EMPLOYING QUANTITATIVE SYSTEMS PHARMACOLOGY TO CHARACTERIZE DIFFERENCES IN IGF1 AND INSULIN SIGNALING PATHWAYS IN BREAST CANCER

by

Cemal Erdem

B.S., Middle East Technical University, 2010

M.S., Koc University, 2012

Submitted to the Graduate Faculty of

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Cemal Erdem

It was defended on

March 21, 2018

and approved by

Adrian V. Lee, PhD, Professor, Department of Pharmacology and Chemical Biology

Russell S. Schwartz, PhD, Professor, CMU Department of Biological Sciences

James R. Faeder, PhD, Associate Professor, Department of Computational & Systems Biology

Dissertation Co-Advisor: Timothy R. Lezon, PhD, Assistant Professor, Department of Computational & Systems Biology

Dissertation Co-Advisor: D. Lansing Taylor, PhD, Professor, Department of Computational & Systems Biology

EMPLOYING QUANTITATIVE SYSTEMS PHARMACOLOGY TO CHARACTERIZE DIFFERENCES IN IGF1 AND INSULIN SIGNALING PATHWAYS IN BREAST CANCER

Cemal Erdem, PhD

Insulin and insulin-like growth factor I (IGF1) have been shown to influence cancer risk and progression through poorly understood mechanisms. Here, new insights on the mechanisms of differential MAPK and Akt activation are revealed by an iterative quantitative systems pharmacology approach. In the first iteration, I combined proteomic screening with computational network inference to uncover differences in IGF1 and insulin induced signaling. Using reverse phase protein array of 21 breast cancer cell lines treated with a time course of IGF1 and insulin, I constructed directed protein expression networks using three separate methods: (i) lasso regression, (ii) conventional matrix inversion, and (iii) entropy maximization. These networks, named here as the time translation models, were analyzed and the inferred interactions were ranked by differential magnitude to identify pathway differences. The two top candidates, chosen for experimental validation, were shown to regulate IGF1/insulin induced phosphorylation events. Both of the knock-down perturbations caused phosphorylation responses stronger in IGF1 stimulated cells compared with insulin. Overall, the time-translation modeling coupled to wet-lab experiments has proven to be powerful in inferring differential interactions downstream of IGF1 and insulin signaling, in vitro. In the second iteration, mechanistic representation of IGF1 and insulin dual signaling cascades by a set of ODEs is generated by rule-based modeling. The mechanistic network modeling provided a framework to elucidate experimental targets downstream of two receptors, which were treated as indistinguishable in previous models. The model included cascades of both mitogen-activated protein kinase

(MAPK) and Akt signaling, as well as the crosstalk and feedback loops in between. The parameter perturbation scanning employed for seven different models of seven cell lines yielded new experimental hypotheses on how differential responses of MAPK and Akt originate. Complementary to the first iteration, the results in this part suggested that regulation of insulin receptor substrate 1 (IRS1) is critical in inducing differential MAPK or Akt activation. Compensation and activating feedback mechanisms collectively depressed the efficacy of anti-IGF1R/InsR therapies. With the quantitative systems pharmacologic approach, the networks of signal transduction constructed in this thesis are aimed to discern novel downstream components of the IGF1R/InsR system, and to direct patients with suitable tumor subclasses to efficient personalized clinical interventions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

First and foremost, I would like to start by thanking my two advisors, Lans Taylor and Tim Lezon. Lans has been my mentor for the last five years and I especially thank him for his continuous encouragement and support for me to become a complete computational biologist, with both computational and experimental skill sets. Tim, no less, has also been there for me and guided me to become a better scientist. I always enjoyed our talks on how should we do things, and how they are wrong. I am truly grateful to them, enlightening my path through these years. I would also like to thank each of my committee members, Adrian Lee, Russell Schwartz, and Jim Faeder. Dr. Lee has been there since the beginning of my thesis, with his contribution of the initial dataset as well as helping me with additional experimental work. My deepest thanks go to Dr. Lee, his graduate student Alison, and the rest of his lab members for teaching me my way around the wet-lab. Drs. Schwartz and Faeder helped shape my work with their valuable comments and suggestions. Although I did not have any lab members, members of CSB and UPDDI have been always there with their smile and helping hands. I sincerely thank Celeste Reese and Laura Vollmer of UPDDI for their help in planning and doing additional experiments. Last but not least, I humbly thank my parents, my sister, my nephew, my brother, and my wife Hülya, and my son Efe for being a part of my life, success, and love. Their presence and patience with me was all I needed to endure the graduate life. This dissertation is dedicated to them.

1.0 INTRODUCTION

1.1 BREAST CANCER

Breast cancer (BC) is the most common type of cancer in women in United States, excluding certain forms of skin cancer[1]. According to the current statistics, one in eight women will be diagnosed with breast cancer during her lifetime. In addition to being common, it is also one the leading of causes of death from cancer among women. Men, too, are diagnosed with BC, almost one percent of the annually reported cases were of men in USA. The study of breast cancer prevention, diagnosis, and therapy is imperative, and National Institutes of Health has allocated almost 715 million USD on average each year in the period 2011-2014 for breast cancer research[2].

Currently, there are five drugs approved by U.S. Food and Drug Administration (FDA) for use to prevent breast cancer and 71 approved drugs for treatment of the disease, as reported by the National Cancer Institute[3]. There are also six drug combinations used to treat breast cancer. One reason that so many drugs are required is the fact that breast cancer is highly heterogeneous and molecular subtypes are quite diverse. The subtypes based on the genetic markers are basal, luminal, claudin-low, and normal-like [1]. One recent study, however,

[1] cdc.gov
[2] report.nih.gov
[3] cancer.gov

concluded that there are at least 10 subtypes based on copy number aberrations in tumor samples [2]. Another classification system is based on the differential expression levels of certain receptors. Such subclasses are triple-negative, hormone receptor positive, and HER2-amplified (human epidermal growth factor receptor 2) [1]. The PAM50 gene expression clustering is used to determine the molecular subtypes of the breast tumors [3].

Apart from these, in patients with familial or hereditary breast cancer, the mutation status of BRCA1 and BRCA2 (breast cancer 1-early onset & breast cancer 2-early onset) genes account for cumulative risks of 60% and 55% for developing breast cancer by the age of 70, respectively [4-6]. Although hormone receptor antagonist treatments, like tamoxifen [7], are shown to decrease mortality from BC, drug therapy resistance mechanisms are becoming one of the major obstacles to effective BC treatment. Such mechanisms are shown to develop mostly by up-regulation of other receptor tyrosine kinase (RTK) cascades [8]. For instance, seventy percent of BC cells express estrogen receptor (ER), for which endocrine therapy resistance has become one major problem [9]. Thus, additional mechanisms involved in tumor growth and proliferation as well as resistance pathways are of vital importance to explore further.

1.2 INSULIN AND INSULIN-LIKE GROWTH FACTOR I (IGF1)

Under normal physiological conditions, IGF1 is a proliferation and development factor whereas insulin has a major role in glucose homeostasis. IGF1 mainly mediates the effects of the growth hormone (GH) and is also important for mammary development [10, 11]. The growth promoting effects of IGF1 include stimulation of proliferation, differentiation and protein synthesis [12]. Upon GH stimulation, IGF1 is produced mainly by liver (75% of circulating IGF1), but it is also

produced by mammary connective tissue stromal cells to increase the signaling response [6]. Thus, IGF1 acts via paracrine and autocrine mechanisms in addition to its endocrine regulatory role. IGF2 is also a growth promoter, which is parentally imprinted where normally only the paternal allele is expressed [13]. IGF1 also regulates the cell cycle events through modulation of cyclins, cyclin-dependent kinases and cyclin-dependent kinase inhibitors [14]. There are six IGF-binding proteins (IGFBP1-6) [15, 16] that bind IGF1 and IGF2 with a higher affinity than the IGF1 receptors. Ninety-nine percent of circulating IGF1 is bound to IGFBPs and 80% thereof to IGFBP-3 [12]. **Figure 1** depicts the binding relationships between the three ligands, two receptors, hybrid receptors, and IGFBPs.

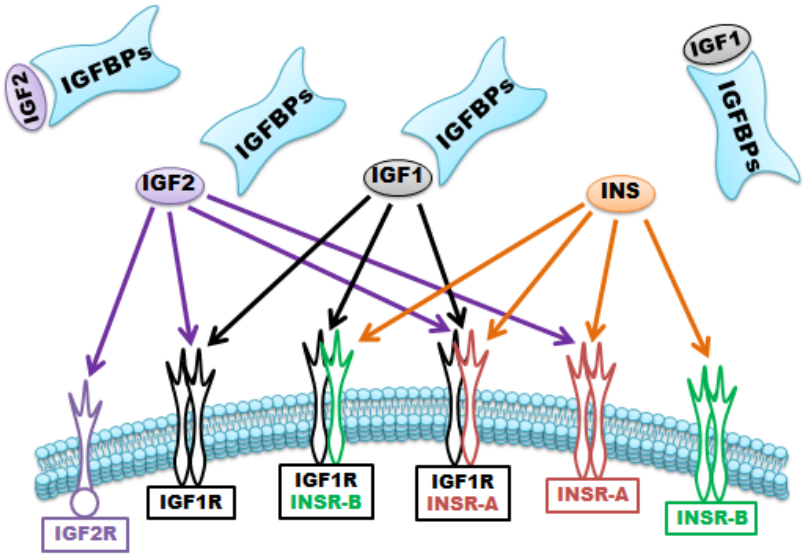

Figure 1. The ligand-receptor binding preferences. In addition to the canonical binding partners, ligands bind to other receptors as well as specific binding proteins. Figure is adapted from [6, 17].

Insulin-like Growth Factor Receptor Type I (IGF1R) and Insulin Receptor (InsR) are receptor tyrosine kinases. These RTKs have 60% amino acid sequence homologs overall. Moreover, they are 84% identical at the kinase domain [6, 9]. The receptors are heterotetramers, or rather a dimer of heterodimers, with two α and two β subunits. Beta subunits contain the

intracellular kinase domains. The α-subunits span the extracellular ligand binding domains. IGF1R has two canonical ligands, insulin-like growth factor 1 (IGF1) and insulin-like growth factor 2 (IGF2), and the InsR has insulin (Ins) as the formal agonist [9, 18]. IGF1R has a homolog called IGF2R, which is shown to bind to and sequester IGF2, with no kinase domains and no downstream intracellular signal transduction reported to date [18]. There are two isoforms of insulin receptor: InsR-A, which is functional in fetal tissues and in human cancer cells, and InsR-B, which is expressed in adults. InsR-A has similar affinity for IGF2 and insulin [6, 13]. IGF1R and InsR can also form hybrid receptors, with one α-β pair from each. These hybrid receptors show differential affinity for the three ligands [18]. Then, depending on the constituting receptor isoforms and the ligand, they initiate different downstream signaling cascades [13, 18].

In BC cells, IGF1R overexpression and elevated IGF1 and IGF2 levels are observed. IGF1 induced by GH is required for mammary terminal end bud (TEB) formation and ductal morphogenesis, where TEBs themselves also express IGF1, IGF1R and IRS signaling adaptors. Moreover, IGF1R content in BC tumors is 14-fold higher than that in normal tissue [13]. Proof-of-principle experiments revealed that IGF receptor blockage can block tumor growth and metastasis. Inhibition of the pathway showed remarkable antineoplastic activity also in pre/early clinical trials [17, 19]. Although it is overexpressed to a lesser extent than other oncogenes, IGF1R is expressed in all BC cells [9]. In estrogen receptor positive BC cells, inhibition of IGF1R blocks double stranded DNA break repair, resulting in an increase of radiation induced apoptosis [13]. IGF1 and estrogen increase BC cell proliferation in MCF7 cells [13]. There are also studies of patient cohorts linking raised blood levels of IGF1 to increased cancer risk [19].

As IGF1R activity is dependent on ligand binding, simple overexpression of this receptor does not mean enhanced activation [9]. The ligand binding promotes a cascade of phosphorylation reactions, including proteins in MAPK (mitogen-activated protein kinase) pathway and IRS1 (insulin receptor substrate 1) (**Figure 2**). The IGF binding results in auto-phosphorylation of intracellular kinase domain of the receptor, resulting in adaptor protein phosphorylation. The PI3K/Akt module is also activated leading to escape from apoptosis [9]. InsR signaling, however, is directed more at the Akt cascade and glucose uptake receptor complex transport to the plasma membrane. A phospho-proteomic study downstream of insulin and InsR showed that these two cascades are indeed triggered [20].

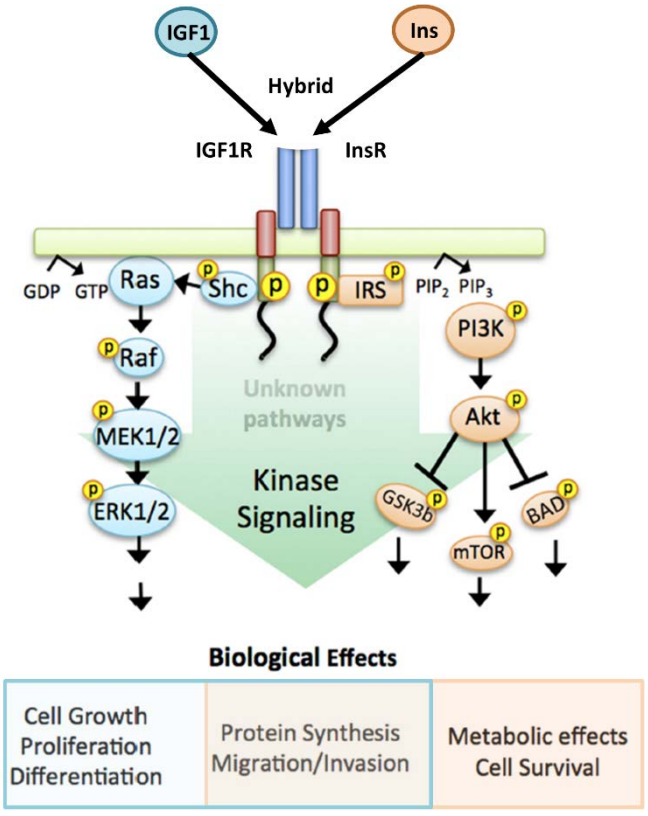

Figure 2. The IGF1R/InsR signaling pathway with Akt and MAPK cascades. The figure is adapted from [21].

Both IGF1R and InsR are correlated with higher breast cancer risk and their roles in neoplasia were reviewed by Pollak in [17, 19] and by others in [14, 22-25]. Inhibiting IGF1R to block tumor growth in breast cancer is pursued actively and there are currently many drugs at different phases of clinical trials [26, 27]. The available data from the first clinical trials have been positive enough to launch several phase II & III trials in various human cancers. There have been several cases of responses in phase I and II trials with anti-IGF1R antibodies, but these agents will most likely not be useful in unselected patient populations [26].

On the structural side, the crystal structures of the full-length receptors are still not available, although there are some *partial* structures of ligand or inhibitor bound states. Two recent studies suggested that the extracellular domains (ECD) of the apo-receptor exert a physical force to keep the intracellular kinase domains apart from each other, enough to prevent auto-phosphorylation [28, 29]. Ligand binding then induces a conformational change that lets the transmembrane and kinase domains to interact and auto-phosphorylate. The changes upon ligand binding are studied by Houde and Demarest in [30]. Kiselyov et al. used modeling approaches to re-capture available ligand binding dynamics of insulin in [31]. This study only considered the ligand-receptor binding events, ignoring the fact that downstream elements of the transduced signal also affect available receptors on the cell surface. These context-dependent activities of IGF1 and insulin in normal or diseased states need to be explored further.

Overall, an understanding of IGF1R signaling is important because it will guide the incorporation of appropriate molecular markers into clinical trial design. This will be a key to patient stratification. Characterizing and representing the pathway will enable generating and testing hypotheses, as well as studying the pathway dynamics and responses.

1.3 QUANTITATIVE SYSTEMS PHARMACOLOGY (QSP)

As the information on disease mechanisms and successful/failed clinical trials have amassed over the last decade, understanding drug mechanisms of action and studying emergence of resistance have become imperative. Based on this observation, a white paper released after a workshop at the National Institutes of Health (NIH) defined a new field of study called *Quantitative Systems Pharmacology* (QSP) [32]. QSP is a systems level approach that aims to combine experimental and clinical tools with computational analyses to elucidate new drugs and to better understand the biology of the drug mechanisms.

Classical drug discovery workflow is target-centric, whereas QSP is network-centric. By analyzing large experimental and clinical data with advanced computational methods and models, QSP studies mechanisms of drugs and test new therapeutic hypotheses *in silico*. By doing so, it offers an alternative to low-efficiency classical drug development process.

The analytical tools employed in QSP are used to integrate patient derived experimental models and dynamical response data to offer personalized precision medicine. Cellular heterogeneity is also captured by QSP computational models. The overall approach will then be utilized to design better treatment schedules for selected set of patients, either as mono or combination therapies.

Computational modeling portion of QSP includes integration of different "-omics" data, drug-target interaction prediction, and comprehensive network modeling. The computational models drive construction and analyses of dynamical networks within a large range of spatial and temporal scales. The methods include both statistical and mechanistic approaches, each with different ways of extracting information from the available data to generate new hypotheses to test for the next iteration of experimentation-computation cycle of QSP platform [33-38].

1.4 STATISTICAL MODELING

In order to gain insight into cellular signaling networks, a method of tracking the evolution of multiple cellular features in time is required. Advances in experimental procedures and instruments have led to the development of proteomics approaches such as mass spectrometry [20] and live cell imaging [39] that enable the quantification of the levels of multiple proteins across large numbers of samples. Similarly, reverse phase protein array (RPPA) is a high-throughput technique that provides quantification of expression levels of total proteins as well as specific phosphorylated forms [40]. Researchers have already used the power of RPPA toward biomarker prediction [41], classification of responses to diverse inputs [42], study of drug responsiveness [43] , and molecular network generation [44].

High throughput data sets have been the subject of computational modeling for well over a decade. Early linear models applied to gene expression data were able to extract clear signals from superficially noisy time-series data. Conventional tools and methods for studying temporal dynamics of genomic data [45] and pathways [46] often rely on sophisticated machine learning techniques that provide high accuracy at the cost of clarity. The utility of simpler linear models is that their results can be interpreted in terms of known biology. Whereas complex models may tend to obfuscate their inner workings, linear models connect components directly to each other in an unambiguous fashion. Application of linear elastic network models (ENM) in structural biology, for instance, enables one to analyze very large systems by lowering complexity [47, 48]. Coarse-grained motions and equilibrium dynamics can be recapitulated with very high accuracy using reductionist models. These and other studies [49-51] have shown that simple, cheaper, and faster approaches are valuable in beginning to understand and apply current knowledge to drive better models of biological systems.

Over the last few decades, the progression from static to time-series data has enabled researchers to infer rates of change of state variables, to find causal interactions in-between, and to study collective system dynamics. The algorithms and protocols to analyze the time-dependent data generated have been expanding in parallel, especially with the invention and development of microarray technology, and a variety of tools and methods are available for studying temporal dynamics [45]. Although most of the techniques are based on the genomic data, they are applicable to other types of temporal data as well. These methods span a large set of approaches, including both linear and non-linear models [52-59]. Linear models, like ones performed in studies focusing on construction of circadian gene regulation or interaction network topologies [57, 60], serve as tools of easy implementation and interpretation.

1.5 MECHANISTIC MODELING

Computational systems biology has gained momentum in application and incorporation into the experimental and clinical workflows mostly after the beginning of the new millennium [61-63]. Study of signaling pathways can employ different methods of formulation comprising one or a combination of logical, probabilistic, and mechanistic approaches. These approaches can further be formulated deterministically or stochastically.

1.5.1 Ordinary Differential Equations

Use of Ordinary Differential Equations (ODEs) is a common strategy because it is particularly easy to implement and interpret. The dynamics of time series data can then be readily studied

with different types and scales of input-output regimes [64, 65]. Simulating ODE systems deterministically provides handling of large systems with low complexity. Fast run times and well-studied analyses types and properties make this type of modeling a first choice.

With ODE modeling, detailed signaling cascades of different pathways (i.e. JAK/STAT [66-68] and EGFR [69]), or coarse-grained molecular mechanisms with systemic parameters can also be studied [70-72]. Furthermore, disease spread and inflammation response have also been explored by modeling [73, 74]. In these studies, building over earlier models is a common strategy. For instance, from models of ligand-receptor interactions to the examination of the role of EGFR in cancer, existing models are exhaustively studied over and modified accordingly in newer models [75-78].

1.5.2 Rule-based Modeling (RBM)

Rule-based modeling utilizes simple rules to construct networks of interactions. The rules are written in languages such as the BioNetGen [79] or Kappa [80]. Model nodes are defined with specific domains, allowing for complex and site-specific interactions. Every biochemical event can be defined by the rules. With the RBM, signaling network interactions can be represented in detail and the combinatorial complexity arising in such systems is handled by the software, by omitting the domains of reactants that do not affect the specificity of a rule. This allows reduction of number of rules to be written as opposed to the number of components can be generated [81]. The reactions can also be written in PySB, a python framework, and the model can be converted into previously mentioned languages [82].

The network defined in BioNetGen and RuleBender software [83] can be simulated both deterministically and stochastically [84]. The network models generated can also be exported into other file formats like SBML [85] and MATLAB (The MathWorks, Inc.) files.

1.6 OVERVIEW

Insulin and insulin-like growth factor I (IGF1) influence cancer risk and progression through poorly understood mechanisms. Multiple investigations have shown association of these two hormones in cancer progression and cell proliferation [17, 19]. The literature, however, lacks enough modeling efforts to delineate differences between IGF1R and InsR signaling.

To better understand the roles of insulin and IGF1 signaling in breast cancer, I combined proteomic screening with computational network inference to uncover differences in IGF1 and insulin induced signaling. Different linear models of directed time translation (TT) networks are constructed from available experimental data. The final TT networks are analyzed to find novel temporal differences between IGF1 and insulin stimulation. The approach selects a set of rationally ranked pathway protein candidates from which perturbation experiments are carried out to check the validity of predictions. These efforts are summarized in **Chapter 2: Statistical Modeling**.

The next step explained here is the mechanistic representation of IGF1 and insulin dual signaling cascades by a set of ODEs, generated by rule-based modeling. I studied the response of the system under different conditions to understand the observed phenotypes in cells of various subtypes. The computationally scalable reverse-engineered models of cellular networks introduced here provide a framework to elucidate experimental targets of pharmacological

importance in a cost effective way. The mechanistic approach is explained in **Chapter 3: Mechanistic Modeling**.

Overall in this thesis, new insights on the mechanisms of differential MAPK and Akt activation are revealed by an iterative systems biology approach. In the second chapter (**Chapter 2: Statistical Modeling**), the statistical computational modeling efforts will be explained. The chapter will also provide experimental evidence for the validation of modeling predictions. The third chapter (**Chapter 3: Mechanistic Modeling**) will go over the computational mechanistic modeling scheme employed. This chapter will also report the experimental validations of particular predictions, in different cell lines tested. The thesis will then end with **Chapter 4: Conclusions**, and a discussion of some future directions.

2.0 STATISTICAL MODELING

The contents of this chapter are from a peer-reviewed publication, where I was the first author [86]. The contents are re-organized for my thesis.

2.1 INTRODUCTION

As mentioned in the previous chapter (**Introduction**), the insulin and IGF1 receptors are tyrosine kinases that are expressed in almost all types of cells [87]. Signaling through InsR and IGF1R initiates a phosphorylation cascade that drives cell growth and proliferation [10-12, 18, 24]. Overexpression of these receptors is correlated with higher breast cancer risk [14, 17, 19, 23-25] and has been shown to influence tumorigenesis, metastasis, and resistance to existing forms of cancer therapy [17, 88]. IGF receptor blockade can slow tumor growth and metastasis, but the receptor has proven to be difficult to target specifically [26, 27, 89]. A confounding factor in developing therapies targeting these receptors is their high sequence (~60%) and structural homology. IGF1R and InsR are able to form functional hybrids, and each can partially compensate for the loss or suppression of the other [8, 9, 13, 90-92]. Moreover, it has been shown that IGF1R signaling is one mechanism of resistance to conventional hormonal therapy [7, 93-96]. Understanding the relationships between IGF1R and InsR signaling cascades and

their combinatorial role in cancer is crucial to developing better diagnostics and personalized treatments for cancer.

Here, I used a simple yet intuitive linear model to explore the time evolution of protein expression profiles in breast cancer cell lines stimulated with IGF1 or insulin. Using the lasso algorithm trained on the time-series RPPA expression profiles, I constructed temporal networks of effective interactions between proteins. Analyzing the resultant networks, I found novel temporal differences in cellular response to IGF1 and insulin stimulation. My approach generated a set of predictions that are validated experimentally by our collaborator's lab. The contents of this chapter are published already.

In short, I employed an iterative experimentation-computation workflow; starting the first iteration with a dataset of time-series protein expression profiles in breast cancer cell lines. The computation step included utilization of different linear models, where directed networks of time translation (TT) are constructed and analyzed to find novel temporal differences between IGF1 and insulin stimulation conditions. With this approach, I have selected a set of rationally ranked protein-pair candidates from which perturbation experiments are carried out to check the validity of the predictions.

The performance and validity of the TT models are tested extensively. The first prediction was that phospho-ACC (pACC) levels have a greater effect on the IGF1-induced phosphorylation of MAPK than on insulin-induced MAPK phosphorylation after 5 and 10 minutes of stimulation. We verified the differential influence of ACC using knock-down experiments (please see the manuscript). The immunoblots showed that upon ACC1/2 siRNA knock-down, phospho-MAPK levels increased. Compared to control cells, IGF1 stimulated cells with no active ACC showed 40% higher ($p<0.05$) MAPK phosphorylation level. No significant

change ($p>0.05$) of relative pMAPK level was observed in insulin stimulated cells, suggesting a higher proliferative IGF1 response compared to insulin induction. Our results are consistent with the previous results such that a decrease in ACC level promotes cell survival.

Another interesting prediction of our model was the differential effect of E-Cadherin expression levels on pAkt level induced by IGF and Ins. This is the top differential effect candidate for the 30 min time point. Leading to higher growth signaling in stimulated cells, the loss of E-Cadherin induced almost 2.5-fold ($p<0.005$) increase in relative pAkt/Akt level in response to IGF1, where the increase was to 2.7-fold ($p<0.05$) with 10 min insulin stimulation. Although the increase in insulin stimulation seems higher, the magnitude of increase in IGF1 stimulation case is higher and is more significant.

There are other interesting perturbation candidates listed, which are yet to be confirmed. For instance, the models predicted that knock-down of ATM protein will have a larger effect on insulin stimulated Akt phosphorylation (both phospho-sites) than in IGF1 stimulation. This effect might be clinically important because at no ATM presence, patients are more sensitive to radiotherapy. Overall, the information gained at this step is also utilized in the next step to learn more on the mechanistic side of these effects.

2.2 THE INITIAL DATASET OF RPPA

2.2.1 Cell culture and reverse phase protein array

The initial dataset used in this study is deposited in NCBI's Gene Expression Omnibus (GEO) [97] database with accession number GSE80233 [86]. The data were generated at the RPPA

Core Facility (MD Anderson Cancer Center, TX). In short, twenty-one breast cancer cell lines (**Table 1**) were cultured in DMEM, DMEM/F-12, RPMI, L15 or McCoys 5A media supplemented with 10% FBS or horse serum according to the conditions provided by Neve *et al.* [98]. Cells were plated in triplicate 6cm dishes per treatment time point and then switched to serum-free medium (SFM). Cells were serum starved for 24 hours, then treated with 10nM IGF1 (Novozymes) or insulin (Sigma) for 5m, 10m, 30m, 6h, 24h, and 48h. Cells treated with vehicle (Vhc), the serum-free control were collected at 5m, 24h, and 48 hr. Cell lysates were harvested by washing in ice-cold PBS twice, and lysing immediately in RPPA lysis buffer with protease inhibitor (1% Triton X-100, 50mM Hepes, 150mM NaCl, 1.5mM $MgCl_2$, 1mM EGTA, 100mM NaF, 10mM NaPPi, 10% Glycerol, 1mM Na_3VO_4, 1mM PMSF, 10μg/ml Aprotinin). Lysates were sonicated and centrifuged at 14,000 rpm at 4^oC for 10 min. Protein concentrations were determined by BCA. Then, lysates were mixed with 4X SDS/2-ME sample buffer to have final concentration of 1μg/ml. The lysates were boiled for 5 min and diluted two-fold five times. The diluted lysates were arrayed on nitrocellulose-coated FAST slides (Whatman, Inc.). Each slide was probed with a primary antibody and a biotin-conjugated secondary antibody. Levels of 134 proteins or phospho-proteins were measured using corresponding antibodies. The signal, amplified by DakoCytomation-catalyzed system (Dako), was visualized by DAB colorimetric reaction. The slides were scanned and quantified using Microvigene (VigeneTech) software. Supercurve fitting was used for loading correction of each sample [99].

2.2.2 Quality control of RPPA antibodies

An internal quality control was routinely employed on the specificity and validity of the antibodies in the RPPA data. The process included a condition of obtaining a single or dominant

Table 1. The cell lines used for the RPPA data and their subtype information.

No	Name	Transcriptional subtype	HER2 status
1	AU565	Luminal	amplified
2	BT-20	Basal	-
3	BT474	Luminal	amplified
4	Cama-1	Luminal	-
5	HCC60	Basal	-
6	HCC1954	Basal	amplified
7	HCC1569	Basal	amplified
8	HS758T	Claudin-low	-
9	MCF7	Luminal	-
10	MCF10A	Normal-like	-
11	MDA-MB-134	Luminal	-
12	MDA-MB-231	Claudin-low	-
13	MDA-MB-361	Luminal	amplified
14	MDA-MB-415	Luminal	-
15	MDA-MB-435	-	-
16	MDA-MB-453	Luminal	-
17	MDA-MB-468	Basal	-
18	SK-BR-3	Luminal	amplified
19	T47D	Luminal	-
20	UACC812	Luminal	amplified
21	ZR75-1	Luminal	-

band for the antibody in question on western blots ran for the dynamic range conditions of RPPA data. Then, a Pearson correlation coefficient of 0.8 between RPPA and WB results was used as cutoff to determine if the antibody is specific enough for the intended target. The RPPA Core Facility (MD Anderson Cancer Center, TX) also considers intra- and inter-slide reproducibility for the antibodies. They update their list of available reliable antibodies regularly based on experiences in hundreds of conditions and cell lines. Finally, 26 of the antibodies were labeled not-specific enough under the specific experimental conditions, and are excluded from all subsequent analyses.

2.2.3 The data

The dataset contains expression levels for 134 proteins, either in total or in a specific phosphorylated form, in 21 breast cancer cell lines stimulated in biological triplicates with IGF1 or insulin for 0 – 48hrs (see **Table 1** and [1] for information on cell line subtypes). A cursory look at the data (**Figure 3**) reveals that IGF1 stimulation induces phosphorylation of IGF1R at Y1135 in all cell lines. Half of the cell lines also exhibit a higher phosphorylation rate of Akt and its downstream protein targets, including ribosomal protein S6. The six time-point data for each cell line show expected responses, with pAkt levels peaking around 5-10 min and decreasing gradually afterwards. The cell lines MCF7, MCF10A, T47D, and ZR751 were found to be highly IGF1 responsive. The pIGF1R_Y1135 antibody also captures corresponding phospho-InsR residue (tyrosine 1150/1151), which is also upregulated upon insulin treatment of almost all cell lines (**Figure 3**).

