2019

# A Heterogeneous And Multiscale Modeling Framework To Develop Patient-Specific Pharmacodynamic Systems Models In Cancer

Alokendra Kumar Ghosh
*University of Pennsylvania*

# A Heterogeneous And Multiscale Modeling Framework To Develop Patient-Specific Pharmacodynamic Systems Models In Cancer


## Abstract

Systems models of key signaling pathways in cancer have been extensively used to understand and explore the mechanisms of action of drugs and growth factors on cancer cell signaling. In general, such models predict the effect of environmental stimuli (both chemical such as for e.g., growth factor and drugs as well as mechanical such as matrix stiffness) in terms of activities of proteins such as ERK or AKT which are important regulators of cell fate decisions. Although such models have helped uncover important emergent properties of signaling networks such as ultrasensitivity, bistability, and oscillations, they miss many key features that would make them useful in a clinical setting. 1) The predictions of activity of proteins such as ERK or AKT cannot be directly translated into a clinically useful parameter such as cell kill rate. 2) They don't work as well when there are multiple biological processes operating under different time and length scales such as receptor-based signaling (4-6 hours) and cell cycle (24-48 hours). 3) The parameter space of such models often exhibits sloppy/stiff character which affects the accuracy of predictions and the robustness of these models. Apart from single-cell systems models of signaling, pharmacokinetic and cell population-based pharmacodynamic models are also extensively used to predict the efficacy of a particular therapy in a clinical setting. However, there are no direct or consistent ways of incorporating patient-specific gene/protein expression data in these models. This thesis describes the development and applications of a multiscale and multiparadigm framework for signaling and pharmacodynamic models that helps us address some of the above shortcomings. First two single scale systems models are described which introduces methods of exploration of parameter space and their effect on model predictions. Then the multiscale framework is described and it is applied to two different cancers - Prostate Adenocarcinoma and Nephroblastoma (Wilm's Tumor). Special mathematical techniques were used to develop algorithms that can integrate models of disparate time scales and time resolutions (continuous vs. discrete-time). Such multiscale modeling frameworks have great potential in the field of personalized medicine and in understanding the physics of cancer taking into account the biology of the cells.

# A HETEROGENEOUS AND MULTISCALE MODELING FRAMEWORK TO DEVELOP PATIENT-SPECIFIC PHARMACODYNAMIC SYSTEMS MODELS IN CANCER

Alok Ghosh

A DISSERTATION

in

Chemical and Biomolecular Engineering

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

---

Ravi Radhakrishnan, Professor of Bioengineering

Graduate Group Chairperson

---

John Crocker, Professor of Chemical and Biomolecular Engineering

Dissertation Committee

Dennis Discher, Professor of Bioengineering

Scott Diamond, Arthur E. Humphrey Professor of Chemical and Biomolecular Engineering

Wei Guo, Professor of Biology

A HETEROGENEOUS AND MULTISCALE MODELING FRAMEWORK TO DEVELOP PATIENT-SPECIFIC PHARMACODYNAMIC SYSTEMS MODELS IN CANCER

*Dedicated to my parents Mr. Jyotirindra Kumar Ghosh and Mrs. Uttara Ghosh*

# ACKNOWLEDGEMENT

I am thankful to many who helped me in my research both directly and indirectly. First and foremost I would like to thank my advisor Dr. Ravi Radhakrishnan for his guidance, patience and understanding throughout the duration of my PhD. It has been a great pleasure and learning experience working with Ravi. Secondly, I would like to thank the Computational Horizons in Cancer (CHIC) Consortium (led by Dr. Georgios Stamatakos and Dr. Norbert Graf) and UPenn Physical Sciences in Oncology (PSOC) group (led by Dr. Dennis Discher) and NCI for supporting my research and inviting me to several great conferences for talks and posters. It has been a great honor collaborating with researchers from these groups particularly Dr. Wei Guo and Dr. Bin Wu from Penn and Dr. Caterina Guiot and Dr. Ilaria Stura from University of Torino. I am also thankful to all present and former members of Radhakrishnan Lab for stimulating discussions and research collaborations. Finally, I would like to thank my parents — my father who passed away 6 years ago just before I started my doctoral studies at Penn and could not see me complete my degree. And my mother for her continued support and patience as we both went through difficult times and came out stronger.

# ABSTRACT

## A HETEROGENEOUS AND MULTISCALE MODELING FRAMEWORK TO DEVELOP PATIENT-SPECIFIC PHARMACODYNAMIC SYSTEMS MODELS IN CANCER

Alok Ghosh

Ravi Radhakrishnan

Systems models of key signaling pathways in cancer have been extensively used to understand and explore the mechanisms of action of drugs and growth factors on cancer cell signaling. In general, such models predict the effect of environmental stimuli (both chemical such as for e.g., growth factor and drugs as well as mechanical such as matrix stiffness) in terms of activities of proteins such as ERK or AKT which are important regulators of cell fate decisions. Although such models have helped uncover important emergent properties of signaling networks such as ultrasensitivity, bistability, and oscillations, they miss many key features that would make them useful in a clinical setting. 1) The predictions of activity of proteins such as ERK or AKT cannot be directly translated into a clinically useful parameter such as cell kill rate. 2) They don't work as well when there are multiple biological processes operating under different time and length scales such as receptor-based signaling (4-6 hours) and cell cycle (24-48 hours). 3) The parameter space of such models often exhibits sloppy/stiff character which affects the accuracy of predictions and the robustness of these models. Apart from single-cell systems models of signaling, pharmacokinetic and cell population-based pharmacodynamic models are also extensively used to predict the efficacy of a particular therapy in a clinical setting. However, there are no direct or consistent ways of incorporating patient-specific gene/protein expression data in these models. This thesis describes the development and applications of a multiscale and multiparadigm framework for signaling and pharmacodynamic models that helps us address some of the above shortcomings. First two single scale systems models are described which introduces methods of

exploration of parameter space and their effect on model predictions. Then the multiscale framework is described and it is applied to two different cancers - Prostate Adenocarcinoma and Nephroblastoma (Wilm's Tumor). Special mathematical techniques were used to develop algorithms that can integrate models of disparate time scales and time resolutions (continuous vs. discrete-time). Such multiscale modeling frameworks have great potential in the field of personalized medicine and in understanding the physics of cancer taking into account the biology of the cells.

LIST OF TABLES

# LIST OF ILLUSTRATIONS

# Part I

# Introduction and Preliminary Concepts

# CHAPTER 1 : INTRODUCTION

Cancer is driven by a complex interplay of processes that operate at multiple length and time scales from molecular, cellular, tissue to organ level. Processes operating at shorter time and length scales such as molecular and cellular regime can influence tissue/organ level behavior and vice versa (1). Cancer initiation, progression, and treatment outcome depend on both *intertumoral* and *intratumoral* heterogeneity. Intertumoral heterogeneity, which is the heterogeneity between patients who have tumors of the same histological types, arises from both germline and somatic mutations as well as environmental factors. It is distinct from intratumoral heterogeneity, which is the heterogeneity in a single tumor within a patient which can manifest itself both in space and time. These heterogeneities cause differences in the genetic makeup of the cells in a tumor and hence a difference in the activities of the underlying signaling pathways and eventually cell fate, response to drugs and development of resistance (2). Many cancers have multiple molecular subtypes characterized by specific molecular markers that determine the progression and outcome of the disease. For most cancers, the success of different treatment strategies is highly dependent on patient-specific gene/protein expression signatures (presence of mutations, overexpression of a gene, etc.) that alters the underlying cellular signaling pathways and the cell fate outcome. Due to this, the current strategy for cancer therapy has shifted from a generic one size fits all approach to a more personalized one based on patient-specific biomarkers or gene expression signatures obtained from tumor samples using technologies such as RNA-seq (3). Databases like The Cancer Genome Atlas (TCGA) are valuable sources of such patient gene/protein expression data and also clinical data such as tumor grades, treatment information, etc. However, present experimental analysis techniques are not suited to deal with such large volumes of data.

Mathematical modeling is emerging as a powerful and flexible tool to extract clinically relevant information from this data and predict treatment outcomes in a patient-specific way. Both statistical data-driven models which analyze clinical data from patients and

predict probabilistic scenarios of recurrence as well as mechanistic models of cell signaling and tumor growth have been employed. Apart from these types of models, pharmacokinetic/pharmacodynamic (PK/PD) models which determine optimum drug dosage/schedule for specific patients have been employed (4). Most of the modeling efforts to date have been disparate and restricted to a specific length or time scale, which is inadequate given the multiscale nature of cancer described above. Mechanistic models at the tissue scale consider a population of different cell types such as cancer cells and healthy cells. Cancer cells are considered as either sensitive to a drug treatment or resistant to it. Such models predict the response to different therapies in terms of change in tumor volumes (5). Mechanistic models of signaling pathways at a single-cell level have also been published for pathways such as Ras-MAPK, PI3K-AKT, which are frequently altered in different cancer types (6; 7). These models contain a large number of nonlinear ordinary differential equations describing the rate of change in expression levels of different proteins as a function of their interactions with other proteins. These equations can be numerically solved to obtain both the transient and steady-state activities of these proteins that characterize the state of the system. However, these steady-state activities of the proteins cannot be used in a straightforward way to determine cellular outcome such as proliferation or apoptosis. However, it is known from literature that the both the steady-state and transient activities of proteins such as ERK or AKT which are activated by Ras-MAPK and PI3K-AKT pathways, are strong regulators of cellular outcomes (8; 9). Cell fate is determined by pathways regulating the cell cycle and operates at a much larger time scale of 24 h to 48 h compared to 4 h to 6 h time scale of receptor-mediated signaling pathways. Also, limited quantitative data is available for these pathways, which makes it impossible to model these using a detailed ODE-based model. Discrete logic-based systems models such as Boolean models which only considers discrete activities of the proteins (on or off) have been more successful here (10).

What has been missing from these modeling efforts is a consistent way to integrate these models operating at different length and time scales and modeled using different techniques (continuous vs discrete models). Such integration is a non-trivial task not only from a

theoretical perspective (no general theory exists that is applicable to all cases) but also challenging computationally because most numerical solvers are not designed to handle such differences in time scales. However, in those cases where such models can be constructed, they are exceedingly valuable not only because they extend the scope of the individual models but they also allow us to incorporate patient-specific gene expression data in one model and predict its effect on another model operating at a different scale.

The objective of my Ph.D. research is to create such an integrated modeling framework that can couple models operating at different time scales. In the future chapters, I'll present my efforts towards this goal and the novel outcomes from the work in different areas of application. First I will describe some of the challenges in multiscale systems modeling as well as some shortcomings of the existing models. Then the existing methods will be described in more detail with some important results from the literature. This will be followed by the very important topic of the effect of parameter spaces in these nonlinear systems models and some related ideas on model sensitivity, robustness, and evolvability. I will apply these simple ideas to two projects with comparatively simple, 'single-scale' models. It will be shown that even in these simple models, considerable insight can be gained by analyzing parameter space sensitivity to assess robustness of the model predictions. After this, I will introduce the hybrid multiscale modeling framework and show its applications in cancers of prostate and nephroblastoma. Finally, I will show how this modeling framework can be extended and applied to different areas such as mechanotransduction by incorporating mechanical interaction of a cell with its environment. Another future area of interest is to incorporate these single-cell models to tissue level pharmacodynamic and pharmacokinetic models, make direct patient-specific predictions of treatment outcome and design optimum combination therapies to overcome drug resistance.

# CHAPTER 2 : CHALLENGES IN SYSTEMS MODELING

In the past decade, there have been massive advances in the fields of sequencing technologies, microscopy and biological imaging. These advances have enabled the accumulation of vast amounts of data at all levels of biological organization. At the molecular and cellular levels, the various genome, proteome, metabolome, and transcriptome projects have provided detailed descriptions of the basic parts of living organisms in unprecedented detail. Although these data provide us access to detailed information about the building blocks of biological systems, experimental complexities still limit the observations to very restricted spatial and/or temporal scales. To be able to use this data and understand how they determine the function of living organisms we need an integrated or systems approach to biology. The main challenge is relating these high dimensional datasets to higher-level phenotypic characteristics which can only be done if we can integrate all relevant information at multiple levels of organizations and recreate the dynamics of the system. This cannot be accomplished just using experimental observations and we need mathematical models that can operate at these levels of organization (11; 12)

## 2.1. Biological Organization Across Multiple Scales

Biological processes span over multiple orders of magnitude in both length and time. The magnitude of this span varies from organism to organism. For e.g. in humans, these processes range from length scales of single proteins (1 nm to 10 nm) to that of whole body (1 m) spanning over 10 orders of magnitude, while in time it can span from events such as receptor activation, ion channel gating (1 µs to 100 µs) to human lifetimes (in the order of $1 \times 10^9$ s) (13). So in the context of modeling biological systems, a natural way to define a *scale* is to split processes up according to their position in the biological hierarchy i.e. whether they represent interactions between proteins inside a cell, between cells, tissues, organs, etc. Although we can specify the location of a biological process in this hierarchy what is more difficult is to give specific values of the lengths at which transitions from one

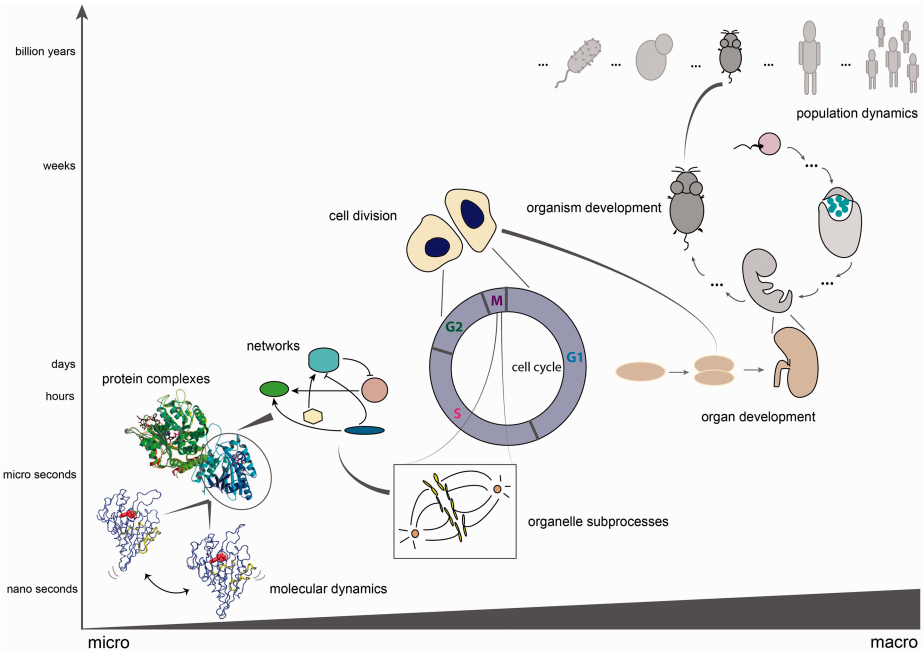


Figure 1: Levels of biological organization across multiple length and time scales spanning from molecular, cellular, tissue to organ and population level (14)

level to the next occurs. As shown in the Fig 1 these levels span from molecular, sub-cellular, cellular, tissue to organism levels (14).

The above scheme of organizing biological processes is not the only one possible. In many areas, an alternative classification scheme is more suitable, for e.g., in a model of tumor which is influenced by the information encoded in the DNA and presence of mutations. This encoded genetic information cannot be put in the above classification scheme and we may need a modified scheme such as genetic, proteomic, transcriptomic and cell and tumor level processes. So the particular classification scheme to be chosen is dependent on the nature of the system and not based on any rigid strategy.

## 2.2. Multiscale Model Challenges

A mathematical model, in its essence, is an approximate representation of a system in a mathematical form that usually contains a system of equations describing the governing laws and a description of the system geometry/domain where they are applied. For most

complex biological processes these systems of equations are too complex to be solved analytically and need numerical methods of solution. The common approach is to identify the basic constituents of the system (such as the cellular compartments, species like genes and proteins) and model their activities as system variables and the governing equations as algorithms/rules and then run the model to obtain its evolution over time and space. A *multi-scale* model would consist of biological processes that belong to two or more levels of organization mentioned above, for e.g., a model of tumor growth that considers subcellular signaling and gene regulatory networks with a model of cell proliferation and apoptosis at the level of a population of cells in the tissue. Apart from multi-scale models, we can also have models consisting of different physical processes for e.g., a model of tumor growth and its mechanical interaction with the extracellular medium. These models are called *multi-physics* models (11). The challenges that arise in these multiscale and multiphysics models are many and arise at different stages of modeling. Firstly, the component "single" scale models often represent the same phenomenon but different levels of description. The challenge lies in how the components were put together and how the variables were linked. For e.g., in the case of the multiphysics model of tumor growth and its interaction with ECM the challenge lies in linking the growth rate of the tumor with the force exerted on the ECM and vice versa. Secondly, a characteristic feature of biological processes is that they are influenced by processes operating at both lower and higher scales. For e.g., activation of receptors such as EGFR by ligand EGF happens at time scales of seconds characteristic of molecular interactions but the particular dynamics of this activation (transient or sustained) influences events like cell cycle progression operating at 24 h to 48 h time scales. Hence a model that consists of these processes must take into account these interactions that propagate across scales. There are also challenges that arise from the computational perspective. At the subcellular scales, the number of particles in a particular domain (such as on membrane surface or cytoplasm) is sufficiently large so that we can treat the system as a continuum and apply a deterministic model. However, in many instances (such as lipid domains on the membrane or in smaller compartments) the number of molecules

of the interacting species becomes too low and the continuum assumption breaks down. A stochastic model is necessary in these cases which is not only more involved computationally but is difficult to parametrize and validate experimentally. Even for deterministic models, when multiple models at different scales are simply coupled together the resulting system of nonlinear equations has a *stiff* character which is challenging to solve numerically especially for large systems. Last but not least, is the problem of available experimental data on these processes that are necessary to validate the models. As mentioned before, experimental measurements, in spite of recent advances, are limited and for many of these complex biological processes, there is often very little quantitative data. This severely limits the type of models we can create. As an example, continuous-time ordinary differential equations can be used when we have detailed time-course measurements for the participating species and their reaction kinetics. In the absence of this, we need to use a different strategy such as discretizing both the time and activity of the species. The number of free parameters in the model also profoundly influences how reliable a model is i.e. how tightly constrained are the model predictions. The larger the number of free parameters more uncertain is the model predictions. Fortunately, as we will see in Chapter 4, because of the *sloppy* nature of these systems models in many cases we can get away with higher uncertainties in specific combinations of system parameters.

## 2.3. Application in Clinic

Apart from the challenges related to bridging differences in length and time scales, there are additional areas that have been relatively unexplored although they are vital for mathematical systems models to be clinically useful. The first is related to modeling the effect of drug treatment and drug resistance particularly in diseases like cancer. The objective here is to predict the effect of a single or a combination of drugs on patients. Traditionally, modeling the effect of drugs has been done through pharmacokinetic/pharmacodynamic (PK/PD) models (15; 3) Most of these models are done without any multiscale considerations and are mostly phenomenological in nature. These models often relate drug type, dosages and

duration of treatment to a macroscopic parameter such as tumor volume reduction and are parameterized using experimental data coming from sources such as MRI. These parameters often have no direct correlation to any physiologically relevant parameters in the patients and cannot say anything about the underlying mechanism. The other quite related area is that of patient-specificity. Diseases like cancer have widely varying progression and treatment outcomes that depend on patient-specific factors such as the presence of specific mutations (such as K-Ras mutations in lung cancer), age and demographic factors. Many of these factors are captured in the specific gene/protein expression profiles obtained from bulk or single-cell sequencing of cells obtained from patient tumor samples. As of now, there has been no systematic approach for integrating this data in multiscale mechanistic models. This is mainly due to the difficulty of mapping patient-specific mRNA or miRNA expression into activities of proteins in a signaling network. Such integrations are the key for these models to be clinically useful and enabling personalized therapies that have been the "Holy Grail" of cancer systems biology.

# CHAPTER 3 : CONTINUOUS AND DISCRETE SYSTEM SOLVERS

Before describing the heterogeneous and multi-scale modeling framework developed in this work for coupling models operating at multiple scales, we will describe two main mathematical formalisms widely used in the field of systems biology. These are not the only formalisms that exist or even appropriate for all situations, but they have surprisingly wide applicability and are very good systems to apply multi-scale modeling strategies such as we have done here. However, this by no means is an exhaustive discussion on the subject. The reader is referred to the many of the cited references for more elaborate discussions.

As mentioned in Section 2.2, one of the key factors, that dictates the type of model we can choose, is the availability and nature of experimental data. Many cellular pathways which are often altered in cancer such as those mediated by Epidermal Receptor Growth Factor (EGFR) family like Ras-MAPK and PI3K/AKT pathways have been explored in-depth using high throughput measurements of time-course changes in the expression of different proteins, the kinetics of activation of receptors by ligands, etc. Due to the availability of such detailed data particularly time course measurements and reaction kinetics, it is possible to use detailed methods such as deterministic continuous-time ordinary differential equations, also called *reaction rate equations* (RRE). On the other hand, pathways such as tumor suppressor p53 mediated DNA damage pathway and cell cycle progression and apoptotic pathways are incompletely known with little quantitative data. Consequently, we need a more coarse-grained approach to model these networks such as discrete-time logic-based modeling. In other problems, one needs to consider the spatial distribution of the network components and how those affect other components of the network. These require us to consider additional dimensions in space and hence we need systems of partial differential equations that consider both reaction and transport of the species in the domain of interest. We will briefly mention this formalism which we use in one of the problems we were interested in — spatial regulation of Rab-GTPases on tubular recycling compartments. Other possibilities include cases where deterministic assumption no longer holds because

of higher levels of fluctuations in the system which requires stochastic models. These are described elsewhere (16).

### 3.1. Systems of Ordinary Differential Equations

Ordinary Differential Equations (ODE) are the most well-known formalism in mathematical modeling which has seen widespread adoption in systems biology. ODEs have been used in many areas from simple models consisting of a few components to understand emergent properties such as bistability and oscillations (17; 18) to large scale gene regulatory and protein interaction networks consisting of hundreds of species and reactions to determine the effect of growth factor and targeted therapy on signaling outcomes (19; 7).

The main advantages of using ODE in computational systems model are:

- Their success in representing the dynamics of real networks.
- Vast literature on both analytical and numerical methods for solving ODEs.
- The wide applicability of the ODE models and ease of integration of different cellular processes.
- Comparatively easier to validate experimentally and at the same time easily provide a mechanistic interpretation of data.

In this formalism, the concentrations of the constituent species such as RNAs, proteins, genes, etc. of a network are modeled as time-dependent variables. Various regulatory interactions are expressed as functional and differential relations between the variables.

The analysis starts with the topology of the underlying network. Let us suppose that, we have such a network of $n$ species and $m$ reactions which is most succinctly captured using the $m \times n$ stoichiometric matrix $\bar{\bar{N}}$ (20). Each row of the matrix corresponds to a particular molecular species and each column corresponds to a reaction.

Each element $j$ of a row $i$ (representing species $i$) indicates the role of species $i$ in the $j$th

reaction. If this element is zero it means that the species $i$ does not participate in reaction $j$. If this is non-zero, then the magnitude and the sign determine the relative quantity of the species that take part in the reaction and whether it is a reactant or a product respectively. It is to be noted that this matrix considers only those species whose concentrations change through the system dynamics and not any source/sink species that are needed to sustain a steady state.

The second factor needed to set up the system of ODEs is the kinetics of the $j$ reactions. This is incorporated in the system of ODEs below that relate the $m \times 1$ concentration vector $\bar{S}$ to the $n \times 1$ rate vector $\bar{\nu}$ as below

$$\frac{\mathrm{d}\bar{S}}{\mathrm{d}t} = \bar{\bar{N}}\bar{\nu} \tag{3.1}$$

In models of real systems, species are constrained by mass conservation relations. The total amount of a particular species is invariant in time and is determined only by the initial conditions of the system. These conserved quantities appear as linearly dependent rows of the stoichiometry matrix $\bar{\bar{N}}$. This allows us to partition the stoichiometry matrix into dependent and independent sets like below ($\bar{\bar{N}}_R$ is the independent and $\bar{\bar{N}}_0$ is the dependent component).

$$\bar{\bar{N}} = \begin{bmatrix} \bar{\bar{N}}_R \\ \bar{\bar{N}}_0 \end{bmatrix} \tag{3.2}$$

Defining a link zero matrix $\bar{\bar{L}}_0$ which satisfies $\bar{\bar{N}}_0 = \bar{\bar{L}}_0\bar{\bar{N}}_R$. Hence we can rewrite Eq 3.1 as

$$\begin{bmatrix} \bar{\bar{I}} \\ \bar{\bar{L}}_0 \end{bmatrix} \bar{\bar{N}}_R\bar{\nu} = \frac{\mathrm{d}\bar{\bar{S}}}{\mathrm{d}t} = \begin{bmatrix} \frac{\mathrm{d}\bar{\bar{S}}_i}{\mathrm{d}t} \\ \frac{\mathrm{d}\bar{\bar{S}}_d}{\mathrm{d}t} \end{bmatrix} \tag{3.3}$$

In 3.3 the terms $\frac{\mathrm{d}\bar{\bar{S}}_i}{\mathrm{d}t}$ and $\frac{\mathrm{d}\bar{\bar{S}}_t}{\mathrm{d}t}$ refer to the independent and dependent rates of change.

The relationship between the dependent and independent species can be shown to be given by

$$\bar{\bar{S}}_d(t) - \bar{\bar{S}}_d(0) = \bar{\bar{L}}_0(\bar{\bar{S}}_i(t) - \bar{\bar{S}}_i(0) \tag{3.4}$$

The equation 3.1 describes the time evolution of the system typically characterized by a *steady state* where net mass flow at the boundary and no concentration changes in time. This steady state equation is obtaied by setting Eq 3.1 to zero or $\bar{\bar{N}}\bar{\nu} = 0$ In general the rate vector $\bar{\bar{\nu}}$ is a nonlinear function.

As an example, in the gene regulatory network the particular form of the equation would be as below for the species $i$ (21):

$$\frac{\mathrm{d}S_i}{\mathrm{d}t} = \kappa_{i,i-1} f(S_{i-1}) - \gamma_i S_i,\ S_i \geq 0,\ 1 < i \leq n \tag{3.5}$$

In Eq 3.5 the rate of change of concentration of a species is made up of two terms. The first term is the rate of production where the parameters $\kappa_{1n}, \kappa_{21}, \cdots \kappa_{n,n-1} > 0$ are the production constants and $\gamma_1, \gamma_2, \cdots \gamma_n > 0$ are degradation constants.

The production term includes a nonlinear regulation function $f$ that represents the effect of the other species in the network on species $i$. In this case, the synthesis of $i$ only depends on the species $i-1$. This function is often linear or quadratic depending on how many input species contribute to the synthesis. Another common regulation function is the so-called *Hill curve*

$$h^+(S_j, \theta_{ij}, m) = \frac{S_j^m}{S_j^m + \theta_{ij}^m} \tag{3.6}$$

The function represents the influence of species $j$ on $i$. The parameter $\theta_{ij} > 0$ is the threshold and $m > 0$ is a steepness parameter. The function increases monotonically from 0 to 1 which expresses the fact that $j$ is an activator for species $i$. If $j$ is a repressor then the regulation function is replaced by $h^-(S_j, \theta_{ij}, m) = 1 - h^+(S_j, \theta_{ij}, m)$. The Hill curve has a characteristic sigmoidal shape for $m > 1$.

### 3.2. Piecewise-Linear Differential Equations

When detailed knowledge about reaction mechanisms are not available, as mentioned before, we need appropriate coarse-graining strategies. One such coarse-grained method that directly arises from the ODE formalism in 3.5 is piecewise linear differential equations (PLDE). The governing equation of PLDE is as shown below (21):

$$\frac{dS_i}{dt} = g_i(\bar{S}) - \gamma_i S_i,\ S_i \geq 0,\ 1 \leq i \leq n \tag{3.7}$$

In Eq 3.7 all parameters other than $g_i(S)$ has similar interpretation as Eq 3.5. For each gene $i$ that is regulated by a set of regulating genes $J$, $g_i$ is a sum consisting of terms like $\kappa_{iJ} \prod_{j \in J} r(S_j)$ where $r(S_j)$ is the regulation function and $\kappa_{iJ}$ is the maximum expression of $i$ under the influence of $J$.

When the regulation function $r(S_j)$ is a Hill function, one can simplify the analysis considerably by discretizing the Hill function into a step function.

$$s^+(S_j, \theta_{ij}) = \begin{cases} 1, & S_j > \theta_{ij} \\ 0, & S_j < \theta_{ij} \end{cases}, \quad s^-(C_j, \theta_{ij}) = 1 - s^+(C_j, \theta_{ij}) \tag{3.8}$$

These eliminate the nonlinearities in $g_i$. This approximation of a continuous sigmoid by a discrete step function is well-established in literature and it is something we use extensively in our hybrid multiscale framework.

As a result of this approximation, the resulting piecewise linear differential equation simplifies to:

$$\frac{dS_i}{dt} = b_i(\bar{\bar{S}}) - \gamma_i S_i,\ S_i \geq 0,\ 1 \leq i \leq n \tag{3.9}$$

In the above $b_i(\bar{\bar{S}})$ is a *piecewise-constant* function which is the weighted sum of products of the step functions. Using the threshold values $\theta_{ij}$ the domain of the variables $S_i$ can be divided into volumes at planes $S_i = \theta_{ij}$.

The piecewise-linear differential equations have two types of steady states (22) — *regular* steady states that lie inside the volume of the $n$-box or singular steady states that lie on the threshold planes.

An important consideration in PLDE is how much information is lost on coarse-graining the nonlinear ODEs. It has been shown that for most systems there is no qualitative difference in the solutions of these systems (23; 24). When the step functions are relaxed to moderately steep sigmoids the steady-state predicted from PLDE was preserved.

### 3.3. Boolean Models

A further level of coarse-graining from the nonlinear ODE system and the piecewise-linear differential equation system is obtained by reducing the state of a network component into a boolean variable. When it is on or has a value of 1 then that component is considered to be active and when it is off or has a value of 0 then that component is inactive. Also, the interaction between elements is represented by Boolean functions of the activities of other genes that interact with that element. Let the state of the Boolean model is represented by the variable $\hat{\bar{S}}$ so that the individual activities $\hat{S}_i$ are determined through a set of boolean functions $\hat{b}_i$. For a network of $n$ components, each component can be on or off and so the state space of the system consists of $2^n$ states. The state of the system at time point $t+1$ is calculated by using the functions $\hat{b}_i$ of a subset $k$ of the $n$ elements at the previous time

point $t$. This can be seen from the equation below

$$\hat{S}_i(t+1) = \hat{b_i}(\hat{S}_{i1}, \hat{S}_{i2}, \cdots \hat{S}_{ik}) \tag{3.10}$$

In the above equation, the update mode is synchronous which means all the variables are updated simultaneously and hence this model is deterministic. The boolean function $\hat{S}_i$ performs logical operations AND, OR, NOT on the $k$ binary inputs.

One can represent the possible trajectories in the state space of the system using state transition diagrams. These types of models eventually reach a steady-state (fixed point) or a set of recurring states. These steady states are called *attractors* and the set of initial states that leads to a specific attractor is called the *basin of attraction*.

# CHAPTER 4 : PARAMETER SPACE SENSITIVITY, SLOPPYNESS AND ROBUSTNESS

## 4.1. Global Sensitivity Analysis

Real biological pathways are nonlinear and various parameters such as initial species expression, kinetic rate constants, etc. can undergo large deviations. Sensitivity analysis techniques are useful to understand how perturbations in these parameters influence the model outputs. Because these systems have a high degree of nonlinearity, simple local sensitivity analysis where each parameter is altered one at a time keeping the others fixed is not sufficient (25). Hence, we use a global sensitivity analysis for this model – more specifically the Sobol Sensitivity Analysis which is based on the analysis of variances in the model parameters and output.

### *4.1.1. Sobol Sensitivity Analysis*

Sobol sensitivity analysis is a variance-based technique and is uniquely well suited for complex nonlinear systems of moderate size. Here we give a very brief description of the method that is relevant to the sensitivity analysis done in this paper. A detailed description can be found in references (26; 27). We use many of the notations from (27) below.

Suppose we have a model with $k$ parameters $X_1, X_2, \ldots, X_k$ which are assumed to be independent random variables. The model output $Y$ is related to these parameters through the relation $Y = f(X_1, X_2, \ldots, X_k)$ where $f$ is a general nonlinear function. In a variance-based sensitivity analysis, we want to understand how variances in the individual and combination of parameters $X_1, X_2, \ldots, X_k$ factor into the variances in $Y$. To determine this, we can first fix a parameter $X_i$ to a value (say $V_i$) and then determine the model output averaged over all remaining parameters $X_j, \quad i \neq j$ (which is denoted with a condensed notation $\bar{X}_{\sim i}$). This average is $E_{\bar{X}_{\sim i}}(Y|X_i = V_i)$ which will be different for different $V_j$. The variance of this average over all possible $V_i$ which is $V_{X_i}(E_{\bar{X}_{\sim i}}(Y|X_i = V_i))$ will give us the net first

order effect of variation in $X_i$ on the variation in $Y$. The first order sensitivity $S_i$ associated with parameter $X_i$ is defined as:

$$S_i = \frac{V_{X_i}(E_{\bar{X}_{\sim i}}(Y|X_i = V_i))}{V(Y)} \tag{4.1}$$

In 4.1 $V(Y)$ is the overall variance in $Y$.

The other sensitivity parameter of interest is the total effect sensitivity $S_{T_i}$ which represents the first and all higher order effects of the parameter $X_i$ on the model output. To determine this, we can start with determining the first-order effect of all parameters except $X_i$ which is denoted by $\bar{X}_{\sim i}$. Again, we first find the value of $Y$ averaged over all $X_i$ keeping all other parameters $\bar{X}_{\sim i}$ fixed which is $E_{X_i}(Y|\bar{X}_{\sim i} = \bar{V}_{\sim i}$. We then find the variance of this quantity over all possible $\bar{X}_{\sim i}$ or $V_{\bar{X}_{\sim i}}(E_{X_i}(Y|\bar{X}_{\sim i} = \bar{V}_{\sim i})$. Then $V(Y) - V_{\bar{X}_{\sim i}}(E_{X_i}(Y|\bar{X}_{\sim i} = \bar{V}_{\sim i})$ must represent the contribution of all terms where $X_i$ appears. Dividing this by $V(Y)$ we get the total effect sensitivity:

$$S_{T_i} = \frac{V(Y) - V_{\bar{X}_{\sim i}}(E_{X_i}(Y|\bar{X}_{\sim i} = \bar{V}_{\sim i})}{V(Y)} = 1 - \frac{V_{\bar{X}_{\sim i}}(E_{X_i}(Y|\bar{X}_{\sim i} = \bar{V}_{\sim i})}{V(Y)} \tag{4.2}$$

This variance-based sensitivity analysis framework is based on a functional decomposition scheme where a square-integrable function $X_1, X_2, \ldots, X_k$ defined over $\Omega$, the k-dimensional unit hypercube, can be expressed as follows

$$f = f_0 + \sum_i f_i + \sum_i \sum_{j>i} f_{ij} + \cdots + f_{12\ldots k} \tag{4.3}$$

The normalization condition here is

$$\int_0^i f_{i_1 \ldots i_s}(x_{i_1}, x_{i_2}, \ldots, x_{i_s})\, \mathrm{d}x_w = 0 \tag{4.4}$$

In the 4.4 we have $w = i_1, \ldots, i_s$. Using this the various terms are calculated as

$$f_0 = \int f(x)\,\mathrm{d}x = E(Y)$$

$$f_i(x_i) = \int f(x) \prod_{w \neq i} \mathrm{d}x_w - f_0 = E_{\bar{X}_{\sim i}}(Y|X_i) - E(Y)$$

$$\begin{aligned} f_{ij}(x_i, x_j) &= \int f(x) \prod_{w \neq i,j} \mathrm{d}x_w - f_0 - f_i(x_i) - f_j(x_j) \\ &= E_{\bar{X}_{\sim ij}}(Y|X_i, X_j) - E_{\bar{X}_{\sim i}}(Y|X_i) - E_{\bar{X}_{\sim j}}(Y|X_j) - E(Y) \end{aligned} \tag{4.5}$$

Taking the variances of both sides of these equations give us

$$\begin{aligned} V_i = V(f_i) \ &= V_{X_i}(E_{\bar{X}_{\sim i}}(Y|X_i)) \\ V_{ij} = V(f_{ij}) &= V_{X_i X_j}(E_{\bar{X}_{\sim ij}}(Y|X_i, X_j)) - V_{X_i}(E_{\bar{X}_{\sim i}}(Y|X_i) - V_{X_j}(E_{\bar{X}_{\sim j}}(Y|X_j) \end{aligned} \tag{4.6}$$

All these variances are linked by:

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \cdots + V_{12\ldots k} \tag{4.7}$$

Dividing the above by $V(Y)$ we get

$$\sum_i S_i + \sum_i \sum_{j>i} S_{ij} + \cdots + S_{12\ldots k} = 1 \tag{4.8}$$

Since the sensitivity coefficients above are multidimensional integrals, the standard way of computing them is by using Monte Carlo type sampling. In Monte Carlo sampling an integral $I[f] = \int f(x)\ \mathrm{d}x$ is computed by generating a sequence of uniformly distributed random numbers and computing their expectation $I_N[f] = \frac{1}{N}\sum_{i=1}^{N} f(x_i)$. In this case, we have a k-dimensional function $Y = f(X_1, X_2, \ldots, X_k)$. Hence, we need to sample $N$ times for each of these $k$ parameters. These samples can be represented by a $N \times k$ matrix. For the calculation of the above sensitivities, the standard procedure is to start with two independent $N \times k$ sampling matrices $\bar{\bar{A}}$ and $\bar{\bar{B}}$. We can compute matrices $\bar{\bar{A}}_B^i$ which is obtained by taking $\bar{\bar{A}}$ and replacing the ith column (for parameter $X_i$) by the corresponding column from $\bar{\bar{B}}$. Similarly, we can define $\bar{\bar{B}}_A^i$. It can be shown that (27) the variances in the equations for $S_i$ and $S_{Ti}$ can be estimated using:

$$V_{X_i}(E_{\bar{X}_{\sim i}}(Y|X_i = V_i)) = \frac{1}{N}\sum_{j=1}^{N} f(\bar{\bar{A}})_j f(\bar{\bar{B}}_A^i)_j - f_0^2$$
$$V_{\bar{X}_{\sim i}}(E_{X_i}(Y|\bar{X}_{\sim i} = \bar{V}_{\sim i})) = \frac{1}{N}\sum_{j=1}^{N} f(\bar{\bar{A}})_j f(\bar{\bar{A}}_B^i)_j - f_0^2 \tag{4.9}$$

The above equations form the basis of the computation of the sensitivity coefficients for the model.

