12-2022

Combining Network Modeling and Experimental Approaches to Predict Drug Combination Responses

Deepraj Sarmah
dsarmah@g.clemson.edu

[COMBINING NETWORK MODELING AND EXPERIMENTAL APPROACHES TO PREDICT DRUG COMBINATION RESPONSES]

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Chemical and Biomolecular Enginnering

---

by
Deepraj Sarmah
December 2022

Accepted by:
Dr. Marc Birtwistle, Committee Chair
Dr. Jessica Larsen
Dr. F. Alex Feltus
Dr. Sarah Harcum

ABSTRACT

Cancer is a lethal disease and complex at multiple levels of cell biology. Despite many advances in treatments, many patients do not respond to therapy. This is owing to the complexity of cancer-genetic variability due to mutations, the multi-variate biochemical networks within which drug targets reside and existence and plasticity of multiple cell states. It is generally understood that a combination of drugs is a way to address the multi-faceted drivers of cancer and drug resistance. However, the sheer number of testable combinations and challenges in matching patients to appropriate combination treatments are major issues.

Here, we first present a general method of network inference which can be applied to infer biological networks. We apply this method to infer different kinds of networks in biological levels where cancer complexity resides-a biochemical network, gene expression and cell state transitions. Next, we focus our attention on glioblastoma and with pharmacological and biological considerations, obtain a ranked list of important drug targets in glioblastoms. We perform drug dose response experiments for 22 blood brain barrier penetrant drugs against 3 glioblastoma cell lines. These methods and experimental results inform a construction of a temporal cell state model to predict and experimentally validate combination treatments for certain drugs. We improve an experimental method to perform high throughput western blots and apply the method to discover biochemical interactions among some important proteins involved in temporal cell state transitions. Lastly, we illustrate a method to investigate potential resistance mechanisms in genome scale proteomic data.

We hope that methods and results presented here can be adapted and improved upon to help in the discovery of biochemical interactions, capturing cell state transitions and ultimately help predict effective combination therapies for cancer.

ACKNOWLEDGMENTS

I must thank providence for having come this far in my PhD. Hailing from a remote village in north-eastern India, it has been almost an unbelievable journey for me to get this PhD from halfway around the world and I must thank the many people involved in the past and present in helping me arrive here.

Firstly, I must thank the people who raised me-my parents, Taranath Sarmah and Tilarupa Devi, my aunt- Bhabhani Devi and my grandparents Jagannath Sarmah and Kaushalya Devi. I believe their encouragement to pursue a scientific career and pushing me along the way was key to the choices I made to pursue this program. Thanks to my uncles-Yuvaraj Sharma, Prakash Sharma and aunts Mandira Sharma, Moon Sharma for welcoming me into their home and assistance during some important times. Thanks to all my cousins for being supportive and being a source of motivation.

Thanks to my colleagues in my previous company, Pluss Advanced Technologies and particularly Dr. Devendra Jain for being a great mentor. Thanks also to my friends-Ashwin Nenmini and Ravi Teja AT who were instrumental in helping me through the process of applying for the PhD and great friends overall.

Thanks to all the other special people in my life who may not have been mentioned but whose presence added meaning to this journey.

Lastly, great thanks to my past and current colleagues in the Birtwistle Lab- Dr. Mehdi Bouhaddou, Caitlin Anglin, Anne Marie Barrette, Alan Stern, Cemal Erdem and Xiaoming Lu. Thanks to the undergraduates who have worked with me in the past- Ian Weber, Madison Price, Zain Kazmi, Wesley Meredith, Matthew Birkley, Micah Jordan,

Nicholas Harold. Thanks to Jonah Huggins for all the good work with Mesowestern in my final year and a half.

A great thanks to Dr. Marc Birtwistle for being his support throughout this PhD. I learnt a lot from him over the years and thank him for being the voice of reason and encouragement when necessary. His enthusiasm and logic were infectious, and I will be grateful for his kindness and friendship.

Finally, thanks to my dissertation committee- Dr. Jessica Larsen, Dr. F. Alex Feltus and Dr. Sarah Harcum for your patience and service.

TABLE OF CONTENTS

Page

CHAPTER ONE

INTRODUCTION

Cancer and Glioblastoma

Cancer is one of the most fatal and tragic maladies that affects human existence today. It is prevalent not only among humans but also most animal species. In recorded human history, symptoms of cancer have been mentioned as far back as 2600 BC in ancient Egypt[1]. Despite several advances in treatment, particularly in the previous century, it remains as one of the leading causes of death in the world[2].

Physiologically, cancer may be classified into several types based on tissue of origin but all of them generally manifest as cells which have obtained an autonomous will to divide. These uncontrolled cell division may invade and take over the body's vital organs and ultimately lead to death. However, the mechanisms employed by these aberrant cells are not foreign at all- they are innate to the human body and its growth and survival[3].

Many advances have been made in the treatments of cancer have been in the past century, even before the role of genes and mutations in cancer were well established. As early as 1896, radiation therapy was used to treat cancers[4]. After the second world war, alkylating agents and antimetabolites were found to suppress growth of certain cancer cells and were established as some of the first chemotherapy regimens[5]. Many of these treatments are still in use today. Hodgkin's lymphoma treated with a combination of such chemotherapy regimens has a five-year survival rate of around 90%[6]. Radiation therapy and temozolomide (an alkylating agent) are the standard for treatment of glioblastoma[7].

A breakthrough in cancer research was to recognize certain genes and mutations as the drivers of cancer. This allowed for innovation of therapies to inhibit these "driver mutations". A success story in such targeted therapy is the use of Imatinib in chronic myeloid leukemia, wherein it inhibits a fused protein BCR-ABL1 caused due to a chromosomal translocation[8]. With a ten-year survival rate of over 80%, average life expectancy of patients treated with Imatinib approaches that of the general population[9]. Following the success of Imatinib, other therapies targeting specific mutations for different cancers were also innovated. However, subsequent application of targeted therapies on many different cancers have not been as successful, even triggering some skepticism about their efficacy[10]. Evidence of "driver mutations" may not always translate to druggable targets and there is a lack of effective screening methods to match patients with treatment regimens.

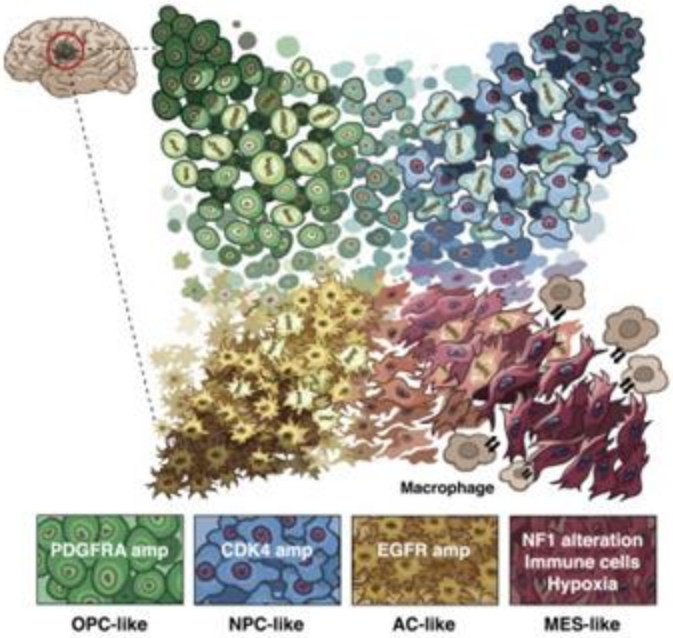
Glioblastoma, a cancer of the brain is one such cancer that has evaded effective therapies and a focus of study in this dissertation. It is marked by a low life expectancy after diagnosis. The standard treatment is surgical resection followed by radiotherapy and temozolomide, which provides a marginal improvement in life expectancy compared to radiotherapy alone[11]. Clinical trials with several other more precise chemotherapy drugs have been performed but have not resulted in significant survival benefits.

The hindrances in the treatment of many cancers may be a testament to the fact that cancer is inherently complex. A brief exploration of the complexity involved in cancer and their implications on drug resistance are mentioned below.

Cell State Heterogeneity in Cancer and Glioblastoma

Cancer cells display highly diverse phenotypic and molecular characteristics both at the inter-tumoral (among different patients) and intra-tumoral levels (within the same tumor). Besides genetic diversity, cancer cells may be divided at multiple omics levels as distinct cell states. Cell states are often defined by their transcriptomics (through for example single cell RNAseq experiments). Cells in different states may have different drug sensitivities[12], and it is becoming appreciated that cells can transition between such states in development-like networks, sometimes called cell state networks[13]. Such plasticity between cell states can contribute to drug resistance[12,14].

In glioblastoma, using the available RNA seq data from The Cancer Genome Atlas (TCGA) 4 cell states were classified in 2010[15] which was later updated to 3 cell states to only include malignant cells[16]. More recently, single cell RNAseq on cells taken from glioblastoma patients has broadly classified glioblastoma into 4 cell states governed by their gene expression and microenvironment and demonstrated plasticity between the cell states (Figure 1.1)[13].



Adapted from Neftel et al. (2019)

Figure 1.1 Licensed from Neftel et. al 2019 (5411941324602). Model for the cellular states of glioblastoma and their genetic and micro-environmental determinants. Lighter or darker tones indicate strength of each program. Intermediate states are shown in between the four states and indicate transitions.

Cell state transition networks have been identified for multiple cancer types generally by combining single cell measurements (e.g. single cell RNAseq), with perturbation time courses, such as enriching for one cell state and then observing the fractional composition dynamics[13]. Because of their inherent plasticity, cell states in cancer have been difficult to pin down but better mapping of cell states should lead to more accurate therapies.

Complexity of Biochemical Networks

Drug targets reside within a multi-variate complex of biochemical networks by which chemotherapy drugs exert their action. These networks can differ between cell

states, adapt to therapy, and also give rise to non-intuitive therapy results, such as feedback loops and compensatory pathways underlying the efficacy of combining Raf and MEK inhibitor combinations, which lie in the same genetic pathway[17].

It stands to reason that a complex disease as cancer would not always be reliant on a single protein or pathway and inhibition of a supposed "driver mutation" may be compensated elsewhere. Glioblastoma cells, where EGFR expression is often amplified, are often not responsive to EGFR inhibition, particularly in cells with loss of PTEN[18]. This is also evident in multiple different cancer cells in cancer cell line encyclopedia[19].

Genetic Variability

Another standard for inter-tumoral and intra-tumoral heterogeneity are the myriad number of genetic mutations. The Cancer Genome Atlas (TCGA) has identified tens of mutations in primary tumors, just among some of the major signaling pathways including the MAPK and PI3K signaling pathways, DNA damage response pathways and cell cycle pathways [20](Figure 1.2). An illustrative summary of such mutations in glioblastoma samples is shown in the Figure 1.2.

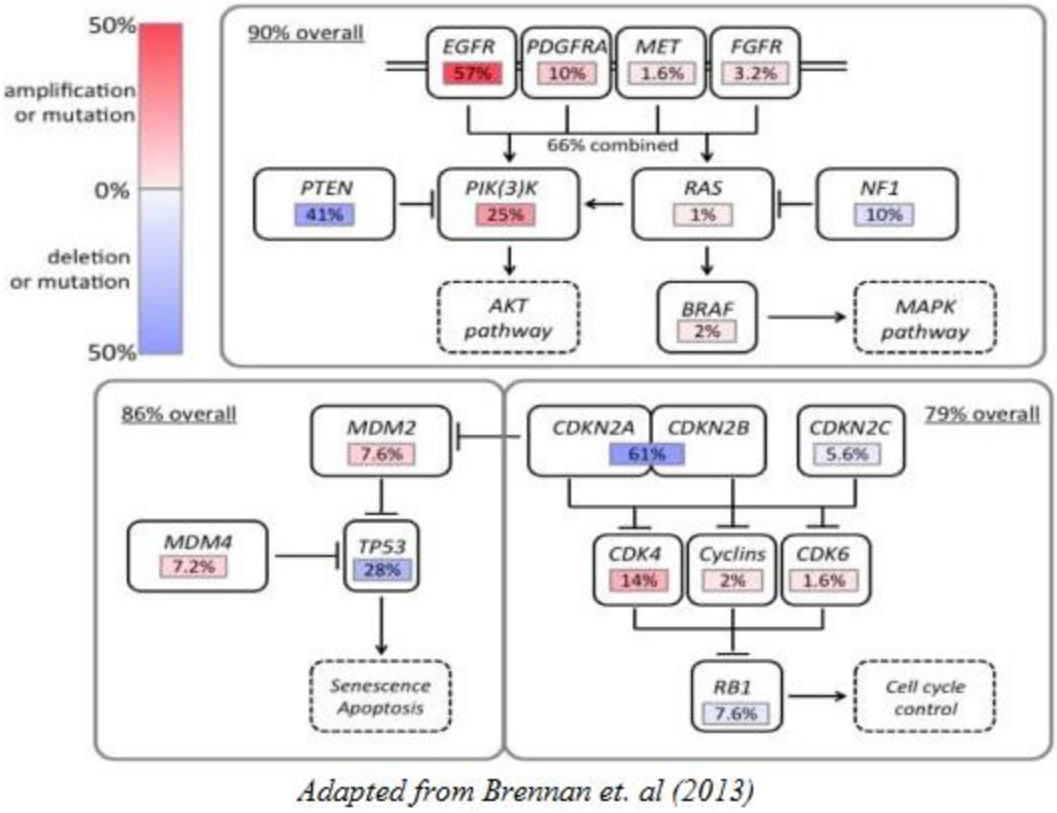


Figure 1.2: Licensed from Brennan et. al 2013 (5412130306368). Alterations affecting canonical signal transduction and tumor suppressor pathways summarized for 251 glioblastoma samples

Besides such inter-patient variability there is also significant evidence for intra-tumoral heterogeneity[12,14]. Such mutational heterogeneity within a tumor can bolster drug resistance as cells with more resistive mutations get selected through Darwinian selection, where the drug may act as a selective pressure.

How do we map the effects of a new mutation? Methods which can map the causal effects of a network in which the mutation resides should help link the effects of the mutation with the rest of the network[21]. Ultimately, better mapping of these mutational profiles should enable appropriate association of patients with treatment regimens[22].

Prediction of Effective Combinations Therapies

To account for the complexities of cancer, combinations of multiple therapies is essential. The idea of combination therapies to counter single drug resistance has been present since 1965[23] and this idea has continued to modern chemotherapy regimens which often include a combination of about 3-4 drugs. The reasons for efficacy of combinations can be diverse – offset pathway feedback effects[17], effective localization of relevant drug to name a few[24].

However, the increased number of FDA approved therapies raises another problem- the sheer space for combinations. For instance, for the 71 kinase inhibitors approved by FDA by May 2022- there can be 2485 two-way and 57115 three-way combinations. Given the manifold heterogeneity of cancer, testing of combinations therapies is infeasible to be performed experimentally and require intervention from faster and relatively inexpensive computational methods to select promising combinations.

Dissertation Overview

In the broadest sense, this dissertation explores methods of mapping biological networks and their experimental application which can help getting us closer to better treatments or combination of treatments for cancer. Although these methods are general, we tested them in glioblastoma cells.

In Chapter 2, we begin by building upon existing methods to present an improved method to infer biological networks, which is more robust to noise and is capable of mapping external parameters to a network. We apply this method to infer a simulated biochemical network, 16 gene regulatory networks and cell state transition networks.

Insights gained from this work inform a subsequent temporal cell state network (Chapter 5). The results from Chapter 2 are in press at npj Systems Biology and Applications.

In Chapter 3, we begin by creating a list of drugs with high blood brain barrier (BBB) penetrance and create a ranked list of kinase targets-scored by two pharmacological and two biological criteria. This chapter is an illustrative exercise for the selection of relevant proteins in a cancer, in absence of drug dose response data.

In Chapter 4, we obtain drug dose responses for a panel of 22 BBB penetrant drugs across three glioblastoma derived cell lines. We contrast these results with the list of ranked kinases in Chapter 3. This work also provides an experimental basis for the work in Chapter 5.

In Chapter 5, we conceive of a temporal cell state model and use it to predict certain drug combinations, which are then validated by experiments. The results from Chapter 5 are available in biorxiv and we have submitted it to PLOS Computational Biology.

In Chapter 6 we describe the various steps we took to validate a version of Mesowestern-a method to enable high throughput western blots which may be used in cancer research or in other biological work. This work has been published at ACS Omega. We further go on to describe the steps we took to come up with an improved design and buffer chemistry and finally used the method to test the time courses of two protein markers in two different glioblastoma cell lines.

In Chapter 7, we describe a collaboration project where we analyzed proteomic data obtained from glioblastoma cell lines with certain overexpressed proto-oncogenes. We explore the use of enrichment analysis and the use of a dimension reduction method with

an intention to uncover signaling pathways involved in resistance mechanisms in these cells. Our collaborator could obtain an experimental validation for one of the results. The work is still ongoing in the Birtwistle Lab.

In Chapter 8, we describe the broad conclusions and future directions of all the work included in this dissertation.

References-

1. Hajdu, S.I. (2016). Pathfinders in oncology from ancient times to the end of the Middle Ages. Cancer *122*, 1638–1646. 10.1002/cncr.29955.

2. Siegel, R.L., Miller, K.D., Fuchs, H.E., and Jemal, A. (2022). Cancer statistics, 2022. CA: A Cancer Journal for Clinicians *72*, 7–33. 10.3322/caac.21708.

3. Cooper, G.M. (2000). The Development and Causes of Cancer. The Cell: A Molecular Approach. 2nd edition.

4. Holsti, L.R. (1995). Development Of Clinical Radiotherapy Since 1896. Acta Oncologica *34*, 995–1003. 10.3109/02841869509127225.

5. DeVita, V.T., Jr., and Chu, E. (2008). A History of Cancer Chemotherapy. Cancer Research *68*, 8643–8653. 10.1158/0008-5472.CAN-07-6611.

6. Shanbhag, S., and Ambinder, R. (2018). Hodgkin Lymphoma: a review and update on recent progress. CA Cancer J Clin *68*, 116–132. 10.3322/caac.21438.

7. Fernandes, C., Costa, A., Osório, L., Lago, R.C., Linhares, P., Carvalho, B., and Caeiro, C. (2017). Current Standards of Care in Glioblastoma Therapy (Codon Publications) 10.15586/codon.glioblastoma.2017.ch11.

8. Druker, B.J., Guilhot, F., O'Brien, S.G., Gathmann, I., Kantarjian, H., Gattermann, N., Deininger, M.W.N., Silver, R.T., Goldman, J.M., Stone, R.M., et al. (2006). Five-Year Follow-up of Patients Receiving Imatinib for Chronic Myeloid Leukemia. New England Journal of Medicine *355*, 2408–2417. 10.1056/NEJMoa062867.

9. Bower, H., Björkholm, M., Dickman, P.W., Höglund, M., Lambert, P.C., and Andersson, T.M.-L. (2016). Life Expectancy of Patients With Chronic Myeloid Leukemia Approaches the Life Expectancy of the General Population. J Clin Oncol *34*, 2851–2857. 10.1200/JCO.2015.66.2866.

10. Prasad, V. (2016). Perspective: The precision-oncology illusion. Nature *537*, S63–S63. 10.1038/537S63a.

11. Perry, J.R., Laperriere, N., O'Callaghan, C.J., Brandes, A.A., Menten, J., Phillips, C., Fay, M., Nishikawa, R., Cairncross, J.G., Roa, W., et al. (2017). Short-Course Radiation plus Temozolomide in Elderly Patients with Glioblastoma. New England Journal of Medicine *376*, 1027–1037. 10.1056/NEJMoa1611977.

12. Marusyk, A., Janiszewska, M., and Polyak, K. (2020). Intratumor heterogeneity: the Rosetta stone of therapy resistance. Cancer Cell *37*, 471–484. 10.1016/j.ccell.2020.03.007.

13. Neftel, C., Laffy, J., Filbin, M.G., Hara, T., Shore, M.E., Rahme, G.J., Richman, A.R., Silverbush, D., Shaw, M.L., Hebert, C.M., et al. (2019). An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. Cell *178*, 835-849.e21. 10.1016/j.cell.2019.06.024.

14. Ramón y Cajal, S., Sesé, M., Capdevila, C., Aasen, T., De Mattos-Arruda, L., Diaz-Cano, S.J., Hernández-Losa, J., and Castellví, J. (2020). Clinical implications of intratumor heterogeneity: challenges and opportunities. J Mol Med (Berl) *98*, 161–177. 10.1007/s00109-020-01874-2.

15. Verhaak, R.G.W., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., et al. (2010). An integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR and NF1. Cancer Cell *17*, 98. 10.1016/j.ccr.2009.12.020.

16. Sidaway, P. (2017). Glioblastoma subtypes revisited. Nat Rev Clin Oncol *14*, 587–587. 10.1038/nrclinonc.2017.122.

17. Flaherty, K.T., Robert, C., Hersey, P., Nathan, P., Garbe, C., Milhem, M., Demidov, L.V., Hassel, J.C., Rutkowski, P., Mohr, P., et al. (2012). Improved Survival with MEK Inhibition in BRAF-Mutated Melanoma. New England Journal of Medicine *367*, 107–114. 10.1056/NEJMoa1203421.

18. Mellinghoff, I.K., Wang, M.Y., Vivanco, I., Haas-Kogan, D.A., Zhu, S., Dia, E.Q., Lu, K.V., Yoshimoto, K., Huang, J.H.Y., Chute, D.J., et al. (2005). Molecular determinants of the response of glioblastomas to EGFR kinase inhibitors. N Engl J Med *353*, 2012–2024. 10.1056/NEJMoa051918.

19. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607. 10.1038/nature11003.

20. Brennan, C.W., Verhaak, R.G.W., McKenna, A., Campos, B., Noushmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al. (2013). The Somatic Genomic Landscape of Glioblastoma. Cell *155*, 462–477. 10.1016/j.cell.2013.09.034.

21. Kholodenko, B.N., Kiyatkin, A., Bruggeman, F.J., Sontag, E., Westerhoff, H.V., and Hoek, J.B. (2002). Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. Proceedings of the National Academy of Sciences of the United States of America *99*, 12841–12846. 10.1073/pnas.192442699.

22. Andre, F., Filleron, T., Kamal, M., Mosele, F., Arnedos, M., Dalenc, F., Sablin, M.-P., Campone, M., Bonnefoi, H., Lefeuvre-Plesse, C., et al. (2022). Genomics to select treatment for patients with metastatic breast cancer. Nature *610*, 343–348. 10.1038/s41586-022-05068-3.

23. Frei, E., Karon, M., Levin, R.H., J. Freireicii, E., Taylor, R.J., Hananian, J., Selawry, O., Holland, J.F., Hoogstraten, B., Wolman, I.J., et al. (1965). The Effectiveness of Combinations of Antileukemic Agents in Inducing and Maintaining Remission in Children with Acute Leukemia. Blood *26*, 642–656.

24. Zhang, Z., Fan, Q., Luo, X., Lou, K., Weiss, W.A., and Shokat, K.M. (2022). Brain-restricted mTOR inhibition with binary pharmacology. Nature *609*, 822–828. 10.1038/s41586-022-05213-y.

CHAPTER TWO

NETWORKS INFERENCE FROM PERTURBATION TIME COURSE DATA

Abstract

Networks underlie much of biology from subcellular to ecological scales. Yet, understanding what experimental data are needed and how to use them for unambiguously identifying the structure of even small networks remains a broad challenge. Here, we integrate a dynamic least squares framework into established modular response analysis (DL-MRA), that specifies sufficient experimental perturbation time course data to robustly infer arbitrary two and three node networks. DL-MRA considers important network properties that current methods often struggle to capture: (i) edge sign and directionality; (ii) cycles with feedback or feedforward loops including self-regulation; (iii) dynamic network behavior; (iv) edges external to the network; and (v) robust performance with experimental noise. We evaluate the performance of and the extent to which the approach applies to cell state transition networks, intracellular signaling networks, and gene regulatory networks. Although signaling networks are often an application of network reconstruction methods, the results suggest that only under quite restricted conditions can they be robustly inferred. For gene regulatory networks, the results suggest that incomplete knockdown is often more informative than full knockout perturbation, which may change experimental strategies for gene regulatory network reconstruction. Overall, the results give a rational basis to experimental data requirements for network reconstruction and can be applied to any such problem where perturbation time course experiments are possible.

Introduction

Networks underlie much cellular and biological behavior, including transcriptional, protein-protein interaction, signaling, metabolic, cell-cell, endocrine, ecological, and social networks, among many others. As such, identifying and then representing their structure has been a focus of many for decades now. This is not just from experimental perspectives alone, but predominantly computational with a variety of statistical methodologies that integrate prior knowledge from interaction databases with new experimental data sets [1–24]. Alternatively, a variety of methods have investigated general ways to infer detailed reaction mechanisms—often a foundation of networks—from experimental data [25–29]. Such tasks may be considered a subset of network inference.

Network structure is usually represented as either an undirected or a directed graph, with edges between nodes specifying the system. There are five main areas where current approaches to reconstructing networks struggle to capture important features of biological networks. The first is directionality of edges [6,8,30,31]. Commonly employed correlational methods predominantly generate undirected edges, which impedes causal and other mechanistic analyses. Second is cycles. Cycles such as feedback or feedforward loops are nearly ubiquitous in biological systems and central to their function [32,33]. This also includes an important type of cycle: self-regulation of a node, that is, an edge onto itself, which is rarely considered [34]. Third is that biological networks are often dynamic. Two notable examples are circadian and p53 oscillators [35,36], where dynamics are key to biological function. Directionality and edge signs (i.e. positive or negative) dictate dynamics. Fourth is pinpointing how external variables impinge on network nodes. For example, is the effect

of a growth factor on a network node direct, or though other nodes in the network? Fifth, the design and method employed should be robust to typical experimental noise levels. The experimental design and data requirements to uniquely identify the ***dynamic, directed and signed*** edge structures in biological networks containing all types of cycles and external stimuli remains a largely open but significant problem. Any such design should ideally be feasible to implement with current experimental technologies.

Modular Response Analysis (MRA) approaches, first pioneered by Kholodenko and colleagues in 2002 [37,38] inherently deal with cycles and directionality by prescribing systematic perturbation experiments followed by steady-state measurements. The premise for data requirements is to measure the entire system response to at least one perturbation for each node. Thus, an *n* node system requires *n* experiments, if the system response can be measured in a global fashion (i.e. all nodes measured at once). The original instantiations struggled with the impact of experimental noise, but total least squares MRA and Monte Carlo sampling helped to improve performance [39–41]. Incomplete and prior knowledge can be handled as well using both maximum likelihood and Bayesian approaches [42–45]. However, these approaches are based on steady-state data, or fixed time point data, limiting abilities to deal with dynamic systems. There is a formal requirement for small perturbations, which are experimentally problematic and introduce issues for estimation with noisy data. Subsequent approaches have recommended the use of large perturbations as a trade off in dealing with noisy data, but the theory still formally requires small perturbations [41]. Lastly, there are two classes of biologically relevant edges that MRA does not comprehensively address. First is self-regulation of a node, which is often normalized

(to -1) causing it to not be uniquely identifiable. The other are the effects of stimuli external to the network (basally present or administered) on the modeled nodes.

In addition to perturbations, another experimental design feature that can inform directionality is a time-series, which have also been integrated into MRA. These work[37,46] use time-series perturbation data to uniquely infer a signed, directed network that can predict dynamic network behavior. In an *n* node open system (e.g. protein levels are not constant), multiple nodes would either be distinctly perturbed more than once, such as both production and degradation of a transcript, or phosphorylation and dephosphorylation of a protein, or the system monitored before and after the perturbation (with one perturbation per node). This can be experimentally challenging both in terms of scale and finding suitable distinct perturbations for a node. Moreover, as is often the case, noise in the experimental data severely limits inference accuracy (due to required estimation of 2nd derivatives). Subsequent work [47], recommends smaller perturbations and difference in timepoints but also does not address noisy data. Further work has demonstrated that larger perturbations produce better results due to inevitable experimental noise [41]. Thus, there remains a need for methods that can infer signed, directed networks from feasible perturbation time course experiments that capture dynamics, can uniquely estimate edge properties related to self-regulation and external stimuli, and finally that function in the presence of typical experimental noise levels.

Here we describe a novel, MRA-inspired approach called Dynamic Least-squares MRA (DL-MRA). For an *n*-node system, *n* perturbation time courses are required, and thus experimental requirements scale linearly as the network size increases. The approach uses

an underlying network model that captures dynamic, directional, and signed networks that include cycles, self-regulation, and external stimulus effects. We test DL-MRA using simulated time-series perturbation data with known network topology under increasing levels of simulated noise. The approach has good accuracy and precision for identifying network structure in randomly generated two and three node networks that contain a wide variety of cycles. For the investigated cases, we find between 7 to 11 evenly distributed time points yielded reasonable results, although we expect this will strongly depend on time point placement. We apply the approach to models describing a cell state switching network [48], a signal transduction network [49], and a gene regulatory network [32]. Although signaling networks are often a focus in network biology, our analysis suggests they have unique properties that render them generally recalcitrant to reconstruction. Results from the gene regulatory network application suggest that incomplete perturbation (e.g. partial knockdown vs. knockout) is more informative than complete inhibition. While challenges remain for expanding to other and larger systems, the proposed algorithm robustly infers a wide range of networks with good specificity and sensitivity using feasible time course experiments, all while making progress on limitations of current inference approaches.

Results

Formulation of Sufficient Experimental Data Requirements for Network Reconstruction

Consider a 2-node network with four directed, weighted edges (Fig. 1a). An external stimulus may affect each of the two nodes differently and its effect is quantified by $S_{1,ex}$ and $S_{2,ex}$, respectively (e.g. Methods, Eq. 15). We also allow for basal/constitutive production in each node ($S_{i,b}$). Let $x_i(k)$ be the activity of node i at time point t_k. The network dynamics can be cast as a system of ordinary differential equations (ODEs) as follows

$$\frac{dx_1}{dt} \equiv f_1(x_1(k), x_2(k), S_{1,ex}, S_{1,b}) \equiv f_1(k); \frac{dx_2}{dt} \equiv f_2(x_1(k), x_2(k), S_{2,ex}, S_{2,b}) \equiv f_2(k) . \quad (1)$$

The network edges can be connected to the system dynamics through the Jacobian matrix $\mathbf{J}$ [37,38,46],

$$\mathbf{J} \equiv \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix} \equiv \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix} \quad (2)$$

The network edge weights (F_{ij}'s) describe how the activity of one node affects the dynamics of another node in a causal and direct sense, given the explicitly considered nodes (though not necessarily in a physical sense). In practice, however, causality can only be approached if every component of the system is included in the model, which is not typical (and even more so, there must be no model mismatch, which is almost impossible to guarantee)[6,30,31,50,51]. In MRA, these nodes may be individual species or "modules". In order to simplify a complex network it may often be separated into "modules" comprising smaller networks of inter-connected species with the assumption that each module is

generally insulated from other modules except for information transfer through so-called communicating species [37]. Cases where such modules may not be completely isolated are explored elsewhere [52].

What experimental data are sufficient to uniquely estimate the signed directionality of the network edges and thus infer the causal relationships within the system? Fundamentally, we know that perturbations and/or dynamics are important for inferring causality [6,37,46,51,52]. Consider a simple setup of three time-course experiments that each measure x_1 and x_2 dynamics in response to a stimulus (Fig. 1b-g). One time course is in the presence of no perturbation (vehicle), one has a perturbation of Node 1, and one has a perturbation of Node 2. Consider further that the perturbations are reasonably specific, such that the perturbation of x_1 has negligible *direct* effects on x_2, and vice versa, and that these perturbations may be large. Experimentally, this could be an shRNA or gRNA that is specific to a particular node, or that a small molecule inhibitor is used at low enough dose to predominantly inhibit the targeted node. A well-posed estimation problem can be formulated (see Methods) that, in principle, allows for unique estimation of the Jacobian elements as a function of time with the following set of linear algebra relations:

$$\begin{bmatrix} y_1(t_{k+1}) \\ y_{1,2}(t_k) \end{bmatrix} = \begin{bmatrix} \Delta_t x_1(t_{k+1}) & \Delta_t x_2(t_{k+1}) \\ \Delta_{p,2} x_1(t_k) & \Delta_{p,2} x_2(t_k) \end{bmatrix} \begin{bmatrix} F_{11}(t_k) \\ F_{12}(t_k) \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} y_2(t_{k+1}) \\ y_{2,1}(t_k) \end{bmatrix} = \begin{bmatrix} \Delta_t x_1(t_{k+1}) & \Delta_t x_2(t_{k+1}) \\ \Delta_{p,1} x_1(t_k) & \Delta_{p,1} x_2(t_k) \end{bmatrix} \begin{bmatrix} F_{21}(t_k) \\ F_{22}(t_k) \end{bmatrix} \quad (4)$$

Here, $y_{i,j}$ refers to a measured first-time derivative of node i in the presence of node j perturbation (if used), and Δ to a difference with respect to perturbation (subscript p) or time (subscript t) (see Methods). Since we do not use data from the perturbation of node i

for estimation of node *i* edges, we do not have to impose assumptions on how the perturbation functionally acts on the system dynamics (see Methods). Moreover, constraints on the perturbation strength can be relaxed, following recent recommendations [41](although accuracy of the underlying Taylor series approximation can affect estimation—see Methods). If these measurements with and without perturbations were each taken in their steady state as is done in MRA, the solution for F_{ij} would be trivial. MRA gets around this by normalizing self-regulatory parameters F_{ii} to -1. Using dynamic data allows unique estimation of self-regulatory parameters without such normalization. Estimation of the node-specific stimulus strengths or basal production rates (*S's*) requires evaluation after specific functional assumptions, but in general these effects are knowable from the data to be generated (see Methods and below results).

Note that this formulation is generalizable to an *n* dimensional network. With n^2 unknown parameters in the Jacobian matrix, *n* equations originate from the vehicle perturbation and *n*-1 equations originate from each of the *n* perturbations (discarding equations from Node *i* with Perturbation *i*). This results in $n + n*(n-1) = n + n^2 - n = n^2$ independent equations.

$$S_{1,b} = -(F_{11}x_{1,ss} + F_{12}x_{2,ss})$$
$$S_{2,b} = -(F_{21}x_{1,ss} + F_{22}x_{2,ss})$$

Using Sufficient Simulated Data to Reconstruct a Network

As an initial test of the above formulation, we used a simple 2 node, single activator network where Node 1 activates Node 2, one node has first-order degradation (-1 diagonal elements), and the other has negative self-regulation (-0.8 diagonal) (Fig. 1a—see Methods for equations). A stimulus at t=0 (time-invariant; $S_{,ex}$ = 1) increases the activity of each

node, which we sample with an evenly spaced 11-point time course. This simulation was done for no perturbation (i.e. vehicle) and for each perturbation (Node 1 and Node 2) to generate the necessary simulation data per the theoretical considerations above (Fig. 1c,e,g, left panel). Here, we modeled perturbations as complete inhibition; for example, a perturbation of Node 1 makes its value 0 at all times. Solving Eqs. 3-4 to infer the Jacobian elements at each time point yielded good agreement between the median estimates and the ground truth values (Fig. 1h, "Analytic Solution", No Noise). Using the node activity data corresponding to the last time point in the time course and the median estimates of Jacobian elements, the external stimuli $S_{1,ex}$ and $S_{2,ex}$ were also determined (Eq.18-19) and reasonably agree with the ground truth values.

How does this approach fare when data are noisy? We performed the estimation with the same data but with a relatively small amount of simulated noise added (10:1 signal-to-noise—Fig. 1c,e,g). The resulting estimates are neither accurate nor precise, varying on a scale more than ten times greater than each parameter's magnitude with median predictions both positive and negative regardless of the ground truth value (Fig. 1i). The stimulus strengths $S_{1,ex}$ and $S_{2,ex}$ are estimated to be negative, while the ground truth is positive.

Although the analytic equations suggest the sufficiency of the perturbation time course datasets to uniquely estimate the edge weights, in practice even small measurement noise corrupts estimates obtained from direct solution of these equations. Therefore, we considered an alternative representation by employing a least squares estimation approach rather than solving the linear equations directly. For a given set of guesses for edge weight

and stimulus parameters, one can integrate to obtain a solution for the dynamic behavior of the resulting model, which can be directly compared to data in a least-squares sense. Least squares methods were shown to improve traditional MRA-based approaches[39,40], but had never been formulated for such dynamic problems. Two hurdles were how to model the effect of a perturbation without (i) adding additional parameters to estimate or (ii) requiring strong functional assumptions regarding perturbation action. We solved these here by using the already-available experimental measurements within the context of the least-squares estimation (see Methods). We applied this approach to the single activator model, 10:1 signal-noise ratio case above where the analytic approach failed. This new estimation approach was able to infer the network structure accurately and precisely (Fig. 1j). We conclude that analytic formulations can be useful for suggesting experimental designs that should be sufficient for obtaining unique estimates for a network reconstruction exercise, but in practice directly applying those equations may not yield precise nor accurate estimates. Alternatively, using a least-squares formulation seems to work well for this application.

Reconstruction of Random 2 and 3 Node Networks

To investigate the robustness of the least-squares estimation approach, we applied it to increasingly complex networks with larger amounts of measurement noise and smaller numbers of time points (Fig. 2). We focused on 2 and 3 node networks. We generated 50 randomized 2 and 3 node models, where each edge weight is randomly sampled from a uniform distribution over the interval [-2,2], and the basal and external strength from [0,2] (Fig. 2a, S1a, S2a). Each random network was screened for stability. Many networks

(29/50 for 2 node and 3 node) displayed potential for oscillatory behavior (non-zero imaginary parts of the eigenvalues of the Jacobian matrix). However, since the real parts of the eigenvalues are non-zero and negative, these oscillations should dampen over time, and no sustained oscillatory behavior was analyzed. For each random model, we generated a simulated dataset based on the prescribed experimental design, using complete inhibition as the perturbation. We considered evenly-spaced sampling within the time interval of 0-10 AU (approximate time to reach steady-state—Fig. S1b, S2b) with different numbers of time points (3, 7, 11 and 21), and added 10:1 signal-to-noise, 5:1 signal-to-noise, and 2:1 signal-to-noise to the data. Non-uniform time point spacing may change inference results but that was not explored at these first investigations.

For each random network model, number of time points, and noise level, we evaluated the fidelity of the proposed reconstruction approach in terms of signed directionality (Fig. 2c-f). We overall found reasonable agreement between inferred and ground truth values, even at the higher noise levels and low number of timepoints. Expectedly, the overall classification accuracy increases with more time points and decreases with higher noise levels. But, surprisingly, even in the worst case investigated of 3 timepoints and 2:1 signal-to-noise ratio, classification accuracy was above 85% for 2 node models and 70% for 3 node models. Increasing the number of nodes decreases performance, with 3-node reconstruction being slightly worse than 2-node reconstruction, other factors held constant.

We wondered whether the magnitude of an edge weight influenced its classification accuracy, since small edge weights may be more difficult to discriminate from noise. We

found that edge weights with greater absolute values, which are expected to have a greater influence on the networks, were more likely to be classified correctly (Fig. S1c-f, S2c-f). Also, for models with damped oscillatory behavior, the classification accuracy is very similar to that of all 50 random models (Fig. S3a-b).

How does this method compare to similar network reconstruction methods? There are limited methods to compare to which also use dynamic data and sequential perturbations. MRA[37], from which this method was inspired, uses steady- state data. However, we could use MRA methods requiring dynamic perturbation data as is used in our method [46,54].To compare, we further generated another set of perturbation data with 50% perturbation (as opposed to 100%). We then used the two sets of perturbation data to estimate the network node edges with dynamic modular response analysis (Fig 2g). Even in absence of noise, for low to medium numbers of timepoints (3-11) the network is not always accurately inferred (Fig. 2g). In the presence of noise, DL-MRA performs better, although the difference between the two methods becomes lower at high number of timepoints. Thus, DL-MRA not only outperforms with half the data, but it also estimates 6 additional parameters-basal production and external stimulus for each node. Although Cho's approach [47] builds upon MRA methods by recommending smaller time point intervals and smaller perturbations, for our purposes, the time intervals and perturbations are fixed and this would not affect the results obtained here. Moreover, further work has actually recommended larger perturbations while dealing with noisy data [41].

To explore a scenario where data from a node might be unavailable, we removed the data from one of the nodes in the 50 random 3 node models and used the remaining

data to reconstruct a 2-node system (Fig. S4). Comparing with corresponding model parameters in the 3 node system, we find a good but expectedly reduced classification accuracy (No Noise-94.75%, 10:1 Signal: Noise-93.75%, 5:1 Signal: Noise-91.25%, 2:1 Signal: Noise-87).

A part of the inference process is performing parameter estimation using multiple starting guesses (i.e. multi-start), and we wanted to determine how robust the estimated parameters were across the multi-start processes. We looked at the distribution of coefficient of variation (CV) among the parameters in the multi-start results in the 50 random 3 node models where either the data generated from the estimated parameters had low sum of squared errors (SSE) compared to the original data ($<10^{-4}$) or with SSE less than twice the minimum SSE. We find that the CVs peak around zero and generally have a small spread, especially for low noise scenarios (Fig. S5). This implies a good convergence of the parameter sets obtained through multi-start.

We conclude that the network parameters of 2 and 3 node systems can be robustly and uniquely estimated using DL-MRA. However, these were ideal conditions where there was no model mismatch that is expected in specific biological applications. How does DL-MRA perform when applied to data reflective of different biological use cases?

Application to Cell State Networks

Cell state transitions are central to multi-cellular organism biology. They are commonly transcriptomic in nature and underlie development and tissue homeostasis and can also play roles in disease, such as drug resistance in cancer [48,55–62]. Could DL-MRA reconstruct cell state transition networks? As the application, we use previous data on

SUM159 cells that transition between luminal, basal and stem-like cells [48]. Pure populations of luminal, basal and stem-like cells eventually grow to a stable final ratio amongst the three. The authors used a discrete time Markov transition probability model to describe the data and estimate a cell state transition network (Fig. 3a). Thus, we seek to compare DL-MRA to such a Markov model in this case.

We hypothesized that perturbations to the system in this case, in contrast to above, did not have to change node activity (i.e. edges). Rather, we thought that perturbing the equilibrium cell state distribution could serve an equivalent purpose. Thus, the data for reconstruction consisted of observing the cell state proportions evolve over time from "pure" populations (Fig. 3b), in addition to equal proportions. DL-MRA is capable of explaining the data (Fig. 3b). Interpretation of the estimated network parameters to DL-MRA depends on the transformation of the original discrete time Markov probabilities to a continuous time formulation (see Methods—there are constraints on self-regulatory parameters), but DL-MRA correctly infers the cell state transition network as well (Fig 3c). Conveniently, DL-MRA is not constrained to 1-day time point spacing as is the original discrete time Markov model.

How does noise and the number of timepoints affect the reconstruction? As above, we generated data for 50 random cell state transition models with 3, 7, 11 and 21 timepoints within 5 days, as the models generally seemed to reach close to equilibrium within 5 days. Noise levels of 10:1, 5:1 and 2:1 were used. All parameters were classified accurately (Fig. 3d) (although additional constraints in the estimation—see Methods—facilitates this classification performance). With 3 timepoints, there was deviation from perfect fit even

with no noise in the data. At 7 and higher number of timepoints, the estimates matched ground truth well, and noise expectedly reduced the accuracy (Fig. 3d). We conclude that DL-MRA can robustly infer cell state networks given perturbation data in the form of non-equilibrium proportions as initial conditions.