2.2.4 Statistical Analysis

Standard one-way ANOVA was performed on the RPPA data, considering each condition separately. The analyses were done for the triplicate measurements of a cell line, with one stimulation condition at a specific time point. All of the simulations and analyses were run in Matlab, versions 2014a and 2015a (The MathWorks, Inc.).

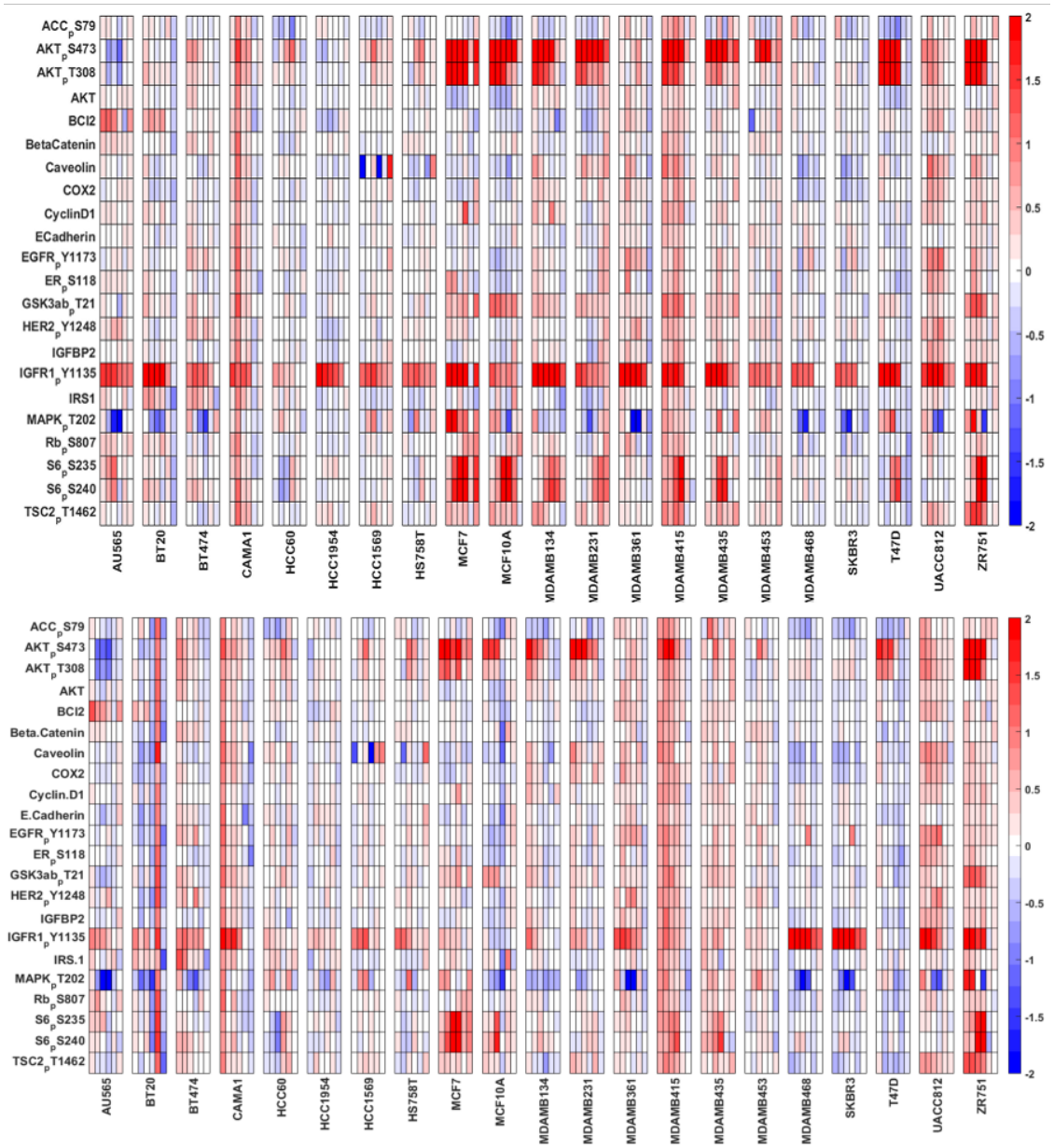

Figure 3. Time-course (phospho)-protein expression profiles of 21 breast cancer cell lines in response to IGF1 (top) and insulin (bottom). The values represent the difference in log2 expression levels (stimulated – serum starved). IGF1R phosphorylation is induced in all cell lines in response to IGF1, and InsR in most of the cell lines in response to insulin (pIGF1R_Y1135 antibody captures both receptors). RAS/MAPK and PI3K/Akt cascades are regulated downstream of these receptors. Six columns are shown for each cell line, each column representing one time point (5, 10, 30 min, 6, 24, 48 hrs). Red color corresponds to upregulation and blue color represents downregulation of the proteins compared to serum starved cells. Row headings indicate proteins or phospho-proteins with phosphorylation sites indicated after the subscript p.

2.3 METHODS AND THE COMPUTATIONAL MODELS

2.3.1 Construction of linear statistical models

2.3.1.1 Data pre-processing

To reduce noise in the initial dataset, I selected only proteins with large expression-level variance. The variance of each (phospho)-protein expression across all conditions are calculated and I identified a variance cutoff that separates the antibody set into 43 high-variance proteins (**Table 2**) and 65 low-variance proteins. The 0.42 cutoff is empirically decided and is stable (**Figure 4**), so that small changes in the cutoff value do not alter the number of proteins retained. I also showed that the data from triplicates are more similar to each other than the mean of triplicates across different condition. The Kolmogorov–Smirnov distance between the standard deviations of the triplicates and the standard deviation of the mean of triplicates across conditions is 0.98 (**Figure 4**). In addition, when one-way ANOVA is performed on the filtered RPPA data for every condition separately, the results indicated that the mean of triplicate measurements across proteins are significantly different. As a result, I only used the mean of triplicate experiments in all subsequent analyses. The individual time point datasets (each of size [43 proteins] × [21 cell lines]) were then column centered, row-centered, and row-normalized.

2.3.1.2 Different computational model construction approaches

The time-translation matrices constructed here satisfy the equation $\mathbf{x}(t_2) = \mathbf{T}\,\mathbf{x}(t_1)$, where the expression vector $\mathbf{x}(t)$ contains the levels of P proteins measured at time point t. The time translation matrix T is a square matrix that transforms protein expression levels from an early time point (t_1) to a later one (t_2). As any matrix that exactly transforms all the data from one

time point to another is considered a valid time translation matrix, the problem of defining T is underdetermined if the number of measured proteins exceeds the number of samples. We investigated three different methods of calculating T from the data, each representing a constrained optimization.

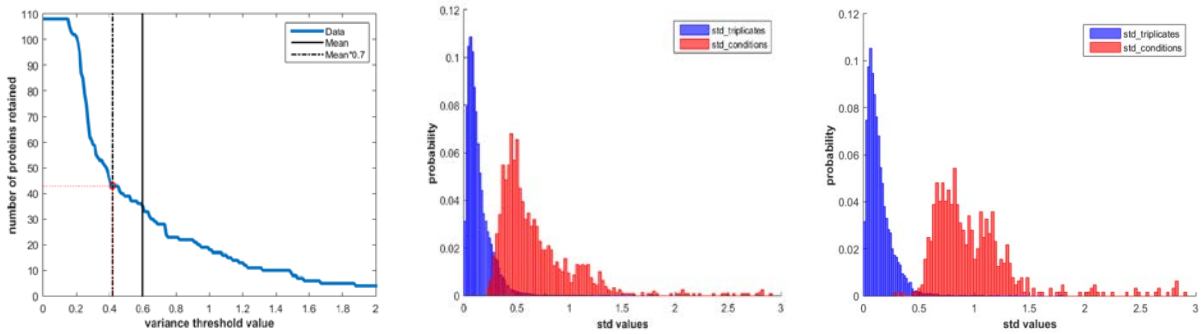

Figure 4. (Left) The empirical determination of high/low variance proteins in the RPPA dataset. High-variance proteins occur above the first plateau. The solid blue line shows the number of proteins with raw variance greater than a threshold value x. The solid black line represents the mean of the variance values calculated for the 108 validated antibodies. The dotted black line represents the empirically determined cutoff value of 0.42. Red dotted line represents the final number of proteins designated as highly variant (n=43), the rest are named as low-variant proteins (n=65). (Middle) The comparison of standard deviation values of triplicates to the standard deviation of the means of triplicates across cell lines. The panel is shown for 108 validated proteins whereas the (right) panel depicts the results for the 43 proteins used in modeling. The use of only the high-variant proteins (n=43) even diminishes the overlapping regions of deviation values. The results shown here recapitulate that the triplicate experiment results are closer to each other than they are to results across different conditions. The Kolmogorov – Smirnov distance of the histograms are 0.8953 and 0.9821 for 108 and 43 protein cases, respectively.

The Pseudo-inversion formalism

If we compile the N expression vectors at time t_1 into the $P \times N$ matrix $\mathbf{X}(t_1)$ and construct $\mathbf{X}(t_2)$ similarly, then $\mathbf{T}$ is given by

$$\boldsymbol{T_C} = \boldsymbol{X}(t_2)\boldsymbol{X^{-1}}(t_1) \qquad \text{Eqn. 1}$$

where $\mathbf{X^{-1}}(t_2)$ is the pseudoinverse of $\mathbf{X}(t_2)$ when *N*<*P*, as in the present case. The time translation matrix calculated this way minimizes the contribution to $\mathbf{T_C}$ from the nullspace of $\mathbf{X}(t_2)$. The steps of model, called *SVD* modeling, construction are depicted in **Figure 5a**.

The Entropy Maximization formalism

We can alternatively assume a distribution that maximizes entropy subject to the constraints of the data. We write the equation of probability for the expression value distributions:

$$p(\boldsymbol{x}(t_1), \boldsymbol{x}(t_2)) = \left({}^{1}/_{Z}\right) exp\left\{-\frac{1}{2}\left(\boldsymbol{x}^T(t_1)\, \boldsymbol{x}^T(t_2)\right)\boldsymbol{M}\begin{pmatrix}\boldsymbol{x}(t_1)\\ \boldsymbol{x}(t_2)\end{pmatrix}\right\} \qquad \text{Eqn. 2}$$

where M is the (pseudo)inverse of the joint-covariance matrix and Z is the partition sum and equal to $(2\pi)^d|\mathbf{M}|^{\left(\frac{-1}{2}\right)}$. M is the matrix of effective interactions between components of the expression matrices. The steps followed for time-translation matrices construction from the M

Table 2. The list of 43 (phospho)protein markers after initial filtering step of the initial data.

1	**'ACC_pS79'**	**12**	**'Caveolin'**	**23**	**'HER2_pY1248'**	**34**	**'PR'**
2	**'AKT_pS473'**	**13**	**'CHK2'**	**24**	**'HER2'**	**35**	**'PTEN'**
3	**'AKT_pT308'**	**14**	**'COX2'**	**25**	**'HSP27'**	**36**	**'Rb_pS807'**
4	**'AKT'**	**15**	**'Cyclin.D1'**	**26**	**'IGFBP2'**	**37**	**'Rb'**
5	**'AR'**	**16**	**'Cyclin.B1'**	**27**	**'IGFR1_pY1135'**	**38**	**'S6_pS235'**
6	**'ATM'**	**17**	**'E-Cadherin'**	**28**	**'IRS1'**	**39**	**'S6_pS240'**
7	**'BCl2'**	**18**	**'EGFR_pY1173'**	**29**	**'MAPK_pT202'**	**40**	**'SRC'**
8	**'Beta-Catenin'**	**19**	**'EGFR'**	**30**	**'p53'**	**41**	**'STAT5'**
9	**'BIM'**	**20**	**'ER_pS118'**	**31**	**'p70S6K'**	**42**	**'TSC2_pT1462'**
10	**'cJUN'**	**21**	**'GATA3_BD'**	**32**	**'PKCα_pS657'**	**43**	**'VEGFR2'**
11	**'cMyc'**	**22**	**'GSK3ab_pT21'**	**33**	**'PKCα'**		

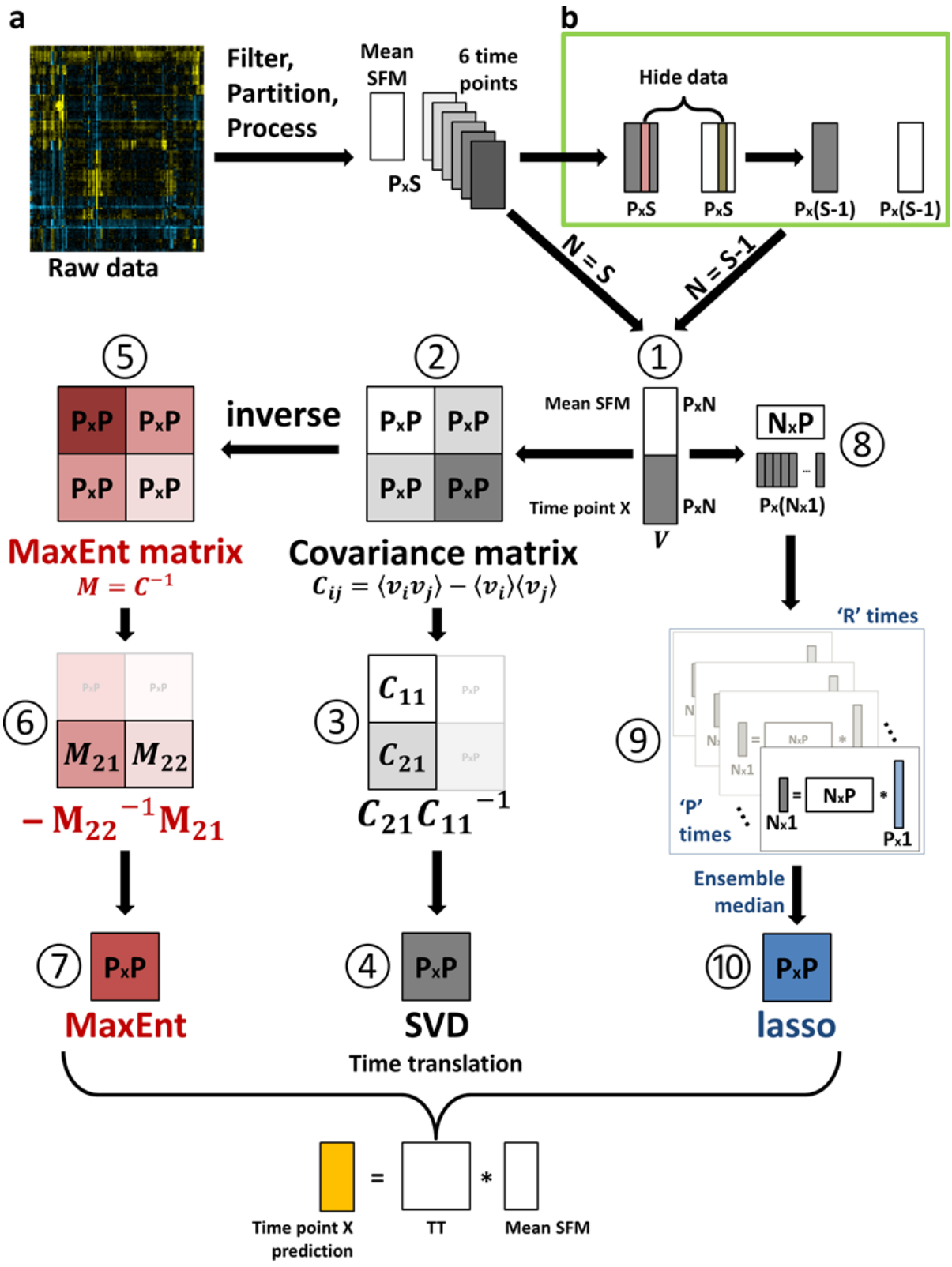

Figure 5. Three time-translation models are constructed. The overview of model construction steps is visualized. First, the dataset was filtered, row-centered, normalized and partitioned into individual time point matrices. Three

time-translation models were constructed for full (a) and leave-one-out analyses (b). (a) The construction steps of the three models. (1, 2, 3, 4) Construction of the TC model matrix using the joint covariance matrix, C. (1, 2, 5, 6, 7) Construction of the TM model matrix using entropy maximization. (1, 8, 9, 10) Construction of TL model matrix using the lasso algorithm. Here, the glmnet package was called for each of the P protein expression vector, and the median matrix from an ensemble of R networks was selected as the model matrix (see Methods for details). These three matrices represent three models of time translation. (b) A leave-one-out cross validation (LOOCV) analysis was performed to check the validity of our approach. The information for test cell lines (tan and red bars) was hidden and the model matrix was constructed via the scheme in (a). After the three time translation matrices are generated in either (a) or (b), they are used to calculate the expression profiles of the future time points using SFM condition data and corresponding T matrix.

matrix are also given in Figure 5a, route 1-2-5-6-7 (the MaxEnt model). The details of model construction are as follows:

Let $\mathbf{x_1} = \mathbf{x}(t_1)$ and $\mathbf{x_2} = \mathbf{x}(t_2)$ be P-component column vectors representing the expression levels of P proteins in one cell line. Then $\mathbf{v} = \begin{bmatrix}\mathbf{x_1}\\ \mathbf{x_2}\end{bmatrix}$ is a $2P$ component column vector of protein expression levels at times t_1 and t_2. We construct the $2P \times N$ matrix $\mathbf{V}$ from the $\mathbf{v}$ vectors, one apiece for each of the N cell lines. Following the procedure described in Lezon et al. [57], we find that the PDF that maximizes entropy while retaining the first and second moments of the data is

$$p(\mathbf{v}) = (2\pi)^{-P}|\mathbf{C}|^{-1/2}\exp\left\{\frac{-1}{2}(\mathbf{v}-\langle\mathbf{v}\rangle)^T\mathbf{C}^{-1}(\mathbf{v}-\langle\mathbf{v}\rangle)\right\}$$

where $\mathbf{C}$ is the two-timepoint covariance matrix with elements

$$C_{ij} = \langle v_i v_j\rangle - \langle v_i\rangle\langle v_j\rangle,$$

and the averages are taken from the data:

$$\langle v_i v_j \rangle = \frac{1}{N}\sum_{k=1}^{N} v_{ik} v_{jk} = \frac{1}{N}(\mathbf{V}\mathbf{V}^T)_{ij}$$

$$\langle v_i \rangle = \frac{1}{N}\sum_{k=1}^{N} \mathbf{V}_{ik}.$$

We define the $2P \times 2P$ interaction matrix **M** as the Moore-Penrose pseudoinverse of **C** (i.e., the inverse in the N-1 dimensional space spanned by the columns of **V**). It is symmetric, positive semi-definite and has the block form

$$\mathbf{M} = \begin{bmatrix} \mathbf{M_{11}} & \mathbf{M_{12}} \\ \mathbf{M_{21}} & \mathbf{M_{22}} \end{bmatrix},$$

where the elements of the $P \times P$ submatrix $\mathbf{M_{ij}}$ are the effective pairwise interactions between proteins at times t_i and t_i. Rewriting the PDF in terms of the individual time points,

$$p(\mathbf{x_1}, \mathbf{x_2}) = (2\pi)^{-P}|\mathbf{C}|^{-1/2}\exp\left\{\frac{-1}{2}[(\mathbf{x_1} - \langle \mathbf{x_1} \rangle)^T\ (\mathbf{x_2} - \langle \mathbf{x_2} \rangle)^T] \begin{bmatrix} \mathbf{M_{11}} & \mathbf{M_{12}} \\ \mathbf{M_{21}} & \mathbf{M_{22}} \end{bmatrix} \begin{bmatrix} (\mathbf{x_1} - \langle \mathbf{x_1} \rangle) \\ (\mathbf{x_2} - \langle \mathbf{x_2} \rangle) \end{bmatrix}\right\}$$

allows us to calculate the marginal PDF for $\mathbf{x_1}$,

$$p(\mathbf{x_1}) = \int d^P \mathbf{x_2}\ p(\mathbf{x_1}, \mathbf{x_2})$$

$$= (2\pi)^{-P/2}|\mathbf{C}|^{-1/2}|\mathbf{M_{22}}|^{-1/2}\exp\left\{\frac{-1}{2}(\mathbf{x_1} - \langle \mathbf{x_1} \rangle)^T\left[\mathbf{M_{11}} - \mathbf{M_{12}}\mathbf{M_{22}}^{-1}\mathbf{M_{21}}\right](\mathbf{x_1} - \langle \mathbf{x_1} \rangle)\right\},$$

and similarly for $\mathbf{x_2}$. Knowing the marginal distribution, we can calculate the conditional distribution of $\mathbf{x_2}$ given $\mathbf{x_1}$,

$$p(\mathbf{x_2}|\mathbf{x_1}) = \frac{p(\mathbf{x_1}, \mathbf{x_2})}{p(\mathbf{x_1})}$$

$$p(\mathbf{x_2}|\mathbf{x_1}) = (2\pi)^{-P/2}|\mathbf{M_{22}}|^{1/2}\exp\left\{\frac{-1}{2}(\mathbf{x_1} - \langle \mathbf{x_1} \rangle)^T\left[\mathbf{M_{12}}\mathbf{M_{22}}^{-1}\mathbf{M_{21}}\right](\mathbf{x_1} - \langle \mathbf{x_1} \rangle)\right.$$

$$\left. -\frac{1}{2}(\mathbf{x_2} - \langle \mathbf{x_2} \rangle)^T\mathbf{M_{22}}(\mathbf{x_2} - \langle \mathbf{x_2} \rangle) - (\mathbf{x_1} - \langle \mathbf{x_1} \rangle)^T\mathbf{M_{12}}(\mathbf{x_2} - \langle \mathbf{x_2} \rangle)\right\}.$$

The expected $\mathbf{x_2}$ associated with an $\mathbf{x_1}$ is

$$\langle \mathbf{x_2}(\mathbf{x_1}) \rangle = \int \mathrm{d}^P \mathbf{x_2}\, p(\mathbf{x_2}|\mathbf{x_1}) \mathbf{x_2}$$

$$= \langle \mathbf{x_2} \rangle - \mathbf{M_{22}}^{-1} \mathbf{M_{12}} (\mathbf{x_1} - \langle \mathbf{x_1} \rangle).$$

The matrix

$$\mathbf{T_M} = -\mathbf{M_{22}}^{-1} \mathbf{M_{21}}$$

translates $\mathbf{x_1} - \langle \mathbf{x_1} \rangle$ to $\mathbf{x_2} - \langle \mathbf{x_2} \rangle$ under the condition of constrained entropy maximization. We note that the same result is arrived at by maximizing the conditional probability $p(\mathbf{x_2}|\mathbf{x_1})$ via

$$\frac{\partial p(\mathbf{x_2}|\mathbf{x_1})}{\partial \mathbf{x_2}} = 0.$$

When the data are mean-centered, the time-translation matrix from the SVD approach, $\mathbf{T_C}$, can be written in terms of the $P \times P$ submatrices that compose the two-timepoint covariance matrix **C**. We can construct the $P \times P$ covariance matrices

$$\mathbf{C_{11}} = \frac{1}{N} \mathbf{X_1} \mathbf{X_1}^T$$

$$\mathbf{C_{21}} = \frac{1}{N} \mathbf{X_2} \mathbf{X_1}^T,$$

which are submatrices of

$$\mathbf{C} = \begin{bmatrix} \mathbf{C_{11}} & \mathbf{C_{12}} \\ \mathbf{C_{21}} & \mathbf{C_{22}} \end{bmatrix}.$$

Inverting $\mathbf{C_{11}}$, we find

$$\mathbf{C_{21}} \mathbf{C_{11}^{-1}} = \frac{1}{N} \mathbf{X_2} \mathbf{X_1}^T \left[N \left(\mathbf{X_1}^T \right)^{-1} \mathbf{X_1}^{-1} \right]$$

$$= \mathbf{X_2} \mathbf{X_1}^{-1},$$

which is $\mathbf{T_C}$ from Eqn. 1. When the two-timepoint covariance matrix **C** has full rank, it is invertible so that **MC** = **CM** = **1**. Writing the matrices in block form,

$$\begin{bmatrix} \mathbf{C_{11}} & \mathbf{C_{12}} \\ \mathbf{C_{21}} & \mathbf{C_{22}} \end{bmatrix} \begin{bmatrix} \mathbf{M_{11}} & \mathbf{M_{12}} \\ \mathbf{M_{21}} & \mathbf{M_{22}} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix},$$

allows us to find the following relationships among the component blocks:

$$\mathbf{1} = \boldsymbol{C_{11}M_{11}} + \boldsymbol{C_{12}M_{21}} \quad \text{Eqn. 3}$$

$$\mathbf{0} = \boldsymbol{C_{11}M_{12}} + \boldsymbol{C_{12}M_{22}} \quad \text{Eqn. 4}$$

$$\mathbf{0} = \boldsymbol{C_{21}M_{11}} + \boldsymbol{C_{22}M_{21}} \quad \text{Eqn. 5}$$

$$\mathbf{1} = \boldsymbol{C_{21}M_{12}} + \boldsymbol{C_{22}M_{22}} \quad \text{Eqn. 6}$$

where **1** and **0** are $P \times P$ identity and zero matrices, respectively. From Eqn. 3 and Eqn. 5, we find

$$\mathbf{M_{21}} = \left(\mathbf{C_{21}{C_{11}}^{-1}C_{12}} - \mathbf{C_{22}}\right)^{-1}\mathbf{C_{21}{C_{11}}^{-1}},$$

and from Eqn. 4 and Eqn. 6,

$$\mathbf{{M_{22}}^{-1}} = -\left(\mathbf{C_{21}{C_{11}}^{-1}C_{12}} - \mathbf{C_{22}}\right).$$

Combining these results, we can see that

$$-\mathbf{{M_{22}}^{-1}M_{21}} = \mathbf{C_{21}{C_{11}}^{-1}},$$

or $\mathbf{T_M} = \mathbf{T_C}$. If C is not invertible, then this is not necessarily true. In the present case, where the number of samples (cell lines) is less than the number of variables (proteins),

$$\mathbf{T_M} = -\left[\mathbf{{T_C}^{T}}\right]^{-1}.$$

The lasso regression formalism

The time translation (TT) matrix $\mathbf{T_L}$ is obtained by an iterative application of the lasso algorithm [100, 101]. For each row *k* of the data matrix, we calculate one row of the time translation matrix **T** and one component of the offset vector δ by numerically minimizing the quantity

$$\sum_{i=1}^{N}\left[X_{ki}(t_2) - \sum_{j=1}^{P} T_{kj}X_{ji}(t_1) - \delta_k\right]^2 + \lambda\sum_{j=1}^{P}\left|T_{kj}\right| \quad \text{Eqn. 7}$$

where the summation over *i* minimizes the mean squared error and the last summation penalizes non-zero entries in the time-translation matrix, dictated by the positive-valued constant λ. The software used here calculates a set of λ values at the start of the cross-validation given the inputs

[102]. By penalizing the L1 norm of $\mathbf{T}$, only a few non-zero coefficients are selected. The procedure is run independently for each row, and the coefficient arrays for every protein are concatenated to obtain $\mathbf{T_L} = \begin{bmatrix} [T_{11} & \cdots & T_{1P}] \\ \vdots & T_{ij} & \vdots \\ [T_{P1} & \ldots & T_{PP}] \end{bmatrix}$, where T_{ij} represents the regression coefficient obtained for the effect of protein j at time t_1 on the level of protein i at time t_2. Ten thousand instances of the time translation matrix $\mathbf{T_L}$, each starting with a different random seed, are generated using the *glmnet* package [103]. Edges that appear in at least half the matrices are considered essential and are used to select a representative matrix from the ensemble. The matrix that contains essential edges the most is selected as the representative the matrix (**Figure 5a**).

2.3.1.3 Computational validation of the lasso models

To validate the lasso approach and the interaction values it selects, I used randomly generated synthetic data. I found that the networks inferred from the RPPA data have fewer (typically about half as many) essential interactions than those generated from random synthetic data (**Figure 6**). It was further found that the distribution of interaction values inferred from synthetic data is normal (**Figure 7**), whereas those inferred from the experimental data contain a few very high values and a long tail. Moreover, it is shown that the number of edges inferred is robust against the number networks generated, and that the number of edges in the representative network converges to the number of accepted edges when 10000 networks are generated (**Figure 8**). The best candidate networks always contain high magnitude interactions both in Ins and IGF1 stimulation cases for full (**Figure 7**) and leave-one-out cross validation (LOOCV) (**Figure 9**) models. Finally, the TT matrices of lasso models for IGF1 and insulin conditions have multiple interaction elements in common (**Figure 10**).

2.3.1.4 Leave-one-out cross validation models

To test the performance of different model formalisms, I employed a leave-one-out cross validation (LOOCV) scheme (**Figure 5b**). In each LOOCV step, one cell line is excluded from the training set and the three TT models are constructed using the remaining cell lines only. Then, the SFM expression profile of the removed cell line is used to predict its profile at the next time point using the TT model learnt by using the remaining 20 cell line data matrices. As a result, twenty one LOOCV models are obtained for each time translation for each model. In the case of lasso models, hundred TT matrix instances are generated per condition, and the ensemble median is selected as the representative model. Comparison of LOOCV model performances of the three methods on the same unseen data enabled us to evaluate the robustness of models (see the chapter **Computational Modeling Results**).

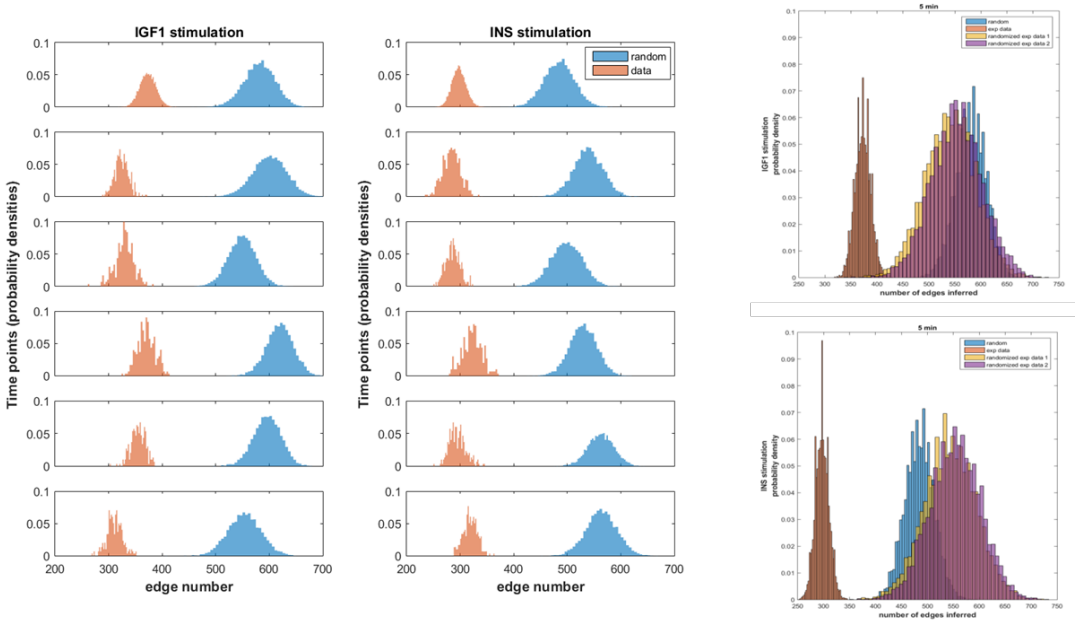

Figure 6. The lasso networks are different from random. (Left) The histograms for the number of interactions inferred by the lasso approach for the RPPA data and histograms for the corresponding random matrices are shown. 10000 networks were generated for each time point pair (one row for one pair). The histograms of random valued input matrices (blue histograms) are visually distinct from the ones for the real data (red histograms). The mean number of interactions (edges) for the random data is almost double the ones for real data, for every time point and

for both hormones. These results indicate that correlations present in the real data reduce the number of required non-zero regression coefficients. Each network in each case is generated starting from a different random seed number. From top row to bottom: for 5 min, 10 min, 30 min, 6 hrs, 24 hrs, and 48 hrs of stimulation. (Right) The lasso networks of the data are different from randomized data inferred networks. In addition to the randomly generated synthetic data (blue histograms), two sets of randomized real data matrices were used to construct 10000 networks (10 TL matrices for each of the 1000 different, shuffled real data matrices) to further show that the inferred interactions are not random. The real data matrices were either first randomized and processed (1st set, yellow histograms) or processed and then randomized (2nd set, purple histograms). These new sets of histograms were also distinct from the real data cases and overlapped highly with the synthetic random data histograms. These results indicate that correlations present in the real data reduce the number of required non-zero regression coefficients and data processing step itself was not the source of the correlations. Each network in each case is generated starting from a different random seed number. Results are shown for 5 min IGF1 (top) and Ins (bottom) stimulation cases.