### *4.1.2. Quasi-random sequences*

Monte Carlo method using a sequence of random or pseudorandom numbers is extensively used to compute multidimensional integrals like above. The Central Limit Theorem of probability shows that the error in the Monte Carlo estimate is equal to the product of the standard deviation of the function and $N^{\frac{1}{2}}$ where $N$ is the number of samples. Hence the convergence of this method is $O(N^{\frac{1}{2}})$ which can be very slow (28). This method of sampling using pseudorandom numbers also suffers from a related problem of clumping where the sample points often tend to clump together and leaves empty spaces in between which is

magnified in higher dimensions. One alternative to obtaining a more uniform distribution of points is by using a stratified sampling method like Latin Hypercube Sampling which divides the intervals into equally spaced points. However, this only works when the dimensionality is low. For integrations in higher dimensions, an alternative sampling technique is applied called quasi-random sampling. A quantitative measure of uniformity of a sequence is a factor termed "discrepancy". Lower the discrepancy, more uniform is the sequence. Suppose we have a sequence of $N$ points $\{x_N\}$ in the $k$ dimensional unit cube $I^k$. For any subset $J \in I^k$ one can define the error in Monte Carlo estimate of the volume of $J$ as (28):

$$R_N(J) = \frac{1}{N}\#\{x_N \in J\} - m(J) \tag{4.10}$$

The discrepancy is then defined as some norm of $R_N(J)$. Formally, if $J$ is restricted to rectangular set and $E$ is all possible such sets then the discrepancy $D_N$ is defined as

$$D_N = \sup_{J \in E} |R_N(J)| \tag{4.11}$$

The Koksma-Hlawka inequality provides an upper bound for a Monte Carlo integration error as a product of the variance of the function and the discrepancy of the sequence. Hence, if one can generate a sequence of points in such a way as to minimize the discrepancy, then it can be used to obtain an improved estimate of the integral. Such sequences are called quasi-random sequences. These are not random at all but are generated deterministically to minimize discrepancy. Since they are not random, quasi-random sequences are more limited in scope than pseudorandom numbers. However, it can be shown that for integration the convergence rate of quasi-random sequences is $O(N^{-1}(\log N)^k)$ where k is any number which is considerably faster than the $O(N^{\frac{1}{2}})$ convergence of the standard Monte-Carlo method using pseudorandom sequences. There are various techniques for determining quasi random sequences. We use Sobol sequences (29) using a method suggested by Saltelli (27). The

software package SALib (30) was used for the computation of the Sobol coefficients along with custom python scripts and Matplotlib (31) for plotting.

## 4.2. Parameter Space Sloppiness and Robustness

Models in systems biology are approximate and often are poorly parametrized because the necessary rate constants have not been measured and/or the absolute cellular concentration of the species is not typically known. Also, these networks are almost always incompletely known as many components and interactions that play a role in determining the final outcome have not been yet identified. Because of this, the predictability of these models is poor in general. Even when we have a model that can explain the experimental data for particular situations it is not sufficient to answer some of the key questions biologists are interested in. These have to do with the concepts of robustness and evolvability. For example:

- If the model is perturbed due to mutations or internal fluctuations, is it robust enough to maintain its functions?
- If there is a change in the environmental conditions or if the system is subjected to external stress/stimuli, can the system adapt itself to cope with the changes?

In this section, we will define these concepts more carefully, give some mathematical interpretations.

### *4.2.1. Definitions*

**Sloppyness**

As mentioned in the previous section, systems biology models are ill-conditioned and the collective behavior cannot be in general used to infer the underlying parameters (32). However, most important models in systems biology have been shown to be *sloppy* (33) in the sense that the feasible parameter space of the model is highly anisotropic with few stiff and

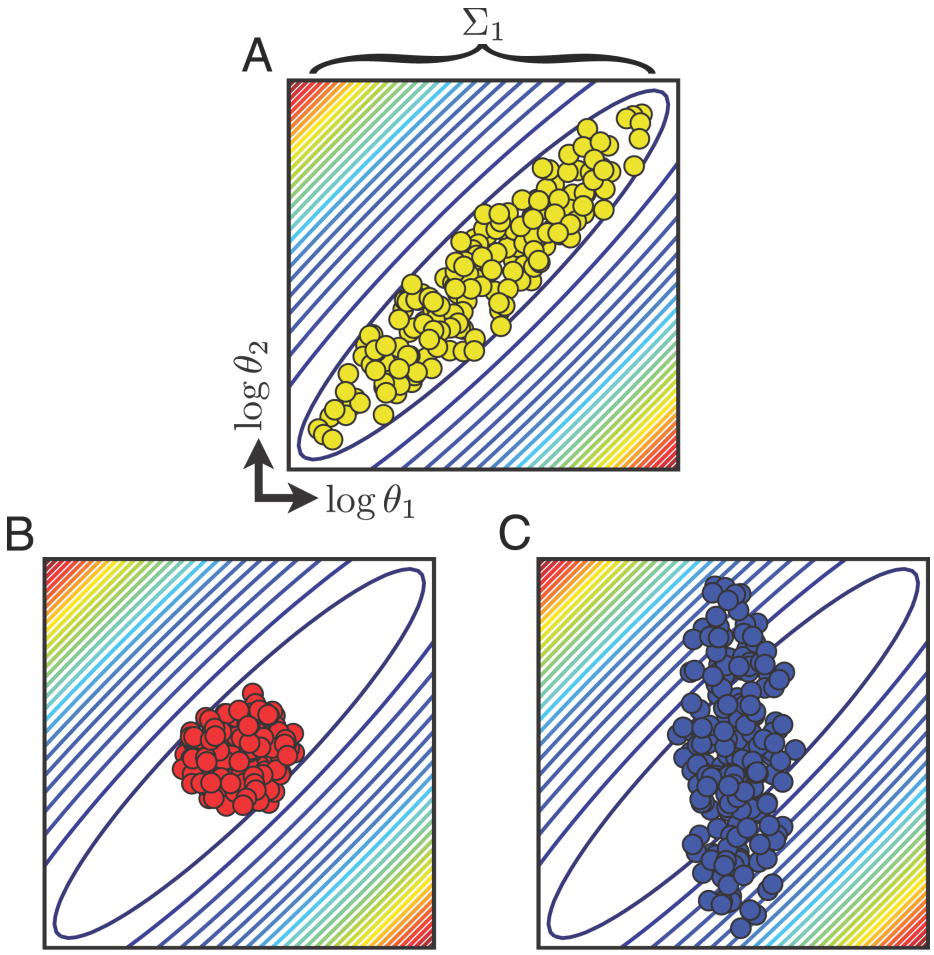


Figure 2: Parameter space spectra and sloppyness of systems models

many soft directions. What this means is that useful model predictions can be made even when a large set of parameters are unknown or known with a high degree of uncertainty. However, in the face of such parameter uncertainties, an attempt to find a single best model that explains the experimental observation is a futile endeavor. What will be more useful is to obtain an ensemble of models which are constrained by experimental data. Such an ensemble makes a model more falsifiable because any new experimental measurement can be compared with the ensemble deviation. The model will be considered defective if new data falls outside this deviation. This is shown in Fig 2 where the idea of an anisotropic parameter space is presented (33)

**Robustness**

There has been much debate over the correct definition of robustness in biology and a true definition has not yet been found. But most agree with the basic notion of robustness. We describe the idea of robustness following the pioneering work of Kitano (34) and Wagner (35). While describing robustness we must keep in mind that it is distinct from related concepts of *homeostasis* and *stability*. The key difference is that in the face of perturbations and uncertainties, homeostasis or stability tries to maintain the *state* of the system whereas robustness tries to maintain *function*. Thus a robust system can (and actually does) move itself to a different stable state or promote instability in the face of external stress or change in environmental conditions to maintain system functions. For e.g., , certain types of bacteria undergo a phenotypic switch in the face of harsh external conditions. Cancer cells in tumors survive chemotherapy by promoting instability — giving rise to a heterogeneous population. The key questions regarding the robustness of a biological system are:

- Is it conserved?
- What trade-offs exist in the biological system between robustness and fragility?

The concept of robustness is illustrated in Fig 3 along with the stability of a dynamical system (34)

Since the mathematical model mimics a biological process that is known to be robust, there should be ways to ascertain the robustness of the model that is consistent with the idea of biological robustness.

*4.2.2. Calculation of Sloppyness and Robustness*

**Kinetic ensemble method and sloppiness**

The basic principle of the ensemble method (36) is to generate an experimentally constrained random sample of the vector $\bar{\Theta}$, which is the set of all parameters and species concentrations

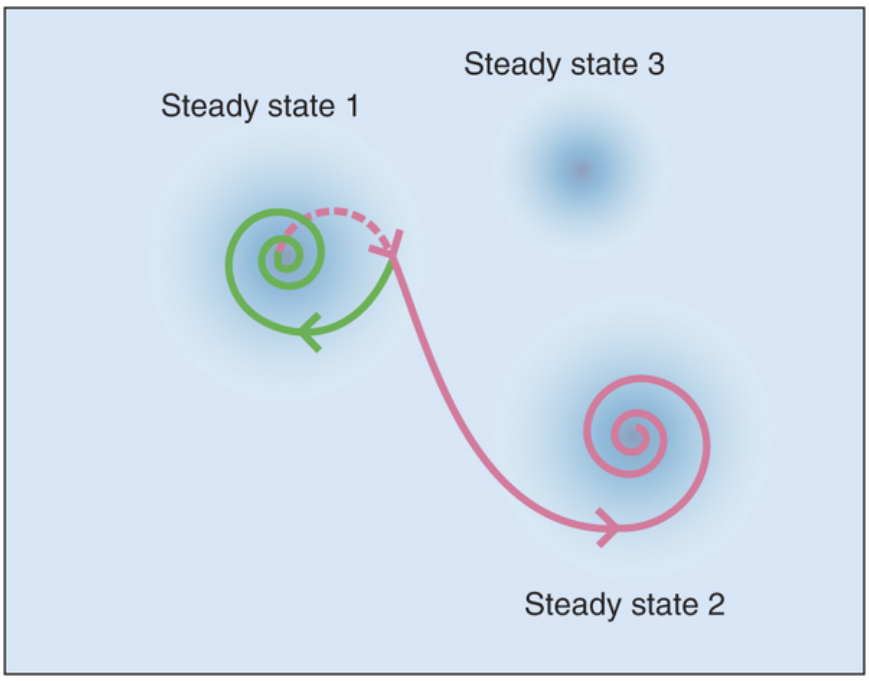


Figure 3: Robustness of dynamical systems

that specifies a model. The degree of consistency of the model with the experimental data can be quantified by calculating a certain figure of merit or a cost function. Let $\bar{Y} := (Y_1, ..., Y_D)$ is the D-tuple of all experimental observables. $D = M_S \times M_T \times M_E$ where:

- $M_E$ different realization of an experiment is performed. Each realization is distinguished by various conditions like external growth factor concentrations.
- $M_E$ different realization of an experiment is performed. Each realization is distinguished by various conditions like external growth factor concentrations.
- $M_S$ distinct species are present whose concentrations are being measured as a function of time.
- Concentrations are recorded for $M_T$ different time points

Now let $F(\bar{\Theta})$ be the vector of values corresponding to Y predicted from the model. Con-

sidering the experimental data is distributed normally (not essential) the probability distribution of the vector of observables $\bar{Y}$ given the mean $\bar{\mu}$

$$P(\bar{Y}|\bar{\mu}) \propto exp[-\sum_{l} \frac{(Y_i - \mu_i)^2}{2\sigma_l^2}] \tag{4.12}$$

Our quantity of interest here the probability distribution $Q(\bar{\Theta})$ of the parameters. Considering the experimental data we want to calculate the conditional probability $Q(\bar{\Theta}|\bar{Y})$ which we can write using a Bayesian interpretation as

$$(\bar{\Theta}|\bar{Y}) = \frac{p(\bar{\Theta})P(\bar{Y}|F(\bar{\Theta}))}{\int p(\bar{\Theta})P(\bar{Y}|F(\bar{\Theta}))} \tag{4.13}$$

If we choose a flat prior the above equation can be modified as

$$Q(\bar{\Theta}|\bar{Y}) = \frac{P(\bar{Y}|F(\bar{\Theta}))}{\int P(\bar{Y}|F(\bar{\Theta}))} = \frac{P(\bar{Y}|F(\bar{\Theta}))}{\Omega} \tag{4.14}$$

The probability $P(\bar{Y}|F(\bar{\Theta}))$ can be written as

$$P(\bar{Y}|F(\bar{\Theta})) \propto exp[-\sum_{l=1}^{D} \frac{(Y_i - F_i(\bar{\Theta}))^2}{2\sigma_l^2}] \tag{4.15}$$

(33) defines the factor inside the exponential as the cost function with a slightly modified expression as below

$$C(\bar{\Theta}) = [\sum_{l=1}^{D} \frac{(Y_i - B_k F_i(\bar{\Theta}))^2}{2\sigma_l^2}] + f(F(\bar{\Theta})) \tag{4.16}$$

Here the factors $B_k$ consists of undetermined multiplicative constants that restrict the interpretation of experimental data and nonlinear function f allows us to enter some fuzzy

data like inequalities.

Hence we can write the expression for conditional probability $Q(\Theta|Y)$ is proportional to

$$Q(\Theta|Y) \propto exp[-\beta C(\bar{\Theta})] \tag{4.17}$$

Hence using this ensemble we can set up a Monte Carlo scheme and perform moves in the parameter space $\bar{\Theta}$ and accept those that decrease the cost function and accept with a probability those that increase the cost function. Hence for any observable $O$ like a time-dependent chemical concentration we can calculate the average and the standard deviation as below

$$< O >= \frac{1}{N_E} \sum_{j=1}^{D} O_j \tag{4.18}$$

$$\sigma_O = \sqrt{< O^2 > - < O >^2} \tag{4.19}$$

From the definition of cost functions, we can identify the stiff and soft directions as below by computing the Hessian matrix defined below

$$H_{ij}(\theta^*) = \frac{\partial^2 C}{\partial \theta_i \partial \theta_j} \tag{4.20}$$

Alternatively in the spirit of the ensemble method, one can construct a covariance matrix using the ensemble data as below.

$$\Theta =< (\theta - < \theta >)(\theta - < \theta >)^T > \tag{4.21}$$

Eigenvalue decomposition of the matrix will also give us the stiff and soft directions.

**Calculation of Robustness**

Kitano (34) defines robustness taking into account his definition - which is the ability of a system to maintain its functions in the face of perturbations. So the robustness $R^s_{aP}$ of a system s with regard to its function can be defined as

$$R^s_{aP} = \int_P \psi(P) D^s_{a,p} dp \tag{4.22}$$

Here P is the space of all perturbations and $\psi(P)$ is the probability of a perturbation p. If all perturbations are equally likely this should be unity. $D^s_{a,p}$ is an evaluation function which gives the viability of the system under a perturbation. This is defined as

$$D^s_{a,p} = \begin{cases} 0 & p \in A \subset P \\ \frac{f_a(p)}{f_a(0)} & p \in P - A \end{cases} \tag{4.23}$$

In the above definition, the evaluation function is zero if p is in a set A which failed to maintain functions and it is given by a viability function when it is able to maintain function. The choice of the viability function is somewhat empirical and system dependent.

Using this definition we can compare two different systems S1 and S2 and define one S1 as more robust than the S2 when we have

$$R^S_{ap}1 > R^S_{ap}2 \tag{4.24}$$

The above comparison holds for a small perturbation space. When we consider a broad enough perturbation space then there will be trade-offs between robustness and fragility and the net contribution to the integral will vanish.

$$\Delta R_{aP}^{S1,S2} = \int_P \psi(P)(D_{a,p}^s 1 - D_{a,p}^s 2)dp = 0 \tag{4.25}$$

which means that $R_{aP}^S 1 = R_{aP}^S 1$

# Part II

# Single-Scale Models

# CHAPTER 5 : SPATIAL REGULATION AND SIGNALING OF RAB11 AND RAB8 GTPASES IN TUBULAR RECYCLING PATHWAY

Adapted from (37)

The first system we studied using the ideas described in the preceding chapters is trafficking and signaling of Rab-GTPases. This model is considered to be a "single-scale" model in the sense that time-scales of the underlying diffusion and reaction processes are similar to one another. However, it is still a suitable system to apply the sensitivity analysis techniques and identify important model parameters that influence both transport and signaling activities of the Rab proteins.

## 5.1. Introduction

Membrane trafficking in eukaryotic cells needs to be precisely regulated to ensure accurate delivery of cargo from one organelle to another. Rab-GTPases have emerged as the key regulators of this process which maintain membrane specificity by physically associating themselves with each organelle as well as their transport vesicles. These Rab proteins function as molecular switches by cycling between the GTP and GDP-bound states. The switch to the GTP-bound state is regulated by guanine nucleotide exchange factors (GEF) and is coupled to Rab activation and membrane association, whereas the switch to the GDP-bound state is regulated by GTPase activation proteins (GAP) and is coupled to Rab inactivation and membrane dissociation (38; 39; 40; 41). The Rabs localize to distinct organelles and mediate the flow of cargo by recruiting the GEFs and GAPs which causes segregation and eventual transition of the organelle from an upstream Rab to a downstream Rab. This is facilitated by a countercurrent cascade operation where a downstream Rab is recruited and activated and an upstream Rab is inactivated and removed. For example, on the exocytotic route, it was shown that the GEFs of the downstream Rabs are recruited by the upstream Rabs along the membrane trafficking pathways (42; 43; 44). In addition to the GEF cascade, it was reported that, in yeast Golgi trafficking, the downstream Rabs recruit

the GAPs of the upstream Rabs to terminate their function at specific Golgi compartment (45; 46; 47), suggesting a negative feedback regulation. In mammalian cells, Rab11 regulates recycling endosome (RE) trafficking from the perinuclear endocytic recycling compartments (ERC) to the plasma membrane (48; 49). The active GTP bound form of Rab11 has been shown to bind and recruit the effector Rabin8 which localizes on Rab11-positive vesicles (44). Rabin8 has specific GEF activity towards Rab8 which indicates that a Rab GEF cascade (analogous to Ypt32-Sec2-Sec4 cascade in yeast) is operational. To facilitate the conversion of a Rab11-positive vesicle to Rab8-positive one a countercurrent cascade needs to be operational. This means another as yet unknown effector of Rab8 must be recruited on these tubules to act as a GAP for Rab11 by facilitating its conversion to the GDP bound form and subsequent removal from the tubules. The protein Ecotropic viral integration site 5 protein homolog (Evi5) (50; 51) has been previously identified as a GAP for Rab11. Here we identify Evi5 as a direct downstream effector of Rab8 which provides strong evidence of the existence of such a countercurrent cascade and establish an important role of Rab11 and Rab8 in trafficking from late endosomal compartments. Evi5 was found to be recruited by Rab8 to the tubular-vesicular carriers (TVCs) where it inactivates Rab11 to complete the Rab conversion on TVCs. To establish the role of Evi5 as a GAP and to show that it promotes depletion of Rab11 from these vesicles where Rab8 is recruited, one needs to measure the spatial distribution of these proteins on these tubules. Unfortunately, this is beyond the scope of existing experimental techniques. Mathematical modeling is often useful under such circumstances. However, since accurate quantitative data is unavailable, in these scenarios the purpose of such models is to obtain qualitative insights into the mechanisms using simple and sound physical principles, generate testable predictions using different alternative hypothesis and guide further experiments. With these goals in mind, we have constructed a spatial reaction-diffusion model of Rab8-Rab11 cascade on the TVCs that incorporates both known and proposed interactions between Rab11, Rab8 and their effectors. We use the model to obtain both local and global behavior of the system and validate it against experimental results (performed globally), identify plausible mechanisms

to guide future experiments. Models that explore the recruitment and transport of Rab proteins and their roles in the budding and fusion of vesicles have been reported (52; 53; 54). However, placing different Rab proteins as part of a regulatory cascade in a spatiotemporal context has not been modeled. On the other hand, such reaction cascades with feedback have been modeled for systems such as the MAP kinase cascades, where the presence of nonlinear feedback profoundly influences the temporal and spatial activity of these proteins (6; 55). Our combined experimental and modeling study strongly suggests the existence of a Rab11-Rab8 cascade and the important role of Evi5 which is recruited by Rab8 on the TVCs. Evi5 was found to deactivate Rab11 and cause a depletion of Rab11 from Rab8 positive vesicles. The spatial model also suggested that assisted diffusion of Rab11-GTP through its interaction with motor proteins which have been reported in literature plays a key role in causing Rab11 depletion along with GDI mediated removal of the inactive form of Rab11.

## 5.2. Methods: Reaction-Diffusion Model of Rab Cascade

The transport of the different Rab proteins and their effectors take place on the TVCs whose longitudinal dimensions ( 1 µm) exceeds the lateral dimension (10 nm to 20 nm) by many orders of magnitude. Hence the domain of our model is a one-dimensional region of 1 µm length which is bounded on two sides by TGN/RE and the plasma membrane which are represented as compartments that hold constant (buffered) concentrations of different species. The general form of the reaction-diffusion equation that governs the concentration of a species $i$ in this one-dimensional domain is as shown below:

$$\frac{\partial c_i(x,t)}{\partial t} = D_i \frac{\partial^2 c_i(x,t)}{\partial x^2} - V_i \frac{\partial c_i(x,t)}{\partial x} - \sum_{\mu} R_{\mu} \tag{5.1}$$

Equation 5.1 is a simple mass balance relation that represents the net accumulation of a species $i$ as a combination of diffusive and reactive transport of the species. The first two terms on the right-hand side represent transport due to random and directed diffusion

respectively where $D_i$ is the random diffusivity of species $i$ and $V_i$ is the velocity of directed diffusion. The last term represents the net flux from all the reactions species i is part of, including those where it is generated or consumed. Since no quantitative measurements exist (to the best of our knowledge) for the reaction rates and diffusion coefficients of these Rab proteins we used corresponding values for Ras superfamily of GTPases as baselines. This is sufficient for our needs, as mentioned before the goal of this modeling effort is not to produce quantitatively accurate predictions of distribution of these Rab proteins (which would be impossible without detailed measurements of the spatial distribution of these proteins) but rather as a tool to explore our basic hypothesis of a countercurrent Rab11-Rab8 cascade. For this purpose, we only need to make sure that these parameters are constrained within biologically relevant ranges. The effect of the parameter space is explored in detail by creating an ensemble of models by sampling over these parameter ranges and calculating the error bounds for all the model predictions. We also perform a global sensitivity analysis to understand how sensitive are the model predictions with respect to the model parameters. We believe such analyses provide more value in these types of situations where quantitative data is lacking and/or experimental measurements are limited in scope because they help us find global behavior of the models that are robust to perturbations in parameters and environmental conditions. A detailed description of the boundary conditions, reaction kinetics, reaction rate and diffusion constants used in the model can be found in the Supplementary Materials of (37). Fig 5.2 shows a schematic diagram of the Rab cascade and the mechanism of diffusion

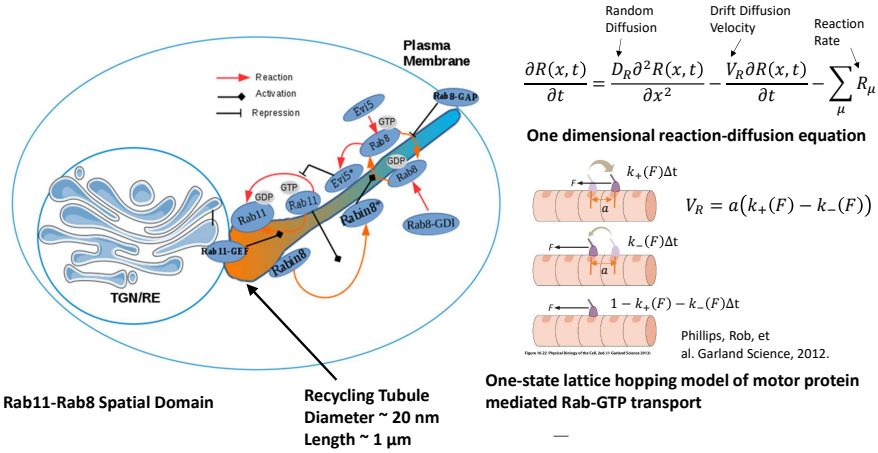


Figure 4: Rab11-Rab8 one dimensional reaction diffusion model with assisted transport: A tubular membrane extends from the compartment TGN/RE to the plasma membrane. Rab11 diffuses in the tubule from TGN/RE and Rab8 is delivered to the tubule by its GDI where it is activated by Rabin8 a downstream effector of Rab11. Evi5 is then recruited by activated Rab8 which deactivates Rab11 and causes its removal from the tubule. On the right we have the basic form of the PDE used for the model and a simple one-step lattice hopping model to simulate interaction of Rab11-GTP and Rab8-GTP with motor proteins which assist their diffusion along the tubules.

This system of reaction-diffusion equations is solved in a custom simulation platform consisting of Matlab (Matlab 9.0 The MathWorks, Inc., Natick, Massachusetts, United States) and python (Python Software Foundation. Python Language Reference, version 2.7 http://www.python.org). The scripts and instructions to run the scripts are available as part of the supplementary materials in (37).

## 5.3. Results

### *5.3.1. Evi5 is localized to the TVC*

TVCs are linear tubular membrane structures generated from ERC that carry cargos such as integrins and MHC-I back to the plasma membrane (48; 56; 57; 58). Rab8 is localized to TVCs, and functions together with EHD1 and MICAL-L1 for endosomal recycling (59; 60; 61). We localized Rab11 to perinuclear ERC, but not to TVCs marked by MICAL-L1 (Fig. 5A). The result is consistent with the previous observation (61). Evi5 was thought to be a GAP for Rab11 that negatively regulates the activity of Rab11 (50; 51; 62; 63).

Evi5 co-localizes with MICAL-L1 and Rab8 on the TVCs, as well as perinuclear ERC (Fig. 5B and C) (59). On the other hand, Evi5 was more enriched in the tubules proximal to the perinuclear region, whereas the MICAL-L1 staining extends from perinuclear ERC towards the plasma membrane (Fig. 5B). This observation suggests that the localization of Evi5 to TVCs is independent of Rab11.

### *5.3.2. Rab8 mediates the recruitment of Evi5 to the TVC*

Since Evi5 is enriched on the tubules marked by Rab8, we investigated whether Rab8 is involved in Evi5 recruitment to the TVCs. Using immunoprecipitation, we found that Evi5 interacts with Rab11 and Rab8, but not with Rab27, the Rab GTPase that functions in lysosomal trafficking (38) (Fig. 6A).

To test if Rab8 and Evi5 interact directly, we used recombinant proteins purified from bacteria to perform the binding assay. Rab8 bound to GST-tagged Evi5, but not to GST-tagged BBS3 or Sec8, which are of similar molecular weights (Fig. 6B). Conversely, GFP-tagged Rab8 bound to Evi5. The constitutively active mutant, Rab8Q67L, showed preferential binding to Evi5 relative to the wild type Rab8 or the dominant-negative mutant, Rab8T22N (Fig. 6C). To map the region of Evi5 that binds to Rab8, we constructed serial C-terminal truncations of Evi5 based on its domain structures (Fig. 6D). Deletion of a.a.714-810 at

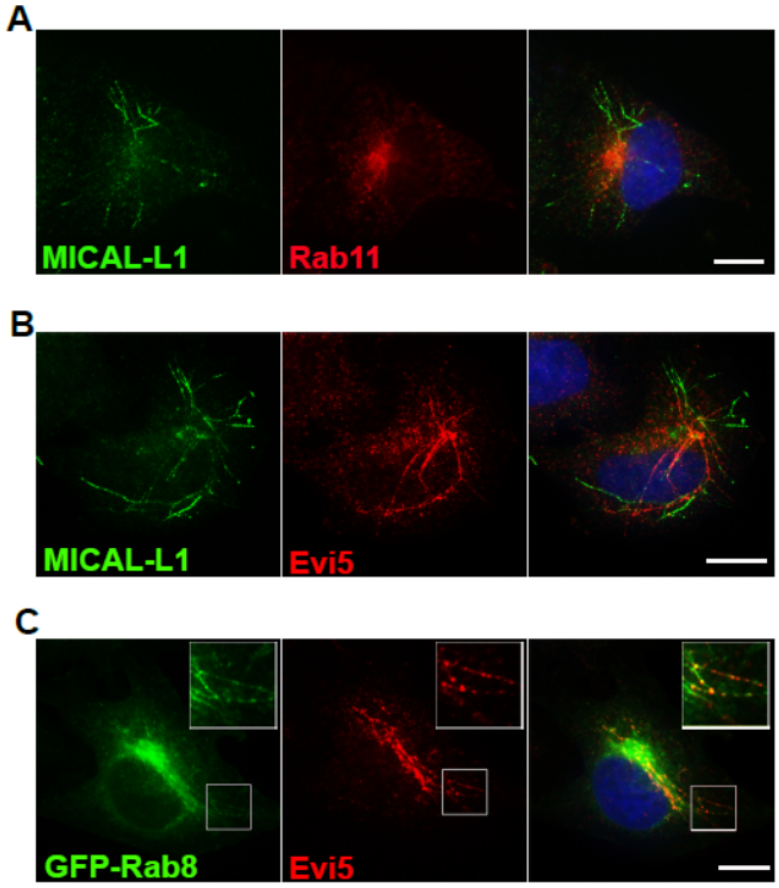


Figure 5: Evi5 is localized to tubular vesicular carriers (A) Evi5 localized to tubular vesicular carriers marked by MICAL-L1. Antibodies against the endogenous Evi5 and MICAL-L1 were used in the immunofluorescence staining. Scale bare=10$\mu$m. (B) Rab11 is mostly localized to perinuclear endosomes. Antibodies against the endogenous Rab11 and MICAL-L1 were used in the immunofluorescence staining. Scale bare=10$\mu$m. (C) Evi5 co-localized with Rab8 on endosomal tubules. Higher-magnification views of the boxed areas are shown in each image. Scale bare=10$\mu$m.

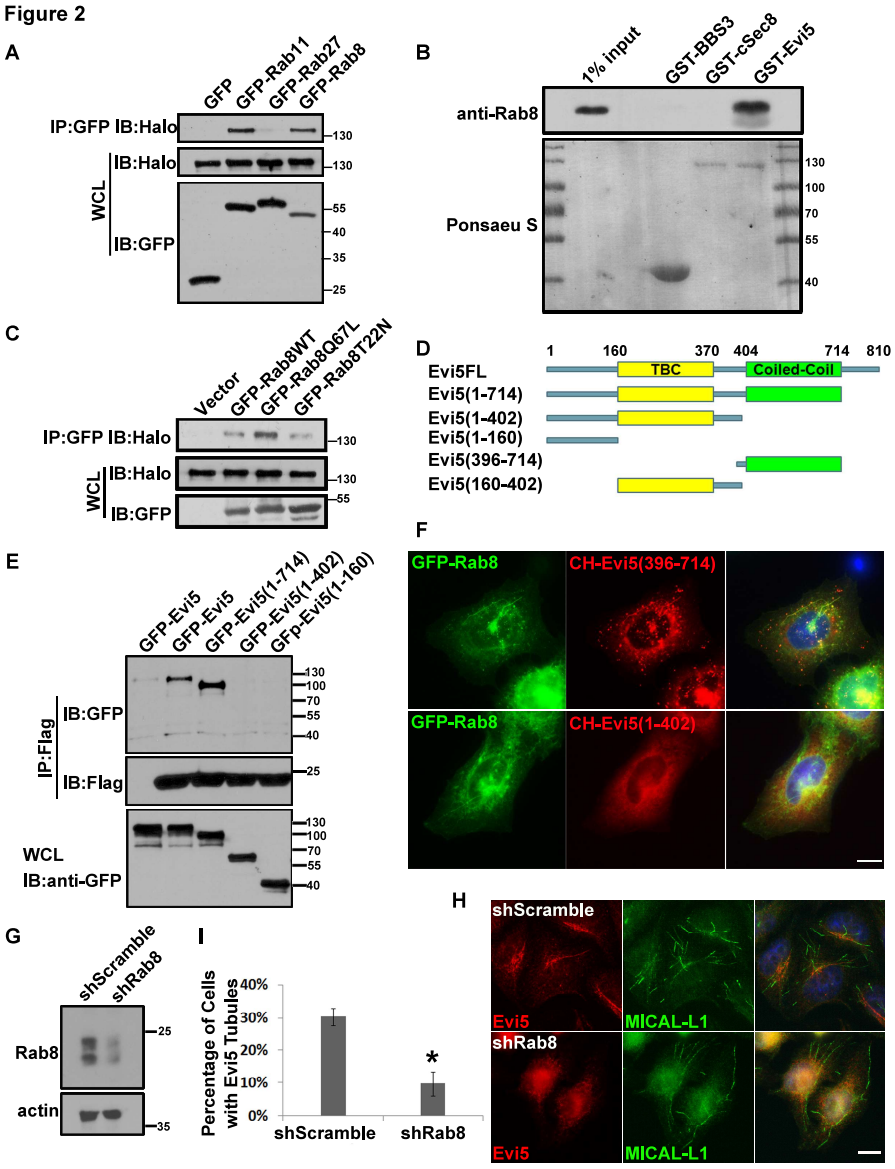


Figure 6: (A) HaloTag-Evi5 co-immunoprecipitated with GFP tagged Rab8, Rab11, and Rab27. Immunoprecipitates ("IP") and total cell lysate ("WCL") were immunoblotted with anti-HaloTag and anti-GFP antibodies. (B) Interaction of Evi5 and Rab8 in vitro. Purified Rab8 interacts with GST-Evi5, but not GST or GST-tagged BBS3 or C. elegan Sec8 ("cSec8"). Ponceau S. staining shows the inputs of GST fusion proteins. Rab8 was detected by western blotting. (C) Co-immunoprecipitation of HaloTag-Evi5 and GFP tagged Rab8 proteins with an anti-GFP antibody. Immunoprecipitates (IP) and total cell lysate (WCL) were analyzed with anti-HaloTag and anti-GFP antibodies. (D) Schematic representation of domains of Evi5 and Evi5 deletion constructs. Full-length Evi5 ("Evi5FL") contains a TBC domain and a coiled-coil region. (E) Co-immunoprecipitation of GFP-tagged Evi5 truncates with Flag-Rab8. Immunoprecipitates and total cell lysate were immunoblotted with anti-GFP and anti-Flag antibodies. (F) mCherry (CH)-tagged Evi5(396-714), but not Evi5(1-402) co-localized with GFP-Rab8. Scale bare=10μm. (G) Immunoblotting of Rab8 in HeLa cells treated with Scramble or Rab8 shRNA. Actin served as the loading control. (H) Immunostaining of Evi5 and MICAL-L1 in HeLa cells treated with Rab8 or Scramble shRNA. Evi5 localization to TVC was reduced in cells treated with Rab8 shRNA. Scale bare=10μm. (I) The percentage of cells containing Evi5 tubules. 100 cells were scored in each of 3 independent experiments. $*p < 0.001$.

the C-terminus did not affect the binding, but further deletion of the coiled-coil domain (a.a.404-714) abolished the interaction (Fig. 6E). Fluorescence microscopy experiments showed that the coiled-coil domain of Evi5 alone was sufficient to associate with TVCs marked by Rab8, while Evi5(a.a.1-402) containing the N-terminus and TBC domain did not (Fig. 6F). We also examined the binding of Evi5 with Rab11. In contrast to its interaction with Rab8, the N-terminus of Evi5 (a.a.1-160) was sufficient to interact with Rab11 (Supplemental Fig 1A in (37)). These results suggest that Evi5 interacts with Rab8 via its coiled-coil domain, while its interaction with Rab11 is mediated via its N-terminal region (Supplemental Fig. 1B in (37)). To test whether Rab8 mediates Evi5 localization to the TVC, we knocked down Rab8 by shRNA (Fig. 6G). The number of Evi5-positive tubules decreased significantly, while tubular staining of MICAL-L1 remained mostly unaffected (Fig. 6H and I). The binding assays and the localization studies suggest that Rab8 recruits Evi5 to the TVC.