Application to Intracellular Signaling Networks

How does the method perform for intracellular signaling networks? The Huang-Ferrell model [49] (Fig. 4a) is a well-known intracellular signaling pathway model and has been investigated by different reconstruction methods, including previous versions of MRA [37,39,41,46,63]. It captures signal flux through a three-tiered MAPK cascade where the 2nd and 3rd tier contain two phosphorylation sites. An important aspect of the Huang-Ferrell model is that although the reaction scheme is a cascade and without obvious feedbacks, there may be hidden feedbacks due to sequestration effects and depending on how the perturbations were performed.

In order to reconstruct the Huang-Ferrell MAPK network through DL-MRA, we first simplified it to a three-node model with p-MAPKKK, pp-MAPKK and pp-MAPK as observable nodes, as is typical for reconstruction efforts (Fig. 4b) [37,39,41,46,52,63]. Second, to model perturbations, we sequentially perturbed the activation parameters of each of the observable species (k3, k15 and k27 respectively). Such perturbations, although hard to achieve experimentally, are important because modules must be "insulated" from one another and perturbations must be specific to the observables [37,52]. Even specific inhibitors do not have such kinetic specificity. Third, in the simplification of the reaction scheme, the observables are shown to influence each other but in the actual scheme, they conduct their

effects through the unphosphorylated and semi-phosphorylated species. We sought to keep the levels of these two species relatively constant between different perturbations, so that they wouldn't add to non-linearities in the estimation. Therefore, we used a stimulus which only activated the observables to a maximum of about 5% of the total forms of the protein [52]

Estimation with DL-MRA under the above conditions fits the data (Fig 4c) and predicts positive node edges down the reaction cascade (F_{21}, F_{32}), negligible direct relation between p-MAPKKK and pp-MAPK (F_{13}, F_{31}), negative self-regulation of each of the observables (F_{11}, F_{22}, F_{33}) negative feedbacks from pp-MAPKK to p-MAPKKK (F_{12}) and from pp-MAPK to pp-MAPKK (F_{23}), and negligible external stimuli on pp-MAPK to pp-MAPKK (F_{13}, F_{31}). All these effects are consistent with the reaction scheme. The negative feedback effects, although not immediately obvious, are consistent with ground truth sequestration effects. For instance, pp-MAPK has an overall negative effect on pp-MAPKK as the existence of pp-MAPK lowers the amount of species MAPK and p-MAPK which sequester pp-MAPKK and makes it avoid deactivation by its phosphatase.

How do the estimation results for the Huang Ferrell model in our method compare with those obtained from other methods? Previous work using MRA also reported negative feedbacks from successive modules to the preceding ones [37,46,52]. Similarly, self-regulation parameters in most preceding MRA based methods are also estimated to be negative but are fixed at -1[37,39,52].

Besides MRA inspired methods, SELDOM is another network reconstruction method which can also deal with dynamic data [63]. SELDOM is a data-driven method which

uses ensembles of logic based dynamic models followed by training and model reduction steps to predict state trajectories under untested conditions. However, when dealing with the Huang-Ferrell network, the true value model of SELDOM does not map the effects of self-regulation, nor feedback effects between nodes (Fig 4e). This may be explained by the fact that although SELDOM uses an extensive number of models to test the data obtained from multiple different stimuli, perturbation data was not included to test the Huang-Ferrell Model. This implies that systematic perturbation of each of the nodes, as prescribed by MRA-based methods, are necessary in order to unearth feedbacks and self-regulation effects.

Although application of DL-MRA to the Huang-Ferrell model was able to unearth latent network structure, the simulation conditions were restrictive. First, the perturbation scheme chosen in this paper, although specifically targeted at the observable species, is hard to produce experimentally. In practice, knock-down/out, overexpression, or specific inhibitors could be used as suitable perturbations, but do not have the preciseness needed to be compatible with MRA-imposed constraints. The feedback effect observed could depend on the perturbation scheme chosen-for instance knockdown of an entire module as a perturbation would likely have manifested as positive feedback to the preceding module. That is because such a knockdown would have reduced the effect of sequestration of the module on the preceding observable and would have made it more available for dephosphorylation. Second, we assumed a low stimulus to avoid effects from the unphosphorylated version of the proteins. A higher activation may increase non-linearities adding to the complexity of the model, whereas a lower stimulus may not activate enough

proteins to be well detected in experiments. The degree of activation needed for an experiment may be hard to predict beforehand. Such specific perturbations and stimulus had to be done to reduce the effects arising from the non-observable species behavior. Hence application of DL-MRA to intracellular signaling networks with multiple physical interactions needs to be carefully considered before modeling or experiments.

Application to Gene Regulatory Networks: Partial Perturbations are More Informative than Full Perturbations

Here, we applied DL-MRA further to a series of well-studied non-linear feed forward loop (FFL) gene regulatory network models that have time-varying Jacobian elements (Fig. 5a, Table 1) [32,64]. Such FFL motifs are strongly enriched in multiple organisms and are important for signaling functions such as integrative control, persistence detection, and fold-change responsiveness [65–67].

The FFL network has three nodes (x_1, x_2, and x_3), and the external stimulus acts on x_1 ($S_{1,ex}$). There is no external stimulus on x_2 and x_3; however, there may be basal production of x_2 ($S_{2,b}$) and x_3 ($S_{3,b}$),. Each node exhibits first-order decay (F_{ii}=-1). The parameters F_{12}, F_{13}, and F_{23} represent connections that do not exist in the model; we call these null edges, but we allow them to be estimated. The relationship between x_1 and x_2 (F_{21}), between x_1 and x_3 (F_{31}), or between x_2 and x_3 (F_{32}) can be either activating or inhibitory. Furthermore, x_1 and x_2 can regulate x_3 through an "AND" gate (both needed) or an "OR" gate (either sufficient) (Fig. 5a). These permutations give rise to 16 different FFL structures (Table 1).

To generate simulated experimental data from these models, we first integrated the system of ODEs starting from a *zero* initial condition to find the steady state in the absence

of stimulus. We then introduced the external stimulus and integrated the system of ODEs (see Methods) to generate time series perturbation data consistent with the proposed reconstruction algorithm, using full inhibitory perturbations. We used 11 evenly spaced timepoints for all 16 non-linear models, based on the random 3-node model analysis above, and also added noise as above.

We first noticed that even in the absence of added noise, a surprising number of inferences were incorrect (Fig. 5b, f). Model #1 (Table 1, Fig. 5b-c) is used as an example, where F_{21}, F_{31} and F_{32} are activators with an AND gate, and F_{31} is incorrectly predicted as null (Fig. 5b—compare ground truth to 100% inhibition). To understand the reason for the incorrect estimation, we looked at the node activity dynamics across the perturbation time courses (Fig. 5d). All three nodes start from an initial steady state of zero, but Node 3 is zero for all three perturbation cases. This is because of the following. Since x_1 is required for the activation of x_2 and x_3, complete inhibition of x_1 completely blocks both x_2 and x_3 activation. But, because both x_1 and x_2 are required for the activation of x_3, completely inhibiting x_2 activity also completely inhibits x_3. Thus, given this experimental setup, it is impossible to discern if x_1 directly influences x_3 or if it acts solely through x_2.

We thus reasoned that full inhibitory perturbation may suppress the information necessary to correctly reconstruct the network, but that a partial perturbation experiment may contain enough information available to make a correct estimate. If this were true, then upon applying partial perturbations (we chose 50% here), Node 3 dynamics should show differences across the perturbation time courses. Simulations showed that this is the case (Fig. 5e). Subsequently, we found that for partial perturbation data, F_{31} is correctly

identified as an activator. More broadly, we obtain perfect classification from noise-free data across all 16 FFL networks when partial perturbation data are used, as opposed to 5/16 networks having discrepancy with full perturbation data (Fig. 5f). The fits to simulated data from the reconstructed model align very closely, despite model mismatch (Fig. S6). We conclude that in these cases of non-linear networks, a partial inhibition is necessary to estimate all the network parameters accurately. Thus, moving forward, we instead applied 50% perturbation to all simulation data and proceeded with least squares estimation.

Application to Gene Regulatory Networks: Performance

The above analysis prompted us to use a partial (50%) perturbation strategy, since it classified each edge for each model in the absence of noise correctly. What classification performance do we obtain in the presence of varying levels of experimental noise? We first devised the following strategy to assess classification performance. We generated 50 bootstrapped datasets for each network structure/signal-to-noise pair, and thus obtained 50 sets of network parameter estimates. To classify the network parameters, we used a symmetric cutoff of a percentile window around the median of these 50 estimates (Fig. 6a). We illustrate this approach with three different example edges and associated estimates, one being positive (Edge 1), one being negative (Edge 2), and one being null (Edge 3). Given the window of values defined by the percentile cutoff being chosen, if the estimates in this window are all positive, then the network parameter would be classified as positive. Similarly, if the estimates in this window are all negative, then the parameter would be classified as negative. Finally, if the estimates in the window cross zero (i.e. span both positive and negative terms), then it would be classified as null. First, consider the case that

the percentile window is just set at the median with no percentile span. Then, the classifications for true positives and negatives are likely to be accurate while the null parameters are likely to be incorrectly categorized as either positive or negative (Fig. 6a). If we increase the percentile window span slightly (e.g. between the 40th and 60th percentile, middle panel), we can categorize null edges better, while maintaining good classification accuracy of both true positive and negative edges. However, if we relax the percentile window too much, (e.g. between the 10th and 90th percentile, far right panel) we may categorize most parameters as null, including the true positive and negatives. Thus, it is clear there will be an optimal percentile cutoff that maximizes true positives and minimizes false positives as the threshold is shifted from the median to the entire range.

Now, we applied this classification strategy to the 16 FFL model estimates from data with different noise levels. We varied the percentile window from the median only (50) to the entire range of estimated values (100) and calculated the true and false positive rates for all edges across all 16 FFL models, which allowed generation of receiver operator characteristic (ROC) curves (Fig. 6b). For each noise level, we chose the percentile window that yielded a 5% false positive rate (13-87 percentile for 10:1 Signal:Noise, 19-81 percentile for 5:1 Signal:Noise, and 21-79 percentile for 2:1 Signal:Noise). Using this simple cutoff classifier, we observed good classification performance across all noise levels according to traditional area under the ROC curve metrics (10:1 AUC=0.99, 5:1 AUC=0.9, 2:1 AUC=0.92).

How does classification accuracy break down by FFL model and edge type? To evaluate the performance for each of the 16 FFL cases, we calculated the fraction of the 12

links in each FFL model that was classified correctly as a function of signal-to-noise, given the percentile windows determined above (Fig. 6c). We also looked at the fraction of the 16 models where each of the 12 links were correctly classified (Fig. 6d). Perfect classification is a value of one, which is the case for no noise, and for many cases with 10:1 signal-to-noise.

In general, as noise level increases, prediction accuracy decreases, as expected. Although for some models and parameters, performance at 2:1 signal-to-noise is poor, in some cases it is surprisingly good. This suggests that the proposed method can yield information even in high noise cases; this information is particularly impactful for null, self-regulatory, and stimulus edges. High noise has strong effects on inference of edges that are either distinct across models, time variant or reliant on other node activities (F_{21}, F_{31}, F_{32}) (Fig. 6c-d, S7). $F_{21,}$ which is reliant on activity of x_1, is inferred better than F_{31} and F_{32}. This may be caused by the fact that x_3 dynamics depend on both x_1 and x_2, whereas x_2 dynamics only depend on x_1.

Comparing across models, we find that Models 1-8 are reconstructed slightly better than Models 9-16 (Fig. 6c) when noise is high. This performance gap is predominantly caused by $S_{3,b}$ misclassification—basal production of Node 3 (Fig. S7). What is the reason for the possible misclassification of $S_{3,b}$ in Models 9-16? We know that $S_{3,b}$ depends on the initial values of x_1, x_2 and x_3 and the estimated values of F_{31}, F_{32} and F_{33} (See Methods, Eq. 19). For Models 1-8, $x_1(t=0)$ and $x_2(t=0)$ are both zero and therefore $S_{3,b}$ is effectively only dependent on estimated value of F_{33} and $x_3(t=0)$ (Fig. S6 and Methods). But for Models 9-16, $x_2(t=0)$ is non-zero and $S_{3,b}$ is dependent on the estimated values of both F_{32} and F_{33}, in

addition to $x_2(t=0)$ and $x_3(t=0)$, which increases the variability of $S_{3,b}$ estimates. Therefore with high levels of noise, $S_{3,b}$ is more likely to be mis-classified in Models 9-16, whereas this does not happen in Models 1-8 (Fig.6c,d, S7). In the future, including stimulus and basal production parameters in the least squares estimations themselves, rather than further deriving algebraic relations to estimate them, will likely help improve reliability.

We conclude that (i) when dealing with non-linear gene regulatory networks, complete perturbations such as genetic knockouts may fundamentally impede one's ability to deduce network architecture and (ii) this class of non-linear networks can be reconstructed with reasonable performance using the proposed strategy employing partial perturbations.

Discussion

Despite intensive research focus on network reconstruction, there is still room to improve discrimination between direct and indirect edges (towards causality), particularly when biologically-ubiquitous feedback and feedforward cycles are present that stymie many statistical or correlation-based methods, and given that experimental noise is inevitable. The presented DL-MRA method prescribes a realistic experimental design for inference of signed, directed edges when typical levels of noise are present. It allows estimation of self-regulation edges as well as those for basal production and external stimuli. For 2 and 3 node networks, the method can successfully handle random linear networks, cell state transition networks, and gene regulatory networks, and, under certain limiting conditions, signaling networks. Prediction accuracy improved with more timepoints, which in our case accounted for more relevant dynamic data. However, we would like to stress that here we did not explore time point placement, which likely underlies the performance increase rather than simply number of timepoints. Prediction accuracy was strong in many cases even with simulated noise that exceeds typical experimental variability (2:1 signal-to-noise). The method presented here is quite general and could be applied not only to cell and molecular biology, but also vastly different fields where perturbation time course experiments are possible, and where network structures are important to determine.

One type of non-linear model that we did not investigate is one with sustained oscillations, such as those found in the cell cycle[68], or sometimes even MAPK signaling pathways[69–71]. We found that in our application to general two and three node linear

models, DL-MRA could reconstruct multiple networks that have damped oscillatory behavior (Fig 1b). However, we expect time point measurement selection and frequency to be much more important for inferring networks that give rise to sustained oscillations, with time points comprehensively covering peaks and troughs, and the frequency high enough to do so. We do expect that the method could infer the structure of such networks given appropriate sampling, but this requires a much deeper investigation.

MRA and its subsequent methods allow for inference of direct edges by prescribing systematic perturbation of each node [37,39,41,43,45] and the idea of directionality has been followed through in DL-MRA. Often, such edge directness is equated to causality, but this is not necessarily the case, especially when the entire system is not explicitly represented. In practice, the causality and strength of an edge may be dependent on how well the model represents the underlying phenomenon and might be affected by simplification of larger networks, non-linearities in the actual model and even by noise in the data. Secondly, in discussions about causal system inferences, consideration of the counterfactuals is important [30,31,50,51]. For a network of nodes going through dynamics, the counterfactuals to intrinsic network edges causing the dynamics would be the environmental factors extrinsic to the network edges. In DL-MRA, by evaluating external stimuli and basal production as well as the network edges, we have mapped some counterfactuals to node dynamics, thus presenting a more complete map of the causal factors to the network dynamics compared to methods which only show network edges. This also allows for a concise mapping of the environmental contexts in which the network edges are reconstructed.

Application of DL-MRA could reconstruct cell state transition networks based on discrete time Markov transition models, with the added benefit of not being constrained to specific time intervals. It can also successfully handle noisy data. The additional constraints in DL-MRA in the context of cell state transitions (summations of transition rates—see Methods) implies that the underlying network may be estimated even with less data requirements than in other cases. This method can be a useful tool to model cell state transitions and predict cell state. Perturbations were modeled as a difference in initial states, and that worked well in this case, suggesting that such modeling of perturbations may work in other cell state transition or biological networks.

Although application of DL-MRA to an intracellular signaling network (Huang-Ferrell MAPK) was able to explain its ground truth, including feedback due to sequestration, the method was constrained to specific, difficult-to-implement perturbations and a low stimulus which may not always be feasible experimentally. Specific inhibitors could be a source of perturbation, but even they influence more kinetic parameters than was required here for a clean solution. In MRA, a larger reaction scheme is often simplified into modules with one species in the module representing the activity of the module. But often, the activity of the other species in the module is implicit and becomes significant in dictating how perturbations and stimulus affect the network dynamics. Moreover, the type of perturbation chosen, such as specific inhibitors versus knock-down, also may yield different network inference results. Therefore, the use of MRA methods on simplified large intracellular signaling networks, especially while dealing with experiments, have significant caveats that should be carefully considered [41,72].

Although complete inhibition is often used for perturbation studies of gene regulatory networks (e.g. CRISPR-mediated gene knockout), we found that partial inhibition is important to fully reconstruct the considered non-linear gene regulatory networks. It is important to distinguish here, however, small perturbations vs. partial perturbations. Small perturbations are formally recommended for both MRA and other techniques [72] where the effects of noise are not extensively explored. In practice however, there is a tradeoff between perturbation strength and feasibility, since the effects of small perturbations are masked by noise [41]. Partial perturbations, as considered in this work (~50%) are much larger than what are typically considered small perturbations. The theoretical formulation of DL-MRA reduces the impact of not having small perturbations, because perturbation data from a particular node is not used for inference of edges connected to that node. Yet, DL-MRA still uses linearizations of the Jacobian which are are always subject to greater inaccuracy the further away from reference points such perturbations take the system. Since many biological networks share the same types of non-linear features contained within the considered FFL models, this is not likely to be the only case when partial inhibition will be important. We are thus inclined to speculate that large partial perturbations may be a generally important experimental design criterion moving forward. Partial inhibition is often "built-in" to certain assay types, such as si/shRNA or pharmacological inhibition that are titratable to a certain extent.

One major remaining challenge is scaling to larger networks. Here, we limited our analysis to 2 and 3 node networks. Conveniently, the number of necessary perturbation time courses needed grows linearly (as opposed to exponentially) with the number of

considered nodes. Furthermore, as long as system-wide or omics-scale assays are available, the experimental workload also grows linearly. This is routine for transcriptome analyses [73], and is becoming even more commonplace for proteomic assays (e.g. mass cytometry [74], cyclic immunofluorescence [75], mass spectrometry [76], RPPA [77]. Thus, the method is arguably *experimentally* scalable to larger networks.

However, the *computational* scaling past 2 and 3 node models remains to be determined and is likely to require different approaches for parameter estimation. Increasing the network size will quadratically increase the number of unknown parameters, which will significantly increase the computational requirements for obtaining robust solutions. Yet, recent work has shown that large estimation problems in ODE models may be broken into several smaller problems[78], which may be applicable here, and is likely to yield large computational speed up by allowing parallelization of much smaller tasks. However, theory on how to merge potentially discrepant results between independently estimated overlapping subnetworks would need to be derived. Importantly, we saw in the linear 2 and 3 node model examples that the impact of experimental noise was larger for 3 node models, implying that increasing the number of nodes past 3 will further increase the impact of experimental noise. Another synergistic avenue could be imposing prior knowledge to improve initial parameter guesses and even reduce the parametric space, such as in Bayesian Modular Response Analysis [45], or with functional database information [79]. Such prior knowledge could also help inform emergent network properties as network size grows, such as degree distributions for scale-free networks [2]. Here, we only investigated dense subnetworks, so sparseness patterns and judicious allocation of non-zero Jacobian

elements could also have great impact on estimation for large networks. Overall, application to larger networks is of great interest but these non-trivial computational roadblocks must be solved first.

In conclusion, the proposed approach to network reconstruction is systematic and feasible, robustly operating in the presence of experimental noise and accepting data from large perturbations. It addresses important features of biological networks that current methods struggle to account for: causality/directionality/sign, cycles (including self-regulation), dynamic behavior and environmental stimuli. It does so while leveraging dynamic data of the network and only requires one perturbation per node for completeness. We expect this approach to be broadly useful not only for reconstruction of biological networks, but to enable using such networks to build more predictive models of disease and response to treatment, and more broadly, to other fields where such networks are important for system behavior.

Methods

Deriving Sufficiency Conditions for Unique Estimation of Jacobian Elements

The first-order partial derivatives comprising **J** (Eq. 2) can be approximated by a first-order Taylor series expansion of Eq. 1 about a time point *k*

$$f_1(k+1) \approx f_1(k) + \frac{\partial}{\partial x_1}\left(f_1(k)\right).\left(x_1(k+1) - x_1(k)\right) + \frac{\partial}{\partial x_2}\left(f_1(k)\right)\cdot\left(x_2(k+1) - x_2(k)\right) \quad (5)$$

$$f_2(k+1) \approx f_2(k) + \frac{\partial}{\partial x_1}\left(f_2(k)\right)\cdot\left(x_1(k+1) - x_1(k)\right) + \frac{\partial}{\partial x_2}\left(f_2(k)\right)\cdot\left(x_2(k+1) - x_2(k)\right) \quad (6)$$

Eq. 5-6 may be written more succinctly as

$$\begin{aligned} y_1(k+1) &\approx F_{11}(k)\cdot\Delta_t x_1(k+1) + F_{12}(k)\cdot\Delta_t x_2(k+1) \\ y_2(k+1) &\approx F_{21}(k)\cdot\Delta_t x_1(k+1) + F_{22}(k)\cdot\Delta_t x_2(k+1) \end{aligned} \quad (7)$$

where

$$y_i(k+1) \equiv f_i(k+1) - f_i(k); \Delta_t x_i(k+1) \equiv x_i(k+1) - x_i(k)\,. \quad (8)$$

The approximation in Eq. 7 becomes more accurate as more time points are measured. Also, the edge weights are potentially time-dependent, although this is rarely considered when describing biological networks.

How do we estimate the edge weights *F* in Eq. 7 and thus reconstruct the network? Time series data can inform x_i's and f_i's as a function of time, following application of a stimulus. Given such stimulus-response data, however, for each time point there are only two equations for four unknowns, an underdetermined system for which more data are needed.

Consider now stimulus-response time course data in the presence of single perturbations. Let p_i be a variable that reflects the strength and/or presence of different potential perturbations: p_1 represents perturbation of x_1 and p_2 represents perturbation of

x_2. If p_j is not explicitly written, its value is zero and/or it has no effect. Now, the ODEs become a function of the perturbation variables

$$f_{i,j}(k) \equiv f_i(k, p_j) = f_i(x_1(k), x_2(k), p_j) \qquad (9)$$

The 1st order Taylor series expansions for cases with perturbations become

$$y_{1,1}(k) \approx F_{11}(k) \cdot \Delta_{p,1} x_1(k) + F_{12}(k) \cdot \Delta_{p,1} x_2(k) + \frac{\partial}{\partial p_1}\left(f_1(k)\right) \cdot p_1 \qquad (10)$$

$$y_{1,2}(k) \approx F_{11}(k) \cdot \Delta_{p,2} x_1(k) + F_{12}(k) \cdot \Delta_{p,2} x_2(k) + \frac{\partial}{\partial p_2}\left(f_1(k)\right) \cdot p_2 \qquad (11)$$

$$y_{2,1}(k) \approx F_{21}(k) \cdot \Delta_{p,1} x_1(k) + F_{22}(k) \cdot \Delta_{p,1} x_2(k) + \frac{\partial}{\partial p_1}\left(f_2(k)\right) \cdot p_1 \qquad (12)$$

$$y_{2,2}(k) \approx F_{21}(k) \cdot \Delta_{p,2} x_1(k) + F_{22}(k) \cdot \Delta_{p,2} x_2(k) + \frac{\partial}{\partial p_2}\left(f_2(k)\right) \cdot p_2 \qquad (13)$$

where

$$y_{i,j}(k) \equiv f_{i,j}(k) - f_i(k); \Delta_{p,j} x_i(k) \equiv x_i(k, p_j) - x_i(k) \qquad (14)$$

Here, we have expanded with respect to the perturbation, rather than with respect to time as previously. However, since the reference point is the same, the Jacobian elements remain identical in these equations. It is also interesting to note that the Jacobian elements, or network, may be affected by the perturbation, but we do not necessarily have to know those effects mathematically, since the reference point is the same. Now we have six potential equations with which to estimate the four Jacobian elements. However, we must make some determination as to how the perturbations p_1 and p_2 directly affect Node 1 and Node 2 dynamics f_1 and f_2 to account for the perturbation variable partial derivatives.

By design, the Node 1 perturbation has significant *direct* effects on Node 1 dynamics, and similarly for the Node 2 perturbation on Node 2 dynamics. Using equations including $\partial f_1/\partial p_1$ and $\partial f_2/\partial p_2$ require precise definition of perturbation strength and their effects on dynamics, which could be difficult to determine experimentally and implement in simulations. Therefore, we do not employ equations involving such terms. On the other hand, if the Node 1 perturbation has negligible *direct* effect on Node 2 dynamics, that is, the effects on Node 2 dynamics are through the network (i.e. p_1) and not explicit in f_2), and similarly the Node 2 perturbation has negligible *direct* effect on Node 1 dynamics, then $\partial f_2/\partial p_1$ and $\partial f_1/\partial p_2$ are approximately zero. This mild condition is often the case experimentally. The only determining factors for the suitability of the Taylor series truncation are the spacing of time points and the accuracy of the expansion about the perturbation difference. From this, the main set of linear equations presented in Eq. 3-4 are obtained.

General Estimation Model Equations

We employ the following general model for a two-node network: -

$$\begin{aligned}\frac{dx_1}{dt} &= f_1(x_1, x_2) = S_1 + F_{11}x_1 + F_{12}x_2 \\ \frac{dx_2}{dt} &= f_2(x_1, x_2) = S_2 + F_{21}x_1 + F_{22}x_2\end{aligned} \qquad (15)$$

Here, S_1 and S_2 are the stimuli strengths on Node 1 and Node 2 respectively, and F_{11}, F_{12}, F_{21} and F_{22} are the network edge weights (Figure 1a). In many systems, there may be a basal or constitutive production driving the node activities, besides an external stimulus. For these cases, the Stimulus term (S_i), may be considered as an addition of these

two effects- the basal production term ($S_{i,b}$) and the external stimulus ($S_{i,ex}$). Then the two-node model can be represented by the following equations-

$$\begin{aligned} \frac{dx_1}{dt} &= S_{1,b} + S_{1,ex} + F_{11}x_1 + F_{12}x_2 \\ \frac{dx_2}{dt} &= S_{2,b} + S_{2,ex} + F_{21}x_1 + F_{22}x_2 \end{aligned} \qquad (16)$$

Or more generally,

$$\frac{dx_i}{dt} = S_{i,b} + S_{i,ex} + \sum_{j=1}^{n} F_{ij}x_j \ , \qquad (17)$$

where n is the total number of nodes.

When a steady state exists, the dx_i/dt terms become zero and it becomes easy to represent the stimulus terms as a function of the node activities (x_i) and network edges (F_{ij}).

$$S_{i,b} + S_{i,ex} = -(\sum_{j=1}^{n} F_{ij}x_{i,ss}) \qquad (18)$$

This is helpful to understand that the perturbation time course data also generally constrains not only the edge weights, but also the stimulus terms. For a system at a steady state without an external stimulus, for example at t=0:

$$S_{i,b} = -(\sum_{j=1}^{n} F_{ij}x_{i,ss}) \qquad (19)$$

The Two-node Single Activator model

The two-node single activator model (Fig. 1a, S1a) is described by

$$\begin{aligned} \frac{dx_1}{dt} &= f_1(x_1, x_2) = 1 - x_1 \\ \frac{dx_2}{dt} &= f_2(x_1, x_2) = 1 + 1.5x_1 - 0.8x_2 \end{aligned} \qquad (20)$$

Here, $S_{1,ex}=1$, $F_{11}=-1$, $F_{12}=0$, $S_{2,ex}=1$, $F_{21}=1.5$, $F_{22}=-0.8$. The basal production terms are both zero, for simplicity, and the initial conditions for $x_1(t=0)$ and $x_2(t=0)$ are zero. The stimulus terms $S_{i,ex}$ are calculated through Eq. 18, using the median values of F_{ij} and the $x_i(t=10)$, when the system reaches near steady state.

Random Two-node and Three-node models

The random 2 node network is described by

$$\begin{aligned}\frac{dx_1}{dt} &= f_1(x_1, x_2) = S_{1,b} + S_{1,ex} + F_{11}x_1 + F_{12}x_2 \\ \frac{dx_2}{dt} &= f_2(x_1, x_2) = S_{2,b} + S_{2,ex} + F_{21}x_1 + F_{22}x_2\end{aligned} \qquad (21)$$

Values for $S_{1,b}$, $S_{2,b}$, $S_{1,ex}$ and $S_{2,ex}$ are sampled from a uniform distribution over the range [0,2] and values for F_{11}, F_{12}, F_{21}, and F_{22} are sampled from a uniform distribution over the range [-2,2] using the MATLAB function `rand`. To capture basal activity, we use a two-step approach. First, starting from node activity values of zero, without the external stimulus on Node 1 and Node 2 ($S_{1,ex}=S_{2,ex}=0$ in Eq 22) we simulate until the network reaches steady-state with just basal production driving the network behavior. Then, we introduce the external stimulus on Node 1 and Node 2, integrate the ODEs, and sample evenly spaced time-points using `ode15s` in MATLAB with default settings. We sample 3,7, 11, and 21 evenly spaced time points across a time course, from 0 to 10 arbitrary time units in all the cases.

The random 3 node networks use the same sampling rules as the 2 node networks with the following equations.

$$\frac{dx_1}{dt} = f_1(x_1, x_2, x_3) = S_{1,b} + S_{1,ex} + F_{11}x_1 + F_{12}x_2 + F_{13}x_3$$
$$\frac{dx_2}{dt} = f_2(x_1, x_2, x_3) = S_{2,b} + S_{2,ex} + F_{21}x_1 + F_{22}x_2 + F_{23}x_3 \qquad (22)$$
$$\frac{dx_3}{dt} = f_3(x_1, x_2, x_3) = S_{3,b} + S_{3,ex} + F_{31}x_1 + F_{32}x_2 + F_{33}x_3$$

Intracellular Signaling Networks

In the simplification of the Huang-Ferrell network to three nodes, p-MAPKKK, pp-MAPKK and pp-MAPK were taken as nodes. Since, in absence of external stimuli, the basal values of the nodes are zero, the basal production was estimated as zero beforehand and not considered in the estimation of the rest of the network. Aside from the basal production edges, a full 3 node network (Fig 4b) was estimated from the simulation data of each of the observables. After estimation, parameters with values less than 1/100th of the largest parameter, were considered negligible.

Cell State Transition Models

The cell transition model [48] is a discrete time Markov probability model. Here, we show how this form is related to the ODE model used in DL-MRA. Starting at any initial value, each next step representing a time difference of one day follows from the previous time point as follows-

$$x_{1,t+1} = M_{11}x_{1,t} + M_{12}x_{2,t} + M_{13}x_{3,t}$$
$$x_{2,t+1} = M_{21}x_{1,t} + M_{22}x_{2,t} + M_{23}x_{3,t} \qquad (23)$$
$$x_{3,t+1} = M_{31}x_{1,t} + M_{32}x_{2,t} + M_{33}x_{3,t}$$

Where M_{ij} denotes the Markov transition probabilities of species *j* into species *i*. In matrix form it may be represented as follows-

$$\begin{bmatrix} x_{1,t+1} \\ x_{2,t+1} \\ x_{3,t+1} \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{bmatrix} \begin{bmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{bmatrix} \quad (24)$$

Representing the Markov parameter matrix as *M* and the species relative concentration variables as vector X, the equation becomes

$$X_{t+1} = MX_t \quad (25)$$

The Markov transition probabilities for a species must add up to 1. In experimental terms, a species can either transition to other species or stay the same and the sum of all those probabilities is 1.

$$\sum_{i=1:3} M_{ij} = 1 \quad (26)$$

As a first step in relating these equations to the ODE form underlying DL-MRA, we put the variables in terms in terms of Δx (with respect to time),

$$X_{t+1} - X_t = MX_t - X_t \quad (27)$$

$$\Delta X_{t+1} = (M - I)X_t \quad (28)$$

$$\Delta X_{t+1} = M' X_t \quad (29)$$

Where ***M'*** is ***M-I***, and ***I*** is the identity matrix. ***M'*** is ***M***, except that 1 is subtracted from all its diagonal elements. Hence Eq. 26 for *M'* becomes

$$\sum_{i=1:3} M'_{ij} = 0 \quad (30)$$

This also implies that the diagonal term for ***M'*** is negative of the sum of the other two terms in the same column. In experimental terms, the amount of reduction of a species is equal to how much it got converted to other species.

The above equations apply for the cases where Δt is 1. We can incorporate arbitrary time steps as

$$\Delta X_{t+\Delta t} = M'_{\Delta t} X_t \Delta t \qquad (31)$$

Where Δt is the scalar value of time difference and $M'_{\Delta t}$ is the matrix of the set of parameters, specific to the time difference chosen. For a case where Δt tends to 0, the equation becomes-

$$\lim_{\Delta t \to 0} (\Delta X_{t+\Delta t} / \Delta t) = M'_{dt} X_t \qquad (32)$$

$$dX / dt = M'_{dt} X_t \qquad (33)$$

Where M'_{dt} is the matrix of the set of parameters specific to the case where Δt is infinitesimally small. Note that Eq. 33 is similar in form to Eq. 22, only without the extra stimulus terms and where M'_{dt} is equivalent to the Jacobian matrix F with terms F_{ij}. There would be an added constraint that the sum of the terms in the same column would add up to zero, or that the diagonal term is the negative of the sum of the other two terms in the same column.

$$\frac{dX}{dt} = FX_t \qquad (34)$$

$$F_{ii} = -\sum_{j=1, j\neq i}^{3} F_{ij} \qquad (35)$$

Non-Linear Models

The non-linear feedforward loop models [32] are described by:

$$\frac{dx_1}{dt} = f_1(x_1,x_2,x_3) = 1 - x_1$$

$$\frac{dx_2}{dt} = f_2(x_1,x_2,x_3) = f(x_1, K_{x_1x_2}) - x_2 \qquad (36)$$

$$\frac{dx_3}{dt} = f_3(x_1,x_2,x_3) = G(x_1, K_{x_1x_3}, x_2, .K_{x_2x_3}) - x_3$$

When an AND gate is present

$$G(x_1, K_{x_1x_3}, x_2, K_{x_2x_3}) = f(x_1, K_{x_1x_3}) * f(x_2, K_{x_2x_3}) \qquad (37)$$

When an OR gate is present

$$G(x_1, K_{x_1x_3}, x_2, K_{x_2x_3}) = fc(x_1, K_{x_1x_3}, K_{x_2x_3}, x_2) + fc(x_2, K_{x_2x_3}, K_{x_1x_3}, x_1) \qquad (38)$$

For a given u, v ϵ {x_1, x_2, x_3} and K, K_u, K_v ϵ { $K_{x_1x_2}$, $K_{x_1x_3}$, $K_{x_2x_3}$ }:

If u activates its target, then:

$$f(u,K) = \frac{\left(\frac{u}{K}\right)^2}{1+\left(\frac{u}{K}\right)^2} \; ; \; fc(u,K_u,K_v,v) = \frac{\left(\frac{u}{K_u}\right)^2}{1+\left(\frac{u}{K_u}\right)^2+\left(\frac{v}{K_v}\right)^2} \qquad (39)$$

If u represses its target, then:

$$f(u,K) = \frac{1}{1+\left(\frac{u}{K}\right)^2} \; ; \; fc(u,K_u,K_v,v) = \frac{1}{1+\left(\frac{u}{K_u}\right)^2+\left(\frac{v}{K_v}\right)^2} \qquad (40)$$

Effectively, an external stimulus of '$S_{1,ex}=1$', acts on Node 1 at $t=0$ and is propagated through the network. There is no external stimulus acting on Node 2 and Node 3. However, in many cases there is basal production in one or both of Node 2 and Node 3.

This leads to a non-zero steady-state of the network before the external stimulus is introduced.

To capture basal activity, we use a two-step approach. First, starting from node activity values of zero, without the external stimulus on Node 1 ($S_{1,ex}=0$), we simulate until the network reaches steady-state. Then, we introduce the external stimulus on Node 1, integrate the ODEs, and sample 11 evenly spaced time-points using `ode15s` in MATLAB with default settings and steady-state node values without the external stimulus as the initial conditions. We chose 11 timepoints because it yields good classification accuracy for the above random 3 node model even in presence of noisy data. For each of the 16 non-linear models, the values of the parameters (*K, Ku, Kv*), were varied and chosen so that the resulting node activity data are responsive to the stimulus and perturbations (Fig. S6, See Supplementary Code for values).

Modeling Perturbations

Precisely modeling perturbations can be a challenge, since experimentally, there may be several ways of causing a perturbation with different mechanisms such as siRNAs, competitive/non-competitive/uncompetitive inhibition, etc. It may be hard to quantify how much a perturbation is affecting a node, in terms of its dynamics (i.e. right-hand sides of the ODEs). Therefore, we employ the following approaches which circumvent the need to model how each perturbation mechanistically manifests in the ODEs during parameter estimation. There are two cases to consider: (i) when we have a perturbation of node *i* and we need to simulate node *i* dynamics; (ii) when we have a perturbation of node *i* and we

need to simulate other node *j* dynamics. To illustrate the approach, we use the above-described 2 node model with an example of a Node 1 perturbation. Recall that

$$\frac{dx_1}{dt} = S_{1,b} + S_{1,ex} + F_{11}x_1 + F_{12}x_2$$
$$\frac{dx_2}{dt} = S_{2,b} + S_{1,ex} + F_{21}x_1 + F_{22}x_2 \qquad (41)$$

For case (i), we have to obtain values for x_1 under perturbation of Node 1. We refer to the perturbed time-course as $x_{1,1}$. In experimental situations, $x_{1,1}$ would be measured directly. To obtain simulation data for $x_{1,1}$ we use the following:

$$x_{1,1}(k) = p_1 \times x_1(k), \qquad (42)$$

where x_1 is obtained from the simulations without perturbations, and recall that *k* refers to time point *k*. For a 50% inhibition, p=0.5 and for a complete inhibition, p=0.

For case (ii), we have to obtain the values for x_2 under perturbation of Node 1, which we refer to as $x_{2,1}$. To do this, we have to integrate the ODE for dx_2/dt, but using $x_{1,1}$ values, as follows

$$\frac{dx_{2,1}}{dt} = S_{2,b} + S_{2,ex} + F_{21}x_{1,1} + F_{22}x_{2,1} \qquad (43)$$

Here, x_2 has been replaced with $x_{2,1}$ to represent x_2 under perturbation of Node 1, for clarity. To solve this equation, we simply use the "measured" $x_{1,1}$ time course directly in the ODE.

When data are generated by simulations, there is little practical limit to temporal resolution, but with real data, to solve Eq. 43 one may need values for $x_{1,1}$ at multiple time points where measurements are not available, depending on the solver being used. We

therefore fit $x_{1,1}$ data to a polynomial using `polyfit` in MATLAB, and use the polynomial to interpolate given a required time point. In this work, we have used an order of 5 to fit the data as well as avoid overfitting, but the functional form is quite malleable so long as it captures the data trends.

For modeling perturbations of the cell transition model, the initial value of the simulated data for the perturbed node was taken as zero during simulation. The estimation was performed in a similar way as a random 3 node network as described above.

For modeling perturbations for the Huang Ferrell model, the parameters k3, k15 and k27 were sequentially set as zero. The estimation was performed in a similar way as a random 3 node network as described above.

Simulated Noise

Normally distributed white (zero mean) noise is added to simulated time courses point-wise with

$$y = x + N(0, d \cdot x) \qquad (44)$$

where x is the simulation data point, y is the noisy data point, and d represents the noise level. Signal-to-noise ratio of 10:1, 5:1 and 2:1 are, respectively d = 0.1, 0.2, and 0.5. Normally distributed samples are obtained using `randn` in MATLAB. While there are many different distributional options for modeling noise, we chose this for simplicity and to capture the effects generically of noisier data. We do not intend to answer questions related to whether specific distributional assumptions about the form of the noise have significant impact of the methods performance.

Parameter Estimation

For the two-node model, the entire network, with and without perturbations, can be explained by the following system of equations

$$\begin{aligned}
\frac{dx_1}{dt} &= S_{1,b} + S_{1,ex} + F_{11}x_1 + F_{12}x_2 \\
\frac{dx_2}{dt} &= S_{2,b} + S_{2,ex} + F_{21}x_1 + F_{22}x_2 \\
\frac{dx_{2,1}}{dt} &= S_{2,b} + S_{2,ex} + F_{21}x_{1,1} + F_{22}x_{2,1} \\
\frac{dx_{1,2}}{dt} &= S_{1,b} + S_{1,ex} + F_{11}x_{1,2} + F_{12}x_{2,2}
\end{aligned} \qquad (45)$$

where $x_{1,1}$ and $x_{2,2}$ are the perturbed node values, from either simulated or experimental data. Eight parameters ($S_{1,b}$, $S_{1,ex}$, F_{11}, F_{12} , $S_{2,b}$, $S_{2,ex}$, F_{21}, F_{22}) need to be estimated to fully reconstruct this network. We seek a set of parameters that minimizes deviation between simulated and measured dynamics.

For an initial guess, the node edge parameters (F_{ij}) are randomly sampled from a uniform distribution over the interval [-2,2] and the stimulus parameters ($S_{i,ex}$) are sampled from a uniform distribution over the interval [0,2]. Using data at $t=0$, which corresponds to a steady-state without $S_{i,ex}$, the $S_{i,b}$ can be estimated during each iteration of the estimation as follows-

$$\begin{aligned}
\hat{S}_{1,b} &= -(\hat{F}_{11}x_1(t=0) + \hat{F}_{12}x_2(t=0)) \\
\hat{S}_{2,b} &= -(\hat{F}_{21}x_1(t=0) + \hat{F}_{22}x_2(t=0))
\end{aligned} \qquad (46)$$

For an n-node model, this equation can be scaled accordingly to obtain each $\hat{S}_{i,b}$.

For these initial guesses we compute the activity data using Eq. 45. The perturbation data $x_{k,k}$ is used in the perturbation equations as detailed above (Eq. 43). Let $\hat{x}_i$, and $\hat{x}_{i,j}$ denote the predicted node activity values for non-perturbed and perturbed cases

respectively. For a total of n nodes and N_t timepoints, the objective function is the sum of squared errors Φ

$$\Phi = \sum_{k=1}^{N_t}\left[\left(\sum_{i=1}^{n}\left(x_i(k)-\hat{x}_i(k)\right)^2\right)+\sum_{i=1}^{n}\sum_{j\neq i}\left(x_{i,j}(k)-\hat{x}_{i,j}(k)\right)^2\right] \quad (47)$$

Note here that we do not use data from node j, when perturbation j was used (per the derivation). The MATLAB function `fmincon` is used to minimize Φ by changing edge weights and stimulus terms within the range [-10,10].

We employ "multi-start" by running the estimation 10 times, starting from different randomly generated initial starting points [80]. The estimated parameters and their respective final sum of squared errors (Φ) are saved and the estimated parameter set corresponding to the minimum Φ is chosen as the final parameter set. Variability of parameter estimates across multi-start runs is explored in Supplementary Figure S5.