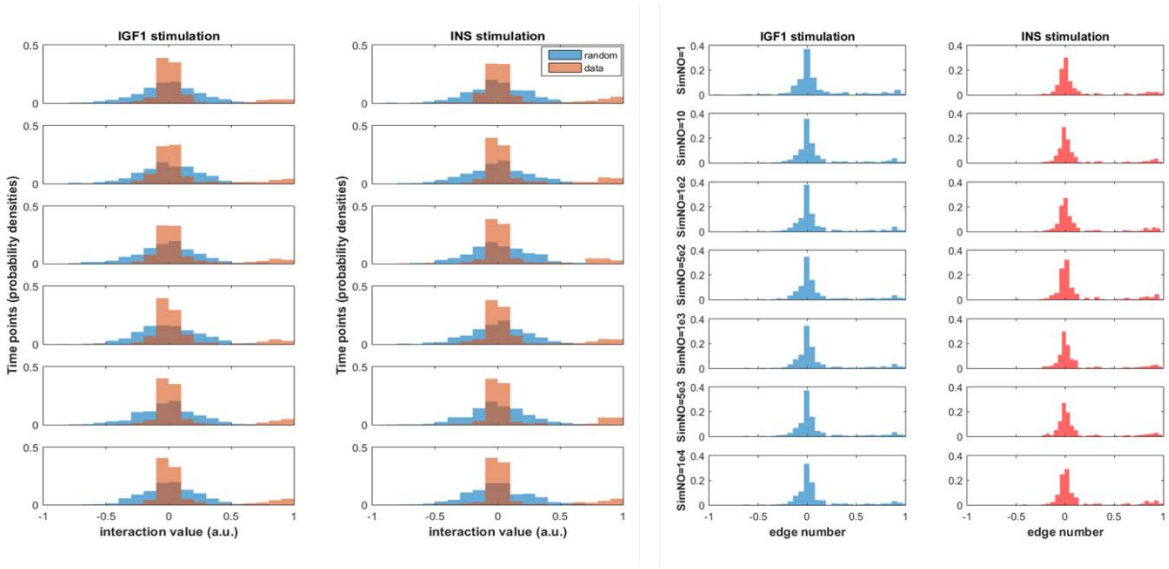

Figure 7. (Left) The lasso inferred magnitudes are normally distributed. The magnitude values of the interactions inferred by the best candidate networks for each time-point pair (from the networks shown in Figure 6 left panel). The candidates from the real data (red histograms) have values in the tails, significantly different from the random case (shown in blue histograms). (Right) The magnitudes of the interactions in the best candidate networks for different number of networks generated (from the ones in Figure 8). The histograms reveal that there are edges with significantly high magnitudes at the tails.

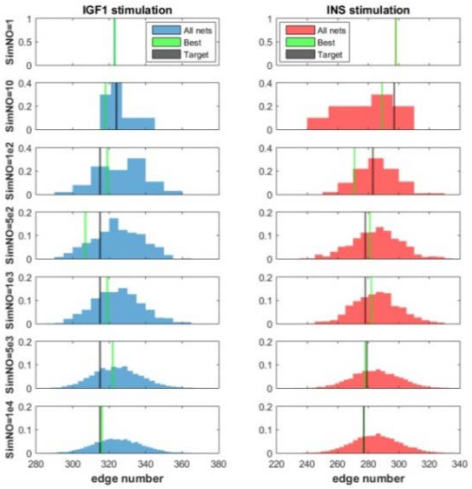

Figure 8. The histograms of number of edges generated by the lasso approach for different number of networks. From top row to bottom: 1, 10, 100, 500, 1000, 5000, and 10000 networks. Left column is for IGF1 stimulation, right column for Ins stimulation. Green lines indicate the number of edges present in the best candidate chosen for that set of networks. The black lines represent the number of edges of target networks. The numbers of present edges in best candidate and target networks converge to have very similar number. The histograms reveal that the number of edges generated is robust so that the mean values are very similar, and different for IGF1 and Ins cases.

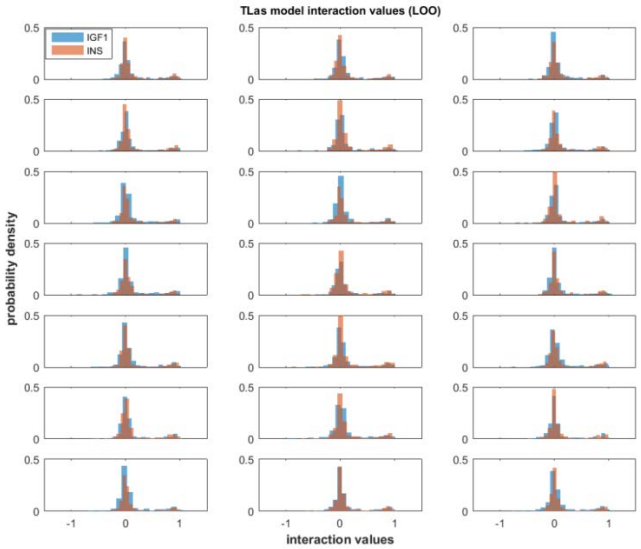

Figure 9. The magnitudes of the interactions in the best candidate networks for each LOOCV analyses. In each case, 100 networks were generated. Blue histograms are for IGF1 and red histograms for the Ins stimulation. The edge weights for the hormones have common as well as distinct valued edges.

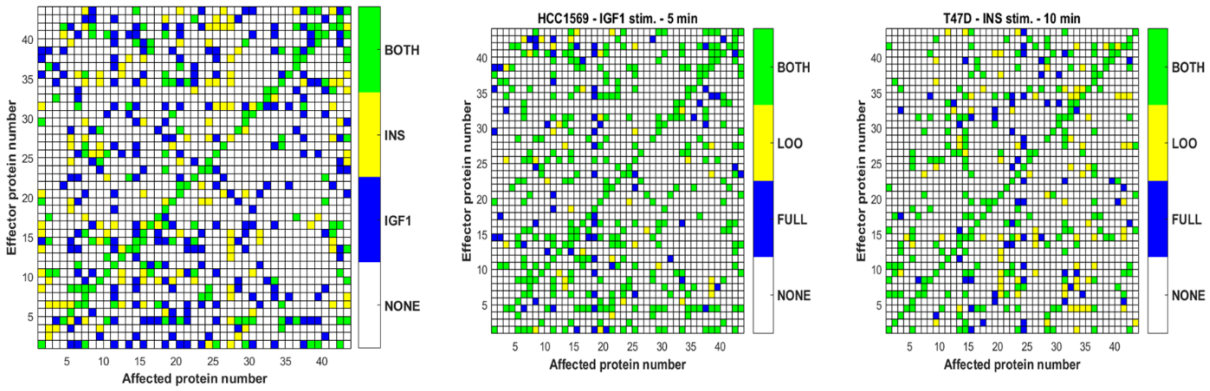

Figure 10. (Left) The sets of inferred interactions for IGF1 and insulin treatments. The coinciding and differential interactions inferred by the best candidates of lasso modeling for IGF1 and Ins stimulation conditions. Green squares indicate edges found for both systems, blue and yellow colors correspond to edges found only in IGF1 or only in Ins treatments, respectively. The empty cells are the possible interactions that are not inferred to be present in either system. The matrices in this plot represent TT models of SFM-5 min. The numbers shown for rows and columns correspond to the numbers shown for protein names in Table 2. The coinciding and differential interactions inferred by the best candidates of lasso modeling for full (green+blue) and LOOCV (green+yellow) modeling schemes. Results are shown for LOOCV models constructed without the information from HCC1569 (middle) or T47D (right) cell lines. There are common edges (Jaccard coefficients of 0.74 and 0.59 for the two respectively) between full model networks and LOOCV networks. The Jaccard coefficients for all cases span the range from 0.45 to 0.79, and above plots are two representatives. The matrices in this plot represent TT models of SFM-5 min for HCC1569 cells and SFM-10 min for T47D cells with IGF1 and insulin stimulation, respectively. The sparsity of the LOOCV networks are 78% and 81% for HCC1569 and T47D cell lines, respectively.

2.4 COMPUTATIONAL MODELING RESULTS

2.4.1 The lasso-based models are accurate and robust

To discover signaling differences from the data, I constructed linear models of time translation (TT) using three different methods as outlined in the previous section (**Construction of linear statistical models**). Each procedure resulted in a time translation matrix, **T**, of size 43×43. The element T_{ij} represents the effect that the level of protein j in unstimulated (serum-free medium, SFM) cells will have on the level of the protein i at a future time (**Figure 5** for details of the model construction).

One can obtain the predicted expression profiles for each time point by multiplying expression profiles from the SFM condition by the corresponding TT matrix. Then, the performance of the TT models are quantified by comparing the predicted expression profiles with the experimental data and through leave-one-out cross validation (LOOCV, **Figure 11**). The SVD model performs perfectly on its own training set, but clearly overfits during LOOCV (**Figure 11**). The lasso model performs relatively well on both the training set and LOOCV. On average (of all time points and conditions), the lasso models showed 0.95 correlation with the data in *full* modeling scheme (**Figure 11** and **Figure 12**). In LOOCV analyses, they also performed well, with Pearson correlation coefficient values of 0.902 ± 0.005 for IGF1 and 0.895 ± 0.006 for Ins stimulated cases (**Figure 12**). The MaxEnt results are not exact for either the training data or LOOCV and are worse than lasso predictions ($R \approx -0.6$). In fact, we notice that the MaxEnt predictions have a negative correlation with experimentally measured values. Based on these performance results, the sparsity of the lasso matrices (**Figure 10**), and the significance

of the inferred interactions, we selected the lasso model as the main model for experimental predictions. Steps 1-8-9-10 in **Figure 5a** schematically depict the lasso model construction.

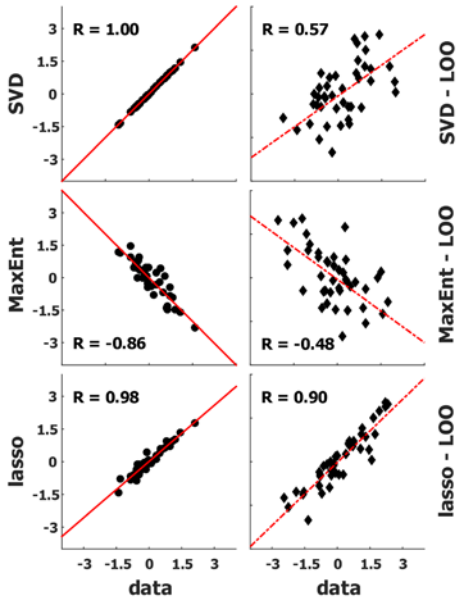

Figure 11. Performance analyses of the three models reveal that lasso models are accurate and robust. Three matrices were constructed for full (Figure 5a) and LOOCV (Figure 5b) models of each time translation. Correlation plots of data and prediction results from the three models for in-training (left) and leave-one-out cases (right) are shown. Black dots in the left panels represent data vs. full model predictions, and black diamonds on the right represent the data vs. LOOCV model predictions. The red lines are the least-square fits. The Pearson correlation coefficients are reported on each subpanel as R. Overall, the lasso models outperformed the other two methods. Modeling results of time translation for SFM to 10 min IGF1 stimulation in MCF7 cells shown as an example.

Differences between lasso-inferred interaction networks in IGF1 and insulin stimulated cells can be visually detected, as shown in **Figure 10** and **Figure 13**. Both networks are sparse, and they have very few interactions in common. To verify that the method is robust, I compared the networks obtained through LOOCV with those constructed from the full data set. In the LOOCV, a TT matrix was constructed for excluding each cell line once, giving a total 21 LOOCV model matrices for each time translation pair. The networks from these models overlap

with the full model networks. Although the number of coinciding edges was dependent on cell lines, these results provide additional support that the lasso model inference was robust. The overlap of inferred interactions from the full and LOO models are presented in **Figure 10**. Two representative LOOCV models constructed without the information for the corresponding cell lines are shown with comparison to full model interaction matrices. The mean Jaccard coefficients for comparison of LOOCV model networks with full model networks are 0.65 (±0.063) and 0.64 (±0.065) for IGF1 and insulin cases, respectively. These numbers mean that out of the interactions inferred for full and LOOCV models, on average 65% of them are found in both of the corresponding time point models and that the lasso models are robust in interaction inference. Additionally, the lasso model matrices are 83% (±2) sparse both in full and LOOCV models.

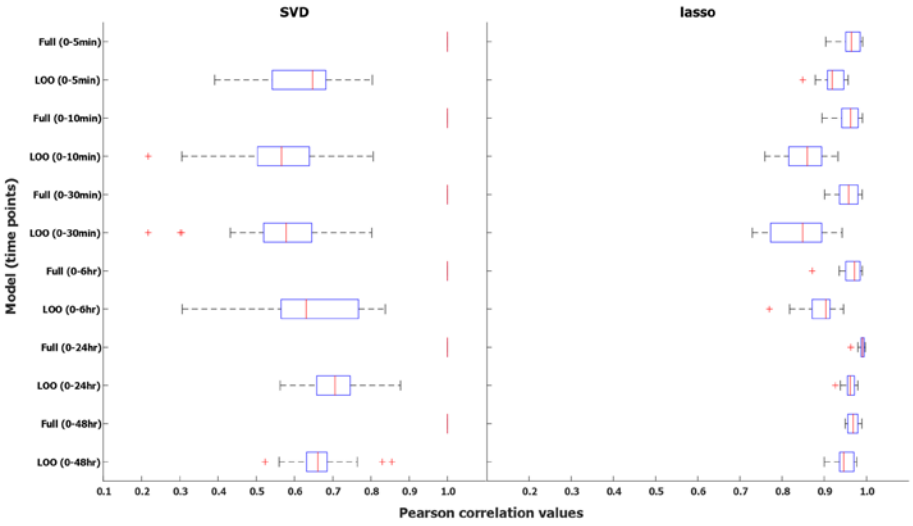

Figure 12. The comparison of correlation values of data vs. model predictions. Left panel is for SVD models, and the right panel for the lasso models. Each row depicts the time translation condition is corresponding plot is for. The boxplots are for the set results for prediction all cell lines separately (n=21). The SVD models are perfect (R=1) for in-training (full) models whereas their performance drops to 0.5 in LOOCV models. The lasso models (right panel) however do perform equally well in test cases (LOO results) as in the in-training cases (full).

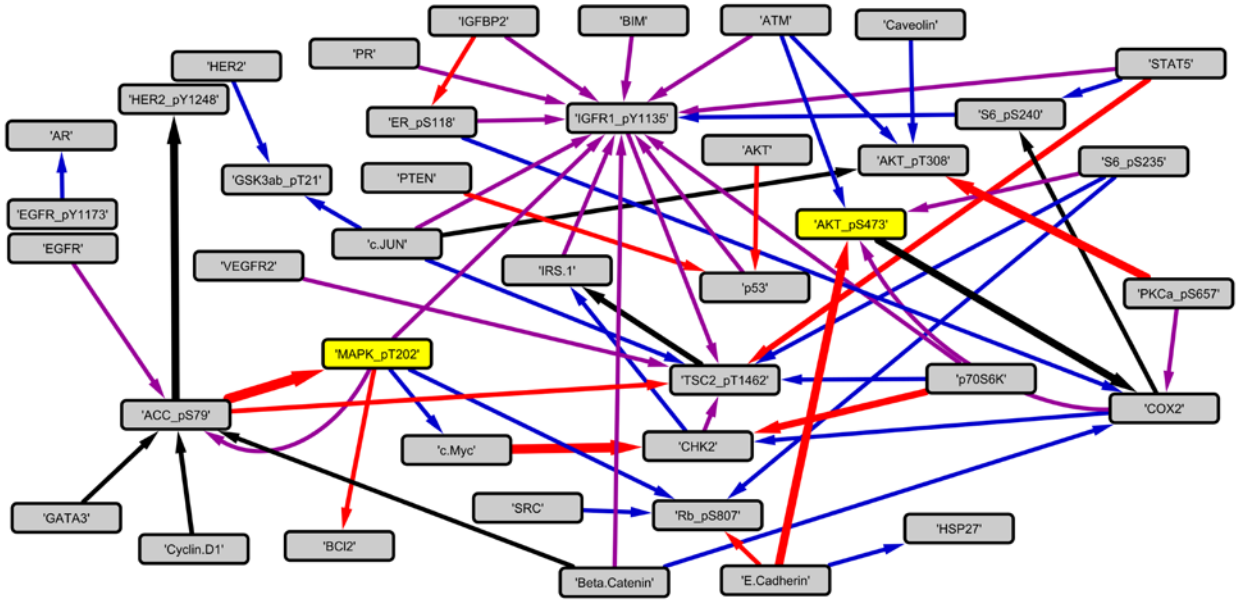

Figure 13. The lasso modeling inferred network of interactions in IGF1/Ins signaling pathways. The interactions inferred by the lasso model for SFM to 10 min stimulation are depicted. The purple edges are the top 20 exclusive interactions present in IGF1 model only. The blue edges are the interactions inferred only in the insulin model. The red edges represent the interactions present in both models while the magnitudes are stronger in the IGF1 model. The black edges are the ones with higher magnitudes in the Ins model. The yellow nodes are the two phospho-proteins taken as the outputs for the experimental validations.

2.4.2 Lasso-inferred and novel interaction predictions

I constructed six TT matrices for IGF1 and six matrices for Ins using the mean SFM condition data as $\mathbf{X}(t_1)$ and data from each stimulation time point as $\mathbf{X}(t_2)$. In this setting, the TT matrices reflect the interactions from time zero to the end of hormone stimulation. I used SFM as the starting point for all TT calculations because SFM protein levels can be experimentally controlled, enabling us to validate model predictions.

The interaction values of the TT matrices reflect the effect of synthetically knocking down a protein. If one perturbs the interaction values of a protein (a column in the TT matrix)

and predicts how the responses deviate from the data, the results are correlated with the magnitudes of the inferred interaction constant. These observations led to the study of the individual interaction values only. First, the set of non-self-interactions present in TT matrices of *both* IGF1 and Ins stimulation cases were determined (96, 87, 98, 78, 62, and 85 interactions found for time points 1-6, respectively). The list was filtered to retain only pairs of interactions having either phospho-Akt (pAkt) or phospho-MAPK (pMAPK) as outputs because these are two of the major downstream mediators of IGF1 and Ins signaling. Finally, for each time point the interactions were ranked by differential response to insulin and IGF1 (i.e., for each interaction in each time point, the quantity $\left|log\frac{TT_{IGF1}}{TT_{Ins}}\right|$ was calculated, where TT_x is the response in condition *x*). The depiction of the top exclusive interactions for IGF1 and insulin cases as well as the top differential interactions revealed how complex our study system is (**Figure 13**).

The lasso models identified a number of previously known interactions. For instance, the magnitudes of the inferred interactions between PKCα – pPKCα and between Rb1 – pRb1 are among the highest strength interactions. These indicate that the level of non-phosphorylated proteins in unstimulated cells influences the level of phosphorylated proteins of the same protein at later times (see **Table 3** for more such interactions). It is also important to note that the existence of these two interactions reflects on the importance of the roles of PKCα and Rb1 in cancer progression and transcriptional regulation, modulated downstream of IGF1 and Ins signaling [104-109]. Additionally, the inferred interaction of Rb1 level affecting pRb1 level is the only non-self interaction present in all eight (four IGF1, four Ins) lasso models up to 6 hrs, pointing out the importance and critical role of Rb1 regulation. Other such interactions are also uncovered by the model, including the levels of phosphorylated Akt [110], EGFR [111, 112], and HER2 [113, 114] depending on the total protein level in SFM condition.

Table 3. The list of intuitive experimental candidates from all three models.

Perturb	Measure	Lasso prediction	SVD prediction	MaxEnt prediction
PKCa	PKCa_pS657	Ins (5, 30 min) IGF1 (10 min)	Ins (5, 10 min) IGF1 (30 min)	Ins (5, 10, 30 min)
Rb1	Rb1_pS807	IGF1 (5 min) Ins (10, 30 min)	Ins (5, 30 min) IGF1 (10 min)	IGF1 (5, 10, 30 min)
EGFR	EGFR_pY1173	Ins (5, 10, 30 min)	Ins (5, 10, 30 min)	Ins (5, 10, 30 min)
HER2	HER2_pY1248	Ins (30 min)	Ins (30 min)	Ins (30 min)
AKT	AKT_pT308	IGF1 (5 min)	IGF1 (5 min)	IGF1 (5 min)
AKT	AKT_pS473	IGF1 (10 min)	IGF1 (10 min)	--
p70S6K	S6_pS235	Ins (30 min)	Ins (30 min)	IGF1 (30 min)
AKT	TSC2_pT1462	Ins (5, 30 min)	Ins (5, 30 min)	Ins (5, 30 min)
AKT	S6_pS240	IGF1 (5 min)	IGF1 (5 min)	IGF1 (5 min)
AKT	S6_pS235	IGF1 (5 min)	IGF1 (5 min)	IGF1 (5 min)
IGFR1	EGFR_pY1173	IGF1 (5 min)	IGF1 (5 min)	IGF1 (5 min)
IGFR1	ER_pS118	IGF1 (10 min)	IGF1 (10 min)	IGF1 (10 min)
S6	S6_pS235	IGF1 (5 min) Ins (10, 30 min)	IGF1 (5 min) Ins (10, 30 min)	IGF1 (5, 10 min) Ins (30 min)

The lasso inference method also predicted several interactions (**Table 4**) that are neither reported in the literature nor in databases like STRING [115]. For example, a STRING search using the 42 gene names corresponding to the 43 proteins used in this study produces a network

Table 4. The ranked list of experimental candidates from all three models. Each perturbation is predicted to have a differential effect in the reported time point. These interaction pairs are either novel or highest ranking predictions with limited literature background. The experimentally validated interactions are highlighted.

Perturb	Measure	Lasso prediction (time point)	SVD prediction (time point)	MaxEnt prediction (time point)
ACC	MAPK_pT202	IGF1 (5, 10 min)	IGF1(5, 10, 30 min)	Ins (10, 30 min)
Cyclin B1	MAPK_pT202	IGF1 (30 min)	--	Ins (30 min)
ATM	AKT_pS473	Ins (5, 10, 30 min)	IGF1 (5, 10, 30 min)	IGF1 (10, 30 min)
E-Cadherin	AKT_pS473	IGF1 (10, 30 min)	IGF1 (10, 30 min)	--
COX2	AKT_pS473	IGF1 (10, 30 min)	IGF1 (5, 10, 30 min)	IGF1 (30 min)
ATM	AKT_pT308	Ins (5, 10 min)	Ins (10 min)	--
COX2	AKT_pT308	--	IGF1 (5, 10, 30 min)	Ins (5, 10 min)

of 49 edges based on a confidence cutoff of 0.9 and evidence only from experiments. None of the interactions reported in **Table 4** are in that list. When the STRING-based network is compared to our lasso models, there are 12 and 10 edges in common for IGF1 and Ins networks, respectively. The probabilities of having at least these many edges correctly assigned in a random network (calculated via the hypergeometric distribution) are $P_{IGF} = 0.021$ and $P_{Ins} = 0.063$, suggesting that the lasso models capture more known interactions than would be expected at random. The list presented in **Table 4** summarizes results from the lasso models. The first line indicates that phospho-ACC (pACC) levels are predicted to have a greater effect on the IGF1-induced phosphorylation of MAPK than on insulin-induced MAPK phosphorylation after 5 and 10 minutes of stimulation. The SVD model makes the same prediction, but extends it to 30 minutes after stimulation. The MaxEnt model also predicts the effect for 5 min of stimulation (**Table 4**), but interestingly predicts that the effect is stronger due to insulin stimulation for 10

and 30 min stimulation. Importantly, all three models predict that altering levels of pACC in cells in SFM will impact levels of pMAPK upon IGF1 and Ins stimulation, and that the extent of the effect will depend on which growth factor is introduced. The interaction was also shown to be robust, appearing in 99.96% of networks (out of 40,000) generated for the first two time points for both insulin and IGF1 stimulation in LOOCV of the HCC1569 cell line. It is also important to note that the models employed in this work do not discriminate between direct and indirect associations, so that these interactions may be mediated by other cell factors that are not measured in the current dataset.

Another interesting prediction of our model is the differential effect of E-Cadherin expression levels on pAkt level induced by IGF and Ins. This is the top differential effect candidate for the 30 min time point, and was shown to be present in 97.35% of LOO networks constructed for 10 and 30 min stimulation of both growth factors.

2.5 EXPERIMENTAL VALIDATION OF THE PREDICTIONS

2.5.1 Acetyl-co-A carboxylase and MAPK

The differential influence of ACC on MAPK phosphorylation is validated using knock-down experiments. The experiments here are carried out by Alison Nagle, a graduate student of Dr. Adrian Lee. The two cell lines for the knock-down, MCF7 and HCC1569, were chosen because they have high pACC protein levels in the data, and because both exhibit an expected MAPK response upon hormone stimulation. The immunoblot shown in **Figure 14a** and the bar graph in **Figure 14c** demonstrate that upon ACC1/2 siRNA knock-down in MCF7 cells, phospho-MAPK

levels increased, and that the extent of the increase was different in IGF1- and Ins-stimulated cells. Compared to Scr cells, IGF1 stimulated cells with no active ACC1/2 showed higher MAPK phosphorylation, amounting to a 40% increase ($P<0.05$) in fold-change level at 5 min. However, a twenty percent decrease in fold-change ($P>0.05$) was observed in insulin stimulated cells, suggesting a higher proliferative IGF1 response compared to insulin induction. At 10 min, there was an insignificant down-regulation of the pMAPK response in ACC knock-down cells in both cases. Additionally, we showed that the ACC knock-down in HCC1569 cell line causes a significant change (70% increase in fold-change, $P<0.05$) in IGF1 stimulated cells compared to an insignificant change (5% increase, $P>0.05$) seen in Ins stimulated cells at 10 min (**Figure 15**). Indeed, the differences in fold-change of relative pMAPK levels were larger in IGF1 stimulated cells for all time points in HCC1569 cells.

2.5.1.1 siRNA knock-down and immunoblotting

The HCC1569 and MCF7 cell lines were acquired from ATCC and grown in RPMI+10% FBS. ACACA and ACACB (gene names for acetyl-coenzyme-A carboxylase isoforms) pooled siRNAs (Dharmacon) were reverse transfected with Lipofectamine RNAi MAX, following the manufacturer's protocol (Life Technologies), with a final concentration of 25nM. Cells were serum starved for 24 hours then stimulated with 10nM IGF1 or insulin for 5 min, 10 min, 30 min, 6 hour, and 24 hour time points. 10nM HCl was used as a vehicle control. Protein lysates were collected with RIPA buffer containing protease and phosphatase inhibitors, and standard immunoblot techniques were used to evaluate protein expression. The pMAPK (Thr202/Tyr204; Cell signaling, 1:1000), MAPK (Cell signaling, 1:1000), ACC (Cell Signaling; 1:500) and β-actin (Sigma; 1:5000) antibodies were used. RNA was isolated using the Illustra RNAspin Mini Kit (GE Healthcare) and reverse transcription was done following the iScript cDNA synthesis kit

protocol (BioRad). The SYBR green supermix protocol (BioRad) was used for qRT-PCR. Primers for ACACA (5'AAGAATCGGACTGGCAGAAG (fwd) and 5' AATGGACAGAGTTGAGAGCAC (rev)), ACACAB (5' ACCACATCTTCCTCAACTTCG (fwd) and 5'TGTTGATCTTGACCTCAGCC (rev)) were used to evaluate siRNA transfection efficiency.

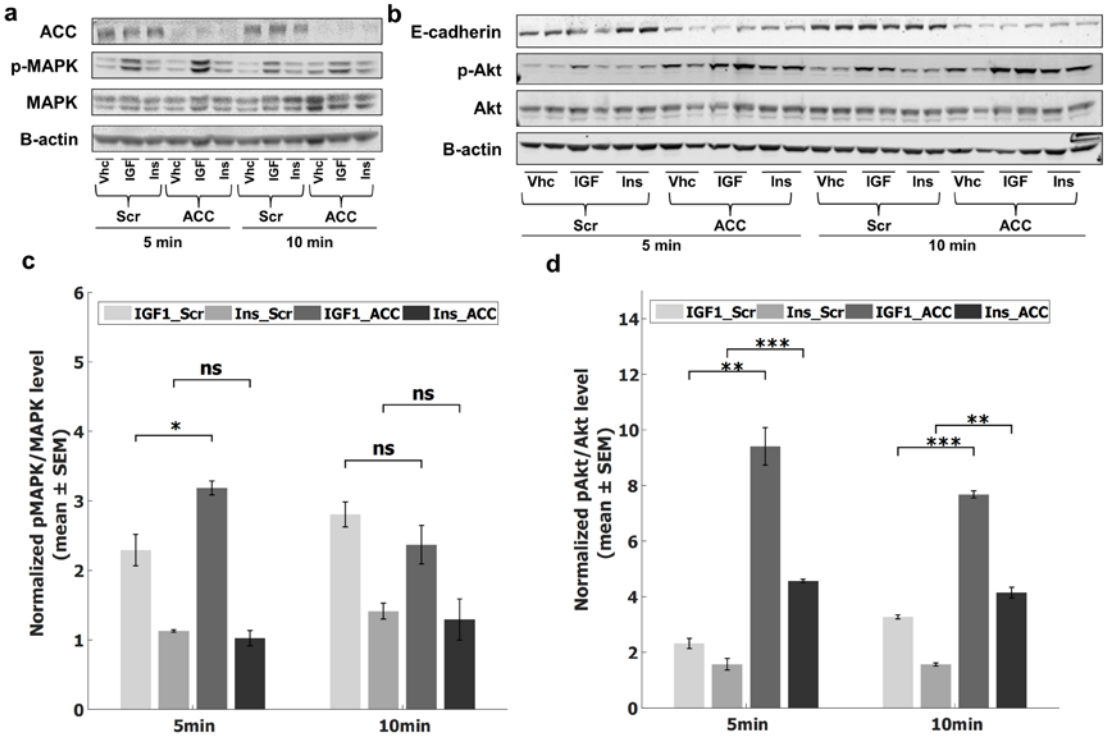

Figure 14. The lasso model predictions are experimentally validated to show a differential effect of ACC knock-down on MAPK phosphorylation, and E-Cadherin knock-down on Akt phosphorylation levels. (a) The representative western blot for ACC knock-down in MCF7 cells at 5 and 10 minutes of stimulation, Scr = control cells, ACC = ACC-knock-down cells. The cells were stimulated with Vhc, IGF1, or Ins. The other experimental replicates are given in Figure 16. (b) The western blot showing the change induced by stable E-Cadherin knock-down in T47D cells in response to 5 and 10 min IGF1 and Ins stimulation. Scr = control cells, CDH1 = E-Cadherin-knock-down cells. The cells were stimulated with Vhc, IGF1, or Ins. The efficiency of E-Cadherin knock-down is reported in Figure 17c. (c) The quantification of western blots given in (a), shown as relative values of

pMAPK/MAPK normalized to Vhc treated Scr condition. (d) The quantification of the western blot in (b). Relative values of pAkt to total Akt, normalized by Vhc treated wild-type cells. The bars represent fold-changes from basal condition and results are shown as mean ± SEM of two or three experimental replicates. The results are compared using unpaired, one-tailed two-sample t-test, and P<0.05 (*), P<0.01 (**), P<0.005 (***), nonsignificant (ns). The bars for each time point, from left-to-right: IGF1 stimulated WT cells, Ins stimulated WT cells, IGF1 stimulated ACC knock-down cells, and Ins stimulated ACC knock-down cells.