#### *5.3.3. Spatiotemporal model of the Rab cascade simulated the regulation of Evi5 on the distribution of Rab11*

The results from the previous section suggest that Rab8, through its recruitment of Evi5 may play a role in the removal of Rab11 from the TVCs. If confirmed, this would establish the vital role of dual Rab11-Rab8 cascade in the trafficking processes from ERC to the plasma membrane along the TVCs. To investigate the operation and the effects of the proposed Rab11-Rab8 cascade we next constructed a spatial reaction-diffusion model. The model-specific details such as the main assumptions, component species/reactions, and values of the parameters can be found in Section 5.2 and Supplementary sections of (37).

For the TVCs, longitudinal dimensions (1 µm) are much larger than the lateral dimension (10 nm to 20 nm) which allows us to restrict the model in one dimension, simplifying the analysis and reducing the complexity of the model considerably. The first thing we want the model to tell us is the effect of Evi5 recruitment on both the amount and distribution of Rab11 (GTP, GDP bound forms and total). To determine this, we performed in-silico

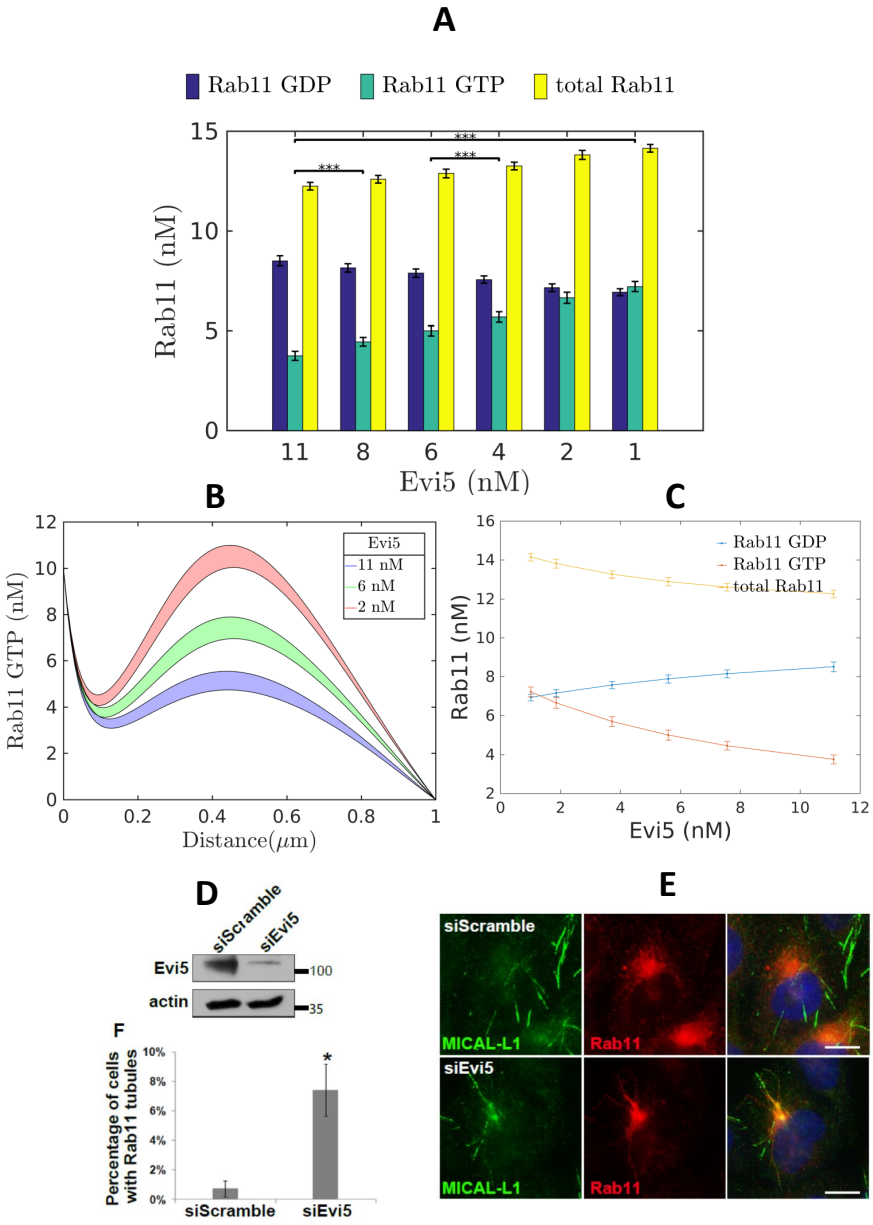


Figure 7: Steady State Rab11-GTP profiles for Evi5 knockdown (A) Model predicted steady-state concentration (nM) of Rab11-GDP, Rab11-GTP and total Rab11 averaged over the tubule length vs Evi5 (nM). This value gives an estimate of average amount of a species on the tubule and can be compared with corresponding experimental measurement. The black lines show the significance levels between different groups with significance level set to 0.01. * implies $p < .01$, each additional * indicates a p-value one order of magnitude lower. (B) Model predicted steady-state concentration profile of Rab11-GTP (nM) over the tubule for different Evi5 levels (nM). This plot shows how the spatial distribution of Rab11-GTP varies over the tubule as the amount of Evi5 levels on the tubule changes. (D) Immunoblotting of Evi5 in HeLa cells treated with Scramble or Evi5 siRNA. Actin served as the loading control. (E) Representative images of Rab11 and MICAL-L1 in HeLa cells transfected with Scramble or Evi5 siRNA. Endogenous Rab11 and MICAL-L1 were stained by antibodies. Scale bare=10μm. (F) The percentage of cells with Rab11 tubules. 200 cells were scored in each of 3 independent experiments. $*p < 0.01$.

Evi5 knockdown by running the model for a different rate of Evi5 recruitment and plotted the corresponding steady-state Rab11 levels. In Fig 7A and C we plot the steady-state concentrations of Rab11-GTP, Rab11-GDP and total Rab11 integrated over the tubule as Bar and Hill plots.

These quantities tell us about the net accumulation of these species on the tubule for different rates of Evi5 recruitment. The bar plots show us that the GTP bound form of Rab11 and total Rab11 increases with Evi5 knockdown at the expense of the GDP bound form. The Hill plots show the slopes (rates) of the corresponding changes and show a characteristic nonlinear slope which is frequently observed in these cascades with negative feedback. Fig 7 shows the steady-state distribution of Rab11-GTP over the tubule for three different Evi5 levels Higher rates of Evi5 recruitment causes localization of Rab11 near the left boundary. Supplementary Fig 3A and 3B in (37) show the profiles for total Rab8 and Evi5 level which shows that Rab11 is effectively "pushed" towards the TGN/RE boundary due to an increase in the Rab8 and Evi5 levels on the tubule. One possible mechanism for the increase in total Rab11 when Evi5 is knocked down comes from the assisted diffusion of the GTP bound form of Rab11 through its interaction with motor proteins which has been documented before (38). To test the model prediction of a decrease in total Rab11 levels with Evi5 knockdown, we performed RNA interference of Evi5 and examined the distribution of Rab11. Rab11 showed tubular localization only in a very small number of cells (Fig. 7E and 7F) under normal Evi5 levels. However, when Evi5 was knocked down by siRNA, the number of cells with Rab11 tubules increased significantly (Fig. 7D and E). This result is consistent with the model predictions in Fig 7A and C which plots the Rab11 levels integrated over the tubule for an ensemble of models (indicated by the error bars). In cells with Evi5 knockdown, Rab11 was partially localized to the MICAL-L1 tubules and was concentrated on the tubules close to the ERC (Fig. 7E), connecting ERC and MICAL-L1 tubules, which was consistent with the distribution of Evi5 (Fig. 5A). This is consistent with the spatial profile of Rab11-GTP predicted by the model in Figure 7B which also shows localization of Rab11 near the TGN/RE boundary.

### *5.3.4. Simulated and experimental knockdown of Rab8 show that Rab8 spatial distribution regulates Evi5 and Rab11 distribution on TVCs*

The Evi5 knockdown results from the previous section establish a necessary condition for an operational Rab11-Rab8 countercurrent cascade by showing that on being recruited, Evi5 can cause depletion of Rab11 from the tubules. This alone is not sufficient and we need to show that such a depletion is caused by Rab8-GTP recruiting Evi5 on the tubule. As mentioned before, activation of Rab8-GDP to Rab8-GTP is catalyzed by Rabin8, which is a downstream effector of Rab11-GTP. We carried out in-silico Rab8 knockdown in a similar fashion as the Evi5 knockdown studies in the section 5.3.3 and again plotted the corresponding steady-state Rab11 levels in Figures 8A and B.

These plots show a similar dependence of Rab11-GTP levels on Rab8 we saw before with Evi5 although with different rates. Taken together with the plots in Figures 7 and 8 confirm that a countercurrent cascade with Rab11 and Rab8 with Evi5 acting as an effector of Rab8-GTP and a GAP for Rab11 can cause a depletion of Rab11 from these TVCs. We next performed Rab11-GTP pulldown to verify the model prediction. Rab11FIP1 is an effector of Rab11, and its Rab-binding domain (RBD) specifically interacts with Rab11 in its GTP-bound form. We purified GST tagged RBD domain of Rab11FIP1 from bacteria to pull down Rab11-GTP from HeLa cell lysates. A higher level of Rab11-GTP was pulled down by GST-RBD from HeLa cells treated with Rab8 shRNA (Fig. 8D and E). This result was consistent with the simulation result and taken together, they confirmed the negative regulation of Rab11 by Rab8 through the GAP Evi5 providing support to the dual Rab8-Rab11 cascade hypothesis.

## 5.4. Discussion

In this work, using fluorescence imaging, biochemistry, and mathematical modeling, we have identified a regulatory loop that spatially controls Rab11 inactivation through Evi5 during endosomal recycling to the plasma membrane. Immunoprecipitation and fluorescence

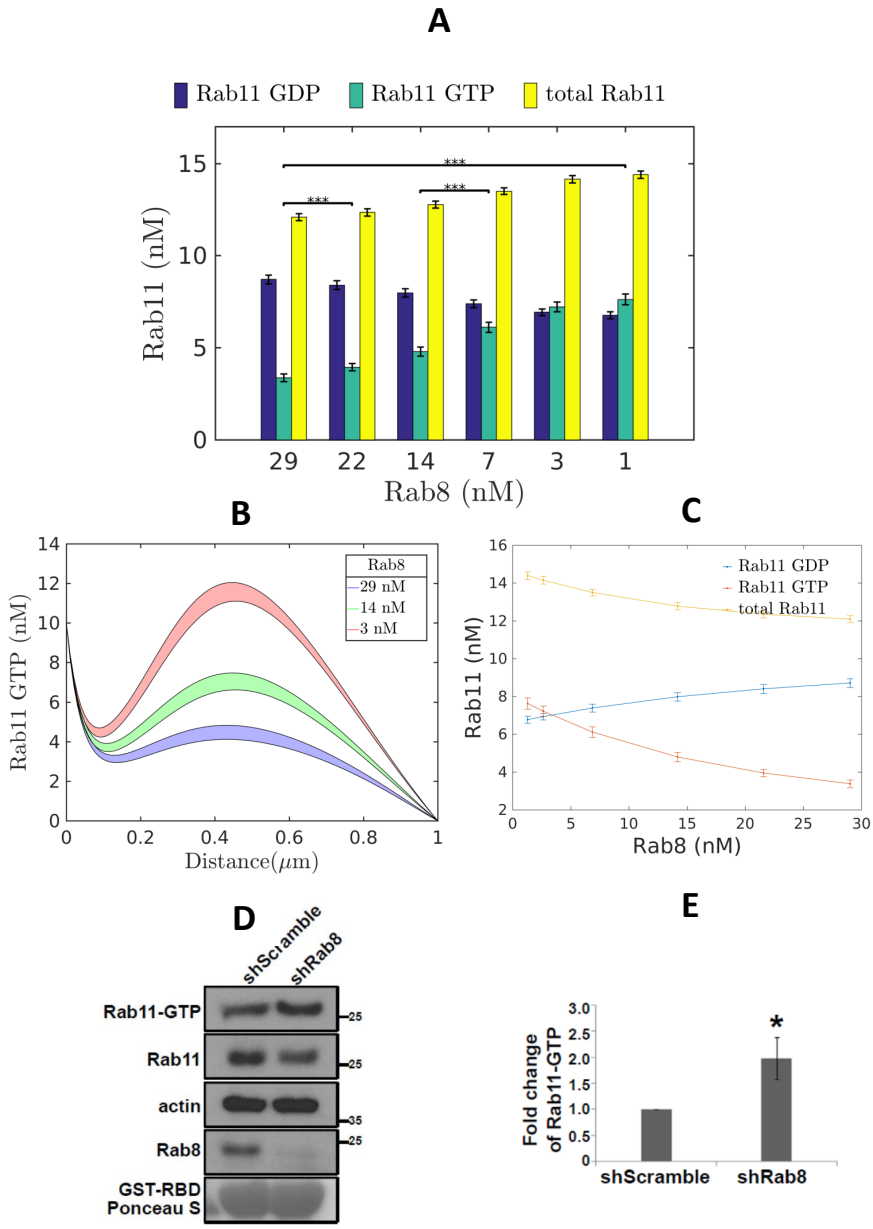


Figure 8: Effect of Rab8 knockdown on Rab11-GTP (A) Model-predicted steady-state concentration (nM) of Rab11-GTP averaged over entire tubule vs. the Rab8 levels (nM). This value gives an estimate of average amount of a species on the tubule and can be compared with corresponding experimental measurement. The black lines show the significance levels between different groups with significance level set to 0.01. * implies $p < .01$, each additional * indicates a p-value one order of magnitude lower. (B) Model-predicted steady-state Hill plot of Rab11-GTP vs. Rab8 concentration (nM) on the tubule highlighting the nonlinear dependence of Rab11-GTP on Rab8 levels. (C) Model predicted steady-state Hill plot of Rab11-GTP (averaged) vs Rab8 level (nM) showing the nonlinear variation of Rab11 with Rab8 knockdown. (D) Immunoblotting of Rab11-GTP in HeLa cells treated with Rab8 or scramble shRNA. The cell lysates were incubated with purified GST-FIP1 RBD fusion protein. GTP-Rab11 that bound to GST-FIP1 RBD fusion protein or in cell lysates were detected with the anti-Rab11 antibody. (E) Quantification of the levels of GTP-Rab11 in (D). The amounts of GTP-Rab11 were normalized to the level of shScramble (n=4). The intensity of the bands was quantified by ImageJ and analyzed using Student's t-test. Values are presented as mean±SD. $*p < 0.05$.

microscopy studies showed that Evi5 directly interacts with Rab11 through its N-terminal region, and interacts with Rab8 through its coiled-coil domain. Evi5 was also found to rely on Rab8 for its localization on the TVC. To obtain a mechanistic understanding of these observations we constructed a reaction-diffusion model of the Rab8-Rab11 cascade. Simulating Evi5 and Rab8 knockdown on the TVC lead to two important predictions:

1. Evi5 depletion on tubules results in an increase in total Rab11.

2. Rab8 depletion on tubules leads to a depletion of Evi5 and a corresponding increase of Rab11-GTP.

Thus, the model suggested an important connection between Rab8 and Rab11 activities on their spatial distribution possibly mediated by their interaction with Evi5. By performing Evi5 and Rab8 siRNA knockdown experiments, we observed an expansion of Rab11 on the tubules and increased the level of GTP-loading, confirming the mathematical predictions. This Rab cascade model is similar to models of Rab systems (52; 53) as well Ras-MAPK systems (55). However, this is one of the first (as far as the authors are aware) to model complete Rab cascade with negative feedback regulation in both space and time. For our model the transient dynamics were short-lived, and the system quickly settled to a steady-state and hence we presented the steady-state results. However, the nature of this steady-state is a very important aspect of these models. Previous studies on Ras-MAPK cascades have shown the emergence of bistable and oscillatory behavior at steady-state (6; 7). We did not observe this in our spatial model, but we are currently exploring this in a non-spatial context. Since the tubule length can influence the diffusion times and we ran out simulations for different lengths of tubules. We found that above 1 µm length there is no significant change in the effect of Evi5 knockdown on Rab11 although the spatial profiles and nature of dependence (slope) changes (Figures 8-9). One major limitation of this kind of model is the multiple free parameters with large uncertainties in their values. It is important to show that such model predictions are robust under perturbations of the parameters. Using Global Sensitivity Analysis (Supplementary Fig. 10 and Supplementary Materials)

we confirmed this and also identified a particular combination of parameters which were important determinants of model outcome like Rab11-GTP deactivation by Evi5. The effect of this reaction was separately explored (Supplementary Fig. 7), where it showed a large influence on the distribution of both Rab11-GTP and Rab11-GDP. Sensitivity analysis also showed a high degree of nonlinearity interaction between the sensitive parameters to determine the model outcome. Membrane traffic in eukaryotic cells is regulated by signaling pathways. The Rab8/Rab11 network is regulated by a number of kinases in response to growth factor signaling (64; 65; 66). Notably, the extracellular signal-regulated kinase (ERK) directly phosphorylates Rabin8 and promotes its GEF activity on Rab8, thereby regulating exocytic trafficking (67). The identification of the Rab11-Rab8 regulatory loop and the establishment of a mathematical framework will allow future studies to be performed to elucidate the molecular mechanisms of compartmental specificity and regulation of membrane trafficking. This study also has important implications for fields such as cancer research. For example, ERK is a downstream target of the Ras-MAPK pathway which is frequently dysregulated in many cancers. The important role of trafficking events such as receptor internalization and recycling on these signaling pathways have been highlighted in previous reviews (68; 69). Through their influence on these trafficking processes, the Rab proteins can determine signaling outcome, cellular fate and consequently the efficacy of different treatment strategies.

# CHAPTER 6 : TIME-DEPENDENT ANTAGONIST-AGONIST SWITCHING IN RECEPTOR TYROSINE KINASE-MEDIATED SIGNALING

Adapted from (70)

## 6.1. Introduction

ErbB4/HER4 belongs to the ErbB receptor tyrosine kinase family (consisting of EGFR, ErbB2, ErbB3, and ErbB4). These receptors regulate several critical signaling pathways that are frequently altered in cancers of lung, breast, prostate, etc. When activated by specific growth factors, they initiate multiple signaling cascades leading to the transcription of genes responsible for the determination of cell fate such as proliferation, differentiation or apoptosis (71). Overexpression of these receptors or the development of domain-specific mutations that allow them to be constitutively activated, causes them to promote various important pathways that drive the cell towards a program of proliferation and suppresses those that lead to apoptosis (cell death) or growth arrest through cell senescence. Compared to the other members of the ErbB family, the role of HER4 is incompletely understood (72). In part, this has to do with the fact that this receptor has several unique features compared to other members of the family. It has four structurally and functionally different isoforms generated by mRNA splicing. Some of these isoforms undergo shedding of their ectodomain by proteolytic cleavage reaction mediated by Tumor Necrosis Factor-alpha converting enzyme (TACE). The remaining 80 kDa fragment is further cleaved by enzyme $\gamma$-secretase which separates from the transmembrane part and transported to the nucleus and other parts of the cell thereby taking part in critical cellular reactions that affect cell fate (73). Unlike the other members of the ErbB family, HER4 has an anti-proliferative role through its activation of the JAK2-STAT5 pathway. Ligand Neuregulin (NRG), which is one of the ten natural ErbB family ligands and is often expressed in human tumors, can bind to and activate HER4. Upon activation and binding with JAK2, activated HER4 s80 domain can activate Signal Transducer and Activator of Transcription 5 (STAT5) causing it to translocate to the nucleus and initiate transcription of genes that promote differen-

tiation (72; 73) Hence it is essential to gain a mechanistic understanding of this pathway, particularly its dynamical behavior in response to NRG. Constructing and solving systems of nonlinear ordinary differential equations is a typical approach of modeling these pathways. The models are constructed based on experimental measurements of protein activities and rate constants and are used to predict the time-dependent dynamics of target gene/protein expression for different ligand stimulations. The predictions are validated through experimental measurements of the time-dependent protein activity through Western Blots or Immunoprecipitation studies (7). Such an approach has been used for HER4 mediated JAK-STAT pathway before (74; 75). However, these models failed to reproduce some interesting recent observations from experiments conducted in HC11 mouse mammary epithelial cell lines which prompted this computational study. In the experiments, it was observed that HER4 mediated transcription of anti-proliferative genes like β-casein milk genes follows a time-dependent switching behavior that cannot be explained by the previously published models. Here we propose an improved model of the HER4-JAK-STAT pathway that incorporates additional interactions that have been previously reported in the literature but were not part of the original models. Some of these reactions include competitive heterodimerization of HER4 receptors with other members of the ErbB family and JAK2 independent activation of HER4. The latter phenomenon of signaling through parallel pathways has some similarities with signaling pathways driven by constitutively activated mutant forms of EGFR (76). Identification of such pathways can be critical to gain a better understanding of the HER4 induced JAK/STAT pathway and leverage its anticancer role by designing an appropriate treatment strategy to treat cancers in cell lines where this receptor is significantly expressed.

## 6.2. Materials and Methods

### *6.2.1. Pathway Description*

The HER4-JAK-STAT system was modeled as a deterministic reaction network using mass action kinetic equations modeled by ordinary differential equations. The system was mod-

eled as a two-compartment system (for cytoplasm and nucleus) with NRG in the extracellular medium. The species were assumed to be in sufficiently large amounts so that the deterministic approximation applies. The model builds on the canonical JAK-STAT pathway model in the literature (74; 75). In this pathway, the receptor-JAK2 complex is activated by the ligand. This activated receptor complex, in turn, activates cytoplasmic STAT5 which then dimerizes and translocates to the nucleus. The STAT5 dimer in the nucleus initiates transcription of various genes of which the SOCS mRNA which translates to SOCS protein and exerts negative feedback on the circuit by deactivating the activated receptor-JAK2 complex is of interest. The other gene of interest for the current HC11 system is, of course, β-casein, which on transcription is transported outside the nucleus. This cytoplasmic β-casein is reported in the experiments. To explain the experimental findings which showed that the canonical pathway is inadequate to explain things like a time-dependent switch in β-casein mRNA levels we incorporated two main modifications of the canonical pathway based on recent literature findings:

a) HER4 can form both homo and heterodimers with other members of the ErbB family on activation by NRG. These reactions compete with the dimerization and activation reactions of HER4- JAK2 complex. Such competition for ligand has been shown to induce an inverse ligand dependence (signaling activity decreases with increasing ligand stimulation) in these pathways which is also observed in the experiments. These additional homodimerization and heterodimerization steps were modeled using three reactions.

b) The requirement of JAK2 for tyrosine phosphorylation of STAT5 was only demonstrated in HB-EGF and Prolactin stimulated pathways (73). There are other shreds of evidence in the literature that suggest STAT5A can directly interact with HER4 s80 when stimulated by HRG/NRG. Hence in the present work, we incorporated this possibility by introducing a JAK2 independent pathway which allows a direct interaction and activation of the s80/4ICD domain of HER4 with STAT5a. The most important effect of adding this pathway is that being JAK independent, this complex will not be negatively

regulated by SOCS to the extent when JAK is present (77) which can increase the overall β-casein gene expression in a time-dependent manner. The rate constants of the activation reactions for this JAK2 independent pathway were taken as smaller than that of the JAK2-dependent pathway since we expect the active kinase domain of JAK2 to catalyze the phosphorylation of STAT5.

The rate constants were taken from published models (74; 75; 19) for the reactions which were common to the current model. For the new reactions, we estimated the rate constants by starting with similar values as the related reactions for which the parameters are known, and then doing sensitivity analysis and comparing with experimental results. Bounds for the rate constants were also confirmed independently using values/estimates of the diffusion coefficient. The model was solved using COPASI (78). A full description of the model along with the initial expressions and rate constants are provided in the SBML `http://sbml.org/Documents/Specifications` format in Additional file 2 in (70). We used Latin Hypercube sampling (79) to sample the parameter space and to create an ensemble of models. Then for each member of the ensemble, we calculate and plot the time-integrated value of β-casein mRNA and average them over the ensemble. This task ensures that the uncertainties in the model parameters are reflected in the predictions to a certain degree and that the model is robust to small perturbations in the model parameters.

## 6.3. Results

### *6.3.1. HER4 signaling through JAK-dependent pathway*

The computational study was motivated by some intriguing results from experiments conducted on HC11 mouse mammary epithelial cells (80) stimulated with different ligands including NRG. In these studies, HC11 cells were grown and maintained at 5 % $CO_2$ medium (RPMI 1640, 10 % FBS, 2 mM L-glutamine, 100 µg/ml penicillin, 100 µg/ml streptomycin, 1 µg/ml hydrocortisone, 10 ng/ml murine EGF and 5 µg/ml insulin). Cells were seeded in 6-well plates at a density of $2 \times 10^4$ cells/cm$^2$ and allowed to grow to 100 % confluence

to induce differentiation. The cells were maintained at confluence for 2 days in serum-free/EGF-free medium to induce competence. The competent cells were then stimulated with NRG to induce differentiation. Three stages of the differentiation process were identified: Stage 1 refers to growing the cells to confluence, Stage 2 refers to maintaining the cells at confluence for 48 hours to induce a state of competence, and Stage 3 refers to cell stimulation with HER4 ligand. RT-PCR and ELISA based Transcription Factor activation assay were performed in differentiating HC11 cells to characterize the signaling dynamics in the HER4-mediated JAK/STAT pathway. Specifically, the following read-outs of HC11 mammary differentiation were assayed:

- Levels of activated STAT5A and glucocorticoid receptor (GR) in the nucleus.
- Expression of the β-casein milk gene mRNA.

STAT5 and GR are the two key transcription factors which are both activated during HC11 cell differentiation and synergize on the β-casein gene promoter (81). In the experimental studies, the steroid hydrocortisone (HC), which signals through GR was included in the medium. It was shown that in cells stimulated with NRG in the absence of HC in the medium, no β-casein was expressed. However, upon addition of HC, HER4 induced robust expression of β-casein. HC was required in Stage 2 of the differentiation process as well as Stage 3. Hence the β-casein expressions obtained by varying NRG levels were normalized by the control case of only HC present in the medium.

In tracking the response of the HC11 cells to HER4 stimulation by NRG in the presence of HC at various time points, it was observed (Fig. 9) that at early time points post-stimulation, NRG does not exhibit any effect on β-casein expression.

However, at 12 h post-stimulation, it was found that increased stimulation with NRG decreased levels of β-casein, relative to the control (HC). At 24 h to 48 h post-stimulation, NRG began to increase β-casein expression levels at sufficiently high ligand concentrations. This result has not been discovered previously, as the few studies examining the HER4-

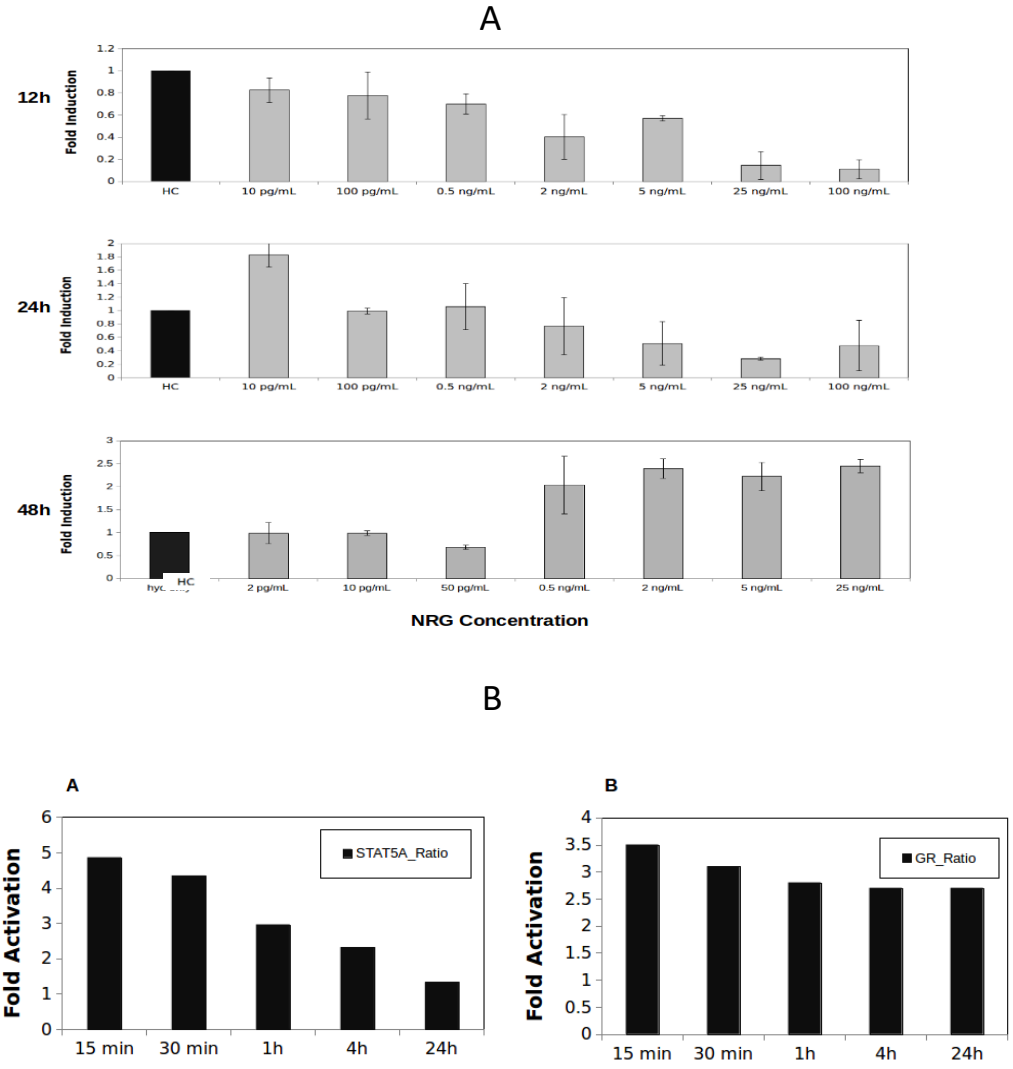


Figure 9: Experimental data from HC-11 murine mammary epithelial cell lines (reproduced with permission from (80)). (A) shows the RT-PCR results of the expression of the differentiation marker β-casein gene at three different time intervals (0-12, 12-24 and 24-48 hours) for different levels of ligand NRG stimulation. The results were compared with β-casein expression with Hydrocortisone (HC) as control. (B) shows the ELISA TF activation assays of the transcription factors STAT5 and Glucocorticoid Receptor (GR) at the nucleus at different time points post ligand stimulation.

STAT5A pathway focused mostly on gene expression at earlier time points post stimulation (81; 82). As a first step towards understanding the mechanism of this time-dependent switching behavior observed in the experiments, we used mathematical models of JAK-STAT pathways reported before in literature (74; 75). In these models, HER4 is activated in a JAK2-dependent fashion, and the HER4-JAK2 dimer activates cytoplasmic STAT5 which dimerizes and translocates to the nucleus. This model was run with increasing levels of ligand (NRG) stimulation in the same range as used in the experiments (80). Although these models were developed for shorter time ranges of 0 h to 12 h we wanted to see to what extent this model can capture the early time behavior observed in experiments. In Fig. 10a and b, we plot the time-integrated β-casein gene expression as a function of ligand NRG stimulation and instantaneous time profile of β-casein gene.

To show how model parameter uncertainties influence its predictions we sampled the parameter space using Latin Hypercube Sampling (see Methods) after setting biologically proper bounds on each parameter. This sampling was used to generate an ensemble of model parameter values which were used to run the model and obtain a set of output values. The average value of this set is reported, and the average dispersion is shown in the form of error bars. All the time-integrated plots in this chapter are are generated using this procedure. We also calculated the statistical significance of a increase/decrease in the β-casein expression for different NRG stimulations. These are indicated by horizontal black lines between two bars (for two different NRG stimulations).

The plots corresponding to the literature model show little sensitivity towards NRG in terms of β-casein expression at all three time intervals (the difference in β-casein expression between successive NRG stimulations are not statistically significant). The literature model has several negative regulators including two cytoplasmic phosphatases for JAK2 and STAT5, one nuclear phosphatase for STAT5 and negative feedback through SOCS protein deactivating HER4-JAK complex. SOCS is another product of STAT5 mediated transcription apart from β-casein. The latter is expected to limit the response and prevent any

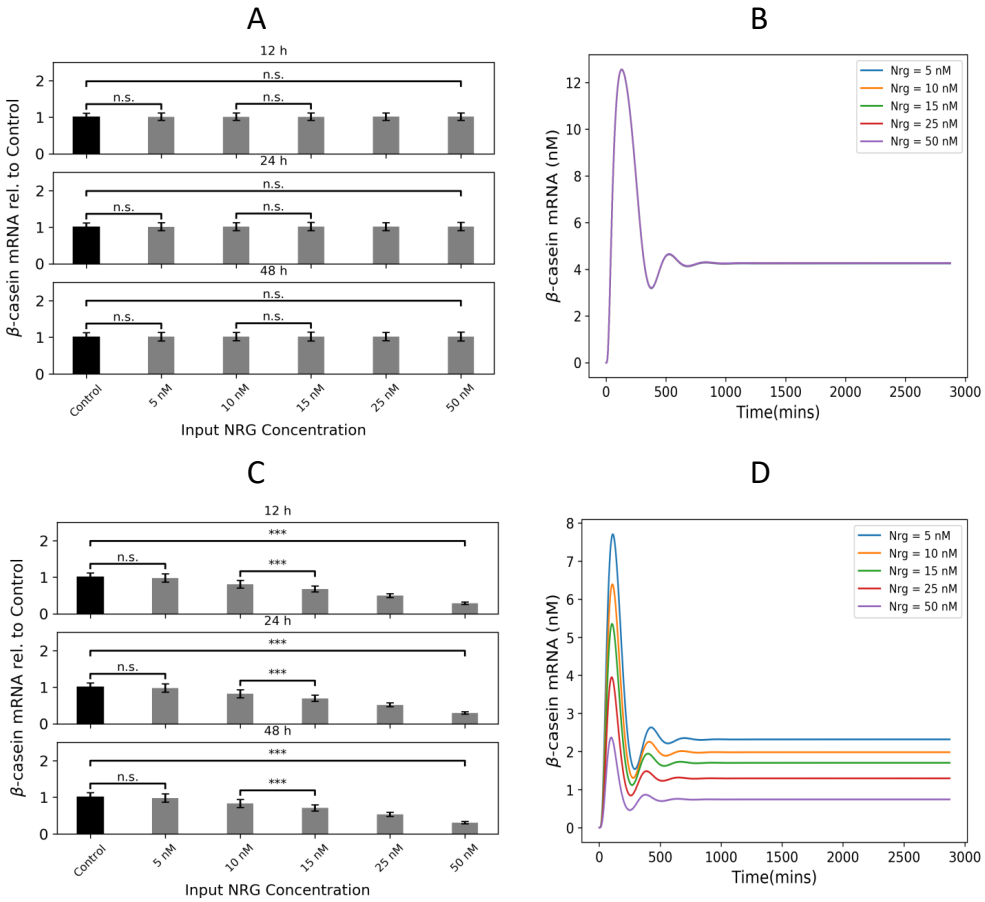


Figure 10: β-casein expressions predicted by the literature model and a modified version with competitive HER4 heterodimerization. Panels A and B show the time integrated β-casein values (obtained by integrating the solutions of the nonlinear system of ODE at different time points over different time intervals - 12,24 and 48 hours in three separate subpanels) and the instantaneous values (obtained by solving the nonlinear systems of ODEs at different time points) respectively as predicted by the literature models of JAK-STAT pathway for different levels of NRG stimulations. The black bars represent the control condition with only Hydrocortisone (which is implicitly assumed to be always present in the model). The black lines show the statistical significance of a one-tailed t-test (greater than) between different groups with significance level set to 0.05. * implies $p < .05$, each additional * indicates a p-value one order of magnitude lower. In these case no significant difference was observed for different NRG stimulations. This shows a statistically significant decrease in β-casein expressions with increasing NRG. Panels C and D show the same time integrated and instantaneous levels of β-casein when a competitive HER4 heterodimerization is included in the model. The time integrated plot contains three subpanels showing the expression at three different time points (12, 24 and 48 hours). The error bars are obtained by generating an ensemble of models by sampling over parameter space. The instantaneous plots are for 0-48 hour range at 5 different NRG stimulations.

increase in transcription due to higher ligand stimulation. However, the insensitivity of β-casein towards NRG is inconsistent with the observations from the experiments. In the 0 h to 12 h interval, the experiments show a decrease in the β-casein gene expression with increasing ligand stimulation. This observation suggested that there are additional mechanisms which are not considered in the canonical JAK-STAT pathways published earlier in the literature.

### *6.3.2. Competitive heterodimerization involving HER4 upon ligand binding*

It has been previously reported in the literature that competitive heterodimerization of HER4 with other members of ErbB family can result in a decrease in HER4 activity for a particular reaction pathway with increasing ligand stimulation (so-called partial agonists) (83; 84; 85; 86). To explore this effect, we introduced additional heterodimerization reactions of HER4 with HER3.