Parameter Estimation for Non-Linear Models

For estimating the Non-Linear models, we start with a prior knowledge that $S_{1,b}$ is always *zero* and $S_{2,ex}$ and $S_{3,ex}$ are always zero as well, which is directly evident from x_1 initial conditions and x_2, x_3 stimulus response in the presence of a complete Node 1 perturbation. The equations for the non-perturbation case become as follows

$$\begin{aligned}
\frac{dx_1}{dt} &= S_{1,ex} + F_{11}x_1 + F_{12}x_2 + F_{13}x_3 \\
\frac{dx_2}{dt} &= S_{2,b} + F_{21}x_1 + F_{22}x_2 + F_{23}x_3 \\
\frac{dx_3}{dt} &= S_{3,b} + F_{31}x_1 + F_{32}x_2 + F_{33}x_3
\end{aligned} \quad (48)$$

Since the system is at steady-state before the external stimulus, the basal production parameter can be estimated during each iteration of the estimation as-

$$\begin{aligned}\hat{S}_{2,b} &= -(\hat{F}_{21}x_1(t=0) + \hat{F}_{22}x_2(t=0) + \hat{F}_{23}x_3(t=0)) \\ \hat{S}_{3,b} &= -(\hat{F}_{31}x_1(t=0) + \hat{F}_{32}x_2(t=0) + \hat{F}_{33}x_3(t=0))\end{aligned} \quad (49)$$

where $\hat{F}_{i,j}$ are the current model parameter estimates and $x_i\,(t=0)$ are the x values at the initial system steady state before the induction of external stimulus.

Bootstrapping Simulated Data for the FFL Model Cases

To generate multiple parameter set estimates to classify edge weights for the FFL model cases, we employ a bootstrapping approach. In an experimental scenario, each data point will have a mean and a standard deviation, and upon a distributional assumption (e.g. normal), one can then resample datasets to obtain measures of estimation uncertainty. We use the simulated data as the mean, and then vary the standard deviation as described above to generate 50 bootstrapped datasets for each of the 16 considered models. Estimation is carried out for each of the 50 datasets using multi-start, which yields 50 best-fitting parameter sets for each model. Uncertainty analysis and classification error is based on these sets.

Data Availability

All relevant simulated data used in the paper are provided and can be accessed along with the code at https://doi.org/10.5281/zenodo.6516238. Any other relevant data can be obtained from the authors.

Code Availability

The code needed to reproduce the data and figures are included and can be accessed at https://doi.org/10.5281/zenodo.6516238 . Parallelization when necessary to generate data was run on Palmetto cluster (372 GB, 48 nodes) and MATLAB 2020a. The code also includes Jupyter notebooks that implement the estimation functions (in python) for a 2-node system, a 3-node system, and a 3-node cell state system. These use simple csv input files where the experimental data are placed.

Acknowledgments

We would like to thank Clemson University and the CCIT team for the generous allotment of time and support in the Palmetto cluster for running the simulations in this paper. MRB acknowledges funding from Mount Sinai, Clemson University, the National Institutes of Health Grants R01GM104184 and R35GM141891 and an IBM faculty award. MB and ADS were supported by a National Institute for General Medical Sciences-funded Integrated Pharmacological Sciences Training Program grant (T32GM062754).

Competing Interests

The authors declare no competing interests.

Author Contributions

MRB, GRS, and DS conceived of the work. DS, GRS, MB, MRB, and JE performed analyses. DS, MB, and GRS made the figures. DS, GRS, MB, and MRB wrote the manuscript.

References

1. Angulo, M. T., Moreno, J. A., Lippner, G., Barabási, A.-L. & Liu, Y.-Y. Fundamental limitations of network reconstruction from temporal data. *Journal of the Royal Society, Interface* **14**, (2017).

2. Barabási, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).

3. Califano, A., Butte, A. J., Friend, S., Ideker, T. & Schadt, E. Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nature genetics* **44**, 841–7 (2012).

4. Calvano, S. E. *et al.* A network-based analysis of systemic inflammation in humans. *Nature* **437**, 1032–7 (2005).

5. Dorel, M. *et al.* Modelling signalling networks from perturbation data. *Bioinformatics* **34**, 4079–4086 (2018).

6. Hackett, S. R. *et al.* Learning causal networks using inducible transcription factors and transcriptome-wide time series. *Mol Syst Biol* **16**, (2020).

7. Hein, M. Y. *et al.* A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163**, 712–23 (2015).

8. Hill, S. M. *et al.* Context Specificity in Causal Signaling Networks Revealed by Phosphoprotein Profiling. *Cell systems* **4**, 73-83.e10 (2017).

9. Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934 (2001).

10. Ideker, T., Ozier, O., Schwikowski, B. & Siegel, A. F. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics (Oxford, England)* **18 Suppl 1**, S233-40 (2002).

11. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Observability of complex systems. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2460–5 (2013).

12. Ma'ayan, A. *et al.* Formation of regulatory patterns during signal propagation in a Mammalian cellular network. *Science (New York, N.Y.)* **309**, 1078–83 (2005).

13. Margolin, A. A. *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC bioinformatics* **7 Suppl 1**, S7 (2006).

14. Mazloom, A. R. *et al.* Recovering protein-protein and domain-domain interactions from aggregation of IP-MS proteomics of coregulator complexes. *PLoS computational biology* **7**, e1002319 (2011).

15. Mehla, J., Caufield, J. H. & Uetz, P. The yeast two-hybrid system: a tool for mapping protein-protein interactions. *Cold Spring Harbor protocols* **2015**, 425–30 (2015).

16. Molinelli, E. J. *et al.* Perturbation biology: inferring signaling networks in cellular systems. *PLoS computational biology* **9**, e1003290 (2013).

17. Nyman, E. *et al.* Perturbation biology links temporal protein changes to drug responses in a melanoma cell line. *PLOS Computational Biology* **16**, e1007909 (2020).

18. Pe'er, D., Regev, A., Elidan, G. & Friedman, N. Inferring subnetworks from perturbed expression profiles. *Bioinformatics (Oxford, England)* **17 Suppl 1**, S215-24 (2001).

19. Pósfai, M., Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Effect of correlations on network controllability. *Scientific reports* **3**, 1067 (2013).

20. Schraivogel, D. *et al.* Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nature Methods* 1–7 (2020) doi:10.1038/s41592-020-0837-5.

21. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498–504 (2003).

22. Stein, R. R., Marks, D. S. & Sander, C. Inferring Pairwise Interactions from Biological Data Using Maximum-Entropy Probability Models. *PLoS computational biology* **11**, e1004182 (2015).

23. Wynn, M. L. *et al.* Inferring Intracellular Signal Transduction Circuitry from Molecular Perturbation Experiments. *Bulletin of mathematical biology* **80**, 1310–1344 (2018).

24. Yuan, B. *et al.* CellBox: Interpretable Machine Learning for Perturbation Biology with Application to the Design of Cancer Combination Therapy. *Cell Systems* **12**, 128-140.e4 (2021).

25. Chevalier, T., Schreiber, I. & Ross, J. Toward a systematic determination of complex reaction mechanisms. *J. Phys. Chem.* **97**, 6776–6787 (1993).

26. Díaz-Sierra, R., Lozano, J. B. & Fairén, V. Deduction of Chemical Mechanisms from the Linear Response around Steady State. *J. Phys. Chem. A* **103**, 337–343 (1999).

27. Hoffmann, M., Fröhner, C. & Noé, F. Reactive SINDy: Discovering governing reactions from concentration data. *J. Chem. Phys.* **150**, 025101 (2019).

28. Kim, J., Bates, D. G., Postlethwaite, I., Heslop-Harrison, P. & Cho, K.-H. Least-squares methods for identifying biochemical regulatory networks from noisy measurements. *BMC Bioinformatics* **8**, 8 (2007).

29. Schmidt, H., Cho, K.-H. & Jacobsen, E. W. Identification of small scale biochemical networks based on general type system perturbations. *The FEBS Journal* **272**, 2141–2151 (2005).

30. Morgan, S. L. & Winship, C. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. (Cambridge University Press, 2014). doi:10.1017/CBO9781107587991.

31. Pearl, J. Structural Counterfactuals: A Brief Introduction. *Cognitive Science* **37**, 977–985 (2013).

32. Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11980–11985 (2003).

33. Reeves, G. T. The engineering principles of combining a transcriptional incoherent feedforward loop with negative feedback. *Journal of Biological Engineering* **13**, 62 (2019).

34. Fournier, T. *et al.* Steady-state expression of self-regulated genes. *Bioinformatics* **23**, 3185–3192 (2007).

35. Bell-Pedersen, D. *et al.* Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nature reviews. Genetics* **6**, 544–56 (2005).

36. Stewart-Ornstein, J., Cheng, H. W. J. & Lahav, G. Conservation and Divergence of p53 Oscillation Dynamics across Species. *Cell systems* **5**, 410-417.e4 (2017).

37. Kholodenko, B. N. *et al.* Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 12841–6 (2002).

38. Santra, T., Rukhlenko, O., Zhernovkov, V. & Kholodenko, B. N. Reconstructing static and dynamic models of signaling pathways using Modular Response Analysis. *Current Opinion in Systems Biology* **9**, 11–21 (2018).

39. Andrec, M., Kholodenko, B. N., Levy, R. M. & Sontag, E. Inference of signaling and gene regulatory networks by steady-state perturbation experiments: Structure and accuracy. *Journal of Theoretical Biology* **232**, 427–441 (2005).

40. Santos, S. D. M., Verveer, P. J. & Bastiaens, P. I. H. Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nature Cell Biology* **9**, 324–330 (2007).

41. Thomaseth, C. *et al.* Impact of measurement noise, experimental design, and estimation methods on Modular Response Analysis based network reconstruction. *Scientific Reports* **8**, 16217 (2018).

42. Gross, T. & Blüthgen, N. Identifiability and experimental design in perturbation studies. *Bioinformatics* **36**, i482–i489 (2020).

43. Halasz, M., Kholodenko, B. N., Kolch, W. & Santra, T. Integrating network reconstruction with mechanistic modeling to predict cancer therapies. *Science Signaling* **9**, (2016).

44. Klinger, B. *et al.* Network quantification of EGFR signaling unveils potential for targeted combination therapy. *Molecular Systems Biology* **9**, (2013).

45. Santra, T., Kolch, W. & Kholodenko, B. N. Integrating Bayesian variable selection with Modular Response Analysis to infer biochemical network topology. *BMC Systems Biology* **7**, (2013).

46. Sontag, E., Kiyatkin, A. & Kholodenko, B. N. Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data. *Bioinformatics (Oxford, England)* **20**, 1877–86 (2004).

47. Cho, K.-H., Choo, S.-M., Wellstead, P. & Wolkenhauer, O. A unified framework for unraveling the functional interaction structure of a biomolecular network based on stimulus-response experimental data. *FEBS Letters* **579**, 4520–4528 (2005).

48. Gupta, P. B. *et al.* Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell* **146**, 633–644 (2011).

49. Huang, C. Y. & Ferrell, J. E. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proceedings of the National Academy of Sciences* **93**, 10078–10083 (1996).

50. Höfler, M. Causal inference based on counterfactuals. *BMC Medical Research Methodology* **5**, 28 (2005).

51. Shipley, B. *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference with R*. (Cambridge University Press, 2016). doi:10.1017/CBO9781139979573.

52. Lill, D. *et al.* Mapping connections in signaling networks with ambiguous modularity. *npj Syst Biol Appl* **5**, 19 (2019).

54. Kholodenko, B. N. & Sontag, E. D. Determination of Functional Network Structure from Local Parameter Dependence Data. Preprint at http://arxiv.org/abs/physics/0205003 (2002).

55. Armond, J. W. *et al.* A stochastic model dissects cell states in biological transition processes. *Sci Rep* **4**, 3692 (2014).

56. Dirkse, A. *et al.* Stem cell-associated heterogeneity in Glioblastoma results from intrinsic tumor plasticity shaped by the microenvironment. *Nat Commun* **10**, 1787 (2019).

57. Hormoz, S. *et al.* Inferring Cell-State Transition Dynamics from Lineage Trees and Endpoint Single-Cell Measurements. *cels* **3**, 419-433.e8 (2016).

58. Larsson, I. *et al.* Modeling glioblastoma heterogeneity as a dynamic network of cell states. *Mol Syst Biol* **17**, e10105 (2021).

59. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849.e21 (2019).

60. Sha, Y., Wang, S., Zhou, P. & Nie, Q. Inference and multiscale model of epithelial-to-mesenchymal transition via single-cell transcriptomic data. *Nucleic Acids Research* **48**, 9505–9520 (2020).

61. Shen, S. & Clairambault, J. Cell plasticity in cancer cell populations. *F1000Res* **9**, F1000 Faculty Rev-635 (2020).

62. Zarkoob, H., Taube, J. H., Singh, S. K., Mani, S. A. & Kohandel, M. Investigating the Link between Molecular Subtypes of Glioblastoma, Epithelial-Mesenchymal Transition, and CD133 Cell Surface Protein. *PLoS One* **8**, e64169 (2013).

63. Henriques, D., Villaverde, A. F., Rocha, M., Saez-Rodriguez, J. & Banga, J. R. Data-driven reverse engineering of signaling pathways using ensembles of dynamic models. *PLOS Computational Biology* **13**, e1005379 (2017).

64. Reeves, G. T. The engineering principles of combining a transcriptional incoherent feedforward loop with negative feedback. *Journal of Biological Engineering* **13**, 62 (2019).

65. Goentoro, L., Shoval, O., Kirschner, M. W. & Alon, U. The Incoherent Feedforward Loop Can Provide Fold-Change Detection in Gene Regulation. *Molecular Cell* **36**, 894–899 (2009).

66. Goentoro, L. & Kirschner, M. W. Evidence that Fold-Change, and Not Absolute Level, of β-Catenin Dictates Wnt Signaling. *Molecular Cell* **36**, 872–884 (2009).

67. Nakakuki, T. *et al.* Ligand-specific c-fos expression emerges from the spatiotemporal control of ErbB network dynamics. *Cell* **141**, (2010).

68. Tyson, J. J., Csikasz-Nagy, A. & Novak, B. The dynamics of cell cycle regulation. *Bioessays* **24**, 1095–1109 (2002).

69. Albeck, J. G., Mills, G. B. & Brugge, J. S. Frequency-Modulated Pulses of ERK Activity Transmit Quantitative Proliferation Signals. *Molecular Cell* **49**, 249–261 (2013).

70. Kholodenko, B. N. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur J Biochem* **267**, 1583–1588 (2000).

71. Ryu, H. *et al.* Frequency modulation of ERK activation dynamics rewires cell fate. *Mol Syst Biol* **11**, 838 (2015).

72. Fuente, A. de la, Brazhnik, P. & Mendes, P. Linking the genes: inferring quantitative gene networks from microarray data. *Trends in Genetics* **18**, 395–398 (2002).

73. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nature Reviews Genetics* **20**, 631–656 (2019).

74. Spitzer, M. H. & Nolan, G. P. Mass Cytometry: Single Cells, Many Features. *Cell* **165**, 780–791 (2016).

76. Aksenov, A. A., da Silva, R., Knight, R., Lopes, N. P. & Dorrestein, P. C. Global chemical analysis of biology by mass spectrometry. *Nature Reviews Chemistry* **1**, 1–20 (2017).

77. Akbani, R. *et al.* Realizing the Promise of Reverse Phase Protein Arrays for Clinical, Translational, and Basic Research: A Workshop Report: The RPPA (Reverse Phase Protein Array) Society. *Molecular & Cellular Proteomics* **13**, 1625–1643 (2014).

78. Stapor, P. *et al.* Mini-batch optimization enables training of ODE models on large-scale datasets. *Nat Commun* **13**, 34 (2022).

79. Wu, G., Feng, X. & Stein, L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol* **11**, R53 (2010).

80. Raue, A. *et al.* Lessons Learned from Quantitative Dynamical Modeling in Systems Biology. *PLOS ONE* **8**, e74335 (2013).

Figure Legends

Figure 1. **Overall DL-MRA Approach. (a)** Two-node network with Jacobian elements labeled. Green arrows are stimuli and basal production terms. **(b,d,f)** Time course experimental design with perturbations: vehicle (b), Node 1 (d), Node 2 (f). The vehicle may be the solvent like DMSO for inhibition with a drug, or a nontargeting si/shRNA for inhibition with si/shRNA. **(c,e,g)** Simulated time course data for Vehicle perturbation (c), Node 1 perturbation (e), Node 2 perturbation (g) from the network in (a). Left Column: no added noise; Right Column 10:1 signal-to-noise added. **(h-j)** Actual versus inferred model parameters *($S_{1,b}$, $S_{1,ex}$, F_{11}, F_{12} , $S_{2,b}$, $S_{2,ex}$, F_{21}, F_{22})* for direct solution of Eq. 3-4 in the absence (h) or presence (i) of noise, or with noise and the least-squares approach (j). In h-i, error bars are standard deviation across time points.

Figure 2. **Application to Linear Two and Three Node Models. (a)** Connections around a Node *i* in an n-Node Model. $S_{i,b}$ and $S_{i,ex}$ are the basal production and external stimulus terms acting on Node *i*, respectively. F_{ii} is the self-regulation term; *Fij* the effect of Node *j* on Node *i* and F_{ji} the effect of Node *i* on Node *j*. **(b)** Example of different signal-to-noise ratio effects on time course data. **(c,d)** Ground truth versus estimated edge weights across all 50 random networks and noise levels for data from four different total timepoints (3,7,11,21) for 2 node (c) and 3 node (d) networks. Quadrant shading indicates edge classification. **(e,f)** Fraction of network parameters correctly classified in 50 randomly generated 2 node networks (e) and 3 node networks (f) with different noise levels and total timepoints. (g) Fraction of network parameters correctly classified in 50 randomly generated 3 node networks with dynamic MRA using two sets of perturbation data.

Figure 3. **Application to Cell State Transition Networks. (a**) Markov transition model of SUM159 cell states. **(b**) Cell proportions over time for SUM159 cells using Markov transition parameters (dots), starting at different initial proportions and respective DL-MRA model fits (lines). **(c)** Parameters from DL-MRA estimates of SUM159 data are similarly classified as transformed Markov parameters (See Methods, Eq. 29-30). **(d)** Ground truth versus estimated edge weights across 50 random cell transition networks and noise levels for data from four different total timepoints (3,7,11,21).

Figure 4. Application to a Signaling Network. (a) Full Reaction scheme for the Huang-Ferrell (HF) Model, depicting the parameters k3, k15 and k27 which were perturbed sequentially to generate the perturbation data. **(b)** Model coarse-graining to a 3-node network. **(c)** Data generated for each node with a small E1 stimulus (2.5×10^{-6} uM) . **(d)** Model parameters estimated as significant (bold) and negligible (dotted lines). **(e)** SELDOM true graph values represented in the 3-node model with parameters considered (bold) and not considered (dotted lines).

Figure 5. **(a)** Feedforward loop (FFL) network models. Across all 16 models (Table 1), F_{11}, F_{22}, and F_{33} values are fixed at -1 and F_{12}, F_{13}, and F_{23} values are fixed at 0. F_{21}, F_{31}, and F_{32} values can be positive or negative depending on the model. The combined effect of x_1 and x_2 on x_3 is described by either an AND gate or an OR gate. There are 16 possible

model structures (Table 1). **(b)** 100% inhibitory perturbations may not provide accurate classification even without noise. In Model #1, F_{31} is positive (ground truth) but is estimated as null. **(c)** Specific structure of Model #1. **(d)** Node activity simulation data for 100% inhibition in Model #1, implying that it is impossible to infer F_{31} from such data. **(e)** Node activity simulation data for 50% inhibition in Model #1, showing potential to infer F_{31}. **(f)** Fraction of model parameters correctly classified in all the 16 non-linear models without noise, for 100% inhibition vs 50% inhibition.

Figure 6. (a) Classification scheme for a distribution of parameter estimates. Going from left to right panels, the same parameter distribution with an actual (ground truth) value of positive (+), negative (-), or null (0), respectively, is estimated using different percentile windows centered on the median. The percentile "window" is the median value for the leftmost panel (rigorous classification), between 40th and 60^{th} percentile in the second panel, and between 10^{th} and 90^{th} percentile in the third panel (conservative classification). Going from rigorous to conservative (left to right), an intermediate between the two gives a good classification performance. **(b)** ROC curves across all parameters for all 16 FFL models. Different color lines are different noise levels. **(c)** Fraction of correctly classified model parameters for different noise levels broken down by FFL model type. **(d)** Fraction of each model parameter correctly classified for different noise levels broken down by parameter type.

FIGURE 2.1

Sarmah et al., 2022. Figure 1

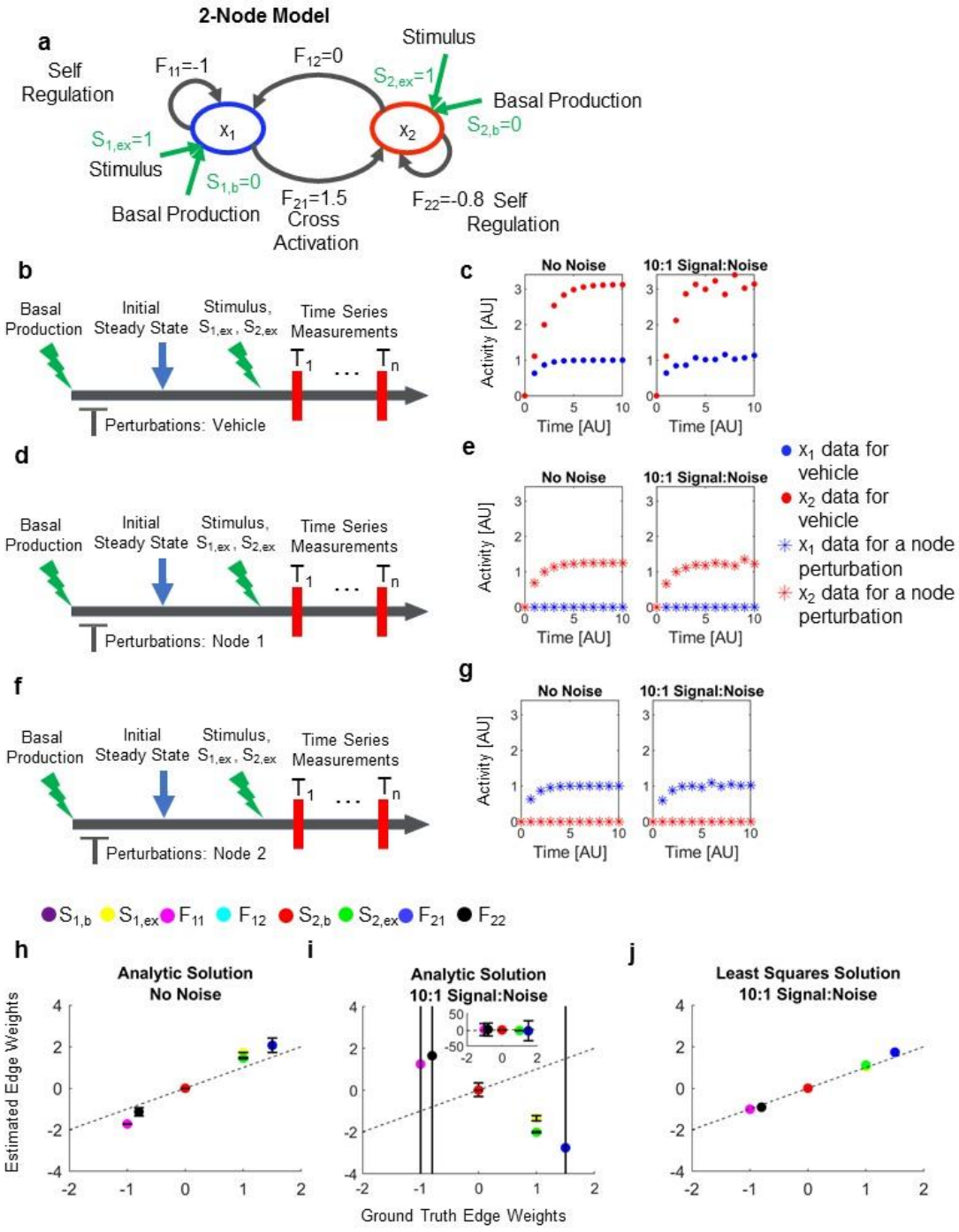


FIGURE 2.2

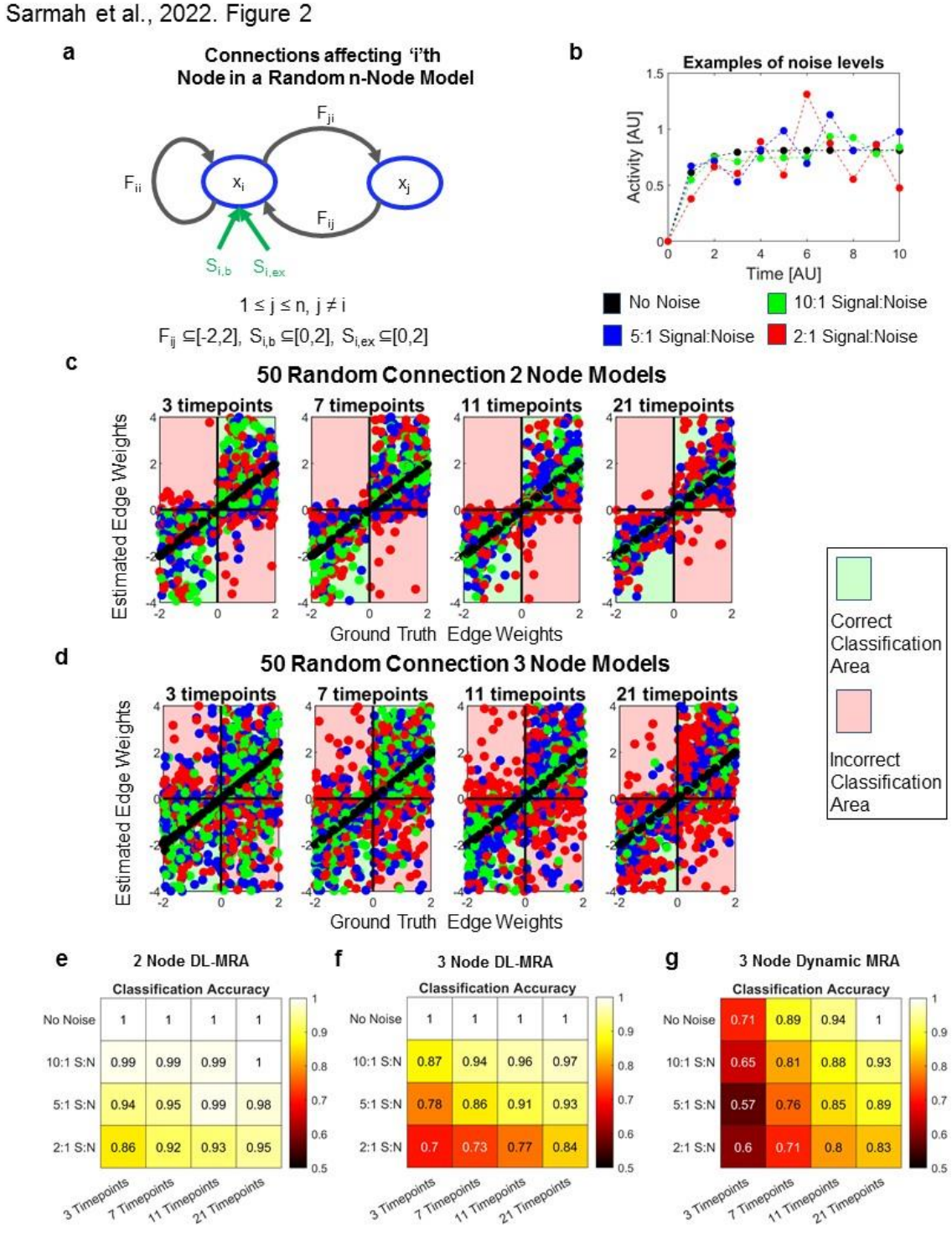

Sarmah et al., 2022. Figure 2
a
Connections affecting 'i'th Node in a Random n-Node Model
F_{ji}
F_{ii}
x_i
x_j
F_{ij}
$S_{i,b}$
$S_{i,ex}$
$1 \le j \le n,\ j \ne i$
$F_{ij} \subseteq [-2,2],\ S_{i,b} \subseteq [0,2],\ S_{i,ex} \subseteq [0,2]$
b
Examples of noise levels
Activity [AU]
Time [AU]
No Noise
10:1 Signal:Noise
5:1 Signal:Noise
2:1 Signal:Noise
c
50 Random Connection 2 Node Models
3 timepoints
7 timepoints
11 timepoints
21 timepoints
Estimated Edge Weights
Ground Truth Edge Weights
d
50 Random Connection 3 Node Models
Correct Classification Area
Incorrect Classification Area
e
2 Node DL-MRA
Classification Accuracy
No Noise 1 1 1 1
10:1 S:N 0.99 0.99 0.99 1
5:1 S:N 0.94 0.95 0.99 0.98
2:1 S:N 0.86 0.92 0.93 0.95
3 Timepoints
7 Timepoints
11 Timepoints
21 Timepoints
f
3 Node DL-MRA
Classification Accuracy
No Noise 1 1 1 1
10:1 S:N 0.87 0.94 0.96 0.97
5:1 S:N 0.78 0.86 0.91 0.93
2:1 S:N 0.7 0.73 0.77 0.84
g
3 Node Dynamic MRA
Classification Accuracy
No Noise 0.71 0.89 0.94 1
10:1 S:N 0.65 0.81 0.88 0.93
5:1 S:N 0.57 0.76 0.85 0.89
2:1 S:N 0.6 0.71 0.8 0.83

FIGURE 2.3

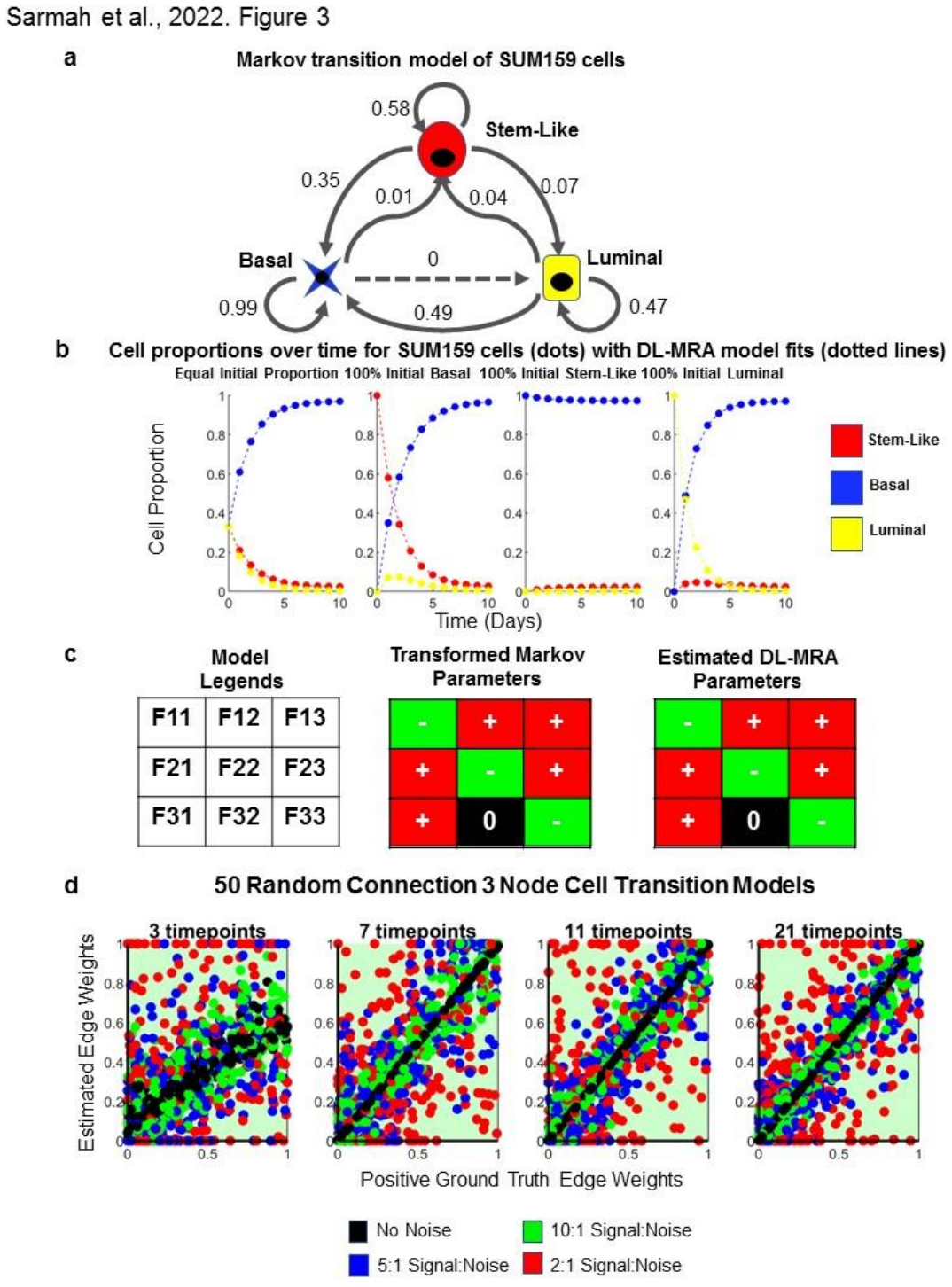

Sarmah et al., 2022. Figure 3
a
Markov transition model of SUM159 cells
0.58
Stem-Like
0.35
0.01
0.04
0.07
Basal
0
Luminal
0.99
0.49
0.47
b
Cell proportions over time for SUM159 cells (dots) with DL-MRA model fits (dotted lines)
Equal Initial Proportion
100% Initial Basal
100% Initial Stem-Like
100% Initial Luminal
Cell Proportion
Time (Days)
Stem-Like
Basal
Luminal
c
Model Legends
F11 F12 F13
F21 F22 F23
F31 F32 F33
Transformed Markov Parameters
Estimated DL-MRA Parameters
d
50 Random Connection 3 Node Cell Transition Models
3 timepoints
7 timepoints
11 timepoints
21 timepoints
Estimated Edge Weights
Positive Ground Truth Edge Weights
No Noise
10:1 Signal:Noise
5:1 Signal:Noise
2:1 Signal:Noise

FIGURE 2.4

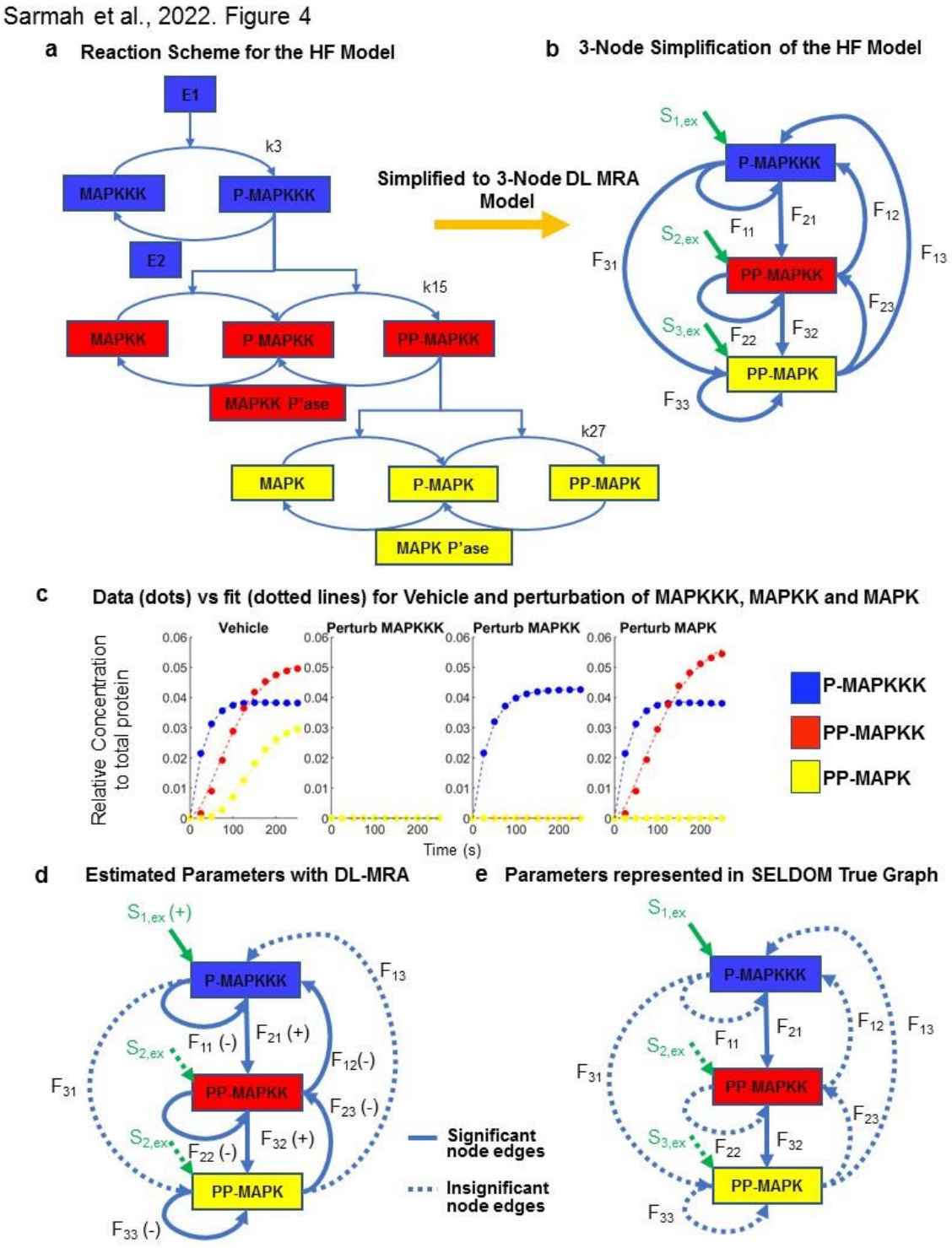

Sarmah et al., 2022. Figure 4
a Reaction Scheme for the HF Model
E1
k3
MAPKKK
P-MAPKKK
E2
k15
MAPKK
P-MAPKK
PP-MAPKK
MAPKK P'ase
k27
MAPK
P-MAPK
PP-MAPK
MAPK P'ase
Simplified to 3-Node DL MRA Model
b 3-Node Simplification of the HF Model
S1,ex
P-MAPKKK
F11
F21
F12
F13
S2,ex
F31
PP-MAPKK
F23
S3,ex
F22
F32
PP-MAPK
F33
c Data (dots) vs fit (dotted lines) for Vehicle and perturbation of MAPKKK, MAPKK and MAPK
Vehicle
Perturb MAPKKK
Perturb MAPKK
Perturb MAPK
Relative Concentration to total protein
Time (s)
P-MAPKKK
PP-MAPKK
PP-MAPK
d Estimated Parameters with DL-MRA
S1,ex (+)
F11 (-)
F21 (+)
F12(-)
F23 (-)
F22 (-)
F32 (+)
F33 (-)
Significant node edges
Insignificant node edges
e Parameters represented in SELDOM True Graph

TABLE 2.1

Sarmah et al., 2022. Table 1

Model#	F_{21}	Gate	F_{31}	F_{32}
1	Activator	AND	Activator	Activator
2	Activator	AND	Activator	Repressor
3	Activator	AND	Repressor	Activator
4	Activator	AND	Repressor	Repressor
5	Activator	OR	Activator	Activator
6	Activator	OR	Activator	Repressor
7	Activator	OR	Repressor	Activator
8	Activator	OR	Repressor	Repressor
9	Repressor	AND	Activator	Activator
10	Repressor	AND	Activator	Repressor
11	Repressor	AND	Repressor	Activator
12	Repressor	AND	Repressor	Repressor
13	Repressor	OR	Activator	Activator
14	Repressor	OR	Activator	Repressor
15	Repressor	OR	Repressor	Activator
16	Repressor	OR	Repressor	Repressor

Table 1. Structure of each of the 16 non-linear models.