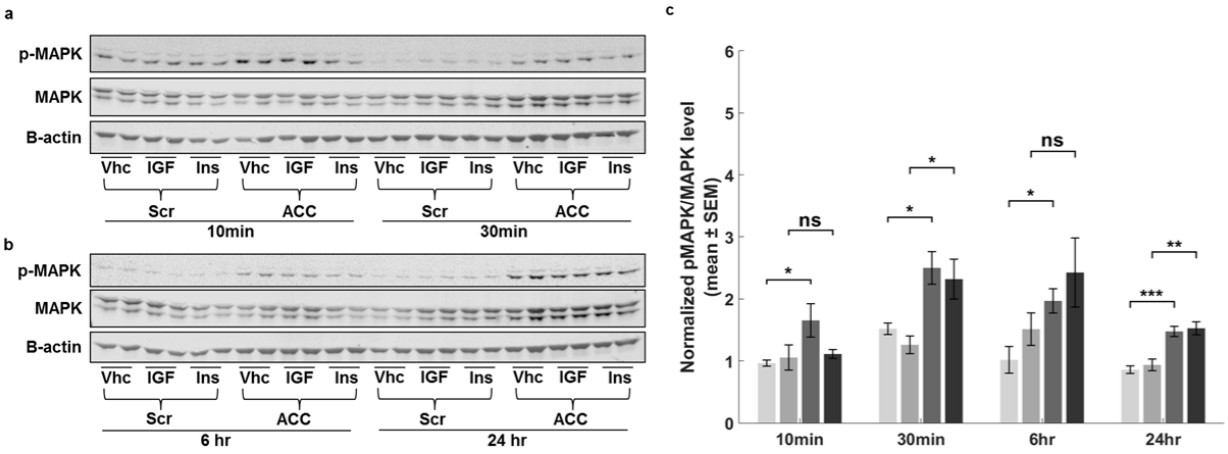

Figure 15. The western blot for ACC knock-down in HCC1569 cells at 5 and 10 min (a), and 6 and 24 hours (b) of stimulation, Scr = control cells, ACC = ACC-knock-down cells. The cells were stimulated with Vhc, IGF1, or Ins. (c) The quantifications of the western blots in (a, b). The results are compared using unpaired, one-tailed two-sample t-test, and P<0.05 (*), P<0.01 (**), P<0.005 (***), nonsignificant (ns). Bars represent from left-to-right (light-to-dark): IGF1 stimulated WT cells, Ins stimulated WT cells, IGF1 stimulated ACC knock-down cells, Ins stimulated ACC knock-down cells. Bars for each time point from left-to-right: IGF1 stimulated WT cells, Ins stimulated WT cells, IGF1 stimulated ACC knock-down cells, Ins stimulated ACC knock-down cells.

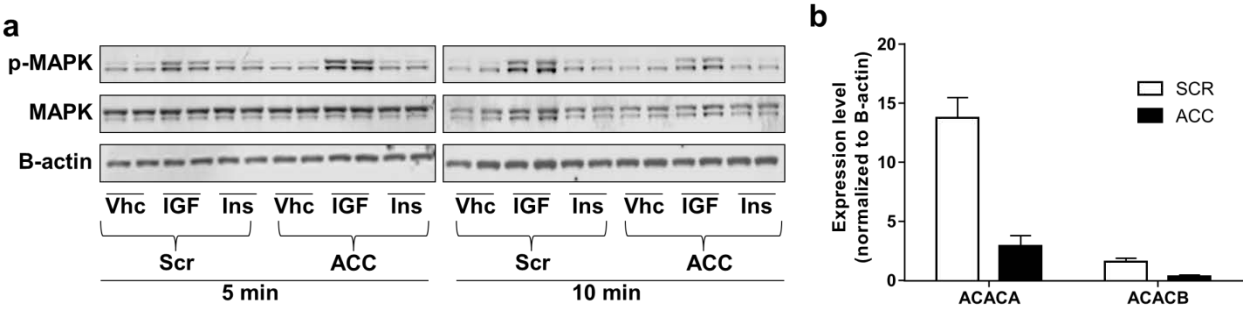

Figure 16. The ACC1/2 knock-down in MCF7 cells showed higher influence in IGF1 stimulated compared to Ins stimulated cells. (a) The western blots for the ACC knock-down in MCF7 cells at 5 and 10 minutes of stimulation with vehicle, IGF1, or Ins. (b) ACC siRNA transfection efficiency quantification by qPCR. The quantification of the levels of ACC1/2 (ACACA/ACACB) in ACC knock-down MCF7 cells compared to Scr transfected cells.

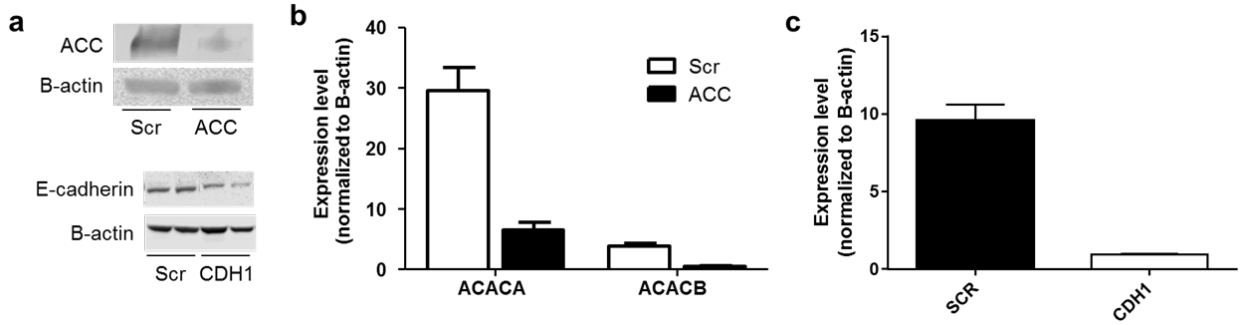

Figure 17. Knock-down efficiencies. (a) The western blot for ACC knock-down efficiency in HCC1569 cells and the blot for E-Cadherin knock-down efficiency in T47D cells. Scr = control cells, ACC = ACC-knock-down cells, CDH1 = E-Cadherin-knock-down cells. Both of the proteins have lesser amounts in knock-down cells. (b) ACC siRNA transfection efficiency quantification by qPCR. The quantification of the levels of ACC1/2 (ACACA/ACACB) in ACC knock-down HCC1569 cells compared to Scr transfected cells. (c) E-Cadherin shRNA transfection efficiency quantification by qPCR. The quantification of the levels of E-Cadherin (CDH1) in knock-down T47D cells compared to Scr transfected cells.

2.5.2 E-Cadherin and Akt

The results of the experiments with E-Cadherin knock-down, **Figure 14b** and **Figure 14d** verify increased Akt activation. The effect was more evident in response to IGF1 than to insulin. Leading to higher growth signaling in stimulated cells, the loss of E-Cadherin induced an almost 2.5-fold increase in relative pAkt/Akt level in response to both IGF1 ($P<0.005$) and Ins ($P<0.01$) at 10 min. At 5 min, the change was significant again in both cases ($P<0.01$ and $P<0.005$, respectively). The magnitude of the increase in fold-change was twice in IGF1 stimulated cells compared to insulin induced cells both at 5 and 10 min.

2.5.2.1 siRNA knock-down and immunoblotting

T47D cells acquired from ATCC were stably transfected with shRenilla and shCDH1 (CDH1: E-cadherin gene name) constructs using a retroviral infection and were grown in RPMI+10% FBS with 1ug/ml puromycin. Cells were serum starved for 24 hours then stimulated with 10nM IGF1 or insulin for 5 and 10 min. 10nM HCl was used as a vehicle control. Protein lysates were collected with RIPA buffer containing protease and phosphatase inhibitors, and standard immunoblot techniques were used to evaluate protein expression. The pAkt (Ser473; Cell signaling; 1:1000), Akt (Cell signaling; 1:1000), E-cadherin (BD Biosciences; 1:1000), and β-actin (Sigma; 1:5000) antibodies were used. RNA was isolated using the Illustra RNAspin Mini Kit (GE Healthcare) and reverse transcription was done following the iScript cDNA synthesis kit protocol (BioRad). The SYBR green supermix protocol (BioRad) was used for qRT-PCR. Primers for CDH1 (5’ GAACAGCACGTACACAGCCCT (fwd) and 5’ GCAGAAGTGTCCCTGTTCCAG (rev)) were used to evaluate siRNA transfection efficiency.

2.5.3 Statistical Analysis

Statistical analysis on the validation experimental results was performed using two-sample t-test with significance level of $P<0.05$. Unless otherwise indicated, the values are presented as means ± SEM (standard error of the mean). All of the simulations and analyses were run in Matlab, versions 2014a and 2015a (The MathWorks, Inc.).

2.6 DISCUSSION

In the present work, I analyzed temporal proteomic data to identify differences in the insulin and IGF1 signaling pathways in a large panel of human breast cancer cell lines. The initial data consisted of protein expression levels generated by RPPA protocol, collected at different time points with IGF1 and Ins stimulation conditions. The transformation matrices obtained by the lasso approach differ significantly from matrices generated using randomized data in the number of interactions (**Figure 6**), and in interaction magnitudes (**Figure 7**). The models predicted that knocking down ACC and E-cadherin would produce differential effects in Ins- and IGF1-stimulated cells, and these predictions are experimentally validated.

RPPA acquisition and analysis is still a growing area of research. Studies utilizing RPPA data have employed a diverse range of data (pre)processing and benchmarking methods, but no single protocol for processing RPPA data has yet been universally accepted [41-43, 116]. Although inclusion of RPPA is still lacking, there are ongoing efforts to standardize and coordinate dissemination of different data types [117]. Here, I mean-centered and standardized the protein arrays before model construction. This pre-processing enabled me to study the

relative shapes of the time-course expression curves rather than of the amplitudes. Each sub-matrix of 43 proteins in 21 cell lines was processed separately, and three different models of time-translation were constructed for each pair (**Figure 5**).

The lasso model TT matrices constructed for each time-point pair were sparse (**Figure 10**). They performed overall best among the three modeling schemes explored (SVD, MaxEnt, and lasso). The lasso algorithm provides a means of selecting a few good descriptor weights given input data, to recapture the output with minimal deviation [100, 102, 118-120]. Although lasso could under-predict interactions when there is redundancy in the network, it was shown in **Figure 11** that the lasso models perform equally well in leave-one-out tests as in training-set tests. When compared to the other two models, the lasso model selected enough good descriptors during model construction to enable better predictions from unseen input data. The lasso models also contain fewer non-zero interactions than the other models, easing the burden of identifying biologically relevant interactions. This lasso procedure follows a gradient descent method that depends on a random seed. To minimize the influence of random effects on the results, a large number of matrices were generated and one representative canonical lasso matrix was selected based on the number of times that each of its elements is non-zero. Selection of this ensemble median network as the model representative is also useful in other applications, such as reconstruction of gene regulatory networks from microarray data [121, 122].

The covariance-matrix derived models (SVD) are analytical solutions of the training problem and suffer from overfitting. These models can recapture the response exactly for the training data, whereas outputs for non-training-set inputs deviate from the real response considerably. The overfitting can be relaxed by invoking maximum entropy (MaxEnt models), but this gives rise to another problem: because the number of proteins is larger than the number

of cell lines the MaxEnt method does not exactly reproduce the training set data. Instead of predicting the training data exactly and the test data poorly, it generates mildly erroneous predictions for all data. Interestingly, we find that the data agree more closely with the inverse of the MaxEnt predictions than with the predictions themselves. This discrepancy between measured and modeled values occurs only when the number of proteins is greater than the number of conditions. If we use 20 proteins instead of 43, the data agree better with the MaxEnt solution than with its inverse. The MaxEnt matrix has been shown in earlier works to be more stable for systems for inputs with random noise [57]. The lasso models outperformed maximum entropy models here, suggesting that the data are not drawn from a maximum entropy distribution.

My computational model predictions were validated in multiple ways. First, existence of post-translational modifications in the translation matrices, i.e. effects of non-phosphorylated forms of PKCα and Rb1 on the levels of respective phosphorylated states, are important measures of validity of the models. Existence of such interactions over the time-course persistently shows that the lasso algorithm selects meaningful, non-random descriptors. Further validation comes from the presence of known interactions between kinase and substrate. Examples from the lasso model include the level of p70S6K affecting the level of phospho-S6 [123] and the predicted dependence of TSC2 phosphorylation level on the level of Akt present. TSC2 is phosphorylated in response to activation of Akt, leading to activation of mTORC1 complex [124]. A third tier of validation is the set of results obtained from the *in vitro* knock-down experiments. The effects of ACC knock-down on pMAPK level and of E-Cadherin knock-down on pAkt level in response to IGF1 and Ins show that the model predictions are accurate. The fact that the aforementioned experimental candidates show differential results in IGF1 and

Ins stimulation conditions validates the overall approach for elucidating novel interactions downstream of the two hormones.

ACC is involved in lipid metabolism, specifically in conversion of acetyl-CoA to malonyl-CoA. The latter is the building block of fatty acids and there is evidence tying ACC activity to the energetic state of the tumor microenvironment and cell survival [125, 126]. Low energy states (high AMP/ATP ratio conditions) promote higher AMPK activity, inhibiting ACC1/2 by phosphorylation. The inhibition of ACC in turn leads to increased fatty acid oxidation and decreased lipid biosynthesis [126]. It has been suggested that high levels of inactive pACC sustain NADPH (nicotinamide adenine dinucleotide phosphate), a reducing agent for biosynthetic reactions, at a high level, providing a better environment for tumor cell growth under stress. It has also been shown that the diabetic drug metformin activates AMPK, leading to an increase in pACC levels and consequently to higher insulin signaling and reduced insulin resistance [125]. Although there is evidence linking pACC to cell growth and insulin signaling, I found no previously reported link between pACC and pMAPK. Our results here are consistent with the previous results such that a decrease in ACC1/2 level promotes cell survival, although through an additional mechanism of action. This differential effect in response to IGF1 or Ins requires further study.

There are controversial results reported so far, some stating that the presence of E-Cadherin junctions between cells recruits PI3K to the cell membrane, resulting in a higher activation of Akt in normal kidney and ovarian cancer cells [127, 128]. Lau and colleagues, on the other hand, stated that endogenous E-Cadherin signaling negatively regulates Akt activation by sequestering beta-catenin leading to increased PTEN transcription and decreased Akt phosphorylation [129]. Loss of E-Cadherin then releases beta-catenin for translocation into the

nucleus leading to inhibition of PTEN transcription and thus upregulation of Akt signaling. Complementarily, previous investigations showed that upon IGF1 treatment, increased pIGF1R and pAkt levels reduced E-Cadherin concentration and functioning [130-133]. It was also shown that high cell-cell contact formation by E-Cadherin disrupted high affinity IGF1 binding to its receptor, resulting in lesser downstream activation [134, 135]. Collectively, our lasso model prediction clearly has foundations in the literature but the differential effect in response to IGF1 and insulin has not been studied. Our results support the notion that loss of E-Cadherin induces pAkt and also coincide with the earlier works showing a relationship between IGF1 signaling and EMT (Epithelial-to-Mesenchymal Transition) through regulation of E-cadherin. Following up on our results presented here, our colleagues recently showed that loss of E-Cadherin sensitizes certain cell lines to IGF1R inhibitors [136]. **Figure 18** summarizes our hypothesized mechanisms of actions for ACC and E-Cadherin within the canonical cascades of IGF and Ins signaling.

Another interesting prediction of the lasso models was the effect of ATM on both phospho-forms of Akt (**Table 4**). Earlier studies showed that ATM mediates full activation and downstream signaling of Akt under insulin and IGF1 stimulation in other tissue types [137-139]. The models are of note to infer edges from ATM to both pAkt proteins, where full activation of Akt is accomplished by dual phosphorylation of these sites. Further study of that interaction might spearhead new combination therapies with IGF1R inhibitors in breast cancer tumors. Indeed, very recently, researchers have shown that presence of kinase-domain mutations in ATM is a biomarker for Topo-isomerase I inhibitor therapy [140].

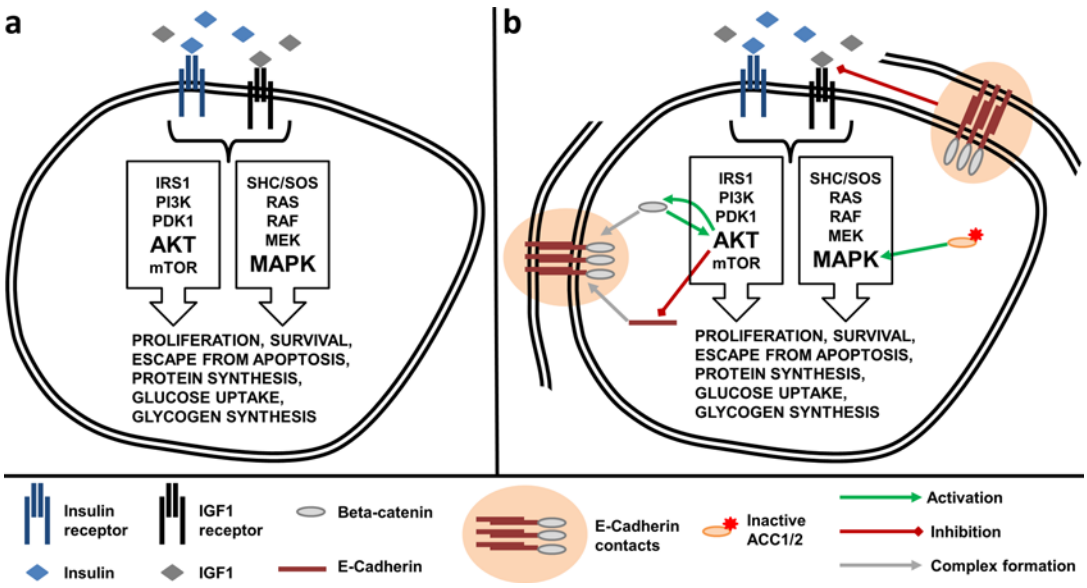

Figure 18. Links between IGF1/Ins signaling, ACC and E-Cadherin. (a) The canonical IGF1R/InsR signaling pathways comprised of MAPK and Akt cascades. (b) The canonical signaling including newly hypothesized interactions based on validated predictions of TT modeling approach. The adherens junction formation sequesters beta-catenin, leading to a repressive effect on Akt activation. When free in cytosol and translocated to nucleus, beta-catenin inhibits PTEN translation, leading to higher activation of Akt. Activated Akt in return phosphorylates GSK3b, and cause a decrease in beta-catenin degradation. Phosphorylated Akt also turns on NF-kB signaling and inhibits E-Cadherin translation. E-Cadherin driven cell-cell contacts also were shown to repress high affinity ligand binding to the receptors, leading to diminished downstream signaling. On the other hand, ACC1 and ACC2 control fatty acid synthesis cascade, but here shown to affect MAPK phosphorylation in response to IGF1 and insulin stimulation.

The main aim of the statistical modeling scheme introduced here was to extract information on important and differential interactions from available experimental data. Any dependence inferred reflects trends in the data enabling us to extrapolate the underlying correlations to new predictions. Such novel findings will potentially lead to new insights and understanding of the underlying mechanisms of IGF and insulin signaling [52]. The experimentally validated interactions can further be explored to find better targets for anti-tumor

response. One option is to incorporate these novel interactions into mechanistic models downstream of IGF1R or InsR signaling to study the system dynamics and pathway dysregulation mechanisms, which will be covered in the next **Chapter**. A similar approach was studied by Iadevaia et al. using RPPA data of IGF1 stimulation in MDA-MB-231 cells [43].

The study of breast cancer prevention, diagnosis, and therapy is imperative. However, compensation and activating feedback mechanisms collectively depress the efficacy of anti-IGF1R/InsR therapies [22]. Only after discerning the dual IGF1R/InsR system dynamics and novel components, will we be able to start to direct patients with suitable tumor subclasses toward more efficient personalized clinical interventions. The computationally scalable reverse-engineered models of cellular networks I introduced here provided us with a framework to elucidate experimental targets of pharmacological importance in a cost effective way.

3.0 MECHANISTIC MODELING

3.1 INTRODUCTION

The dynamic responses of time series data are exploited by use of computational modeling approaches. The representation and then simulation of biological systems is important to test new hypotheses and learn more of the hidden biology within these systems. Ranging from statistical modeling (see previous **Chapter 2**) to mechanistic and to more complicated whole cell models [141, 142]; the method differs with the question at hand. Toward computationally identifying differences between insulin and IGF1 signaling pathways in breast cancer, I developed and refined methods for combining proteomics data to reveal details of aberrant signaling in cancer.

Prior to this work there were models with only IGF1R [43, 143-145] or with only InsR [146, 147]. These previous models assumed IGF1R and InsR as indistinguishable and omitted the opportunity to discover differences for targeted therapies. The main aim of the modeling scheme here is to extract information on important and differential interactions from available experimental data. As outlined previously, the ligand binding to insulin-like growth factor receptor type I and insulin receptor promotes cascades of phosphorylation reactions [9, 20]. To recapitulate the downstream dynamics of these two receptor-tyrosine-kinase systems, the topology of the mechanistic pathway model using knowledge from already existing models is designed and modeled using a rule-based approach [83, 148, 149]. Here, I wanted to represent

the IGF1R and InsR signaling system together in a simpler computational model where I would be able to test mechanistic details in downstream signalin. Thus, I built ODE models, primarily for the MCF7 cell line, and then parametrized the model for six additional cell lines.

The initial computational mechanistic model constructed in this chapter is comprised of the two receptors, IGF1R and InsR, where no hybrid receptors are allowed. Having no hybrid receptors, I ought to seek what is different between only IGF1 and only insulin activated cascades. Moreover, hybrid receptor formation and function are still under studied, with a very recent publication from our collaborator [150]. The model includes cascades of both mitogen-activated protein kinase (MAPK) and Akt signaling, as well as the crosstalk and feedback loops in between. The constructed model has 14 proteins and two ligands, and 66 parameters, of which 16 are the total protein counts. The protein counts, and 34 of the rate parameters, are common between IGF1 and insulin models, where each model has eight specific parameters. The parameter estimation is performed using Markov Chain Monte Carlo (MCMC) sampling, and the model performance is checked using test data of dose responses of IGF1 and insulin, and PI3K inhibition data with IGF1 stimulation. All of these fits are within acceptable range and convey qualitatively well agreement.

In the analysis of the final computational model for MCF7 cell line data, parameter perturbation scanning yielded new experimental hypotheses on how ACC and E-Cadherin knock-downs result in distinct responses upon IGF1 and insulin administration as explained in previous **Chapter 2: Statistical Modeling**. The results here suggested that SOS activation through IRS is critical to have the greater MAPK activation in IGF1 stimulation case compared to insulin stimulation. For the E-Cadherin knock-down data, the negative feedback from Akt to IRS is important to obtain the differential Akt activation.

Finally, the feedback from ribosomal protein S6 kinase (p70S6K) on IRS is predicted to differentially affect Akt activation under IGF1 and insulin stimulated cells. The experimental validation of the last prediction showed that there indeed is a difference in the regulation of Akt activity in response to different stimuli, with a greater change induced by insulin rather than by IGF1. The results presented in this chapter, computational and experimental together, showed the importance of interactions of the IRS protein in activating specific cascades in breast cancer cells in response to different stimulus. The results of this chapter provide insights on the mechanistic details of the novel interactions predicted and explained earlier in the previous chapter (**Statistical Modeling**).

3.2 METHODS

3.2.1 Rule-based modeling

The new models are constructed using rule base modeling and BioNetGen [83, 148]. In the RBM scheme, a rule can represent binding, unbinding, modification, production, degradation, catalysis, or any other biochemical or physical event. The rest of this section below will explain the RBM scheme with an example.

The molecules or proteins in the example system can be defined as:

$$L(rec) \qquad \text{Eqn. 8}$$

$$R(lig, phos\sim U\sim P, dim\sim Y\sim N) \qquad \text{Eqn. 9}$$

The Eqn. 8 represents the ligand with a receptor binding domain of “rec”. The Eqn. 9 defines the receptor with three domains of ligand binding “lig”, phosphorylation site of “phos”,

and dimerization domain of “dim”. The “phos” domain is defined to have two different states of un-phosphorylated “U” and phosphorylated “P”. Likewise, the dimerization domain specifies two states of dimerized (Y) or not dimerized (N) receptors.

A rule for binding between these two proteins can then be written as:

$$L(rec) + R(lig) -> L(rec!1).R(lig!1) \quad \text{Eqn. 10}$$

The rule exemplified above in Eqn. 10 is defined by using only specific domains of the two proteins. In the reactions above, the exclamation point, following number, and the dot in between indicate the formation of a physical bond between the two reactants, where the number is only an arbitrary name for the bond. The simple rule encapsulates *all possible* combinations of binding interactions, saving one from manually writing every state variable possible in the system. In that simple rule, one free ligand molecule binds to one, ligand-unbound receptor molecule. That rule applies to all molecules, such that any free, complexed, phosphorylated or un-phosphorylated receptor with an available unbound *lig* domain will bind a receptor-free ligand. So, all of the below interactions occur:

$$L(rec) + R(lig, phos{\sim}U, dim{\sim}N) -> L(rec!1).R(lig!1, phos{\sim}U, dim{\sim}N) \quad \text{Eqn. 11}$$

$$L(rec) + R(lig, phos{\sim}P, dim{\sim}N) -> L(rec!1).R(lig!1, phos{\sim}P, dim{\sim}N) \quad \text{Eqn. 12}$$

$$L(rec) + R(lig, phos{\sim}U, dim{\sim}Y) -> L(rec!1).R(lig!1, phos{\sim}U, dim{\sim}Y) \quad \text{Eqn. 13}$$

$$L(rec) + R(lig, phos{\sim}P, dim{\sim}Y) -> L(rec!1).R(lig!1, phos{\sim}P, dim{\sim}Y) \quad \text{Eqn. 14}$$

All of the four reactions in Eqns. 11-14 occur following the Eqn. 10 and produce the four species of ligand-bound receptors. One can modify the rule in Eqn. 10 to specify only one of reactions to occur. The software can then be used to generate the desired network topology and to run the simulations on-site, both deterministically and stochastically. The network information is exported in desired file format, like SBML [85, 151] or Matlab “.m” file. In this work, all of

the ODEs corresponding to the state variables are obtained from BioNetGen, and the rest of the simulations and analyses are done in Matlab. See Error! Bookmark not defined. for model settings.

3.2.2 Parameter estimation

The dataset utilized in this chapter is the same as in **Chapter 2: Statistical Modeling** [86]. The values of the parameters are estimated using the available reverse phase protein assay (RPPA) data, consisting of the relative expression levels of four phospho-proteins: pReceptor, pAkt, pRPS6K, and pMAPK up to 30 min. The network model outputs are the relative values of phosphorylated protein counts upon stimulation to the total number of that protein. They are named as *observables*. Then, the experimental fold-changes in expression intensities from serum-free media condition to the time point of stimulation are used to fit the model parameters. Three discrete time points of 5, 10, and 30 min are used to calculate the fitting error.

The resulting ODE network model is run in Matlab R2015a [The MathWorks, Inc.] and the parameter estimation is done using Markov chain Monte Carlo (MCMC) sampling (see Error! Bookmark not defined. for the custom script). Also a method called PTempEst is used, available on the web (https://github.com/RuleWorld/ptempest). The latter is a parallel programming running MCMC at different temperatures to search the parameter space both locally and globally. The parallel tempering approach samples Bayesian posterior distribution of each parameter [121,122]. For the prior distributions of each parameter, uniform priors are used. By running multiple chains in parallel at different temperatures provided better sampling. High temperature chains scan the parameter space more globally and the swaps among different chains help avoid getting stuck in local minima. The estimation procedure outputs parameter ensembles

for each chain. Then, the minimum fitting error parameter set is defined as the “best-fit” and is used for all subsequent analyses.

3.2.3 Parameter perturbation scanning

Using the model with the “best-fit” parameter set as a base-line, simulations are run to analyze the response of the system. Each parameter is perturbed individually and for every different value of each parameter, one simulation is run. The model output of the levels of pMAPK and pAkt in the perturbed system is compared to their levels in the un-perturbed model output. The computational results here are comparable to experimental knock-down (up-regulation) of proteins or reaction rates. Based on the results of the simulations here, differential ones between IGF1 and insulin stimulation conditions are selected for further experimental exploration. The associated results are summarized in section “***In silico* experimental predictions**”.

3.2.4 Global sensitivity analysis

In addition to the perturbation scanning, sensitivity analysis of the parameters is also done: (i) change the value of a single parameter by two-folds up or down, (ii) simulate the model, (iii) calculate the ratio of model response to un-perturbed model result, and (iv) take the average of two perturbation results. The procedure is repeated twice for two different output measures. First is the observable pMAPK or pAkt count at 5 min, where the logarithm of the ratios of perturbed/unperturbed are averaged. The second output measure is the ratio of area-under-the-curve (AUC) of each observable in perturbed parameter simulation result to that in original

model result. The integral for AUC is calculated for 0-30 minutes. The results of this analysis are reported in section "**Global sensitivity analysis**".

3.2.5 Parametrization of the network model for other cell lines

The original computational model of MCF7 cell line is used as a starting point to fit parameters to data from six additional cell lines. The data used is from the same study, so there is little concern for comparability [86]. In this part, only some the parameters are changed from the original model. Fourteen of these parameters are the total protein counts. The rest of them are reaction rate constants. The parameter estimation is the same as described earlier in section **Parameter estimation**.

3.2.6 Cell culture and immunoblotting

MCF7 (ATCC) cells are cultured in DMEM (ThermoFisher) with 10% FBS, plated on six well plates at 400,000 cells/well density. The cells, rested overnight, are serum starved for 16-24 hours. Then, the cells are treated with DMSO control or ribosomal protein S6 kinase inhibitor LY2584702 (500 nM, Selleckchem) for three hours. Next, the cells are stimulated with HCl control, IGF1 (10 nM), or insulin (10 nM) for 10 and 30 min. The cells are harvested and protein concentration is quantified by BCA. Samples are collected using RIPA buffer (50mM Tris pH 7.4, 150mM NaCl, 1mM EDTA, 0.5% Nonidet P-40, 0.5% NaDeoxycholate, 0.1% SDS) with 1x HALT protease & phosphatase cocktail (ThermoFisher). The immunoblotting is done using 12% acrylamide gels and PVDF membrane transfer (Millipore #IPFL00010, 0.45µm). Membranes are blocked in Odyssey PBS Blocking Buffer (LiCor), and incubated in primary antibodies

overnight: Akt S473 (Cell Signaling #4060; 1:1000), total Akt (Cell Signaling #2920; 1:1000), S6 S235/236 (Cell Signaling #4858, 1:1000), total S6 (Cell Signaling #2217, 1:1000), and β-actin (Sigma #A5441; 1:5000). Membranes were incubated in LiCor secondary antibodies for 1 hour (anti-rabbit 800CW, LiCor #926-32211 or anti-mouse 680LT, LiCor #925-68020; 1:10,000). The imaging is done at Odyssey Infrared Imager. The quantification of the blots is done in LiCoR Image Studio Lite v5.2 software.