Introduction of such reactions resulted in a decrease in the β-casein activity with increasing NRG stimulation consistent with experimental observations. The reduction is evident from Fig. 10c and d where we have again plotted both integrated and instantaneous β-casein mRNA concentrations. Since there are several competing reactions possible for a fixed number of free HER4 receptors at the cell surface, increasing ligand stimulation promotes the more favorable (heterodimerization) reaction at the expense of the less favorable (HER4-JAK dimerization). Addition of competitive heterodimerization, however, does not increase β-casein activity in the 24 h to 48 h interval as observed in the experiments. This observation suggests additional reactions are at play during this late period contributing to an increase in the β-casein gene expression.

### *6.3.3. Combined model with JAK-dependent and JAK-independent HER4 signaling*

It has been reported before in the literature (87; 77) that HER4 can activate STAT5 independent of JAK2 although with lower rates compared to the canonical JAK-dependent pathway. We hypothesized that such a slower JAK-independent pathway might likely become active

in the later stages (24 h to 48 h) where it can contribute to an increase in β-casein expression with increasing NRG stimulation after the faster and more favorable heterodimerization reactions have equilibrated. Also, this pathway is not affected by negative feedback through SOCS, which can only act on the JAK-bound complex. We incorporated such a pathway in our model to explore late time behavior (see the method section for more details). The combined pathway is shown in the schematic diagram in Fig. 11 where the JAK-dependent and JAK-independent pathways are highlighted.

When this combined pathway was stimulated with increasing NRG stimulations (plotted in Fig. 12), we were able to reproduce the time-dependent switching behavior seen in the experiments: that is, in the 0 h to 24 h period the β-casein activity decreases with ligand stimulation, and it increases in the later stages (24 h to 48 h).

From the β-casein mRNA time profiles, we see that after the 24 h mark there is a transition from negative to positive dependence of β-casein mRNA towards ligand stimulation. In this regime, the β-casein levels continue to increase.

*6.3.4. Time delay in β-casein mRNA transcription*

In the HC11 cell line experiments mentioned above additional ELISA-based transcription factor (TF) activation assays were also performed to measure the activity of STAT5A and Glucocorticoid receptor (GR) which are the two transcription factors necessary for transcription of β-casein mRNA. It was observed that even though STAT5A and GR activity was highest in the nucleus 15 min to 30 min post ligand stimulation, a significant activity of these transcription factors persisted even 24 h post stimulation. These findings were consistent with previous ChIP-Seq studies (88; 89) which assayed for binding of various TFs (including STAT5A and GR) to the β-casein gene promoter at different time points following stimulation with prolactin (PRL). These studies showed that although several of the TFs assemble on the promoter at early time points, the RNA polymerase does not bind and commence transcription until 24 h post-stimulation. To incorporate this delay in

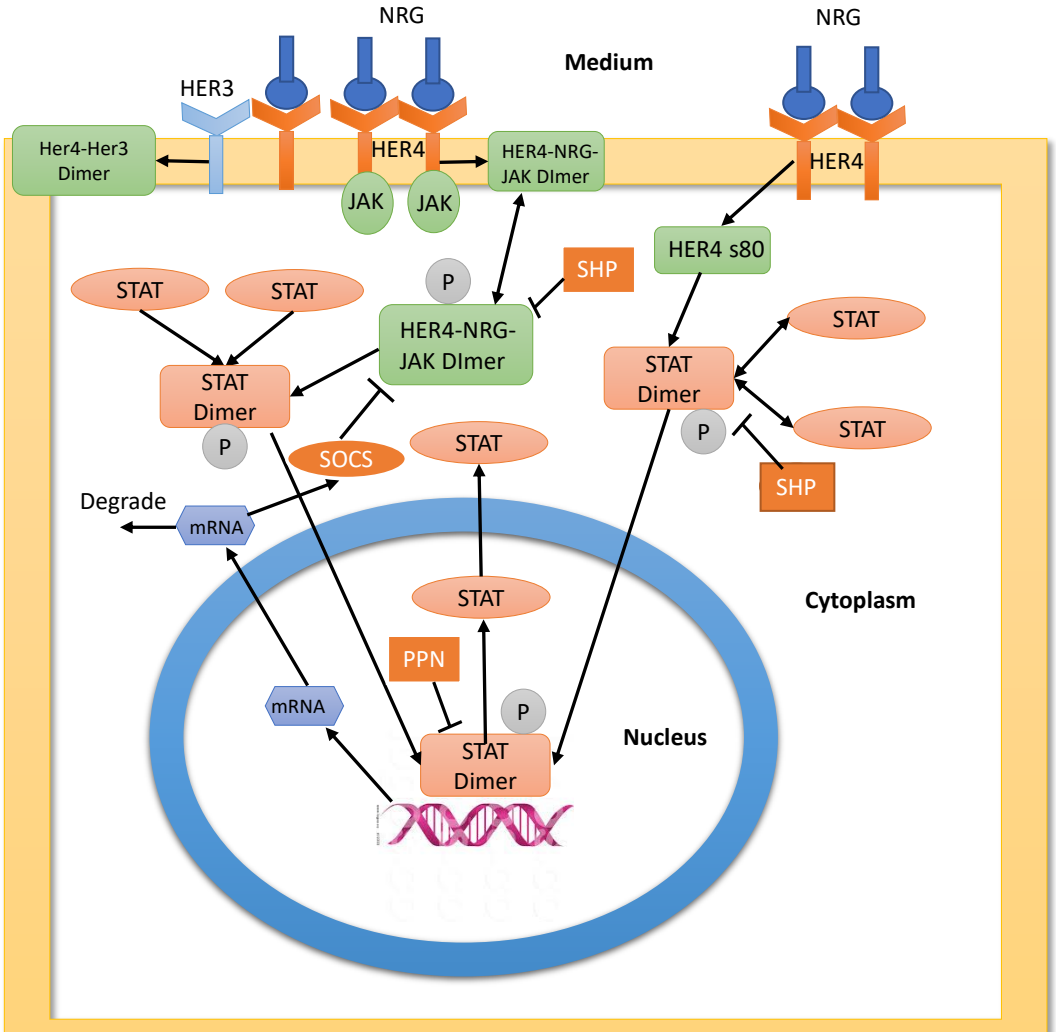


Figure 11: Schematic diagram of the HER4-JAK2-STAT5 pathway modeled in this paper. This is based on the canonical HER4-JAK-STAT pathway with two additional modules proposed in this work. First one involves competitive heterodimerization of HER4 with HER3 at the plasma membrane. The other module is the JAK-independent activation of STAT5 by HER4. We introduced cytoplasmic and nuclear compartments and transport reactions between the compartments as indicated. The heterodimerization reactions are indicated with a side branch from HER4 and not shown explicitly. More details are available in the Methods section.

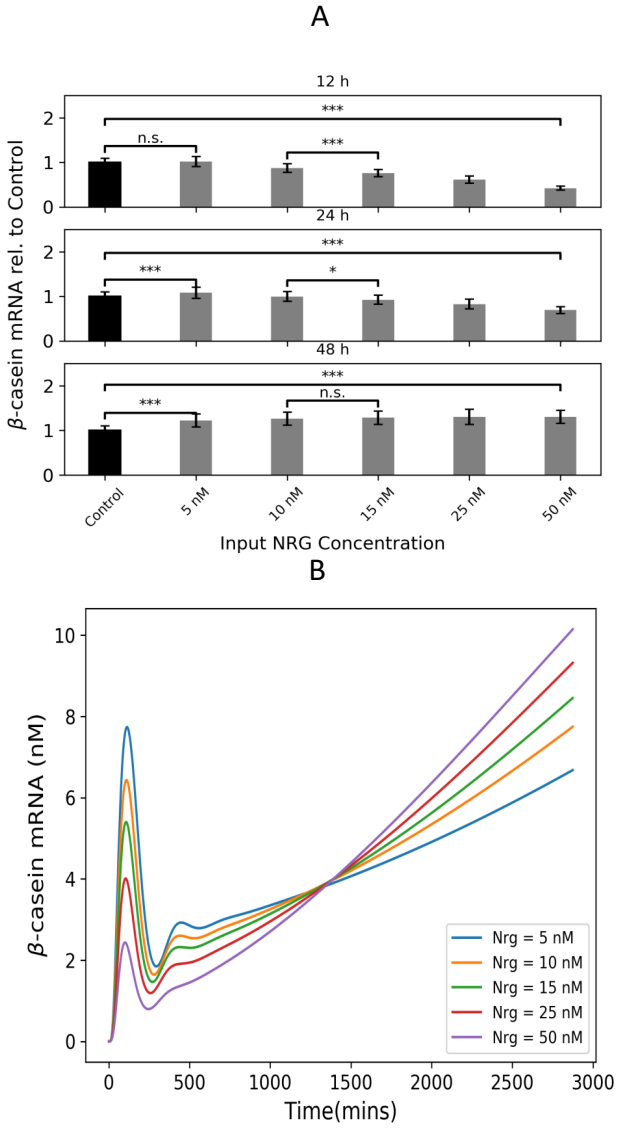


Figure 12: β-casein expressions for model with competitive heterodimerization and parallel JAK-independent activation of STAT5. Panels A and B show the time integrated and instantaneous levels of β-casein for this model. The time integrated plot contains three subpanels showing the expression at three different time points (12, 24 and 48 hours). The black bars represent the control condition with only Hydrocortisone (which is implicitly assumed to be always present in the model). The error bars are obtained by generating an ensemble of models by sampling over parameter space. The instantaneous plots are for 0-48 hour range at 5 different NRG stimulations. The black lines show the statistical significance of a one-tailed t-test (greater than) between different groups with significance level set to 0.05. * implies $p < .05$, each additional * indicates a p-value one order of magnitude lower. This shows a statistically significant decrease in β-casein expressions with increasing NRG in the 0-12 hour and 12-24 hour intervals but an increase in β-casein expression in the 24-48 hour interval.

the transcription and detection of β-casein mRNA, we introduced two significant transport rates in our model – nuclear import of activated STAT5 dimer and nuclear export of β-casein mRNA. The plots of instantaneous β-casein mRNA and activated STAT5 dimer at the nucleus between 0 and 12 h time points for two different mRNA transport rates show that with low rates of mRNA export we get a significant delay in the mRNA peak activity.

These results are consistent with previous experimental and modeling studies (74; 77) which identified these transport rates as essential determinants of the signaling activity of this pathway. Also, such delay in transcription did not affect the time-dependent switching behavior (Fig. 13e and f) in the β-casein expression which occurs over a much longer time scale. These results complement the findings from the Global Sensitivity Analysis (next section).

*6.3.5. Global sensitivity analysis*

To systematically explore the effect of model parameters on various output quantities of interest (such as β-casein expression, time-dependent switch in β-casein expression, β-casein transcription time delay, etc.), we performed sensitivity analysis. The HER4-JAK-STAT signaling network is a highly nonlinear system, and most of the parameters can vary over a wide range in the corresponding biological system. Hence, a Global Sensitivity Analysis (GSA) which considers the combined effect of the model parameters rather than one at a time is a more appropriate method here (25). We use a particular type of global sensitivity analysis called Sobol Sensitivity Analysis (26) method for this study (details in the Methods section). The analysis procedure involves:

- Sampling over the space of model parameters to create an ensemble of models.
- Running the simulations for each member of the ensemble.
- Calculating the output quantity of interest from the results for each member of the ensemble.

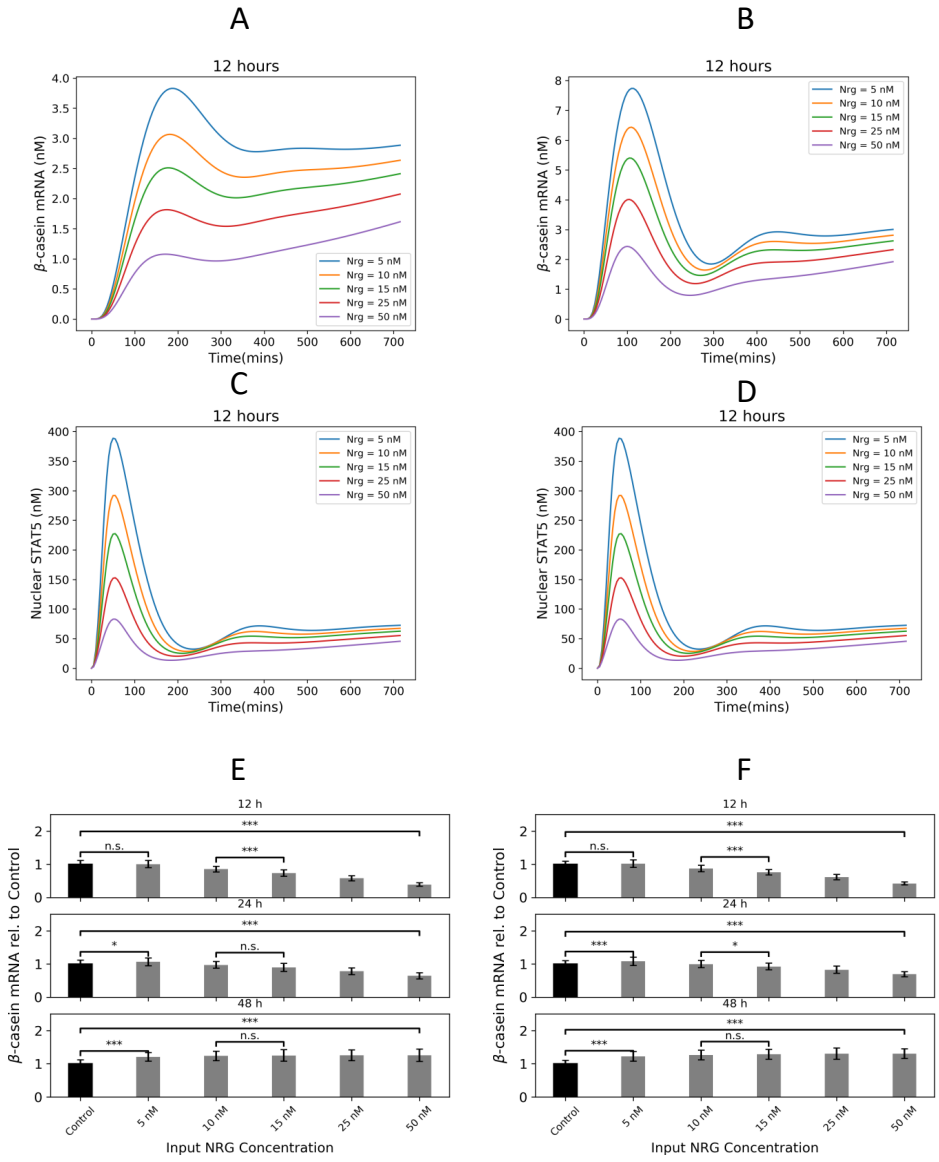


Figure 13: Effect of the mRNA transport rates on the transcription time delay. Figure 5A and 5B shows the β-casein mRNA expression time profiles at low and baseline values of the rate of mRNA export. Figures 5C and 5D shows the corresponding nuclear STAT5 levels. By decreasing the transport rate 10 fold we are able to see almost 3 hours delay between STAT5 and β-casein mRNA peaks, showing that these rates contribute to a large extent towards the observed time delay in experiments. Figures 5E and 5F shows the integrated β-casein levels at the three time points which shows that the temporal switch is preserved at low mRNA transport rates. The error bars are obtained by generating an ensemble of models by sampling over parameter space. Again the horizontal lines indicate a statistical significance of a one-tailed t-test (greater than) between successive NRG stimulations. $*p < 0.05$

- Determining the effect of the model parameter variation on the variation in the output quantity using Sobol Sensitivity Analysis.

The input parameters here are the initial concentrations of all the species and the kinetic rate constants of the reactions. Three main output quantities were calculated from the results obtained by running the ensemble of models, and the Sobol sensitivity coefficients were calculated for each of these three quantities. These are:

**Integrated β-casein levels**

β-casein mRNA levels integrated over the whole time interval (0 h to 48 h) is a direct output of the model which we have used to present the results in the previous section. This quantity is an automatic choice for the sensitivity analysis.

**Time-dependent switch in the β-casein expression**

We consider the relative contributions to the time-integrated β-casein levels of the JAK-dependent and the JAK-independent parts of the pathway. Since we know from the above results that the JAK-dependent pathway is operational in the earlier part (0 h to 12 h) while the JAK-independent pathway in the later part (24 h to 48 h) we can calculate the integrated mRNA levels separately for these time intervals and calculate their ratio. When this ratio is near unity, it indicates equal contributions of both these pathways to the net β-casein expression. On the other hand, a significant value of this ratio will indicate that the main contribution comes from JAK-dependent pathway and the model does not show a time-dependent switch. A similar interpretation can be made when the ratio is small.

**Transcription time delay**

We get the delay between the times when activated STAT5 dimer reaches the nucleus and when β-casein mRNA transcription starts by getting the difference between the peak times of nuclear STAT5 dimer and cytoplasmic β-casein mRNA. The previous results also showed

that the dynamical behavior of the combined pathway is much dependent on the level of ligand NRG stimulation. So, we decided to conduct the global sensitivity analysis at three different NRG stimulation – low (10 nM), medium (20 nM) and high (50 nM).

### *6.3.6. Time-integrated β-casein levels*

We performed the global sensitivity analysis concerning the total integrated β-casein mRNA levels in the cytoplasm. The results are presented separately for the initial amounts of species and the kinetic rate constants (Fig. 14a and b).

The plots show both the first order ($S_1$) and the total effect sensitivities ($S_T$) of the top 10 most sensitive species and parameters of the model. Table **??** below (left column) shows the names of the parameters and the reactions they belong to. From the results, we see the β-casein mRNA expression is sensitive to initial numbers of Her4 and Her2 receptors. It is also sensitive concerning the phosphatases SHP and PPX and JAK which are expected. As for the reactions the sensitivity is high for the transcription parameter (including the Hill coefficient), suggesting that the β-casein expression is sensitive to the mechanism of the transcription reaction. We want to find if the same holds when we consider our main output quantities of interests – time-dependent switch and transcription delay. Among other reaction parameters, we see that the rates of JAK-independent activation of STAT also features among the top sensitive parameters. The sensitivity plots also show that the total effect sensitivity $S_T \approx S_1$ here for all the sensitive parameters which suggests that there is not much higher order effect of parameters on the absolute β-casein levels.

### *6.3.7. Time-dependent switch in β-casein expression*

The next output parameter we considered is the ratio of β-casein mRNA expression during the JAK-dependent pathway (0 h to 12 h) to that of JAK-independent pathway (12 h to 48 h). As explained previously, this quantity can be an indicator of whether and to what extent there is a time-dependent switch in β-casein mRNA expression. This output quantity was found (Fig. 14c and d) to be particularly sensitive to HER4 heterodimerization and

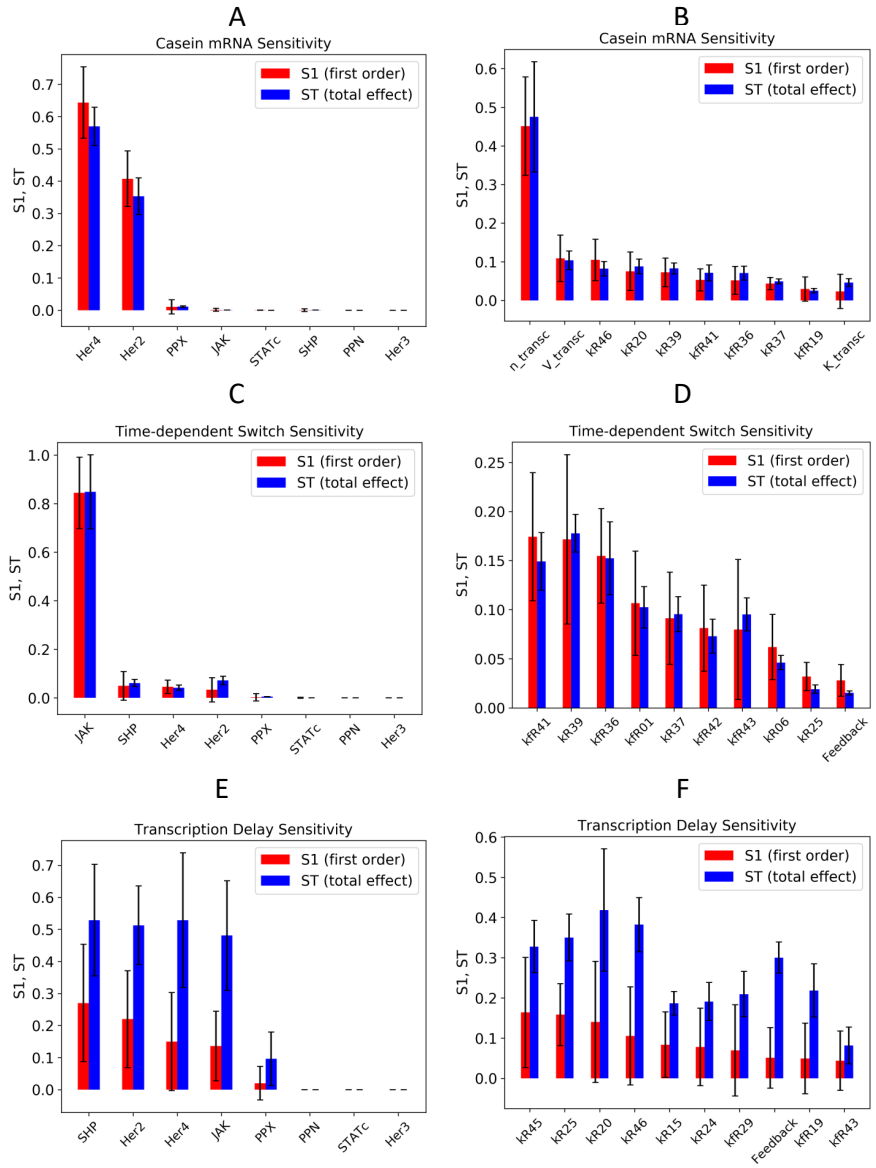


Figure 14: First order (S1) and total effect (ST) Sobol Sensitivity Coefficients for the most sensitive species and kinetic rate constants for different model outputs. See Methods section for the definition of these sensitivity coefficients. A and B shows the sensitivities with respect to the species and parameters for absolute β-casein levels as output. C and D shows sensitivities with respect to ratio of contributions to β-casein transcription by the JAK-dependent and JAK-independent parts of the pathway. E and F shows the sensitivities with respect to the time delay between STAT5 import into nucleus and β-casein transcription. A sensitivity coefficient $ST >> S1$ indicates significant higher order contributions to the sensitivity.

Table 1: Table summarizes the top 10 most sensitive system parameters for the global sensitivity analysis concerning integrated β-casein (left), transcription time delay (middle) and time-dependent switch (right). For each output parameter, some of the hypothesized mechanisms are highlighted. For example, for transcription time delay the mRNA and STAT transport rates are in bold.

| **Casein** | | **Delay** | | **Ratio** | |
|---|---|---|---|---|---|
| Parameter | Name | Parameter | Name | Parameter | Name |
| $n_{transc}$ | Transcription Hill Coefficient | $k_{R45}$ | mRNA Nuclear Export Casein | $k_{fR41}$ | HER4 Heterodimerization |
| $V_{trasnc}$ | Transcription Half-Maximum | $k_{R25}$ | Translation | $k_{R39}$ | STATc Phosphorylation |
| $k_{R46}$ | mRNA degradation rate | $k_{R20}$ | Nuclear STAT dimer dephosphorylation | $k_{fR36}$ | IFNR(s80) formation |
| $k_{R20}$ | Nuclear STAT dimer dephosphorylation | $k_{R46}$ | mRNA degradation rate | $k_{fR01}$ | HER4-JAK Binding |
| $k_{fR41}$ | HER4 Heterodimerization | $k_{R24}$ | mRNA Nuclear Export | $k_{fR42}$ | HER4 Heterodimerization Constitutive |
| $k_{fR36}$ | IFNR (HER4 s80) Formation | $k_{fR29}$ | Binding of STATc-SOCS complex | $k_{fR43}$ | HER4 Heterodimer Ligand Binding |
| $k_{R37}$ | Nuclear STAT-dimer PPN Binding | $k_{fR19}$ | Nuclear STAT dimer PPN binding | $k_{R25}$ | Translation |

homodimerization reactions (both ligand-dependent and constitutive) as well as the JAK-independent HER4 activation. The results confirm our hypothesis that both these additions to the literature JAK-dependent pathway are sufficient for producing a time-dependent switch. Interestingly the transcription reaction parameters which influenced the absolute time integrated mRNA levels to a high degree (from the previous section) did not appear in the list of top parameters for the time-dependent switch. These insights highlight the importance of selecting a proper output quantity of interest in Global Sensitivity Analysis as the results can vary depending on the choice of the output parameter. Among the species we see HER4, and HER2 numbers are again significant, along with JAK and cytoplasmic JAK phosphatase SHP2. Here also we see that the total effect sensitivity $S_T \approx S_1$ here for all the sensitive parameters which suggests that there is not much higher order effect of parameters on the time-dependent switch.

*6.3.8. Delay in STAT nuclear translocation and β-casein transcription*

The next output parameter of interest is the time delay between STAT5 nuclear translocation and β-casein transcription and transport outside the nucleus. Here we make several interesting observations (Fig. 14e and f ). The various transport rates in the model especially the rate of activated STAT5 dimer nuclear import and rate of β-casein mRNA nuclear export were prominent in the list of sensitive parameters. The effect of these transport rates on the pathway activity has been reported before in literature (74; 90) and is expected from the model. The SOCS mediated negative feedback rate also features on the sensitive parameters list. The negative feedback does not directly influence the β-casein mRNA levels but has a more indirect effect. The result shows that in a nonlinear system like HER4-JAK-STAT pathway parallel/side reactions can influence the dynamic behavior and can be exploited as therapeutic targets. In contrast to the absolute β-casein sensitivity we see here that $S_T >> S_1$ which suggests that there is a high degree of nonlinear effect of the sensitive parameters to the overall sensitivity of transcription time delay.

## 6.4. Discussion

Epidermal growth factor family of receptors plays a crucial role in different cancers by activating critical signaling pathways that control cell fate decisions. HER4, a member of this family has received attention due to its several distinctive properties. Some isoforms of this receptor can undergo enzyme-mediated cleavage reactions resulting in shedding of its ectodomain and cytoplasmic domain leaving an 80 kDa transmembrane domain which can translocate to the nucleus and promote transcription of various genes. On activation by ligand NRG, it can activate the anti-proliferative JAK2-STAT5A pathway which results in activation, dimerization and nuclear localization of the transcription factor STAT5A which promotes transcription of genes mediating differentiation in particular breast cancer cell lines. This study was motivated by experimental work conducted on mouse HC11 mammary epithelial cell lines stimulated by various ligands including NRG using RT-PCR to measure the expression of the important differentiation marker gene β-casein. One of the intriguing

results from the study was that the response of these cells to the NRG stimulation (agonistic versus antagonistic) is time-dependent. It was observed that whereas NRG suppressed the transcription of β-casein at early time points (0 h to 12 h), at later time points (24 h to 48 h), it promoted transcription. Using ELISA based TF activation assays of STAT5A and GR two transcription factors required for transcription of β-casein, the same study also found that the activity of these transcription factors persisted even 24 h after NRG stimulation. These observations mirrored similar findings in the literature from ChIP-seq studies that showed there is a significant time delay between the STAT5A entry to the nucleus and transcription of the β-casein gene. To obtain a mechanistic understanding of these observations we turned to mathematical modeling. Although the JAK-STAT pathway has been extensively modeled before, the literature models failed to reproduce these experimental findings particularly the time-dependent switch in β-casein gene transcription. The failure of the existing models led us to develop a new model for the HER4-JAK2-STAT5 system that retains the core reactions of the literature models but adds two essential components which have been reported before in the literature but have not been included in the models. These are:

- Competitive binding and heterodimerization of HER4 receptor with other members of the ErbB family like HER3.
- A HER4 mediated JAK-independent activation of STAT5 which proceeds at a lower rate than the canonical JAK-dependent activation.

Including these reactions in our model, we were able to reproduce the experimental findings. By systematically turning these reactions on and off we showed how the signaling dynamics shifted to the pattern seen in the experiments at late time points. We performed extensive parameter sweep studies to understand how different individual components of the model influenced the signaling dynamics. We also conducted global sensitivity analysis to test the robustness of model predictions and to obtain the most sensitive parameters of the model for different output dynamical quantities. These studies show that the competitive HER4 heterodimerization reactions have a profound impact on the sensitivity of the pathway to

NRG stimulation at earlier time points. This reaction is a necessary condition for the observed suppression of transcription by NRG.

Along with this competitive heterodimerization, the addition of a slower JAK-independent mechanism of activation of HER4 was sufficient to reproduce the time-dependent switch in the transcription of the β-casein gene observed in the experiments. We also found that the various transport rates in the model such as STAT5 dimer nuclear import and β-casein mRNA export influences the time delays associated with transcription. The Global sensitivity analysis results confirm these model findings by showing that the parameters of the above reactions are most sensitive to the corresponding model output such as delay. This study highlights the effect of competitive and parallel reaction pathways on both short and long-term dynamics of receptor-mediated signaling. Such time-dependent alterations in the signaling behavior of these pathways are highly consequential in cancer and are one of the main ways tumor cells develop resistance to targeted inhibitors. Identification of such time-dependent changes in signaling dynamics can also help in designing optimal treatment strategies with different dose interval and durations (91; 9). By obtaining a deeper understanding of the dynamics of such pathways, we will be able to design more efficient drug dosing regimens that can target and exploit the differential dynamics. However, because of the uncertainties inherent in these mechanistic models, there may be alternative mechanisms that might explain the observed data. In those cases, a data-driven clustering based approach [25, 26] which has received attention recently can be used to perform system identification and map dynamical behavior the observed data.

## 6.5. Appendix: A model for β-casein transcription

Our simple ODE model of HER4 mediated JAK/STAT signalling pathway successfully reproduced many key experimental observations for HC11 mammary epithelial cells including a temporal switching of β-casein gene expression.

However the model did not show the significant delay observed between the time activated

STAT5 dimer translocated to the nucleus and the time when significant expression of β-casein gene was observed. The ODE model uses a simple Hill type kinetics used in the previous attempts at modeling the JAK-STAT pathway (75) Here the transcription rate is a Hill function of STAT5 concentration with a Hill coefficient of unity. Recent ChIP studies have shown (89) that β-casein transcription mechanism is very elaborate and requires the assembly of a complex transcriptional machinery and structural modification of DNA like an open chromatin structure through histone acetylation and DNA looping. Here we perform some simple calculations based on ideas from equilibrium statistical mechanics to estimate the probability of such assembly process. This is by no means meant to serve as an accurate model for such transcription processes, rather to show that such elaborate assembly processes can account for the delays observed in experiments.

We use a simplified version of the β-casein transcription model presented in (89) to calculate the probability of the events leading to transcription. As described in (89) the STAT5A and GR mediated transcription of β-casein gene requires binding of three transcription factors — phosphorylated dimer of STAT5A, GR and C/EBPβ. Two of these binding sites (one each for STAT5A and GR) are located in β-casein promoter region and the other one (STAT5A) in the enhancer region which is approximately 6 kb downstream of the promoter site.

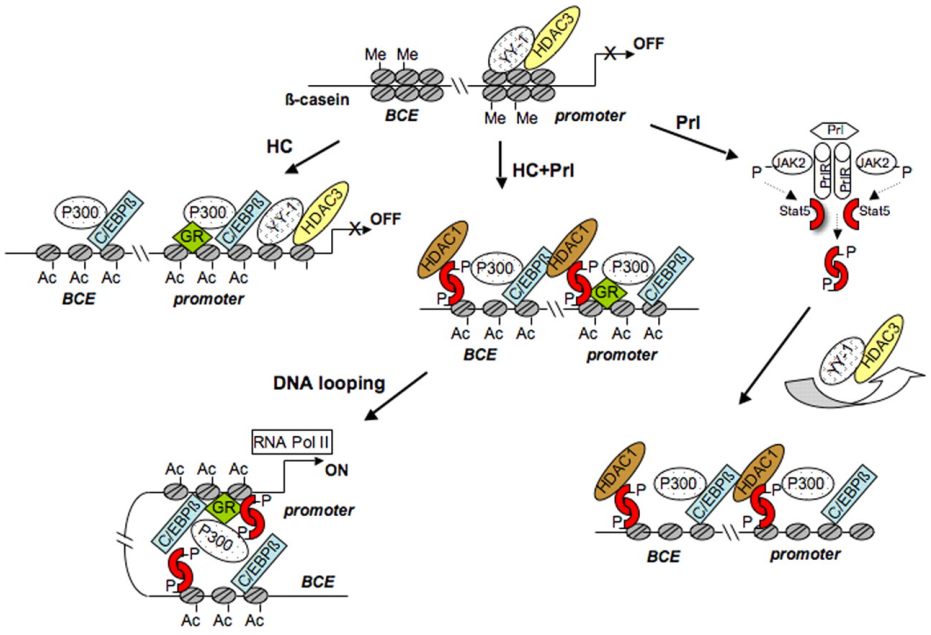


Figure 15: A model to explain the sequential formation of complexes leading to the activation of β-casein gene expression following hormonal stimulation (89)

To build the simplest possible transcription model we only consider binding of STAT5A and GR as the necessary event for transcription to begin. In order for these transcription factors to bind to their respective binding sites, these sites must first be released by unwinding of the DNA from the histone octamer core which has a free energy cost that we can calculate. After the occupation of all three binding sites the transcription factors recruits p300 which facilitates histone acetylation and modifies the chromatin organization. It eventually leads to formation of STAT5A tetramer through looping of the intervening DNA segment and the subsequent assembly of the transcription complex which then recruits RNA-Polymerase II and this initiates transcription. Some key experimental findings from (89) which are of relevance to the current model are:

- Both transcription factors are seen to accumulate rapidly near their respective DNA binding sites, within 15 min of ligand stimulation
- For cells treated with both PRL and HC (which activates GR) the gene expression

was appreciable within 3-4 hours and was very high by about 24 hours

- The assembled machinery recruits RNA pol II and it was observed that the accumulation of pol II occurred in both promoter and enhancer region within about 4 hours post stimulation. This suggests that DNA looping is rate limiting and this is what delays the transcription start event from 15 mins to about 4 hours.

To estimate the transcription rate, we first calculate the probability of event(s) leading to RNA-Polymerase II recruitment — binding of both transcription factors and looping of intervening DNA. This event can be broken into several sub-events as below:

- Unwinding of DNA from the histone octamer core to release the three binding sites - two at the promoter region and one at enhancer.
- Binding of the three vacant sites by the TFs STAT5A and GR from the adjacent medium
- Recruitment of p300 and subsequent looping of the DNA to assemble the transcription complex (Modeled as a single event)

We also assume that PRL is present in the medium at sufficient amount necessary to initiate transcription even though it is not explicitly included in the model.

To be able to write down the probability of the transcription assembly process using equilibrium statistical mechanics we first need to write down the partition function. Choosing a grand canonical ensemble, we write the expression for the partition function which is a sum over all possible microstates. For this model the various states are listed below with corresponding mathematical expressions.

Table 2: Table of various biophysical parameters in transcription model

| **Name** | **Formula** |
|---|---|
| Unbound | $\frac{e^{\beta N \Delta\gamma a}}{1-e^{-\beta\Delta\gamma a}}$ |
| STAT1 binding | $\frac{C_{STAT}}{C_{STAT0}} e^{\beta(\epsilon_{STAT}-\Delta\gamma ax)}$ |
| GR binding | $\frac{C_{GR}}{C_{GR0}} e^{\beta(\epsilon_{GR}-\Delta\gamma ay)}$ |
| STAT2 binding | $\frac{C_{STAT}}{C_{STAT0}} e^{\beta(\epsilon_{STAT}-\Delta\gamma ax)}$ |
| STAT1-GR binding | $\frac{C_{STAT}C_{GR}}{C_{STAT0}C_{GR0}} e^{\beta(\epsilon_{STAT}+\epsilon_{GR}-\Delta\gamma ay)}$ |
| STAT1-STAT2 binding | $\frac{C_{STAT}^2}{C_{STAT0}^2} e^{\beta(2\epsilon_{STAT}-2\Delta\gamma ax)}$ |
| GR-STAT2 binding | $\frac{C_{STAT}C_{GR}}{C_{STAT0}C_{GR0}} e^{\beta(\epsilon_{STAT}+\epsilon_{GR}-\Delta\gamma ax-\Delta\gamma ay)}$ |
| TF binding | $\frac{C_{STAT}^2C_{GR}}{C_{STAT0}^2C_{GR0}} e^{\beta(2\epsilon_{STAT}+\epsilon_{GR}-2\Delta\gamma ax-\Delta\gamma ay)}$ |
| TF binding and Looping | $\frac{C_{STAT}^2C_{GR}}{C_{STAT0}^2C_{GR0}} e^{\beta(2\epsilon_{STAT}+\epsilon_{GR}-\Delta F_{loop}-\Delta\gamma ax-\Delta\gamma ay)}$ |

Summing up the above terms the total partition function can be written as:

$$\begin{aligned} Z = Z(0) &+ \frac{C_{STAT}}{C_{STAT0}} e^{\beta\epsilon_{STAT}} Z(x) + \frac{C_{GR}}{C_{GR0}} e^{\beta\epsilon_{GR}} Z(y) \\ &+ \frac{C_{STAT}}{C_{STAT0}} e^{\beta\epsilon_{STAT}} Z(x) + \frac{C_{STAT}C_{GR}}{C_{STAT0}C_{GR0}} e^{\beta(\epsilon_{STAT}+\epsilon_{GR})} Z(y) \\ &+ \frac{C_{STAT}^2}{C_{STAT0}^2} e^{2\beta\epsilon_{STAT}} Z(x)^2 + \frac{C_{STAT}C_{GR}}{C_{STAT0}C_{GR0}} e^{\beta(\epsilon_{STAT}+\epsilon_{GR})} Z(x)Z(y) \\ &+ \frac{C_{STAT}^2C_{GR}}{C_{STAT0}^2C_{GR0}} e^{\beta(2\epsilon_{STAT}+\epsilon_{GR})}(1+e^{-\beta\Delta F_{loop}})Z(x)Z(y) \end{aligned} \tag{6.1}$$

$C_{STAT}$ and $C_{GR}$ are the nuclear concentration of STAT5A and GR. The other parameters are described in Table 3.