FIGURE 2.5

Sarmah et al., 2022. Figure 5

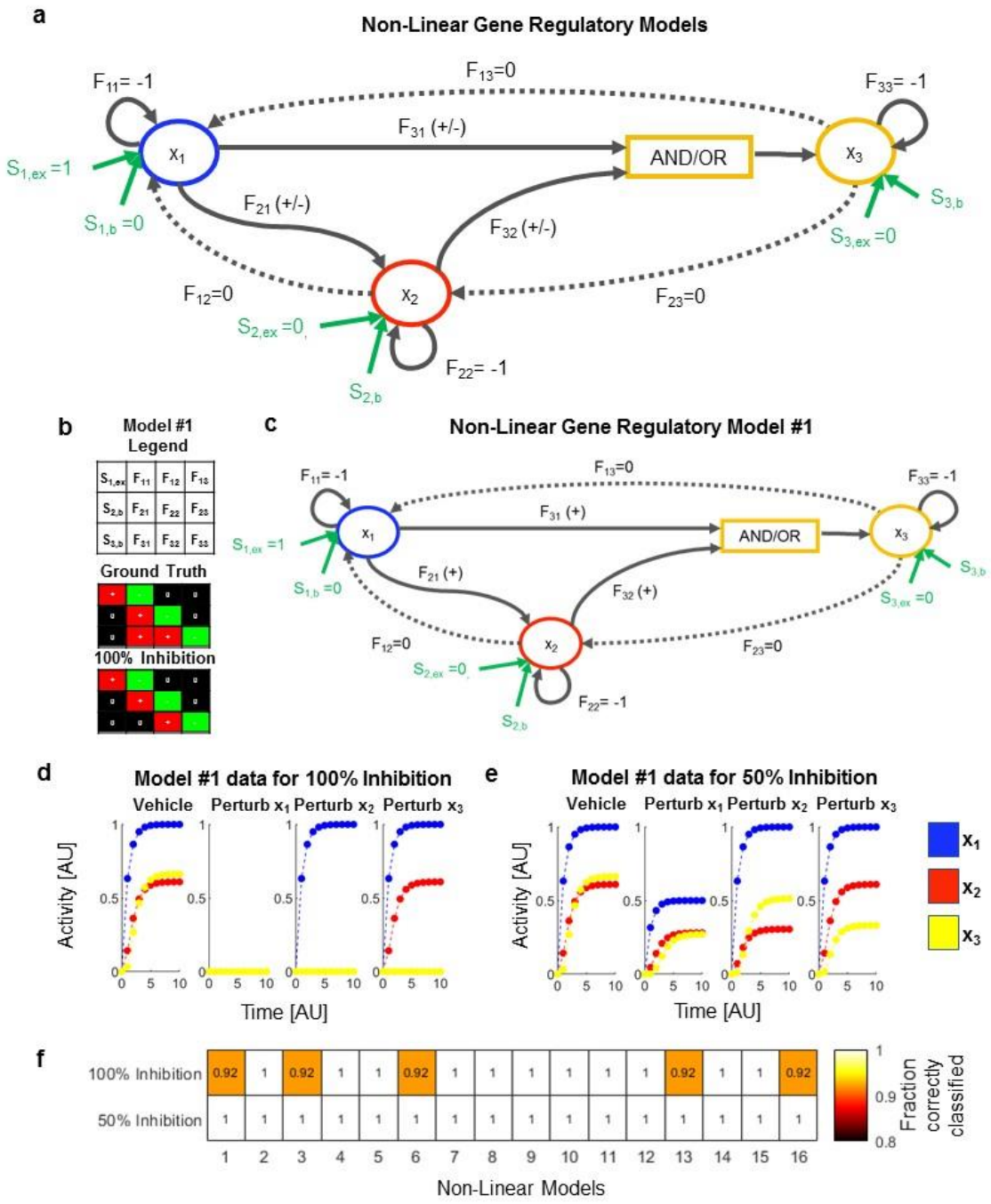
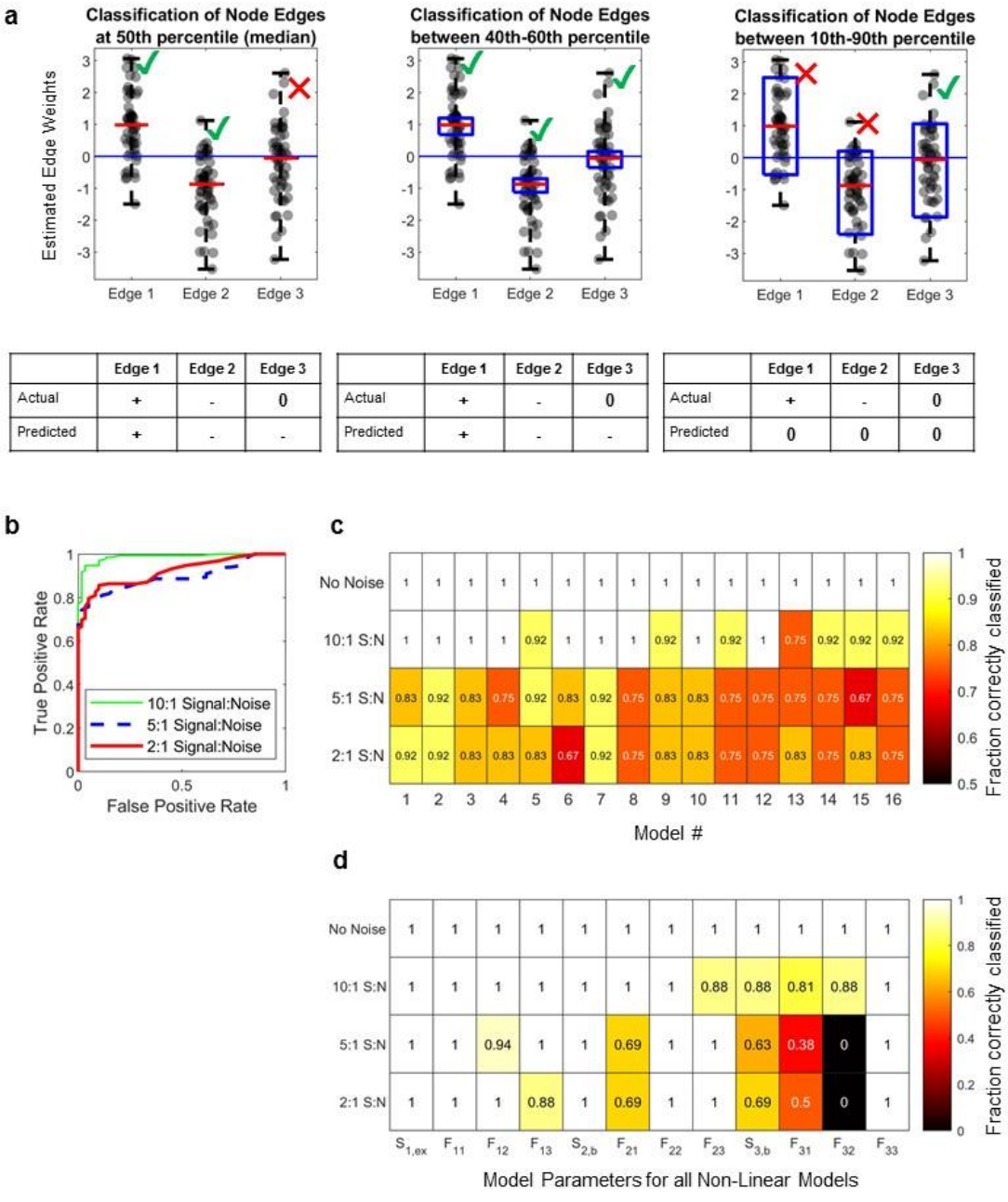


FIGURE 2.6

Sarmah et al., 2022. Figure 6



CHAPTER THREE

RANKING DRUG TARGETS FOR GLIOBLASTOMA

Overview

Targeting glioblastoma using small molecule kinase inhibitors requires these drugs to be able to cross the blood brain barrier. Hence, for subsequent studies in this chapter and experiments in the following two chapters, we chose to focus our attention to drugs which are known to have significant blood brain barrier penetrance. We made a list of 27 drugs (FDA approved or undergoing trials) which were reported to have blood brain barrier penetrance[1]. These are listed in Table 3.1 below. The blood brain barrier penetrance was either indicated by their free brain or cerebro-spinal-fluid (CSF) ratio to free plasma concentrations (>0.3)[1] or reported elsewhere as high blood brain barrier penetrant for a minority of the drugs[2–5].

The number of kinases which are parts of signaling networks involved in cancer and glioblastoma is numerous. Often, for making meaningful predictions about drug effects, which is our eventual goal (Chapter 5) it is reasonable to trim the list of drug targets to the most relevant proteins in cancer[6]. We chose to follow this exercise for the kinases involved in glioblastoma, with an added objective to serve as a starting point for students in the future interested in ranking relevant proteins or genes with respect to cancer or other diseases.

First considering a pharmacological viewpoint, we reasoned that listing the top targets of the blood brain barrier penetrant drugs would help us begin to rank the most important kinases to target in glioblastoma. These can help rank the kinases in terms of

their average affinity with the drugs (Criteria-drug affinity) as well as help us map the recurring kinase targets for all the drugs (Criteria-drug recurrence). These two pharmacological criteria were considered by listing the kinase targets of the drugs (see Methods).

However only considering the pharmacological targets does not give us a complete picture of whether the drugs will be effective in the biology of glioblastoma. Therefore, we considered two more categories focused on biological relevance. First, overexpression of a kinase can be an indicator of its role as a driver of cancer. Secondly, the number of scientific papers and clinical trials relating a protein to a certain disease can be an important indicator of its importance in treating the disease. Therefore, besides the previous two pharmacological criteria, we also considered these two biological criteria (Expression fold change and relevant studies, see Methods). We weighed these four criteria equally and the final ranking is based on the sum of all four criteria (see Methods).

Results and Discussion

The top 10 ranked kinases in each category are mentioned in Table 3.3 below.

The top five ranked kinases for the criteria “Drug Affinity” (MET, DDR1, TRKA, ROS, EPHA3) were all strong targets of the drug cabozantinib and ROS additionally was the principal target of Lorlatinib, besides them being potent off targets of other drugs. In general, they may be considered as strongly inhibited by their relevant drugs.

In the list of ranked kinases for the criteria “Drug Recurrence” two SRC family kinases-LCK and SRC were in the top five and three more- FYN, YES and HCK were

among the top ten. This indicates that SRC family kinases are highly likely to be off targets of multiple kinase inhibitors.

The top five most studied kinases with relation to glioblastoma (see Methods) were EGFR, KDR, MET, ROS and SRC. EGFR and KDR (VEGFR) have been also subject to clinical trials for glioblastoma although unsuccessful so far[7,8]. In the top ten list of overexpressed kinases, nine (except EGFR) were absent from the list of top ten most studied kinases with relation to glioblastoma and most (except AURKA and IGFR) had fewer than 15 citations (see Methods). There might be potential for future studies relating these overexpressed kinases to glioblastoma, if overexpression is an indicator of a kinase being an important driver of cancer.

The top five overall ranked kinases were EGFR, KDR, YES, ERBB2 and LCK. It is not known at this point if inhibiting these kinases would result in significant outcomes in glioblastoma cells, (although targeted trials on glioblastoma patients have been unsuccessful) and we explore the effects of the kinase inhibitors on glioblastoma cells in the next chapter. As mentioned before, an objective of this ranking exercise is to serve as a starting point for students in the future interested in ranking relevant proteins or genes with respect to cancer or other diseases. At the same time, it is worth noting that such lists can only indicate a ranked list of kinases-this may not always translate to these kinases being potent drug targets with respect to inhibition of cancer cell proliferation. Further data on the effect of drugs or other kinase inhibitors (for eg. siRNA, shRNA, CRISPR) on the growth of cancer cells may be needed before conclusively rating these kinases as relevant drug/therapy targets for cancer or other diseases.

Methods-

Ranking Methodology

The kinase targets for the mentioned drugs were obtained from kinomescan assay data at LINCS[9], proteomicsdb[10,11] and other papers focusing on kinome wide assays for specific drugs[12–15]. These kinase targets were listed by their disassociation constant (K_d) against the drugs. In cases where K_d was not reported and a percentage inhibition at a specific dose of the drug was available instead, the K_d was be measured by fitting, under the assumption that the drug and the kinase targets follow Michaelis-Mentan kinetics. The kinases were then ranked in terms of their affinity with the drug, with the maximum K_d taken as 2000 nM (Table 3.2). The average of these values across all drugs, for a particular kinase was taken as its average "drug affinity" score and the number of drugs it showed up for is taken as its "drug recurrence" and was assigned a score accordingly (Table 3.2).

In order to rank proteins by the expression fold change criteria, the we used the web application called Gene Expression Profiling Interactive Analysis (GEPIA)[16]. This enables us to compare the normalized (Transcripts per million) expression data of a gene in a particular cancer as found in TCGA (The Cancer Genome Atlas)[17] to the expression data of the same gene in normal gene tissue as found in GTEx (The Genotype-Tissue Expression)[18]. In order to compare the expression our list of drug targets, we used the "Expression DIY" module followed by "Multiple Gene Comparison" module in GEPIA. The "Log Scale" was turned off, the "Dataset" chosen was GBM and "Matched Normal Data" was set to "Match TCGA normal and GTEx data". The relevant drug targets were input in the "Gene List" box and the expression data of the gene in both TCGA and GTEx

was noted. These values were compiled their relative fold change is noted (log of (TCGA/GTEx) expression data with base 2). Ranking criteria (1-20) was set for all the expression fold change scores and the drug targets were assigned a score as described in Table 3.2.

In order to rank the kinases by relevant studies, we listed the number of results on the search term "Glioblastoma xx" in Pubmed where xx is the name of the gene. Ranking criteria (1-20) was set for all the expression fold change scores and the drug targets were assigned a ranking score as described in Table 3.2. This gives us a list of potential kinase drug targets ranked.

TABLE 3.1

LIST OF BLOOD BRAIN BARRIER PENETRANT DRUGS

Abemiciclib	Alectinib	ARQ 087
ASP3026	AZD3759	Bosutinib
Brigatinib	Brivanib	Buparlisib
Cabozantinib	E6201	Entrectinib
GDC0084	GNE-317	Ibrutinib
Lorlatinib	NT113	Palomid
PBI-05204	PD0325901	Pimasertib
PQR309	Saracatinib	TAK-285
Tak-960	Tucatinib	URMC-099

TABLE 3.2

SCORING CRITERIA FOR EACH CATEGORY

Drug Affinity (K_d in nM)	Score	Drug Recurrence	Score	Expression Fold Change	Score	Relevant Studies	Score
<0.1	1	>8	1	>4	1	>2000	1
0.1 to 0.3	2	8	3.4	3.5 to 4	2	1500 to 2000	2
0.3 to 0.5	3	7	5.8	3 to 3.5	3	900 to 1500	3
0.5 to 1	4	6	8.2	2.5 to 3	4	600 to 900	4
1 to 3	5	5	10.5	2 to 2.5	5	300 to 600	5
3 to 5	6	4	12.9	1.8 to 2	6	200 to 300	6
5 to 10	7	3	15.3	1.6 to 1.8	7	150 to 200	7
10 to 20	8	2	17.7	1.4 to 1.6	8	100 to 150	8
20 to 40	9	1	20	1.2 to 1.4	9	80 to 100	9
40 to 65	10			1 to 1.2	10	65 to 80	10
65 to 100	11			0.8 to 1	11	50 to 65	11
100 to 150	12			0.6 to 0.8	12	40 to 50	12
150 to 250	13			0.45 to 6	13	30 to 40	13
250 to 400	14			0.3 to 0.45	14	20 to 30	14
400 to 600	15			0.15 to 0.3	15	15 to 20	15
600 to 800	16			0 to 0.15	16	10 to 15	16
800 to 1000	17			-0.1 to 0	17	5 to 10	17
1000 to 1300	18			-0.25 to -0.1	18	3,4	18
1300 to 1600	19			-0.5 to -0.25	19	1,2	19
1600 to 2000	20			< -0.5	20	0	20

TABLE 3.3

TOP 10 RANKED KINASES IN EACH CATEGORY

Drug Affinity	Drug Recurrence	Expression Fold Change	Relevant Studies	Overall Rank
MET	LCK	AURKB	EGFR	EGFR
DDR1	SRC	EGFR	KDR	KDR
TRKA	ALK	HCK	MET	YES
ROS	ABL1	Syk	ROS	ERBB2
EPHA3	FLT3	BLK	SRC	LCK
SRMS	EGFR	AURKA	KIT	SRC
TRKC	YES	EPHB4	ERBB2	ABL1
ERBB3	HCK	IGF1R	PDGFRB	IGF1R
HIPK4	MAP2K1	DDR1	FAK	HCK
RIPK2	FYN	LCK	PDGFRA	PDGFRA

References-

1. Heffron, T.P. (2016). Small Molecule Kinase Inhibitors for the Treatment of Brain Cancer. J. Med. Chem. *59*, 10030–10066. 10.1021/acs.jmedchem.6b00618.

2. Goodfellow, V.S., Loweth, C.J., Ravula, S.B., Wiemann, T., Nguyen, T., Xu, Y., Todd, D.E., Sheppard, D., Pollack, S., Polesskaya, O., et al. (2013). Discovery, Synthesis, and Characterization of an Orally Bioavailable, Brain Penetrant Inhibitor of Mixed Lineage Kinase 3. J. Med. Chem. *56*, 8032–8048. 10.1021/jm401094t.

3. Fischer, H., Ullah, M., de la Cruz, C.C., Hunsaker, T., Senn, C., Wirz, T., Wagner, B., Draganov, D., Vazvaei, F., Donzelli, M., et al. (2020). Entrectinib, a TRK/ROS1 inhibitor with anti-CNS tumor activity: differentiation from other inhibitors in its class due to weak interaction with P-glycoprotein. Neuro Oncol *22*, 819–829. 10.1093/neuonc/noaa052.

4. Naito, T., Shiraishi, H., and Fujiwara, Y. (2021). Brigatinib and lorlatinib: their effect on ALK inhibitors in NSCLC focusing on resistant mutations and central nervous system metastases. Japanese Journal of Clinical Oncology *51*, 37–44. 10.1093/jjco/hyaa192.

5. Koul, D., Wang, S., Wu, S., Saito, N., Zheng, S., Gao, F., Kaul, I., Setoguchi, M., Nakayama, K., Koyama, K., et al. (2017). Preclinical therapeutic efficacy of a novel blood-brain barrier-penetrant dual PI3K/mTOR inhibitor with preferential response in PI3K/PTEN mutant glioma. Oncotarget *8*, 21741–21753. 10.18632/oncotarget.15566.

6. Menden, M.P., Wang, D., Mason, M.J., Szalai, B., Bulusu, K.C., Guan, Y., Yu, T., Kang, J., Jeon, M., Wolfinger, R., et al. (2019). Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. Nat Commun *10*, 2674. 10.1038/s41467-019-09799-2.

7. Reardon, D.A., Turner, S., Peters, K.B., Desjardins, A., Gururangan, S., Sampson, J.H., McLendon, R.E., Herndon, J.E., Jones, L.W., Kirkpatrick, J.P., et al. (2011). A review of VEGF/VEGFR-targeted therapeutics for recurrent glioblastoma. J Natl Compr Canc Netw *9*, 414–427. 10.6004/jnccn.2011.0038.

8. Reardon, D.A., Brandes, A.A., Omuro, A., Mulholland, P., Lim, M., Wick, A., Baehring, J., Ahluwalia, M.S., Roth, P., Bähr, O., et al. (2020). Effect of Nivolumab vs Bevacizumab in Patients With Recurrent Glioblastoma: The CheckMate 143 Phase 3 Randomized Clinical Trial. JAMA Oncology *6*, 1003–1010. 10.1001/jamaoncol.2020.1024.

9. Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017). A Next Generation

Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell *171*, 1437-1452.e17. 10.1016/j.cell.2017.10.049.

10. Samaras, P., Schmidt, T., Frejno, M., Gessulat, S., Reinecke, M., Jarzab, A., Zecha, J., Mergner, J., Giansanti, P., Ehrlich, H.-C., et al. (2020). ProteomicsDB: a multi-omics and multi-organism resource for life science research. Nucleic Acids Research *48*, D1153–D1163. 10.1093/nar/gkz974.

11. Schmidt, T., Samaras, P., Frejno, M., Gessulat, S., Barnert, M., Kienegger, H., Krcmar, H., Schlegl, J., Ehrlich, H.-C., Aiche, S., et al. (2018). ProteomicsDB. Nucleic Acids Research *46*, D1271–D1281. 10.1093/nar/gkx1029.

12. Zeng, Q., Wang, J., Cheng, Z., Chen, K., Johnström, P., Varnäs, K., Li, D.Y., Yang, Z.F., and Zhang, X. (2015). Discovery and Evaluation of Clinical Candidate AZD3759, a Potent, Oral Active, Central Nervous System-Penetrant, Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor. J. Med. Chem. *58*, 8200–8215. 10.1021/acs.jmedchem.5b01073.

13. Heffron, T.P., Ndubaku, C.O., Salphati, L., Alicke, B., Cheong, J., Drobnick, J., Edgar, K., Gould, S.E., Lee, L.B., Lesnick, J.D., et al. (2016). Discovery of Clinical Development Candidate GDC-0084, a Brain Penetrant Inhibitor of PI3K and mTOR. ACS Med. Chem. Lett. *7*, 351–356. 10.1021/acsmedchemlett.6b00005.

14. Uitdehaag, J.C.M., Roos, J.A.D.M. de, Doornmalen, A.M. van, Prinsen, M.B.W., Man, J. de, Tanizawa, Y., Kawase, Y., Yoshino, K., Buijsman, R.C., and Zaman, G.J.R. (2014). Comparison of the Cancer Gene Targeting and Biochemical Selectivities of All Targeted Kinase Inhibitors Approved for Clinical Use. PLOS ONE *9*, e92146. 10.1371/journal.pone.0092146.

15. Ishikawa, T., Seto, M., Banno, H., Kawakita, Y., Oorui, M., Taniguchi, T., Ohta, Y., Tamura, T., Nakayama, A., Miki, H., et al. (2011). Design and Synthesis of Novel Human Epidermal Growth Factor Receptor 2 (HER2)/Epidermal Growth Factor Receptor (EGFR) Dual Inhibitors Bearing a Pyrrolo[3,2-d]pyrimidine Scaffold. J. Med. Chem. *54*, 8030–8050. 10.1021/jm2008634.

16. Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res *45*, W98–W102. 10.1093/nar/gkx247.

17. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet *45*, 1113–1120. 10.1038/ng.2764.

18. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The Genotype-Tissue Expression (GTEx) project. Nat Genet *45*, 580–585. 10.1038/ng.2653.

CHAPTER FOUR

DRUG DOSE RESPONSES FOR GLIOBLASTOMA CELL LINES

Introduction

Although targeted inhibition of cancer biomarkers has been a cornerstone in cancer, patients often don't respond to such treatment. Combination therapy is generally associated with better outcomes in cancer and or combination regimens of 3-4 drugs have become the standard in cancer treatment[1]. But given the large number of potential drugs (FDA approved or in phase trials), all possible combinations are hard to test experimentally because of the sheer number of such experiments. Specific to glioblastoma, even if we only consider the 27 blood brain barrier penetrant drugs we discussed in chapter 3, the total number of 2 way combinations of these drugs is 351 and 3 way combinations is 2925. Considering that these combinations may have to be tested across multiple cell lines, an attempt to test all of them would be a Herculean task and there is a need for smarter interventions to predict combination therapies.

However, it is relatively easier to perform single drug dose responses. Single drug dose responses, along with information about biological networks (explored in chapter 2 and further implemented in chapter 5) may potentially help us to predict biological outcomes for combination therapy which then may be verified with experiments. Moreover, single drug dose responses, by themselves can provide insight into which kinase targets, when inhibited have the most effect on cellular outcome -thus helping focus our attention on the most promising drug targets, which would effective in drug combinations.

Therefore, we sought to perform experiments to obtain single drug dose responses with blood brain barrier penetrant drugs. Out of 27 blood brain barrier penetrant drugs discussed in chapter 3, 22 drugs were bought, either from Selleckchem (most drugs) or Tocris (TAK-960). The 5 that were not bought were either unavailable (E6201, NT113, PBI-05204) at that time or cost prohibitive and redundant in their principal targets (ARQ 087, PQR309). These were tested on three cell lines representing glioblastoma (U87, U251 and LN229-EGFRvIII) between the doses 1 nM and 10 μM, which is a usual range of cancer drugs when used in the human body. LN229 was also procured but we were unable to thaw live cells from the frozen vials.

Results

The drug dose responses experiments were performed, and the data obtained were fit into a sigmoidal hill-type equation (see Methods). The compiled results for 20 drugs (2 drugs were not able to be quantified due to reasons explained below) are shown in a single panel (Figure 4.1). The parameters of these fits for each drug and cell line and their respective IC_{50} doses (drug doses where the cell number was inhibited by 50% with respect to control) are further shown in Tables 4.1 and 4.2. By comparing the average IC_{50} of the drugs across the three cell lines, we can begin to analyze the potency of each drug and their respective targets. Principal targets and off targets (although not an exhaustive list) of the respective drugs are also tabulated in Table 4.3 for ease of further analysis.

First, we observe that 8 out of the 20 drugs are have minimal inhibitive effect on cell numbers and the inhibition (*Y*, see methods) is less than 50% even at the maximal dose

of drugs. We can call these drugs non-responsive and they include the drugs AZD3759, Brivanib, GNE-317, Ibrutinib, Lorlatinib, Palomid, TAK-285, Tucatinib.

Secondly, there were 6 drugs where the inhibition was effective at relatively high doses (1 μM-10 uM) but not at low doses. These may be called mildly responsive drugs and include Alectinib, ASP3026, Brigatinib, Buparlisib, Entrectinib, GDC0084. For these drugs, their Hill coefficient, maximal inhibition Y_{max}, and EC50 (halfway point of the curve) and calculated IC_{50} are shown in Table 4.4.

Lastly, for 6 more drugs, the inhibition was effective even at lower doses (1 nM-1 μM). These may be called highly responsive drugs and include Abemaciclib, Bosutinib, PD0325901, Pimasertib, TAK-960 and URMC-099. TAK-960 was the most potent drug with IC_{50} between 3-11 nM among the three cell lines. The Hill coefficient, maximal inhibition Y_{max}, and EC_{50} (halfway point of the curve) and calculated IC_{50} are shown in Table 4.3.

Based on the drug dose responses, the most potent responses were to the inhibition of the following kinases- PLK1, CDK4/6, MEK1/2 and ABL1. Intermediate responses were to the inhibition of ALK, TRKA and potentially the SRC family of kinases. Inhibition of EGFR, ERBB2 and KDR(VEGFR) the most important kinases according to the ranked list in Chapter 3 had no noticeable effects.

For two more drugs for which the single dose response experiments were performed, we could not create any dose response curves. First, the drug cabozantinib was relatively insoluble at higher doses and the drug aggregates showed up in the final assay. Including the cell counts at these higher doses of cabozantinib would not have been prudent

(as incomplete dissolution implies incorrect final concentration) nor feasible (as the drug aggregates can have their own fluorescence and upset the actual cell counts). Secondly, for higher doses of saracatinib (1 µM-10µM), we were unable to use our cell counting software Cell Profiler (see Methods) to accurately count distinct cells as at these doses, cells tended to cluster closely together. This may have implications in cell migration but further investigations into the roles of saractinib in cell migration was out of scope for our current study which was limited to obtaining single drug dose responses.

Discussion-

What may cause cells to be more responsive to the increase of drugs dose? There are possible ways this may occur. First, targeted kinases which may have important roles in cell proliferation, are increasingly inhibited at higher doses. Secondly a larger number of off targets begin to be inhibited at higher doses of drugs which may include kinases with key roles in cell proliferation. These two causes are not exclusive to one another and both of them are likely to happen at the same time for most drugs, although the first factor is likely to be more important in highly selective drugs and the second factor in drugs with more promiscuity towards their targets. Three of the highly responsive drugs- TAK960, PD0325901 and Pimasertib are highly selective in their targets and three are more promiscuous-Abemaciclib, Bosutinib and URMC-099. Also given that many drugs are non-responsive even at the highest dose (10µM), which may inhibit many kinases for multiple drugs we can observe that cell proliferation may generally be unaffected by the inhibition of many kinases. This is assuming that drugs inhibit kinases within cells at a

similar rate as they do in kinome-wide assays (chapter 3), which may be a fair assumption in absence of more information.

The case of cells being clustered at higher doses of saracatinib could indicate that the cells division may be ongoing, but their migration is limited by the drug. This role of saracatinib in inhibiting cell migration has been reported in prior literature[3,4]. This could be important in controlling cell migration in aggressive cancers in future research.

Methods-

Experiments- Cell Culture

U87, U251 and LN229-EGFRvIII cells were cultured in full growth medium comprising DMEM (Gibco #10313039) supplemented with 10% FBS (Corning #35-011-CV) and 2 mM L-Glutamine (Corning #25- 005-CI). The cells were cultured at 37ºC in 5% CO_2 in a humidified incubator and passaged every 2-3 days with 0.05% trypsin (Corning #25- 052-Cl) to maintain sub confluency.

Experiments-Drug Dilutions

Each drug was diluted in sterile filtered DMSO (except Brigatinib which was dissolved in sterile filtered 100% ethanol since it is insoluble in DMSO) as shown in Table 4.4 and 4.5 to bring the final concentration to 10 mM for each drug. These dilutions were then aliquoted into 10 µL batches.

Experiments-Drug Dose Assays

Before adding to cells, 990µL of full growth media was added to a 10 µL drug aliquot, diluting it to 100 µM, or 10X times the highest desired dose. This concentration was further

serially diluted 8 more times in full growth media containing 1% DMSO (to maintain the same DMSO concentration in each dilution) and by a factor of 3.16 each time. This results in 9 dilutions with the drug concentrations between 10 nM-100 µM. In 9 wells with cells seeded overnight in 90 µL media, 10 µL of the serially diluted drugs are added. In the 10^{th} well, 10 ul of full growth media containing 1% DMSO was added as the vehicle control dose.

Staining and Imaging

After 72 hours of treatment with the drugs, the cells were stained with Hoechst (BDBiosciences #BD 561908) and Propidium Iodide (Millipore Sigma #P4170) at a final concentration of 1 µg/ml and 2 µg/ml to stain all cells and dead cells respectively. After 30 minutes, the wells were imaged using the TagBFP (Excitation- 390nm, Emission- 447nm) and RFP filters (Excitation- 531/40 nm, Emission- 593/40 nm) in Cytation 5 (Biotek). Each image was flatfield corrected and background subtracted using CellProfiler. The nuclei were then identified using the `IdentifyPrimaryObjects` feature, and a pseudo image was generated. The number of all counts of cell nuclei stained with Hoechst and Propidium Iodide were compiled by CellProfiler and exported as csv files. The Propidium Iodide stained nuclei counts were subtracted from the Hoechst stained nuclei counts for each well, and this was taken as the live cell counts.

Drug Dose Response Fitting and IC_{50}

The drug dose responses were fitted in a sigmoidal hill-type equation[5] as follows

$$Y = Y_{\max} \times \frac{(D / EC_{50})^n}{1 + (D / EC_{50})^n}$$

Where Y is the inhibition, which in this case is 1 minus the relative cell number with respect to the control, Y_{max} is the maximal possible inhibition effect, EC_{50} is the half maximal drug dose for the sigmoid and n is the Hill coefficient and D is the dose of the drug in µM. We used `curve_fit` in Jupyter Notebook to obtain least-squares estimates for the three parameters for each of the drug dose responses.

We define IC_{50} as the drug dose at which the inhibition caused by the drug is 50% relative to control. Therefore, in the above equation, we can replace Y with 0.5 and solve for D, which should give us the IC_{50} dose. Solving the equation in this manner gives us the form of IC_{50} dose as follows

$$IC_{50} = EC_{50} \times \frac{(0.5)^{1/n}}{(Y_{\max} - 0.5)^{1/n}}$$

In case the maximal inhibition Y_{max} equals 1, then the IC_{50} dose equals the EC_{50} dose, since in this case, the dose for 50% inhibition of cell number also equals the half maximal dose of the sigmoid.

FIGURE 4.1: DOSE RESPONSE FITS FOR 20 DRUGS

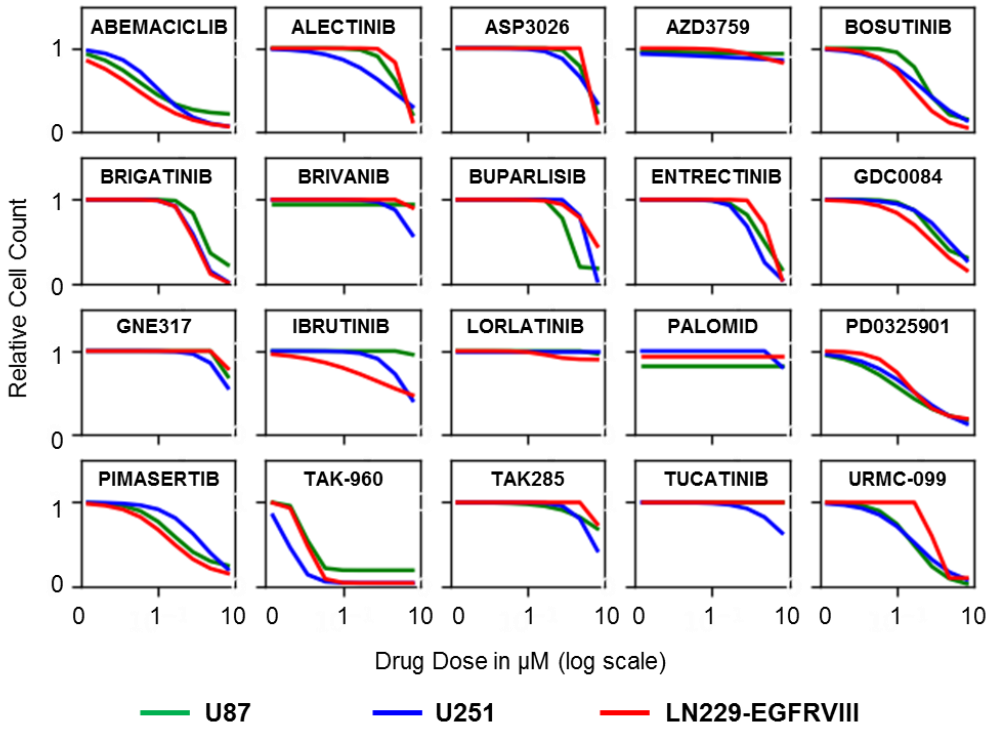


Table 4.1 Fitting parameters for highly responsive drugs

Drug	n-U87	Ymax-U87	EC50-U87	IC50-U87	n-U251	Ymax-U251	EC50-U251	IC50-U251	n-LN229-EGFRVIII	Ymax-LN229-EGFRVIII	EC50-LN229-EGFRVIII	IC50-LN229-EGFRVIII
Abemiciclib	0.682± 0.084	0.794± 0.031	0.028± 0.006	0.062	0.781± 0.104	0.952± 0.045	0.096± 0.021	0.109	0.535± 0.057	0.966± 0.037	0.023± 0.005	0.026
Bosutinib	1.538± 0.494	0.854± 0.079	0.622± 0.163	0.778	0.647± 0.158	1.000± 0.168	0.612± 0.432	0.612	0.919± 0.15	0.98±0. 061	0.282± 0.067	0.295
PD0325901	0.584± 0.064	0.879± 0.046	0.112± 0.031	0.18	0.551± 0.068	1.000± 0.084	0.34±0. 14	0.34	1.046± 0.147	0.82±0. 036	0.217± 0.037	0.332
Pimasertib	0.871± 0.146	0.784± 0.053	0.279± 0.074	0.533	0.818± 0.25	1.000± 0.256	1.909± 1.527	1.909	0.736± 0.178	0.218± 0.104	0.895± 0.097	0.3
TAK-960	2.776± 0.831	0.807± 0.018	0.009± 0.001	0.011	1.653± 0.343	0.959± 0.025	0.003± 0.0	0.003	2.377± 0.302	0.959± 0.012	0.009± 0.0	0.01
URMC-099	0.968± 0.208	1.0±0.0 78	0.302± 0.089	0.302	0.725± 0.233	1.000± 0.166	0.357± 0.247	0.357	2.283± 1.023	0.884± 0.079	0.81±0. 169	0.91

Table 4.2 Fitting parameters for mildly responsive drugs

Drug	n-U87	Ymax-U87	EC50-U87	IC50-U87	n-U251	Ymax-U251	EC50-U251	IC50-U251	n-LN229-EGFRVIII	Ymax-LN229-EGFRVIII	EC50-LN229-EGFRVIII	IC50-LN229-EGFRVIII
Alectinib	1.530± 0.528	1.000± 0.256	4.329± 1.94	4.329	0.582± 0.231	2.465± 4.692	1.000± 0.489	2.465	3.025± 8.918	1.000± 1.609	5.355± 11.811	5.355
ASP3026	2.126± 1.794	1.000± 0.744	5.842± 5.787	5.842	1.126± 0.752	1.000± 0.845	5.730± 9.806	5.73	25.741 ±10^6	0.887± 10^5	7.350± 10^8	7.424
Brigatinib	2.427± 0.876	0.787± 0.089	1.749± 0.414	2.197	1.830± 0.804	1.22±0. 347	1.000± 0.125	1.22	1.925± 0.345	1.000± 0.045	1.109± 0.113	1.109
Buparlisib	4.159± 6.081	0.818± 0.089	1.265± 0.486	1.41	4.107± 25.147	1.000± 1.082	4.480± 11.031	4.48	1.307± 0.953	1.000± 1.32	8.567± 18.088	8.567
Entrectinib	1.335± 0.411	1.000± 0.204	3.101± 1.303	3.101	1.606± 0.485	1.000± 0.116	1.614± 0.42	1.614	3.322± 6.83	1.000± 0.427	4.106± 2.995	4.106
GDC0084	1.398± 0.522	0.715± 0.093	0.944± 0.325	1.727	0.866± 0.21	0.969± 0.23	2.882± 1.921	3.1	0.726± 0.128	1.000± 0.128	1.013± 0.476	1.013

Table 4.3 List of drugs and their targets

S. No	Drug	Principal targets	Off targets	Specificity
1	Abemiciclib	CDK 4/6	GSK3A, GSK3B, CLK1	low
2	Alectinib	ALK	EPHA1, AAK1	high
3	ASP3026	TNK2	TNK1, HIPK, PLK4	low
4	AZD3759	EGFR	YES, EPHB4, LYN	low
5	Bosutinib	ABL1	LCK, ERBB3, FRK	low
6	Brigatinib	ALK	FLT3, ROS	low
7	Brivanib	KDR	PDGFRA, FLT1, KIT, STK35	low
8	Buparlisib	PIK3CA	mTOR, RSK4	low
9	Cabozantinib	KDR, DDR1, MET, FRK, HIPK4, EPHA6	LYN, PDGFRA, BLK	low
10	Entrectinib	TRKA	TRKB, TRKC, ROS	low
11	GDC0084	SRC	Syk, PI3KC2B,	low
12	GNE-317	PI3K	unknown	unknown
13	Ibrutinib	BTK, ERBB2, ERBB4	BLK, ERBB3	low
14	Palomid	mTOR	unknown	unknown
15	Lorlatinib	TNK2, ROS, ALK	_	high
16	PD0325901	MEK1/2	MAPK13	high
17	Pimasertib	MEK1/2	_	high
18	Saracatinib	LCK	ABL1, GAK, YES, LYN, HCK	low
19	Tak-285	ERBB2	EGFR	low
20	Tak-960	PLK1	PLK2, PLK3	high
21	Tucatinib	ERBB2	unknown	unknown
22	URMC-099	ABL1	MLK2, TRKA	low

Table 4.4 List of drugs and dilutions

Drug	Catalog Number	Quantity (mg)	Molar Mass(g)	DMSO added
Abemiciclib	S5716	25	506.6	4.93
Alectinib	S2762	10	482.6	2.07
ASP3026	S8054	50	580.7	8.61
AZD3759	S7971	25	459.9	5.44
Bosutinib	S1014	50	530.4	9.43
Brigatinib	S8229	25	584.1	4.28
Brivanib	S1084	50	370.4	13.50
Buparlisib	S2247	10	410.4	2.44
Cabozantinib	S1119	50	501.5	9.97
Entrectinib	S7998	25	560.6	4.46
GDC0084	S8163	25	382.4	6.54

Table 4.5 List of drugs and dilutions (contd.)

Drug	Catalog Number	Quantity (mg)	Molar Mass(g)	DMSO (ml)
GNE-317	S7798	25	414.5	6.03
Ibrutinib	S2680	10	440.5	2.27
Lorlatinib	S7536	25	406.4	6.15
Palomid	S2238	50	406.4	12.30
PD0325901	S1036	25	482.2	5.18
Pimasertib	S1475	50	431.2	11.60
Saracatinib	S1006	25	542	4.61
TAK-285	S2784	10	548	1.82
TAK-960	5403	10	598.06	1.67
Tucatinib	S8362	25	480.5	5.20
URMC-099	S7343	25	421.5	5.93

References-

1. Bayat Mokhtari, R., Homayouni, T.S., Baluch, N., Morgatskaya, E., Kumar, S., Das, B., and Yeger, H. (2017). Combination therapy in combating cancer. Oncotarget *8*, 38022–38043. 10.18632/oncotarget.16723.

2. Mellinghoff, I.K., Wang, M.Y., Vivanco, I., Haas-Kogan, D.A., Zhu, S., Dia, E.Q., Lu, K.V., Yoshimoto, K., Huang, J.H.Y., Chute, D.J., et al. (2005). Molecular determinants of the response of glioblastomas to EGFR kinase inhibitors. N Engl J Med *353*, 2012–2024. 10.1056/NEJMoa051918.

3. Gucalp, A., Sparano, J.A., Caravelli, J., Santamauro, J., Patil, S., Abbruzzi, A., Pellegrino, C., Bromberg, J., Dang, C., Theodoulou, M., et al. (2011). Phase II trial of saracatinib (AZD0530), an oral SRC-inhibitor for the treatment of patients with hormone receptor-negative metastatic breast cancer. Clin Breast Cancer *11*, 306–311. 10.1016/j.clbc.2011.03.021.

4. Nam, H.-J., Im, S.-A., Oh, D.-Y., Elvin, P., Kim, H.-P., Yoon, Y.-K., Min, A., Song, S.-H., Han, S.-W., Kim, T.-Y., et al. (2013). Antitumor Activity of Saracatinib (AZD0530), a c-Src/Abl Kinase Inhibitor, Alone or in Combination with Chemotherapeutic Agents in Gastric Cancer. Molecular Cancer Therapeutics *12*, 16–26. 10.1158/1535-7163.MCT-12-0109.

5. Yadav, B., Wennerberg, K., Aittokallio, T., and Tang, J. (2015). Searching for Drug Synergy in Complex Dose–Response Landscapes Using an Interaction Potency Model. Comput Struct Biotechnol J *13*, 504–513. 10.1016/j.csbj.2015.09.001.

CHAPTER FIVE

PREDICTING ANTI-CANCER DRUG COMBINATION RESPONSES WITH A TEMPORAL CELL STATE NETWORK MODEL

Abstract

Cancer chemotherapy combines multiple drugs but predicting the effects of drug combinations on cancer cell proliferation remains challenging. We hypothesized that by combining knowledge of single drug dose responses and cell state transition network dynamics, we could predict how a population of cancer cells will respond to drug combinations. We tested this hypothesis here using three targeted inhibitors of different cell cycle states in two different cell lines. We formulated a Markov model to capture temporal cell state transitions between different cell cycle phases, with single drug data constraining how drug doses affect transition rates. This model was able to predict the landscape of all three different pairwise drug combinations across all dose ranges for both cell lines with no additional data. This work shows how currently available or attainable information could be combined to predict how cancer cell populations respond to drug combinations.

Introduction

Matching chemotherapy regimens to cancer patients remains a grand challenge of oncology and personalized medicine. Targeted drugs often have genetic biomarkers, such as BRAFV600E for vemurafenib[1,2], EGFR mutations and copy number amplification for gefitinib[3,5,7], BCR-ABL fusion for imatinib[9,11], and HER2 copy number amplification for trastuzumab[13,14]. However, such matched patients often do not respond to therapy and/or eventually develop resistance. Why? One major driver is tumor heterogeneity; cells in different "states" that have different drug sensitivities. Cell states are often either defined by their histology or transcriptomics (through for example single cell RNAseq experiments)[15–21], and it is becoming appreciated that cells can transition between such states in development-like networks, sometimes called cell state networks[17,22]. Such plasticity between cell states can contribute to drug resistance[23–25], and combinations of drugs targeting different pathways and factors involving phenotype transition have been proposed to prevent such resistance[23]. Another is the multi-variate complexity of biochemical networks within which drug targets reside and by which chemotherapy drugs exert their action[26–31]. These networks can differ between cell states, adapt to therapy, and also give rise to non-intuitive therapy results, such as feedback loops and compensatory pathways underlying the efficacy of combining Raf and MEK inhibitor combinations, which lie in the same genetic pathway[32–35].

Massive agnostic efforts have screened thousands of cancer cell lines for sensitivity to hundreds of anti-cancer drugs, with matched multi-omic data to mine for biomarkers predictive of drug response[36–43]. These efforts, while substantial, still have not solved the

problem of how to match patients to drugs. Moreover, many chemotherapy regimens comprise combinations of 3-4 drugs. Comprehensive experimental exploration of just 2-way drug combinations for hundreds of anti-cancer drugs across a representative cohort is infeasible clinically, and currently unreachable even in cell culture systems.

The inability to obtain an experimental solution to the problem of matching drug combinations to patients has motivated computational modeling approaches. In principle, more comprehensive exploration of drug combination space could be achieved *in silico*. Various computational methods including mechanistic models and machine learning approaches have shown promise in predicting drug combination responses, especially taking into consideration context specific pathology and omics data as well as identifying specific biomarkers and drug-targets [44–49]. Regardless of the modeling methods being used, there is a widespread focus on using information about biochemical networks to facilitate drug combination response prediction [31,50–52]. Despite advanced methods being applied to integrate such information into models, building predictive drug combination response models remains an unsolved challenge. Any solution to this problem must invariably rely on experimental data that is already existing or is realistically attainable, such as single drug dose responses.

In this paper, rather than focus on modeling biochemical networks, we test the hypothesis that by combining knowledge of single drug dose responses and cell state transition network dynamics, we could predict how a population of cancer cells will respond to drug combinations. Although this hypothesis runs contrary to the predominant biochemical network-centered view of this problem, cell state transitions are largely

governed by biochemical networks in which drug targets are embedded, so in a sense this idea is encompassing prior logic. We test this hypothesis by focusing on three drugs that target cell cycle transitions in two different cell lines. A Markov model is developed to capture population growth and single drug dose responses, and then this model is used to predict all two-way drug combination responses with no further adjustment. Comparison of these model predictions to experimental tests shows surprisingly good agreement. These results suggest a sufficient formulation for predicting how cell population growth dynamics respond to drug combinations that relies on currently available and attainable information, and as such could have widespread impact for precision oncology.

Results

To test our hypothesis that knowledge of single drug dose responses combined with cell state transition network dynamics could enable prediction of drug combination responses **(Fig. 1a),** a model system is needed. There are a variety of choices for cell state transition networks and drugs which modulate them; here we focus on the cell cycle and three targeted kinase inhibitors (**Fig. 1b**). Specifically, we focus on a MEK1/2 inhibitor (PD0325901) that primarily blocks transition of G_0/G_1 cells [53], a CDK4/6 inhibitor (abemaciclib) that primarily blocks transition of (late)G_1/S cells [54–57], and a PLK1 inhibitor (TAK-960) that primarily blocks transition of G_2/M cells [58–61]. Drug dose response experiments evaluating cell number after 3 days of treatment show that both U87 and U251 cells are responsive to these drugs as single agents (**Fig. 1c**).