T47D (ATCC) and ZR75-1 (ATCC) cells are cultured in RPMI-1640 (HyClone, GE) with 10% FBS, 1% glutamine, and 1% Penicillin-Streptomycin, plated on six well plates at 500,000 cells/well density. The cells, rested overnight, are serum starved for 16-24 hours. Then, the cells are treated with DMSO control or ribosomal protein S6 kinase inhibitor LY2584702 (1 μM) overnight. Next, the cells are stimulated with HCl control, IGF1 (10 nM), or insulin (10 nM) for 10 and 30 min. The cells are harvested and protein concentration is quantified by Bradford absorbance assay. Samples are collected using HEPES buffer (1% Triton X-100, 10% Glycerol, 5mM $MgCl_2$, 25mM NaF, 1mM EGTA, 10mM NaCl) with 1x HALT protease & phosphatase cocktail (ThermoFisher). The immunoblotting is done using 12% acrylamide gels (ThermoFisher #XP00125BOX) and PVDF membrane transfer (ThermoFisher #LC2002, 0.2 μm). Membranes are blocked in 5% milk in 1x TBST solution (TBST: Tris Buffered Saline (Sigma # T6664) with 0.1% Tween20), and incubated in primary antibodies overnight: Akt S473 (Cell Signaling #4060; 1:1000), total Akt (Cell Signaling #2920; 1:1000), S6 S235/236 (Cell Signaling #4858, 1:1000), total S6 (Cell Signaling #2217, 1:1000), and β-actin (Sigma #A5441; 1:5000). Membranes were incubated in HRP secondary antibodies for 45 min (anti-rabbit, Jackson #111-035-003 or anti-mouse, Jackson #115-035-003; 1:8,000). The imaging is done by

chemiluminescence using ECL substrates (BioRad #170-5060). Imaging is done at Philipps L4000 Imager. The quantification of the blots is done in LiCor Image Studio Lite v5.2 software.

3.3 RESULTS

3.3.1 Results for MCF7 cell line

3.3.1.1 The final network topology

Based on insights from current knowledge, a mechanistic model of IGF1 and insulin signaling pathways is constructed. The mechanistic approach of representing both IGF1R and InsR signaling cascades by a set of ODEs generated by rule-based modeling enabled us to study the response of the system under different conditions.

The final model, depicted in **Figure 19**, includes 16 proteins (two ligands, two receptors, and twelve downstream proteins), 38 species, 16 parameters for total protein numbers, and 50 parameters for rate constants. Thirty-four of the rate parameters are for cascades downstream of IGF1 and insulin signaling. There are eight specific parameters for specific receptor events.

3.3.1.2 Computational model simulation vs experimental training data

The resulting parameter set of the network model (**Figure 19**) is estimated and simulation results convey qualitative and quantitative agreement to experimental data for both IGF1R and InsR signaling. **Figure 20** shows the performance of the "best-fit" (see section **3.2.2. Parameter estimation** for details) parameter set model on training data. The computational model outputs

fall into the range of experimental standard deviations. These plots convey overall good performance of the MCF7 model parametrization.

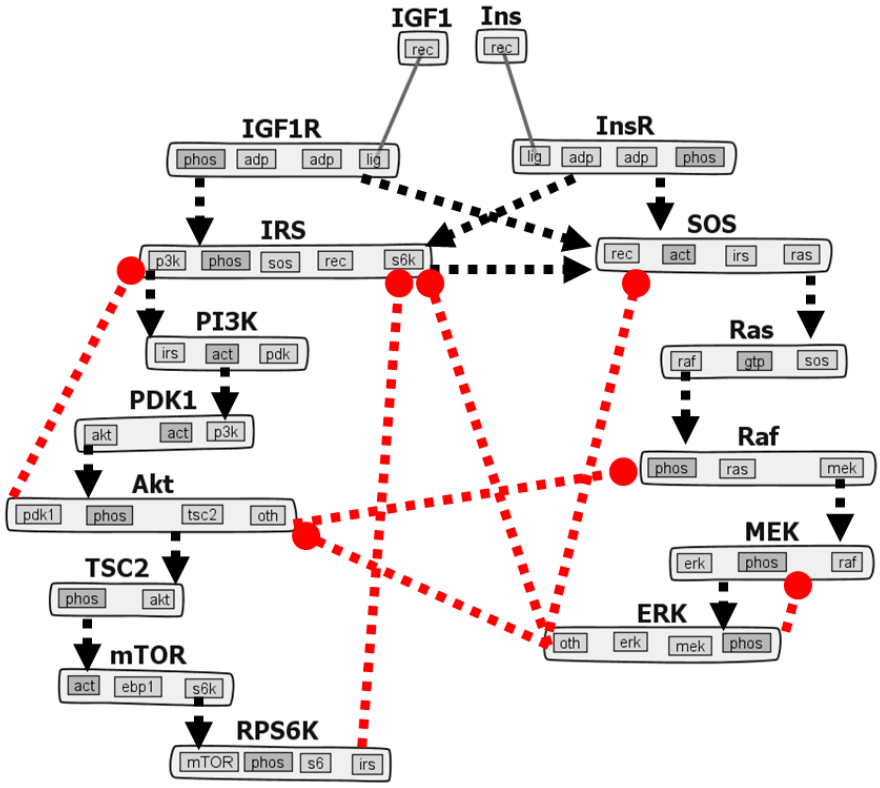

Figure 19. The computational mechanistic model representation. The topology of dual IGF1R/InsR signaling network is illustrated. The model includes 16 proteins. Black arrows represent activation and red lines indicate inhibition of the corresponding active molecule. The figure is generated in RuleBender software [83].

In addition to the best-fit models of IGF1 and insulin stimulation, the ODE models are simulated for an ensemble of parameter sets. Each parameter set has different numbers of total protein numbers, spanning a large range of possible "cell" conditions. **Figure 21** shows that within that range of parameters, the models fit to the data and indicate a decent set of rate parameters are estimated.

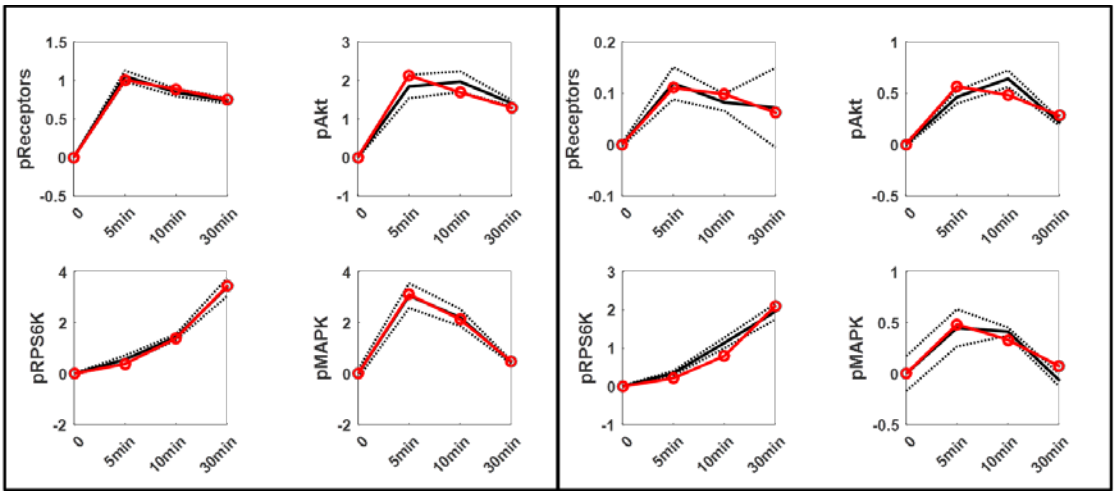

Figure 20. The ODE model simulation. The simulations are up to 30 min for IGF1 (left) and insulin (right) stimulation. The experimental data for phospho-receptors, pAkt, pS6K, and pMAPK are represented with black lines. The solid black line is the mean of triplicate experiments and dashed lines are the S.E.M. Red curves are the simulation outputs. The values on the y-axis represent scaled protein numbers.

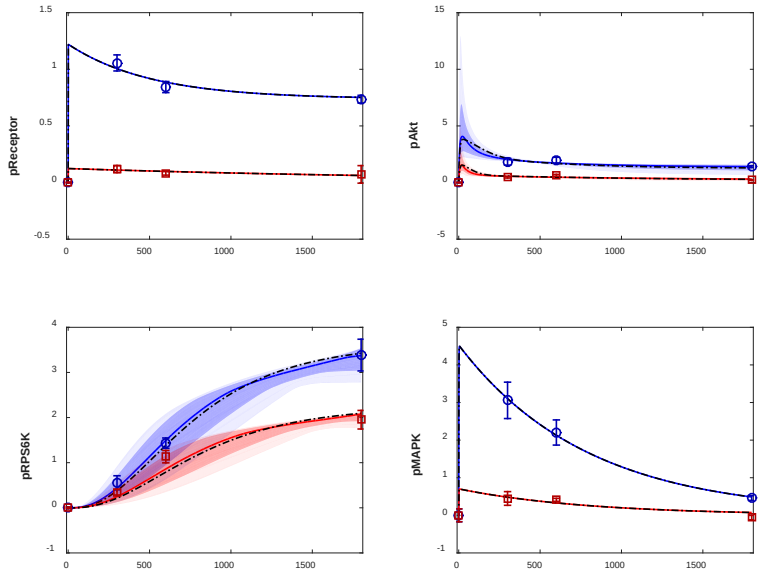

Figure 21. The ODE model simulation ensemble. The ODE model is simulated 10000 times more with an ensemble of parameter sets, with different protein numbers. The initial counts for the proteins (14 in total) are sampled using Latin Hypercube Sampling (LHS) and is model output is scaled individually. The plots show 5%-95% (light) and 15%-85% (dark) confidence intervals for IGF1 (blue) and insulin (red) models. The circles with error bars are the corresponding RPPA data points. The dashed black lines are the trajectories for "best-fit" parameter sets.

3.3.1.3 IGF1 dose response and PI3K inhibition in MCF7 cells

After the best-fit model parameters are obtained, the model performance is tested using dose response data of IGF1 and PI3K inhibition data with IGF1 stimulation. The results for IGF1 dose response are obtained by only changing the level of IGF1 input into the system. The results are shown in **Figure 22**, where most of the results are within acceptable range and show qualitatively good agreement. Though, there is a discrepancy in MAPK response, which is mostly due to the simplifying assumptions of the model topology. There are many details of the MAPK phosphorylation, like scaffold proteins and double-phosphorylation events, which are omitted here. The single cascade of phosphorylation studied in this model is able to recapitulate the system dynamics at the higher concentrations but not at lower IGF1 concentrations. However, the model results fall into the experimental replicate error range for IGF1 concentration of 10 nM, which the actual model parameters are trained at. The model overall thus provided adequate performance on IGF1 dose response.

A second tier of MCF7 cell line computational model test is done by testing the model on PI3K and mTOR inhibition data. One specific PI3K inhibitor, called LY294002, and another inhibitor of dual PI3K and mTOR named BEZ235 are used. The inhibitors are administered alone or in combination with 15 min of IGF1 (10 nM) stimulation.

In the modeling part, the rate parameter of PI3K action on the activation of PDK1 is decreased by 90% to recapitulate the PI3K inhibitor LY294002 mechanism of action. The rate of S6K activation through mTOR and the previous rate constant are both decreased by 90% for the second inhibitor action of BEZ235. Without any ligand input, neither of the inhibitors affects any downstream signaling, as shown in **Figure 23**. Addition of IGF1 into the system after the inhibitor does activate signaling and increase the phosphorylation of Akt, although to a lesser

extent. The S6 kinase phosphorylation is diminished in the experiments, however the model results in less activation but not full. This might be due to the fact that these inhibitors are dirtier than the simulated conditions. All in all, the model performs adequately to capture these inhibition effects with only one or two rate constant alterations.

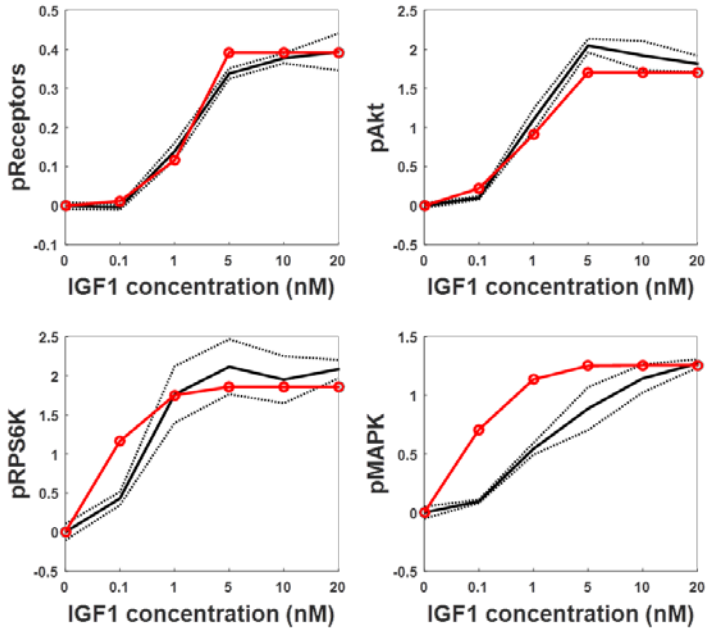

Figure 22. The computational model recapitulates IGF1 dose response data. The solid black line is the mean of triplicate experiments and dashed lines are the S.E.M. Red curves are the simulation outputs. The values on the y-axis represent scaled protein numbers.

3.3.2 In silico experimental predictions

The estimated parameter set is tested with IGF1 dose response and inhibitor datasets. The next step undertaken is then to generate new hypotheses. First, to utilize earlier results of validated differential interactions explained in **Chapter 2: Statistical Modeling**, I employed a parameter scan to search for specific perturbations capturing the effects seen in the knock-down

experiments of ACC and E-Cadherin. The parameter perturbation explained earlier (**3.2.3. Parameter perturbation scanning**) is utilized and the scanning yielded new experimental hypotheses on how differential responses of MAPK and Akt might originate.

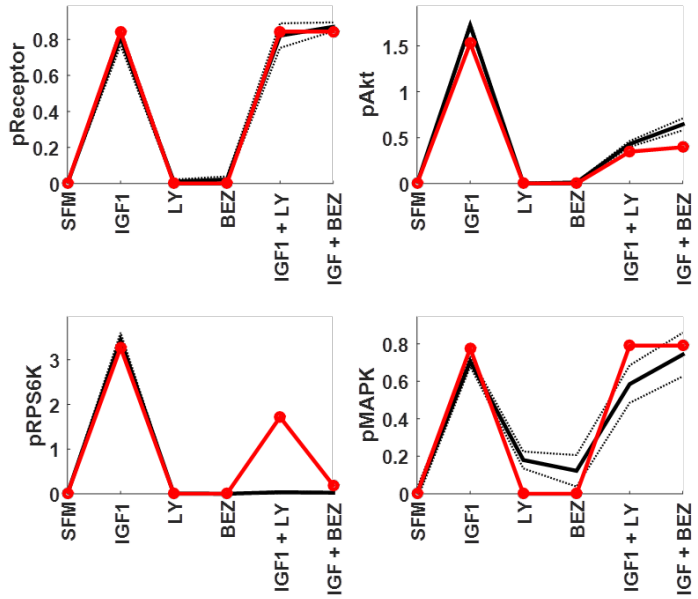

Figure 23. The computational model recapitulates PI3K inhibition data. The columns of x-axis correspond to control, IGF1 (10 nM) stimulation, first inhibitor only, second inhibitor only, first inhibitor and IGF1 (10 nM), and second inhibitor with IGF1 (10 nM) stimulation. The receptor and MAPK phosphorylation are not affected with PI3K and mTOR inhibition whereas Akt and S6 kinase phosphorylation is decreased. The inhibitions are simulated in the computational models as a 90% reduction in the corresponding rate constant(s). The solid black line is the mean of triplicate experiments and dashed lines are the S.E.M. Red curves are the simulation outputs. The values on the y-axis represent scaled protein numbers.

First, the MCF7 mechanistic model is analyzed and important interactions leading to differential responses are determined. Changing the value of each parameter individually and analyzing the resulting changes in Akt phosphorylation levels showed that the negative feedback of Akt on upstream adaptor protein IRS is important for E-Cadherin function on Akt activation.

Figure 24 shows that the rate parameter 63, or kf208, causes up-regulation Akt activation and the effect is differential between IGF1 and insulin stimulation. It is further shown that the result is different at 30 min, with a larger increase indicated for IGF1 stimulation. This recapitulates the earlier E-Cadherin knock-down experimental results.

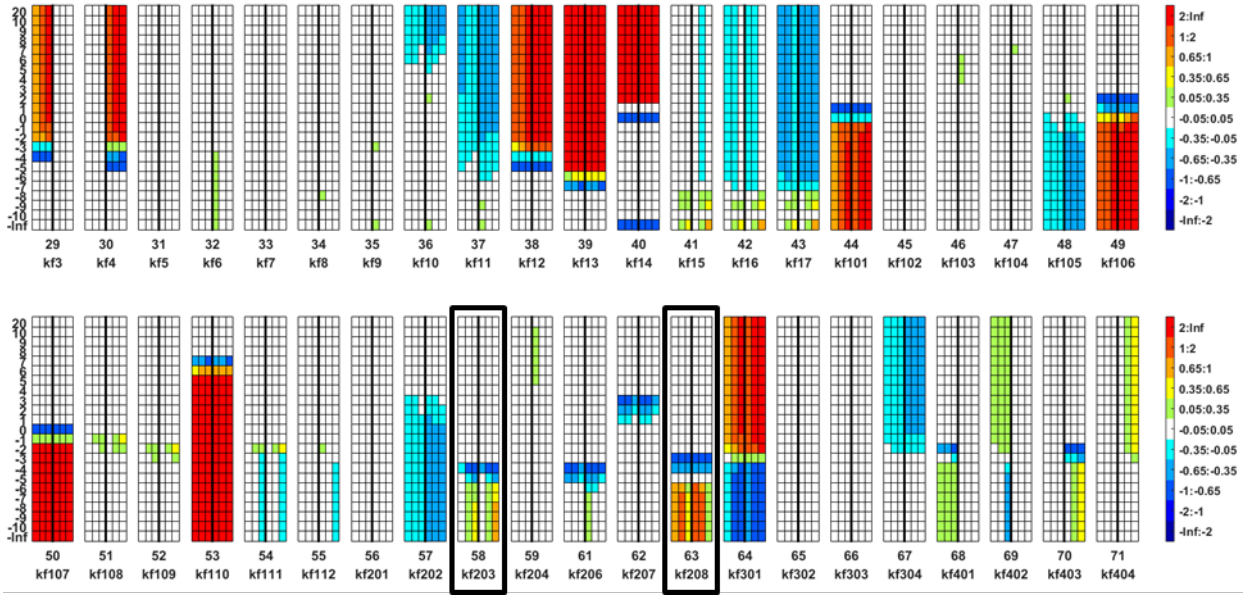

Figure 24. *In silico* perturbation of rate constants reveal differential effects on Akt phosphorylation. The ODE model output of pAkt levels are determined by setting and changing the value of each parameter individually, from zero to infinity. Each box above is for a parameter, and each column is for one time point response of 5, 10, and 30min. Left three columns are from IGF1 stimulation results and the right three columns are from insulin stimulated simulations. Rows represent the log10 of set parameter value. The colors represent the log2 fold change from un-perturbed model output. Red and blue respectively indicate up and down regulation of the Akt phosphorylation.

The second analysis is carried out the determine clues for ACC action on MAPK phosphorylation. Previously, it was shown that the ACC knock-down causes an increase in MAPK activation, with a larger change induced by IGF1 than by insulin. The parameter perturbation scan results showed that the up-regulation of the rate of SOS activation by IRS

protein should be explored further to pinpoint the mechanism of action of ACC on MAPK. **Figure 25** depicts the results of the parameter perturbation scan. The differential MAPK activation upon ACC knock-down is best captured by the rate parameter 33, or k7. It is of note that the differential response is captured by "up-regulation" of a rate, rather than a knock-down. It is non-trivial and considerably hard to experimentally validate such a prediction.

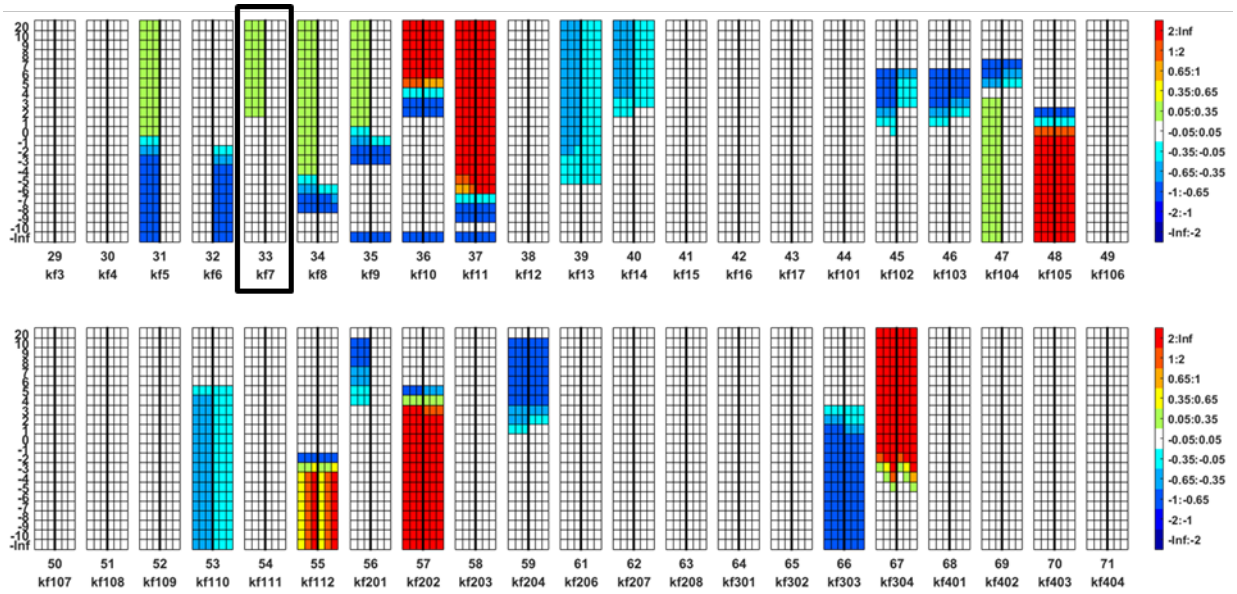

Figure 25. *In silico* perturbation of rate constants reveal differential effects on MAPK phosphorylation. The ODE model output of pMAPK levels are determined by setting and changing the value of each parameter individually, from zero to infinity. Each box above is for a parameter, and each column is for one time point response of 5, 10, and 30min. Left three columns are from IGF1 stimulation results and the right three columns are from insulin stimulated simulations. Rows represent the log10 of set parameter value. The colors represent the log2 fold change from un-perturbed model output. Red and blue respectively indicate up and down regulation of the MAPK phosphorylation.

Finally, other hypotheses of differential regulation of Akt and MAPK phosphorylation are explored based on the results of parameter perturbation scanning. One of the predictions with a differential response from IGF1 and insulin is the knock-down of ribosomal protein S6 kinase.

The prediction is that upon inhibiting S6 kinase, the insulin stimulated cells would have increased Akt phosphorylation at 30 min, and that the magnitude of the increase will be larger than that in the IGF1 stimulated cells. This is shown in **Figure 24** with the rate parameter 58 or kf203. This prediction is experimentally validated and the results will be reported in section **3.4. Experiments**.

3.3.2.1 Global sensitivity analysis

Parameter perturbation scanning showed that some rate parameters can differentiate MAPK and Akt responses, between IGF1 and insulin stimulation conditions. It is also apparent from **Figure 24** that there are certain ranges of parameter values where knock-down or up-regulation experiments show differential signaling.

To complement the perturbation scanning, a global sensitivity analysis of the model is carried out. Following the steps explained in section **3.2.4. Global sensitivity analysis**, the computational model parameters are shown to play a critical role on the network dynamics. The two different measures of sensitivity, the ratio of observed protein numbers with and without the perturbation, and the ratio of the AUC of the response variable in perturbed and un-perturbed systems showed similar results, see **Figure 26** and **Figure 27**. The parameters detected as important for each response are indeed the rate parameters of the corresponding cascade.

3.3.3 Model parameter estimation for other cell lines

The mechanistic model parameters are re-estimated with experimental data from other cell lines. The list of these cell lines are given in **Table 5** and the parameters that are different from the MCF7 model are reported in **Table 6**. The six additional cell lines are from an array of different

subtypes of breast cancer. The parameter estimation procedure is the same as in MCF7 model. **Figure 28-30** show the model fits for the six additional cell lines. The model fitting of these cell lines will provide additional information on what are the differences between sub-types of cells and how accurately they can be modeled. Exploration of differences in the model parameters as well as model responses under different perturbations will provide comparison to MCF7 cell line.

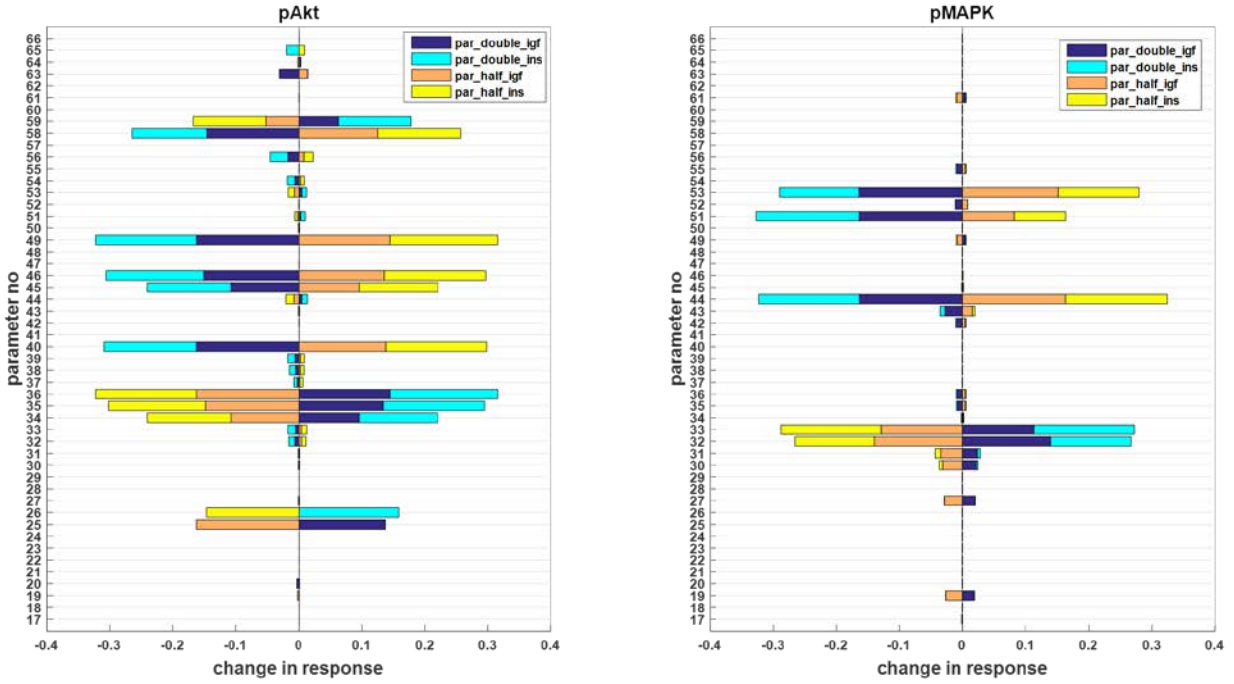

Figure 26. The parameter sensitivity analyses of Akt and MAPK responses. Changing the individual parameter values two times up or down causes a change in Akt and MAPK phosphorylation levels. The values reported indicate the ratio of perturbed model response protein level to the level of the protein of un-perturbed model. Dark and light blue colors represent doubled parameter rate conditions with IGF1 and insulin stimulations, respectively. Orange and yellow colors represent halved parameter rate conditions with IGF1 and insulin stimulations, respectively. These two opposite manipulations result in opposite changes in the phosphorylation levels.

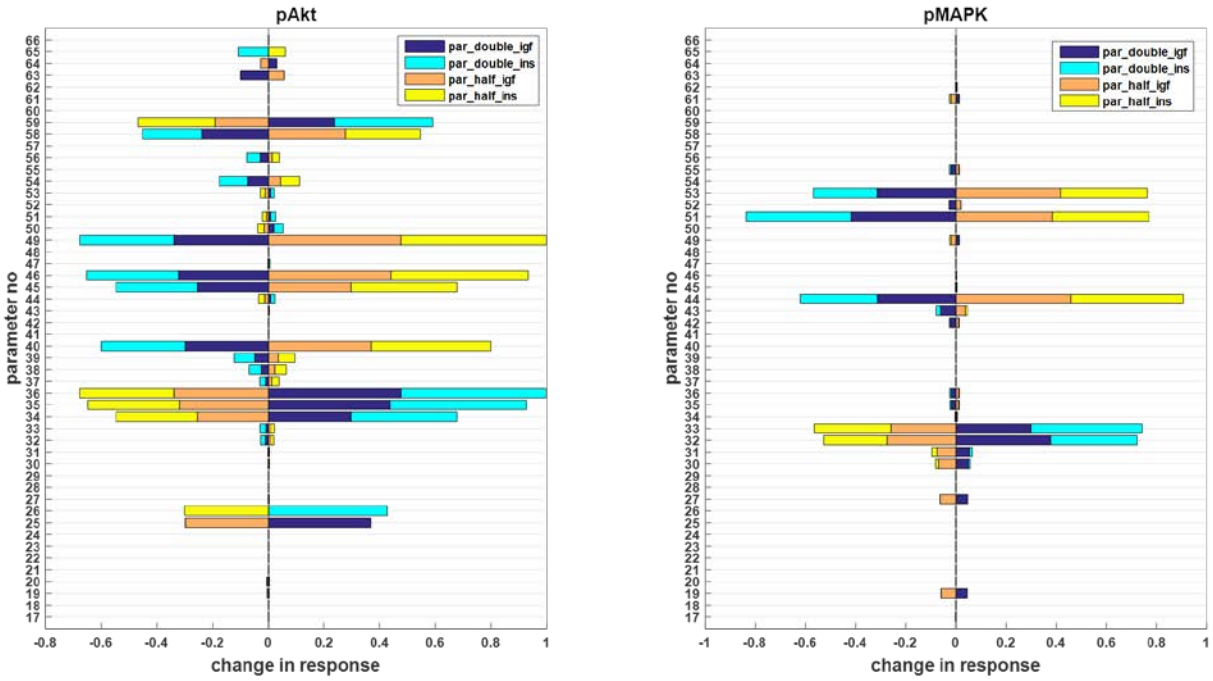

Figure 27. The parameter sensitivity analyses of Akt and MAPK responses. Changing the individual parameter values two times up or down causes a change in pAkt and pMAPK levels. The values reported indicate the ratio of the AUC of the corresponding protein in perturbed model to the AUC of that protein in un-perturbed model. Dark and light blue colors represent doubled parameter rate conditions with IGF1 and insulin stimulations, respectively. Orange and yellow colors represent halved parameter rate conditions with IGF1 and insulin stimulations, respectively. These two opposite manipulations result in opposite changes in the phosphorylation levels.