In the above equation $Z(x)$ is the partition function of DNA unwrapping where we have x contacts broken. This is defined as below:

$$Z(x) = \sum_{j=x}^{N} e^{\beta(N-j)\Delta\gamma a} = e^{\beta(N-x)\Delta\gamma a} \frac{1 - e^{-\beta(N-x+1)\Delta\gamma a}}{1 - e^{-\beta\Delta\gamma a}} \tag{6.2}$$

Separating the x independent part we get

$$Z(x) = \frac{e^{\beta N\Delta\gamma a}}{1 - e^{-\beta\Delta\gamma a}} (e^{-\beta x\Delta\gamma a} - e^{-\beta(N+1)\Delta\gamma a}) \tag{6.3}$$

If N is large compared to x the second term within parentheses can be neglected compared to the first term. We can also take out the x independent term which is common to all the terms of the partition factor and hence will eventually get cancelled in the expression of probability.

$$Z(x) \propto e^{-\beta x\Delta\gamma a} \tag{6.4}$$

The partition function thus can be simplified to:

$$\begin{aligned} Z \propto 1 &+ \frac{C_{STAT}}{C_{STAT0}} e^{\beta(\epsilon_{STAT} - \Delta\gamma a x)} + \frac{C_{GR}}{C_{GR0}} e^{\beta(\epsilon_{GR} - \Delta\gamma a y)} + \frac{C_{STAT}}{C_{STAT0}} e^{\beta(\epsilon_{STAT} - \Delta\gamma a x)} \\ &+ \frac{C_{STAT} C_{GR}}{C_{STAT0} C_{GR0}} e^{\beta(\epsilon_{STAT} + \epsilon_{GR} - \Delta\gamma a y)} + \frac{C_{STAT}^2}{C_{STAT0}^2} e^{\beta(2\epsilon_{STAT} - 2\Delta\gamma a x)} \\ &+ \frac{C_{STAT} C_{GR}}{C_{STAT0} C_{GR0}} e^{\beta(\epsilon_{STAT} + \epsilon_{GR} - \Delta\gamma a x - \Delta\gamma a y)} \\ &+ \frac{C_{STAT}^2 C_{GR}}{C_{STAT0}^2 C_{GR0}} e^{\beta(2\epsilon_{STAT} + \epsilon_{GR} - 2\Delta\gamma a x - \Delta\gamma a y)} (1 + e^{-\beta\Delta F_{loop}}) \end{aligned} \tag{6.5}$$

We want to evaluate the probability of the last event when all TF are bound and the intervening DNA segment has formed a loop. This probability is given by:

$$p_{trans} = \frac{\frac{C_{STAT}^2 C_{GR}}{C_{STAT0}^2 C_{GR0}} e^{\beta(2\epsilon_{STAT}+\epsilon_{GR}+\Delta F_{loop}-\Delta\gamma az)}}{Z} \tag{6.6}$$

*Estimation of free energy of looping of unwrapped DNA*

The DNA looping free energy can also be estimated by assuming a simple circular loop model (92). We assume that the DNA is completely unbound from all the intervening histone octamers. The energy required the bend the DNA into a circular loop is given by:

$$\Delta E_{loop} = \frac{2\pi^2}{N_{bp}}(\frac{\xi_p}{a}) \tag{6.7}$$

In the above equation $N_{bp}$ is the number of base pairs along the circumference of the loop and $\xi_p$ is the persistence length of DNA. Assuming a semi-stiff DNA take the persistence length as approximately 50 nm (93) and the length per base pair $a$ is approximately 0.34 nm. Using these values the energy of looping can be rewritten as:

$$\Delta E_{loop} = \frac{3000}{N_{bp}} \tag{6.8}$$

The entropic contribution to the total free energy can be estimated by considering a simple random walk model of the polymer where the probability of loop formation $p_0 \propto N_{bp}^{\frac{3}{2}}$. This gives us the entropy as:

$$\Delta S_{loop} = -k_B \frac{3}{2} ln(N_{bp}) \tag{6.9}$$

The free energy $\Delta F$ as we know is a sum of the energy and entropic contribution:

$$\Delta F = \Delta E - T\Delta S \tag{6.10}$$

For this system the number of base pairs $N_{bp}$ between the promoter and enhancer region is approximately 6 kb which is what we use to estimate the looping free energy. Using $N_{bp} = 6000$ the free energy of looping is calculated to be about 13.55 $k_BT$.

*Estimation of free energy of looping of partially wrapped DNA*

Now we consider the DNA architecture where in order to form a loop we only need a partial unwinding of DNA, limited to the particular histone octamer at respective promoter and enhancer site.

Again in the spirit of keeping things uncomplicated we consider a circular loop of DNA with N histone octamer cores arranged in the form of a regular polygon and the intervening linkers bent in the form of a circular arc (see 16). The free energy of looping for such a structure can be calculated by considering the bending energy for each individual linker which is in the form of a circular arc. The histone octamer is assumed to be rigid. Because the length of a linker is typically small (on the order of about 50 bp) we can neglect the entropic contribution to the free energy of bending.

Each linker is a circular arc which subtends an angle given by $2 * Pi/N_oct$ where $N_oct$ is the number of intervening histone octamers (and also linkers). If $N_bp$ is the total number of base pairs between the the promoter and enhancer site and d is the diameter of each histone octamer then the length of each linker segment is given by:

$$L_{link} = \frac{N_{bp}a - \pi N_{oct}d}{N_{oct}} \tag{6.11}$$

The bending energy of a circular arc of radius R and angle $\frac{2\pi}{N}$ at the center with persistence length $\psi_p$ is given by

$$E_{bend} = \frac{\xi_p \pi}{R} \frac{2\pi}{N_{oct}} \tag{6.12}$$

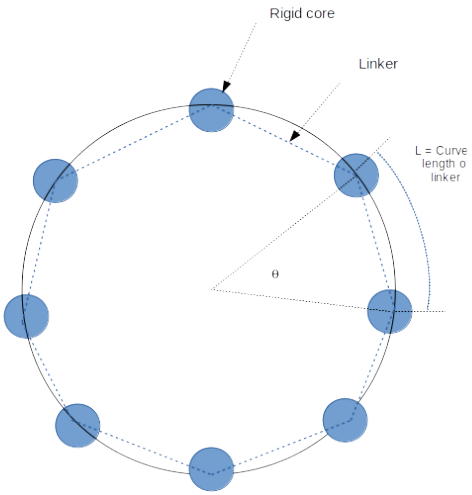


Figure 16: Cartoon model of DNA looping

The radius of curvature is given by

$$R = \frac{L_{link}}{\frac{2\pi}{N_{oct}}} \tag{6.13}$$

Substituting the expression for $L_{link}$ and substituting R in the energy expression we get

$$E_{bend} = \frac{\xi_p \pi}{\frac{N_{bp} a - \pi N_{oct} d}{N_{oct}}} \frac{2\pi}{N_{oct}} \tag{6.14}$$

This is simplified to:

$$E_{bend} = \frac{\xi_p \pi}{N_{bp} a - \pi N_{oct} d} \frac{4\pi^2}{N_{oct}} \tag{6.15}$$

If we know the number of histone octamers between the promoter and the enhancer $N_{bp}$ and the diameter $d$ of a typical octamer core then we can calculate the energy of bending using the expression above. The values for length per base pair $a$ and persistence length $\xi_p$ can be same as before. Using typical numbers of 147 bp per histone and 50 bp per linker (92) the number of octamers can be estimated as below. The diameter of each octamer is roughly 7 nm. Using these estimates the bending energy can be calculated, we simultaneously multiply the energy by the number of octamers which gets canceled out. This gives us a binding free energy of about 5.44 $k_BT$ which is substantially less compared to the free energy of looping of the unwrapped DNA.

*Estimation of transcription rate and time*

Using the value of probability of transcription factor binding $p_{trans}$ we can calculate the rate of transcription and the time necessary for transcription as below. In the HER4 network model we assume a Hill-type dependence of transcription rate on the STAT5A concentration as in (6.16)

$$r_{trans} = V_{max} * \frac{C_{STAT}}{K_D + C_{STAT}} \tag{6.16}$$

In the above equation $V_{max}$ is the maximal rate of transcription and the remaining term is the probability of STAT5A binding. Here $C_{STAT}$ refers to the activated STAT5 dimer concentration in the nucleus. Using our results from the transcription model we can replace the STAT5 binding probability with the probability $p_{trans}$ as calculated above. Hence the new transcription rate is as below where $p_{trans}$ is given by (6.6):

$$r_{trans} = V_{max} * p_{trans} \tag{6.17}$$

The time necessary to transcribe detectable levels of mRNA ($> 1.0 \quad nM$) is calculated simply by dividing the concentration with the rate and converting it to hours.

The values of the different parameters in the above equation are given in Table 3:

Table 3: Table of various biophysical parameters in transcription model

| **Symbol** | **Description** | **Value** | **Units** |
|---|---|---|---|
| β | Temperature | 1 | $k_BT$ |
| $\epsilon_{STAT}$ | STAT binding energy | 15 | $k_BT$ |
| $\epsilon_{GR}$ | GR binding energy | 12 | $k_BT$ |
| $\Delta\gamma$ | DNA unwinding parameter | 1.2 | $k_BT$ |
| x | STAT binding site 1 distance from reference | 3 | kb |
| y | GR binding site 1 distance from reference | 5 | kb |
| z | STAT binding site 2 distance from reference | 20 | kb |
| a | Length of base pair | 0.3 | nm |
| $\Delta F_{Loop}$ | Looping Free Energy | 1.6 | $k_BT$ |

**Results:**

To assess the model predictions we first conducted a scan over two different ranges of concentrations of the transcription factors STAT5A and GR at the nucleus.

*Scan over transcription factor concentrations - Part 1*

To understand the effect of transcription factor concentration on the bound probability and various other terms of the partition function, we performed an order of magnitude scan over the concentration of the transcription factors. Although, many of these concentrations are not physically relevant, they give us insight on the contribution of the various terms on the partition function and the overall transcription rate and helps us in performing sanity checks of our model.

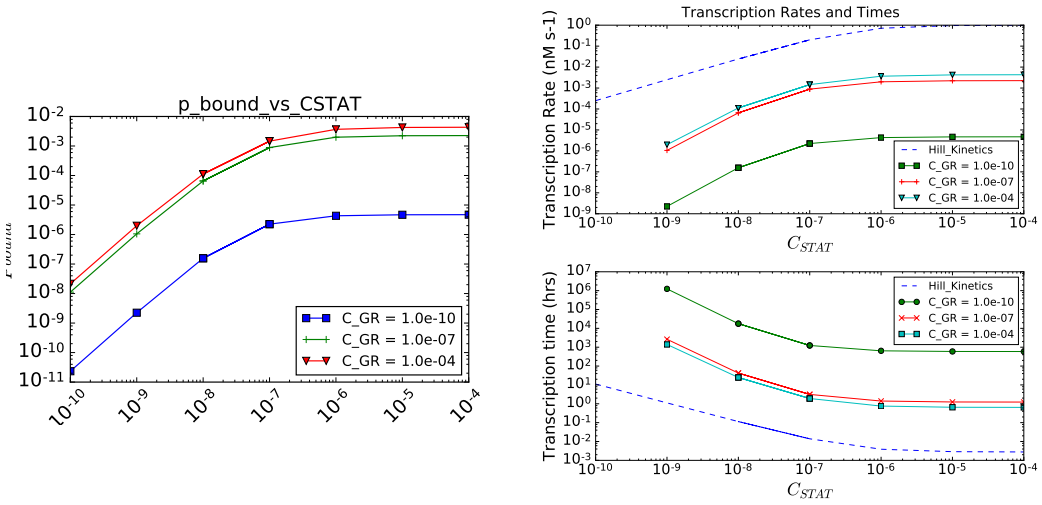


Figure 17: a: Transcription probability vs. STAT5A concentration with GR concentration as parameter; b: Transcription Rate and Time vs STAT5A Concentration with GR concentration as parameter

As Fig 17 shows, physiologically relevant the calculated transcription rates are observed at significantly high concentrations of transcription factors. This suggests our initial parameter set must be off. So we want to investigate the region of the parameter set that gives us physiologically relevant results.

To investigate the effect of the various terms in the partition function on the overall value we also separately plotted them for different transcription factor concentration.

*Scan over transcription factor concentrations - Part 2*

Using the insight obtained from the order of magnitude parameter sweep, we now focus on STAT5A and GR concentration ranges which are physiologically relevant - about 1-100 nM of each. The results are as shown below:

**Bound probability vs STAT concentration (GR concentration as parameters)**

To evaluate the functional dependence of the bound probability on the STAT5A concentration we plotted $p_{bound}$ as a function of STAT5A concentration (nM) for different values of GR concentrations (nM). This is shown below (y axis is in logarithmic scale)

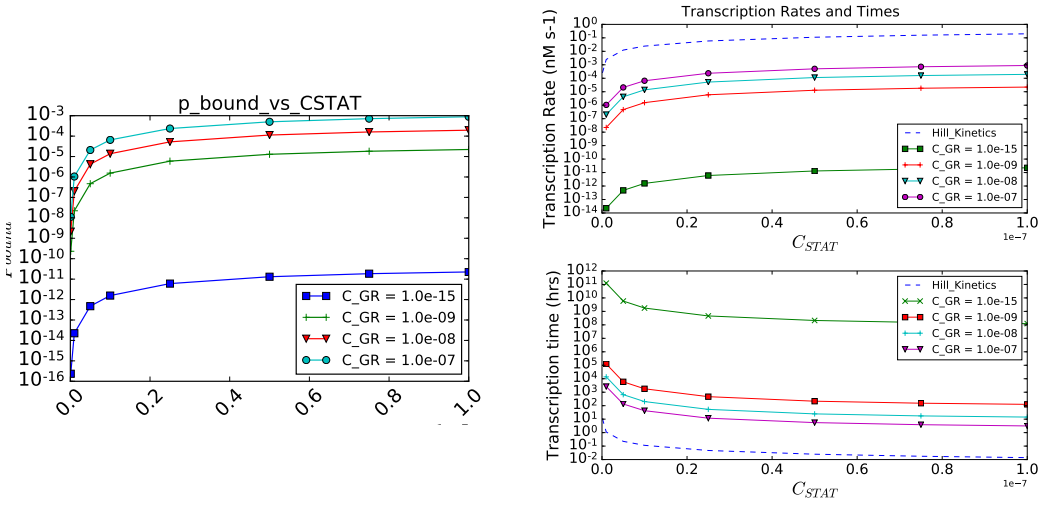


Figure 18: a: Transcription probability vs. STAT5A concentration with GR concentration as parameter; b: Transcription Rate and Time vs STAT5A Concentration with GR concentration as parameter

The plots above essentially offers us a magnified view of the behavior observed from the larger concentration scan above. Here we are zooming in at a physiologically relevant range of concentrations. The transcription probability follows a typical trend observed in the binding behavior with a rapid initial rise followed by saturation. We also see that in this range the GR concentration has an almost linear influence on the transcription probability - an order of magnitude increase in the GR concentration increases the transcription probability by the same amount.

The following plot shows the relative effect of various terms in the partition function

Figure 19: Breakdown of various terms in the partition function to illustrate the contribution of each in the overall partition function

The experimental results (89) on β-casein gene expression showed that the rate limiting step in the transcription is the DNA looping event. The plots in Figure (19) confirms this by showing that at this range of STAT5A and GR concentration the contribution of the looping event is the lowest. An increase in the GR concentration (second figure on the right) increases this somewhat which results in the elevated probabilities we saw earlier.

*Scan over other parameters of the model*

Next we performed a sweep over various key parameters of the model like the binding energies of each transcription factor and the DNA looping energy. The results are shown below:

**Binding Energies** $\epsilon_{STAT}$**,** $\epsilon_{GR}$

We first varied the STAT5A binding energy while keeping the other parameters including the GR binding energy as constant. Such variations in binding energy can occur due to mutations and also between different STAT isoforms and various low affinity binding sites have been shown (94) to have different binding affinities owards these isoforms.

In the physiologically relevent range of concentration for STAT5 it is observed that a higher binding affinity increases the probability of transcriptional assembly and correspondingly decreases the transcription time. However as we found earlier, the DNA looping step is rate limiting. This is reflected in the probability curves approaching an upper limit as binding energy is increased. The same trends are observed when we repeat the steps with GR binding energy.

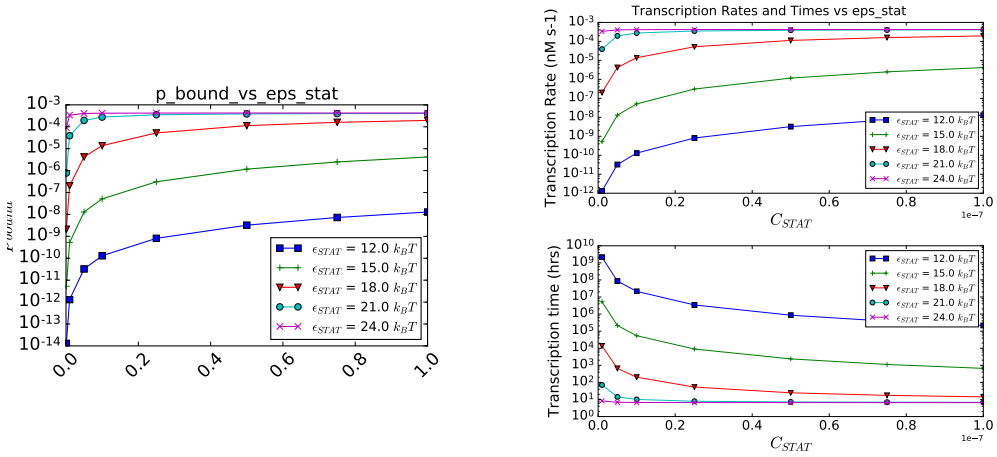


Figure 20: a: Transcription probability vs. STAT5A concentration with STAT binding energy as parameter; b: Transcription Rate and Time vs STAT5A Concentration with STAT binding energy concentration as parameter

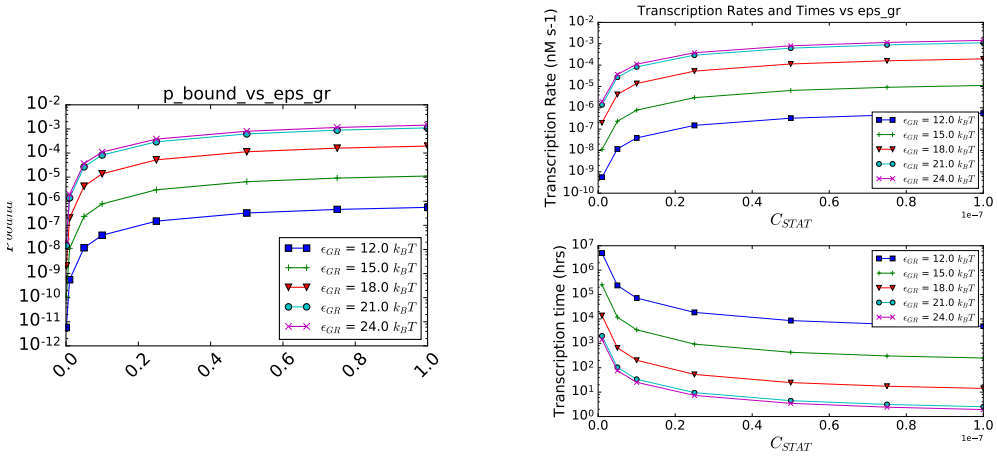


Figure 21: a: Transcription probability vs. STAT5A concentration with GR binding energy as parameter; b: Transcription Rate and Time vs STAT5A Concentration with GR binding energy concentration as parameter

### DNA Looping Energy $\Delta F_{loop}$

In a previous section we obtained a very rough estimate of the looping energy considering a circular loop with known number of base pairs separating the promoter and enhancer site

of the DNA. A better model has to take into account various looped configurations of the DNA and also the entropic contribution to the total free energy which we neglected here. We can get a preview of such differences through a parameter scan over the looping free energy which is shown below.

The first thing we notice is that the bound probability and the transcription time is not bounded like binding eneergy scans before in this concentration range. We expect this because we know that the DNA looping is rate limiting so it is expected to have a very pronounced effect on the probability of transcription and overall transcription time.

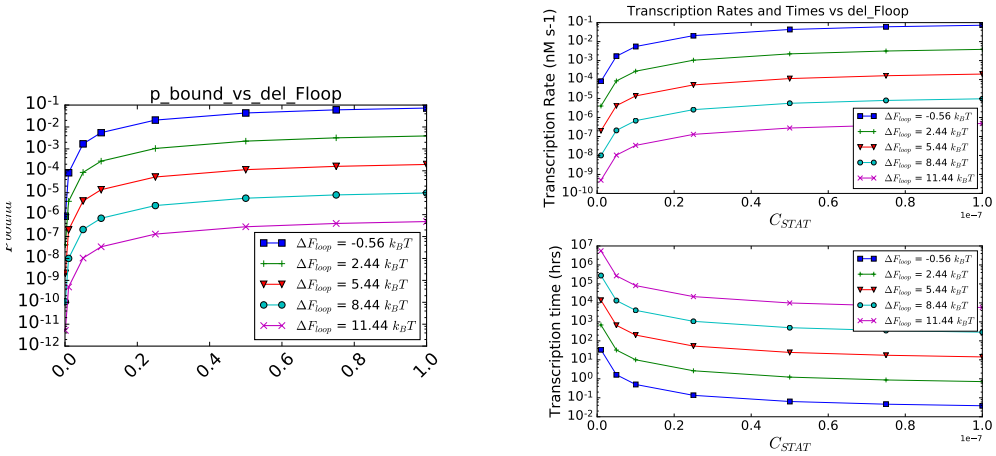


Figure 22: a: Transcription probability vs. STAT5A concentration with DNA looping energy as parameter; b: Transcription Rate and Time vs STAT5A Concentration with DNA looping energy concentration as parameter

# Part III

# Multi-Scale Models

# CHAPTER 7 : HYBRID MULTISCALE MODELING FRAMEWORK

## 7.1. Bridging the gap: Strategies for integrating multiscale models

When combining two or more models operating at different scales, one is often tempted to simply put the models together by combining the code. However, there are several pitfalls associated with this approach.

- Each individual model have their own uncertainties/errors in their variables. Simply combining the models will result in errors from one model affecting the other in unpredictable ways.

- When the processes in a model have different time scales, it results in a stiff system of equations that is difficult to compute. Implicit methods are needed to guarantee stable solutions, but they are often impractical for large systems. Hence one needs to combine implicit and explicit methods which makes the computation significantly more complex.

When the models to be combined are widely separated in time/length scales, it is possible to simplify the calculations using some approximations. We can consider the process with a shorter time/length scale to be fast from the perspective of the other process with a longer time/length scale. So for the *fast* process we only need to consider the steady-state values thereby eliminating the corresponding differential equation. In such cases where there is a separation between the scales, it is also possible to use the results of the faster process as a boundary or initial condition of the slower process (11).

The above ideas form the basis of most multiscale modeling techniques in the literature. Two complementary approaches, Heterogeneous Multiscale Method (95) and Equation free method (96) provide a rigorous formal description. The basic concept is that in complex systems, macroscopic coherent behaviors emerge from the interaction of microscopic components. So one can deduce the macroscopic rules from the microscopic ones by making

appropriately initialized calls to the microscopic simulator for short times (97). Hence if we can identify the interfaces between the models then these can be used to transfer data between the models. The degree of coupling can be *weak* on one extreme where one process is simulated over a long time frame and the results are used as inputs to the second process, or *tight* where both models are solved simultaneously. There are also approaches that are intermediate in that the models are run separately but the information transfer is *bidirectional* meaning there is feedback from the longer time scale process to the shorter one. We use this latter approach in our framework.

## 7.2. Description of the framework

### *7.2.1. Basic principles*

Models of biological pathways exist in different forms and used in different contexts. A modeling framework that aims to incorporate these existing models and possibly combining them with new ones, should have few specific requirements that enable such integration. It is to be noted that this integration applies not only to the core equations defining the model but also their input and output information such as initial and boundary conditions. These requirements are described below:

1. The framework should be able to incorporate two or more systems models that operate at different scales and run them together, passing information in a way that is both mathematically and biologically consistent.

2. The framework should allow incorporating input data from various sources such as tumor sequencing data for cancer patients, treatment information (such as drug dosage and schedule) and incorporate them into the component models.

3. The framework should have simple and clearly defined interfaces so that individual models can be added, modified or deleted without affecting other models in the assembly.

4. One should be able to test the effect of parameter space (for e.g., through sensitivity analysis) and variabilities in external inputs (for e.g., variation in growth factor concentration for a model of receptor-based signaling pathway).

5. The framework should allow extracting output data from individual models and model combinations and also allow calculation of biologically and clinically relevant parameters such as cell fates or production rate of a specific biomarker that is the end product of the signaling pathway.

6. The framework should be able to use models in standard formats such as SBML (98), JSON and allow the modeler to use different solvers for individual models.

*7.2.2. Model setup and algorithm*

There are two basic principles we use in our framework in order to integrate models operating at different scales.

- When two processes are well-separated in their characteristic time scales, then from the perspective of the slower process the faster process is at a steady state. This observation allows us to couple dynamics of these processes by evolving the slower process using steady-state information from the faster process.

- When two processes are modeled by different mathematical representations for e.g., continuous-time ordinary differential equations and discrete-time logical equations, a meaningful coupling of the models can be achieved by identifying a set of species which are common to both and modifying the governing equations and initial state of one model by using information obtained by running the other model for a specified amount of time.

In the Heterogeneous Multi-Scale Model formalism (95; 96), the model which has a greater characteristic time scale than the other is called a *macroscopic* the other is called *microscopic*. We use observation (1) from above to pass steady-state concentrations of interface

species between the two models. Let us say we have two models $A$ and $B$ which have $N_A$ and $N_B$ species respectively. The characteristic time scales of these models are $\tau_A$ and $\tau_B$ and $\tau_B >> \tau_A$ so model B is macroscopic and A is microscopic. Model A is governed by the following system of ODEs and initial conditions as described in Section 3.1

$$\frac{dC_i^A}{dt} = f(C_{i_1}^A, C_{i_2}^A, \cdots C_{i_{k\,i}}^A; \bar{p}),\ 1 \leq i \leq N_A \tag{7.1}$$

In Eq 7.1 the right hand side is a nonlinear function of species $i_1, i_2, \cdots i_{k_i}$ which are the $k$ species that interact with species $i$ including itself and $\bar{p}$ is the set of parameters such as rate constants. Each of these species satisfies an initial condition of the following form (where $C_{i0}$ is the initial expression of species $i$)

$$C_i^A(t=0) = C_{i0}^A,\ C_{i0}^A \geq 0 \tag{7.2}$$

The set of initial conditions $\bar{C}_0$ and $\bar{p}$ represent the information that is necessary for the model to be solved and predict its states at future times $t > 0$.

The model $B$ can be governed by a similar system of ODEs as above in which case it will have a similar form as Eq 7.1. However, to illustrate our full integration strategy let us consider that this is instead governed by a system of discrete equations as given below in the form of piecewise-linear equations described in section 3.2

$$\frac{dC_i^B}{dt} = b_i(\bar{C^B}) - \gamma_i C_i^B,\ C_i^B \geq 0,\ 1 \leq i \leq n \tag{7.3}$$

In above the function $b_i$ is a piecewise linear function that is made up of weighted products of step functions which have the following form

$$s^+(C_j^B, \theta_{ij}) = \begin{cases} 1, & C_j^B > \theta_{ij}^B \\ 0, & C_j^B < \theta_{ij}^B \end{cases}, \quad s^-(C_j^B, \theta_{ij}) = 1 - s^+(C_j^B, \theta_{ij}) \tag{7.4}$$

The initial conditions for the models are as below

$$C_i^B(t=0) = C_{i0}^B,\ C_{i0}^B \geq 0 \tag{7.5}$$

For a boolean model which is a special case of piecewise-linear differential equations, the governing equation 7.3 simply reduces to a rule based expression $\hat{C}_i(t+1) = \hat{b}_i(\hat{C}_{i1}, \hat{C}_{i2}, \cdots \hat{C}_{ik})$ where the concentration $\hat{C}$ can only assume values 0 and 1.

With these two governing equations, the integration strategy is the following

- Find the set of species $AB$ which are common to both models. Let there are $N_{AB}$ such species.
- Run model A for its characteristic equilibration times $\tau_A$ and record the steady-state concentrations of these $N_{AB}$ species $C_{i\infty}$ where $0 < i \leq N_{AB}$.
- If the model B is governed by PLDE then we use these concentrations directly to initialize B. If it is boolean then we use the following strategy (for step $n$)

$$\hat{C}_{i0}^B(n) = \begin{cases} 0 & \text{if } C_{i\infty}^A(n) < \theta_{i,l} \\ 1 & \text{if } C_{i\infty}^A(n) > \theta_{i,u} \\ \hat{C}_i^B(n-1) & \text{if } \theta_{i,l} < C_{i\infty}^A(n) < \theta_{i,u} \end{cases} \tag{7.6}$$

  In the above $\theta_{i,l}$ and $\theta_{i,u}$ are the lower and upper thresholds for the species $i$ which are obtained from literature.
- Run model B by modifying above initial conditions for the $N_{AB}$ species and record the steady state concentrations (or on/off status for Boolean model).

- Use these steady-state concentrations to initialize the concentrations of model $A$ depending on whether the model retains the memory of the previous iteration or not (more on this later). If we have a Boolean model then these discrete ON/OFF states are converted to continuous using the inverse of transformation 7.6

$$C_{i0}^{A}(n) = \begin{cases} \theta_{i,l} & \text{if } \hat{C}_i^B(n) = 0 \text{ and } \hat{C}_i^B(n-1) \neq \hat{C}_i^B(n) \\ \theta_{i,u} & \text{if } \hat{C}_i^B(n) = 1 \text{ and } \hat{C}_i^B(n-1) \neq \hat{C}_i^B(n) \\ C_{i\infty}^{A}(n-1) & \text{if } \hat{C}_i^B(n-1) = \hat{C}_i^B(n) \end{cases} \tag{7.7}$$

- Both the models are run until the final results converge or till a maximum number of steps are exceeded.

The above algorithm is illustrated in the flowchart 23.

### *7.2.3. Model Input and Output*

The type of input and output is dependent on the type of model. To show the flexibility of the framework we will again use the example of a microscopic ODE model and a macroscopic boolean model as in the previous section. The ODE models we have used in this work have been models of cell surface receptor-mediated signaling pathways such as EGFR mediated Ras-MAPK or PI3K/AKT pathway. The input to these models is the initial expressions of all the species in the pathway and the rate constants of the modeled interactions. These are obtained from values published in the literature. The boolean model we have used is a model of DNA damage repair and cell cycle driven by tumor suppressor p53. This model has two different initial configurations depending on whether a drug (or radiation) is being dosed or not which triggers a DNA damage response. The species thresholds, that determine their on/off status as explained in the previous section, are obtained from the literature. The model that is initialized with literature values is our control or base model. To make our predictions patient-specific, we get the miRNA, mRNA and protein expression data. The type and availability of this data are variable and can come from sources such as TCGA or

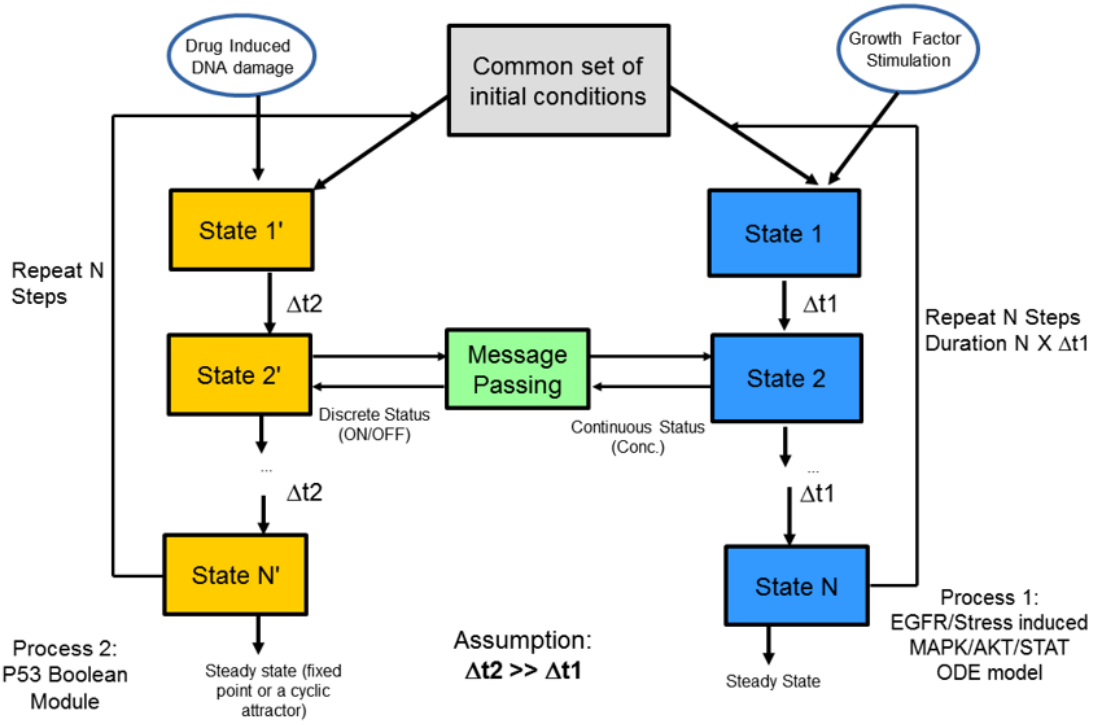


Figure 23: Flowchart showing the algorithm of the hybrid simulator. The yellow boxes represent the macroscopic models and the blue boxes represent the microscopic models. Each model has their own characteristic time scales $\Delta t_1$ and $\Delta t_2$ and here $\Delta t_1 << \Delta t_2$. Each model can have a common set of input condition such as patient specific mRNA/protein expressions. We can also incorporate other inputs such as growth factors and chemotherapeutic drug induced DNA damage. The green box represents the message passing interface that uses the common interface species to pass the information between these models. Each model is evolved in this way until they settle to a common steady state

directly from a clinician. These data are preprocessed separately and the top genes which are over or under-expressed with respect to control (patients without disease) are obtained. Through network analysis, the downstream targets of these species that are present in our model are found and their initial expression is adjusted by a similar proportion as their difference from control data. For a boolean model where only two states are possible, we constrain the target nodes to an ON/OFF state (depending on whether the gene is over or under-expressed with respect to control) for all members of the initial state space (explained in the next section). Thus we get two instances of the model initialized with two different gene and protein expression signatures which can be run to obtain the predictions for the patient and compare it with control.

The model output is also dependent on the type of model. When we have one or more pathways such as EGFR mediated Ras-MAPK pathway the output is typically the steady-state expression of the species in the pathway such as ERK. For a boolean model of the cell cycle, the output is steady-state status (on/off) of the proteins which can be both a *fixed point* (a single value that does not change over subsequent steps) or *limit cycle* (a set of states that repeat over time). For the p53 driven DNA damage and cell cycle pathway, these steady states have been correlated with specific cellular fates like cell death (apoptosis) or cell proliferation/growth (10). We calculate probabilities of three cell fates (cell death, growth, and senescence) by running our hybrid model for different combinations of initial states of the boolean model nodes which are not constrained by initial expression or drug dosage. This combination of initial states obtained from the unconstrained nodes form the *initial state space* of the model. The cell fate probability can be used to calculate clinically relevant parameters such as cell kill rate of a specific drug or production rate of a particular biomarker such as Prostate Specific Antigen (PSA). These will be shown in the next two chapters.