Before accounting for drug effects, we first constructed and parameterized a temporal cell state network model based on Markov formalisms (see Methods) that describes cell population dynamics in the absence of drug for U87 and U251 cells. Cells in the G_0/G_1 state can transition to the (late)G_1/S state, which can then transition to the G_2/M state. Upon transition from G_2/M to G_0/G_1, cell division occurs, increasing cell number by one. For this case without drug, we consider cell death transitions (which decrease cell number by one) to be negligible (we include them below to capture high dose features of some single drug responses). In each time step (chosen to be 1 hr), cells can either remain in their current state, or transition. We set the three unknown transition probabilities for each cell line by requiring agreement with population doubling time (30 hours-U87[62–64]

and 24 hours-U251[65–67] and the steady-state cell state ratios (64:19:17 for U87[68–72] and 58:23:19 for U251[73–75]) (**Fig. 1c**). Simulations recapitulate these features.

Drug action was modeled by assuming the transition parameters are a sigmoidal function of drug dose (see Methods). Fitting to the single drug dose response data yielded excellent agreement between model and data (**Fig. 1d**). At high doses for some drugs / cell lines, small cell death transition terms were included to account for the fact that observed cell numbers were lower than the initial number of cells (see Methods). Overall, these results demonstrate that the Markov model of cell state transition dynamics can capture cell population growth and single drug responses for the investigated system.

Now that we had a model that could take as input any of the three drugs at any dose and simulate cell population dynamics, we could predict how drug combinations would affect cell number for all pairwise combinations of the three drugs (**Fig. 2a-b, left**). These predictions demonstrate good agreement with independent experimental data for every drug combination for each cell line (**Fig. 2a-b-right**). Notably, no modifications were made to the model—only information about the cell state transition network dynamics and single drug dose responses were needed to perform this prediction.

Analysis of drug combination responses often includes assessment of drug synergy or antagonism, a more qualitative and categorical analysis. A common metric that we use here is excess over Bliss [76], which quantifies how much of the observed drug response is beyond statistically independent action by each drug. In particular, we used a variation of excess over Bliss that is more robust and reproducible because it uses sigmoidal fits to data to mitigate the impact of experimental noise in any single data point [77]. Values close to

zero indicate non-interacting combinations, positive values indicate synergistic combinations and negative values indicate antagonistic combinations. We stratified the data and model predictions into four dose quadrants (high/high, low/high, high/low, and low/low) and evaluated agreement between the two (**Fig. 2c-d; Fig. S1**). Both model and experiment show drug combinations were predominantly mildly antagonistic or non-interacting. Overall, these results provide support for the hypothesis that responses to anti-cancer drug combinations can be predicted with a model of cell state transition dynamics and knowledge of single drug dose responses.

Discussion

Predicting how varied drug combinations control cancer cell population growth is key to improving cancer precision medicine. Experimental solutions alone cannot cover the vast combinatoric space comprising drug combinations and different cancer cell types, necessitating computational approaches. Any computational approach should rely only on data that is available and/or feasibly attainable. Here, we explore the use of a computational approach that, rather than focus on biochemical networks in which drug targets reside, focuses on cell state networks where drugs influence transitions. By combining information about the cell state network dynamics with single cell drug dose responses, we were able to predict combination responses for three different targeted anti-cancer drugs in two different cell lines with no additional model modifications. We expect this finding to be impactful as it informs expansion to different drugs and cell types.

While we do not explicitly consider the role of biochemical networks in drug combination response, in a sense, they are implicitly accounted for in the mapping of drug concentration to cell state transition rates. In the investigated case, there was an arguably clean mapping of drug concentration to single transition rates, which simplified the effort. In other cases, such mapping may not be known *a priori* and/or more complex, i.e., a single drug may influence multiple transition rates. Biochemical network models that capture such complexities or mapping may prove useful in such situations[26–30,45,78]. Assumptions regarding the additivity (or not) of multi-drug action on transition rates would have to be asserted. The current availability of drug combination response data sets [79–82]could facilitate the testing of such methods. Such future work could explore drug combination

features we did not here, such as combining drugs that are not effective as single agents. They could also explore conditions that lead to drug combination synergy; the systems chosen here exhibited predominantly antagonistic behavior. Avoiding antagonism, however, is an important goal. It is thought that a small fraction of all drug combinations lead to synergistic behavior, but finding them, and how synergy is controlled by cell type, is of critical importance for precision oncology.

Application of this approach to other systems requires identification of cell state network models. Again, in our case the cell cycle is well established in terms of structure, but other such networks may not be. Studies have also confirmed the factors behind certain other cell state transitions-for instance, the transcriptomic factors and signaling molecules in different epithelial to mesenchymal transitions[30,83–85]. Cell state transition networks have been identified for multiple cancer types[15–19,21,22,86–88], generally by combining single cell measurements (e.g. single cell RNAseq), with perturbation time courses, such as enriching for one cell state and then observing the fractional composition dynamics. Recently, we proposed a general theory built upon modular response analysis[89–92] that allows one to reconstruct cell state networks from such perturbation time course data[78]. This theory is compatible with the Markov formalisms used here. Such Markov formulation may have further applicability to other cell state systems[22,87,88,93–95], but other approaches have been used[96–99]. Cell state transitions are subject to inherent stochasticity and describing the cell transitions as a Markov process is a common tool to capture this probabilistic aspect. However, this also relies on the assumption that the transition probabilities and the underlying variables are known and that the cell states are properly sampled and well

classified. There could be several knowledge gaps in these assumptions including that cells may be transcriptomically intermediate between canonically defined states[100] and biological data may be sparse[101]. Inclusion of methods such as lineage tracing and methods able to handle sparse data[98] may help address some of these gaps.

Overall, we have tested a relatively simple hypothesis that knowledge of single drug dose responses combined with cell state network dynamics is sufficient for prediction of drug combination responses. This hypothesis seems to hold true at least for the three drugs and two cell lines studied here, providing a potentially powerful rationale for guiding drug combination response modeling efforts. Expansion to more cell lines, cell state systems, and drugs will of course be important for further testing. Our findings here provide an important step towards being able to predict how cancer cell populations will respond to combinations of anti-cancer drugs, a key capability for cancer precision medicine.

Data Availability

The code needed to reproduce the data and figures are included and can be accessed at DOI: 10.5281/zenodo.7150564

Acknowledgments

MRB acknowledges funding from the National Institutes of Health Grants R35GM141891 and support from the Clemson Creative Inquiry program.

Competing Interests

The authors declare no competing interests.

Author Contributions

MRB and DS conceived of the work. DS made the figures. DS, WOM, IW and MRB performed the experiments. DS and MRB wrote the manuscript.

References

1. Chapman, P.B., Hauschild, A., Robert, C., Haanen, J.B., Ascierto, P., Larkin, J., Dummer, R., Garbe, C., Testori, A., Maio, M., et al. (2011). Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation. New England Journal of Medicine *364*, 2507–2516. 10.1056/NEJMoa1103782.

2. Sosman, J.A., Kim, K.B., Schuchter, L., Gonzalez, R., Pavlick, A.C., Weber, J.S., McArthur, G.A., Hutson, T.E., Moschos, S.J., Flaherty, K.T., et al. (2012). Survival in BRAF V600–Mutant Advanced Melanoma Treated with Vemurafenib. New England Journal of Medicine *366*, 707–714. 10.1056/NEJMoa1112302.

3. Herbst, R.S., Fukuoka, M., and Baselga, J. (2004). Gefitinib — a novel targeted approach to treating cancer. Nat Rev Cancer *4*, 956–965. 10.1038/nrc1506.

4. Roy, S., Lane, T., Allen, C., Aragon, A.D., and Werner-Washburne, M. (2006). A Hidden-State Markov Model for Cell Population Deconvolution. Journal of Computational Biology *13*, 1749–1774. 10.1089/cmb.2006.13.1749.

5. Paez, J.G., Jänne, P.A., Lee, J.C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F.J., Lindeman, N., Boggon, T.J., et al. (2004). EGFR Mutations in Lung Cancer: Correlation with Clinical Response to Gefitinib Therapy. Science *304*, 1497–1500. 10.1126/science.1099314.

6. Bar-Joseph, Z., Farkash, S., Gifford, D.K., Simon, I., and Rosenfeld, R. (2004). Deconvolving cell cycle expression data with complementary information. Bioinformatics *20*, i23–i30. 10.1093/bioinformatics/bth915.

7. Takano, T., Ohe, Y., Sakamoto, H., Tsuta, K., Matsuno, Y., Tateishi, U., Yamamoto, S., Nokihara, H., Yamamoto, N., Sekine, I., et al. (2005). Epidermal Growth Factor Receptor Gene Mutations and Increased Copy Numbers Predict Gefitinib Sensitivity in Patients With Recurrent Non–Small-Cell Lung Cancer. JCO *23*, 6829–6837. 10.1200/JCO.2005.01.0793.

8. Rees, M.G., Brenan, L., do Carmo, M., Duggan, P., Bajrami, B., Arciprete, M., Boghossian, A., Vaimberg, E., Ferrara, S.J., Lewis, T.A., et al. (2022). Systematic identification of biomarker-driven drug combinations to overcome resistance. Nat Chem Biol *18*, 615–624. 10.1038/s41589-022-00996-7.

9. Druker, B.J., and Lydon, N.B. (2000). Lessons learned from the development of an Abl tyrosine kinase inhibitor for chronic myelogenous leukemia. J Clin Invest *105*, 3–7. 10.1172/JCI9083.

10. Jaaks, P., Coker, E.A., Vis, D.J., Edwards, O., Carpenter, E.F., Leto, S.M., Dwane, L., Sassi, F., Lightfoot, H., Barthorpe, S., et al. (2022). Effective drug

combinations in breast, colon and pancreatic cancer cells. Nature *603*, 166–173. 10.1038/s41586-022-04437-2.

11. Druker, B.J., Guilhot, F., O'Brien, S.G., Gathmann, I., Kantarjian, H., Gattermann, N., Deininger, M.W.N., Silver, R.T., Goldman, J.M., Stone, R.M., et al. (2006). Five-Year Follow-up of Patients Receiving Imatinib for Chronic Myeloid Leukemia. New England Journal of Medicine *355*, 2408–2417. 10.1056/NEJMoa062867.

12. Pantaleón García, J., Kulkarni, V.V., Reese, T.C., Wali, S., Wase, S.J., Zhang, J., Singh, R., Caetano, M.S., Kadara, H., Moghaddam, S.J., et al. (2022). OBIF: an omics-based interaction framework to reveal molecular drivers of synergy. NAR Genomics and Bioinformatics *4*, lqac028. 10.1093/nargab/lqac028.

13. Hudis, C.A. (2007). Trastuzumab — Mechanism of Action and Use in Clinical Practice. New England Journal of Medicine *357*, 39–51. 10.1056/NEJMra043186.

14. Slamon, D., Eiermann, W., Robert, N., Pienkowski, T., Martin, M., Press, M., Mackey, J., Glaspy, J., Chan, A., Pawlicki, M., et al. (2011). Adjuvant Trastuzumab in HER2-Positive Breast Cancer. New England Journal of Medicine *365*, 1273–1283. 10.1056/NEJMoa0910383.

15. Bailey, P., Chang, D.K., Nones, K., Johns, A.L., Patch, A.-M., Gingras, M.-C., Miller, D.K., Christ, A.N., Bruxner, T.J.C., Quinn, M.C., et al. (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. Nature *531*, 47–52. 10.1038/nature16965.

16. Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., et al. (2015). The consensus molecular subtypes of colorectal cancer. Nat Med *21*, 1350–1356. 10.1038/nm.3967.

17. Neftel, C., Laffy, J., Filbin, M.G., Hara, T., Shore, M.E., Rahme, G.J., Richman, A.R., Silverbush, D., Shaw, M.L., Hebert, C.M., et al. (2019). An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. Cell *178*, 835-849.e21. 10.1016/j.cell.2019.06.024.

18. Rudin, C.M., Poirier, J.T., Byers, L.A., Dive, C., Dowlati, A., George, J., Heymach, J.V., Johnson, J.E., Lehman, J.M., MacPherson, D., et al. (2019). Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. Nat Rev Cancer *19*, 289–297. 10.1038/s41568-019-0133-9.

19. Wang, Q., Hu, B., Hu, X., Kim, H., Squatrito, M., Scarpace, L., deCarvalho, A.C., Lyu, S., Li, P., Li, Y., et al. (2017). Tumor Evolution of Glioma-Intrinsic Gene

Expression Subtypes Associates with Immunological Changes in the Microenvironment. Cancer Cell *32*, 42-56.e6. 10.1016/j.ccell.2017.06.003.

20. Yersal, O., and Barutca, S. (2014). Biological subtypes of breast cancer: Prognostic and therapeutic implications. World Journal of Clinical Oncology *5*, 412–424. 10.5306/wjco.v5.i3.412.

21. Rukhlenko, O.S., Halasz, M., Rauch, N., Zhernovkov, V., Prince, T., Wynne, K., Maher, S., Kashdan, E., MacLeod, K., Carragher, N.O., et al. (2022). Control of cell state transitions. Nature *609*, 975–985. 10.1038/s41586-022-05194-y.

22. Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic State Transitions Give Rise to Phenotypic Equilibrium in Populations of Cancer Cells. Cell *146*, 633–644. 10.1016/j.cell.2011.07.026.

23. Boumahdi, S., and de Sauvage, F.J. (2020). The great escape: tumour cell plasticity in resistance to targeted therapy. Nat Rev Drug Discov *19*, 39–56. 10.1038/s41573-019-0044-1.

24. Brown, R., Curry, E., Magnani, L., Wilhelm-Benartzi, C.S., and Borley, J. (2014). Poised epigenetic states and acquired drug resistance in cancer. Nat Rev Cancer *14*, 747–753. 10.1038/nrc3819.

25. Easwaran, H., Tsai, H.-C., and Baylin, S.B. (2014). Cancer Epigenetics: Tumor Heterogeneity, Plasticity of Stem-like States, and Drug Resistance. Molecular Cell *54*, 716–727. 10.1016/j.molcel.2014.05.015.

26. Csermely, P., Korcsmáros, T., Kiss, H.J.M., London, G., and Nussinov, R. (2013). Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review. Pharmacology & Therapeutics *138*, 333–408. 10.1016/j.pharmthera.2013.01.016.

27. Hornberg, J.J., Bruggeman, F.J., Westerhoff, H.V., and Lankelma, J. (2006). Cancer: A Systems Biology disease. Biosystems *83*, 81–90. 10.1016/j.biosystems.2005.05.014.

28. Kolch, W., and Pitt, A. (2010). Functional proteomics to dissect tyrosine kinase signalling pathways in cancer. Nat Rev Cancer *10*, 618–629. 10.1038/nrc2900.

29. Krogan, N.J., Lippman, S., Agard, D.A., Ashworth, A., and Ideker, T. (2015). The Cancer Cell Map Initiative: Defining the Hallmark Networks of Cancer. Molecular Cell *58*, 690–698. 10.1016/j.molcel.2015.05.008.

30. Moustakas, A., and Heldin, C.-H. (2007). Signaling networks guiding epithelial–mesenchymal transitions during embryogenesis and cancer progression. Cancer Science *98*, 1512–1520. 10.1111/j.1349-7006.2007.00550.x.

31. Yin, N., Ma, W., Pei, J., Ouyang, Q., Tang, C., and Lai, L. (2014). Synergistic and antagonistic drug combinations depend on network topology. PLoS One *9*, e93960. 10.1371/journal.pone.0093960.

32. Flaherty, K.T., Infante, J.R., Daud, A., Gonzalez, R., Kefford, R.F., Sosman, J., Hamid, O., Schuchter, L., Cebon, J., Ibrahim, N., et al. (2012). Combined BRAF and MEK Inhibition in Melanoma with BRAF V600 Mutations. New England Journal of Medicine *367*, 1694–1703. 10.1056/NEJMoa1210093.

33. Long, G.V., Stroyakovskiy, D., Gogas, H., Levchenko, E., de Braud, F., Larkin, J., Garbe, C., Jouary, T., Hauschild, A., Grob, J.J., et al. (2014). Combined BRAF and MEK Inhibition versus BRAF Inhibition Alone in Melanoma. New England Journal of Medicine *371*, 1877–1888. 10.1056/NEJMoa1406037.

34. Poulikakos, P.I., Zhang, C., Bollag, G., Shokat, K.M., and Rosen, N. (2010). RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. Nature *464*, 427–430. 10.1038/nature08902.

35. Rukhlenko, O.S., Khorsand, F., Krstic, A., Rozanc, J., Alexopoulos, L.G., Rauch, N., Erickson, K.E., Hlavacek, W.S., Posner, R.G., Gómez-Coca, S., et al. (2018). Dissecting RAF Inhibitor Resistance by Structure-based Modeling Reveals Ways to Overcome Oncogenic RAS Signaling. Cell Systems *7*, 161-179.e14. 10.1016/j.cels.2018.06.002.

36. Amundson, S.A., Myers, T.G., Scudiero, D., Kitada, S., Reed, J.C., and Fornace, A.J., Jr. (2000). An Informatics Approach Identifying Markers of Chemosensitivity in Human Cancer Cell Lines1. Cancer Research *60*, 6101–6110.

37. Bansal, M., Yang, J., Karan, C., Menden, M.P., Costello, J.C., Tang, H., Xiao, G., Li, Y., Allen, J., Zhong, R., et al. (2014). A community computational challenge to predict the activity of pairs of compounds. Nat Biotechnol *32*, 1213–1222. 10.1038/nbt.3052.

38. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607. 10.1038/nature11003.

39. Costello, J.C., Heiser, L.M., Georgii, E., Gönen, M., Menden, M.P., Wang, N.J., Bansal, M., Ammad-ud-din, M., Hintsanen, P., Khan, S.A., et al. (2014). A community

effort to assess and improve drug sensitivity prediction algorithms. Nat Biotechnol *32*, 1202–1212. 10.1038/nbt.2877.

40. Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature *483*, 570–575. 10.1038/nature11005.

41. Menden, M.P., Iorio, F., Garnett, M., McDermott, U., Benes, C.H., Ballester, P.J., and Saez-Rodriguez, J. (2013). Machine Learning Prediction of Cancer Cell Sensitivity to Drugs Based on Genomic and Chemical Properties. PLOS ONE *8*, e61318. 10.1371/journal.pone.0061318.

42. Menden, M.P., Wang, D., Mason, M.J., Szalai, B., Bulusu, K.C., Guan, Y., Yu, T., Kang, J., Jeon, M., Wolfinger, R., et al. (2019). Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. Nat Commun *10*, 2674. 10.1038/s41467-019-09799-2.

43. Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Research *41*, D955–D961. 10.1093/nar/gks1111.

44. Bouhaddou, M., Barrette, A.M., Stern, A.D., Koch, R.J., DiStefano, M.S., Riesel, E.A., Santos, L.C., Tan, A.L., Mertz, A.E., and Birtwistle, M.R. (2018). A mechanistic pan-cancer pathway model informed by multi-omics data interprets stochastic cell fate responses to drugs and mitogens. PLOS Computational Biology *14*, e1005985. 10.1371/journal.pcbi.1005985.

45. Erdem, C., Mutsuddy, A., Bensman, E.M., Dodd, W.B., Saint-Antoine, M.M., Bouhaddou, M., Blake, R.C., Gross, S.M., Heiser, L.M., Feltus, F.A., et al. (2022). A scalable, open-source implementation of a large-scale mechanistic model for single cell proliferation and death signaling. Nat Commun *13*, 3555. 10.1038/s41467-022-31138-1.

46. Fröhlich, F., Kaltenbacher, B., Theis, F.J., and Hasenauer, J. (2017). Scalable Parameter Estimation for Genome-Scale Biochemical Reaction Networks. PLOS Computational Biology *13*, e1005331. 10.1371/journal.pcbi.1005331.

47. Li, J., Xu, H., and McIndoe, R.A. (2022). A novel network based linear model for prioritization of synergistic drug combinations. PLOS ONE *17*, e0266382. 10.1371/journal.pone.0266382.

48. Narayan, R.S., Molenaar, P., Teng, J., Cornelissen, F.M.G., Roelofs, I., Menezes, R., Dik, R., Lagerweij, T., Broersma, Y., Petersen, N., et al. (2020). A cancer drug

atlas enables synergistic targeting of independent drug vulnerabilities. Nat Commun *11*, 2935. 10.1038/s41467-020-16735-2.

49. Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., et al. (2019). Applications of machine learning in drug discovery and development. Nat Rev Drug Discov *18*, 463–477. 10.1038/s41573-019-0024-5.

50. Chen, D., Liu, X., Yang, Y., Yang, H., and Lu, P. (2015). Systematic synergy modeling: understanding drug synergy from a systems biology perspective. BMC Syst Biol *9*, 56. 10.1186/s12918-015-0202-y.

51. Cheng, F., Kovács, I.A., and Barabási, A.-L. (2019). Network-based prediction of drug combinations. Nat Commun *10*, 1197. 10.1038/s41467-019-09186-x.

52. Tang, J., Gautam, P., Gupta, A., He, L., Timonen, S., Akimov, Y., Wang, W., Szwajda, A., Jaiswal, A., Turei, D., et al. (2019). Network pharmacology modeling identifies synergistic Aurora B and ZAK interaction in triple-negative breast cancer. npj Syst Biol Appl *5*, 1–11. 10.1038/s41540-019-0098-z.

53. Tai, Y.-T., Kim, K., Li, X.-F., Fulciniti, M., Song, W., Nahar, S., Burger, P., Rumizen, M.J., Podar, K., Chauhan, D., et al. (2009). Targeting MEK1/2 Signaling Cascade by AS703026, a Novel Selective MEK1/2 Inhibitor, Induces Pleiotropic Anti-Myeloma Activity in Vitro and In Vivo. Blood *114*, 3848. 10.1182/blood.V114.22.3848.3848.

54. Burkhart, D.L., and Sage, J. (2008). Cellular mechanisms of tumour suppression by the retinoblastoma gene. Nat Rev Cancer *8*, 671–682. 10.1038/nrc2399.

55. Kent, L.N., and Leone, G. (2019). The broken cycle: E2F dysfunction in cancer. Nat Rev Cancer *19*, 326–338. 10.1038/s41568-019-0143-7.

56. Malumbres, M., and Barbacid, M. (2009). Cell cycle, CDKs and cancer: a changing paradigm. Nat Rev Cancer *9*, 153–166. 10.1038/nrc2602.

57. Otto, T., and Sicinski, P. (2017). Cell cycle proteins as promising targets in cancer therapy. Nat Rev Cancer *17*, 93–115. 10.1038/nrc.2016.138.

58. Gheghiani, L., Loew, D., Lombard, B., Mansfeld, J., and Gavet, O. (2017). PLK1 Activation in Late G2 Sets Up Commitment to Mitosis. Cell Reports *19*, 2060–2073. 10.1016/j.celrep.2017.05.031.

59. Lan, R., Lin, G., Yin, F., Xu, J., Zhang, X., Wang, J., Wang, Y., Gong, J., Ding, Y.-H., Yang, Z., et al. (2012). Dissecting the phenotypes of Plk1 inhibition in cancer

cells using novel kinase inhibitory chemical CBB2001. Lab Invest *92*, 1503–1514. 10.1038/labinvest.2012.114.

60. Pezuk, J.A., Brassesco, M.S., Morales, A.G., de Oliveira, J.C., de Oliveira, H.F., Scrideli, C.A., and Tone, L.G. (2013). Inhibition of Polo-Like Kinase 1 Induces Cell Cycle Arrest and Sensitizes Glioblastoma Cells to Ionizing Radiation. Cancer Biother Radiopharm *28*, 516–522. 10.1089/cbr.2012.1415.

61. Zhang, Z., Su, W.-H., Feng, C., Yu, D.-H., Cui, C., Xu, X.-Y., and Yu, B.-Z. (2007). Polo-like kinase 1 may regulate G2/M transition of mouse fertilized eggs by means of inhibiting the phosphorylation of Tyr 15 of Cdc2. Mol Reprod Dev *74*, 1247–1254. 10.1002/mrd.20703.

62. Jin, S.-G., Ryu, H.-H., Li, S.-Y., Li, C.-H., Lim, S.-H., Jang, W.-Y., and Jung, S. (2016). Nogo-A inhibits the migration and invasion of human malignant glioma U87MG cells. Oncology Reports *35*, 3395–3402. 10.3892/or.2016.4737.

63. Lane, R., Simon, T., Vintu, M., Solkin, B., Koch, B., Stewart, N., Benstead-Hume, G., Pearl, F.M.G., Critchley, G., Stebbing, J., et al. (2019). Cell-derived extracellular vesicles can be used as a biomarker reservoir for glioblastoma tumor subtyping. Commun Biol *2*, 1–12. 10.1038/s42003-019-0560-x.

64. Oraiopoulou, M.-E., Tzamali, E., Tzedakis, G., Vakis, A., Papamatheakis, J., and Sakkalis, V. (2017). In Vitro/In Silico Study on the Role of Doubling Time Heterogeneity among Primary Glioblastoma Cell Lines. BioMed Research International *2017*, e8569328. 10.1155/2017/8569328.

65. Cowley, G.S., Weir, B.A., Vazquez, F., Tamayo, P., Scott, J.A., Rusin, S., East-Seletsky, A., Ali, L.D., Gerath, W.F., Pantel, S.E., et al. (2014). Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. Sci Data *1*, 140035. 10.1038/sdata.2014.35.

66. Lee, D.W., Lee, M.-Y., Ku, B., and Nam, D.-H. (2015). Automatic 3D Cell Analysis in High-Throughput Microarray Using Micropillar and Microwell Chips. J Biomol Screen *20*, 1178–1184. 10.1177/1087057115597635.

67. Weller, M., Rieger, J., Grimmel, C., Van Meir, E.G., De Tribolet, N., Krajewski, S., Reed, J.C., von Deimling, A., and Dichgans, J. (1998). Predicting chemoresistance in human malignant glioma cells: The role of molecular genetic analyses. International Journal of Cancer *79*, 640–644. 10.1002/(SICI)1097-0215(19981218)79:6<640::AID-IJC15>3.0.CO;2-Z.

68. Anandharaj, A., Cinghu, S., and Park, W.-Y. (2011). Rapamycin-mediated mTOR inhibition attenuates survivin and sensitizes glioblastoma cells to radiation therapy. Acta Biochimica et Biophysica Sinica *43*, 292–300. 10.1093/abbs/gmr012.

69. Hu, G., Cun, X., Ruan, S., Shi, K., Wang, Y., Kuang, Q., Hu, C., Xiao, W., He, Q., and Gao, H. (2016). Utilizing G2/M retention effect to enhance tumor accumulation of active targeting nanoparticles. Sci Rep *6*, 27669. 10.1038/srep27669.

70. Pisapia, L., Barba, P., Cortese, A., Cicatiello, V., Morelli, F., and Del Pozzo, G. (2015). EBP1 protein modulates the expression of human MHC class II molecules in non-hematopoietic cancer cells. International Journal of Oncology *47*, 481–489. 10.3892/ijo.2015.3051.

71. Metabolic Impact of Anti-Angiogenic Agents on U87 Glioma Cells (2014). PLOS ONE *9*, e99198. 10.1371/journal.pone.0099198.

72. Xu, P.-F., Yang, J.-A., Liu, J.-H., Yang, X., Liao, J.-M., Yuan, F.-E., Liu, B.-H., and Chen, Q.-X. (2019). PI3Kβ inhibitor AZD6482 exerts antiproliferative activity and induces apoptosis in human glioblastoma cells. Oncol Rep *41*, 125–132. 10.3892/or.2018.6845.

73. He, X., Maimaiti, M., Jiao, Y., Meng, X., and Li, H. (2018). Sinomenine Induces G1-Phase Cell Cycle Arrest and Apoptosis in Malignant Glioma Cells Via Downregulation of Sirtuin 1 and Induction of p53 Acetylation. Technol Cancer Res Treat *17*, 1533034618770305. 10.1177/1533034618770305.

74. Liu, Y., Wang, F., Liu, Y., Yao, Y., Lv, X., Dong, B., Li, J., Ren, S., Yao, Y., and Xu, Y. (2016). RNF135, RING finger protein, promotes the proliferation of human glioblastoma cells in vivo and in vitro via the ERK pathway. Sci Rep *6*, 20642. 10.1038/srep20642.

75. ZHANG, Y., CHEN, X., LIAN, H., LIU, J., ZHOU, B., HAN, S., PENG, B., YIN, J., LIU, W., and HE, X. (2014). MicroRNA-503 acts as a tumor suppressor in glioblastoma for multiple antitumor effects by targeting IGF-1R. Oncol Rep *31*, 1445–1452. 10.3892/or.2013.2951.

76. Bliss, C.I. (1956). THE CALCULATION OF MICROBIAL ASSAYS. Bacteriol Rev *20*, 243–258.

77. Yadav, B., Wennerberg, K., Aittokallio, T., and Tang, J. (2015). Searching for Drug Synergy in Complex Dose–Response Landscapes Using an Interaction Potency Model. Comput Struct Biotechnol J *13*, 504–513. 10.1016/j.csbj.2015.09.001.

78. Sarmah, D., Smith, G.R., Bouhaddou, M., Stern, A.D., Erskine, J., and Birtwistle, M.R. (2022). Network Inference from Perturbation Time Course Data. 341008. 10.1101/341008.

79. Holbeck, S.L., Camalier, R., Crowell, J.A., Govindharajulu, J.P., Hollingshead, M., Anderson, L.W., Polley, E., Rubinstein, L., Srivastava, A., Wilsker, D., et al.

(2017). The National Cancer Institute ALMANAC: A Comprehensive Screening Resource for the Detection of Anticancer Drug Pairs with Enhanced Therapeutic Activity. Cancer Research *77*, 3564–3576. 10.1158/0008-5472.CAN-17-0489.

80. Liu, H., Zhang, W., Zou, B., Wang, J., Deng, Y., and Deng, L. (2020). DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. Nucleic Acids Research *48*, D871–D881. 10.1093/nar/gkz1007.

81. O'Neil, J., Benita, Y., Feldman, I., Chenard, M., Roberts, B., Liu, Y., Li, J., Kral, A., Lejnine, S., Loboda, A., et al. (2016). An Unbiased Oncology Compound Screen to Identify Novel Combination Strategies. Molecular Cancer Therapeutics *15*, 1155–1162. 10.1158/1535-7163.MCT-15-0843.

82. Shtar, G., Azulay, L., Nizri, O., Rokach, L., and Shapira, B. (2022). CDCDB: A large and continuously updated drug combination database. Sci Data *9*, 263. 10.1038/s41597-022-01360-z.

83. Bhowmick, N.A., Ghiassi, M., Bakin, A., Aakre, M., Lundquist, C.A., Engel, M.E., Arteaga, C.L., and Moses, H.L. (2001). Transforming Growth Factor-β1 Mediates Epithelial to Mesenchymal Transdifferentiation through a RhoA-dependent Mechanism. MBoC *12*, 27–36. 10.1091/mbc.12.1.27.

84. Bolós, V., Peinado, H., Pérez-Moreno, M.A., Fraga, M.F., Esteller, M., and Cano, A. (2003). The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors. Journal of Cell Science *116*, 499–511. 10.1242/jcs.00224.

85. Tania, M., Khan, Md.A., and Fu, J. (2014). Epithelial to mesenchymal transition inducing transcription factors and metastatic cancer. Tumor Biol. *35*, 7335–7342. 10.1007/s13277-014-2163-y.

86. Schlauch, D., Glass, K., Hersh, C.P., Silverman, E.K., and Quackenbush, J. (2017). Estimating drivers of cell state transitions using gene regulatory network models. BMC Systems Biology *11*, 139. 10.1186/s12918-017-0517-y.

87. Sehl, M.E., Shimada, M., Landeros, A., Lange, K., and Wicha, M.S. (2015). Modeling of Cancer Stem Cell State Transitions Predicts Therapeutic Response. PLOS ONE *10*, e0135797. 10.1371/journal.pone.0135797.

88. Su, Y., Wei, W., Robert, L., Xue, M., Tsoi, J., Garcia-Diaz, A., Homet Moreno, B., Kim, J., Ng, R.H., Lee, J.W., et al. (2017). Single-cell analysis resolves the cell state transition and signaling dynamics associated with melanoma drug-induced resistance. Proceedings of the National Academy of Sciences *114*, 13679–13684. 10.1073/pnas.1712064115.

89. Kholodenko, B.N., Kiyatkin, A., Bruggeman, F.J., Sontag, E., Westerhoff, H.V., and Hoek, J.B. (2002). Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. Proceedings of the National Academy of Sciences *99*, 12841–12846. 10.1073/pnas.192442699.

90. Lill, D., Rukhlenko, O.S., Mc Elwee, A.J., Kashdan, E., Timmer, J., and Kholodenko, B.N. (2019). Mapping connections in signaling networks with ambiguous modularity. npj Syst Biol Appl *5*, 1–11. 10.1038/s41540-019-0096-1.

91. Sontag, E., Kiyatkin, A., and Kholodenko, B.N. (2004). Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data. Bioinformatics *20*, 1877–1886. 10.1093/bioinformatics/bth173.

92. Thomaseth, C., Fey, D., Santra, T., Rukhlenko, O.S., Radde, N.E., and Kholodenko, B.N. (2018). Impact of measurement noise, experimental design, and estimation methods on Modular Response Analysis based network reconstruction. Sci Rep *8*, 16217. 10.1038/s41598-018-34353-3.

93. Armond, J.W., Saha, K., Rana, A.A., Oates, C.J., Jaenisch, R., Nicodemi, M., and Mukherjee, S. (2014). A stochastic model dissects cell states in biological transition processes. Sci Rep *4*, 3692. 10.1038/srep03692.

94. Buder, T., Deutsch, A., Seifert, M., and Voss-Böhme, A. (2017). CellTrans: An R Package to Quantify Stochastic Cell State Transitions. Bioinform Biol Insights *11*, 1177932217712241. 10.1177/1177932217712241.

95. Jagannathan, N.S., Ihsan, M.O., Kin, X.X., Welsch, R.E., Clément, M.-V., and Tucker-Kellogg, L. (2020). Transcompp: understanding phenotypic plasticity by estimating Markov transition rates for cell state transitions. Bioinformatics *36*, 2813–2820. 10.1093/bioinformatics/btaa021.

96. Sáez, M., Blassberg, R., Camacho-Aguilar, E., Siggia, E.D., Rand, D.A., and Briscoe, J. (2022). Statistically derived geometrical landscapes capture principles of decision-making dynamics during cell fate transitions. Cell Systems *13*, 12-28.e3. 10.1016/j.cels.2021.08.013.

97. Stumpf, P.S., Smith, R.C.G., Lenz, M., Schuppert, A., Müller, F.-J., Babtie, A., Chan, T.E., Stumpf, M.P.H., Please, C.P., Howison, S.D., et al. (2017). Stem Cell Differentiation as a Non-Markov Stochastic Process. Cell Systems *5*, 268-282.e7. 10.1016/j.cels.2017.08.009.

98. Wang, S.-W., Herriges, M.J., Hurley, K., Kotton, D.N., and Klein, A.M. (2022). CoSpar identifies early cell fate biases from single-cell transcriptomic and lineage information. Nat Biotechnol *40*, 1066–1074. 10.1038/s41587-022-01209-1.

99. Zhou, P., Wang, S., Li, T., and Nie, Q. (2021). Dissecting transition cells from single-cell transcriptome data through multiscale stochastic dynamics. Nat Commun *12*, 5609. 10.1038/s41467-021-25548-w.

100. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F.D., and Klein, A.M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. Science *367*, eaaw3381. 10.1126/science.aaw3381.

101. Mulas, C., Chaigne, A., Smith, A., and Chalut, K.J. (2021). Cell state transitions: definitions and challenges. Development *148*, dev199950. 10.1242/dev.199950.

102. Prinz, H. (2009). Hill coefficients, dose–response curves and allosteric mechanisms. J Chem Biol *3*, 37–44. 10.1007/s12154-009-0029-3.

Methods

Experimental Methods

Cell Culture. U87 and U251 cells (both STR profiled internally) were cultured in full growth medium comprising DMEM (Gibco #10313039) supplemented with 10% FBS (Corning #35-011-CV) and 2 mM L-Glutamine (Corning #25- 005-CI). The cells were cultured at 37^{o}C in 5% CO_2 in a humidified incubator and passaged every 2-3 days with 0.05% trypsin (Corning #25- 052-Cl) to maintain sub confluency.

Drug Dose Response Experiments

U87 and U251 cells were seeded in 96 well plates (Corning-Falcon #353072) with 500 cells per well, counted with a hemocytometer. Cells were seeded in 90 µl full growth media and cultured overnight. The next day, 10 µl of media containing 10X the final drug concentration was added and the plates cultured for 72 hours.

The three drugs were procured from the following sources - PD0325901 (Selleckchem #S1036), Abemaciclib (Selleckchem #S5716) and TAK960 (Tocris #5403). The quantities of each drug-PD0325901 (25 mg, molar mass-482.19g), Abemaciclib (25 mg, molar mass-506.59g) and TAK960 (10 mg, molar mass-598.06g) corresponded to 0.0518 millimoles, 0.0493 millimoles and 0.0167 millimoles respectively and were diluted in 5.18 mls, 4.18 mls and 1.67 mls of sterile filtered DMSO to bring the final concentration to 10 mM for each drug. These dilutions were then aliquoted into 10 µL batches. Before adding to cells, 990µL of full growth media was added to a 10 µL drug aliquot, diluting it to 100 µM, or 10X times the highest desired dose. This concentration was further serially diluted 8 more times in full growth media containing 1% DMSO (to maintain the same

DMSO concentration in each dilution) and by a factor of 3.16 each time. This results in 9 dilutions with the drug concentrations between 10 nM-100 µM. In 9 wells with cells seeded overnight in 90 µL media, 10 µL of the serially diluted drugs are added. In the 10^{th} well, 10 ul of full growth media containing 1% DMSO was added as the vehicle control dose.

For combination dose responses, U87 and U251 cells were seeded in 96 well plates (Corning-Falcon #353072) with 500 cells in each well. Eight by eight wells were seeded in 150 µl full growth media and cultured overnight. The next day, 25 µl of media was added twice to each well, each containing 8x of the final drug concentration and cultured for 72 hours. The final drug concentrations were chosen to reflect their responsive range for the cell lines (1.22 nM-5µM for PD0325901 and Abemaciclib, 0.0122nM-50nM for TAK960). Before adding to cells, full growth media was added to a 10 µL drug aliquot, diluting it to 8X times the highest desired dose. This concentration was further serially diluted 6 more times in full growth media containing 1% DMSO (to maintain the same DMSO concentration in each dilution) and by a factor of 4 each time. This results in 7 dilutions with the desired drug concentrations.

Staining and Computational Image Analysis

After 72 hours of treatment with the drugs, the cells were stained with Hoechst (BDBiosciences #BD 561908) and Propidium Iodide (Millipore Sigma #P4170) at a final concentration of 1 µg/ml and 2 µg/ml to stain all cells and dead cells respectively. After 30 minutes, the wells were imaged using the TagBFP (Excitation- 390nm, Emission- 447nm) and RFP filters (Excitation- 531/40 nm, Emission- 593/40 nm) in Cytation 5 (Biotek). Each image was flatfield corrected and background subtracted using CellProfiler. The

nuclei were then identified using the `IdentifyPrimaryObjects` feature, and a pseudo image was generated. The number of all counts of cell nuclei stained with Hoechst and Propidium Iodide were compiled by CellProfiler and exported as csv files. The Propidium Iodide stained nuclei counts were subtracted from the Hoechst stained nuclei counts for each well, and this was taken as the live cell counts.

Model and Computational Methods

A Markov Model of Temporal Cell State Transitions

Consider a Markov state model comprising three nodes, representing cell states G_0/G_1, late G_1/S and G_2/M, taken as 1, 2, and 3 in short. M_1, M_2 and M_3 are the proportion of cells transitioning from states 1-2, 2-3 and 3-1 respectively, within a given timestep-one hour for our considerations. M_{11}, M_{22} and M_{33} are the proportion of cells that did not transition from states 1, 2 and 3 respectively. A cell in state 3 undergoes cell division which gives rise to two cells in state 1. We formulate this scenario using a jump Markov process model as follows:

$$\begin{aligned} x_{1,t+1} &= M_{11}x_{1,t} + 2\times M_3 x_{3,t} \\ x_{2,t+1} &= M_{22}x_{2,t} + M_1 x_{1,t} \\ x_{3,t+1} &= M_{33}x_{3,t} + M_2 x_{2,t} \end{aligned} \qquad (1)$$

where, $x_{i,t}$, are the numbers of cells in state i at time point t. The sum of cell numbers in each state at a particular time gives the total number of cells at that time. We set the time interval between two Markov jumps at 1 hour and simulate the model for a total of 72 hours.

These equations are subject to the constraints that the proportion of cells within a state must add to 1. Therefore-

$$M_{ii} + M_i = 1 \qquad (2)$$

Or $$M_{ii} = 1 - M_i \qquad (3)$$

Incorporating this in the above equation enables us to represent the system in terms of M_1, M_2 and M_3

$$\begin{aligned} x_{1,t+1} &= (1 - M_1)x_{1,t} + 2 \times M_3 x_{3,t} \\ x_{2,t+1} &= (1 - M_2)x_{2,t} + M_1 x_{1,t} \\ x_{3,t+1} &= (1 - M_3)x_{3,t} + M_2 x_{2,t} \end{aligned} \qquad (4)$$

In short this may be represented as follows-

$$\begin{aligned} & x_{i,t+1} = (1 - M_i)x_{i,t} + f \times M_j x_{j,t} \\ & if\ i = 1 : f = 2, j = 3 \\ & if\ i = 2,3 : f = 1, j = i - 1 \end{aligned} \qquad (5)$$

We estimated the unknown transition rate parameters based on experimental data for cell doubling times of 30 and 24 hours and cell state ratios of 64:19:17 and 58:23:19 for U87 and U251 cells respectively (data references in Results). We used `fmincon` in MATLAB with a least squares formulation giving values of M_1, M_2 and M_3 as 0.05 hr^{-1}, 0.14 hr^{-1}, 0.14 hr^{-1} for U87 cells and 0.075 hr^{-1}, 0.16 hr^{-1}, 0.16 hr^{-1} for U251 cells. We repeated this estimation 5 times with random initial guesses, each converging to the same values, demonstrating uniqueness of the estimates.

Single Drug Dose Response Modeling

Each of the drugs- PD0325901 (MEK1/2 inhibitor), Abemaciclib (CDK4/6 inhibitor), TAK960 (PLK1 inhibitor) were modeled as having an inhibitory effect on the

respective Markov parameters M_1, M_2 and M_3. We used a sigmoidal hill-type function to describe the effect of drug concentration on transition rates as follows

$$M_{pi,d} = M_i \times (1 - \frac{(D/EC_{50})^n}{1+(D/EC_{50})^n}) \qquad (6)$$

Where, $M_{pi,d}$ is the perturbed/inhibited Markov parameter M_i after inhibition with the respective drug of dose D. The remaining parameters in the equation- half maximal 'EC_{50}' and hill coefficient *'n'* were initially taken as unknowns and were estimated by fitting the respective drug dose response for each drug. We used `fmincon` in MATLAB to obtain the set of parameters that minimized the sum of squares error relative to the data.