The parameter perturbation scanning for these cell lines showed that the S6 kinase inhibition causes different differential regulation of Akt dynamics in some of the cell lines. The predictions for cell lines are summarized in **Table 5**. It is seen that the same clinical sub-type of cell lines might have different regulatory mechanisms in the Akt cascade and this prediction is followed further with wet-lab experimentation, explained in the next section "**Experiments**".

Table 5. The information of modeled cell lines.

CELL LINE	SUBTYPE	RECEPTOR	MUTATIONS
MCF7	luminal	HR+	PIK3CA, GATA3
HCC70	basal	TNBC	PTEN, TP53
HCC1954	basal	Her2amp	PIK3CA, TP53
MM415	luminal	HR+	MAP3K1, PTEN, TP53, MAP2K4
T47D	luminal	HR+	MLL3, PIK3CA, TP53
UACC812	luminal	Her2amp	CDH1
ZR751	luminal	HR+	PTEN

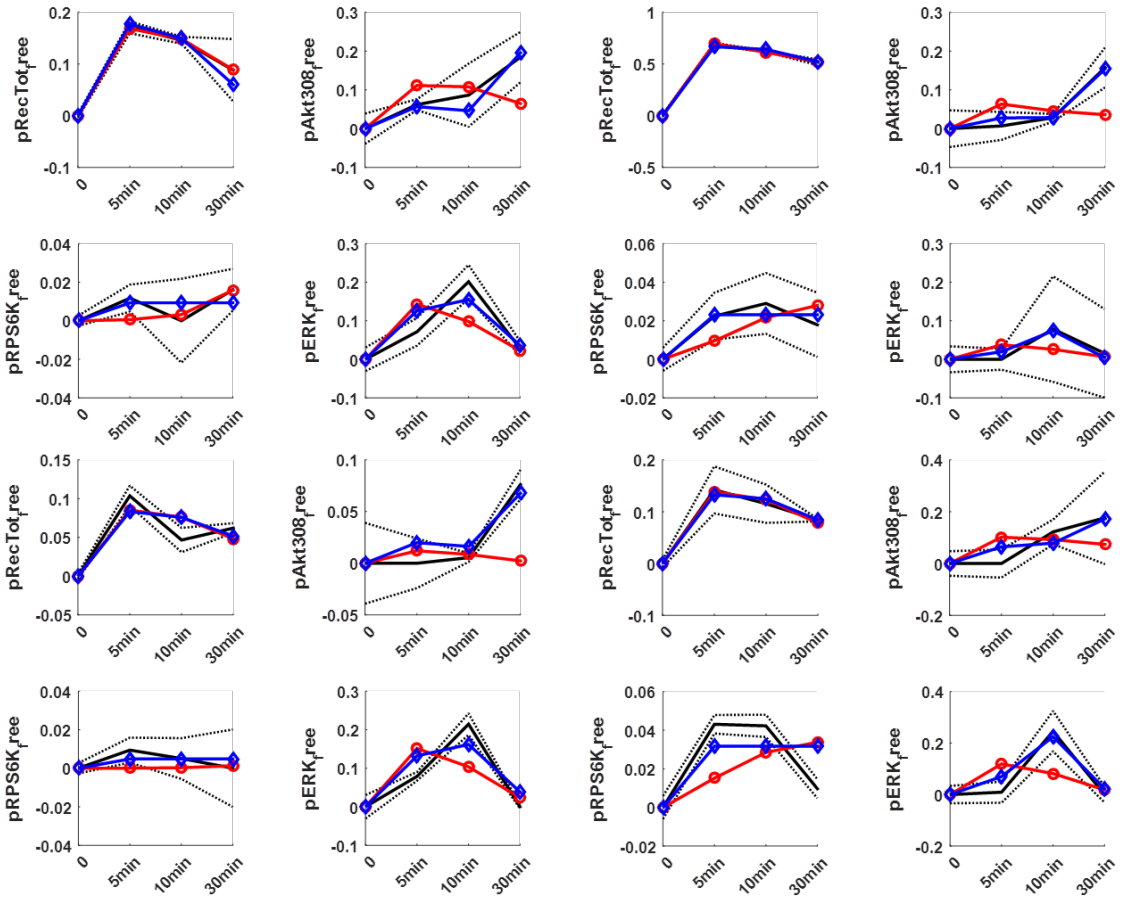

Figure 28. The ODE model performance of HCC70 and HCC1954 cell lines. IGF1 (left) and insulin (right) stimulation responses are shown. The experimental data for phospho-receptors, pAkt, pS6K, and pMAPK are represented with black lines. The solid black line is the mean of triplicate experiments and dashed lines are the

S.E.M. Red curves are the simulation outputs of the models with MCF7 “best-fit” parameter set. Blue curves are the simulation outputs of the models with “best-fit” parameter set of the corresponding cell line. The values on the y-axis represent scaled protein numbers.

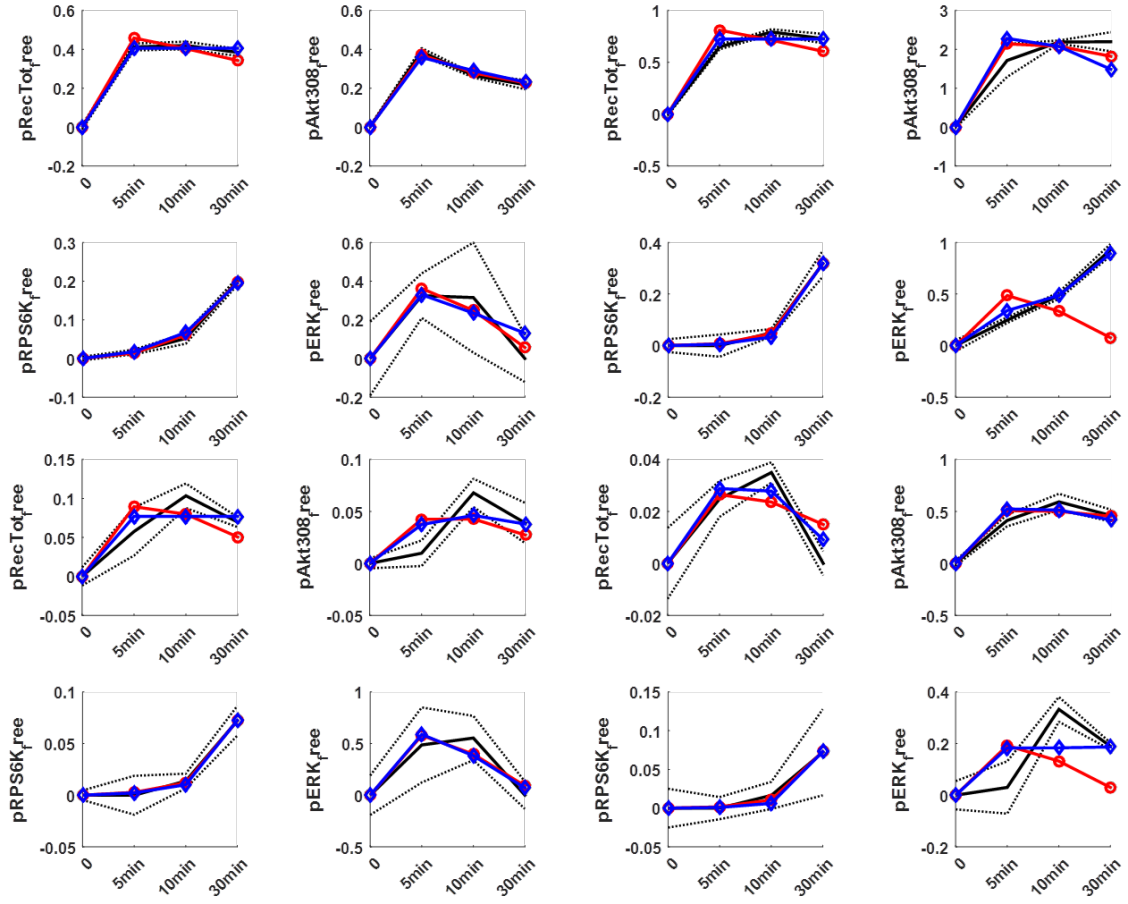

Figure 29. The ODE model performance of MDA-MB-415 and T47D cell lines. IGF1 (left) and insulin (right) stimulation responses are shown. The experimental data for phospho-receptors, pAkt, pS6K, and pMAPK are represented with black lines. The solid black line is the mean of triplicate experiments and dashed lines are the S.E.M. Red curves are the simulation outputs of the models with MCF7 “best-fit” parameter set. Blue curves are the simulation outputs of the models with “best-fit” parameter set of the corresponding cell line. The values on the y-axis represent scaled protein numbers.

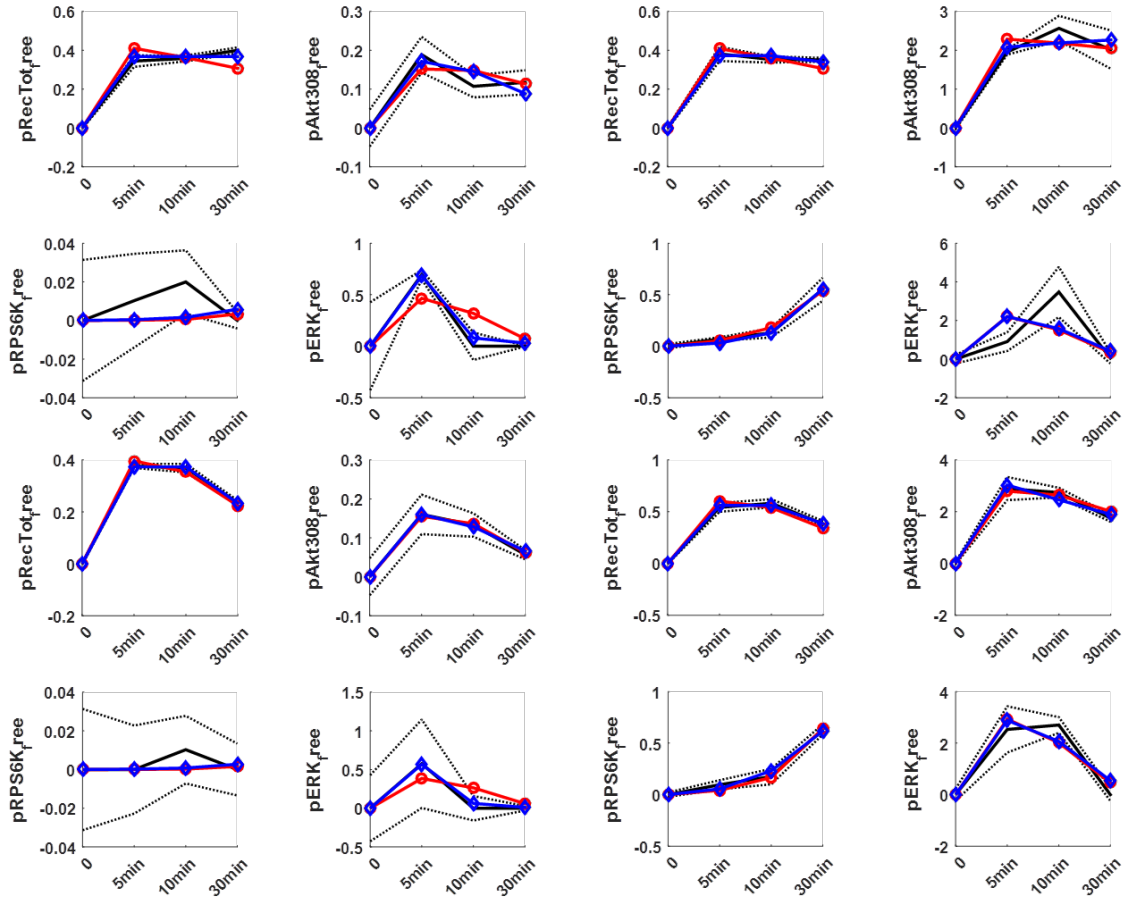

Figure 30. The ODE model performance of UACC812 and ZR75-1 cell lines. IGF1 (left) and insulin (right) stimulation responses are shown. The experimental data for phospho-receptors, pAkt, pS6K, and pMAPK are represented with black lines. The solid black line is the mean of triplicate experiments and dashed lines are the S.E.M. Red curves are the simulation outputs of the models with MCF7 "best-fit" parameter set. Blue curves are the simulation outputs of the models with "best-fit" parameter set of the corresponding cell line. The values on the y-axis represent scaled protein numbers.

Table 6. Best-fit parameter set for each cell line. Red numbers indicate parameters different from the MCF7 model.

		MCF7	HCC70	HCC1954	MDA-MB-415	T47D	UACC812	ZR75-1
1	'IGF1R_0'	**25000**	**77690**	**48434**	**40461**	**1057**	**13613**	**2708**
2	'INSR_0'	**25000**	**98718**	**90648**	**1275**	**65698**	**67016**	**81608**
3	'IRS_0'	**92766**	**29732**	**223611**	**22739**	**163403**	**38752**	**28112**
4	'SOS_0'	**90075**	**24404**	**80625**	**96586**	**49408**	**229260**	**130811**
5	'RAS_0'	**230642**	**163585**	**35381**	**71208**	**103342**	**125283**	**202775**
6	'RAF_0'	**126069**	**26005**	**175763**	**236131**	**222812**	**28043**	**206997**
7	'MEK_0'	**1098164**	**1157715**	**891267**	**920260**	**1009279**	**426163**	**1121731**
8	'ERK_0'	**763172**	**1105929**	**1088177**	**1184028**	**601763**	**1148490**	**1106150**
9	'PI3K_0'	**64009**	**132505**	**100200**	**59910**	**102132**	**107743**	**125705**
10	'PDK1_0'	**186081**	**52093**	**135844**	**184542**	**239159**	**138702**	**127893**
11	'AKT_0'	**432907**	**424072**	**794767**	**675586**	**537954**	**573200**	**598046**
12	'TSC2_0'	**131339**	**186051**	**170078**	**66281**	**43600**	**64359**	**122720**
13	'MTOR_0'	**83469**	**125539**	**116430**	**67028**	**56110**	**156525**	**30685**
14	'RPS6K_0'	**121978**	**214285**	**58190**	**153225**	**95568**	**226526**	**24365**
15	'kf1'	**0.484**	0.484	**-5.813**	0.484	0.484	0.484	0.484
16	'kf1b'	**-2.915**	-2.915	**4.552**	-2.915	-2.915	-2.915	-2.915
17	'kf1c'	**2.987**	2.987	**4.114**	2.987	2.987	2.987	2.987
18	'kf1d'	**1.205**	1.205	**-1.685**	1.205	1.205	1.205	1.205
19	'kf2'	**4.631**	4.631	**0.887**	4.631	4.631	4.631	4.631
20	'kf2b'	**-0.867**	-0.867	**-4.800**	-0.867	-0.867	-0.867	-0.867
21	'kf2c'	**4.876**	4.876	**-1.695**	4.876	4.876	4.876	4.876
22	'kf2d'	**-2.653**	-2.653	**-4.321**	-2.653	-2.653	-2.653	-2.653
23	'kf3'	**-2.791**	-2.791	**-0.746**	-2.791	-2.791	-2.791	-2.791
24	'kf4'	**-3.190**	-3.190	**-0.790**	-3.190	-3.190	-3.190	-3.190
25	'kf5'	**-0.692**	**-7.820**	**-6.371**	**-4.520**	**-6.733**	**2.765**	**0.715**
26	'kf6'	**4.125**	**0.650**	**-6.532**	**1.253**	**-3.982**	**-1.718**	**-6.009**
27	'kf7'	**-3.040**	**-1.558**	**-7.377**	**-3.560**	**-2.879**	**-2.663**	**-3.623**
28	'kf8'	**-4.576**	**-2.085**	**-7.576**	-4.576	-4.576	-4.576	-4.576
29	'kf9'	**0.353**	**0.312**	**-1.121**	0.353	0.353	0.353	0.353
30	'kf10'	**4.331**	**4.484**	**-7.919**	4.331	4.331	4.331	4.331
31	'kf11'	**-6.932**	**-7.058**	**1.549**	-6.932	-6.932	-6.932	-6.932
32	'kf12'	**-3.802**	-3.802	**-7.683**	-3.802	-3.802	-3.802	-3.802
33	'kf13'	**-6.248**	**-1.892**	**-2.536**	**-6.167**	**-7.710**	**-6.312**	**-6.974**
34	'kf14'	**1.000**	1.000	**4.154**	1.000	1.000	1.000	1.000
35	'kf15'	**-6.779**	-6.779	**-0.877**	-6.779	-6.779	-6.779	-6.779
36	'kf16'	**-7.789**	**3.291**	**0.279**	**-7.612**	**-7.998**	**-7.976**	**-7.304**
37	'kf17'	**-7.712**	-7.712	**-1.210**	-7.712	-7.712	-7.712	-7.712
38	'kf101'	**-0.083**	**-2.215**	**2.628**	**-0.505**	**-0.177**	**-0.289**	**-0.033**
39	'kf102'	**-6.473**	**-1.750**	**-3.181**	**-1.868**	**-5.759**	**4.118**	**1.634**

40	'kf103'	**0.048**	**-6.620**	**3.484**	**-2.624**	**0.368**	**-3.289**	**2.464**
41	'kf104'	**3.904**	**-0.567**	**-2.153**	**4.498**	**3.308**	**2.069**	**-0.794**
42	'kf105'	**0.759**	**-1.151**	**-2.445**	**0.606**	**1.957**	**-1.009**	**0.932**
43	'kf106'	**0.514**	**4.950**	**-5.818**	0.514	0.514	0.514	0.514
44	'kf107'	**-0.868**	**-1.387**	**-3.695**	-0.868	-0.868	-0.868	-0.868
45	'kf108'	**-3.345**	**2.567**	**-2.730**	-3.345	-3.345	-3.345	-3.345
46	'kf109'	**-4.510**	**-1.519**	**-3.954**	-4.510	-4.510	-4.510	-4.510
47	'kf110'	**6.393**	**-6.089**	**1.056**	6.393	6.393	6.393	6.393
48	'kf111'	**-2.796**	**-1.399**	**-2.257**	-2.796	-2.796	-2.796	-2.796
49	'kf112'	**-2.900**	**-1.615**	**-2.078**	**-2.840**	**-5.860**	**-2.133**	**-2.954**
50	'kf201'	**2.763**	2.763	**-5.692**	2.763	2.763	2.763	2.763
51	'kf202'	**4.077**	4.077	**-6.653**	4.077	4.077	4.077	4.077
52	'kf203'	**-5.444**	-5.444	**3.117**	-5.444	-5.444	-5.444	-5.444
53	'kf204'	**0.227**	0.227	**-1.840**	0.227	0.227	0.227	0.227
54	'kf206'	**-6.351**	-6.351	**-1.219**	-6.351	-6.351	-6.351	-6.351
55	'kf207'	**-6.375**	**4.545**	**3.128**	**1.519**	**2.643**	**-0.282**	**-6.584**
56	'kf208'	**-5.006**	**-7.987**	**-5.042**	**-4.912**	**-4.974**	**-5.105**	**-4.857**
57	'kf301'	**-3.094**	**-3.092**	**-0.934**	-3.094	-3.094	-3.094	-3.094
58	'kf302'	**-0.981**	**3.987**	**-6.873**	-0.981	-0.981	-0.981	-0.981
59	'kf303'	**4.279**	**-4.902**	**-6.485**	4.279	4.279	4.279	4.279
60	'kf304'	**-5.656**	**-5.180**	**-4.108**	-5.656	-5.656	-5.656	-5.656
61	'kf401'	**-3.102**	**-3.111**	**-3.085**	**-3.410**	**-2.199**	**-2.971**	**-1.690**
62	'kf402'	**-2.925**	**0.258**	**3.070**	**1.551**	**2.058**	**0.303**	**0.838**
63	'kf403'	**-3.419**	**-3.496**	**-3.481**	**-1.621**	**-3.032**	**-3.272**	**-3.473**
64	'kf404'	**-7.081**	**1.456**	**2.075**	**2.071**	**2.077**	**0.112**	**2.952**

3.4 EXPERIMENTS

3.4.1 MCF7 cell line and ribosomal protein S6 kinase (RPS6K) inhibition

The prediction from parameter scanning of MCF7 model parameters is experimentally validated. The ribosomal protein S6 kinase is inhibited in MCF7 cells and standard western blot procedures are followed, as described in section **Cell culture and immunoblotting**. The experiments are

carried out four times, and the representative blots and the final quantifications are shown in **Figure 31** and in **Figure 32**, respectively.

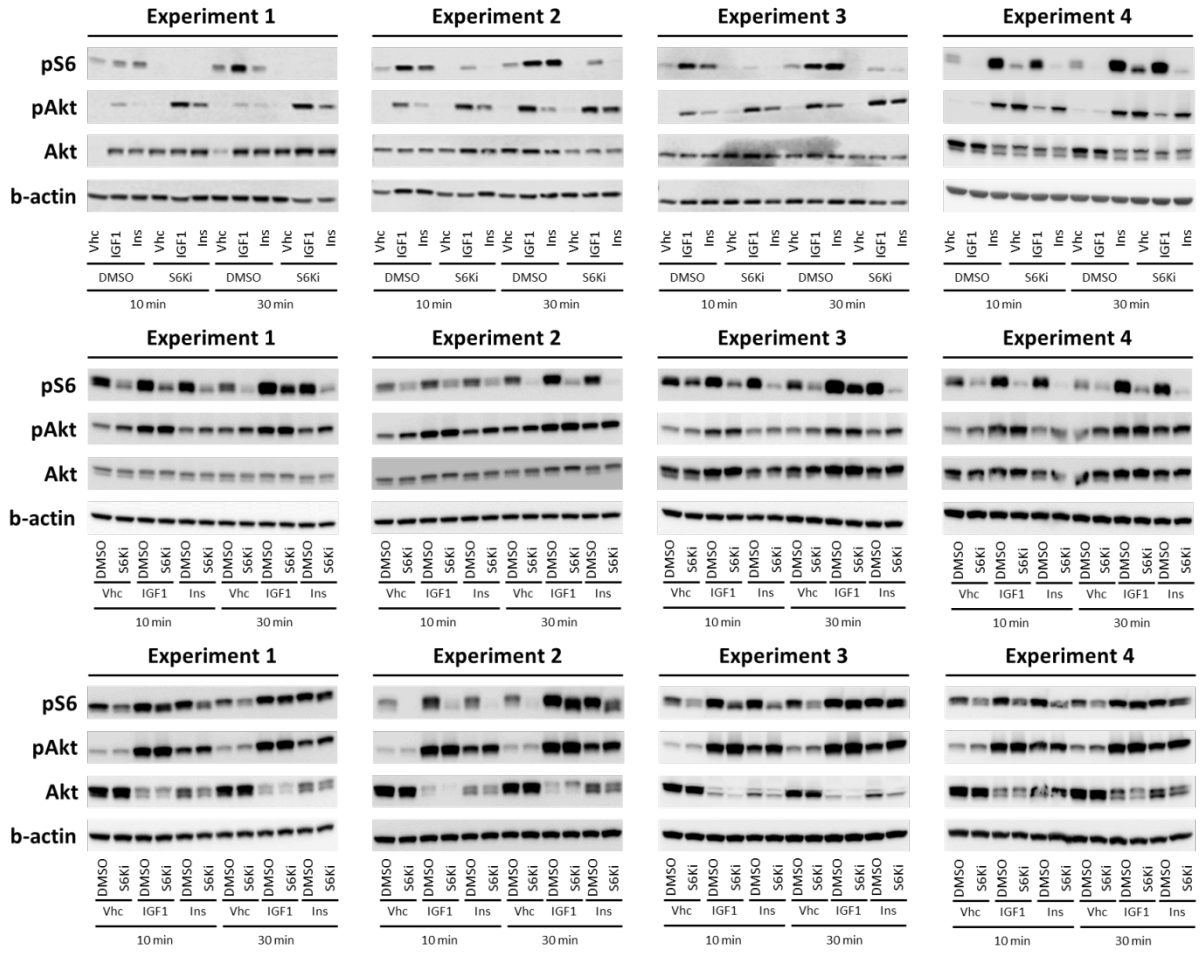

Figure 31. Ribosomal S6 kinase inhibition up-regulates Akt phosphorylation. The RPS6K is inhibited as described in section **3.2.6. Cell culture and immunoblotting**. In all three cell lines of MCF7 (top row), T47D (middle row), and ZR75-1 (bottom row), the inhibition increases pAkt levels at 10 and 30 min. The pS6 levels are used as the proxy for S6K inhibition efficiency in un-stimulated cells.

The knock-down of the S6 kinase is captured by the decrease in S6 phosphorylation levels (**Fig. 31**). In MCF7 cell line, almost all of the S6 proteins become un-phosphorylated upon inhibition (top row in **Fig. 32**), even at the 10 and 30 min stimulation with IGF1 and insulin.

With the inhibition, the level of Akt phosphorylation (S473) increases in insulin stimulated cells, compared to insignificant increase in IGF1 stimulated cells. This result is in compliance with the computational prediction of the MCF7 mechanistic model (**Table 5**).

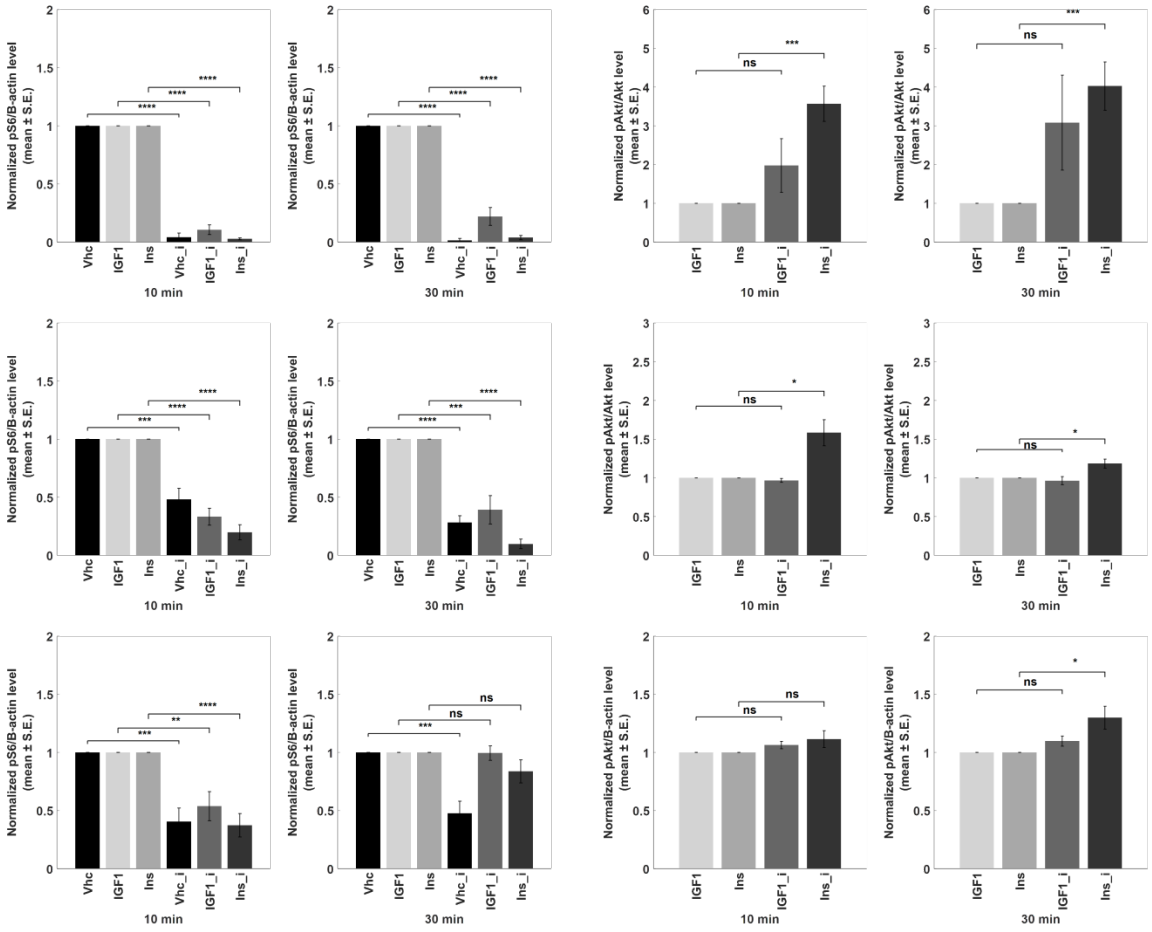

Figure 32. Quantification of ribosomal S6 kinase inhibition. The values are reported as normalized to the corresponding no-inhibition control. S6 phosphorylation quantification (left columns) at 10 and 30 min show inhibition efficiency and pAkt levels (right columns) represent the response of the cells to the perturbation. MCF7 (top row) and T47D (middle row) cell lines showed higher up-regulation of pAkt in insulin stimulated cells at 30 min and ZR75-1 (bottom row) showed no significant increase at 30 min. The results are compared using unpaired, one-tailed two-sample t-test, and $P<0.05$ (*), $P<0.01$ (**), $P<0.005$ (***), nonsignificant (ns). Results shown are mean ± s.e.m. of four independent replicates.

3.4.2 RPS6 kinase inhibition in T47D and ZR75-1 cells

The RPS6K is inhibited in T47D and ZR75-1 cell lines. The predictions for the effect of the inhibition in these cell lines are different from that in MCF7 cell line (see **Table 5**). The representative immunoblots and the quantifications are given in **Fig. 31** & **Fig. 32**, respectively. The figures show that the S6 kinase inhibition efficiency is lower in these cell lines than in MCF7 (see left column plots in **Fig. 32**). Especially at 30min IGF1 and insulin stimulated ZR75-1 cells, the phospho-S6 levels almost return to un-inhibited levels, however, the pAkt levels still show an increase in insulin stimulated cells. Overall, these cell lines also show a differential up-regulation of Akt activation in 30min insulin stimulated cells, as shown true for MCF7 cells.

The results in this section are not in compliance with the corresponding predictions for the T47D and ZR75-1 cell lines (**Table 5**). This requires iteration of computational model refinement, and further to do another set of experiments to validate the refined models.

3.5 DISCUSSION

While no recurrent cancer-specific mutations of the IGF1R or its ligands have been described to date, a plethora of studies have provided evidence for a link between this signaling pathway and the risk of developing cancer. IGF1R signaling leads to both proliferative and anti-apoptotic signaling by employing RAS/MAPK and PI3K/Akt cascades. These two RTKs are 60% amino acid sequence homologs overall. Moreover, they are 84% identical at the kinase domain [6, 9]. Recent studies of the available partial structures of InsR and IGF1R, proposed mechanisms of activation of receptors upon ligand binding, where the latter attributed the role of extra-cellular

domain to keep the kinase domains apart in the absence of ligand [28, 29]. The strategies to inhibit the IGF1R signaling are: (1) using monoclonal antibodies against extracellular portion of the receptor to prevent ligand binding, (2) using small molecule inhibitors to act on tyrosine kinase portion of the receptor, (3) having siRNA and antisense approaches to reduce receptor expression levels, and (4) expressing dominant-negative truncated IGF1R proteins that interfere with receptor function. Moreover, targeting IGF1R results in: (1) impressive antineoplastic activity, (2) prevention of downstream signaling, (3) decrease in tumor cell proliferation, and (4) decrease in xenograft and tumor growth. Clinical trials of IGF1R targeting received some positive responses, however compensation mechanisms emerge and decrease the efficacy of such drugs [26, 27, 152]. Thus, an understanding of IGF1R signaling is important because it will guide the incorporation of appropriate molecular markers into clinical trial design and patient stratification.