*7.2.4. Effect of parameter space and environment*

There are two important sources of variability in models of biological pathways like the ones described in the previous section. The first source of variability is from the set of free parameters (initial expression and kinetic rate constants for ODE models and threshold values for the Boolean model). As explained in Chapter 4 the characteristics of the parameter space have profound consequences for the model output. Because of the hybrid nature of the models, the traditional sensitivity and sloppiness/robustness analysis techniques mentioned in the previous sections do not apply. We use a simpler strategy to account for these variabilities. First, we identify biologically relevant thresholds for these parameters from the literature. Then using techniques such as Latin Hypercube Sampling or Quasi Random sampling (Section 4.1.2) we create an ensemble of parameters (and hence model). The model is run for each member of this ensemble and the ensemble output is presented along with average and dispersion measures. To highlight the differences from the base set of model parameters one can calculate a parameter called Normalized Deviation of cell fate (for example cell growth) for the $i^{th}$ member of the ensemble as below.

$$\bar{Pr}^{i}(growth) = \frac{Pr^{i}(growth) - Pr^{base}(growth)}{Pr^{base}(growth)} \tag{7.8}$$

The second source of variability comes from the external environment. For a model of a receptor-mediated signaling pathway, some examples of variabilities in the external environment would be variations in growth factors, drugs, etc in the extracellular medium. For a cell cycle model, the cell cycle time could also vary which is related to the *memory* of the cell. In the context of our model, memory means whether the model remembers its state at the end of step $n-1$ when we start the simulation for step $n$. We consider both possibilities of the model remembering or forgetting its previous state and include these values as part of our predictions.

# CHAPTER 8 : HETEROGENEOUS, MULTISCALE MODEL FOR PROSTATE ADENOCARCINOMA AND EFFECT OF PTEN DELETION ON ANDROGEN DEPRIVATION THERAPY

Adapted from (99)

## 8.1. Introduction

In this work, we have created a heterogeneous and multiscale modeling framework that allows one to combine two or more models of cellular pathways operating under different time scales and time descriptions. We have shown the utility of this framework by applying it to Prostate Cancer (PCa), currently the third most lethal form of cancer (100). There are multiple molecular subtypes and the treatment outcome is highly patient and tumor-specific (101). The disease progresses through multiple stages starting from benign prostatic intraepithelial neoplasia which progresses to Localized prostate cancer depending on the presence of specific mutations (102). Localized prostate cancer (PCa) is commonly treated with radical prostatectomy (RP). About 20 % to 25 % of patients undergoing RP experiences biochemical recurrence within 5 years as indicated by significant ($\geq$ 0.2 ng/mL) levels of Prostate Specific Antigen (PSA) in serum which is expected to be absent after primary tumor resection. PCa is driven by androgen receptor (AR) signaling, which is dependent on androgens like testosterone and dihydrotestosterone (DHT). Hence, one treatment option for biochemically recurrent PCa is Androgen Deprivation Therapy (ADT). Various pre- and post-treatment prognostic factors such as absolute baseline PSA, tumor stage and pathologic findings such as Gleason Score allow physicians to risk stratify a particular patient and in high-risk cases, ADT is administrated to prevent a recurrence, in the so-called adjuvant setting (103). These therapies have a temporary benefit manifested by decreased tumor cell proliferation and reduced PSA levels. However, due to aberrant AR signaling, genomic and epigenetic instabilities, PCa cells are thought to become resistant and grow independent of androgen. Multiple AR specific factors such as point mutations, overexpression, splicing,

etc. and alterations in the AR signaling pathway have been identified as factors that can lead to resistance. Another key mechanism of resistance to ADT is the cross-talk of AR pathway with other proliferative pathways like PI3K/AKT which do not explicitly depend on androgens (104). Loss of PTEN (phosphatase and tensin homolog) which negatively regulates AKT has been shown to be sufficient to induce tumors in mouse models (105). Tumors with PTEN protein loss has been found to be more likely to be upgraded at radical prostatectomy with respect to the previous biopsy than those without it (106).

We have used our multiscale modeling framework to combine the AR signaling pathway with Ras-MAPK, PI3K-AKT and tumor suppressor p53 mediated DNA damage repair and cell cycle pathways. The model predicted cell fate outcomes were validated by converting them into serum PSA values using an exponential tumor growth model and comparing them with actual patient postoperative PSA levels during ADT. The model was used to analyze the outcome of ADT for a population of PCa patients in TCGA who were classified based on the presence or absence of PTEN deletion mutation and their response to ADT by incorporating the differences in gene/protein expression data between these groups. Heterogeneities in tumor microenvironment and effect of perturbations in the model parameters were also considered. The model showed several intriguing and non-intuitive results for these patient groups. In particular, it showed how heterogeneity can be an important means that tumors can exploit to evade treatments like ADT. We believe such a multiscale framework is invaluable both as a clinical tool and for designing/interpreting experiments that can enhance our understanding of the progression of cancer, response to therapies and development of resistance.

## 8.2. Materials and Methods

### *8.2.1. Population Description*

Tumor sequencing and clinical information for prostate cancer patients were obtained from The Cancer Genomic Atlas (TCGA) http://cancergenome.nih.gov/ and GDC Data Por-

tal http://cancergenome.nih.gov/. They were classified based on molecular and clinical parameters such as biochemical recurrence (occurrence and timing), adjuvant ADT therapy, PTEN deletion status, pathological Gleason Score, positive margins, tumor status evaluated by biopsy and imaging.

From the initial set of 501 patients, we first excluded patients who had adjuvant and neo-adjuvant radiotherapy since the effect of radiation is not included in our model. A final cohort of 250 patients was selected based on their Gleason Grade ($> 7$), presence/absence of PTEN deletion mutation and their response to ADT. Patients who responded to ADT (no recurrence) were put in the Control (CNT) group, the patients who showed biochemical recurrence (BR) based on serum PSA measurements and the patients who showed tumor recurrence (TR) based on biopsy were put in two other groups with the same name.

The selected cohort was further categorized first based on the presence of PTEN deletion mutation. Figure 24 shows the classification scheme, including the size of the combined groups) and response to ADT. Differential gene expression analysis was performed between the latter subgroups (CNT, BR and TR) using the R2 genomics platform http://r2.amc.nl to obtain a list of genes that are differentially expressed in the groups with recurrence compared to control (no recurrence).

## 8.3. Molecular Model Description

### *8.3.1. Biological Pathways*

Our modeling framework (as described in Chapter 7) can combine two or more biological pathways relevant to a particular disease type taking into account the differences in their time scales and time descriptions. For PCa, apart from the Androgen Receptor pathway, many other pathways have been identified to play a role in disease progression and development of resistance to ADT. In this work, we mainly focus on the cross-talks of the AR pathway with Ras-MAPK, PI3K-AKT pathways and p53 driven DNA damage repair and cell cycle pathway. These cross-talks have been identified as one of the key mechanisms of

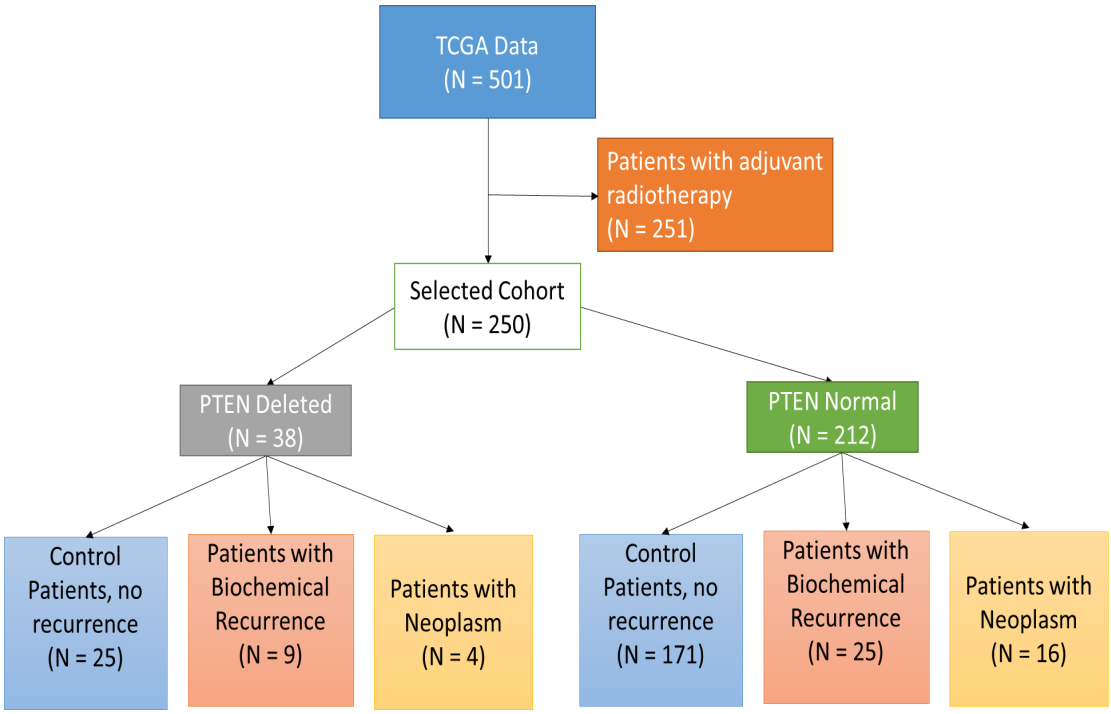


Figure 24: Flow chart of the cohort selection. From the initial 501 patient samples we first removed patients who underwent adjuvant radiotherapy. This cohort of patient was further classified into two groups based on whether or not PTEN deletion mutation was present. Each of these groups were further classified as control or patients with no recurrence (CNT), patients with biochemical recurrence (BR) based on PSA values and patients with tumor recurrence (TR).

development of resistance to ADT and progression to Androgen-independent Prostate Cancer. The protein PTEN, which is an important component of the PI3K/AKT pathway has been found to be particularly important in these interactions between the three pathways. We constructed a simple continuous-time ODE based model of the Androgen Receptor pathway based on literature which has a characteristic time scale of about 6-8 hours. The different components of the model, the reactions and the parameter values are described in the Supplementary section of (99). Continuous-time ODE based model of Ras-MAPK and PI3K/AKT pathways have already been published in the literature which has been adapted in this work (7). This pathway has a characteristic time scale of about 4-6 hours. In order to predict cell fates, we need a model for cell cycle driven by tumor suppressor proteins such as p53. Unfortunately, detailed continuous-time ODE based models of these pathways are not available due to lack of quantitative data and highly complex nature of these pathways. A logic-based discrete modeling approach has been employed with some degree of success. We have adapted such a discrete model of cell cycle driven by p53 from literature which has been shown to generate three distinct outcomes corresponding to cell cycle progression (proliferation), cell death (apoptosis) and cell senescence (growth arrest) (10). We have chosen this pathway because both the AR signaling pathway and PI3K/AKT pathway have been known to interact with p53 and its effectors to alter cell fate decisions particularly during the progression of the disease into Androgen Independent form (104). The first step to combine these models with different time scales and time descriptions (continuous vs. discrete) is to identify interface proteins which are common to two or more pathways. These interface proteins were identified by performing a literature search (107; 108; 109; 110). The androgen receptor signaling pathway used in this work is shown in Figure 25

In our model, these three pathways interact through certain interfaces that are well-known in the literature. Some of these interactions are shown in the Figure 26.

We used all these interactions to design model interfaces through which the individual models communicated with each other based on the protocol we established in Chapter 7.

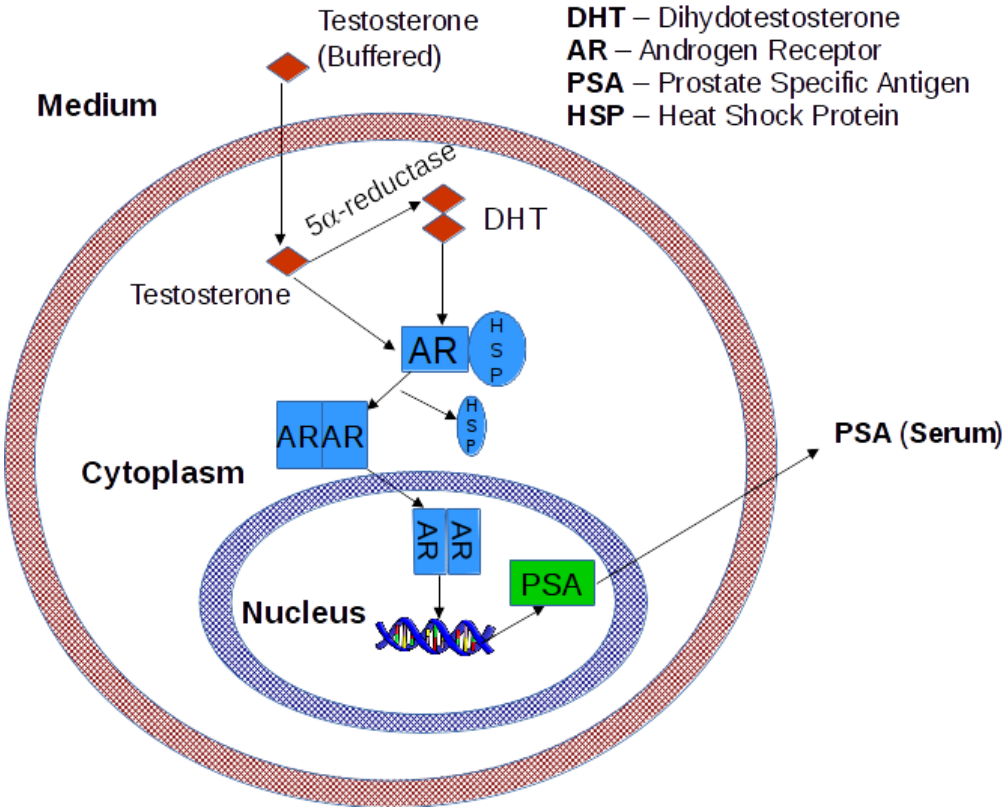


Figure 25: The androgen receptor signaling pathway based on the description in (104). The androgen receptors are activated by either testosterone or dihydrotestosterone. Activated AR dimerizes and translocates to the nucleus to initiate transcription of many proliferative genes including PSA.

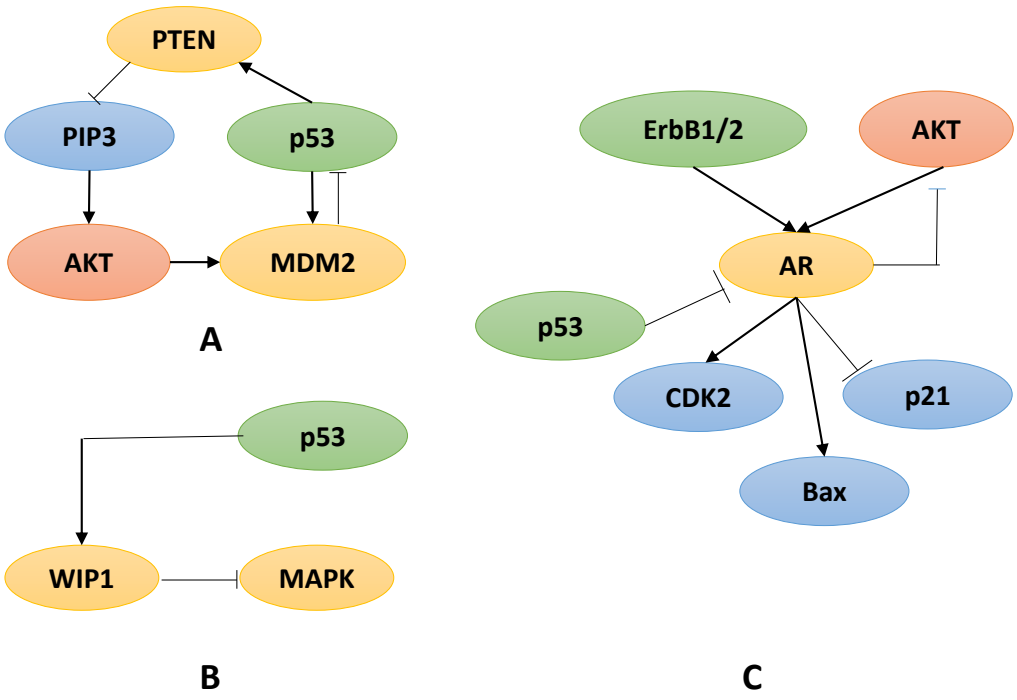


Figure 26: The interactions between the key proteins in the pathways considered in the model.

*8.3.2. Differential Gene Expression Analysis and Model Initialization*

As described in the first section our selected TCGA cohort was classified into two main groups depending on the presence or absence of PTEN deletion mutation. Each of these groups (referred here by PTEN del and PTEN normal) were further classified into three groups based on their response to ADT:

1. A control group (CNT) with no recurrence events.
2. A group with biochemical recurrence (BR) based on PSA levels in the serum.
3. A group with tumor recurrence (TR) measured with a biopsy.

The way we replicate this classification in our model population is as follows:

1. We run two instances of our model with high and low initial PTEN numbers.
2. For each of the above instances, we generate three different instances by performing a differential gene expression analysis of the above groups in R2 genomics platform for the TCGA dataset and find which mRNA and proteins are differentially expressed in the BR and TR groups compared to CNT.
3. We modify the initial expression of the target of these mRNA and proteins in our network and create the three model sub instances.

*8.3.3. Effect of ADT and Tumor Microenvironment Heterogeneity*

To simulate ADT we ran our model for different levels of androgen (Testosterone) stimulations. The resting testosterone concentration in untreated subjects was reported in the range of 1-20 nM in previous studies (111) and we have chosen the testosterone levels in the same range for subjects under ADT. To consider variations due to heterogeneity in the tumor microenvironment (in a crude fashion) we create four sub instances where we consider two different levels of a growth factor (EGF) and two different memory conditions. Mem-

Table 4: Table showing the main simulation conditions. There were two main classification schemes, initial PTEN expression and Testosterone levels. We selected 5 different testosterone levels from 1 nM to 20 nM and two different PTEN expressions.

| Case | Testosterone (nM) | PTEN (#) |
|---|---|---|
| 1 | 20 | 56000 |
| 2 | 15 | 56000 |
| 3 | 10 | 56000 |
| 4 | 5 | 56000 |
| 5 | 1 | 56000 |
| 6 | 20 | 5600 |
| 7 | 15 | 5600 |
| 8 | 10 | 5600 |
| 9 | 5 | 5600 |
| 10 | 1 | 5600 |

| Condition | Value |
|---|---|
| EGF | 0.1 nM |
| EGF | 100 nM |
| Memory | Yes |
| Memory | No |

ory here indicates to what extent the model remembers its previous state. We consider two extremes - *with memory* where the model remembers its previous state and *without memory* where the model does not remember its previous state. These memory conditions are representative of cell cycle time variations. So for each class of patient (CNT, BR or TR), we have two different PTEN expressions, five different testosterone levels and four different heterogeneity conditions giving a total of 40 different instances of the base model. These conditions and their corresponding labels are summarized in Table 4 (left). In addition, the effect of intra-tumoral heterogeneity is factored by running each case in Table 4 for four conditions summarized in Table 4 (right).

### *8.3.4. Calculation of Cell Fate and PSA*

The main predictions of a specific instance of the model are probabilities of cell fates - proliferation/growth, apoptosis/death, and senescence/arrest. These probabilities represent the likelihood that a cell will commit itself to a specific fate based on the activities of the constituent pathways. However, for clinical use, the parameters of interest are tumor growth rates or PSA values. These probabilities can be converted very simply to growth rates using an exponential or Gompertzian growth model.

We use a simple population model of tumor growth. Since we are considering PSA values in the early stages (12-14 months post-surgery), it is assumed that the tumor is in an initial phase where it grows with an exponential law,

$$\frac{dS}{dt} = rS(t) \tag{8.1}$$

In 8.1 equation, S is the number of tumor cells after time t and r is the growth rate. The growth rate r is typically testosterone dependent and has a linear relation with testosterone concentration T

$$r = r_0 + pT \tag{8.2}$$

Since during ADT the testosterone levels are low we can assume that the second term is much smaller than the first so that $r \approx r_0$.

The predicted net cell growth probability NCG is directly related to $r_0$ as $r_0 = k.NCG$ where $k$ is some proportionality constant. Integrating the above differential equation with respect to time and using the above approximations we get

$$S = S_0 e^{kNCGt} \tag{8.3}$$

The PSA values are directly proportional to the number of tumor cells and so the final expression relating PSA to NCG is as below

$$PSA(t) = S_{coeff} S_0 e^{kNCGt} \tag{8.4}$$

The value of k can be obtained from literature estimates of $r_0$ and $S_{coeff}$ can also be

obtained from the literature (112). We used the NCG values averaged over the different tumor heterogeneity conditions and the average deviations to compute an upper and lower bound of the PSA vs. time curves.

### *8.3.5. Effect of Parameter Variations*

A common method of checking the sensitivity and robustness of non-linear systems models in Global Sensitivity Analysis. These techniques are (25) helpful in understanding what parameter or set of parameters influence the model outcome and how uncertainties in these parameters manifest themselves as uncertainties in the model output. Common variance-based global sensitivity analysis techniques determine a sensitivity coefficient for each model parameter. This coefficient quantifies the contribution of the uncertainties/variations in the corresponding parameter to the uncertainties/variation in the model outcome. For a multiscale and heterogeneous model like ours, a single sensitivity coefficient is not very meaningful as the model coarse grains the output of the microscopic ODE based models when passing it to the macroscopic model. So we used a different and simpler strategy to investigate the effect of parameter variations. We generated an ensemble of models by sampling the parameter space. Since the model has many parameters we need efficient means of sampling from a space of such high dimensions. One such method that is effective for parameter spaces of large dimensions is Latin Hypercube Sampling (25). If we have $N$ parameters of the system and we want to draw $M$ number of samples then first we divide the range of each of $N$ parameters into $M$ equally spaced intervals. Then from a particular parameter $i$ we randomly select a value from each interval. Each of these values is paired randomly with the $M$ values of all other $N-i$ samples. We run the model for each of the $M$ resultant samples to generate $M$ NCG values. Then we calculate the normalized deviation from the base values. For the ith sample, this can be defined as

$$ND_i = \frac{NCG_i - NCG_{base}}{NCG_{base}} \tag{8.5}$$

This enables us to compare the net cell growth for different samples under different conditions of tumor heterogeneity. Further, the sign of $ND$ tells us whether the sample shows higher or lower proliferative potential compared to the base.

## 8.4. Results

### *8.4.1. Model Validation*

The molecular model was validated using the EUREKA 1 (113) data collection which contains a subset of prostatectomized patients considered high risk based on their Gleason Scores ($\geq 8$) and positive surgical margins. Three patients were selected from this group who underwent adjuvant ADT and for whom at least three post-operative serum PSA values were available. Of these three patients, two had an increase in their serum PSA levels above the baseline value of 0.2 ng/mL over a time period of 14 months following prostatectomy. For the third patient, the PSA levels stayed below baseline in this period and we selected this patient as control. Then we used the molecular model, initialized it based on the mRNA and protein expression profile for each of these patients and ran the model to obtain predicted net cell growth probabilities. Using a simple exponential tumor growth model (described in methods) we converted these net cell growth probabilities into PSA values and compared them with the observed values. The PSA levels (ng/mL) vs. time in months for the control patient and the patients with tumor recurrence are plotted in Figure 27 along with the molecular model predictions.

As explained in methods, the model-predicted net cell growth probabilities are averaged over different conditions that mimic tumor heterogeneity. Hence, we included both the upper and lower bounds of predicted PSA vs. time curves using the average deviations due to tumor heterogeneity. The plots in Figure 27 demonstrate that the molecular model can distinguish the different genetic signatures of control patients and patients with recurrence and the corresponding predictions of net cell growth and PSA values qualitatively match the changes in the observed PSA values.

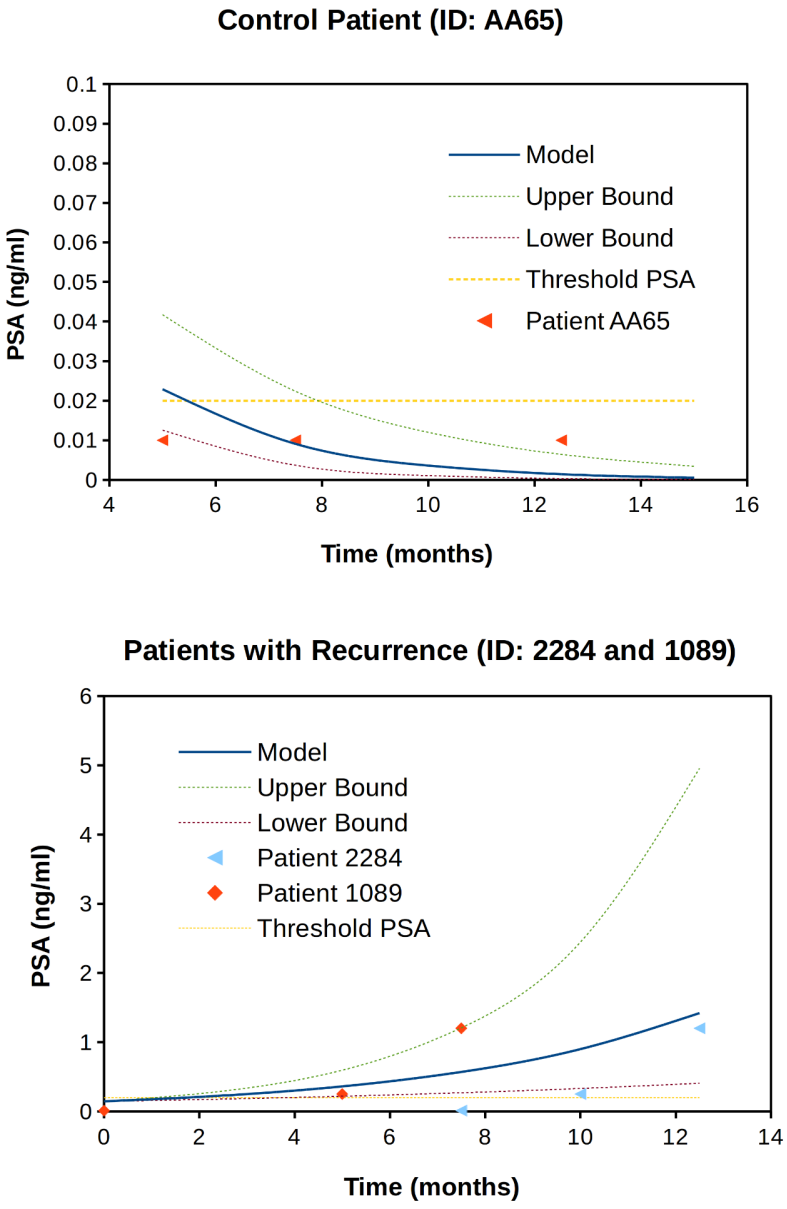


Figure 27: Validation of molecular model by comparing model predicted PSA time profiles with the actual post-operative PSA values in control (no recurrence) and patient with tumor recurrence from EUREKA 1 database. The PSA time profile was obtained using a simple one parameter exponential tumor growth model (SI) which used to predict net cell growth probabilities as input. To highlight the uncertainties in model output due to tumor heterogeneity upper and lower bounds of these curves were also included.

### *8.4.2. Predicting the Effect of ADT on TCGA Prostate Cancer Patients with and without PTEN Deletion*

The results in the previous section showed that the molecular model can distinguish two sets of patients - control group with no tumor recurrence vs. those with recurrence based on their gene and protein expression data. Our next objective was to apply this model to a larger cohort of patients and systematically explore the effect of ADT on different subgroups of this cohort. We used the TCGA database where we selected a dataset for 501 prostate cancer patients. This population of the patient was classified based on different markers such as presence or absence of PTEN deletion and their response to ADT (detailed in Methods). Unlike the dataset used for model validation in the previous section, here we do not have postoperative PSA values at different time points. We only have records of whether a patient showed no recurrence, biochemical (through PSA values) recurrence or tumor (through biopsies) recurrence. Hence instead of predicting PSA values here, we focus mainly on the differences in predicted cell fate values between the control, biochemical and tumor recurred groups for both PTEN deleted and PTEN normal patients for different levels of androgen (testosterone) stimulations. Using differential gene expression analysis we obtained the specific mRNA/proteins in the three pathways which are differentially expressed in the groups of patients with biochemical and tumor recurrence with respect to the control group with no recurrence. Based on the differential gene expression analysis, the initial expressions of these target mRNA/proteins were adjusted in the models. We want to see how the model predictions of net cell growth probability, which is an indicator of the proliferative potential of the cell, changes between these groups for different PTEN expression and different androgen stimulation (effect of ADT) and whether these are consistent with what is known in the literature.

The box plots in Figure 28 show the model predicted cell kill, cell growth and net cell growth probabilities (NCG) which is the difference between cell growth and cell death probability for the control (non-relapsed) patients. The labels on the horizontal axis correspond to

the simulation conditions shown in Table 4 (left) whereas each point in a particular box represents conditions in Table 4 (right). Each plot has two sub-regions. Labels 1-5 show the probabilities for PTEN normal patients whereas labels 6-10 show the probabilities for PTEN deleted patients. The androgen stimulation decreases from 20 nM to 1 nM for each group of PTEN normal (1-5) and PTEN deleted (6-10). For patients with no PTEN deletion, the plots show that cell growth and net cell growth probabilities decrease as androgen levels are decreased, suggesting that these patients will respond to ADT. However, for the PTEN deleted group we see little dependence on androgen levels and the average net cell growth probability is higher compared to the PTEN normal group which suggests that these patients will respond poorly to ADT. For the Net Cell Growth (NCG) values we used linear regression to calculate the coefficient of determination $R^2$ for each group of patient and plotted the trendlines (corresponding equations and values are given in the caption of Figure 28). These trendlines and $R^2$ values confirm the better response of PTEN normal patients compared to PTEN deleted patients. To confirm the difference in the mean NCG between PTEN normal and PTEN deleted patients we also calculated the p-values between these groups. These are shown in Table 5 which shows that the PTEN normal patients have a lower average NCG compared to PTEN deleted patients within a significance threshold of 0.05. This is consistent with the literature and also the fact that there is a higher percentage of patients with recurrence (52 % to 23.9 %) for the PTEN deleted group compared to PTEN normal group in the TCGA dataset. Apart from the differences in the average values of net cell growth, we also observed differences in the spread (variance) of the net cell growth for different tumor heterogeneity conditions (growth factor levels and memory conditions, explained in the Method section) between the PTEN deleted and PTEN normal patients. For PTEN normal group we see a higher spread compared to PTEN deleted group. This difference in variability between the PTEN deleted and PTEN normal groups can have important consequences so far as the effect of ADT is concerned. Higher variability implies that a part of the tumor population will survive the treatment (as shown by higher net cell growth) due to tumor microenvironment heterogeneity. This is also something that has been

Table 5: Comparison between PTEN deleted and PTEN normal groups for different patients. Significant values are shown in bold $*p < 0.05$. The up and down arrows indicate the type of test (greater or less). So a downward arrow for the CNT patient indicates mean PTEN normal net cell growth is lower than mean PTEN deleted net cell growth.

| Patients | p-value (PTEN Norm vs. PTEN Del) |
|---|---|
| CNT | **2.012e-04**↓ |
| BR | **1.174e-04**↓ |
| TR | **0.038**↓ |

Table 6: Comparison of p-values between different groups of patients. Significant values are highlighted. Significant values are shown in bold $*p < 0.05$. The up and down arrows indicate the type of test (greater or less). So an upward arrow for the CNT patient and TR patient (first row third column) for the table at the right (PTEN deleted group) indicates the mean net cell growth of control patients is higher than the tumor recurred patients

| PTEN Norm | Control | BR | TR |
|---|---|---|---|
| Control | 1 | 0.9165 | 0.7888 |
| BR | 0.9165 | 1 | 0.7243 |
| TR | 0.7888 | 0.7243 | 1 |

| PTEN Del | Control | BR | TR |
|---|---|---|---|
| Control | 1 | 0.5904 | **0.0343**↑ |
| BR | 0.5904 | 1 | **0.016**↑ |
| TR | **0.0343**↓ | **0.016**↓ | 1 |

reported in the literature for different cancers including PCa. As we will see, the nature of this variability differs between control patients and patients with tumor recurrence. This variability due to tumor heterogeneity is also distinct from those induced by perturbations in the mRNA/protein expression levels and reaction rates.

Next, we ran our model for the patients with biochemical recurrence and compared the net cell growth probabilities with the control (no recurrence) group from the previous section. The results are shown in Figure 29 where we plot the same three probabilities (cell death, cell growth, and net cell growth) under the simulation conditions in Table 4. In Table 5 and 6 we give a summary of the comparison between the different patients and PTEN groups in terms of p-values.

From these plots, we see that apart from minor differences in the variabilities of net cell growth there are no significant differences in the changes of average net cell growth probability compared to the control group. This observation is confirmed from the p-values

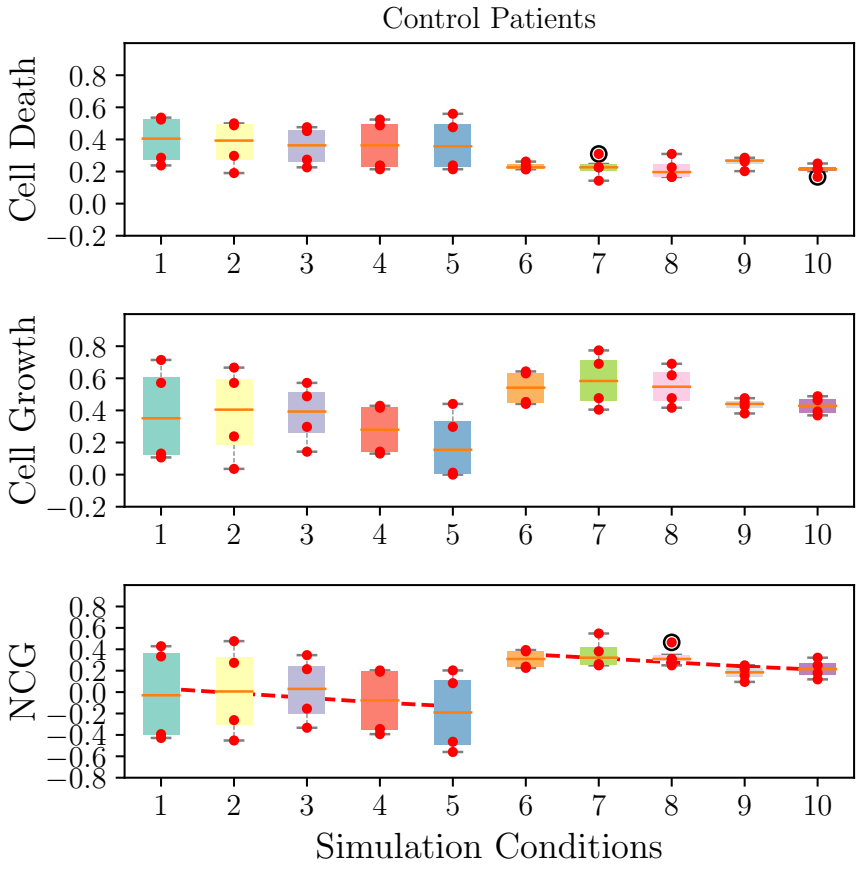


Figure 28: Calculations of cell kill, cell growth and net cell growth probabilities for control (non-relapsed patients). The ten conditions (1 to 10) are the conditions summarized in Table 4 (left). The four data points for each condition are the four states summarized in Table 4 (right). The boxes **1–5** indicate the net cell growth probabilities **without PTEN deletion** for **decreasing** testosterone stimulations 20 nM to 1 nM respectively. The boxes **6–10** indicate net cell growth probabilities **with PTEN deletion** for **decreasing** testosterone stimulations 20 nM to 1 nM respectively. The lower panel which shows the Net Cell Growth probabilities also includes a trendline drawn between the means to indicate their change with ADT (decreasing testosterone levels). For the PTEN normal patient the trendline is $NCG = 0.008838C_T + -0.1419$ where $C_T$ is the testosterone concentration. The least square fit parameter is $R^2 = 0.628$. For the PTEN deleted group the trendline is $NCG = 0.00766C_T + 0.2016$ and $R^2 = 0.556$.