This fit model could explain the drug dose responses reasonably well. However, in some cases, the model fits at higher doses of the drugs were higher than the experimentally obtained cell numbers. The experimentally observed cell numbers at some high doses were lower than the initial number of cells loaded, indicating some cell death. Therefore, an additional parameter $M_{\varphi i,d}$ was introduced for each drug, with a value of zero for no drug and a higher value at greater doses of the respective drug

$$M_{\Phi i,d} = \frac{E_{\max,\Phi}(D/EC_{50,\Phi})}{1+(D/EC_{50,\Phi})} \qquad (7)$$

where, $E_{max,\varphi}$ is the maximum cell death possible by a high dose of the drug and $EC_{50,\varphi}$ is the half maximal drug dose related to cell death.

Eq.5 then changes as follows.

$$x_{i,t+1} = (1 - M_{pi,d} - M_{\Phi i,d})x_{i,t} + f \times M_{pj,d} x_{j,t}$$
$$if\ i = 1: f = 2, j = 3$$
$$if\ i = 2,3: f = 1, j = i - 1 \quad (8)$$

The respective parameters obtained to fit the single dose response parameters for U87 and U251 cells are as follows-

Table #1

U87 cells	PD0325901	Abemaciclib	TAK-960
M	0.05	0.14	0.14
EC_{50} *(μM)*	0.0769	0.011	0.0074
n	0.609	0.686	2.67
E_{max}	0.086	0.0012	3.9×10^{-6}
$EC_{50,\Phi}$*(μM)*	1.514	0.0274	662.4

Table #4

U251 cells	PD0325901	Abemaciclib	TAK-960
M	0.075	0.16	0.16
EC_{50}*(μM)*	0.268	0.121	0.0024
n	0.565	0.655	1.663
Emax	0.016	0.022	0.018
$EC_{50,\Phi}$*(μM)*	5.7	0.098	0.009

Combination Drug Dose Modeling and Synergy Score

To model combination drug effects for two drugs, drug effects at evaluated doses were simulated together using the above Markov models. In order to quantify combination response in both the model and experimental data, we used a robustexcess over Bliss[77].Consider each row in the combination drug dose response matrix. One of the drug's doses would be constant across the row (Drug#1, Fig S1a) but the dose of the other drug (Drug#2, Fig S1a) increases from left to right. This can be captured by a 4-parameter logistic model[77,102]-

$$y = \frac{E_{\min} + E_{\max}(D/EC_{50})^n}{1+(D/EC_{50})^n} \qquad (9)$$

where, y is the inhibition effect which in this case is 1 minus the relative cell number with respect to the control, E_{min} is the minimal possible inhibition effect , E_{max} is the maximal possible inhibition effect , EC_{50} is the half maximal drug dose and n is the Hill coefficient .

We used `lsqcurvefit` in MATLAB to obtain least-squares estimates for the four-parameters for each of the 8 rows and 8 columns. For a particular drug combination point, the average of the two fitted inhibition values across it's row and column was taken as the final fitted inhibition value (y_{AB}).Consider the fitted inhibition values for particular doses of drug A alone (y_A), drug B alone (y_B) and their combination (y_{AB}). The Bliss independence scores[37,76,77] are calculated by

$$y_{Bliss,AB} = y_A + y_B - y_A y_B$$

and the excess over Bliss (EOB) scores are calculated by

$$EOB_{AB} = y_{AB} - y_{Bliss,AB}$$

These excess over Bliss scores are calculated for all the three drug combination model predictions for both U87 and U251 cell lines. This exercise is also performed for each of the individual triplicates of the corresponding experimental data.

In order to compare the excess over Bliss scores in the model vs. experiments, the drug combinations are divided into four quadrants (Fig S1a) and the average *EOB* scores across each quadrant are plotted for experimental data vs model predictions (Fig 2c,d).

Figure Legends

Figure 1. Modeling temporal cell states and single drug dose responses (a) Graphical abstract. Schematic showing that integration of single drug dose response experiments with temporal cell state network can predict drug combination responses **(b)** Schematic of the temporal cell state network comprising G0/G1, late G1/S and G2/M states and the activity of the drugs PD0325901, Abemaciclib and TAK960 in each state **(c)** Time courses of cell number in the temporal cell state model for U87 and U251 cells starting with 100 cells for 72 hours. Cell proportions at G0/G1, late G1/S and G2/M states are also mapped which remain constant. **(d)** Single drug dose responses for PD0325901, Abemaciclib and TAK960 in U87 and U251 cells at 72 hours. The temporal cell state model can capture the single drug dose responses reasonably well.

Figure 2. Combination dose responses: Model prediction vs experiments (a,b) Predicted and measured combination drug dose responses for Abemaciclib/ PD0325901, Abemaciclib/TAK960 and PD0325901/ TAK960 for U87 cells(a) and U251 cells (b) **(c,d)** Experiment vs model excess over Bliss scores for Abemaciclib/ PD0325901, Abemaciclib/TAK960 and PD0325901/ TAK960 for U87 cells(c) and U251 cells (d). For a pair of drugs(say Drug A/Drug B), the combinations of the lowest 4 doses of both drugs (besides zero doses) are categorized as low/low, combinations of the highest 3 doses of Drug B with the lowest 4 doses of Drug A are categorized as high/low, combinations of the lowest 4 doses of Drug B with the highest 3 doses of Drug A are categorized as low/high, and combinations of the highest 3 doses of both the drugs are categorized as high/high.

FIGURE 5.1

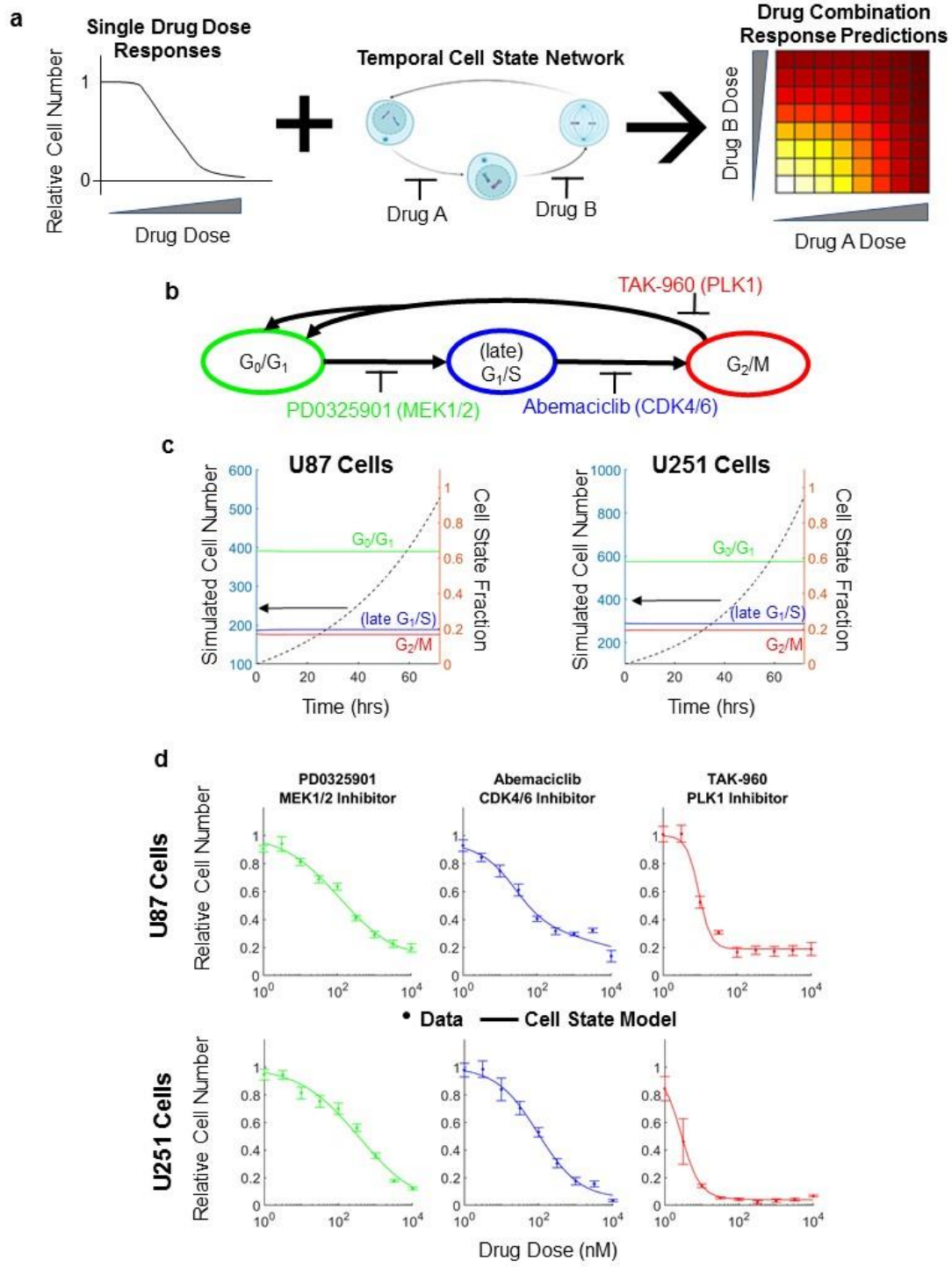
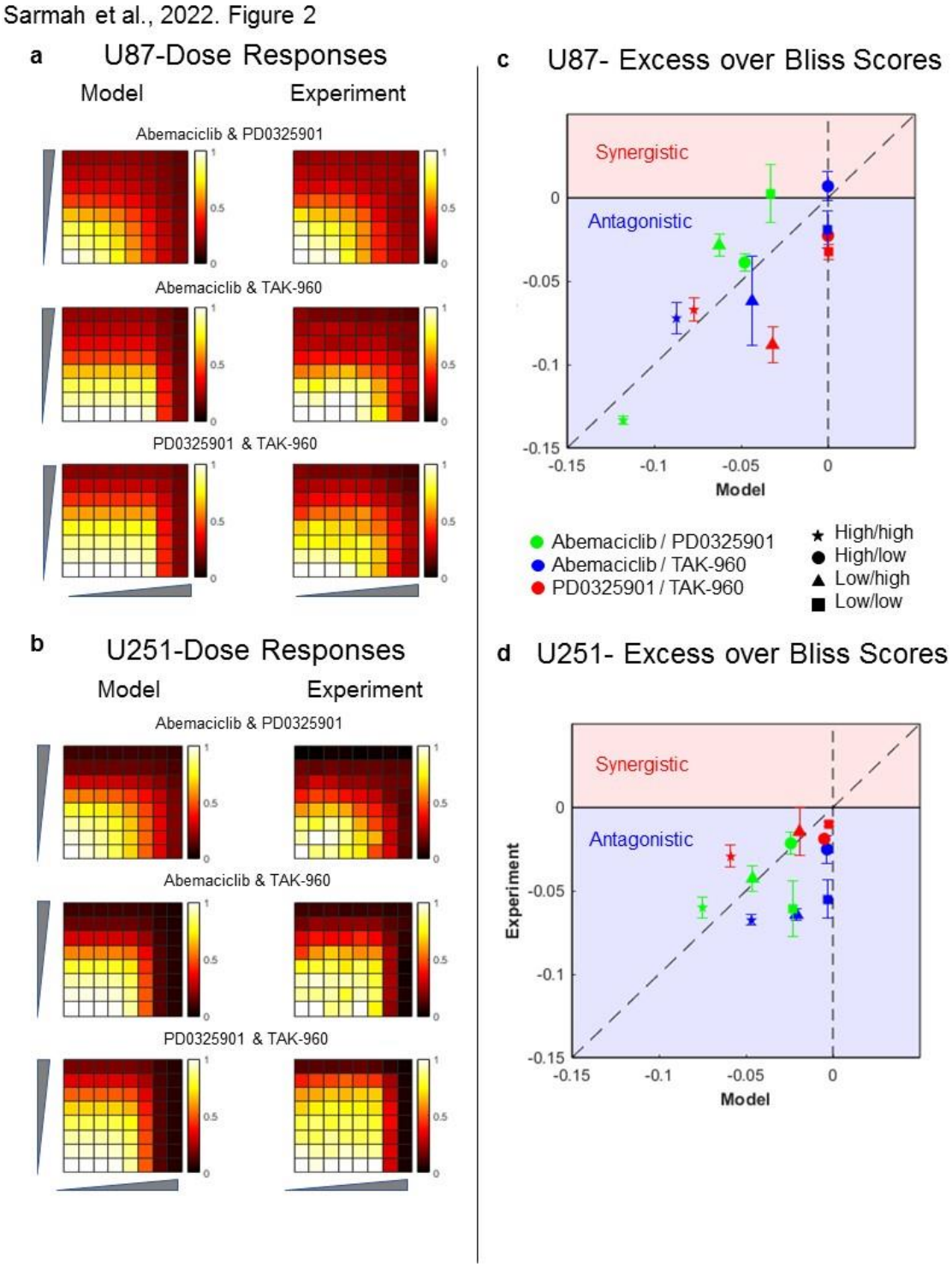

Sarmah et al., 2022. Figure 1
a
Single Drug Dose Responses
Relative Cell Number
Drug Dose
Temporal Cell State Network
Drug A
Drug B
Drug Combination Response Predictions
Drug B Dose
Drug A Dose
b
TAK-960 (PLK1)
G0/G1
(late) G1/S
G2/M
PD0325901 (MEK1/2)
Abemaciclib (CDK4/6)
c
U87 Cells
U251 Cells
Simulated Cell Number
Cell State Fraction
Time (hrs)
G0/G1
(late G1/S)
G2/M
d
PD0325901 MEK1/2 Inhibitor
Abemaciclib CDK4/6 Inhibitor
TAK-960 PLK1 Inhibitor
U87 Cells
U251 Cells
Relative Cell Number
Data
Cell State Model
Drug Dose (nM)

FIGURE 5.2


Sarmah et al., 2022. Figure 2
a
U87-Dose Responses
Model
Experiment
Abemaciclib & PD0325901
Abemaciclib & TAK-960
PD0325901 & TAK-960
b
U251-Dose Responses
Model
Experiment
Abemaciclib & PD0325901
Abemaciclib & TAK-960
PD0325901 & TAK-960
c
U87- Excess over Bliss Scores
Synergistic
Antagonistic
Model
Abemaciclib / PD0325901
Abemaciclib / TAK-960
PD0325901 / TAK-960
High/high
High/low
Low/high
Low/low
d
U251- Excess over Bliss Scores
Synergistic
Antagonistic
Experiment
Model

CHAPTER 6

MESOWESTERN: AN EXPERIMENTAL FRAMEWORK TO MAP BIOCHEMICAL NETWORKS

Western blot is a widely used technique for separation and quantification of proteins based on their molecular weight. Higher throughput versions of western blot especially micro western are cost prohibitive requiring a piezo electric equipment. In concurrence with the other projects in this dissertation, the Birtwistle lab had been working on a relatively affordable approach- the Mesowestern apparatus requiring a 3D-printable gel casting mold. We refer to this system as Mesowestern (v1) in this chapter and it was developed by Cameron Zadeh and Jonah Huggins with contributions from other students. However, an important step for validation of the Mesowestern (v1) remained- to compare the results of Mesowestern (v1) to a regular western. We stepped in to complete this validation and the overall results on Mesowestern (v1) have been published at ACS Omega[1]. The sections containing the background and need for Mesowestern (v1) (Introduction to Mesowestern (v1)) and the comparison of Mesowestern (v1) with regular western in this chapter are adapted from this paper. These two sections were largely written by Dr. Marc Birtwistle with experiments for the second section performed by Deepraj Sarmah, Micah Jordan, Wesley Meredith and Nicholas Harold. In Methods- the sections describing Mesowestern(v1) processes were adapted from protocols written by Cameron Zadeh. Deepraj Sarmah wrote the remainder of the sections dealing with the improvements, methods and results from Mesowestern(v2). A more thorough description of the results and operating procedures for Mesowestern (v1) can be obtained in the ACS Omega paper[1].

However, there still was a need to improve Mesowestern (v1) to make it more conducive for technology transfer and ease of use. We made several improvements in this regard with a new design and the results from this version is referred subsequently in this chapter as Mesowestern (v2). We finally used Mesowestern (v2) to validate and test some protein time courses, following the results in chapters 2 and 5.

Introduction to Mesowestern (v1)

The Western blot has been a staple of molecular biology research for decades since its first description in 1979. [2] It uses vertical immersed tank-based polyacrylamide gel electrophoresis (PAGE) to separate proteins by molecular weight, followed by transfer to a nitrocellulose or poly(vinylidene fluoride) (PVDF) membrane, and finally the application of antibodies to sensitively detect levels of proteins, posttranslational modifications, and even protein complexes. [3–5] Detection modalities include the enzyme-mediated generation of colorimetric molecules or light, or direct conjugation of fluorescent molecules to antibodies, [6,7] which, when combined with carefully designed experiments, can be quantitative. [8–10] Western blotting is still widely ingrained in biomedical research as a protein analytic tool, even perhaps the most used technique in protein-related publications in the last 10 years. [11] In fact, the use of Western blotting, despite falling "out of fashion", seems stable according to publication metrics. [11]

Although Western blot usage remains high, there are notable limitations. Reliance on antibodies for detection is increasingly criticized, [12,13] although the separation of proteins by molecular weight is a strong indicator of antibody validity not typically available to other antibody-based technologies, and Western blotting is often used as a confirmatory

assay to bolster the support generated by other protein assays. Multiplexing is limited to a handful of analytes per gel, which can be increased slightly by stripping antibodies from the membrane and reprobing with new antibodies, [6,14] cutting the membrane into targeted molecular weight range strips for incubation of each with different antibodies,[15,16] or orthogonal detection methods.[17,18] Lastly, traditional Westerns are limited by throughput and sample size; typical gels contain only ~10 wells for the analysis of 10 samples simultaneously, and each sample usually requires ~10 ug of total protein content from cell or tissue lysates. It is this latter limitation of throughput and sample size that we focus on in this paper with the Mesowestern blot.

Before describing the Mesowestern blot, it is instructive to review the myriad of other related protein analytic technologies that address the shortcomings of the Western blot. Reverse-phase protein arrays (RPPAs) use lysates similar to Western blotting but greatly increase multiplexing by spotting lysates on chips so that hundreds of antibodies can be used simultaneously.[19,20] However, lysates are not separated by molecular weight, which causes increased stringency for antibody quality; in fact, antibodies are often validated for use in RPPA by Western blot. Luminex xMAP technology,[21] although technologically distinct from RPPA as it uses barcoded, antibody-conjugated beads, also offers increased multiplexing from cell lysates but does not separate proteins by molecular weight. Enzyme-linked immunosorbent assay (ELISA) has been established even longer than the Western blot and uses two antibodies, one to capture the analyte from a lysate and the other to detect the captured analyte, with detection modalities similar to Western blots.[22,23] Although ELISA does not separate analytes by molecular weight, the use of two different antibodies

for the same target can, in some cases, compensate for specificity issues with one, although obviously the need for two antibodies can be a drawback itself. ELISA enables high-throughput implementation in multiwell plates for the simultaneous analysis of hundreds of samples. Mass spectrometry-based proteomics is antibody-free and can analyze virtually any protein present in a lysate so long as it is ionizable.[24–28] However, specific posttranslational modifications are not always observable.[6] Moreover, findings from mass spectrometry experiments often require orthogonal validation with antibody-based techniques such as a Western blot.[29] Thus, there remains space for a more high-throughput Western blot for analytes that are not amenable to mass spectrometry or when increased specificity is needed for antibodies. Moreover, Western blotting is likely to remain broadly useful as a complementary and confirmatory assay.

There have been advances in Western blotting itself that have improved on the limitations. Single-cell Western blotting using PAGE (Protein Simple) has been developed, greatly reducing sample size requirements.[30,31] Other innovative Protein Simple apparati use capillary electrophoresis rather than a slab gel to allow analysis of up to 96 samples in a single loading (12 simultaneously) in a streamlined manner.[32–34] Digiwest combines Luminex technology with Western blotting by completing electrophoresis and transfer but then cutting the membrane into molecular weight sections to be analyzed by separate spectrally distinct beads.[35] This provides the multiplexing capabilities of Luminex with the molecular weight separation of Western. The Microwestern blot[36–38] uses a piezoelectric pipetting apparatus to spot nL amounts of lysate onto a typical-sized slab gel, followed by semidry horizontal electrophoresis (as opposed to tank-based), and finally, a gasket system

for incubating different parts of the resultant membrane with up to 96 different antibodies. Thus, the Microwestern addresses both throughput and, to some extent, multiplexing limitations. However, the piezoelectric pipetting apparatus is not available to many labs.[36,38] There is not yet a slab gel-based Western technology that addresses throughput and sample size limitations that is micropipette-loadable.

Here, we present the Mesowestern blot that, similar to the Microwestern, allows for high-throughput analysis of hundreds of small samples in a typical-sized slab gel but does not require piezoelectric pipetting because it is micropipette-loadable. To do this, we designed and 3D-printed a gel casting mold that produces a polyacrylamide gel with 336, ~0.5 μL sample wells arranged with 8 rows by 42 columns in a microplate footprint. The main tradeoff is molecular weight resolution because samples have less distance to migrate. However, the format is flexible because the cast is 3D-printed and gel acrylamide % can be adjusted. Proof-of-concept experiments using both infrared-fluorescent molecular weight ladder and cell lysates demonstrate that proteins loaded in Mesowestern gels are amenable to the standard Western blotting steps of gel electrophoresis followed by the transfer to a membrane for imaging. These experiments also show another main tradeoff that sample-to-sample CV is high, making the technique more suited for qualitative screening applications. The main difference from Western blotting is horizontal electrophoresis as opposed to tank-based electrophoresis, and, as mentioned above, reduced molecular weight resolution. Because the gel mold is 3D-printable, users with access to institutional additive manufacturing cores (which are relatively commonplace) have significant design freedom for custom layouts. We expect that the technique could be

easily adopted by any typical cell and molecular biology laboratory already performing Western blots. *(Adapted from ACS Omega paper)*

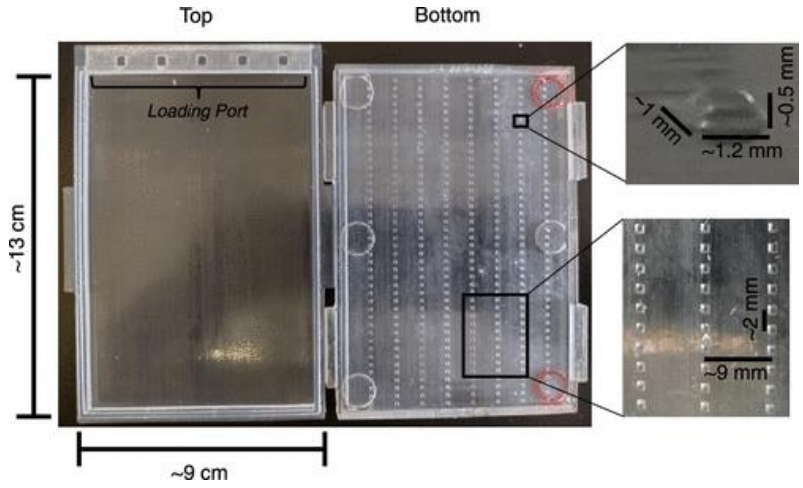


Figure 6.1. Gel casting mold for Mesowestern(v1). The mold consists of two pieces, which we refer to as "top" and "bottom". The top contains the loading port for the unpolymerized gel solution, whereas the bottom contains the raised regions, which become wells in the Mesowestern gel. The gel dimensions are approximately 9 cm by 13 cm in width and length and is 1.2 mm thick. Each well is a rectangle that is 1 mm by 1.2 mm in width and length and is 0.5 mm deep. Wells are spaced 1.8 mm apart and have 8.7 mm to run in their lane before the next well is reached.

Results-

Comparing Mesowestern(v1) with Western

We wanted to investigate sensitivity and linear range for Mesowestern(v1) in direct comparison to Western. While this will invariably be dependent on the epitope of interest, its abundance in the cell lysate, and the antibody being used, we started by investigating this for β-actin. Specifically, we performed a 6-point, 2-fold serial dilution of lysate from exponentially growing U87 cells and replicated this dilution curve 6 times on a portion of a Mesowestern(v1) gel (Figure 6.2A─we used a mold with 1 μL wells for this experiment; replicates in Figure 6.8). Simultaneously, 20 μL of the same lysates were loaded into a

regular Western, and the same blocking and antibody solutions were used to compare Western to Mesowestern(v1) side by side (Figure 6.1 B,C; replicates in Figure 6.2). Bands were observable from each technology in each sample. We had expected the lowest concentration lysate to be below the limit of detection for Mesowestern(v1), but to our surprise, this was not the case. In both cases, the linear range of detection was at least ~33-fold, with R2 ~ 0.99 (Figure 6.1D). The lowest lysate concentration loaded in Mesowestern(v1) is at least equivalent to 29 ng of total protein, which is approximately 100–300 cells (0.1–0.3 ng of protein yield/cell)[39–41] Based on the estimates for absolute expression levels of β-actin in mammalian cells (~106 copies/cell),[42] this is at least ~500 attomol sensitivity, but as mentioned above, the serial dilution curve did not find the lower limit of detection. For regular Western, a 15 attomol limit of detection using the same infrared fluorescence modality has been reported for transferrin (no such data could be found for β-actin).[43] More broadly, the manufacturers report limits of detection for various modalities ranging from 500 fg to 500 pg of protein, which is consistent with ~10 attomol for proteins of typical molecular weight ranges.[44]*(Adapted from ACS Omega paper).*

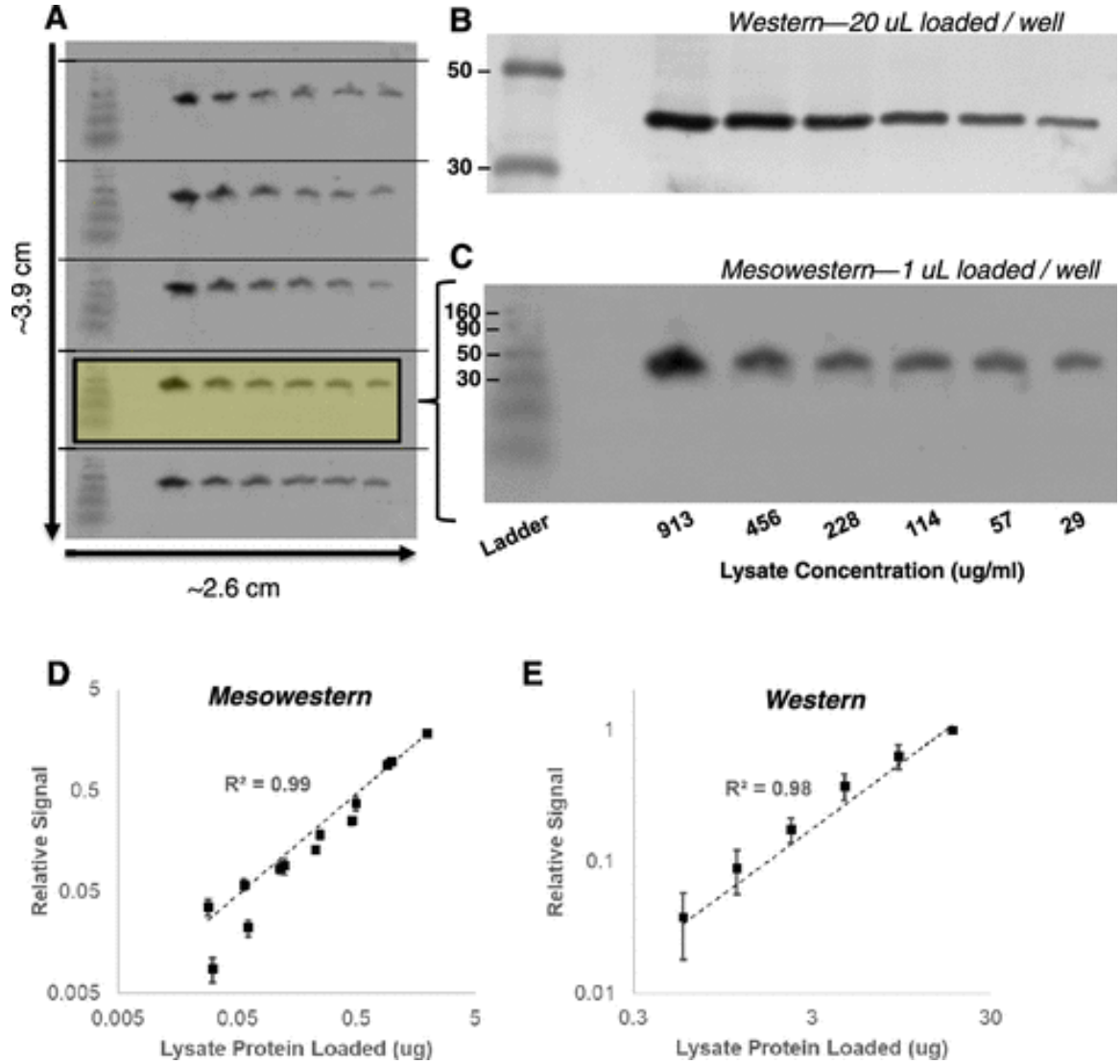


Figure 6.2. Comparison of Mesowestern and regular Western. A–C. Cell lysate from exponentially growing U87 cells was prepared at a range of protein concentrations (twofold serial dilution) and subjected to Mesowestern and Western analyses, as indicated. Full scans are shown in Figure 6.8. The PVDF membrane was incubated with anti-β-actin antibodies and a secondary antibody for detection. D, E. The signal derived from the image analysis of each band (N = 78 bands for Mesowestern from 2 blots and 14 dilution curves; N = 18 bands for Western from 3 blots/dilution curves) was plotted versus the known amount of total protein mass loaded. For Mesowestern (D), data were normalized such that 1 mg/mL lysate corresponds to a relative signal intensity of 1. For Western, data were normalized such that the maximum normalized signal was 1. Error bars are the standard error of the mean for each sample ($N \geq 3$).

Shortcomings of Mesowestern (v1)

Although Mesowestern (v1) could handle low amounts and dilutions of liquids, it had a few shortcomings. Broadly, it came down to two key factors, both related to ease of use-

(i) Need for a more ordered approach- Mesowestern (v1) requires multiple steps where the gel is relocated- taken off the cast, placed on a thick blotting paper in a proper alignment with the flow of current, and into the transfer cassette after electrophoresis (see Methods). An alternate method which can minimize moving the gel would be more user friendly.

Besides, Mesowestern (v1) also requires electrophoresis to be performed with the gel placed on a buffer-soaked blotter paper. This is a somewhat amateurish technique, with potential problems in technology transfer and most contemporary electrophoresis for either DNA or proteins are rightly performed in a buffer filled tank.

(ii) Option for automated pipetting option- Since high throughput is one of the chief goals of Mesowestern, there needs to be an option, where the pipetting of hundreds of samples can be automated. It is difficult to achieve that with Mesowestern (v1), since the gel is not in a fixed cast suitable for loading in such an automated setup. The requirement for the gel to be moved multiple times can also change the location and dimensions of the wells, which need to be known more precisely in order to create automated loading protocols, or even manually loaded with a multichannel pipet.

In order to address these concerns, we attempted a vertical setup, which had some shortcomings of its own and finally came up with a horizontal tank setup- Mesowestern(v2). This addresses some of the above concerns. The steps we followed, and the results obtained are in the following sections.

Attempt at a Vertical Mesowestern

In order to address some of the issues with Mesowestern(v1), particularly related to having a more ordered and conventional procedure, we attempted a vertical design. This was in part inspired by the fact that most regular westerns have a vertical design and a Mesowestern setup with a similar electrophoresis procedure would be relatively easy to understand and use. It comprised of a gel cast and electrophoresis setup designed for that purpose (Figure 6.3). In brief, the procedure involved the gel to be cast between the two plastic casts (see Methods for casting gel), the half of the cast with the well protrusions removed(with the gel still adhered to the other half of the cast), the samples loaded into the wells, the cast with the loaded gel vertically attached to a electrophoresis setup within a Mini Trans-Blot Cell (BIO-RAD, 1703930) and the electrophoresis run.

However, our results with the vertical Mesowestern were largely unsuccessful for the following reasons-

i. Buffer leakage into the gel area during electrophoresis- Unlike a regular western, where the gel, once cast, remains within the glass/plastic casts till the end of electrophoresis, in the vertical Mesowestern, the half of the cast with the

well protrusions, has to be removed and after the gel has been loaded with samples-reattached to the electrophoresis setup. This often causes an uneven seal and the buffer leaks through the Mesowestern gel space. It requires constant monitoring, and some buffer added every few minutes. If the buffer level gets too low, the current would no longer flows through the setup and electrophoresis would stop.

ii. Sample retention issues- This was due to two reasons. First, when the gel gets loaded with samples and reattached to the electrophoresis setup, the contact with the new surface may pull samples out of the well due to surface tension effects. Secondly, the buffer leakage into the gel area also tends to wash some samples away from the wells.

iii. Uneven electrophoresis across the gel- This may be caused by an uneven seal of the gel with the electrophoresis setup. Current may be uneven across the gel and can cause different run times for the same sample (shown in Figure 6.3b).

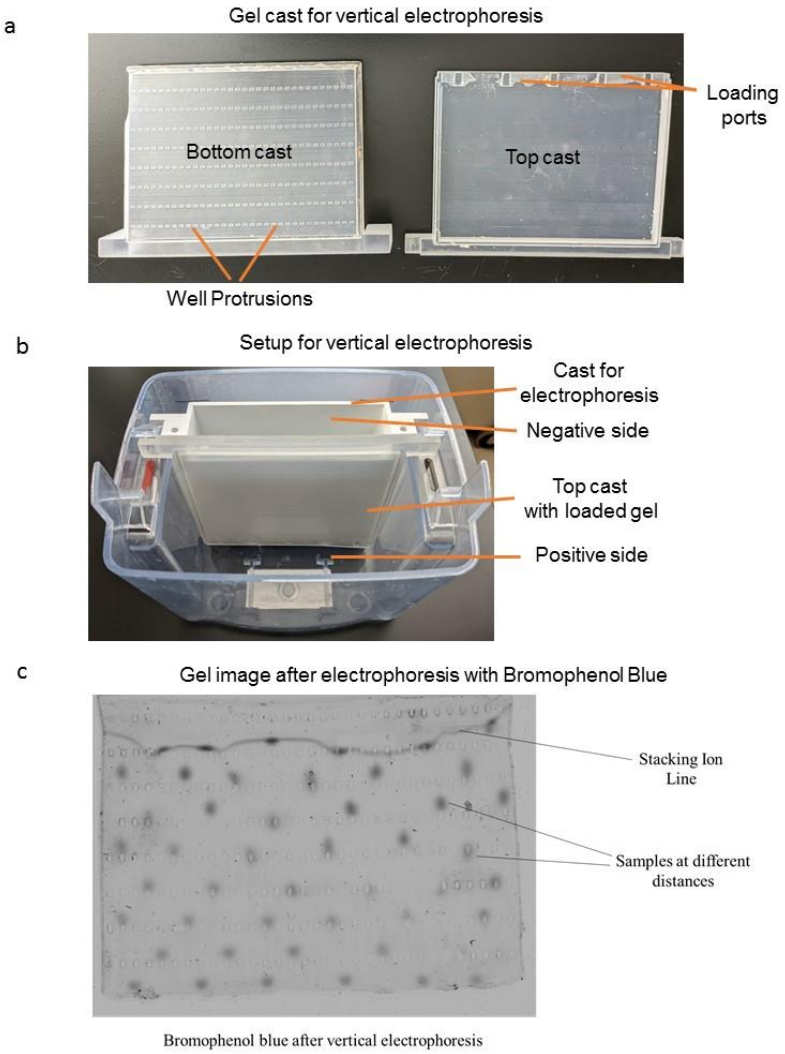


Figure 6.3 Vertical Mesowestern Setup and Results (A) Top and bottom gel casts for a vertical Mesowestern. (B) Top cast with loaded gel attached to the electrophoresis setup within a Mini Trans-Blot Cell (BIO-RAD, 1703930 (C) Results with a electrophoresis run with bromophenol blue.

Hortizontal Tank Design-Mesowestern(v2)

Since most of the shortcomings in vertical electrophoresis have to do with problems in adjusting the gel with sample filled wells to a new surface, we decided to do away with

designs that require an additional surface over the wells. This could be enabled by going back to a horizontal electrophoresis design, where the samples would remain in the wells, due to gravity. However, to make it more accessible to users compared to Mesowestern(v1), we tried a tank electrophoresis design. The bottom cast was also designed to be fit in an automated pipettor (Opentrons V2). This design, which we call Mesowestern(v2) in this chapter, has parallels to the design of a DNA agarose gel electrophoresis, although the wells in Mesowestern(v2) are relatively smaller and more numerous. The images of a cast and tank setup are shown in Figure 6.4 although the exact dimensions may be proprietary.

We tried two kinds of gels with Mesowestern(v2)-A 9.5% Tris-Acetate gel and a 10% Tris HCl gel. We noted that if the gel is fully immersed in the electrophoresis buffer after being loaded with samples, there some samples may be washed off by the buffer. This can be due to the wells not being sufficiently deep, as well as the sample density not being high enough to remain sunk in the wells. In our experiments with Mesowestern(v2), we tried a semi-submerged procedure, where the buffer is just in contact with the gel sides but does not go over the gel. This may be a fair design in a lab setting for a few experiments but not necessarily transferable. Therefore, we expect subsequent designs of the Mesowestern(v2) to have deeper wells and higher sample density by incorporating more glycerol in the sample buffer, so that it would be conducive to run in a fully immersed setting. Potentially, designs where the gel can be loaded while being submerged (as is done for a DNA agarose gel) either manually, with a single channel pipette, multichannel or through an automated device should also be contemplated.

We also noticed in initial experiments that if electrophoresis is run for too long in a Mesowestern(v2) setup (>45 minutes), the gel tends to come off the top cast during electrophoresis. This is more of a problem in the Tris-Acetate gel, where the gel tends to swell as well. For subsequent experiments with Mesowestern(v2) in this chapter, we used a Tris HCl gel-generally a 10% gel, and occasionally an 8% gel when resolution of higher molecular weights was required. Electrophoresis time was generally about 35 minutes and was stopped when the bromophenol blue dye in the sample buffer reached close to the next set of wells.

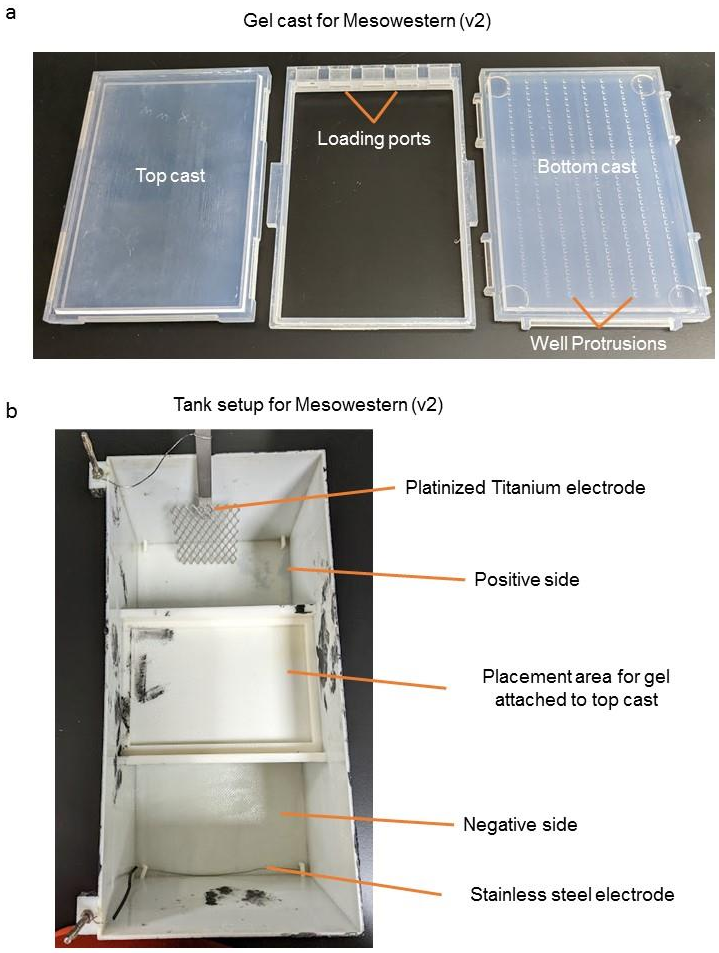


Figure 6.4 Mesowestern (v2) Setup and Results (A) Top, middle and bottom gel casts for a Mesowestern(v2). (B) Tank setup to run Mesowestern(v2)

Arriving at a suitable buffer chemistry for Mesowestern(v2)

When we used a Tris-Glycine buffer with a Tris HCl gel, as is done in the electrophoresis of many regular western blots, we noticed a gradually moving line in the gel, where electrophoresis speeds were different before and after the line. This can also be seen in the attempt with a vertical Mesowestern (Figure 6.3C). This is not desirable for uniformity of

electrophoresis across the gel. We reasoned that this is caused due to depletion of Cl^- ions in the gel and glycine taking over as the dominant ion, as happens at the stacking line of a regular western[45]. In order to ensure ion uniformity throughout electrophoresis, we switched to a Tris HCl based electrophoresis buffer instead of Tris Glycine (see Methods). We no longer noticed the "stacking line", but this change came with a minor drawback that higher concentrations, Cl^- can be potentially corrosive to the anode (positive electrode) during electrophoresis. To ensure protection to the electrode, we switched to a relatively inert (platinized titanium) electrode for the anode (Figure 6.4). The Cl^- concentration is also kept relatively low as the buffer consists of 0.1 M of Tris HCl pH 7.5. More than 50 electrophoresis runs have been performed since these changes without any noticeable corrosion to the electrode.

Increasing Pipetting Options

Conventionally pipetting for western blot is done with hand, generally with a pipette with a maximum range ~20 ul P20. We can do the same for meso western(v2) with a P2 pipette with maximum range ~2 ul. This version of meso western is not totally compatible with multichannel pipetting although future versions might be (wells in alignment with 96 well plate design). However, the horizontal setup does allow us more options for pipetting. The cast was made to fit into the loading site of an Opentrons V2 automated robot pipetting device. Since the gel is supposed to remain stuck to one side of the cast during loading and electrophoresis, the dimensional locations of the wells are known in advance. We can use this property to automate a pipetting robot to load a unique sample or multiple different samples.

Although this allowed us to pipette many samples, it was unsuccessful on occasion -the Opentrons pipette missed the wells and this needed manual intervention. This may be attributed to variability in the robot and well positions in the gel. Future designs should allow for more variability in the robot pipetter operation as well as gel position. A design with wider and longer wells should help reduce such loading errors.

Ladder Results with Mesowestern (v2)

The next logical step for Mesowestern(v2) was to test whether it could resolve different molecular weights and the range of the resolution. For a 10% Tris HCl gel we loaded the gel with ladder (Chameleon Duo, Licor #928-60000, diluted 1:10) and ran electrophoresis for all 8 rows and 14 adjacent columns for ~35 minutes (see Methods) and transferred to a nitrocellulose membrane. The results are shown in Figure 6.5. We could observe resolved bands from between 8 to 70 kDa (Figure 6.5). Qualitatively, the results looked quite uniform across the membrane. This was also later seen in ladder and protein runs (Figure 6.6). We reasoned that a lower percentage gel and longer runs would enable us to see higher molecular weight resolutions. In a later experiment, (pRb/β-actin, Figure 6.6) we ran an 8% gel for a longer time (~45 mins) and bands between 15-90 kDa were able to be resolved and protein bands of 110 kDa (pRb) were also observed.

