99. Academic Press, 2013. ISBN 9780123944306. doi: 10.1016/B978-0-12-394430-6.00005-4.

J. X. Zhou, L. Brusch, and S. Huang. Predicting pancreas cell fate decisions and reprogramming with a hierarchical multi-attractor model. *PLoS ONE*, 6, 2011. ISSN 19326203. doi: 10.1371/journal.pone.0014752.

J. X. Zhou, M. D. S. Aliyu, E. Aurell, and S. Huang. Quasi-potential landscape in complex multi-stable systems. *Journal of The Royal Society Interface*, 9(77):3539–3553, 2012.

A. Zimmer, I. Katzir, E. Dekel, A. E. Mayo, and U. Alon. Prediction of multidimensional drug dose responses based on measurements of drug pairs. *Proceedings of the National*

Academy of Sciences of the United States of America, 113(37):10442–7, 9 2016. ISSN 1091-6490. doi: 10.1073/pnas.1606301113.