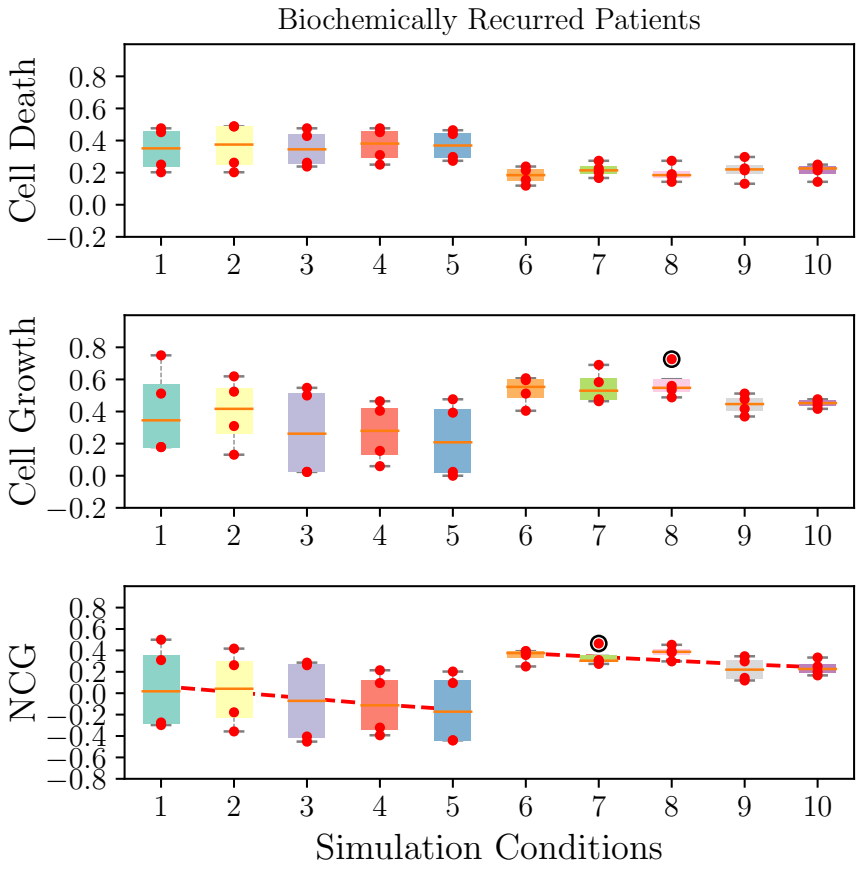


Figure 29: Calculations of cell kill, cell growth and net cell growth probabilities for patients with biochemical recurrence (PSA values). The ten conditions (1 to 10) are the conditions summarized in Table 4 (left). The four data points for each condition are the four states summarized in Table 4 (right). The boxes **1–5** indicate the net cell growth probabilities **without PTEN deletion** for **decreasing** testosterone stimulations 20 nM to 1 nM respectively. The boxes **6–10** indicate net cell growth probabilities **with PTEN deletion** for **decreasing** testosterone stimulations 20 nM to 1 nM respectively. The lower panel which shows the Net Cell Growth probabilities also includes a trendline drawn between the means to indicate their change with ADT (decreasing testosterone levels). For the PTEN normal patient the trendline is $NCG = 0.01143C_T + -0.1624$ where $C_T$ is the testosterone concentration. The least square fit parameter is $R^2 = 0.948$. For the PTEN deleted group the trendline is $NCG = 0.006863C_T + 0.236$ and $R^2 = 0.563$.

calculated for the differences in the mean net cell growth between the control and BR patients as shown in Table 6. The p-values between Control and BR patients are greater than the significance threshold for both PTEN normal and PTEN deleted groups. This suggests that the differentially expressed genes between control patients and patients with biochemical recurrence did not produce significant differences in the activities of the signaling networks to produce a different cell fate outcome. Literature studies on different PCa patients have suggested that isolated PSA values are not a good indicator of PCa recurrence or effect of ADT. Our results which distinguished these groups based on their gene/protein expression signatures, are consistent with these results and at the same time provides a mechanistic basis based on the interactions of the AR pathway with Ras-MAPK and PI3K-AKT pathway and the cell cycle pathway.

Finally, we ran the simulations for the third group of patients who showed tumor relapse as measured through biopsies. The results are shown in Figure 30 where we plot the same three probabilities (cell death, cell growth and net cell growth) under the simulation conditions in Table 4.

The bar plots show that for this group of patients there are significant differences in the predicted net cell growth probabilities compared with control, particularly among those patients who have PTEN deletion mutation. For the PTEN deleted group we observe a bimodal distribution of cell fate values where a population of patients show consistently low net cell growth. This is again confirmed from the calculated p-values for the difference in the mean net cell growth between Control and TR patients shown in Table 6 which shows a statistically significant difference in the mean values based on a significance threshold of 0.05. This suggests that the patients in this category may respond to androgen deprivation initially, but the tumor can recur at lower androgen concentrations due to the emergence of heterogeneous populations of cell types in the tumor microenvironment.

Overall, the results from all three groups indicate that PTEN deletion generally leads to elevated net cell proliferation and an escape from androgen dependence. We are aware that

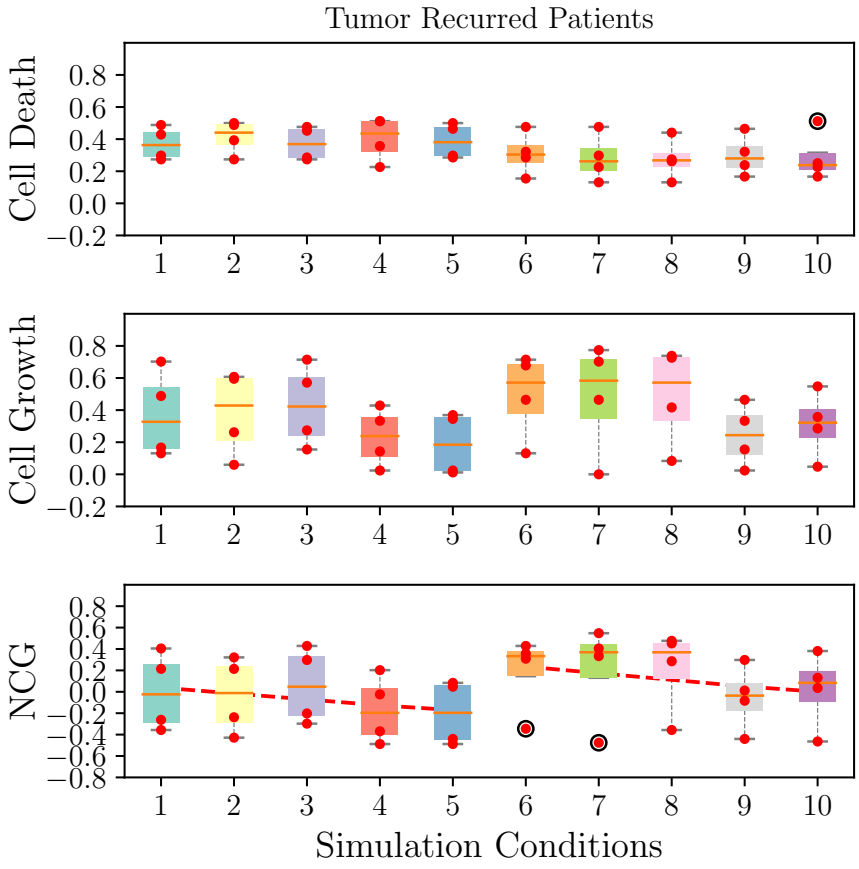


Figure 30: Calculations of cell kill, cell growth and net cell growth probabilities for patients with tumor recurrence. The ten conditions (1 to 10) are the conditions summarized in Table 4 (left). The four data points for each condition are the four states summarized in Table 4 (right). The boxes **1**–**5** indicate the net cell growth probabilities **without PTEN deletion** for **decreasing** testosterone stimulations 20 nM to 1 nM respectively. The boxes **6**–**10** indicate net cell growth probabilities **with PTEN deletion** for **decreasing** testosterone stimulations 20 nM to 1 nM respectively. The lower panel which shows the Net Cell Growth probabilities also includes a trendline drawn between the means to indicate their change with ADT (decreasing testosterone levels). For the PTEN normal patient the trendline is $NCG = 0.01104C_T + -0.1817$ where $C_T$ is the testosterone concentration. The least square fit parameter is $R^2 = 0.574$. For the PTEN deleted group the trendline is $NCG = 0.01236C_T + -0.0118$ and $R^2 = 0.587$.

these results are still preliminary and would require more detailed experimental validation through in-vitro and in-vivo studies to confirm these predictions. However we want to highlight the strength of such an integrated multiscale modeling framework through this example showing that not only it can detect changes in different gene/protein expression profiles between different patients and predict different outcomes, but it also allows us to capture the effect of tumor heterogeneity in a systematic way and give novel results which are worth exploring in more detail.

### *8.4.3. Effect of Perturbations in the Model Parameters on its predictions*

In the previous section, we saw that heterogeneities in the tumor environment resulted in higher deviations from the mean values of the cell fate probabilities, in particular the net cell growth probability, in patients without PTEN deletion compared to patients with PTEN deletion. We now want to explore how perturbations in model parameters like kinetic rate constants (for e.g. through mutations) can influence this deviation in the net cell growth and whether these are distinct from those due to tumor heterogeneity. To this end, we generated an ensemble of models by using Latin Hypercube Sampling to sample the parameter values from biologically relevant ranges. Then we ran the simulations for each member of the ensemble and calculated the normalized deviation (described in Methods) of the NCG for each member of the ensemble with respect to the base. This normalized deviation is plotted in Figure 31 for both control patients and patients with recurrence for different PTEN initial expressions, EGF and testosterone levels.

Two important observations we can immediately make are:

1. The average deviation is higher at normal PTEN for both groups of patients.
2. The average deviation is higher for patients with tumor recurrence than the control group.

Both these observations were made for the base set of parameters as well (Figures 28, 29 and

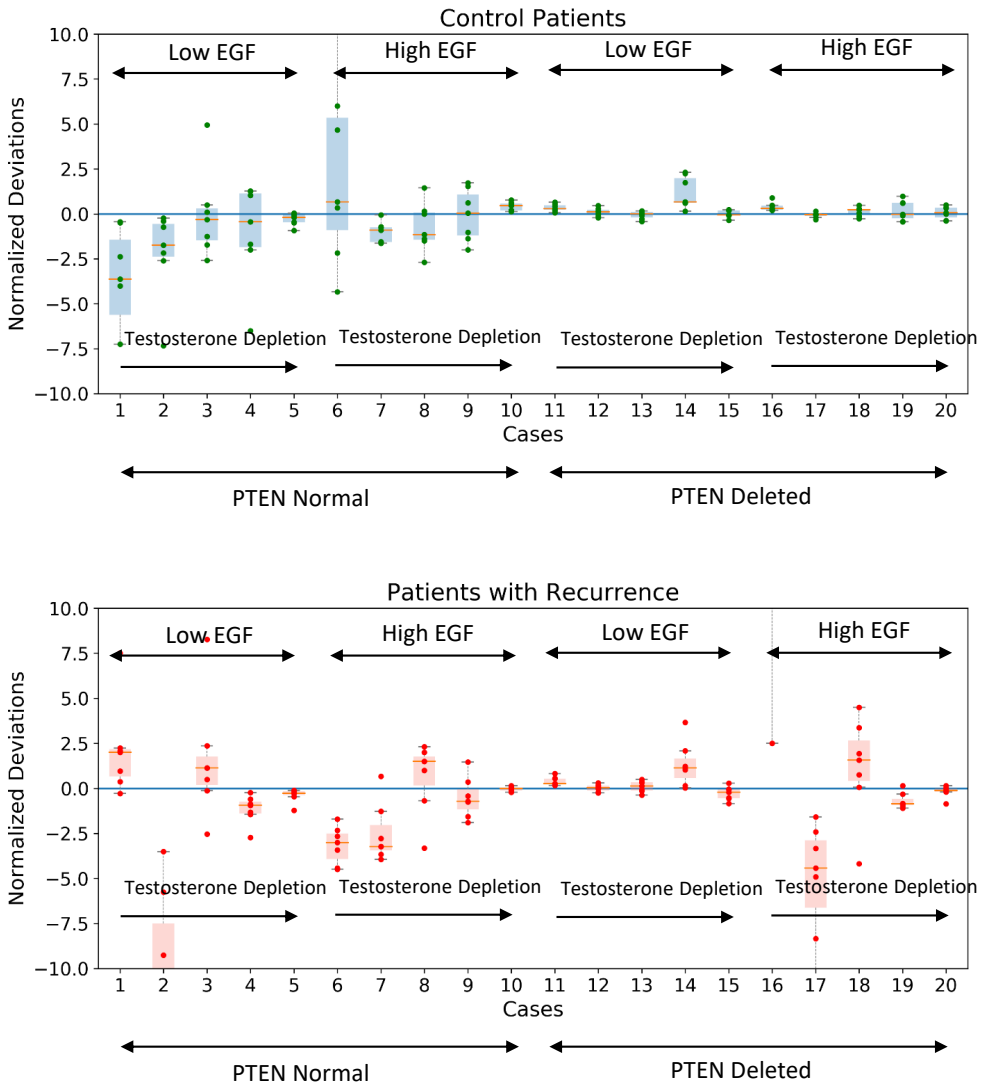


Figure 31: The normalized deviation (ND) of net cell growth probability in control patients and patients with tumor recurrence for different PTEN, androgen (testosterone) and growth factor (EGF) levels. The line through zero deviation represents the base set of parameters. $ND > 0$ indicates a higher NCG in the sample and $ND < 0$ indicates lower NCG in the sample compared to base.

30) This tells us that these variations are attributable to differences in expressions of PTEN and other genes/proteins which differentiate the control group from patients with recurrence. We also see some differences under high and low growth factors for both groups of patients and a testosterone dependent increase in the deviations. These observations suggest that one of the ways tumor cells can evade androgen deprivation is by controlling the variations due to heterogeneities in the tumor microenvironment through random perturbations in parameters such as kinetic rate constants which could be altered due to mutations from genomic instabilities when a tumor cell is subjected to external stress. If there are higher fluctuations in cell cycle activity at low androgen conditions that means a part of the tumor population can evade therapy and persist which can eventually lead to recurrence.

## 8.5. Discussion

In this work, we have introduced a heterogeneous and multiscale modeling framework to combine two or more systems models taking into account the differences in their time scales and time descriptions. This framework allows us to combine a receptor-based cell signaling pathway modeled by ordinary differential equations with a cell cycle pathway modeled by boolean/piecewise-linear differential equations and predict the combined outcome as a cell fate probability. This framework can couple multiple models together by defining interfaces containing common proteins of the component networks and takes advantage of:

a) Large separation in the time scales which results in the model with a larger time scale (macroscopic model) detect only changes in the steady-state concentrations of the model with a shorter time scale (microscopic model).

b) Sigmoidal activity of the interface proteins which allows us to convert a continuous concentration-time profile into discrete one and vice-versa.

A key advantage of such an integrated framework is that it can incorporate patient-specific mRNA/protein expression data and tumor microenvironment heterogeneities and determine their effect on cell fate decisions. We demonstrated the utility of such a framework

by applying it to prostate cancer patient data and predicting the effect of Androgen Deprivation Therapy (ADT). This model was built by combining Ras-MAPK and PI3K/AKT pathway, Androgen Receptor signaling pathway with p53 driven cell cycle pathway. Our model combination was able to successfully distinguish patients with cancer recurrence from the control patients. We applied this to a population of prostate cancer patient data from TCGA where we first classified the patients into groups based on the presence or absence of PTEN deletion mutation and response to androgen deprivation therapy. The model showed that PTEN deleted patients show poor response to ADT as reflected in higher average net cell growth probability which is consistent with the literature. A novel prediction of the model was about the variability induced due to tumor microenvironment heterogeneity which was incorporated into the model by varying the growth factor levels and memory conditions to mimic cell cycle time variations. The model could not distinguish between a control group of patients and patients showing biochemical recurrence (BR) based on PSA values. The fact that BR may not reflect the effective level of the disease leads to some speculations. Namely, since BR is determined by measuring serum PSA levels, the development of less PSA producing tumor clones would escape detection. As a matter of fact, castration-resistant PCa (CRPC), which may develop also in consequence of testosterone depletion, shows a complex PSA dynamics not adequately monitored by BR (114). As already pointed out in (115) the timing of the BR, in addition to its occurrence, may be useful to improve the prediction of PCa recurrence. The model also predicted a difference in the variabilities in the NCG of patients who had tumor recurrence (measured through biopsies) from the control group. In particular, it showed a bimodality in the NCG values where a population of cells had a consistently high value compared to others. This difference was traced to a difference in the dynamics of AKT which is an important component in all three signaling pathways. Lastly, we also investigated how variations in model parameters such as kinetic rate constants of the reactions affected the model outcome by generating an ensemble of models and calculating Normalized Deviation of Net Cell Growth (NCG). This showed similar changes in the variability suggesting a possible strategy that cancer cells

may adopt by producing fluctuations in these rate constants through mutations induced by genetic instability and thereby producing variabilities in outcomes where a part of the population will escape the effect of therapy and can lead to recurrence. Although these results are preliminary and need more validation through experiments (both in-vivo and in-vitro), they give us some interesting insights on how tumor microenvironment heterogeneity can influence the response of different patients. This also addresses an important shortcoming of the existing systems models where the differences in time scales, effect of tumor microenvironment heterogeneity and variations in the model parameters are not considered thereby significantly limiting the scope of the model predictions. Our framework provides a powerful way to a couple of different models and allows one to perform a more integrated and patient-specific exploration of the effect of treatments and designing an optimum treatment strategy.

# CHAPTER 9 : HETEROGENEOUS, MULTISCALE MODEL FOR NEPHROBLASTOMA

## 9.1. Introduction

Nephroblastoma is currently the most common childhood renal cancer and the fifth most common pediatric malignancy, representing approximately 6 % of all pediatric cancer (116). Treatment protocols vary by geographical location but are usually a combination of surgery and chemotherapy with radiation therapy also used for high-risk patients. This cancer is also a success story for modern medicine as substantial progress has been made by refining the risk stratification and using the existing chemotherapy schedule which has improved the overall survival from 30 % to greater than 90 % for localized disease and 75 % for metastatic disease (117; 118). There are two major groups that have studied the optimal management of nephroblastoma which is Children's Oncology Group (COG) and International Society of Paediatric Oncology (SIOP). The main difference between the approach of treatment for these groups is that COG recommends surgery before any adjuvant therapy whereas SIOP favors preoperative chemotherapy for all cases except very young infants (119). In the SIOP classification (which is relevant for this study) the tumors are classified and risk-stratified based on a percentage of blastemal cells (in a population of blastemal, epithelial and stromal cells) which are distinguished only histologically from the remaining cell population. Survival of a high proportion of blastemal cells and/or presence of diffuse anaplasia are considered high risk. The risk stratification and treatment are entirely based on imaging and no biopsies are done. So there is an urgent need to identify molecular biomarkers and the mechanisms that regulate the disease in high-risk variants and better optimize the treatment to prevent toxicity for the low-risk group.

MicroRNAs have been identified as important molecular markers in many diseases (120). MicroRNAs are small non-coding RNAs that negatively regulate gene expression at the post-transcriptional level. They play a significant role in many biological processes and their

aberrant expression has been found to play a key role in cancers of the lung, breast and gastric (121; 122). An increasing number of recent studies have identified miRNA playing an important part in nephroblastoma. One recent study (123) found fourteen miRNAs that were differentially expressed in nephroblastoma compared to control and two of them were present independent of subtypes or amount of sample.

Apart from miRNAs, mutations other proteins such as tumor suppressor p53 have been shown to have a significant correlation with diffuse anaplasia and survival suggesting its role as a prognostic biomarker for poor outcome and unfavorable histology (124; 125). The main tumor suppressor protein that has been linked to the etiology of nephroblastoma is WT1 Wilm's Tumor suppressor. It has been shown that the hormones IGF-1 regulate the transcription of WT1 through the MAPK pathway and reduced level of WT1 facilitates IGF-1 mediated cell cycle progression (126).

In the light of these findings, it has become clear that we need a detailed mechanistic understanding of nephroblastoma, which signaling pathways are dysregulated and how they impact cell cycle. Another important question is how the particular miRNA, mRNA and protein expression profiles of a particular patient determine the outcome of chemotherapy and radiotherapy. Addressing these questions will not only help us find optimized therapies for the subset of high-risk patients but also would help us identify biomarkers that can improve the risk stratification providing benefits to a large group of patients.

This is certainly an ideal scenario to use our hybrid multiscale modeling framework that can not only incorporate patient-specific expression data but can simulate the effect of a single or a combination of drugs on the patient. We have built a multiscale model for nephroblastoma in this work and used it with patient miRNA expression data, prescribed drug dosages and predicted the response.

## 9.2. Description of Biological Pathways

As mentioned in the introduction some of the key pathways that are frequently altered in the more aggressive forms of nephroblastoma are p53 driven pathway and MAPK pathway. To illustrate the interaction between such pathways we chose two models for our hybrid simulator. The first one is a continuous-time ODE model of receptor drive Ras-MAPK pathway and PI3K/AKT and the other is a boolean model of p53 driven DNA damage repair and cell cycle pathways. These models were adapted from the published model in the literature (7; 10). Unlike the prostate hypermodel in the previous chapter, here we are mainly focusing on the p53 driven cell cycle pathway. Our objective is mainly to understand whether and how a receptor-mediated cell signaling pathway which typically operates at much faster time scales, can influence the cell fate decisions of this pathway in the presence and absence of drugs. The exact components of the faster pathway (the type of receptor, MAPK etc) is not important as we mainly want to ascertain how the evolution of a slower cell cycle model to a specific steady state (which is indicative of a particular cellular fate) is changed due to events that happen at much faster scale. For e.g., we want to find how events such as receptor activation, internalization/recycling and even steady-state activities of proteins like ERK or AKT which have time scales between few minutes to 4 h to 6 h can affect events like G1-S transition or apoptosis through activation of caspases which happen at time scales of 24 h to 48 h. Since this slower pathway can only sense steady-state changes in the faster pathway and so it smooths out a lot of transient behavior, we do not need the faster pathway to be very specific or with tightly constrained parameters. With this in mind, we integrate these two models in exactly the same manner described in the previous section. The input to the model comes from the miRNA expression data for different patients and drug dosage information. In the next sections, we describe how these are incorporated into the model.

## 9.3. Using miRNA expression data

As was mentioned in 9.1, micro-RNAs are short non-coding RNAs that regulate gene expression post-transcriptionally by either inhibiting translation by binding to the 3' UTR of target genes or by promoting mRNA degradation. Tumor-derived miRNA circulating in the serum or plasma has significant prognostic value because of multiple reasons (127)

1. miRNA expression is frequently dysregulated in cancer
2. Expression patterns of miRNA are tissue-specific
3. They are bound to protein species in the serum/plasma or are transported in exosomes/microvesicles which makes them very stable.

It has been shown recently (123) miRNAs are significantly overexpressed in nephroblastoma. For this work, we obtained pre and post-operative serum and tissue miRNA levels from patients of nephroblastoma. We were able to utilize this expression data by identifying the specific targets of the top 20 miRNAs which were overexpressed relative to control using miRTarBase (128). For those mRNAs which are present in our network, we decreased the initial expression level by a set amount before each model run. For the ODE model, this corresponded to a decrease in the initial concentrations of the target species while for the Boolean these constrained the target nodes as explained in section 7.2.3.

Since each model was thus "personalized" using specific expression data of the patient, the predictions were specific to those groups of patients who shared similar initial expression profiles. The molecular hypermodel takes as an input, the miRNA profiles of a given patient from the CHIC data repository and then uses a database mapping (miRTarBase) to map the enriched miRNA in either tissue or serum of a patient to the corresponding mRNA.

## 9.4. Effect of Treatment

### *9.4.1. Chemotherapy*

In the SIOP protocol (129), the standard pre-operative treatment for localized tumor is 4 weeks of Vincristine + Actinomycin and for metastatic tumor 6 weeks of Vincristine + Actinomycin + Doxorubicin. Additional dosages are prescribed post-surgery depending on the risk stratification. For each of these drugs there exist literature published cell kill rates that are uniformly applied for all patients. Our aim is to calculate an adjusted cell kill rate that takes into consideration the probability of cell kill predicted by our hybrid model initialized with patient-specific expression data.

To do this we assume that the effect of the drug on cell survival follows a Poisson distribution so that the fraction of cells killed $R$ is given by $R = 1 - e^{-kt}$ where $k$ is the rate constant and is proportional to the cell kill probability. Using subscripts *lit* and *adj* for literature and adjusted cell kill rates we get

$$\begin{aligned} R_{adj} &= 1 - e^{-k_{adj}t} \\ R_{lit} &= 1 - e^{-k_{lit}t} \end{aligned} \tag{9.1}$$

Then the cell kill rates $R_{adj}$ and $R_{lit}$ a related through the following equation:

$$\frac{k_{adj}}{k_{lit}} = \frac{\ln\left(1 - R_{adj}\right)}{\ln\left(1 - R_{lit}\right)} \tag{9.2}$$

The ratio $k_{adj}/k_{lit}$ is obtained from simulation for a patient and control where control indicates no miRNA based initialization of the model.

For a combination of drugs, we assume additivity of rate constants (probabilities) instead of additivities in cell kill rates which is commonly used in literature.

So if $R_1$ and $R_2$ are the cell kill rates for two different drugs additivity of rate constants give us the following relation

$$\ln(1 - R_1) + \ln(1 - R_2) = -(k_1 + k_2)t = \ln(1 - R_{1+2}) \tag{9.3}$$

From this, we see that the survival fraction of the two-drug combination is the product of the individual.

## 9.5. Results

For each patient selected we had the following data

- miRNA expression.
- Prescribed chemotherapeutic drug dosage and schedule.
- Tumor volume pre and post-therapy.

For each patient, we ran the hybrid model after initializing the model using patient miRNA expression data using the procedure mentioned in the previous section. We do not use only the base model for our calculations. We create an ensemble of models by choosing different members of the initial state space of the Boolean model (p53 driven cell cycle model). If the Boolean model has $N$ nodes and of them $M < N$ nodes are constrained (which means their activities are set by the effect of drugs, miRNA or interaction with the other model) then the initial state space is generated from the $N - M$ nodes. For a boolean model with two possible states for every node, there are $2^{N-M}$ possible states which form the initial state space. In general, this number is too large for our boolean model which has more than 25 nodes with about 10-15 unconstrained nodes. So we select a subset of this initial state space by randomly sampling $K$ initial states out of $N - M$. To take into consideration the effect of tumor heterogeneity, we follow a similar approach as in prostate cancer in the previous chapter. We select two different growth factor concentrations and memory conditions (with

and without memory). This creates an ensemble of size $4 \times 2^K$. Cell fate is reported at every step of the model run and there are a total of $N_{step}$. So the final cell fate probabilities are calculated over $4 \times 2^k \times N_{step}$ different simulations. The result is presented as a box plot showing the average and distribution of the points around mean. This allows us to make a meaningful comparison between the groups.

We ran our simulation for a control patient and three different nephroblastoma patients (with pseudo-anonymized ids as indicated in the figures) Based on the SIOP protocol of treatment the simulations were run with drugs a) Vincristine + Actinomycin b) Doxorubicin and c) All three together.

In Fig 32 and Fig 33 the model predicted cell growth (proliferation), cell death (apoptosis) and net cell growth which is the difference between cell growth and cell death probability is plotted for a control patient (no cancer) and three other nephroblastoma patient (with pseudo-anonymized ids 4L3YB6, 5XIHQG and 6Z34IQ).

In Fig 32 looking at the mean cell fate probabilities (black lines) we see there is a clear correlation with drug dosage. Both the control group and nephroblastoma patient 4L3YB6 has a similar proliferation rate in the absence of the drugs. However, they show a different response to chemotherapeutic drugs and the nephroblastoma patient shows a more pronounced response to the combination of Vincristine+Actinomycin and Doxorubicin than either one alone. The mean cell fate probabilities are however incomplete to understand the effect of these drugs. Tumor cells in-vivo are subjected to a wide variety of conditions due to changes in the extracellular environment and factors like cell cycle time. We mentioned how we incorporate this in a very simplistic fashion by considering two different growth factor (EGF) concentration and cell memory conditions. Even in this limited set, we see considerable differences in the variability of the cell fates both for a single patient under different drug combinations and across different patients. To see the effect of the same drug combination on other patients of nephroblastoma we ran the simulations for two other patients (5XIHQG and 6Z34IQ). These patients again show a very different response to

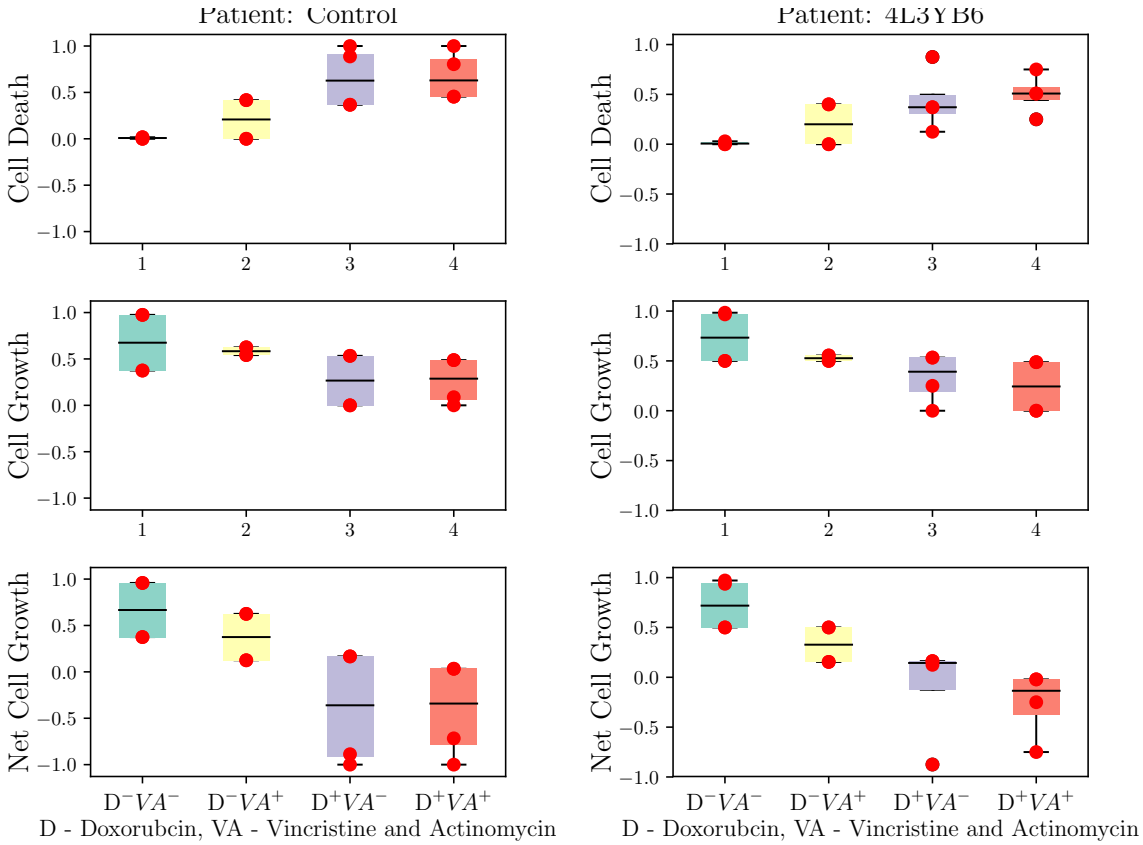


Figure 32: Model predicted cell fate probabilities: (A) Left panel shows the predicted cell death, growth and net cell growth probabilities of control patient (with no cancer) and B) Right panel shows the same parameters for a nephroblastoma patient (anonymized id 4L3YB6)

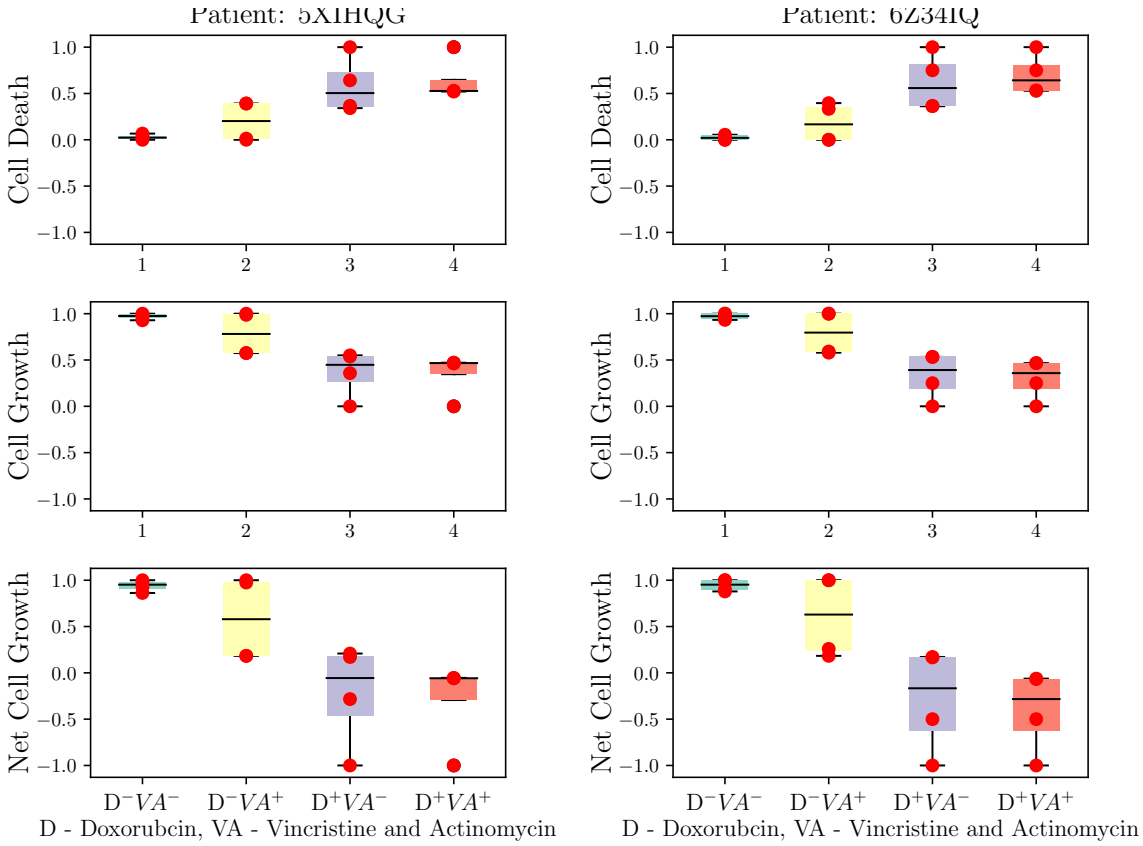


Figure 33: Model predicted cell fate probabilities: Left panel shows the predicted cell death, growth and net cell growth probabilities of patient with anonymized id 5XIHQG and the right panel shows the same parameters for another patient with id 6Z34IQ

these drug combinations in terms of both the mean cell fate values and the variability in cell fate values. Since as far as the model is concerned the only difference between these four patients comes in the miRNA and gene expression data which was used to initialize the model, these drug responses can be thought to capture the effect of these gene/miRNA expression profiles on the cellular response. This is shown in Figure 33

Finally, we wanted to find out how our predicted cell fate probabilities compared with the measured tumor volumes. For this, we plot the net cell growth parameters with the tumor volume measurement pre and post-treatment. The cell fate predictions pre and post-treatment were approximated by running the simulation for a patient with and without the drug. It is to be noted that this will not always correspond to the actual treatment the

Table 7: The patient ids, pre and post operative tumor volumes and the drugs they were given

| Patient ID | Treatment Stage | Tumor Vol (ml) | Drugs Given |
|---|---|---|---|
| 5XIHQG | pre | 78.55 | Vincristine ($1\,mg/m^2$) and |
| | post | 7.32 | Actinomycin ($650\,\mu g/kg$) |
| 6Z34IQ | pre | 754.75 | Dox ($34\,mg\,m^{-2}$), Vincristine ($1\,mg/m^2$) |
| | post | 147.68 | Actinomycin ($650\,\mu g/kg$) |
| 4L3YB6 | pre | 287.2 | Dox ($34\,mg\,m^{-2}$), Vincristine ($1\,mg/m^2$) |
| | post | 37.48 | Actinomycin ($650\,\mu g/kg$) |

patients received which is normally starting with Vincristine and Actinomycin and then following up with Doxorubicin or combination at different dosage intervals (from days to weeks). Handling such general cases is beyond the scope of this simple model. Instead, we calculate the predicted reduction in net cell growth probability for different drugs and compare them with the actual tumor volume measurements. So individually these may not represent a real drug dosing schedule but together they can help us understand how a particular patient will respond to individual or combination and can give us a starting point to designing optimum combination therapies. The table below shows the tumor volumes and the net cell growth for the patient pre and post-treatment

Using Table 7 we calculate the reduction in tumor volume for each patient and then plot it against the corresponding decrease in net cell growth probability predicted for the patient. This is shown in Figure 34.

The Figure 34 shows that we get a monotonic and almost quadratically increasing correlation between the predicted change in net cell growth and a corresponding change in tumor volume. This gives us some confidence in the model predictions particularly since it was able to differentiate the patient which showed a maximum response to the drug.

We also predicted the net cell growth change for other drug combinations for each patient and compared them with those in SIOP protocol. These are summarized in Table 8.

The table shows that for the patient 5XIHQG using Dox can significantly improve the reduce

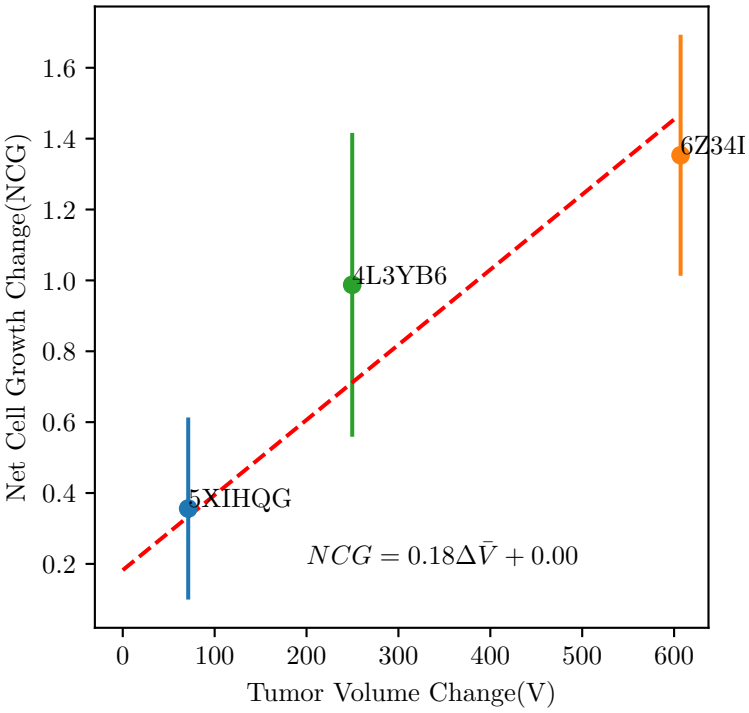


Figure 34: The figure shows the changes in predicted net cell growth for the three patient vs their observed change in tumor volumes. These change quantities are calculated between pre and post treatment values so a positive value indicates a reduction in these quantities and hence a positive outcome of treatment

the net cell cell growth. It also shows that the combination of all three drugs produces only a slight improvement over Dox alone.