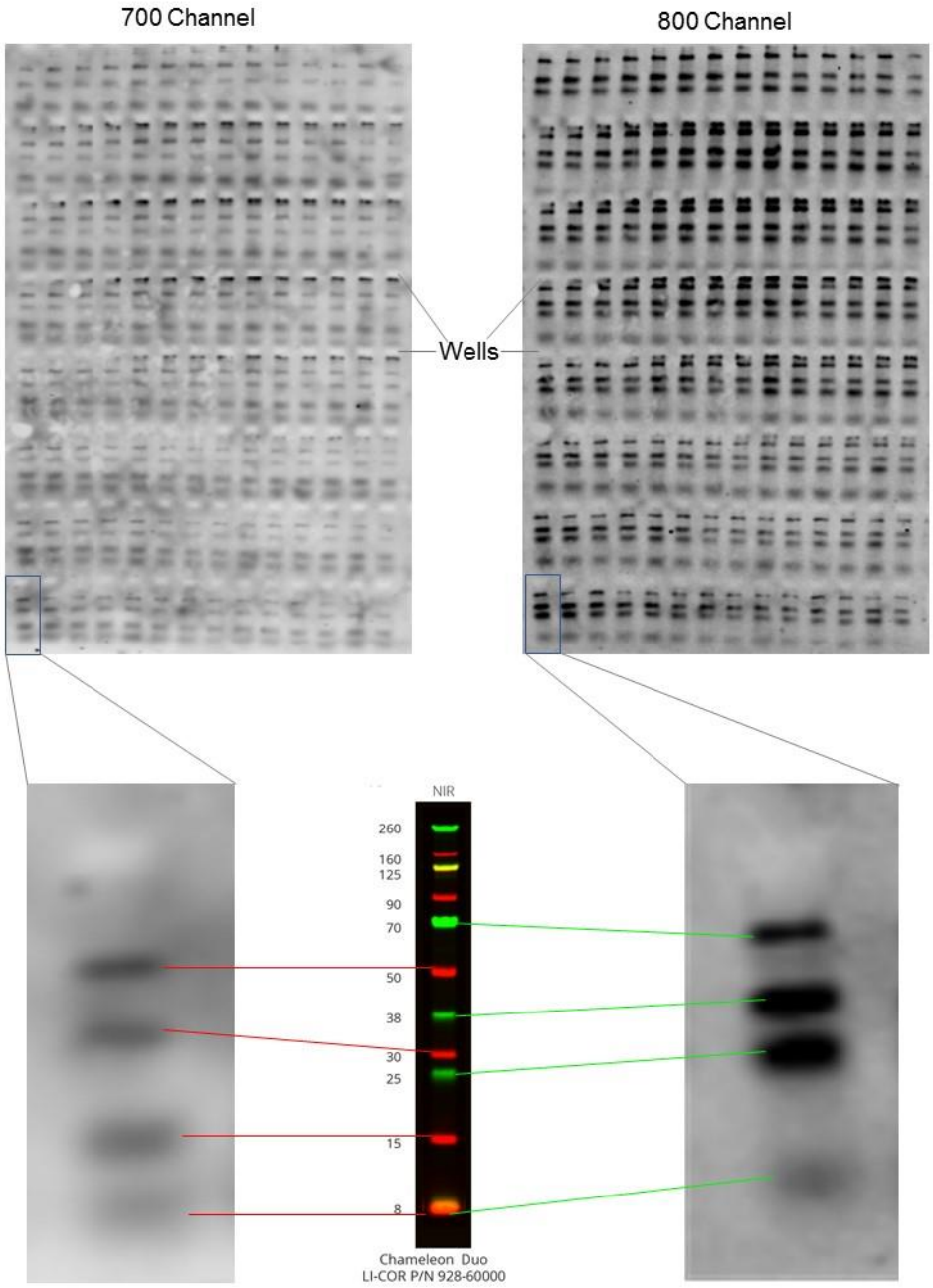


Figure 6.5 Resolution of ladder bands in the 700 and 800 channel and comparison with reference molecular weight ladder

Protein validation with Mesowestern(v2)

For the kinases involved in cell states (MEK1/2, CDK4/6 and PLK1), discussed in chapter 5, we wanted to see if we could perform western blot experiments to obtain their network structure. For the high number of samples for this experiment, Mesowestern could be useful, because of its high throughput. Downstream markers, which can represent the activity of each protein are essential. And lastly, in order to uncover network edges (chapter 2), one perturbation per node is essential. We wanted to test potential downstream markers for the respective proteins and if validated use them subsequently in time course data(chapter 2).

For the perturbation of MEK1/2, CDK4/6 and PLK1, we decided to use a relatively high dose of their respective inhibitors PD0325901, Abemaciclib and TAK-960 corresponding to IC75 dose for U87 (2500 nM for PD0325901, Abemaciclib, 25 nM for TAK-960). Validation would be if we see a conspicuous band of the marker at the DMSO treatment and a significantly reduced band for the respective drug treatment.

pMAPK (Cell Signaling, #4370S) has been known to be a marker for MEK1/2 and has been used in the Birtwistle lab in the past for this purpose[38]. We reasoned that a phosphorylated form of the protein retinoblastoma, Rb (Cell Signaling, #D59B7) would be a suitable marker for CDK4/6 and this is informed by literature[46] as well as our understanding of canonical cell pathways. However, literature was sparse on downstream

targets of PLK. We considered TCTP[47] (Cell Signaling, #5251) phosphorylated at Ser46 as the most likely marker for PLK1 and included it for validation.

We could observe the bands for pMAPK and pRb at DMSO treatment and inhibition at the respective dose of the drug (Figure 6.6). We could also see the bands albeit faint for pTCTP at DMSO treatment. However, there was no significant inhibition for treatment with TAK-960. Therefore, we did not consider it a suitable marker for subsequent analysis

We subsequently tested a few more possible downstream markers for PLK1-CDC25c(Cell Signaling, #5251, no noticeable bands), Cyclin E (Cell Signaling, #D7T3U, no inhibition)and MLF1 (Rockland #600-401-A92, no noticeable bands) but without success.

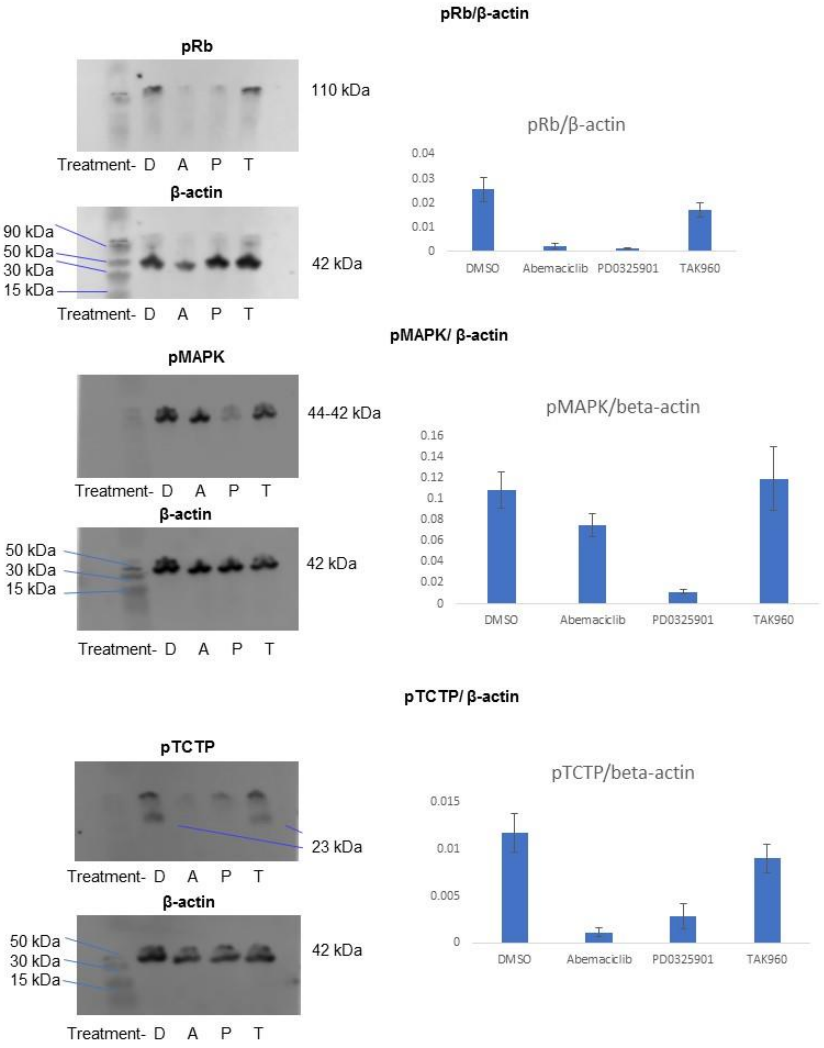


Figure 6.6 Protein Marker Validation Results for pRb, pMAPK and pTCTP. Results on the right are normalized by respective β-actin signals.

Protein quantification with Mesowestern (v2)

We ran time course experiments (see Methods) at 0,0.5,2,9,24,48 and 72 hours for each of the treatments(PD0325901-Perturb MEK1/2, Abemaciclib-Perturb CDK4/6 and TAK-960-Perturb PLK1) and against the two antibodies validated for MEK1/2 and CDK4/6. The results (normalized by β-actin and then by respective mean DMSO treatment signals) are shown in the Figure 6.7 and the full Mesowestern blots are shown in Figures 6.10 and 6.11.

Although we did not have enough data to construct a 3-node network, we can note that when MEK1/2 is inhibited, CDK4/6 levels go down as well-as expected by canonical pathways (chapter 5). When CDK4/6 is inhibited, MEK1/2 levels seem to rise in both U87 and U251 cells indicating a negative feedback from CDK4/6 to MEK1/2.

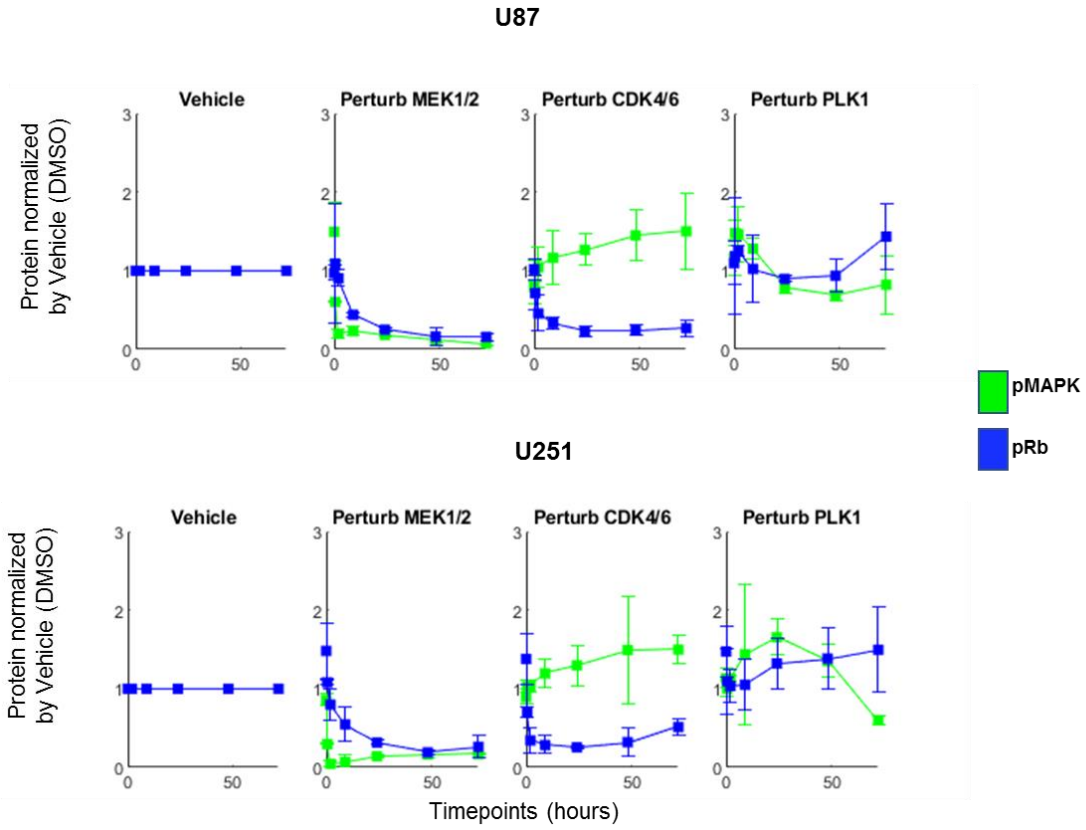


Figure 6.7 Normalized time courses of pMAPK and pRb with vehicle control, MEK1/2 inhibition, CDK4/6 inhibition and PLK1 inhibition, for U87 and U251 cells respectively

Methods-

Printing the Mold

Molds were printed in the Clemson Additive Manufacturing Lab with the Stratasys Connex 350 and Veroclear as the material (Stratasys, OBJ-03271-RGD810). Following printing, a self-forming valve packing (Danco, #80794) was inlayed into the outer edge of the well perforation unit (bottom).

Casting a Gel

Gel casting was completed through a process of silanization of surfaces coming into contact with the gel, clamping to ensure a tight leak-proof fit, and serological pipetting

of unpolymerized solution into the mold. Briefly, a 2.5% v/v silane solution was prepared by combining 1.25 mL of dichlorodimethylsilane (Sigma-Aldrich, #40140) and 48.75 mL of 100% ethanol (Fisher, #04-355-22) in a 50 mL conical tube (Fisher, 14-432-22). We then applied 250 μL of the silane solution to the interior surfaces of both the top and bottom mold pieces, gently spread it across the surface by rocking, and wicked excess with a kimwipe. After assembling the top and bottom pieces together (and middle piece for Mesowestern(v2)), four C-clamps (Irwin #1901235) were tightened onto the assembly at the designated corner locations (indented circles). At this point, the assembly is ready for loading.

Tris Acetate gel for Mesowestern (v1)-A 9.5% gel solution was prepared by combining 47.5 mL of 30% bis/acrylamide solution 29:1 (BIO-RAD, #161-0156) with 41 mL of MilliQ water, 30 mL of glycerol (Sigma-Aldrich, #G5516-500 mL), 30 mL of 5× tris-acetate buffer (recipe as follows), and 1.5 mL of 10% sodium dodecyl sulfate (SDS) (Fisher, #BP 2436) together. Preparation of the 5× tris-acetate buffer was completed by dissolving 145.4g of tris base (BIO-RAD, 161-0719) in 700 mL of MilliQ water (pH expected between 11.0 and 11.4). The pH was adjusted by adding 65 mL of glacial acetic acid (Sigma-Aldrich, #320099) and allowing the solution to sit overnight. Then, 0.5 mL of glacial acetic acid was pipetted into the solution and allowed to sit for an hour at room temperature. This was repeated until the solution reached pH 6.9. Finally, the volume of the solution was brought up to 1 L with MilliQ water and stored at 4 °C. Polymerizing gel solution was made by combining 15 mL of 9.5% gel solution with 133 μL of 10% ammonium persulfate solution (APS) and 13.3 μL of TEMED (BIO-RAD #161-0700) into

a beaker under a fume hood. The 10% APS solution was prepared by dissolving 0.2 g of ammonium persulfate (BIO-RAD #161-0800) into 2 mL of MilliQ water. Quickly after preparation, 15 mL of gel solution was dispensed by serological pipetting into the mold assembly via the loading port. The assembly was kept still under the fume hood for 30 min at room temperature to achieve full polymerization.

To remove the gel from the mold, first the C-clamps were removed. Then, the top and bottom mold pieces were carefully separated using a gel releaser (BIO-RAD, #165330) on the lateral protrusions, followed by carefully moving the releaser around the internal face of the top. After splitting the top and bottom pieces, the gel was removed by inverting the mold so that the gel is facing thick blotter paper (BIO-RAD #1703958) that is presoaked in running buffer. The blotter paper was approximately 5 cm larger than the gel on the top and bottom and about 1 cm larger than the gel on each side. The gel is slowly peeled away from a corner using the gel releaser until gravity facilitates the remaining gel to gently fall onto the soaked blotter paper support. The gel can be used immediately or be stored in a sealed bag at 4 °C for several months (at least).

Tris HCl gel For Mesowestern(v2)- A 10% gel solution was prepared by combining 50 mL of 30% bis/acrylamide solution 29:1 (BIO-RAD, #161-0156) with 60.3 mL of MilliQ water, 37.5 mL of 1.5 M Tris HCl buffer (recipe as follows), and 1.5 mL of 10% sodium dodecyl sulfate (SDS) (Fisher, #BP 2436) together. . Preparation of the 1.5 M Tris HCl buffer was completed by dissolving 18.15 g of tris base (BIO-RAD, 161-0719) in 80 mL of MilliQ water (pH expected between 11.0 and 11.4). The pH was adjusted by a few drops of hydrochloric acid (Sigma-Aldrich, #258148) until the solution reached pH 8.8. Finally,

the volume of the solution was brought up to 100 mL with MilliQ water and stored at 4 °C. Polymerizing gel solution was made by combining 25 mL of 10% gel solution with 125 µL of 10% ammonium persulfate solution (APS) and 125 µL of TEMED (BIO-RAD #161-0700) into a beaker under a fume hood. The 10% APS solution was prepared by dissolving 0.2 g of ammonium persulfate (BIO-RAD #161-0800) into 2 mL of MilliQ water. Quickly after preparation, 15 mL of gel solution was dispensed by serological pipetting into the mold assembly via the loading port. The assembly was kept still under the fume hood for about an hour at room temperature to achieve full polymerization.

To remove the gel from the mold, first the C-clamps were removed. Then, the top piece with the well protrusions were carefully separated using a gel releaser (BIO-RAD, #165330) on the lateral protrusions, followed by carefully moving the releaser around the internal face of the top. The gel can be loaded with the samples either manually or in an automated system (Opentrons v2).

Cell Culture

MCF10A cells (from LINCS Consortium and STR verified internally) were cultured in DMEM/F12 (Gibco #11330032) medium containing 5% (v/v) of horse serum (Gibco #16050122), 20 ng/mL of EGF (PeproTech #AF-100-15), 0.5 mg/mL of hydrocortisone (Sigma-Aldrich #H-0888), 10 µg/mL of insulin (Sigma-Aldrich #I-1882), 100 ng/mL of cholera toxin (Sigma-Aldrich #C-8052), and 2 mM of L-glutamine (Corning #25-005-CI). U87 and U251 cells (STR verified internally) were cultured in DMEM (Gibco#10313021) medium containing 10% (v/v) fetal bovine serum (Corning#35-011-CV) and L-glutamine (2 mM) (Corning #25-005-CI). The cells are kept at 37 °C in 5%

CO_2 in a humidified incubator. To maintain subconfluency, the cells are passaged every 2–3 days, washing once with phosphate-buffered saline (PBS), lifting with 0.25% trypsin (Corning #25-053-CI), and reseeding in full growth media.

Lysate and Sample Preparation

Cells growing in full growth media were collected, counted, and seeded (150,000 cells/well for MCF10a, 50000 for U87 and U251) in tissue culture-treated 6-well plates (Corning # 08-772-1B). MCF10a cells were kept at 37 °C in 5% CO_2 in a humidified incubator for ~48 h, while U87 and U251 were treated with respective drugs after seeding overnight (Abemaciclib-2500 nM for U87, 500 nM for U251, PD0325901-2500 nM for both cell lines and TAK960-25 nM for both cell lines) and samples were collected at time points of 0,0.5,2,9,24,48 and 72 hours. The plates were removed from the incubator, and media in the wells was aspirated. The wells were washed with ice-cold PBS once and placed on ice. Freshly prepared, ice-cold RIPA buffer (110 μL, 50 mM tris, pH 7–8 (Acros Organics #14050-0010), 150 mM NaCl (Fluka #71383), 0.1% SDS (v/v from 10% stock, Fisher #46040CI), 0.5% sodium deoxycholate (g/mL Alfa Aesar, J62288), 1% Triton-X-100 (v/v, Fisher, BP151) with protease and phosphatase inhibitors (1 μg/mL aprotinin (MP Biomedicals #0219115801), 1 μg/mL leupeptin (MP Biochemicals #0215155301), 1 μg/mL pepstatin A (MP Biochemicals #0219536801), 10 mM β-glycerophosphate (Santa Cruz Biotechnology #sc203323), and 1 mM sodium orthovanadate (Sigma-Aldrich # S6508)) were added into each well. The plates were kept on a rocker (slow) in the cold room for 15 min. Then, the lysates were scraped off with a cell scraper (Stellar Scientific TC-CS-25), and 100 μL lysate from each well was transferred into labeled Eppendorf tubes

on ice. Each tube was vortexed three times for ~30 s to homogenize cell debris. Next, the tubes were centrifuged at 4 °C for 15 min at ~21,000*g* (max speed). Finally, 80 μL of the supernatant from each tube was transferred into a new Eppendorf tube, being careful not to disturb the debris pellet. Lysates were stored at −20 °C for short-term storage and transferred to −80 °C for long-term storage.

Protein Quantification

Total protein quantification of lysates was done using either the BCA-Pierce 660 Assay (Thermo Scientific #23225) or Pierce Rapid Gold BCA (Thermo Scientific A53225), and BSA stock (Thermo Scientific #23209) was used as a reference according to the manufacturer's protocol. In short, 10 μL of the lysate sample or BSA standards were loaded into 96-well plates (Corning #3370), in triplicate. Then, the BCA Protein Assay Reagent was loaded into each nonempty well. The plate was covered with the lid and incubated at room temperature for 5 min. The absorbance readings at 660 or 480 nm were obtained in a plate reader (BioTek #Epoch2). The average of blank wells was subtracted from each reading to calculate blank-corrected averages for each condition. The standard curve is fitted to a line using blank-corrected mean values of each standard condition versus its BSA concentration. The protein concentration in each sample was calculated using the standard curve formula.

Sample Preparation

Lysate stocks are thawed on ice (if applicable). Then, a 5× sample buffer was prepared (5 mL of glycerol (Sigma-Aldrich #G5516), 0.5 mL of 10% SDS (Fisher #BP 2436), 0.01 g of bromophenol blue (Calbiochem #2830), 2.1 mL of 5× tris-acetate buffer

(for Mesowestern v1)/ 2.1 mL of 1.5 M Tris HCl buffer pH 8.8 (for Mesowestern (v2)), 0.5 mL of β-mercaptoethanol (Sigma-Aldrich #M6250), then the total volume brought to 10 mL with MilliQ water). This was mixed with lysates in a 1:4 (v/v) ratio. Next, the tubes were heated at 95 °C for 5 min in a dry heating block and then briefly spun in benchtop microcentrifuge before loading (below).

Loading the Gel for Mesowestern(v1)

Following the release of the gel onto the soaked blotter paper, the assembly was placed down on a flat surface with the wells facing up. If folds and stretching of the gel are evident, light rolling was used to flatten. A p2 micropipette with 10 μL tips was used to load the prepared lysates and/or molecular weight ladder (LI-COR, 928-60000) into wells as desired. We have found that wells less than 2 mm away from the gel boundaries may be subject to inconsistent electrophoresis and transfer and therefore avoid them when possible. Care was taken not to adjust the gel on the blotter paper after any loading and also to transport the gel with a spatula support underneath.

Loading the Gel for Mesowestern (v2)

We automated an Opentrons (OT-2) robot to run several experiments using diluted ladder across the gel or different protein samples. At first a code was written in Opentrons which specified the dimensions and locations of the wells (with the help of the REDDI Lab). Secondly, we specify within the Opentrons loading area, the locations of the gel, samples, pipettes, etc. We can then design a loading protocol as a json file (https://designer.opentrons.com/) which specifies the samples and the wells into which they are supposed to be added. There is another option in Opentrons which lets us create

the protocol as a python code, but we did not do so because we could design all our loading protocols as json files in the protocol designer. For more complex protocols, the protocol may be designed as a python code.

Horizontal Electrophoresis in Mesowestern (v1)

Horizontal electrophoresis was carried out using the Flatbed Professional (Gel Company Store, FC-EDCProf-2836). The apparatus was maintained at 10 °C during electrophoresis. First, ∼10 mL of cooled running buffer was poured onto the center of the apparatus, followed by transfer of the blotter paper/loaded gel by spatula onto this buffer. Running buffer was made by combining 20 mL of 5× tris-acetate buffer (see above) with 29.5 mL of MilliQ water and 0.5 mL of 10% SDS. The gel should be oriented to have the red bar at the bottom, where the proteins will migrate toward. Additionally, the wells should be aligned with the apparatus gridlines, and excess running buffer should be wiped up with no buffer accumulated outside of the blotter paper. Then, the anode and cathode wires were placed over the blotter paper, about 3 cm from the gel. Finally, the glass plate was placed on top of the anode and cathode and the lid was closed. Electrophoresis was conducted at 100 V for ∼2 h, although each run should be individually monitored. Samples should be visible as blue dots in the gel after ∼30 min, and ideally, the run should be stopped when it reaches the top edge of the next well. After 30 min, we paused the run, lifted the blotter paper and gel with a spatula, and rehydrated by placing another 10 mL of cool running buffer as previously.

Horizontal Tank Electrophoresis in Mesowestern(v2)

Running buffer was made by adding 12.1 g of tris to 900 mL of MilliQ water and 10 mL of 10% SDS. pH was adjusted to 7.5 by adding concentrated HCl dropwise and final volume adjusted to 1000 mL. Ice packs were placed in the positive and negative buffer chambers in the electrophoresis tank to keep the system cool during electrophoresis (Figure 6.4B). The gel with samples loaded into the wells and attached to the cast, was placed in the gel cast placement area and the buffer was gradually added in both the positive and negative chambers to make contact with the gel but not flow over the gel. The electrophoresis can be started and run for about 30-35 minutes till the bromophenol blue reaches close to the subsequent wells.

Transfer to Membrane

Transfer buffer was prepared by first making 10× tris-glycine buffer (600 mL of MilliQ water with 30.3 g of tris base (BIO-RAD #161-0719) and 144 g of glycine (VWR #0167), then MilliQ water was added to a final volume of 800 mL). Transfer buffer (~2 L, 1×) was made by taking 160 mL of 10× tris-glycine buffer, adding MilliQ water up to a final volume of 1600 mL, and finally, adding ~400 mL of methanol (Fisher #A412-in a fume hood) to 2 L. Transfer buffer is stored at 4 °C.

For quarter gels, we used a Mini Trans-Blot Cell (BIO-RAD, 1703930), and for full gels, we used a Criterion Blotter (BIO-RAD, 1704070). We have successfully used both nitrocellulose (Licor Odyssey #926-31092) and PVDF (BIO-RAD, 1620264) membranes for Mesowestern. For Mesowestern(v1) the reported results are from PVDF and for Mesowestern(v2) the reported results are with nitrocellulose.

To prepare the gel and membrane for transfer, cold transfer buffer was poured into a pyrex dish to a depth of ~3 cm. Blotter paper, cut to the size of the transfer cassette but larger than the gel, was placed into the pyrex dish to soak. After soaking, the blotter paper was placed on a clean, flat benchtop. Then, the gel could soak in the same transfer buffer for ~15 min, making sure to keep track of which side of the gel has the well indentations. The gel was then placed onto the soaked blotter paper, with the wells facing down on the paper. A spatula was always used to transport the gel. The gel was then gently rolled flat, and air pockets were removed using a roller (BIO-RAD, 1651279). The membrane was cut to the same size as the gel, being careful never to touch the membrane except with clean tools. After wetting with methanol (if PVDF is used), the membrane was then placed to soak in transfer buffer. Forceps were used to gently place the membrane onto the gel. If the membrane is not aligned, we did not move it, rather, we got a new membrane. Then, the membrane was rolled as previously. A second piece of transfer buffer-soaked blotter paper was then placed on top of the membrane in line with the first piece of blotter paper and rolled as previously. Finally, a spatula was used to lift the "sandwich" onto a fiber pad (BIO-RAD, 1703933), and another fiber pad was placed on top. This fiber pad-surrounded sandwich was moved to the transfer cassette, making sure that the side of the sandwich closest to the membrane was on the clear/positive side of the cassette (BIO-RAD, 1703931). This also means that the side of the sandwich closest to the gel is on the black/negative side of the cassette. The cassette was then placed into the transfer apparatus (negative to negative/black to black, positive to positive/clear to red). If desired, a second sandwich was made and placed into the apparatus.

With the cassettes in the transfer apparatus, a cold transfer buffer was added until it reached the indicated volume line. The apparatus was moved to a 4 °C room, and then transfer was carried out at 30 V for 16 h (usually overnight). After the transfer, the membrane was removed with clean forceps and was placed in a clean incubation box (Li-Cor, 929-97201), with the side of the membrane that was in contact with the gel facing up.

Antibody Incubation

First, TBS and TBST buffers were prepared. Briefly, 10× TBS was made by dissolving a 24 g tris base (BIO-RAD #161-0719) and 88 g NaCl (CAS 7647-14-5) in 1 L of MilliQ water. The pH was monitored with continuous magnetic stirring while adding HCl dropwise to bring the pH to 7.6. To make 1× TBS, 50 mL of 10× TBS was added to 450 mL of MilliQ water and stored at 4 °C (stable for several months). To make 1× TBST, 2.5 mL of 10% Tween 20 (BIO-RAD #161-0781) was added to 500 mL of 1× TBS and similarly stored at 4 °C.

All membrane incubations were done in the dark (sealed black box or covered in aluminum foil). The membrane was incubated first in ~20 mL of blocking buffer (1 g BSA (Fisher, BP1600) in 20 mL of 1× TBS) for at least 30 min at room temperature with gentle rocking. After blocking, the blocking buffer was removed, and the membrane was directly incubated with a primary antibody solution (10 mL blocking buffer, 50 μL of 10% Tween 20, v/v dilution of primary antibody to the desired working concentration) for at least 2 h at room temperature or overnight at 4 °C, all with gentle rocking. After primary antibody incubation, the membrane was washed with ~10 mL of 1× TBST three times, 5 min for nitrocellulose, and four times, 15 min for PVDF. After washing, the secondary antibody

solution (10 mL 1× TBST with 1:20,000 v/v; see below) was added to the membrane and incubated with gentle rocking for 1 h at room temperature. After incubation, the secondary antibody solution was discarded, and the membrane was washed as previously with 1× TBST. After the last TBST wash, a final TBS wash was done. The membrane was then scanned with the side that was facing up (closest to gel during transfer) now facing down on the clean surface of a LI-COR Odyssey infrared fluorescence scanning instrument (LI-COR model number 9140).

Antibodies were obtained from and used with working concentrations as follows: p-MAPK (Cell Signaling, #4370S, 1:1000), α-tubulin (Novus, #NB100-690, 1:1000), β-actin (Cell Signaling, #3700, 1:1000), pRb(Cell Signaling, #D59B7), pTCTP(Cell Signaling, #5251), antirabbit (800CW LI-COR #926-32211, 1:20,000), and antimouse (680LT LI-COR #925-68070, 1:20,000).

Imaging and Quantification

Placement of the membrane on the scanning surface was set in Image Studio. Both 700 nm and 800 nm wavelength channels were set to be scanned. Resolution was set to generally 42 μm (some exceptions for speed at times), and the focus offset was set to 0.0 mm. After the membrane finished scanning, the image and the associated zip file were exported from the Li-Cor Odyssey scanner and imported into Image Studio Lite for analysis. In Image Studio, boxes were drawn around protein bands and the "signal" metric generated by the software was used as the quantification.[8,38]

Discussion and Future Work

Mesowestern (v2) is a significant improvement over Mesowestern(v1) in its more streamlined approach and ease of use but requires more changes to be a more transferable technique, especially regarding the gel being immersed in the buffer. In order to minimize sample, wash off in a fully immersed setting, we expect subsequent designs of the Mesowestern(v2) to have deeper wells and increased sample density by incorporating more glycerol in the sample buffer. Designs where the gel can be loaded while being submerged (as is done for a DNA agarose gel) either manually, with a single channel pipette, multichannel or through an automated device should also be contemplated. In addition, a greater area for the wells should also be considered, to reduce errors in loading samples.

The Mesowestern(v2) results with pMAPK and pRb have a noticeable variability, especially at high signal values. We hope such variance should go down with designs which allow for a larger quantity of sample per well.

The results with pMAPK and pRb also indicate a negative feedback from CDK4/6 to MEK1/2 in both U87 and U251 cells. This is an interesting result and may need to be validated by a more robust experiment. While negative feedback loops are ubiquitous in signaling pathways, potential feedbacks from CDK4/6 to MEK1/2 are relatively underreported and may have some implications in drug efficacy-if not in glioblastoma cells for other types of cancer as well.

Acknowledgements

Cameron O. Zadeh, Jonah R. Huggins, Baylee C. Westbury, William R. Interiano, S. Ashley Phillips and William B. Dodd worked to develop Mesowestern (v1). Micah C.

Jordan, Wesley O. Meredith and Nicholas J. Harold were involved with the experiments to compare Mesowestern (v1) with regular western. Jonah R. Huggins was involved in every new cast design, design of tank and its 3D printing. Thanks to all the undergraduates involved for help and support in this project. Clemson Additive Manufacturing Lab and in particular-Tim Pruett printed the casts and we are very grateful to their support in this project.

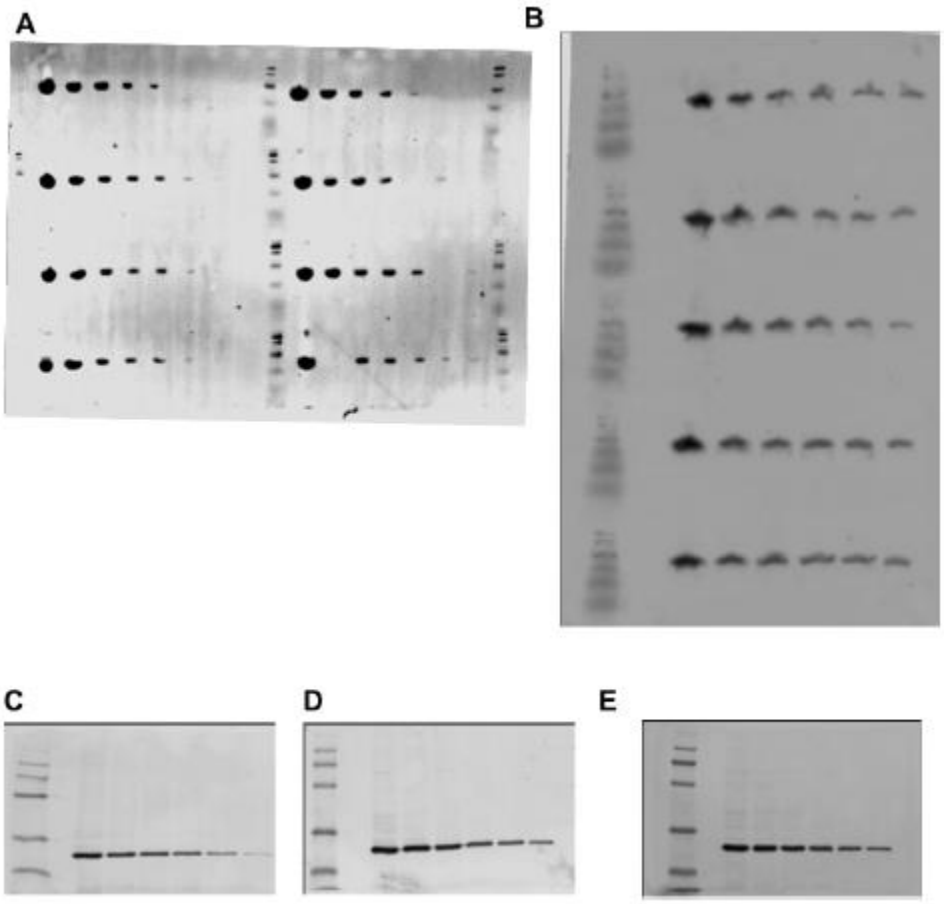


Figure 6.8 Full scans of the results shown in Figure 6.2. Operating parameters and other details are contained in Figure 6.2 and associated text. **A-B.** Mesowestern. In A, MCF10A lysates were used and the maximum lysate concentration was ~2 mg/mL, giving an extra datapoint. **C-E.** Western blot replicates.

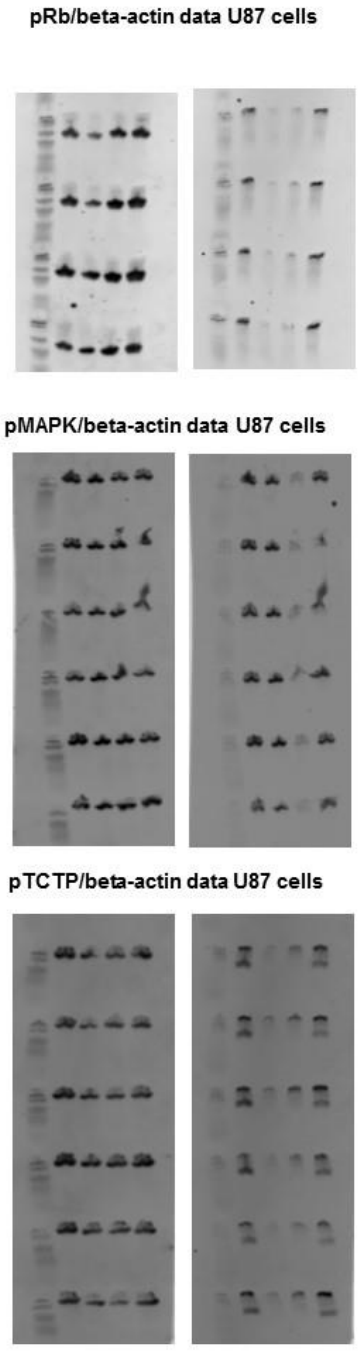
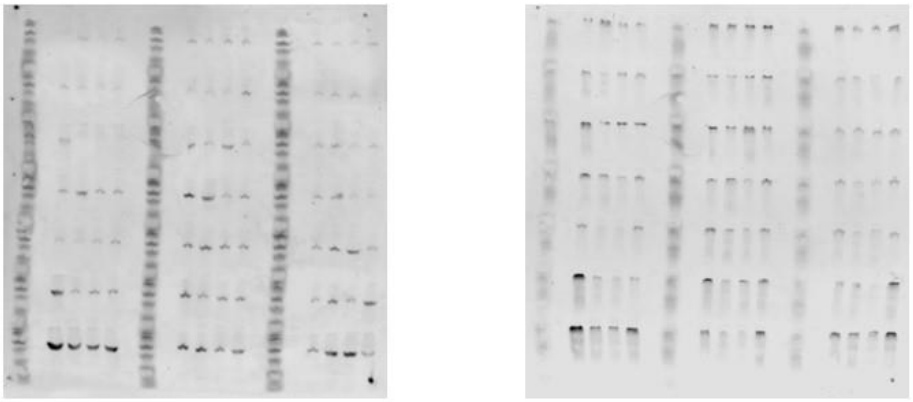


Figure 6.9 Replicates for validation of pMAPK, pRb and pTCTP antibodies. Quantification in Figure 6.6

a Beta actin/pRb data for time courses (top to bottom) for three replicates of U87 cells



b Beta actin/pMAPK data for time courses (top to bottom) for three replicates of U87 cells

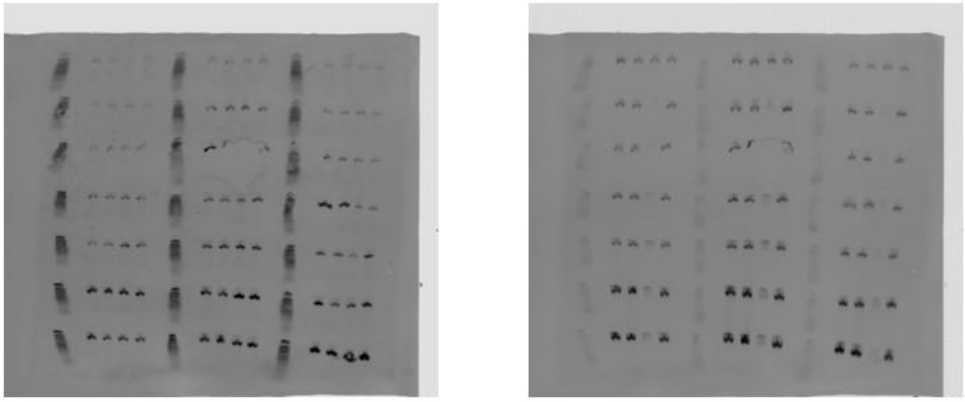
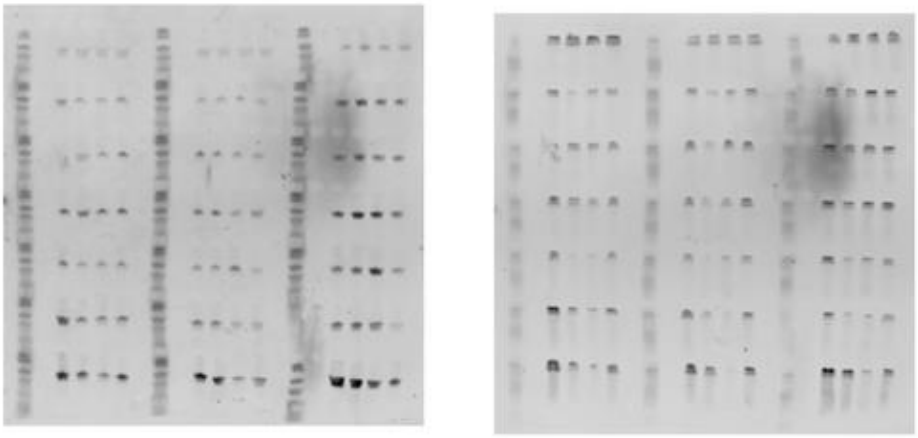


Figure 6.10 Full time course and 3 replicates for pMAPK and pRb antibodies in U87 cells. Top to bottom are time courses and left to right are three sets of replicates. Blot on the left is β-actin and blot on the right is pRb (above) and pMAPK(below).

a Beta actin/pRb data for time courses (top to bottom) for three replicates of U251 cells



b Beta actin/pMAPK data for time courses (top to bottom) for three replicates of U251 cells

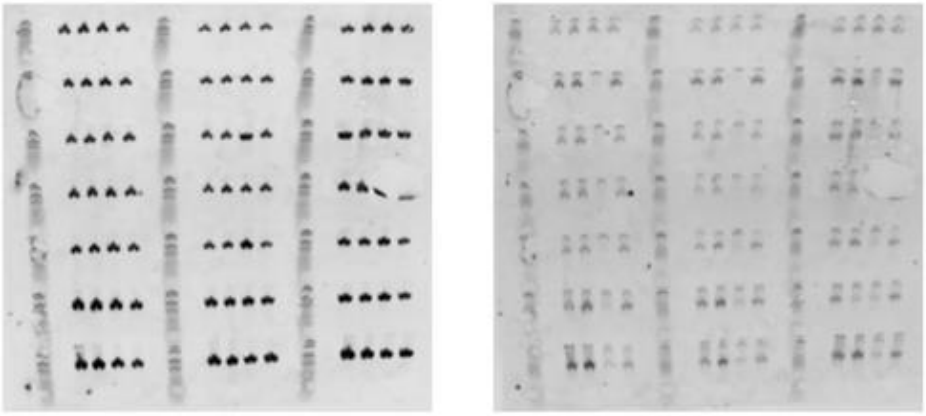


Figure 6.11 Full time course and 3 replicates for pMAPK and pRb antibodies in U251 cells. Top to bottom are time courses and left to right are three sets of replicates. Blot on the left is β-actin and blot on the right is pRb (above) and pMAPK(below).