The study of system biology encompasses the employment of tools and techniques to extract information about such networks from data. However, complexity of protein interaction networks hinders elucidation of structural and functional relations. Network topologies and interplay between the signal transduction cascades through activation-inhibition cycles determine biological responses in cells. It is not trivial to define which cell employs which cascade in which temporal order if they even do so. As a result, computational analyses and protocols are definitely needed considering the infeasibility of combinatorially large numbers of possible experiments.

Certainly there are other methods and approaches in the literature, but the use of system of ODEs is proven useful many times over the years in addition to its intuitive and powerful nature of explaining time-series phenomena. Other methodologies like probabilistic graphical

models [153, 154], Bayesian modeling structures [155, 156] or Boolean models [157, 158] could also be tested. The use of BioNetGen is also not a must for rule-based modeling but it is one of the few, and is locally developed and supported. The topology of the initial network also contains some assumptions but they are plausible and are employed frequently in most models in BioModels database [144].

The overall approach [149] and the mechanistic models here utilized the power of rule-based modeling with the time-series performance of ODE models. Networks of signal transduction are constructed to generate and test hypotheses of hormone stimulation in breast cancer cell lines. The study of these models, together with novel interactions deducted from the statistical modeling approach, helps us to understand the observed phenotypes under IGF1 or insulin stimulation.

Characterizing and representing the pathway enables us to generate and test hypotheses *in silico*, potentially aiding in selecting better therapies without extensive screening or wet-lab experimentation. To characterize the differences of the two aforementioned signaling pathways, I employed mechanistic modeling of different breast cancer cell lines. These efforts enable us to study combined interventions to bypass resistance development due to alterations induced by the primary drugs, and will help answer the question of why the two hormones have very distinct primary roles in normal but do cause growth-promoting actions in cancer cells. Additionally, studies like this work will also result in elucidation of novel interplays downstream of the two hormones.

The results in this chapter showed that the mechanisms of action of ACC and E-Cadherin are linked to IRS protein regulation. The parameter scans of MCF7 computational model revealed that pMAPK upregulation following ACC knock-down is similar to increasing the rate

of SOS activation through IRS binding. The E-Cadherin knock-down is also captured by IRS regulation, primarily by the negative-feedback of Akt on IRS activation. The combined analysis of statistical (**Chapter 2**) and mechanistic modeling (**Chapter 3**) resulted in insights on how pAkt and pMAPK are differentially regulated downstream of IGF1 and insulin.

Another result of this chapter was that S6 kinase inhibition leads to increased Akt phosphorylation, when stimulated with IGF1 and insulin. The MCF7 computational model suggested that the increase would be larger in insulin stimulated cells at 30 min. The experiments (**Figure 31**) also verified that prediction. In addition to this, the ODE models of other cell lines, specifically of T47D and ZR75-1, also resulted in differential pAkt regulation predictions. However, the predictions for T47D and ZR75-1 cell lines cannot be verified by experiments, rather they showed that the result of the S6K inhibition is similar to that of MCF7 cells.

Overall, seven ODE models of seven breast cancer cell lines are constructed and analyzed to find differential signaling cues in the downstream of insulin-InsR and IGF1-IGF1R pathways. The investigation of parameter set perturbations elucidated new experimental predictions, which are then verified by immunoblotting. The iterative quantitative systems pharmacologic approach utilized in this thesis demonstrates the power in finding novel interactions among canonical cascades of Akt and MAPK signaling.

4.0 CONCLUSIONS

4.1 OVERVIEW OF RESULTS

In cancer cells, insulin and IGF1 functions converge and both can activate proliferation and survival cascades. In this thesis, I employed a quantitative systems pharmacologic approach to understand IGF1R/InsR signaling, as an attempt to learn the details of tumor progression and drug resistance acquisition in breast cancer cells.

QSP combines experimental and clinical tools with computational analyses to better understand the biology, where iterations of experimentation-computation are employed. The first iteration of my thesis work started with RPPA data of insulin and IGF1 stimulated breast cancer cells. Then, I constructed a statistical network inference framework to elucidate underlying correlations between protein levels across time points. Analysis of the final networks of time translation yielded differential interactions between IGF1 and insulin conditions. Two of the top ranked (most differential) interactions are experimentally validated. Results showed that acetyl-CoA carboxylase knock-down increases MAPK phosphorylation while E-cadherin knock-down promotes higher Akt activation in IGF1 stimulated cells. These are novel findings that show how insulin and IGF1 downstream signaling cascades differ in cancer cells.

At the end of the first iteration, two novel interactions are resolved, with a focus on cascades of Akt and MAPK. To delineate the finer details of such interactions, I constructed

computational mechanistic models, within the second computational iteration. The model topology is based on the literature and model ODEs are generated by BioNetGen. A total of seven ODE models (of seven cell lines) are parametrized and parameter perturbation analyses revealed cues of differences in the downstream of insulin-InsR and IGF1-IGF1R pathways. The mechanisms of action of ACC and E-Cadherin functions are hypothesized to focus on the regulation of IRS1 protein. IRS1 is an adapter protein and one of the bottlenecks of signaling activation [110, 159]. Structural analysis of the two RTKs also suggests a differential binding of IRS1 [160].

Additionally, one of the mechanistic modeling predictions is experimentally validated for MCF7 cell line. Ribosomal protein S6 kinase inhibition is shown to increase Akt activation in insulin stimulated cells. The RPS6K inhibition in two other cell lines did not match the predictions, so that another round of computational model refinement is being carried out for mechanistic models.

4.2 FUTURE DIRECTIONS

The use of cancer cell line cultures is well suited to study the biology of signaling cascades. The tumors, on the other hand, are heterogeneous systems with many different cell and tissue types. The study and quantification of heterogeneity [161] and acquisition and analysis of clinical tumor samples is important, where recent advancements also sustain tens of measurements taken [162]. Analysis of paraffin-embedded samples by multiplexing has many advantages over classical methods and can further be employed in the context of computational modeling of

signaling cascades. Further efforts can be directed to model the distribution of single-cell responses, to match the many protein levels measured by multiplexing.

The experimentation portion of the QSP approach in the thesis included an RPPA dataset and western blots as validation of computational model predictions. Although immunoblotting experiments can show what happens to the levels of proteins when the cells are perturbed, they do not convey any information on single cell state, microenvironment, or three-dimensional (3D) organization. To circumvent the weaknesses of 2D experimental models, 3D cell cultures and organoid systems with extracellular matrix structure have been developed [163-166]. These cultures capture the morphology and physical properties of living tissues better, also offering better assessment of drug effectiveness. There are studies showing that the cell line responses differ between 2D and 3D experimental models [167-169]. This really indicates that after learning the details of the system behavior in detail with 2D models, higher order analyses, like drug development, should be done with the 3D systems. The 3D cultures require use of extracellular matrix proteins and other molecules to sustain the three-dimensional conformation. The 3D models can also be utilized to study the interactions of different cell types, matrix elements, and gradients of extracellular signaling molecules.

These 3D systems can mimic developmental processes, providing information of how and when certain structures form [170]. However, such technologies also require additional advancements in engineering matrix materials [171, 172], imaging protocols and equipment [173, 174], and analyses methods.

The use of imaging cells over time is a powerful tool. It encompasses the benefit of measuring many features over the course of experiment time instead of endpoint measurements like in western blots. Although limitations like resolution, antibody specificity, and limited

number of tag-able proteins do exist, imaging technologies has the advantage of dynamical temporal data. In combination with 3D model systems, they do offer even better advantages [39].

In addition to the 3D cell cultures, organ-on-a-chip models are also available [175]. Organ-on-chip models offer another layer of biological relevance with designed architectures. Similar to other 3D experimental models, chips also have synthetic materials as components, and this requires further study. The chips provide the ability to format different cell lines and tissue structures to study screening and drug development. Use of patient-derived cells in these systems offers personalized models. Complimentary to animal models, patient-derived xenograft (PDX) systems can be used in QSP approach to iteratively refine the computational models with heterogeneity and resistance emergence information, in a personalized manner [176].

Going from 2D to 3D and to organ-on-chip systems, these efforts are in-line with the idea of having cell-to-organism level computational models [72, 141]. The computational modeling can aid in simulation of normal functions as well as disease modes. Within the QSP framework, the computational models are supplied with laboratory and clinic level results and are iteratively refined. The role of the models is then to simulate drug regimens *in silico* and analyze the response with any possible off-target effects. From the dosage and timing of the drugs to the overall tendency of tumor mass can be pinpointed with well-trained computational models. So, with the simultaneous advent of better 3D experimental models and advanced computational models, quantitative systems pharmacology can help to accelerate translational research in cancer biology. In doing so, we can start stratifying patients to suitable personalized medicine treatments after recognizing and distinguishing that the IGF1R and InsR systems have different dynamics and novel signaling components.

APPENDIX A MECHANISTIC MODEL BIONETGEN SCRIPT

```
# CMM for MCF7 cell line
# 03/01/18 by Cemal ERDEM
# MODEL details:
# 1) IGF1-IGF1R and insulin-InsR binding only
# 2) ONE phospho sites at each receptor
# 3) NO basal phosphorylation
# 4) Rate parameters are in log10
# 5) FOUR observables

begin model
begin parameters
 IGF1_0 100000
 INS_0 0
 IGF1R_0 25000
 INSR_0 25000
 IRS_0 92766
 SOS_0 90075
 RAS_0 230642
 RAF_0 126069
 MEK_0 1098164
 ERK_0 763172
 PI3K_0 64009
 PDK1_0 186081
 AKT_0 432907
 TSC2_0 131339
 MTOR_0 83469
 RPS6K_0 121978

# Ligand-receptor interactions
 kf1     0.4837
 kf1b    -2.9153
 kf1c    2.9865
 kf1d    1.2052
 kf2     4.6312
 kf2b    -0.8667
 kf2c    4.8758
```

```
 kf2d   -2.6526

# Downstream interactions
 kf3    -2.7913
 kf4    -3.1902
 kf5    -0.6920
 kf6    4.1250
 kf7    -3.0400
 kf8    -4.5760
 kf9    0.3532
 kf10   4.3309
 kf11   -6.9315
 kf12   -3.8016
 kf13   -6.2483
 kf14   1
 kf15   -6.7787
 kf16   -7.7887
 kf17   -7.7124

# -P rates
 kf101  -0.0831
 kf102  -6.4728
 kf103  0.0477
 kf104  3.9039
 kf105  0.7593
 kf106  0.5137
 kf107  -0.8678
 kf108  -3.3447
 kf109  -4.5098
 kf110  6.3930
 kf111  -2.7962
 kf112  -2.8996

# Feedback events
 kf201  2.7628
 kf202  4.0772
 kf203  -5.4445
 kf204  0.2274
 kf206  -6.3512
 kf207  -6.3752
 kf208  -5.0059

# Re-sensitization events
 kf301  -3.0943
 kf302  -0.9807
 kf303  4.2786
```

```
 kf304  -5.6559

## Recepter recycling
 kf401  -3.1016
 kf402  -2.9249
 kf403  -3.4193
 kf404  -7.0807
end parameters

begin molecule types
 IGF1(rec)
 Ins(rec)
 IGF1R(lig,adp,adp,phos~U~P,int~N~Y)
 InsR(lig,adp,adp,phos~U~P,int~N~Y)
 IRS(rec,sos,p3k,s6k,phos~U~P,inh~N~Y)
 SOS(rec,irs,ras,act~N~Y,inh~N~Y)
 Ras(sos,raf,gtp~N~Y)
 Raf(ras,mek,phos~U~P,inh~N~Y)
 MEK(raf,erk,phos~U~P,inh~N~Y)
 PI3K(irs,pdk,act~N~Y)
 PDK1(p3k,akt,act~N~Y)
 TSC2(akt,phos~U~P)
 mTOR(s6k,ebp1,act~N~Y)
 Akt(pdk1,tsc2,oth,phos~U~P)
 RPS6K(mTOR,s6,irs,phos~U~P)
 ERK(mek,erk,oth,phos~U~P)
end molecule types

begin seed species
 IGF1(rec) IGF1_0
 Ins(rec) INS_0
 IGF1R(lig,adp,adp,phos~U,int~N) IGF1R_0
 InsR(lig,adp,adp,phos~U,int~N) INSR_0
 IRS(rec,sos,p3k,s6k,phos~U,inh~N) IRS_0
 SOS(rec,irs,ras,act~N,inh~N) SOS_0
 Ras(sos,raf,gtp~N) RAS_0
 Raf(ras,mek,phos~U,inh~N) RAF_0
 MEK(raf,erk,phos~U,inh~N) MEK_0
 PI3K(irs,pdk,act~N) PI3K_0
 PDK1(p3k,akt,act~N) PDK1_0
 TSC2(akt,phos~U) TSC2_0
 mTOR(s6k,ebp1,act~N) MTOR_0
 Akt(pdk1,tsc2,oth,phos~U) AKT_0
 RPS6K(mTOR,s6,irs,phos~U) RPS6K_0
 ERK(mek,erk,oth,phos~U) ERK_0
end seed species
```

```
begin observables
 Species       pRecTot              IGF1R(int~N,phos~P),InsR(int~N,phos~P)
 Species       pAkt308              Akt(pdk1,tsc2,oth,phos~P)
 Species       pRPS6K               RPS6K(mTOR,s6,irs,phos~P)
 Species       pERK                 ERK(mek,erk,oth,phos~P)
end observables

begin reaction rules
# Initial ligand-receptor binding
 IGF1(rec) + IGF1R(lig,adp,adp,int~N,phos~U)  <->
IGF1(rec!1).IGF1R(lig!1,adp,adp,int~N,phos~U)     10^kf1,10^kf1b
 IGF1R(lig!+,adp,adp,int~N,phos~U) -> IGF1R(lig!+,adp,adp,int~N,phos~P)       10^kf1c
 IGF1R(lig!+,adp,adp,int~N,phos~P) -> IGF1R(lig,adp,adp,int~N,phos~U)         10^kf1d
 Ins(rec) + InsR(lig,adp,adp,int~N,phos~U)  <->  Ins(rec!1).InsR(lig!1,adp,adp,int~N,phos~U)
        10^kf2,10^kf2b
 InsR(lig!+,adp,adp,int~N,phos~U) -> InsR(lig!+,adp,adp,int~N,phos~P)          10^kf2c
 InsR(lig!+,adp,adp,int~N,phos~P) -> InsR(lig,adp,adp,int~N,phos~U)            10^kf2d

# pReceptor-IRS binding and activation
 IGF1R(adp,adp,int~N,phos~P) + IRS(rec,sos,p3k,s6k,inh~N,phos~U) ->
IGF1R(adp,adp,int~N,phos~P) + IRS(rec,sos,p3k,s6k,inh~N,phos~P)       10^kf3
 InsR(adp,adp,int~N,phos~P) + IRS(rec,sos,p3k,s6k,inh~N,phos~U) ->
InsR(adp,adp,int~N,phos~P) + IRS(rec,sos,p3k,s6k,inh~N,phos~P)         10^kf4

# pReceptor-SOS binding and activation
 IGF1R(adp,adp,int~N,phos~P) + SOS(rec,irs,ras,inh~N,act~N) ->
IGF1R(adp,adp,int~N,phos~P) + SOS(rec,irs,ras,inh~N,act~Y)              10^kf5
 InsR(adp,adp,int~N,phos~P) + SOS(rec,irs,ras,inh~N,act~N) -> InsR(adp,adp,int~N,phos~P) +
SOS(rec,irs,ras,inh~N,act~Y)                  10^kf6

# SOS activation by IRS1
 IRS(sos,s6k,p3k,inh~N,phos~P) + SOS(rec,irs,ras,inh~N,act~N) ->
IRS(sos,s6k,p3k,inh~N,phos~P) + SOS(rec,irs,ras,inh~N,act~Y)            10^kf7

# Ras activation by SOS
 SOS(rec,irs,ras,inh~N,act~Y) + Ras(sos,raf,gtp~N) -> SOS(rec,irs,ras,inh~N,act~Y) +
Ras(sos,raf,gtp~Y)     10^kf8

# Raf activation by Ras
 Ras(sos,raf,gtp~Y) + Raf(ras,mek,inh~N,phos~U) -> Ras(sos,raf,gtp~Y) +
Raf(ras,mek,inh~N,phos~P)            10^kf9

# MEK activation by Raf
 Raf(ras,mek,inh~N,phos~P) + MEK(raf,erk,inh~N,phos~U) -> Raf(ras,mek,inh~N,phos~P) +
MEK(raf,erk,inh~N,phos~P)            10^kf10
```

```
# ERK activation by MEK
 MEK(raf,erk,inh~N,phos~P) + ERK(mek,erk,oth,phos~U) -> MEK(raf,erk,inh~N,phos~P) +
ERK(mek,erk,oth,phos~P)              10^kf11

# PI3K activation by IRS1
 IRS(sos,s6k,p3k,inh~N,phos~P) + PI3K(irs,pdk,act~N) -> IRS(sos,s6k,p3k,inh~N,phos~P) +
PI3K(irs,pdk,act~Y)              10^kf12

# PDK1 activation PI3K
 PI3K(irs,pdk,act~Y) + PDK1(p3k,akt,act~N) -> PI3K(irs,pdk,act~Y) + PDK1(p3k,akt,act~Y)
          10^kf13
# Akt activation by PDK1
 PDK1(p3k,akt,act~Y) + Akt(pdk1,tsc2,oth,phos~U) -> PDK1(p3k,akt,act~Y) +
Akt(pdk1,tsc2,oth,phos~P)              10^kf14

# TSC2 inactivation by Akt
 Akt(pdk1,tsc2,oth,phos~P) + TSC2(akt,phos~U) -> Akt(pdk1,tsc2,oth,phos~P) +
TSC2(akt,phos~P)              10^kf15

# mTOR activation by inactive TSC2
 TSC2(akt,phos~P) + mTOR(s6k,ebp1,act~N) -> TSC2(akt,phos~P) + mTOR(s6k,ebp1,act~Y)
          10^kf16
# RPS6K activation by mTOR
 mTOR(s6k,ebp1,act~Y) + RPS6K(mTOR,s6,irs,phos~U) -> mTOR(s6k,ebp1,act~Y) +
RPS6K(mTOR,s6,irs,phos~P)                    10^kf17

# De-phosphorylation (-P) events
 IRS(phos~P) -> IRS(phos~U)                  10^kf101
 SOS(act~Y) -> SOS(act~N)                    10^kf102
 Ras(gtp~Y) -> Ras(gtp~N)                    10^kf103
 Raf(phos~P) -> Raf(phos~U)                  10^kf104
 MEK(phos~P) -> MEK(phos~U)                  10^kf105
 PI3K(act~Y) ->          PI3K(act~N)         10^kf106
 PDK1(act~Y) -> PDK1(act~N)                  10^kf107
 TSC2(phos~P) -> TSC2(phos~U)                10^kf108
 mTOR(act~Y) -> mTOR(act~N)                  10^kf109
 Akt(phos~P) -> Akt(phos~U)                  10^kf110
 RPS6K(phos~P) -> RPS6K(phos~U)              10^kf111
 ERK(phos~P) -> ERK(phos~U)                  10^kf112

#### Negative feedbacks
# SOS inactivation by pERK
 ERK(mek,erk,oth,phos~P) + SOS(rec,irs,ras,inh~N,act~N) -> ERK(mek,erk,oth,phos~P) +
SOS(rec,irs,ras,inh~Y,act~N)              10^kf201

# MEK inactivation by pERK
```

```
 ERK(mek,erk,oth,phos~P) + MEK(raf,erk,inh~N,phos~U) -> ERK(mek,erk,oth,phos~P) +
MEK(raf,erk,inh~Y,phos~U)            10^kf202

# IRS1 inhibition by pRPS6K
 RPS6K(mTOR,s6,irs,phos~P) + IRS(sos,p3k,s6k,inh~N,phos~U) ->
RPS6K(mTOR,s6,irs,phos~P) + IRS(sos,p3k,s6k,inh~Y,phos~U)   10^kf203

# Raf inactivation by pAkt
 Akt(pdk1,tsc2,oth,phos~P) + Raf(ras,mek,inh~N,phos~U) -> Akt(pdk1,tsc2,oth,phos~P) +
Raf(ras,mek,inh~Y,phos~U)            10^kf204

# IRS inactivation by pERK
 ERK(mek,erk,oth,phos~P) + IRS(sos,s6k,p3k,inh~N,phos~U) -> ERK(mek,erk,oth,phos~P) +
IRS(sos,s6k,p3k,inh~Y,phos~U)       10^kf206

# Akt inactivation by pERK
 ERK(mek,erk,oth,phos~P) + Akt(pdk1,tsc2,oth,phos~P) -> ERK(mek,erk,oth,phos~P) +
Akt(pdk1,tsc2,oth,phos~U)    10^kf207

# IRS inactivation by pAkt
 Akt(pdk1,tsc2,oth,phos~P) + IRS(sos,s6k,p3k,inh~N,phos~U) -> Akt(pdk1,tsc2,oth,phos~P) +
IRS(sos,s6k,p3k,inh~Y,phos~U)                10^kf208

#### Re-sensitization
 IRS(inh~Y) -> IRS(inh~N)            10^kf301
 SOS(inh~Y) -> SOS(inh~N)            10^kf302
 Raf(inh~Y) -> Raf(inh~N)            10^kf303
 MEK(inh~Y) -> MEK(inh~N)        10^kf304

## Recepter recycling
 IGF1R(lig!+,adp,adp,int~N,phos~P) -> IGF1R(lig!+,adp,adp,int~Y,phos~P)        10^kf401
 IGF1(rec!1).IGF1R(lig!1,adp,adp,int~Y,phos~P) -> IGF1R(lig,adp,adp,int~N,phos~U)10^kf402
 InsR(lig!+,adp,adp,int~N,phos~P) -> InsR(lig!+,adp,adp,int~Y,phos~P)            10^kf403
 Ins(rec!1).InsR(lig!1,adp,adp,int~Y,phos~P) -> InsR(lig,adp,adp,int~N,phos~U)  10^kf404
end reaction rules
end model

begin actions
 generate_network({overwrite=>1})
 simulate({method=>"ode",t_start=>0,t_end=>1800,n_steps=>2e4})
 writeMfile_all()
end actions
```

APPENDIX B PARAMETER ESTIAMTION SCRIPT

```
clear all;
close all;
load('CMMnet_MCF7_init.mat')
clc;
rng(6); % set random number generator seed

%%%%%%%%%% CHANGE THESE AS DESIRED
sim_num = 5e4; % number of simulations/iterations
dxx = 0.05; % step size for perturbation
parIDs1 = []; % IDs of parameters to change in every iteration - of Initial protein counts only
parIDs3 = [18:20]; % IDs of parameters to change in every iteration - of rate parameters only
parIDsTC = []; % IDs of parameters to set constant
parValsTC = [50000,50000]; % the IDs of parameters to set constant
selectIDss = [5:16,21,22,25,27,29,26,28,32,35,30,38,31]; % IDs of parameters to plot

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

ScaleFacsIGF = {}; % array of scaling factors for IGF1 simulations
ScaleFacsINS = {}; % array of scaling factors for IGF1 simulations
lenIDs1 = length(parIDs1);
lenIDs3 = length(parIDs3);
lenIDsTC = length(parIDsTC);
acct = 0; % acceptance rate parameter, will store number of accepted moves
Accpt_conds = zeros(1,sim_num); % array of the types of accceptances
counter1 = 2; % counter for simulations
counter2 = 1; % counter for unacceptable perturbations
Errs = []; % will store objective values (fitting errors) of each iteration
Errs(1) = 0;

Params = param_defaults; % array of parameter sets
parsNew = (Params(1,:)); % temporary parameter set for the iterations

Y0 = initSpecies_CMM_mcf7( parsNew ); % Initialize species array
Y0IGF = Y0;
Y0IGF(1,1) = expdataigf.IGF0; % IGF1 stimulation data for IGF1 number
Y0IGF(1,2) = expdataigf.INS0; % IGF1 stimulation data for insulin number
```

```
[~, ~, Y1igf, obsY1igf] = CMM_mcf7(expdataigf.time, Y0IGF, parsNew, 1); % model trajectory
for ss = 1:lenObss
   ScaleFacsIGF{1,1}(ss) = obsY1igf(:,ss)\expdataigf.mean(:,ss); % calculate scaling factors
end
tempFitigf = (repmat(ScaleFacsIGF{1,1},size(obsY1igf,1),1)).*obsY1igf; % scale observable protein numbers
tempFitErrigf = ((expdataigf.mean-tempFitigf).^2)./(2.*(expdataigf.stdev.^2)); % calculate fitting error
tempFitErrigf = expdataigf.weights.*tempFitErrigf; % weights on fitting erros
Errs(1) = Errs(1) + sum(sum(abs(tempFitErrigf))); % store IGF1 fitting error

Y0Ins = Y0;
Y0Ins(1,1) = expdatains.IGF0; % Insulin stimulation data for IGF1 number
Y0Ins(1,2) = expdatains.INS0; % Insulin stimulation data for insulin number
[err, ~, Y1ins, obsY1ins] = CMM_mcf7(expdataigf.time, Y0Ins, parsNew, 1);
for ss = 1:numsfits
   ScaleFacsINS{1,1}(ss)  = obsY1ins(:,ss)\expdatains.mean(:,ss);
end
tempFitins = (repmat(ScaleFacsINS{1,1},size(obsY1ins,1),1)).*obsY1ins;
tempFitErrins = ((expdatains.mean-tempFitins).^2)./(2.*(expdatains.stdev.^2));
tempFitErrins = expdatains.weights.*tempFitErrins;
Errs(1) = Errs(1) + sum(sum(abs(tempFitErrins))); % store insulin fitting error

randstate = rng; % store the system random number generator state (caution for unexpected termination of the run)
while counter1 <= sim_num % simulation loop (of (5e5)-1 steps)
%     rng(randstate) % set the random number generator to stored one
   parsNew = Params(end,:); % temporary parameter set for the current iteration

   ScFacts = []; % temporary scaling factor array - IGF1
   ScFacts2 = []; % temporary scaling factor array - insulin

   %%%% Change total molecule counts
   for qq = 1:lenIDsTC
      parsNew(parIDsTC(qq)) = parValsTC(qq);
   end

   %%%% Initial condition parameter perturbation
   for qq = 1:lenIDs1
      lim1 = param_bounds(parIDs1(qq),1)-1;
      lim2 = param_bounds(parIDs1(qq),2)-lim1;
      randith = randi(lim2)+lim1;
      parsNew(parIDs1(qq)) = randith;
   end

   %%%% Rate constant parameter perturbation
```

```
for mm = 1:lenIDs3
    lim1 = param_bounds(parIDs3(mm),1);
    lim2 = param_bounds(parIDs3(mm),2);
    randith = (parsNew(parIDs3(mm)))+(randn(1).*dxx); % trial move on parameters
    if (lim1 > randith) || (lim2 < randith)
        randith = random('Uniform',lim1,lim2);
    end
    parsNew(parIDs3(mm)) = randith;
end

%%%% Run the new trajectories
Y0 = initSpecies_CMM_mcf7( parsNew );
Y0IGF = Y0;
Y0IGF(1,1) = expdataigf.IGF0;
Y0IGF(1,2) = expdataigf.INS0;
[~, ~, Y1igf, obsY1igf] = CMM_mcf7(expdataigf.time, Y0IGF, parsNew, 1);
for ss = 1:lenObss
    ScFacts{1,1}(ss) = obsY1igf(:,ss)\expdataigf.mean(:,ss);
end
Y1select = Y1igf(2,selectIDss);
Y0Ins = Y0;
Y0Ins(1,1) = expdatains.IGF0;
Y0Ins(1,2) = expdatains.INS0;
[~, ~, Y1ins, obsY1ins] = CMM_mcf7(expdataigf.time, Y0Ins, parsNew, 1);
for ss = 1:numsfits
    ScFacts2{1,1}(ss)  = obsY1ins(:,ss)\expdatains.mean(:,ss);
end
Y1selectIns = Y1ins(2,selectIDss);

%%%% Check if the trajectories are run without erros and with large enough number of
proteins
if (all(ScFacts{1,1}<1e-3)) && (all(ScFacts{1,1}>0)) && (all(obsY1igf(end,:)>=1e3)) && ...
   (all(ScFacts2{1,1}<1e-3)) && (all(ScFacts2{1,1}>0)) && (all(obsY1ins(end,:)>=1e3))

    if (size(obsY1igf,1)==uptotime) && (size(obsY1ins,1)==uptotime) % Check if trajectories
are run upto 30min
        tempFitigf = (repmat(ScFacts{1,1},size(obsY1igf,1),1)).*obsY1igf;
        tempFitErrigf = ((expdataigf.mean-tempFitigf).^2)./(2.*(expdataigf.stdev.^2));
        tempFitErrigf = expdataigf.weights.*tempFitErrigf;
        next_errigf = sum(sum(abs(tempFitErrigf)));
        tempFitins = (repmat(ScFacts2{1,1},size(obsY1ins,1),1)).*obsY1ins;
        tempFitErrins = ((expdatains.mean-tempFitins).^2)./(2.*(expdatains.stdev.^2));
        tempFitErrins = expdatains.weights.*tempFitErrins;
        next_errins = sum(sum(abs(tempFitErrins)));
        next_err = next_errigf + next_errins;
```

```
        if ~isnan(next_err)
            if (next_err < Errs(end)) % accept new parameter set with a lower fitting error than the
current one
                Params(counter1,:) = parsNew;
                Errs(end+1) = next_err;
                acct = acct + 1;
                Accpt_conds(counter1) = 1;
                ScaleFacsIGF{end+1,1} = ScFacts;
                ScaleFacsINS{end+1,1} = ScFacts2;
            else
                rr = rand(1);
                D_err = next_err - Errs(end);
                hh = min(1,exp(-D_err));
                if rr < hh % Metropolis acceptance
                    Params(counter1,:) = parsNew;
                    Errs(end+1) = next_err;
                    acct = acct + 1;
                    Accpt_conds(counter1) = 2;
                    ScaleFacsIGF{end+1,1} = ScFacts;
                    ScaleFacsINS{end+1,1} = ScFacts2;
                else % keep last parameter set as the new one also
                    Params(counter1,:) = Params(end,:);
                    Errs(end+1) = Errs(end);
                    Accpt_conds(counter1) = 3;
                    ScaleFacsIGF{end+1,1} = ScaleFacsIGF{end,1};
                    ScaleFacsINS{end+1,1} = ScaleFacsINS{end,1};
                end
            end
        else % if the error cannot be calculated, continue with the current parameter set
            Params(counter1,:) = Params(end,:);
            Errs(end+1) = Errs(end);
            Accpt_conds(counter1) = 4;
            ScaleFacsIGF{end+1,1} = ScaleFacsIGF{end,1};
            ScaleFacsINS{end+1,1} = ScaleFacsINS{end,1};
        end
        randstate = rng;
        counter1 = counter1 + 1;
    end
  elseif counter2<=100 % if first check is violated try a new perturbation
    % (note that this does not increase iteration counter)
    counter2 = counter2 + 1;
  end
end
```

BIBLIOGRAPHY

1. Heiser, L.M., et al., Subtype and pathway specific responses to anticancer compounds in breast cancer. Proc Natl Acad Sci U S A, 2012. **109**(8): p. 2724-9.