References

1. Zadeh, C.O., Huggins, J.R., Sarmah, D., Westbury, B.C., Interiano, W.R., Jordan, M.C., Phillips, S.A., Dodd, W.B., Meredith, W.O., Harold, N.J., et al. (2022). Mesowestern Blot: Simultaneous Analysis of Hundreds of Submicroliter Lysates. ACS Omega *7*, 28912–28923. 10.1021/acsomega.2c02201.

2. Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci U S A *76*, 4350–4354. 10.1073/pnas.76.9.4350.

3. Kurien, B.T., and Scofield, R.H. (2015). Western Blotting: An Introduction. In Western Blotting: Methods and Protocols Methods in Molecular Biology., B. T. Kurien and R. H. Scofield, eds. (Springer), pp. 17–30. 10.1007/978-1-4939-2694-7_5.

4. Hnasko, T.S., and Hnasko, R.M. (2015). The Western Blot. In ELISA: Methods and Protocols Methods in Molecular Biology., R. Hnasko, ed. (Springer), pp. 87–96. 10.1007/978-1-4939-2742-5_9.

5. Kurien, B.T., and Scofield, R.H. eds. (2009). Protein Blotting and Detection: Methods and Protocols (Humana Press) 10.1007/978-1-59745-542-8.

6. Alegria-Schaffer, A., Lodge, A., and Vattem, K. (2009). Chapter 33 Performing and Optimizing Western Blots with an Emphasis on Chemiluminescent Detection. In Methods in Enzymology Guide to Protein Purification, 2nd Edition., R. R. Burgess and M. P. Deutscher, eds. (Academic Press), pp. 573–599. 10.1016/S0076-6879(09)63033-0.

7. Mathews, S.T., Plaisance, E.P., and Kim, T. (2009). Imaging Systems for Westerns: Chemiluminescence vs. Infrared Detection. In Protein Blotting and Detection: Methods and Protocols Methods in Molecular Biology., B. T. Kurien and R. H. Scofield, eds. (Humana Press), pp. 499–513. 10.1007/978-1-59745-542-8_51.

8. Janes, K.A. (2015). An analysis of critical factors for quantitative immunoblotting. Science Signaling *8*, rs2–rs2. 10.1126/scisignal.2005966.

9. Degasperi, A., Birtwistle, M.R., Volinsky, N., Rauch, J., Kolch, W., and Kholodenko, B.N. (2014). Evaluating Strategies to Normalise Biological Replicates of Western Blot Data. PLOS ONE *9*, e87293. 10.1371/journal.pone.0087293.

10. Pillai-Kastoori, L., Schutz-Geschwender, A.R., and Harford, J.A. (2020). A systematic approach to quantitative Western blot analysis. Analytical Biochemistry *593*, 113608. 10.1016/j.ab.2020.113608.

11. Moritz, C.P. (2020). 40 years Western blotting: A scientific birthday toast. Journal of Proteomics *212*, 103575. 10.1016/j.jprot.2019.103575.

12. Uhlen, M., Bandrowski, A., Carr, S., Edwards, A., Ellenberg, J., Lundberg, E., Rimm, D.L., Rodriguez, H., Hiltke, T., Snyder, M., et al. (2016). A proposal for validation of antibodies. Nat Methods *13*, 823–827. 10.1038/nmeth.3995.

13. Baker, M. (2015). Reproducibility crisis: Blame it on the antibodies. Nature *521*, 274–276. 10.1038/521274a.

14. Krajewski, S. (Stanislaw), Tsukamoto, M.M., Huang, X., and Krajewski, S.B. (2015). Nonstripping "Rainbow" and Multiple Antigen Detection (MAD) Western Blotting. In Detection of Blotted Proteins: Methods and Protocols Methods in Molecular Biology., B. T. Kurien and R. H. Scofield, eds. (Springer), pp. 287–301. 10.1007/978-1-4939-2718-0_30.

15. Aksamitiene, E., Hoek, J.B., Kholodenko, B., and Kiyatkin, A. (2007). Multistrip Western blotting to increase quantitative data output. ELECTROPHORESIS *28*, 3163–3173. 10.1002/elps.200700002.

16. Aksamitiene, E., Hoek, J.B., and Kiyatkin, A. (2015). Multistrip Western Blotting: A Tool for Comparative Quantitative Analysis of Multiple Proteins. In Western Blotting: Methods and Protocols Methods in Molecular Biology., B. T. Kurien and R. H. Scofield, eds. (Springer), pp. 197–226. 10.1007/978-1-4939-2694-7_23.

17. Krajewski, S., Huang, X., and Krajewska, M. (2009). Multiple antigen detection (MAD) western blotting. In Protein Blotting and Detection: Methods and Protocols Methods in Molecular Biology., B. T. Kurien and R. H. Scofield, eds. (Humana Press), pp. 473–481. 10.1007/978-1-59745-542-8_48.

18. Krajewski, S. (2009). "Rainbow" Western Blotting. In Protein Blotting and Detection: Methods and Protocols Methods in Molecular Biology., B. T. Kurien and R. H. Scofield, eds. (Humana Press), pp. 463–472. 10.1007/978-1-59745-542-8_47.

19. Tibes, R., Qiu, Y., Lu, Y., Hennessy, B., Andreeff, M., Mills, G.B., and Kornblau, S.M. (2006). Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. Molecular Cancer Therapeutics *5*, 2512–2521. 10.1158/1535-7163.MCT-06-0334.

20. Hennessy, B.T., Lu, Y., Gonzalez-Angulo, A.M., Carey, M.S., Myhre, S., Ju, Z., Davies, M.A., Liu, W., Coombes, K., Meric-Bernstam, F., et al. (2010). A Technical Assessment of the Utility of Reverse Phase Protein Arrays for the Study of the Functional Proteome in Non-microdissected Human Breast Cancers. Clin Proteom *6*, 129–151. 10.1007/s12014-010-9055-y.

21. Earley, M.C., Vogt Jr., R.F., Shapiro, H.M., Mandy, F.F., Kellar, K.L., Bellisario, R., Pass, K.A., Marti, G.E., Stewart, C.C., and Hannon, W.H. (2002). Report from a workshop on multianalyte microsphere assays. Cytometry *50*, 239–242. 10.1002/cyto.10140.

22. He, J. (2013). Chapter 5.1 - Practical Guide to ELISA Development. In The Immunoassay Handbook (Fourth Edition), D. Wild, ed. (Elsevier), pp. 381–393. 10.1016/B978-0-08-097037-0.00025-7.

23. Engvall, E., and Perlmann, P. (1972). Enzyme-linked immunosorbent assay, Elisa. 3. Quantitation of specific antibodies by enzyme-labeled anti-immunoglobulin in antigen-coated tubes. J Immunol *109*, 129–135.

24. Aebersold, R., and Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. Nature *537*, 347–355. 10.1038/nature19949.

25. Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A.M., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., et al. (2014). Mass-spectrometry-based draft of the human proteome. Nature *509*, 582–587. 10.1038/nature13319.

26. Shevchenko, A., Tomas, H., Havli, J., Olsen, J.V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc *1*, 2856–2860. 10.1038/nprot.2006.468.

27. Bittremieux, W., Tabb, D.L., Impens, F., Staes, A., Timmerman, E., Martens, L., and Laukens, K. (2018). Quality control in mass spectrometry-based proteomics. Mass Spectrometry Reviews *37*, 697–711. 10.1002/mas.21544.

28. Yates, J.R., Ruse, C.I., and Nakorchevsky, A. (2009). Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. Annual Review of Biomedical Engineering *11*, 49–79. 10.1146/annurev-bioeng-061008-124934.

29. Handler, D.C., Pascovici, D., Mirzaei, M., Gupta, V., Salekdeh, G.H., and Haynes, P.A. (2018). The Art of Validating Quantitative Proteomics Data. PROTEOMICS *18*, 1800222. 10.1002/pmic.201800222.

30. Hughes, A.J., Spelke, D.P., Xu, Z., Kang, C.-C., Schaffer, D.V., and Herr, A.E. (2014). Single-cell western blotting. Nat Methods *11*, 749–755. 10.1038/nmeth.2992.

31. Kang, C.-C., Yamauchi, K.A., Vlassakis, J., Sinkala, E., Duncombe, T.A., and Herr, A.E. (2016). Single cell–resolution western blotting. Nat Protoc *11*, 1508–1530. 10.1038/nprot.2016.089.

32. Sanders, B.J., Kim, D.C., and Dunn, R.C. (2016). Recent advances in microscale western blotting. Anal. Methods *8*, 7002–7013. 10.1039/C6AY01947A.

33. O'Neill, R.A., Bhamidipati, A., Bi, X., Deb-Basu, D., Cahill, L., Ferrante, J., Gentalen, E., Glazer, M., Gossett, J., Hacker, K., et al. (2006). Isoelectric focusing technology quantifies protein signaling in 25 cells. Proceedings of the National Academy of Sciences *103*, 16153–16158. 10.1073/pnas.0607973103.

34. Dengler-Crish, C.M., Smith, M.A., and Wilson, G.N. (2017). Early Evidence of Low Bone Density and Decreased Serotonergic Synthesis in the Dorsal Raphe of a Tauopathy Model of Alzheimer's Disease. Journal of Alzheimer's Disease *55*, 1605–1619. 10.3233/JAD-160658.

35. Treindl, F., Ruprecht, B., Beiter, Y., Schultz, S., Döttinger, A., Staebler, A., Joos, T.O., Kling, S., Poetz, O., Fehm, T., et al. (2016). A bead-based western for high-throughput cellular signal transduction analyses. Nat Commun *7*, 12852. 10.1038/ncomms12852.

36. Ciaccio, M.F., and Jones, R.B. (2017). Microwestern Arrays for Systems-Level Analysis of SH2 Domain-Containing Proteins. In SH2 Domains: Methods and Protocols Methods in Molecular Biology., K. Machida and B. A. Liu, eds. (Springer), pp. 453–473. 10.1007/978-1-4939-6762-9_27.

37. Ciaccio, M.F., Wagner, J.P., Chuu, C.-P., Lauffenburger, D.A., and Jones, R.B. (2010). Systems analysis of EGF receptor signaling dynamics with microwestern arrays. Nat Methods *7*, 148–155. 10.1038/nmeth.1418.

38. Koch, R.J., Barrette, A.M., Stern, A.D., Hu, B., Bouhaddou, M., Azeloglu, E.U., Iyengar, R., and Birtwistle, M.R. (2018). Validating Antibodies for Quantitative Western Blot Measurements with Microwestern Array. Sci Rep *8*, 11329. 10.1038/s41598-018-29436-0.

39. Wiśniewski, J.R., Hein, M.Y., Cox, J., and Mann, M. (2014). A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. Mol Cell Proteomics *13*, 3497–3506. 10.1074/mcp.M113.037309.

40. Macromolecular Components of E. coli and HeLa Cells | Thermo Fisher Scientific - US https://www.thermofisher.com/us/en/home/references/ambion-tech-support/rna-tools-and-calculators/macromolecular-components-of-e.html.

41. HeLa cell volume - Human Homo sapiens - BNID 103725 https://bionumbers.hms.harvard.edu/bionumber.aspx?id=103725&ver=14.

42. Fan, B., Li, X., Liu, L., Chen, D., Cao, S., Men, D., Wang, J., and Chen, J. (2018). Absolute Copy Numbers of β-Actin Proteins Collected from 10,000 Single Cells. Micromachines (Basel) *9*, E254. 10.3390/mi9050254.

43. Schutz-Geschwender, A., Zhang, Y., Holt, T., McDermitt, D., and Olive, D.M. Quantitative, Two-Color Western Blot Detection With Infrared Fluorescence. 8.

44. Detection Methods Bio-Rad Laboratories. https://www.bio-rad.com/en-us/applications-technologies/detection-methods?ID=LUSQ6KKG4.

45. Kurien, B.T., and Scofield, R.H. (2012). Common artifacts and mistakes made in electrophoresis. Methods Mol Biol *869*, 633–640. 10.1007/978-1-61779-821-4_58.

46. Zhang, Z., Golomb, L., and Meyerson, M. (2022). Functional Genomic Analysis of CDK4 and CDK6 Gene Dependency across Human Cancer Cell Lines. Cancer Res *82*, 2171–2184. 10.1158/0008-5472.CAN-21-2428.

47. Cucchi, U., Gianellini, L.M., De Ponti, A., Sola, F., Alzani, R., Patton, V., Pezzoni, A., Troiani, S., Saccardo, M.B., Rizzi, S., et al. (2010). Phosphorylation of TCTP as a marker for polo-like kinase-1 activity in vivo. Anticancer Res *30*, 4973–4985.

CHAPTER SEVEN

MAPPING DRUG RESISTANCE MECHANISMS DUE TO OVEREXPRESSION OF A KINASE

Overview

Biological markers of cancers often include mutations resulting in overexpressed or hyperactive genes. Targeted therapy to these markers is thought to be a solution but are often accompanied by drug resistance. Drug resistance in cancer can be a daunting in its complexity but activation of alternate survival pathways is thought to be an important contributing factor to drug resistance in cases where a driver pathway is inhibited[1]. In such cases, combinations treatments, targeting the alternate pathways can be potentially more effective than single treatment alone, but the inherent resistance mechanisms need to be understood first.

A start to understanding the different mechanisms of drug resistance can be through analysis of different omics data in cases of drug treatment vs control. Specifically, the differences in omics data in drug treated cells with an overexpressed oncogene vs drug treated cells may give us an idea about the different pathways being employed for drug resistance in the overexpressed cells.

Over the years, various gene set enrichment approaches have been helpful in analyzing genome wide data- Gene Set Enrichment Analysis (GSEA)[2], Enrichr[3] and GeneTrail[4], among others. These approaches can employ various databases of gene ontologies to match what we are looking for-for instance pathways, cellular component etc. Such gene enrichment may help bringing forward pathways potentially involved in resistance mechanisms. Statistical methods such as principal component analysis (PCA)

may also potentially help curate the omics data before being processed for gene set enrichments by aligning the data along some principal axes, representative of overall effect.

Here we employed one gene set enrichment method-GSEA to understand resistance mechanisms in two cases in proteomics data generated through mass spectrometry by the Kriegsheim Lab at the University of Edinburgh. These were U87 cells with overexpression of two proto oncogenes- (i) mTOR and (ii) MEK2 and treated with their respective inhibitors. By scoring the genes through a method employing Euclidian distance in PCA, and using GSEA and one of the largest pathway databases (Reactome)[5], we could begin to uncover some potential pathways involved in the resistance mechanisms, when these proto oncogenes were inhibited. Further work on this project, concerning incorporation of the proteomic data in mechanistic modeling is currently underway in the lab.

Results-

Data-

Three stable U87 cell lines were generated in the Kriegheim lab with three different CRISPR gRNAs of the respective proto oncogene. It was found that the cell lines with mTOR overexpression were more resistant to the mTOR inhibitor eCF309 (Figure 7.1).

The proteomic dataset involved log2 normalized proteomics data generated through mass spectrometry four kinds of treatments for the respective proto oncogene.

i. Protein Overexpression(-), Drug Treatment(-)
ii. Protein Overexpression(-), Drug Treatment(+)
iii. Protein Overexpression(+), Drug Treatment(-)

iv. Protein Overexpression(+), Drug Treatment(+)

Each of the four treatments involved three replicates of drug treatment, and in case of overexpression, three different gRNAs. We took the median values of the replicates, for simplicity and to eliminate any large variations.

We reasoned that the following difference between these treatment medians could allow us to uncover each of the following mechanisms-

i. Mapping effects of drug without involvement of overexpression (Drug Effect) = Cells with drug treatment-Cells without drug treatment (Both without Overexpression)

ii. Adaptation of cellular pathways to overexpression (Overexpression Adaptation) = Cells with overexpression-cells without overexpression (Both without drug treatment)

iii. Mapping effects of drug when overexpression is present (Drug Effect with Overexpression) = Cells with drug treatment-Cells without drug treatment (Both with mTOR overexpression)

iv. Mapping resistive pathways in overexpressed cells compared to cells without overexpression (Resistome)= Cells with overexpression-Cells without Overexpression (Both with drug treatment)

These differences were first taken through a direct difference method (see Methods) and subsequently through a method involving Euclidean distance in PCA (Figure 7.2)

GSEA-mTOR overexpressed cells

We first used a direct difference method (see Methods) to generate a scored list of proteins. This list is then processed through GSEA to obtained enriched pathways in Reactome- 2016 dataset (see Methods). The positively enriched pathways, with a nominal p value less than 0.01 are shown in Table 7.1. These enriched pathways, particularly in the "Resistome" did not indicate any evidence of enriched signaling pathways.

We then decided to employ the Euclidean PCA distance method to generate a scored list of proteins. This list is again processed through GSEA to obtained enriched pathways in Reactome- 2016 dataset. The positively enriched pathways, with a nominal p value less than 0.01 are shown in Table 7.2. This time we saw enrichment of signaling pathways, particularly involving FGFR, SCF, KIT, and two death signaling pathways involving NRAGE and JNK. By using the mapping feature in GSEA (based on fraction of common genes between pathways), we observed mapping between FGFR and SCF/KIT pathways and similar mapping between the two death signaling pathways

FGFR associated with resistance to mTOR inhibitor

Based on the above results, the Kriegheim lab decided to experimentally test whether FGFR pathways were involved in resistance to mTOR inhibition. Different concentrations of the mitogen FGF2 was added to U87 cells, also being treated with the mTOR inhibitor. Increasing FGF2 concentration seems to increase cell survival, even at high doses of the mTOR inhibitor (Figure 7.3).

GSEA results-MEK2 overexpressed cells

We employed the Euclidean PCA distance method to generate a scored list of proteins for the MEK2 overexpressed U87 cells. This list is processed through GSEA to

obtained enriched pathways in Reactome- 2016 dataset. However, the number of positively enriched pathways, with a nominal p value less than 0.01, were too many in each category and only the relevant list regarding drug resistance-"Resistome" is shown in Table 7.3. No special signaling pathways seem to stand out in GSEA data except those involving cell cycle and the WNT pathway. Further experimentation was not performed to verify the role of WNT in resistance in MEK2 signaling pathway.

Methods

Direct Difference Method

In this method, the list of median proteins is first normalized (z-scores across all rows) and then the difference taken as explained before (Results-Data).

PCA method-

In this method, the list of median proteins is first quantile normalized and further z-scored across all columns. A PCA is created with each of the treatments occupying a location in the PCA plane (PC1 and PC2). The Euclidean distance between each relevant treatment is measured. Then, each protein is sequentially dropped from all the treatments and the new Euclidean distance between each relevant treatment is measured again. This is followed for all proteins and the difference between the old and new Euclidean distance is reasoned to score the importance of a protein in the PCA plane. The method is illustrated in Figure 7.2.

GSEA

The scored lists of proteins have to be converted to a .rnk file (insert .rnk in the file name and save as txt file). Then they are put through GSEA software (v 4.0.3. Newer

versions are now available). We use the RunGSEAPreranked feature and upload the file with the protein scores. The Gene set database is selected to “Reactome-2016” (Reactome-2022 should be now available). After a minute or so, the enrichment should be complete, and the enrichment data will be downloaded.

Discussion

We found GSEA to be the very conducive to pre-scored data because of its feature to be able to take in pre-scored gene names. There was no need to select a certain number of top gene and the whole dataset could be uploaded. Combining GSEA with the pre-scored lists Euclidean PCA method may have potential to bring out relevant pathways compared to the direct difference method.

FGFR pathways may have a role in resistance to mTOR inhibitors. Previous studies had mentioned potential synergistic effects of targeting both FGFR and mTOR in different cancer models. It seems reasonable that this could be true for glioblastoma cells as well[6–8].

The current status of this study is ongoing in the lab, particularly for incorporating the proteomic data for MEK2 inhibition and possibly gaining further insight into predicting better combination therapies.

Table 7.1

DIRECT DIFFERENCE-GSEA RESULTS FOR MTOR OVEREXPRESSED CELLS

Drug Effect
REACTOME_MRNA_SPLICING
REACTOME_PROCESSING_OF_CAPPED_INTRON_CONTAINING_PRE_MRNA
REACTOME_MRNA_PROCESSING
REACTOME_INFLUENZA_LIFE_CYCLE
REACTOME_MRNA_SPLICING_MINOR_PATHWAY
REACTOME_OPIOID_SIGNALLING
REACTOME_METABOLISM_OF_RNA
REACTOME_METABOLISM_OF_CARBOHYDRATES
REACTOME_SIGNAL_AMPLIFICATION
REACTOME_NEUROTRANSMITTER_RECEPTOR_BINDING_AND_DOWNSTREAM_TRANSMISSION_IN_THE_POSTSYNAPTIC_CELL
REACTOME_G_ALPHA_I_SIGNALLING_EVENTS
REACTOME_RNA_POL_II_TRANSCRIPTION
REACTOME_REGULATION_OF_INSULIN_SECRETION_BY_GLUCAGON_LIKE_PEPTIDE1
REACTOME_GLUCOSE_METABOLISM
REACTOME_TRANSPORT_OF_RIBONUCLEOPROTEINS_INTO_THE_HOST_NUCLEUS
REACTOME_G_ALPHA1213_SIGNALLING_EVENTS
mTOR Overexpression Adaptation
REACTOME_SIGNALING_BY_SCF_KIT
REACTOME_ASPARAGINE_N_LINKED_GLYCOSYLATION
REACTOME_MEIOTIC_SYNAPSIS
REACTOME_COSTIMULATION_BY_THE_CD28_FAMILY
REACTOME_INSULIN_RECEPTOR_SIGNALLING_CASCADE
REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS
REACTOME_CYTOSOLIC_TRNA_AMINOACYLATION
Drug Effect with mTOR Overexpression
REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM
REACTOME_SIGNALING_BY_ILS
REACTOME_AXON_GUIDANCE
REACTOME_NCAM_SIGNALING_FOR_NEURITE_OUT_GROWTH
REACTOME_DEVELOPMENTAL_BIOLOGY
REACTOME_INTERFERON_GAMMA_SIGNALING
Resistome
REACTOME_MEIOTIC_RECOMBINATION
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_UNCOUPLING_PROTEINS_
REACTOME_RNA_POL_I_RNA_POL_III_AND_MITOCHONDRIAL_TRANSCRIPTION
REACTOME_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT
REACTOME_RNA_POL_I_TRANSCRIPTION
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT
REACTOME_CITRIC_ACID_CYCLE_TCA_CYCLE

Figure 7.1

MTOR OVEREXPRESSION CONFERS RESISTANCE TO MTOR INHIBITOR

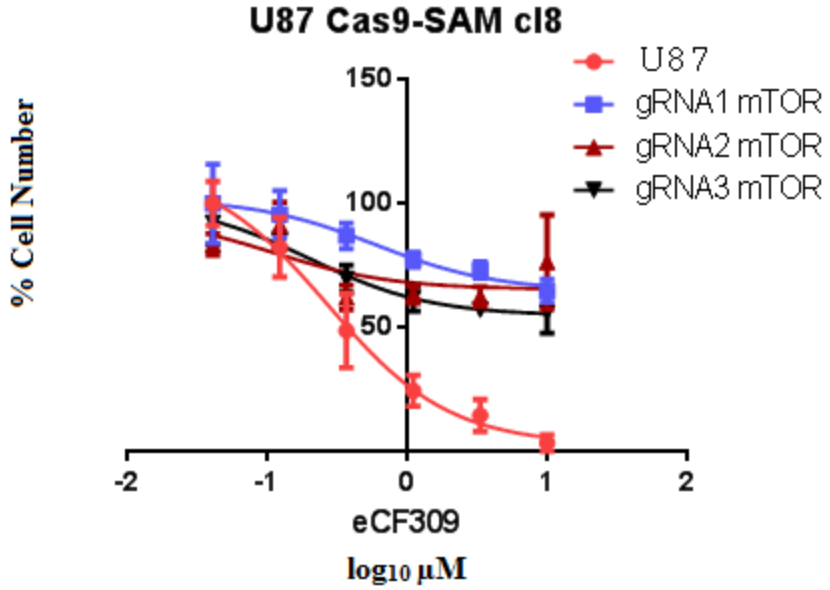


Figure 7.2

EUCLIDEAN DISTANCE PCA METHOD

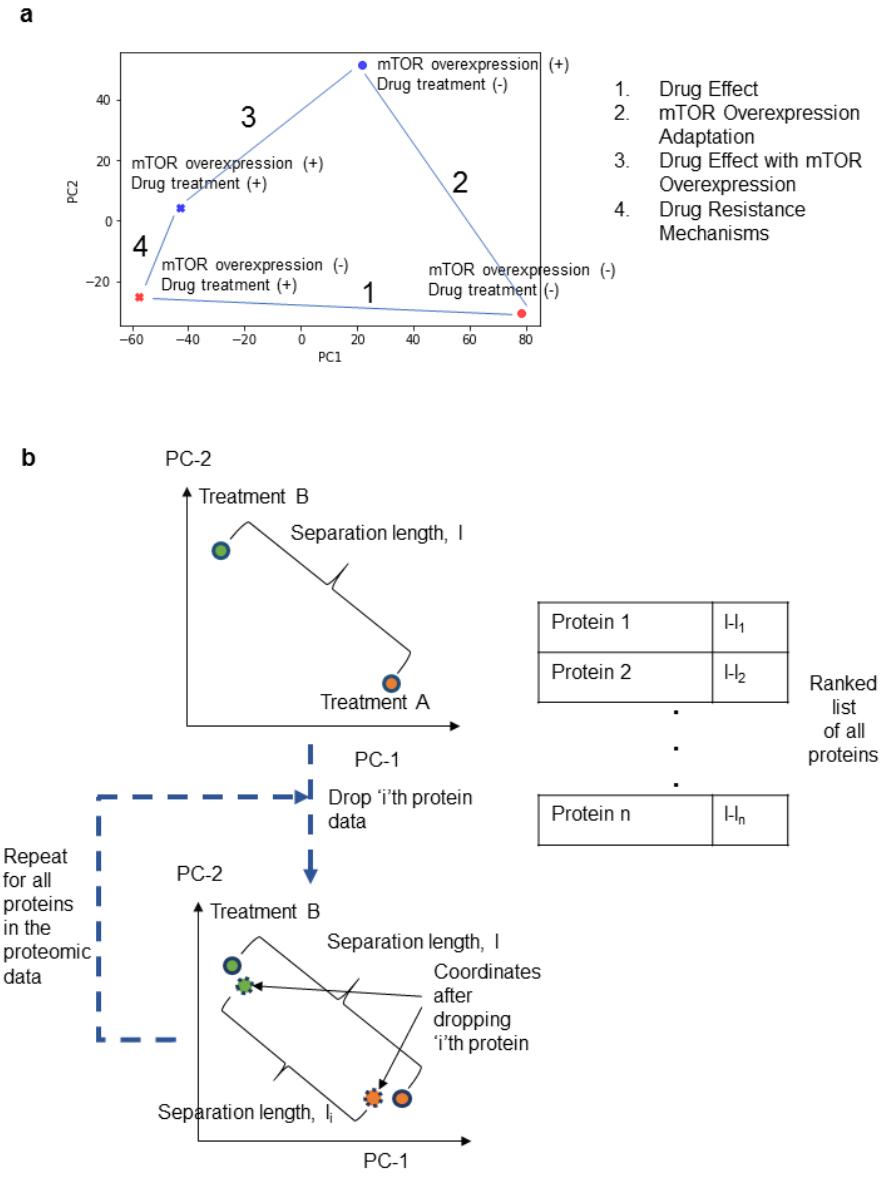


Table 7.2

EUCLIDEAN PCA DISTANCE-GSEA RESULTS FOR MTOR OVEREXPRESSED CELLS

Drug effect
REACTOME_REGULATION_OF_INSULIN_SECRETION_BY_GLUCAGON_LIKE_PEPTIDE1
REACTOME_OPIOID_SIGNALLING
REACTOME_METABOLISM_OF_CARBOHYDRATES
REACTOME_G_ALPHA_I_SIGNALLING_EVENTS
REACTOME_INTEGRATION_OF_ENERGY_METABOLISM
REACTOME_G_ALPHA1213_SIGNALLING_EVENTS
REACTOME_G_ALPHA_Q_SIGNALLING_EVENTS
REACTOME_SIGNAL_AMPLIFICATION
REACTOME_REGULATION_OF_INSULIN_SECRETION
REACTOME_LATE_PHASE_OF_HIV_LIFE_CYCLE
REACTOME_HIV_LIFE_CYCLE
REACTOME_PLATELET_HOMEOSTASIS
REACTOME_MRNA_SPLICING_MINOR_PATHWAY
REACTOME_GPCR_DOWNSTREAM_SIGNALING
mTOR Overexpression Adaptation
REACTOME_SIGNALING_BY_TGF_BETA_RECEPTOR_COMPLEX
Drug Effect with mTOR Overexpression
REACTOME_MEIOTIC_RECOMBINATION
REACTOME_SIGNALING_BY_PDGF
REACTOME_SIGNALLING_BY_NGF
REACTOME_NGF_SIGNALLING_VIA_TRKA_FROM_THE_PLASMA_MEMBRANE
REACTOME_SIGNALING_BY_GPCR
REACTOME_SIGNALING_BY_FGFR_IN_DISEASE
Resistome
REACTOME_CELL_DEATH_SIGNALLING_VIA_NRAGE_NRIF_AND_NADE
REACTOME_SIGNALING_BY_FGFR_IN_DISEASE
REACTOME_SIGNALING_BY_SCF_KIT
REACTOME_NRAGE_SIGNALS_DEATH_THROUGH_JNK

Figure 7.3

GSEA RESULTS FOR MTOR OVEREXPRESSED CELLS

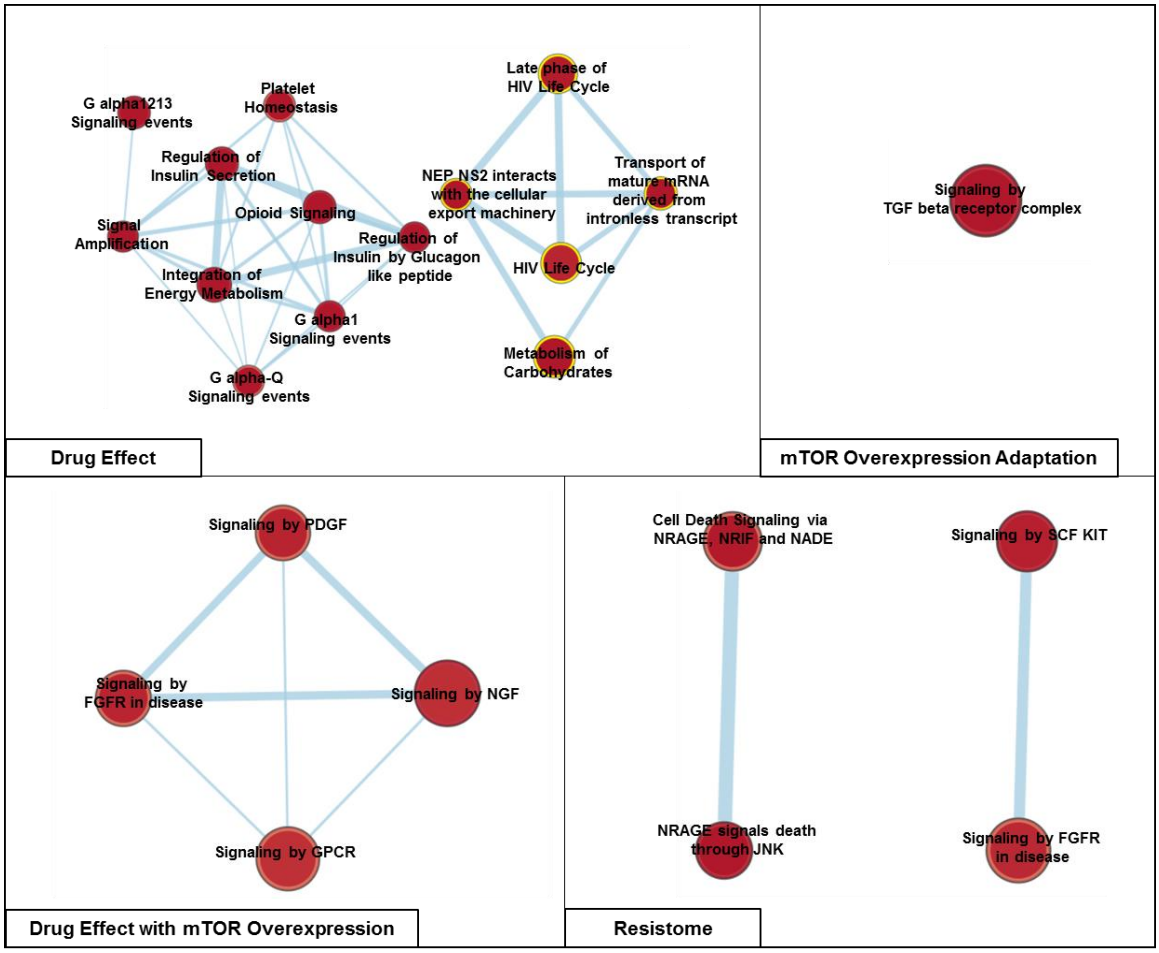


Figure 7.3

FGF2 CONFERS RESISTANCE TO MTOR INHIBITOR

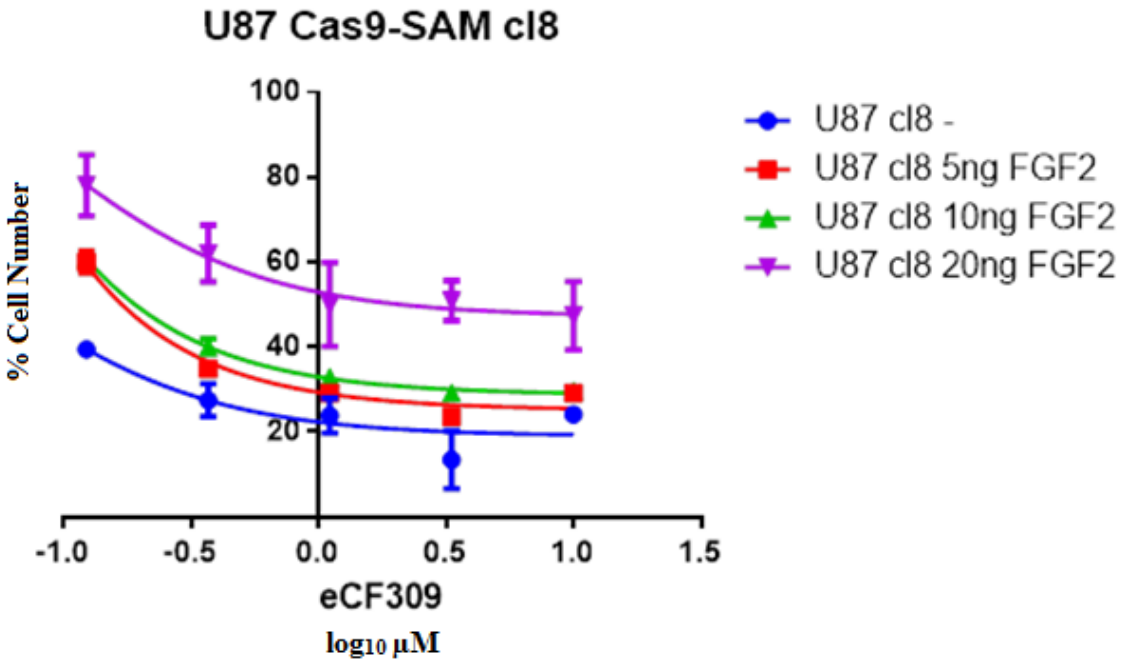


Table 7.3

PCA DIFFERENCE-GSEA RESULTS FOR MEK2 OVEREXPRESSED CELLS

Resistome
REACTOME_SCF_BETA_TRCP_MEDIATED_DEGRADATION_OF_EMI1
REACTOME_CYCLIN_E_ASSOCIATED_EVENTS_DURING_G1_S_TRANSITION_
REACTOME_ACTIVATION_OF_NF_KAPPAB_IN_B_CELLS
REACTOME_AUTODEGRADATION_OF_THE_E3_UBIQUITIN_LIGASE_COP1
REACTOME_CDT1_ASSOCIATION_WITH_THE_CDC6_ORC_ORIGIN_COMPLEX
REACTOME_P53_DEPENDENT_G1_DNA_DAMAGE_RESPONSE
REACTOME_CDK_MEDIATED_PHOSPHORYLATION_AND_REMOVAL_OF_CDC6
REACTOME_SCFSKP2_MEDIATED_DEGRADATION_OF_P27_P21
REACTOME_SIGNALING_BY_WNT
REACTOME_P53_INDEPENDENT_G1_S_DNA_DAMAGE_CHECKPOINT
REACTOME_AUTODEGRADATION_OF_CDH1_BY_CDH1_APC_C
REACTOME_G1_S_TRANSITION
REACTOME_MITOTIC_G1_G1_S_PHASES
REACTOME_ASSEMBLY_OF_THE_PRE_REPLICATIVE_COMPLEX
REACTOME_APC_C_CDH1_MEDIATED_DEGRADATION_OF_CDC20_AND_OTHER_APC_C_CDH1_TARGETED_PROTEINS_IN_LATE_MITOSIS_EARLY_G1
REACTOME_CROSS_PRESENTATION_OF_SOLUBLE_EXOGENOUS_ANTIGENS_ENDOSOMES
REACTOME_REGULATION_OF_MITOTIC_CELL_CYCLE
REACTOME_GOLGI_ASSOCIATED_VESICLE_BIOGENESIS
REACTOME_REGULATION_OF_ORNITHINE_DECARBOXYLASE_ODC
REACTOME_ORC1_REMOVAL_FROM_CHROMATIN
REACTOME_APC_C_CDC20_MEDIATED_DEGRADATION_OF_MITOTIC_PROTEINS
REACTOME_G1_PHASE
REACTOME_S_PHASE
REACTOME_HOST_INTERACTIONS_OF_HIV_FACTORS
REACTOME_HIV_INFECTION
REACTOME_VIF_MEDIATED_DEGRADATION_OF_APOBEC3G
REACTOME_DESTABILIZATION_OF_MRNA_BY_AUF1_HNRNP_D0
REACTOME_REGULATION_OF_APOPTOSIS

References

1. Holohan, C., Van Schaeybroeck, S., Longley, D.B., and Johnston, P.G. (2013). Cancer drug resistance: an evolving paradigm. Nat Rev Cancer *13*, 714–726. 10.1038/nrc3599.

2. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences *102*, 15545–15550. 10.1073/pnas.0506580102.

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

4. GeneTrail—advanced gene set enrichment analysis | Nucleic Acids Research | Oxford Academic https://academic.oup.com/nar/article/35/suppl_2/W186/2923179?login=false.

5. Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., et al. (2011). Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res *39*, D691-697. 10.1093/nar/gkq1018.

6. Cai, W., Song, B., and Ai, H. (2019). Combined inhibition of FGFR and mTOR pathways is effective in suppressing ovarian cancer. Am J Transl Res *11*, 1616–1625.

7. Gozgit, J.M., Squillace, R.M., Wongchenko, M.J., Miller, D., Wardwell, S., Mohemmad, Q., Narasimhan, N.I., Wang, F., Clackson, T., and Rivera, V.M. (2013). Combined targeting of FGFR2 and mTOR by ponatinib and ridaforolimus results in synergistic antitumor activity in FGFR2 mutant endometrial cancer models. Cancer Chemother Pharmacol *71*, 1315–1323. 10.1007/s00280-013-2131-z.

8. Scheller, T., Hellerbrand, C., Moser, C., Schmidt, K., Kroemer, A., Brunner, S.M., Schlitt, H.J., Geissler, E.K., and Lang, S.A. (2015). mTOR inhibition improves fibroblast growth factor receptor targeting in hepatocellular carcinoma. Br J Cancer *112*, 841–850. 10.1038/bjc.2014.638.

CHAPTER EIGHT

CONCLUSION

Conclusions

This dissertation was broadly inspired by the need to map various the various levels of complexity in cancer in order to predict effective therapies. Due to the multi-faceted contexts of cancer and increasing number of possible combinations, we propose that proper computational models should guide effective combination therapies. An increasingly important facet of cancer is cell state networks and their role in drug resistance. This dissertation explores different methods to understand and map the inherent biology at different levels of cancer biology and use these techniques to predict drug combinations. Although a lot of work remains to be done in expanding the scope of these techniques, this dissertation lays a foundation for mapping relevant cancer biology for prediction of better therapies.

In Chapter 2, we build upon existing methods to create Dynamic Least-Squares Modular Response Analysis (DL-MRA)- a method to help map different kinds of biological networks. We apply this method to infer a simulated biochemical network, 16 gene regulatory networks and cell state transition networks and we found satisfactory mapping in all cases, albeit some concerns about experimental application to biochemical networks. Insights gained from this work informed a subsequent temporal cell state network in Chapter 5.

In Chapter 3, we created a ranked list of kinase targets-scored by two pharmacological and two biological criteria. Begun as an illustrative exercise for the

selection of relevant proteins in a cancer, in absence of drug dose response data, it was later shown that such ranking may not necessarily match cancer cell's response to drugs.

In Chapter 4, we obtained drug dose responses for a panel of 22 BBB penetrant drugs across three glioblastoma derived cell lines. This work also provideed an experimental basis for the work in Chapter 5 and can provide useful information about cancer cell's response to inhibition of drug targets.

In Chapter 5, we conceive of a temporal cell state model and use it to predict certain drug combinations, which are then validated by experiments. We hope that such modeling and prediction becomes more prevalent and useful, especially in cases where the cell state transitions are not properly understood.

In Chapter 6 we described the various steps we took to validate a version of Mesowestern-a method to enable high throughput western blots. Although we made significant improvements, there is more scope for improvement in the design as well as variability of results.

In Chapter 7, we analyzed proteomic data obtained from glioblastoma cell lines with certain overexpressed proto-oncogenes. We explore the use of enrichment analysis and the use of a dimension reduction method with an intention to uncover signaling pathways involved in resistance mechanisms in these cells. We could obtain an experimental result hinting at validation for one of the results.

Future Work

There can be several improvements in the techniques and future avenues of the methods we explored in this dissertation.

In Chapter 2, an improvement for DL-MRA would be to be able to handle larger networks. Sparcer networks integrated with by canonical information may be a way to accomplish this. Another improvement would be to accommodate multiple perturbations. This may be approached in different ways and a simple way may be to frame perturbations as a positive or negative stimulus. This can be tested in the basic DL-MRA framework and also simulated biological networks. During mapping of cell state transitions, we observed that discrete Markov models are computationally faster than differential equation-based model. It is worth exploring whether a discrete version of DL-MRA can be used without significant loss in prediction accuracy. This can significantly improve the computation time required to test multiple random networks.

The ideas explored in Chapter 5 have multiple potential avenues. First, instead of temporal cell states, could we map actual cell states and use it to predict better drug combinations? Another way to increase the scope of the model would be to integrate biochemical networks and perhaps even cancer cell states along with temporal cell state transitions. A potential test case to integrate biochemical networks with temporal cell state with might be the inclusion of Abl1 (or PI3K, mTOR) and its cross talk with MEK1/2 and test its effects on drug dose response predictions.

We also expect future versions of the Mesowestern be more robust in design and results and more suitable for automated pipetting or through multi-channel pipette.

Although the methods explored in this dissertation were tested in glioblastoma cells, in theory they are general and can be tested across different types of cancer. The

complexity of cancer can be daunting but is inevitable. I believe that through better mapping of cancer biology with the help of computational tools, we can get ever closer to very significantly improving the life outcomes of millions who are or will be diagnosed with this malady.