2. Curtis, C., et al., The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature, 2012. **486**(7403): p. 346-52.

3. Perou, C.M., et al., Molecular portraits of human breast tumours. Nature, 2000. **406**(6797): p. 747-52.

4. Mavaddat, N., et al., Cancer risks for BRCA1 and BRCA2 mutation carriers: results from prospective analysis of EMBRACE. J Natl Cancer Inst, 2013. **105**(11): p. 812-22.

5. Miki, Y., et al., A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science, 1994. **266**(5182): p. 66-71.

6. Boone, D.N. and A.V. Lee, Targeting the insulin-like growth factor receptor: developing biomarkers from gene expression profiling. Crit Rev Oncog, 2012. **17**(2): p. 161-73.

7. Jordan, V.C., Tamoxifen: A most unlikely pioneering medicine. Nature Reviews Drug Discovery, 2003. **2**(3): p. 205-213.

8. Niederst, M.J. and J.A. Engelman, Bypass mechanisms of resistance to receptor tyrosine kinase inhibition in lung cancer. Sci Signal, 2013. **6**(294): p. re6.

9. Casa, A.J., et al., The type I insulin-like growth factor receptor pathway: a key player in cancer therapeutic resistance. Front Biosci, 2008. **13**: p. 3273-87.

10. Liu, J.L. and D. LeRoith, Insulin-like growth factor I is essential for postnatal growth in response to growth hormone. Endocrinology, 1999. **140**(11): p. 5178-5184.

11. Braun, S., K. Bitton-Worms, and D. LeRoith, The Link between the Metabolic Syndrome and Cancer. International Journal of Biological Sciences, 2011. **7**(7): p. 1003-1015.

12. Kooijman, R., Regulation of apoptosis by insulin-like growth factor (IGF)-I. Cytokine & Growth Factor Reviews, 2006. **17**(4): p. 305-323.

13. Casa, A., et al., Insulin-Like Growth Factor Signaling in Normal Mammary Gland Development and Breast Cancer Progression, in Breast Cancer: Prognosis, Treatment and Prevention. 2008. p. 303-322.

14. Gallagher, E.J. and D. LeRoith, The proliferating role of insulin and insulin-like growth factors in cancer. Trends in Endocrinology and Metabolism, 2010. **21**(10): p. 610-618.

15. Baxter, R.C., Insulin-like growth factor (IGF)-binding proteins: interactions with IGFs and intrinsic bioactivities. American Journal of Physiology-Endocrinology and Metabolism, 2000. **278**(6): p. E967-E976.

16. Clemmons, D.R., et al., Modifications of insulin-like growth factor binding proteins and their role in controlling IGF actions. Endocrine Journal, 1998. **45**: p. S1-S8.

17. Pollak, M., Insulin and insulin-like growth factor signalling in neoplasia. Nature Reviews Cancer, 2008. **8**(12): p. 915-928.

18. Sachdev, D. and D. Yee, The IGF system and breast cancer. Endocr Relat Cancer, 2001. **8**(3): p. 197-209.

19. Pollak, M.N., E.S. Schernhammer, and S.E. Hankinson, Insulin-like growth factors and neoplasia. Nature Reviews Cancer, 2004. **4**(7): p. 505-518.

20. Kruger, M., et al., Dissection of the insulin signaling pathway via quantitative phosphoproteomics. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(7): p. 2451-2456.

21. Crudden, C., A. Girnita, and L. Girnita, Targeting the IGF-1R: The Tale of the Tortoise and the Hare. Front Endocrinol (Lausanne), 2015. **6**: p. 64.

22. Farabaugh, S.M., D.N. Boone, and A.V. Lee, Role of IGF1R in Breast Cancer Subtypes, Stemness, and Lineage Differentiation. Front Endocrinol (Lausanne), 2015. **6**: p. 59.

23. Maki, R.G., Small is beautiful: insulin-like growth factors and their role in growth, development, and cancer. J Clin Oncol, 2010. **28**(33): p. 4985-95.

24. Siddle, K., Signalling by insulin and IGF receptors: supporting acts and new players. J Mol Endocrinol, 2011. **47**(1): p. R1-10.

25. Clayton, P.E., et al., Growth hormone, the insulin-like growth factor axis, insulin and cancer risk. Nat Rev Endocrinol, 2011. **7**(1): p. 11-24.

26. Arcaro, A., Targeting the insulin-like growth factor-1 receptor in human cancer. Frontiers in Pharmacology, 2013. **4**.

27. Olmos, D., B. Basu, and J.S. de Bono, Targeting Insulin-Like Growth Factor Signaling: Rational Combination Strategies. Molecular Cancer Therapeutics, 2010. **9**(9): p. 2447-2449.

28. Menting, J.G., et al., How insulin engages its primary binding site on the insulin receptor. Nature, 2013. **493**(7431): p. 241-U276.

29. Kavran, J.M., et al., How IGF-1 Activates its Receptor. Elife, 2014. **3**.

30. Houde, D. and S.J. Demarest, Fine Details of IGF-1R Activation, Inhibition, and Asymmetry Determined by Associated Hydrogen/Deuterium-Exchange and Peptide Mass Mapping. Structure, 2011. **19**(6): p. 890-900.

31. Kiselyov, V.V., et al., Harmonic oscillator model of the insulin and IGF1 receptors' allosteric binding and activation. Molecular Systems Biology, 2009. **5**.

32. P. K. Sorger and S.R.B. Allerheiligen. Quantitative and Systems Pharmacology in the Postgenomic Era: New Approaches to Discovering Drugs and Understanding Therapeutic Mechanisms. [An NIH White Paper by the QSP Workshop Group] 2011 [cited 2018 March1]; http://www.nigms.nih.gov/News/reports/Documents/SystemsPharmaWPSorger2011.pdf.

33. Stern, A.M., et al., A Perspective on Implementing a Quantitative Systems Pharmacology Platform for Drug Discovery and the Advancement of Personalized Medicine. J Biomol Screen, 2016. **21**(6): p. 521-34.

34. Zhao, S. and R. Iyengar, Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. Annu Rev Pharmacol Toxicol, 2012. **52**: p. 505-21.

35. Hopkins, A.L., Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol, 2008. **4**(11): p. 682-90.

36. Musante, C.J., et al., Quantitative Systems Pharmacology: A Case for Disease Models. Clin Pharmacol Ther, 2017. **101**(1): p. 24-27.

37. Liu, J., et al., Prediction of Efficacy of Vabicaserin, a 5-HT2C Agonist, for the Treatment of Schizophrenia Using a Quantitative Systems Pharmacology Model. CPT Pharmacometrics Syst Pharmacol, 2014. **3**: p. e111.

38. Leil, T.A. and R. Bertz, Quantitative Systems Pharmacology can reduce attrition and improve productivity in pharmaceutical research and development. Front Pharmacol, 2014. **5**: p. 247.

39. Senutovitch, N., et al., Fluorescent protein biosensors applied to microphysiological systems. Exp Biol Med (Maywood), 2015. **240**(6): p. 795-808.

40. Tibes, R., et al., Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells. Mol Cancer Ther, 2006. **5**(10): p. 2512-21.

41. Daemen, A., et al., Modeling precision treatment of breast cancer. Genome Biol, 2013. **14**(10): p. R110.

42. Toettcher, J.E., O.D. Weiner, and W.A. Lim, Using optogenetics to interrogate the dynamic control of signal transmission by the Ras/Erk module. Cell, 2013. **155**(6): p. 1422-34.

43. Iadevaia, S., et al., Identification of optimal drug combinations targeting cellular networks: integrating phospho-proteomics and computational network analysis. Cancer Res, 2010. **70**(17): p. 6704-14.

44. Sheehan, K.M., et al., Use of reverse phase protein microarrays and reference standard development for molecular network analysis of metastatic ovarian carcinoma. Mol Cell Proteomics, 2005. **4**(4): p. 346-55.

45. Bar-Joseph, Z., A. Gitter, and I. Simon, STUDY DESIGNS Studying and modelling dynamic biological processes using time-series gene expression data. Nature Reviews Genetics, 2012. **13**(8): p. 552-564.

46. Jin, L., et al., Pathway-based analysis tools for complex diseases: a review. Genomics Proteomics Bioinformatics, 2014. **12**(5): p. 210-20.

47. Bahar, I., A.R. Atilgan, and B. Erman, Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Folding & Design, 1997. **2**(3): p. 173-181.

48. Chennubhotla, C., et al., Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. Phys Biol, 2005. **2**(4): p. S173-80.

49. Wang, G., Singularity analysis of the AKT signaling pathway reveals connections between cancer and metabolic diseases. Phys Biol, 2010. **7**(4): p. 046015.

50. Poirel, C.L., et al., Top-Down Network Analysis to Drive Bottom-Up Modeling of Physiological Processes. Journal of Computational Biology, 2013. **20**(5): p. 409-418.

51. Niepel, M., et al., Profiles of Basal and Stimulated Receptor Signaling Networks Predict Drug Response in Breast Cancer Lines. Science Signaling, 2013. **6**(294).

52. Huang, S.S. and E. Fraenkel, Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. Sci Signal, 2009. **2**(81): p. ra40.

53. Vinayagam, A., et al., A directed protein interaction network for investigating intracellular signal transduction. Sci Signal, 2011. **4**(189): p. rs8.

54. White, F.M., Quantitative phosphoproteomic analysis of signaling network dynamics. Curr Opin Biotechnol, 2008. **19**(4): p. 404-9.

55. Alon, U., Network motifs: theory and experimental approaches. Nat Rev Genet, 2007. **8**(6): p. 450-61.

56. Holter, N.S., et al., Dynamic modeling of gene expression data. Proc Natl Acad Sci U S A, 2001. **98**(4): p. 1693-8.

57. Lezon, T.R., et al., Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. Proc Natl Acad Sci U S A, 2006. **103**(50): p. 19033-8.

58. Friedman, N., Inferring cellular networks using probabilistic graphical models. Science, 2004. **303**(5659): p. 799-805.

59. Ernst, J., et al., Reconstructing dynamic regulatory maps. Mol Syst Biol, 2007. **3**: p. 74.

60. Friedman, J., T. Hastie, and R. Tibshirani, Sparse inverse covariance estimation with the graphical lasso. Biostatistics, 2008. **9**(3): p. 432-41.

61. Kitano, H., Systems biology: a brief overview. Science, 2002. **295**(5560): p. 1662-4.

62. Kitano, H., Computational systems biology. Nature, 2002. **420**(6912): p. 206-10.

63. Ideker, T., T. Galitski, and L. Hood, A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet, 2001. **2**: p. 343-72.

64. Machado, D., et al., Modeling formalisms in Systems Biology. AMB Express, 2011. **1**: p. 45.

65. Bartocci, E. and P. Lio, Computational Modeling, Formal Analysis, and Tools for Systems Biology. PLoS Comput Biol, 2016. **12**(1): p. e1004591.

66. Singh, A., A. Jayaraman, and J. Hahn, Modeling regulatory mechanisms in IL-6 signal transduction in hepatocytes. Biotechnol Bioeng, 2006. **95**(5): p. 850-62.

67. Bachmann, J., et al., Division of labor by dual feedback regulators controls JAK2/STAT5 signaling over broad ligand range. Mol Syst Biol, 2011. **7**: p. 516.

68. Gambin, A., et al., Computational models of the JAK1/2-STAT1 signaling. JAKSTAT, 2013. **2**(3): p. e24672.

69. Wiley, H.S., S.Y. Shvartsman, and D.A. Lauffenburger, Computational modeling of the EGF-receptor system: a paradigm for systems biology. Trends Cell Biol, 2003. **13**(1): p. 43-50.

70. Erdem, C., et al., Mathematical Modeling of Behcet's Disease: A Dynamical Systems Approach. Journal of Biological Systems, 2015. **23**(2): p. 231-257.

71. Nyman, E., et al., A Hierarchical Whole-body Modeling Approach Elucidates the Link between in Vitro Insulin Signaling and in Vivo Glucose Homeostasis. Journal of Biological Chemistry, 2011. **286**(29): p. 26028-26041.

72. Man, C.D., R.A. Rizza, and C. Cobelli, Meal simulation model of the glucose-insulin system. Ieee Transactions on Biomedical Engineering, 2007. **54**(10): p. 1740-1749.

73. Lukens, S., et al., A large-scale immuno-epidemiological simulation of influenza A epidemics. BMC Public Health, 2014. **14**: p. 1019.

74. Reynolds, A., et al., A reduced mathematical model of the acute inflammatory response: I. Derivation of model and analysis of anti-inflammation. J Theor Biol, 2006. **242**(1): p. 220-36.

75. Kholodenko, B.N., et al., Quantification of short term signaling by the epidermal growth factor receptor. J Biol Chem, 1999. **274**(42): p. 30169-81.

76. Schoeberl, B., et al., Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. Nat Biotechnol, 2002. **20**(4): p. 370-5.

77. Blinov, M.L., et al., A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. Biosystems, 2006. **83**(2-3): p. 136-51.

78. Villasenor, R., et al., Regulation of EGFR signal transduction by analogue-to-digital conversion in endosomes. Elife, 2015. **4**.

79. Harris, L.A., et al., BioNetGen 2.2: advances in rule-based modeling. Bioinformatics, 2016. **32**(21): p. 3366-3368.

80. Danos, V. and C. Laneve, Formal molecular biology. Theoretical Computer Science, 2004. **325**(1): p. 69-110.

81. Chylek, L.A., et al., Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems. Wiley Interdiscip Rev Syst Biol Med, 2014. **6**(1): p. 13-36.

82. Lopez, C.F., et al., Programming biological models in Python using PySB. Molecular Systems Biology, 2013. **9**.

83. Xu, W., et al., RuleBender: a visual interface for rule-based modeling. Bioinformatics, 2011. **27**(12): p. 1721-1722.

84. Gillespie, D.T., Stochastic simulation of chemical kinetics. Annu Rev Phys Chem, 2007. **58**: p. 35-55.

85. Hucka, M., et al., The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics, 2003. **19**(4): p. 524-31.

86. Erdem, C., et al., Proteomic Screening and Lasso Regression Reveal Differential Signaling in Insulin and Insulin-like Growth Factor I (IGF1) Pathways. Molecular & Cellular Proteomics, 2016. **15**(9): p. 3045-3057.

87. Law, J.H., et al., Phosphorylated insulin-like growth factor-i/insulin receptor is present in all breast cancer subtypes and is related to poor survival. Cancer Res, 2008. **68**(24): p. 10238-46.

88. Habibi, G., et al., Redefining prognostic factors for breast cancer: YB-1 is a stronger predictor of relapse and disease-specific survival than estrogen receptor or HER-2 across all tumor subtypes. Breast Cancer Res, 2008. **10**(5): p. R86.

89. Sachdev, D. and D. Yee, Inhibitors of insulin-like growth factor signaling: a therapeutic approach for breast cancer. J Mammary Gland Biol Neoplasia, 2006. **11**(1): p. 27-39.

90. Lu, Y., et al., Insulin-like growth factor-I receptor signaling and resistance to trastuzumab (Herceptin). J Natl Cancer Inst, 2001. **93**(24): p. 1852-7.

91. Buck, E., et al., Compensatory insulin receptor (IR) activation on inhibition of insulin-like growth factor-1 receptor (IGF-1R): rationale for cotargeting IGF-1R and IR in cancer. Mol Cancer Ther, 2010. **9**(10): p. 2652-64.

92. Lemmon, M.A. and J. Schlessinger, Cell signaling by receptor tyrosine kinases. Cell, 2010. **141**(7): p. 1117-34.

93. Cancer Genome Atlas, N., Comprehensive molecular portraits of human breast tumours. Nature, 2012. **490**(7418): p. 61-70.

94. Chakravarti, A., J.S. Loeffler, and N.J. Dyson, Insulin-like growth factor receptor I mediates resistance to anti-epidermal growth factor receptor therapy in primary human glioblastoma cells through continued activation of phosphoinositide 3-kinase signaling. Cancer Res, 2002. **62**(1): p. 200-7.

95. Jones, H.E., et al., Insulin-like growth factor-I receptor signalling and acquired resistance to gefitinib (ZD1839; Iressa) in human breast and prostate cancer cells. Endocr Relat Cancer, 2004. **11**(4): p. 793-814.

96. Fitzgerald, J.B., et al., MM-141, an IGF-IR- and ErbB3-directed bispecific antibody, overcomes network adaptations that limit activity of IGF-IR inhibitors. Mol Cancer Ther, 2014. **13**(2): p. 410-25.

97. Edgar, R., M. Domrachev, and A.E. Lash, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res, 2002. **30**(1): p. 207-10.

98. Neve, R.M., et al., A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell, 2006. **10**(6): p. 515-27.

99. Hu, J., et al., Non-parametric quantification of protein lysate arrays. Bioinformatics, 2007. **23**(15): p. 1986-94.

100. Tibshirani, R., Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society Series B-Methodological, 1996. **58**(1): p. 267-288.

101. Tibshirani, R., Regression shrinkage and selection via the lasso: a retrospective. Journal of the Royal Statistical Society Series B-Statistical Methodology, 2011. **73**: p. 273-282.

102. Friedman, J., T. Hastie, and R. Tibshirani, Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw, 2010. **33**(1): p. 1-22.

103. Qian, J., Hastie, T., Friedman, J., Tibshirani, R. and Simon, N. Glmnet for Matlab. 2013; Available from: http://www.stanford.edu/~hastie/glmnet_matlab/.

104. Yamasaki, T., et al., Phosphorylation of Activation Transcription Factor-2 at Serine 121 by Protein Kinase C Controls c-Jun-mediated Activation of Transcription. J Biol Chem, 2009. **284**(13): p. 8567-81.

105. Tsao, H.K., P.H. Chiu, and S.H. Sun, PKC-dependent ERK phosphorylation is essential for P2X7 receptor-mediated neuronal differentiation of neural progenitor cells. Cell Death Dis, 2013. **4**: p. e751.

106. Dick, F.A. and S.M. Rubin, Molecular mechanisms underlying RB protein function. Nature Reviews Molecular Cell Biology, 2013. **14**(5): p. 297-306.

107. Lamber, E.P., et al., Structural Insights into the Mechanism of Phosphoregulation of the Retinoblastoma Protein. Plos One, 2013. **8**(3).

108. Lapenna, S. and A. Giordano, Cell cycle kinases as therapeutic targets for cancer. Nature Reviews Drug Discovery, 2009. **8**(7): p. 547-566.

109. Ren, S.J. and B.J. Rollins, Cyclin C/Cdk3 promotes Rb-dependent G0 exit. Cell, 2004. **117**(2): p. 239-251.

110. Taniguchi, C.M., B. Emanuelli, and C.R. Kahn, Critical nodes in signalling pathways: insights into insulin action. Nat Rev Mol Cell Biol, 2006. **7**(2): p. 85-96.

111. Citri, A. and Y. Yarden, EGF-ERBB signalling: towards the systems level. Nat Rev Mol Cell Biol, 2006. **7**(7): p. 505-16.

112. Yarden, Y. and M.X. Sliwkowski, Untangling the ErbB signalling network. Nat Rev Mol Cell Biol, 2001. **2**(2): p. 127-37.

113. Baselga, J. and S.M. Swain, Novel anticancer targets: revisiting ERBB2 and discovering ERBB3. Nat Rev Cancer, 2009. **9**(7): p. 463-75.

114. Emde, A., et al., Therapeutic strategies and mechanisms of tumorigenesis of HER2-overexpressing breast cancer. Crit Rev Oncol Hematol, 2012. **84 Suppl 1**: p. e49-57.

115. Szklarczyk, D., et al., STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Research, 2015. **43**(D1): p. D447-D452.

116. Koboldt, D.C., et al., Comprehensive molecular portraits of human breast tumours. Nature, 2012. **490**(7418): p. 61-70.

117. Vempati, U.D., et al., Metadata Standard and Data Exchange Specifications to Describe, Model, and Integrate Complex and Diverse High-Throughput Screening Data from the Library of Integrated Network-based Cellular Signatures (LINCS). J Biomol Screen, 2014. **19**(5): p. 803-816.

118. Efron, B., et al., Least angle regression. Annals of Statistics, 2004. **32**(2): p. 407-451.

119. Tibshirani, R.J., The lasso problem and uniqueness. Electronic Journal of Statistics, 2013. **7**: p. 1456-1490.

120. Lockhart, R., et al., A Significance Test for the Lasso (Vol 42, Pg 518, 2014). Annals of Statistics, 2014. **42**(5): p. 2138-2139.

121. Gustafsson, M., M. Hornquist, and A. Lombardi, Constructing and analyzing a large-scale gene-to-gene regulatory network - Lasso-constrained inference and biological validation. Ieee-Acm Transactions on Computational Biology and Bioinformatics, 2005. **2**(3): p. 254-261.

122. Gustafsson, M., et al., Reverse Engineering of Gene Networks with LASSO and Nonlinear Basis Functions. Challenges of Systems Biology: Community Efforts to Harness Biological Complexity, 2009. **1158**: p. 265-275.

123. Fenton, T.R. and I.T. Gout, Functions and regulation of the 70 kDa ribosomal S6 kinases. International Journal of Biochemistry & Cell Biology, 2011. **43**(1): p. 47-59.

124. Shimobayashi, M. and M.N. Hall, Making new contacts: the mTOR network in metabolism and signalling crosstalk. Nature Reviews Molecular Cell Biology, 2014. **15**(3): p. 155-162.

125. Fullerton, M.D., et al., Single phosphorylation sites in Acc1 and Acc2 regulate lipid homeostasis and the insulin-sensitizing effects of metformin. Nat Med, 2013. **19**(12): p. 1649-54.

126. Jeon, S.M., N.S. Chandel, and N. Hay, AMPK regulates NADPH homeostasis to promote tumour cell survival during energy stress. Nature, 2012. **485**(7400): p. 661-5.

127. Pece, S., et al., Activation of the protein kinase Akt/PKB by the formation of E-cadherin-mediated cell-cell junctions. Evidence for the association of phosphatidylinositol 3-kinase with the E-cadherin adhesion complex. J Biol Chem, 1999. **274**(27): p. 19347-51.

128. De Santis, G., et al., E-cadherin directly contributes to PI3K/AKT activation by engaging the PI3K-p85 regulatory subunit to adherens junctions of ovarian carcinoma cells. Oncogene, 2009. **28**(9): p. 1206-17.

129. Lau, M.T., C. Klausen, and P.C. Leung, E-cadherin inhibits tumor cell growth by suppressing PI3K/Akt signaling via beta-catenin-Egr1-mediated PTEN expression. Oncogene, 2011. **30**(24): p. 2753-66.

130. Kim, E.Y., et al., Inhibition of mTORC1 induces loss of E-cadherin through AKT/GSK-3beta signaling-mediated upregulation of E-cadherin repressor complexes in non-small cell lung cancer cells. Respir Res, 2014. **15**: p. 26.

131. Lau, M.T. and P.C. Leung, The PI3K/Akt/mTOR signaling pathway mediates insulin-like growth factor 1-induced E-cadherin down-regulation and cell proliferation in ovarian cancer cells. Cancer Lett, 2012. **326**(2): p. 191-8.

132. Hong, K.O., et al., Inhibition of Akt activity induces the mesenchymal-to-epithelial reverting transition with restoring E-cadherin expression in KB and KOSCC-25B oral squamous cell carcinoma cells. J Exp Clin Cancer Res, 2009. **28**: p. 28.

133. Kim, H.J., et al., Constitutively active type I insulin-like growth factor receptor causes transformation and xenograft growth of immortalized mammary epithelial cells and is accompanied by an epithelial-to-mesenchymal transition mediated by NF-kappaB and snail. Mol Cell Biol, 2007. **27**(8): p. 3165-75.

134. Qian, X., et al., E-cadherin-mediated adhesion inhibits ligand-dependent activation of diverse receptor tyrosine kinases. EMBO J, 2004. **23**(8): p. 1739-48.

135. de-Freitas-Junior, J.C., et al., Insulin/IGF-I signaling pathways enhances tumor cell invasion through bisecting GlcNAc N-glycans modulation. an interplay with E-cadherin. PLoS One, 2013. **8**(11): p. e81579.

136. Nagle, A., et al., Loss of E-cadherin enhances IGF1-IGF1R pathway activation and sensitizes breast cancers to anti-IGF1R inhibitors. bioRxiv, 2018.

137. Ching, J.K., et al., Ataxia telangiectasia mutated impacts insulin-like growth factor 1 signalling in skeletal muscle. Exp Physiol, 2013. **98**(2): p. 526-35.

138. Halaby, M.J., et al., ATM protein kinase mediates full activation of Akt and regulates glucose transporter 4 translocation by insulin in muscle cells. Cell Signal, 2008. **20**(8): p. 1555-63.

139. Viniegra, J.G., et al., Full activation of PKB/Akt in response to insulin or ionizing radiation is mediated through ATM. J Biol Chem, 2005. **280**(6): p. 4029-36.

140. Yamamoto, K., et al., Kinase-dead ATM protein is highly oncogenic and can be preferentially targeted by Topo-isomerase I inhibitors. eLife, 2016. **5**: p. e14709.

141. Karr, J.R., et al., A whole-cell computational model predicts phenotype from genotype. Cell, 2012. **150**(2): p. 389-401.

142. Goldberg, A.P., et al., Emerging whole-cell modeling principles and methods. Curr Opin Biotechnol, 2017. **51**: p. 97-102.

143. Bianconi, F., et al., Computational model of EGFR and IGF1R pathways in lung cancer: A Systems Biology approach for Translational Oncology (vol 30, pg 142, 2012). Biotechnology Advances, 2013. **31**(2): p. 358-360.

144. Le Novere, N., et al., BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. Nucleic Acids Research, 2006. **34**: p. D689-D691.

145. Halasz, M., et al., Integrating network reconstruction with mechanistic modeling to predict cancer therapies. Sci Signal, 2016. **9**(455): p. ra114.

146. Sedaghat, A.R., A. Sherman, and M.J. Quon, A mathematical model of metabolic insulin signaling pathways. Am J Physiol Endocrinol Metab, 2002. **283**(5): p. E1084-101.

147. Borisov, N., et al., Systems-level interactions between insulin-EGF networks amplify mitogenic signaling. Molecular Systems Biology, 2009. **5**.

148. Blinov, M.L., et al., BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. Bioinformatics, 2004. **20**(17): p. 3289-3291.

149. Azeloglu, E.U. and R. Iyengar, Good practices for building dynamical models in systems biology. Science Signaling, 2015. **8**(371).

150. Chen, J., et al., Controlled dimerization of insulin-like growth factor-1 and insulin receptors reveal shared and distinct activities of holo and hybrid receptors. J Biol Chem, 2018.

151. Finney, A. and M. Hucka, Systems biology markup language: Level 2 and beyond. Biochem Soc Trans, 2003. **31**(Pt 6): p. 1472-3.

152. Simpson, A., et al., Insulin-Like Growth Factor (IGF) Pathway Targeting in Cancer: Role of the IGF Axis and Opportunities for Future Combination Studies. Targeted Oncology, 2017. **12**(5): p. 571-597.

153. Nyman, E., G. Cedersund, and P. Stralfors, Insulin signaling - mathematical modeling comes of age. Trends Endocrinol Metab, 2012. **23**(3): p. 107-15.

154. Sonntag, A.G., et al., A modelling-experimental approach reveals insulin receptor substrate (IRS)-dependent regulation of adenosine monosphosphate-dependent kinase (AMPK) by insulin. FEBS J, 2012. **279**(18): p. 3314-28.

155. Azad, A., A. Lawen, and J.M. Keith, Prediction of signaling cross-talks contributing to acquired drug resistance in breast cancer cells by Bayesian statistical modeling. BMC Syst Biol, 2015. **9**: p. 2.

156. Stingo, F.C., et al., Incorporating Biological Information into Linear Models: A Bayesian Approach to the Selection of Pathways and Genes. Ann Appl Stat, 2011. **5**(3): p. 1978-2002.

157. Wu, M., X. Yang, and C. Chan, A dynamic analysis of IRS-PKR signaling in liver cells: a discrete modeling approach. PLoS One, 2009. **4**(12): p. e8040.

158. Wang, R.S., A. Saadatpour, and R. Albert, Boolean modeling in systems biology: an overview of methodology and applications. Physical Biology, 2012. **9**(5).

159. Shi, T., et al., Conservation of protein abundance patterns reveals the regulatory architecture of the EGFR-MAPK pathway. Sci Signal, 2016. **9**(436): p. rs6.

160. Cai, W., et al., Domain-dependent effects of insulin and IGF-1 receptors on signalling and gene expression. Nat Commun, 2017. **8**: p. 14892.

161. Gough, A., et al., Biologically Relevant Heterogeneity: Metrics and Practical Insights. SLAS Discov, 2017. **22**(3): p. 213-237.

162. Gerdes, M.J., et al., Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. Proc Natl Acad Sci U S A, 2013. **110**(29): p. 11982-7.

163. Pampaloni, F., E.G. Reynaud, and E.H. Stelzer, The third dimension bridges the gap between cell culture and live tissue. Nat Rev Mol Cell Biol, 2007. **8**(10): p. 839-45.

164. Shamir, E.R. and A.J. Ewald, Three-dimensional organotypic culture: experimental models of mammalian biology and disease. Nat Rev Mol Cell Biol, 2014. **15**(10): p. 647-64.

165. Yamada, K.M. and E. Cukierman, Modeling tissue morphogenesis and cancer in 3D. Cell, 2007. **130**(4): p. 601-10.

166. Griffith, L.G. and M.A. Swartz, Capturing complex 3D tissue physiology in vitro. Nat Rev Mol Cell Biol, 2006. **7**(3): p. 211-24.

167. Riedl, A., et al., Comparison of cancer cells in 2D vs 3D culture reveals differences in AKT-mTOR-S6K signaling and drug responses. J Cell Sci, 2017. **130**(1): p. 203-218.

168. Edmondson, R., et al., Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors. Assay Drug Dev Technol, 2014. **12**(4): p. 207-18.

169. Imamura, Y., et al., Comparison of 2D- and 3D-culture models as drug-testing platforms in breast cancer. Oncol Rep, 2015. **33**(4): p. 1837-43.

170. Clevers, H., Modeling Development and Disease with Organoids. Cell, 2016. **165**(7): p. 1586-1597.

171. Kessenbrock, K., M. Brown, and Z. Werb, Measuring matrix metalloproteinase activity in macrophages and polymorphonuclear leukocytes. Curr Protoc Immunol, 2011. **Chapter 14**: p. Unit14 24.

172. Gu, L. and D.J. Mooney, Biomaterials and emerging anticancer therapeutics: engineering the microenvironment. Nat Rev Cancer, 2016. **16**(1): p. 56-66.

173. Miller, M.A. and R. Weissleder, Imaging of anticancer drug action in single cells. Nat Rev Cancer, 2017. **17**(7): p. 399-414.

174. Megason, S.G. and S.E. Fraser, Imaging in systems biology. Cell, 2007. **130**(5): p. 784-95.

175. Esch, E.W., A. Bahinski, and D. Huh, Organs-on-chips at the frontiers of drug discovery. Nat Rev Drug Discov, 2015. **14**(4): p. 248-60.

176. Byrne, A.T., et al., Interrogating open issues in cancer precision medicine with patient-derived xenografts. Nat Rev Cancer, 2017. **17**(4): p. 254-268.