It has to be mentioned that tumor volume data for only three patients is not sufficient for making any specific conclusions. The purpose of this was to show how our mechanistic model can incorporate the effect of the different drugs alone or in combination and compare them with clinical data. As a future extension of this project, this model will be applied to a larger population of patients to make more definitive predictions.

## 9.6. Discussion

The present work we describe an integrated cellular framework to model key cell-signaling pathways operating at different time scales — a well-recognized challenge in the field. Here

Table 8: The predicted changes in net cell growth for the three patients for different drug combinations. A positive value indicates that there was a net reduction in net cell growth after drug treatment.

| Patient | Vincristine & Actinomycin | Dox | Dox, Vincristine & Actinomycin |
|---|---|---|---|
| 5XIHQG | **0.36 (±0.13)** | 1.17 (±0.19) | 1.23 (±0.18) |
| 4L3YB6 | 0.40 (±0.205) | 0.83 (±0.245) | **0.99 (±0.205)** |
| 6Z34IQ | 0.34 (±0.125) | 1.24 (±0.19) | **1.35 (±0.17)** |

we model the ErbB receptor-mediated Ras-MAPK and PI3K/AKT pathway and integrate it with p53 mediated DNA damage response pathway to obtain a cell kill rate under specific drug dosing and patient-specific miRNA expression levels. The obtained cell kill rate was directly used as an input to phenomenological tumor growth models. The aim of such an integrated molecular model is to provide a mechanistic foundation to the more empirical models used in the field to obtain cell kill and growth rates under particular dosage conditions. The integration of the models was accomplished by identifying model interfaces and passing information between runs of the two models. This framework has been tested for the lung cancer and the nephroblastoma scenarios and the future work will be focused on performing a detailed sensitivity analysis to simulate the inherent tumor heterogeneity and also the effect of various mutations and subject the framework to clinical validation.

# Part IV

# Conclusion and Future Work

# CHAPTER 10 : FUTURE WORK

## Cell Mechanics and Exosomes in Systems Models

Hepatocellular carcinoma is one of the leading cause of cancer related deaths in the world. Extracellular matrix stiffness has been strongly correlated to proliferative outcomes (130). The extracellular matrix is increasingly recognized to be a key regulator of the hallmarks of cancer (131). One of the main ways cells can sense changes in stiffness of the extracellular matrix is through proteins like Integrins which serve as mechano-chemical signal converters and thus couple mechanical and biochemical signaling networks.

Recent experimental results point to two distinct (and possibly complementary) mechanisms of mechanotransduction in HCC. Signaling outcome of receptors like EGFR are strongly dependent on the stability of the dimer (less stable dimers elicits a sustained response) (132). ECM stiffness can influence this stability.

It has been also been observed that increased stiffness in HCC cells promotes secretion of exosomes (133). Exosomes are 30-100 nm extracellular vesicles produced by intraluminal budding of multivesicular bodies (MVB) and fusion of MVB and plasma membrane. They have been implicated in promoting tumor growth by changing the microenvironment in different cancers (134). In Huh7 cells, increased stiffness have been found to increase exosome secretion through AKT. This increased exosome secretion is hypothesized to change the ligand composition of the ECM to affect cellular outcomes.

There may be multiple connections between the two proposed method.

a) Recent work (135) in Cornell in breast cancer cell lines show that Sirt1 a member of Sirtuin family of NAD dependent deacetylases and deacylases, plays an important role in lysosomal degradation of ubiquitinated EGFR. Decreased Sirt1 activity may inhibit EGFR degradation leading to more recycling and increased time at the cell surface

b) Significant cross-talk between Integrin and EGFR signaling pathways have been identified

in literature both downstream and more recently upstream through Src family kinases (136)

Based on the above literature evidence here are the main questions we want to answer:

- How does stiffness dependent activation and stabilization of cell surface receptors like Integrins influence Akt activation?
- How does the positive feedback through Akt mediated exosome secretion and change in ligand composition influence the AKT-Rabin8-Rab8 pathway?
- How does AKT-Rabin8-Rab8 mediated exosome secretion and resultant inhibition of receptor degradation affected signaling through RTKs like EGFR?
- Can EGFR-Integrin interaction influence exosome secretion?

Integrin mediated mechanotransduction typically occurs in multiple steps. These are listed below

a) Receptor activation: Mechanical forces first cause a change in Integrin conformation from inactive to active. The forward and backward rate constants $k_f$ and $k_r$ are proportional to the corresponding probabilities and the ratio of these probabilities are given by ($E_T$ is energy of barrier and $E_i$ is the energy of inactive configuration

$$\frac{k_f}{k_r} = e^{\frac{E_T - E_i}{k_B T}} = e^{-\frac{E_{ac}}{k_B T}} \tag{10.1}$$

Using a linear approximation of the elastic behavior of the substrate the activation energy can be shown to be is force required to switch conformation and $k$ is the effective stiffness (137)

$$E_{ac} = \frac{f_b^2}{2k} \tag{10.2}$$

The effective stiffness $k$ is related to substrate stiffness $k_s$ and the receptor stiffness $k_r$ as below. These individual stiffnesses can be calculated from mechanical properties like Young's modulus and Poisson's ratio and also obtained from molecular dynamics simulations.

$$1/k_f = 1/k_r + 1/k_s \tag{10.3}$$

b) Ligand Receptor Binding: The next step is binding of activated receptor with ligand on the substrate. The off and rebinding rates can both be substrate dependent. For slip bonds the dissociation rate increases with force according to the following relation

$$k_{off} = k^*_{off} e^{-f/f_\beta} \tag{10.4}$$

Here $f_\beta = k_B T/x_\beta$ is the intrinsic force scale where $x_\beta$ is the displacement to dissociate the bond.

The overall model schematic is shown in Figure 35. Here the receptors (Integrins) are activated and bound to ligands in a stiffness dependent manner. On activation and clustering with other receptors it can activate AKT through a downstream signaling pathway directly or indirectly through EGFR (138). AKT then increases secretion of exosomes through Rabin8-Rab8 pathway. We will consider both cases where:

1. Through release of exosomes the extracellular ligand concentration is altered.

2. Through increases receptor recycling and less degradation the net number of cell surface receptors are increased.

The AKT activation and increased exosome secretion will be expressed as a function of substrate stiffness.

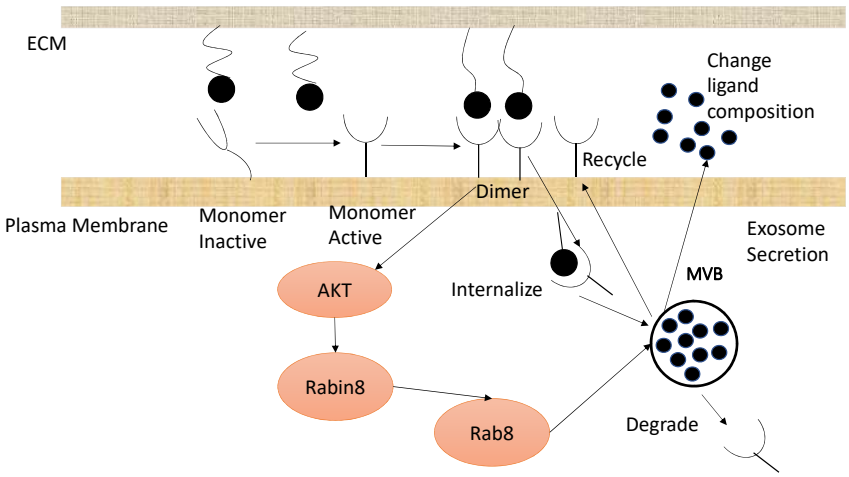


Figure 35: Schematic of the proposed stiffness dependent AKT activation and increased exosome secretion in HCC

# APPENDIX

# BIBLIOGRAPHY


[1] PK Maini and ARA Anderson. Mathematical oncology. *Bulletin of Mathematical Biology*, 80(5), 2018.

[2] Ibiayi Dagogo-Jack and Alice T Shaw. Tumour heterogeneity and resistance to cancer therapies. *Nature reviews Clinical oncology*, 15(2):81, 2018.

[3] Laura J Van't Veer and René Bernards. Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature*, 452(7187):564, 2008.

[4] Dominique Barbolosi, Joseph Ciccolini, Bruno Lacarelle, Fabrice Barlési, and Nicolas André. Computational oncology—mathematical modelling of drug regimens for precision medicine. *Nature reviews Clinical oncology*, 13(4):242, 2016.

[5] Harsh Vardhan Jain, Steven K Clinton, Arvinder Bhinder, and Avner Friedman. Mathematical modeling of prostate cancer progression in response to androgen ablation therapy. *Proceedings of the National Academy of Sciences*, 108(49):19701–19706, 2011.

[6] Marc R Birtwistle, Mariko Hatakeyama, Noriko Yumoto, Babatunde A Ogunnaike, Jan B Hoek, and Boris N Kholodenko. Ligand-dependent responses of the erbb signaling network: experimental and modeling analyses. *Molecular systems biology*, 3(1), 2007.

[7] William W Chen, Birgit Schoeberl, Paul J Jasper, Mario Niepel, Ulrik B Nielsen, Douglas A Lauffenburger, and Peter K Sorger. Input–output behavior of erbb signaling pathways as revealed by a mass action model trained against dynamic data. *Molecular systems biology*, 5(1), 2009.

[8] Jia-Yun Chen, Jia-Ren Lin, Karlene A Cimprich, and Tobias Meyer. A two-dimensional erk-akt signaling code for an ngf-triggered cell-fate decision. *Molecular cell*, 45(2):196–209, 2012.

[9] Michael J Lee, S Ye Albert, Alexandra K Gardino, Anne Margriet Heijink, Peter K Sorger, Gavin MacBeath, and Michael B Yaffe. Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell*, 149(4):780–794, 2012.

[10] Minsoo Choi, Jue Shi, Sung Hoon Jung, Xi Chen, and Kwang-Hyun Cho. Attractor landscape analysis reveals feedback loops in the p53 network that control the cellular response to dna damage. *Sci. Signal.*, 5(251):ra83–ra83, 2012.

[11] James Southern, Joe Pitt-Francis, Jonathan Whiteley, Daniel Stokeley, Hiromichi Kobashi, Ross Nobes, Yoshimasa Kadooka, and David Gavaghan. Multi-scale computational modelling in biology and physiology. *Progress in biophysics and molecular biology*, 96(1-3):60–89, 2008.

[12] Filippo Castiglione, Francesco Pappalardo, Carlo Bianca, Giulia Russo, and Santo Motta. Modeling biology spanning different scales: an open challenge. *BioMed research international*, 2014, 2014.

[13] Peter J Hunter and Thomas K Borg. Integration from proteins to organs: the physiome project. *Nature reviews Molecular cell biology*, 4(3):237, 2003.

[14] Maria Secrier and Reinhard Schneider. Visualizing time-related data in biology, a review. *Briefings in bioinformatics*, 15(5):771–782, 2013.

[15] Hartmut Derendorf and Bernd Meibohm. Modeling of pharmacokinetic/pharmacodynamic (pk/pd) relationships: concepts and perspectives. *Pharmaceutical research*, 16(2):176–185, 1999.

[16] Haluk Resat, Linda Petzold, and Michel F Pettigrew. Kinetic modeling of biological systems. In *Computational Systems Biology*, pages 311–335. Springer, 2009.

[17] Boris N Kholodenko. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *European journal of biochemistry*, 267(6):1583–1588, 2000.

[18] Nick I Markevich, Jan B Hoek, and Boris N Kholodenko. Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *The Journal of cell biology*, 164(3):353–359, 2004.

[19] Birgit Schoeberl, Emily A Pace, Jonathan B Fitzgerald, Brian D Harms, Lihui Xu, Lin Nie, Bryan Linggi, Ashish Kalra, Violette Paragas, Raghida Bukhalid, et al. Therapeutically targeting erbb3: A key node in ligand-induced activation of the erbb receptor–pi3k axis. *Sci. Signal.*, 2(77):ra31–ra31, 2009.

[20] Herbert M Sauro. Network dynamics. In *Computational Systems Biology*, pages 269–309. Springer, 2009.

[21] Hidde De Jong. Modeling and simulation of genetic regulatory systems: a literature review. *Journal of computational biology*, 9(1):67–103, 2002.

[22] El Houssine Snoussi and Rene Thomas. Logical identification of all steady states: the concept of feedback loop characteristic states. *Bulletin of Mathematical Biology*, 55(5):973–991, 1993.

[23] Leon Glass and Stuart A Kauffman. The logical analysis of continuous, non-linear biochemical control networks. *Journal of theoretical Biology*, 39(1):103–129, 1973.

[24] Leon Glass and Rafael Pérez. Limit cycle oscillations in compartmental chemical systems. *The Journal of Chemical Physics*, 61(12):5242–5249, 1974.

[25] Zhike Zi. Sensitivity analysis approaches applied to systems biology models. *IET systems biology*, 5(6):336–346, 2011.

[26] Sobol IM. Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp*, 1(4):407–414, 1993.

[27] Andrea Saltelli, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto, and Stefano Tarantola. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010.

[28] Russel E Caflisch. Monte carlo and quasi-monte carlo methods. *Acta numerica*, 7:1–49, 1998.

[29] Ilya M Sobol. Uniformly distributed sequences with an additional uniform property. *USSR Computational Mathematics and Mathematical Physics*, 16(5):236–242, 1976.

[30] Jonathan D Herman and Will Usher. Salib: An open-source python library for sensitivity analysis. *J. Open Source Software*, 2(9):97, 2017.

[31] John D Hunter. Matplotlib: A 2d graphics environment. *Computing in science & engineering*, 9(3):90, 2007.

[32] Bryan C Daniels, Yan-Jiun Chen, James P Sethna, Ryan N Gutenkunst, and Christopher R Myers. Sloppiness, robustness, and evolvability in systems biology. *Current opinion in biotechnology*, 19(4):389–395, 2008.

[33] Ryan N Gutenkunst, Joshua J Waterfall, Fergal P Casey, Kevin S Brown, Christopher R Myers, and James P Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS computational biology*, 3(10):e189, 2007.

[34] Hiroaki Kitano. Towards a theory of biological robustness. *Molecular systems biology*, 3(1), 2007.

[35] Stefano Ciliberti, Olivier C Martin, and Andreas Wagner. Robustness can evolve gradually in complex regulatory gene networks with varying topology. *PLoS computational biology*, 3(2):e15, 2007.

[36] Dorjsuren Battogtokh, David K Asch, Mary E Case, Jonathan Arnold, and H-B Schüttler. An ensemble method for identifying regulatory circuits with special reference to the qa gene cluster of neurospora crassa. *Proceedings of the National Academy of Sciences*, 99(26):16904–16909, 2002.

[37] Bin Wu, Alokendra Ghosh, Ravi Radhakrishnan, and Wei Guo. Spatial control of rab11 activation conferred by a gef and gap regulatory loop during exocytosis. *Proceedings of the National Academy of Sciences*, in preparation.

[38] Stenmark H. Rab gtpases as coordinators of vesicle traffic. *Nat Rev Mol Cell Biol*, 2009.

[39] Hutagalung A.H and Novick P.J. Role of rab gtpases in membrane traffic and cell physiology. *Physiol Rev*, 2011.

[40] Mizuno-Yamasaki E, Rivera-Molina F, and Novick P. Gtpase networks in membrane traffic. *Annu Rev Biochem*, 2012.

[41] Suzanne R Pfeffer. Rab gtpase regulation of membrane identity. *Current opinion in cell biology*, 25(4):414–419, 2013.

[42] Rink J, Ghigo E, Kalaidzidis Y, and Zerial M. Rab conversion as a mechanism of progression from early to late endosomes. *Cell*, 2005.

[43] Ortiz D, Medkova M, Walch-Solimena C, and Novick P. Ypt32 recruits the sec4p guanine nucleotide exchange factor, sec2p, to secretory vesicles; evidence for a rab cascade in yeast. *J Cell Biol*, 2002.

[44] Knodler A, Feng S, Zhang J, Zhang X, Das A, Peränen J, and Guo W. Coordination of rab8 and rab11 in primary ciliogenesis. *Proc Natl Acad Sci U S A*, 2010.

[45] Rivera-Molina F.E and Novick P.J. A rab gap cascade defines the boundary between two rab gtpases on the secretory pathway. *Proc Natl Acad Sci U S A*, 2009.

[46] Suda Y, Kurokawa K, Hirata R, and Nakano A. Rab gap cascade regulates dynamics of ypt6 in the golgi traffic. *Proc Natl Acad Sci U S A*, 2013.

[47] Novick P. Regulation of membrane traffic by rab gef and gap cascades. *Small GTPases*, 2016.

[48] Grant B.D and Donaldson J.G. Pathways and mechanisms of endocytic recycling. *Nat Rev Mol Cell Biol*, 2009.

[49] Maxfield F.R and McGraw T.E. Endocytic recycling. *Nat Rev Mol Cell Biol*, 2004.

[50] Dabbeekeh J.T, Faitar S.L, Dufresne C.P, and Cowell J.K. The evi5 tbc domain provides the gtpase-activating protein motif for rab11. *Oncogene*, 2007.

[51] Westlake C.J., Junutula J.R., Simon G.C., Pilli M., Prekeris R., Scheller R.H., Jackson P.K., and Eldridge A.G. Identification of rab11 as a small gtpase binding protein for the evi5 oncogene. *Proc Natl Acad Sci U S A*, 2007.

[52] Bernd Binder and Hermann-Georg Holzhütter. A hypothetical model of cargo-selective rab recruitment during organelle maturation. *Cell biochemistry and biophysics*, 63(1):59–71, 2012.

[53] Stephanie Voss, Fu Li, Andreas Ratz, Matthias Roger, and Yao-Wen Wu. Spatial cycling of rab gtpase, driven by the gtpase cycle, controls rab's subcellular distribution. *Biochemistry*, 58(4):276–285, 2019.

[54] Perla Del Conte-Zerial, Lutz Brusch, Jochen C Rink, Claudio Collinet, Yannis Kalaidzidis, Marino Zerial, and Andreas Deutsch. Membrane identity and gtpase cascades regulated by toggle and cut-out switches. *Molecular systems biology*, 4(1), 2008.

[55] Boris N Kholodenko, John F Hancock, and Walter Kolch. Signalling ballet in space and time. *Nature reviews Molecular cell biology*, 11(6):414, 2010.

[56] Steve Caplan, Naava Naslavsky, Lisa M Hartnell, Robert Lodge, Roman S Polishchuk, Julie G Donaldson, and Juan S Bonifacino. A tubular ehd1-containing compartment involved in the recycling of major histocompatibility complex class i molecules to the plasma membrane. *The EMBO journal*, 21(11):2557–2567, 2002.

[57] Naava Naslavsky and Steve Caplan. Ehd proteins: key conductors of endocytic transport. *Trends in cell biology*, 21(2):122–131, 2011.

[58] James B Reinecke, Dawn Katafiasz, Naava Naslavsky, and Steve Caplan. Regulation of src trafficking and activation by the endocytic regulatory proteins mical-l1 and ehd1. *J Cell Sci*, 127(8):1684–1698, 2014.

[59] Mahak Sharma, Sai Srinivas Panapakkam Giridharan, Juliati Rahajeng, Naava Naslavsky, and Steve Caplan. Mical-l1 links ehd1 to tubular recycling endosomes and regulates receptor recycling. *Molecular biology of the cell*, 20(24):5181–5194, 2009.

[60] Katarina Hattula, Johanna Furuhjelm, Jaana Tikkanen, Kimmo Tanhuanpää, Pirjo Laakkonen, and Johan Peränen. Characterization of the rab8-specific membrane traffic route linked to protrusion formation. *Journal of cell science*, 119(23):4866–4877, 2006.

[61] Joseph T Roland, Anne K Kenworthy, Johan Peranen, Steve Caplan, and James R Goldenring. Myosin vb interacts with rab8a on a tubular network containing ehd1 and ehd3. *Molecular biology of the cell*, 18(8):2828–2837, 2007.

[62] Carl Laflamme, Gloria Assaker, Damien Ramel, Jonas F Dorn, Desmond She, Paul S Maddox, and Gregory Emery. Evi5 promotes collective cell migration through its rab-gap activity. *J Cell Biol*, 198(1):57–67, 2012.

[63] Heidi Hehnly, Chun-Ting Chen, Christine M Powers, Hui-Lin Liu, and Stephen Doxsey. The centrosome regulates the rab11-dependent recycling endosome pathway at appendages of the mother centriole. *Current Biology*, 22(20):1944–1950, 2012.

[64] Juanfei Wang, Jinqi Ren, Bin Wu, Shanshan Feng, Guoping Cai, Florin Tuluc, Johan Peränen, and Wei Guo. Activation of rab8 guanine nucleotide exchange factor rabin8 by erk1/2 in response to egf signaling. *Proceedings of the National Academy of Sciences*, 112(1):148–153, 2015.

[65] Susan Kane, Hiroyuki Sano, Simon CH Liu, John M Asara, William S Lane, Charles C

Garner, and Gustav E Lienhard. A method to identify serine kinase substrates akt phosphorylates a novel adipocyte protein with a rab gtpase-activating protein (gap) domain. *Journal of Biological Chemistry*, 277(25):22115–22118, 2002.

[66] Danièle Stalder, Emi Mizuno-Yamasaki, Majid Ghassemian, and Peter J Novick. Phosphorylation of the rab exchange factor sec2p directs a switch in regulatory binding partners. *Proceedings of the National Academy of Sciences*, 110(50):19995–20002, 2013.

[67] Patrick T Caswell, May Chan, Andrew J Lindsay, Mary W McCaffrey, David Boettiger, and Jim C Norman. Rab-coupling protein coordinates recycling of $\alpha 5\beta 1$ integrin and egfr1 to promote cell migration in 3d microenvironments. *J Cell Biol*, 183(1):143–155, 2008.

[68] Mark A Lemmon, Daniel M Freed, Joseph Schlessinger, and Anatoly Kiyatkin. The dark side of cell signaling: positive roles for negative regulators. *Cell*, 164(6):1172–1184, 2016.

[69] Lai Kuan Goh and Alexander Sorkin. Endocytosis of receptor tyrosine kinases. *Cold Spring Harbor perspectives in biology*, 5(5):a017459, 2013.

[70] Alokendra Ghosh and Ravi Radhakrishnan. Time-dependent antagonist-agonist switching in receptor tyrosine kinase-mediated signaling. *BMC bioinformatics*, 20(1):242, 2019.

[71] Ami Citri and Yosef Yarden. Egf–erbb signalling: towards the systems level. *Nature reviews Molecular cell biology*, 7(7):505, 2006.

[72] Ann D Thor, Susan M Edgerton, and Frank E Jones. Subcellular localization of the her4 intracellular domain, 4icd, identifies distinct prognostic outcomes for breast cancer patients. *The American journal of pathology*, 175(5):1802–1809, 2009.

[73] Christopher C Williams, June G Allison, Gregory A Vidal, Matthew E Burow, Barbara S Beckman, Luis Marrero, and Frank E Jones. The erbb4/her4 receptor tyrosine kinase regulates gene expression by functioning as a stat5a nuclear chaperone. *J Cell Biol*, 167(3):469–478, 2004.

[74] Ira Swameye, TG Müller, Jens Timmer, O Sandra, and Ursula Klingmüller. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proceedings of the National Academy of Sciences*, 100(3):1028–1033, 2003.

[75] Satoshi Yamada, Satoru Shiono, Akiko Joo, and Akihiko Yoshimura. Control mechanism of jak/stat signal transduction pathway. *FEBS letters*, 534(1-3):190–196, 2003.

[76] Heidi Greulich, Tzu-Hsiu Chen, Whei Feng, Pasi A Jänne, James V Alvarez, Mauro Zappaterra, Sara E Bulmer, David A Frank, William C Hahn, William R Sellers,

et al. Oncogenic transformation by inhibitor-sensitive and-resistant egfr mutants. *PLoS medicine*, 2(11):e313, 2005.

[77] Lothar Hennighausen and Gertraud W Robinson. Interpretation of cytokine signaling through the transcription factors stat5a and stat5b. *Genes & development*, 22(6):711–721, 2008.

[78] Stefan Hoops, Sven Sahle, Ralph Gauges, Christine Lee, Jürgen Pahle, Natalia Simus, Mudita Singhal, Liang Xu, Pedro Mendes, and Ursula Kummer. Copasi—a complex pathway simulator. *Bioinformatics*, 22(24):3067–3074, 2006.

[79] Michael D McKay, Richard J Beckman, and William J Conover. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[80] Shannon E Telesco. *Multiscale modeling of the ErbB receptor tyrosine kinase signaling network through theory and experiment*. PhD thesis, University of Pennsylvania, 2011.

[81] Rebecca S Muraoka-Cook, Melissa Sandahl, Carty Husted, Debra Hunter, Leah Miraglia, Shu-mang Feng, Klaus Elenius, and H Shelton Earp III. The intracellular domain of erbb4 induces differentiation of mammary epithelial cells. *Molecular biology of the cell*, 17(9):4118–4129, 2006.

[82] Rebecca S Muraoka-Cook, Melissa Sandahl, Debra Hunter, Leah Miraglia, and H Shelton Earp III. Prolactin and erbb4/her4 signaling interact via janus kinase 2 to induce mammary epithelial cell gene expression differentiation. *Molecular endocrinology*, 22(10):2307–2321, 2008.

[83] David J Riese. Ligand-based receptor tyrosine kinase partial agonists: New paradigm for cancer drug discovery? *Expert opinion on drug discovery*, 6(2):185–193, 2011.

[84] Jennifer L Gilmore, II Riese, and J David. secerbb4-26/549 antagonizes ligand-induced erbb4 tyrosine phosphorylation. *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics*, 14(11-12):589–602, 2004.

[85] Jennifer L Gilmore, Jeffrey A Scott, Zhor Bouizar, Alex Robling, Sarah E Pitfield, David J Riese, and John Foley. Amphiregulin-egfr signaling regulates pthrp gene expression in breast cancer cells. *Breast cancer research and treatment*, 110(3):493–505, 2008.

[86] Nicole E Willmarth, Andrea Baillo, Michele L Dziubinski, Kristy Wilson, David J Riese II, and Stephen P Ethier. Altered egfr localization and degradation in human breast cancer cells with an amphiregulin/egfr autocrine loop. *Cellular signalling*, 21(2):212–219, 2009.

[87] Diane E Clark, Christopher C Williams, Tamika T Duplessis, Kimberly L Moring, Amy R Notwick, Weiwen Long, William S Lane, Iwan Beuvink, Nancy E Hynes,

and Frank E Jones. Erbb4/her4 potentiates stat5a transcriptional activity by regulating novel stat5a serine phosphorylation events. *Journal of Biological Chemistry*, 280(25):24175–24180, 2005.

[88] Elena B Kabotyanski, Markus Huetter, Wa Xian, Monique Rijnkels, and Jeffrey M Rosen. Integration of prolactin and glucocorticoid signaling at the $\beta$-casein promoter and enhancer by ordered recruitment of specific transcription factors and chromatin modifiers. *Molecular Endocrinology*, 20(10):2355–2368, 2006.

[89] Elena B Kabotyanski, Monique Rijnkels, Courtneay Freeman-Zadrowski, Adam C Buser, Dean P Edwards, and Jeffrey M Rosen. Lactogenic hormonal induction of long distance interactions between $\beta$-casein gene regulatory elements. *Journal of Biological Chemistry*, 284(34):22815–22824, 2009.

[90] Anna Gambin, Agata Charzyńska, Aleksandra Ellert-Miklaszewska, and Mikołaj Rybiński. Computational models of the jak1/2-stat1 signaling. *Jak-stat*, 2(3):e24672, 2013.

[91] Walter Kolch, Melinda Halasz, Marina Granovskaya, and Boris N Kholodenko. The dynamic control of signal transduction networks in cancer cells. *Nature Reviews Cancer*, 15(9):515, 2015.

[92] Rob Phillips, Jane Kondev, Julie Theriot, and Hernan Garcia. *Physical biology of the cell*. Garland Science, 2012.

[93] John F Marko and Eric D Siggia. Stretching dna. *Macromolecules*, 28(26):8759–8770, 1995.

[94] Elisabetta Soldaini, Susan John, Stefano Moro, Julie Bollenbacher, Ulrike Schindler, and Warren J Leonard. Dna binding site selection of dimeric and tetrameric stat5 proteins reveals a large repertoire of divergent tetrameric stat5a binding sites. *Molecular and Cellular Biology*, 20(1):389–401, 2000.

[95] E Weinan, Bjorn Engquist, Xiantao Li, Weiqing Ren, and Eric Vanden-Eijnden. Heterogeneous multiscale methods: a review. *Commun. Comput. Phys*, 2(3):367–450, 2007.

[96] Ioannis G Kevrekidis, C William Gear, and Gerhard Hummer. Equation-free: The computer-aided analysis of complex multiscale systems. *AIChE Journal*, 50(7):1346–1355, 2004.

[97] Joseph O Dada and Pedro Mendes. Multi-scale modelling and simulation in systems biology. *Integrative Biology*, 3(2):86–96, 2011.

[98] Michael Hucka, Andrew Finney, Herbert M Sauro, Hamid Bolouri, John C Doyle, Hiroaki Kitano, Adam P Arkin, Benjamin J Bornstein, Dennis Bray, Athel Cornish-Bowden, et al. The systems biology markup language (sbml): a medium for represen-

tation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[99] Alokendra Ghosh, Ravi Radhakrishnan, Caterina Guiot, and Ilaria Stura. Heterogeneous, multiscale model for prostate adenocarcinoma and effect of pten deletion on androgen deprivation therapy. *Biotechnology Progress*, in preparation.

[100] Rebecca L. Siegel, Kimberly D. Miller, and Ahmedin Jemal. Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians*, 67(1):7–30, 2017.

[101] Adam Abeshouse, Jaeil Ahn, Rehan Akbani, Adrian Ally, Samirkumar Amin, Christopher D Andry, Matti Annala, Armen Aprikian, Joshua Armenia, Arshi Arora, et al. The molecular taxonomy of primary prostate cancer. *Cell*, 163(4):1011–1025, 2015.

[102] William G. Nelson, Angelo M. De Marzo, and William B. Isaacs. Prostate cancer. *New England Journal of Medicine*, 349(4):366–381, 2003.

[103] Channing J Paller and Emmanuel S Antonarakis. Management of biochemically recurrent prostate cancer after local therapy: evolving standards of care and new directions. *Clinical advances in hematology & oncology: H&O*, 11(1):14, 2013.

[104] Ayesha A Shafi, Aihua E Yen, and Nancy L Weigel. Androgen receptors in hormone-dependent and castration-resistant prostate cancer. *Pharmacology & therapeutics*, 140(3):223–238, 2013.

[105] Shunyou Wang, Jing Gao, Qunying Lei, Nora Rozengurt, Colin Pritchard, Jing Jiao, George V Thomas, Gang Li, Pradip Roy-Burman, Peter S Nelson, et al. Prostate-specific deletion of the murine pten tumor suppressor gene leads to metastatic prostate cancer. *Cancer cell*, 4(3):209–221, 2003.

[106] Tamara L Lotan, Filipe LF Carvalho, Sarah B Peskoe, Jessica L Hicks, Jennifer Good, Helen L Fedor, Elizabeth Humphreys, Misop Han, Elizabeth A Platz, Jeremy A Squire, et al. Pten loss is associated with upgrading of prostate cancer from biopsy to radical prostatectomy. *Modern Pathology*, 28(1):128, 2015.

[107] Meng-Lei Zhu and Natasha Kyprianou. Androgen receptor and growth factor signaling cross-talk in prostate cancer cells. *Endocrine-related cancer*, 15(4):841, 2008.

[108] Hui-Kuan Lin, Shuyuan Yeh, Hong-Yo Kang, and Chawnshang Chang. Akt suppresses androgen-induced apoptosis by phosphorylating and inhibiting androgen receptor. *Proceedings of the National Academy of Sciences*, 98(13):7200–7205, 2001.

[109] Zoran Culig. New possibilities for urinary molecular diagnostics. *BJU international*, 117(4):547–548, 2016.

[110] Marcus V Cronauer, Wolfgang A Schulz, Tatjana Burchardt, Rolf Ackermann, and Martin Burchardt. Inhibition of p53 function diminishes androgen receptor-mediated signaling in prostate cancer cell lines. *Oncogene*, 23(20):3541, 2004.

[111] Charles Huggins and Clarence V Hodges. Studies on prostatic cancer: I. the effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. *The Journal of urology*, 167(2 Part 2):948–951, 2002.

[112] Jason D Morken, Aaron Packer, Rebecca A Everett, John D Nagy, and Yang Kuang. Mechanisms of resistance to intermittent androgen deprivation in patients with prostate cancer identified by a novel computational method. *Cancer research*, 74(14):3673–3683, 2014.

[113] D Gabriele, F Porpiglia, G Muto, P Gontero, C Terrone, S Annoscia, D Randone, S Benvenuti, G Arena, I Stura, et al. Eureka-1 database: an epidemiogical analysis. *Minerva Urol Nefrol*, 1(1):9–15, 2015.

[114] Atsushi Mizokami, Kouji Izumi, Hiroyuki Konaka, Yasuhide Kitagawa, Yoshifumi Kadono, Kazutaka Narimoto, Takahiro Nohara, Amit K Bahl, and Mikio Namiki. Understanding prostate-specific antigen dynamics in monitoring metastatic castration-resistant prostate cancer: implications for clinical practice. *Asian journal of andrology*, 19(2):143, 2017.

[115] Ilaria Stura, Domenico Gabriele, and Caterina Guiot. A simple psa-based computational approach predicts the timing of cancer relapse in prostatectomized patients. *Cancer research*, 76(17):4941–4947, 2016.

[116] Andrew M Davidoff. Wilms tumor. *Current opinion in pediatrics*, 21(3):357, 2009.

[117] Filippo Spreafico and Franca Fossati Bellani. Wilms' tumor: past, present and (possibly) future. *Expert review of anticancer therapy*, 6(2):249–258, 2006.

[118] K Pritchard-Jones. Controversies and advances in the management of wilms' tumour. *Archives of disease in childhood*, 87(3):241–244, 2002.

[119] Elwira Szychot, John Apps, and Kathy Pritchard-Jones. Wilms' tumor: biology, diagnosis and treatment. *Translational pediatrics*, 3(1):12, 2014.

[120] Xin Yu, Zheng Li, Matthew TV Chan, and William Ka Kei Wu. The roles of micrornas in wilms' tumors. *Tumor Biology*, 37(2):1445–1450, 2016.

[121] Z Yang, Y Han, K Cheng, G Zhang, and X Wang. mir-99a directly targets the mtor signalling pathway in breast cancer side population cells. *Cell proliferation*, 47(6):587–595, 2014.

[122] Guangfeng Niu, Bin Li, Jianmin Sun, and Li Sun. mir-454 is down-regulated in osteosarcomas and suppresses cell proliferation and invasion by directly targeting c-met. *Cell proliferation*, 48(3):348–355, 2015.

[123] Nicole Ludwig, Nasenien Nourkami-Tutdibi, Christina Backes, Hans-Peter Lenhof, Norbert Graf, Andreas Keller, and Eckart Meese. Circulating serum mirnas as po-

tential biomarkers for nephroblastoma. *Pediatric blood & cancer*, 62(8):1360–1367, 2015.

[124] AJ Beniers, T Efferth, L Füzesi, B Granzen, R Mertens, and G Jakse. p53 expression in wilms' tumor: a possible role as prognostic factor. *International journal of oncology*, 18(1):133–142, 2001.

[125] Mariana Maschietto, Richard D Williams, Tasnim Chagtai, Sergey D Popov, Neil J Sebire, Gordan Vujanic, Elizabeth Perlman, James R Anderson, Paul Grundy, Jeffrey S Dome, et al. Tp53 mutational status is a potential marker for risk stratification in wilms tumour with diffuse anaplasia. *PloS one*, 9(10):e109924, 2014.

[126] Itay Bentov, Derek LeRoith, and Haim Werner. The wt1 wilms' tumor suppressor gene: a novel target for insulin-like growth factor-i action. *Endocrinology*, 144(10):4276–4279, 2003.

[127] Patrick S Mitchell, Rachael K Parkin, Evan M Kroh, Brian R Fritz, Stacia K Wyman, Era L Pogosova-Agadjanyan, Amelia Peterson, Jennifer Noteboom, Kathy C O'Briant, April Allen, et al. Circulating microrna as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences*, 105(30):10513–10518, 2008.

[128] Sheng-Da Hsu, Feng-Mao Lin, Wei-Yun Wu, Chao Liang, Wei-Chih Huang, Wen-Ling Chan, Wen-Ting Tsai, Goun-Zhou Chen, Chia-Jung Lee, Chih-Min Chiu, et al. mirtarbase: a database curates experimentally validated microrna–target interactions. *Nucleic acids research*, 39(suppl_1):D163–D169, 2010.

[129] Gordan M Vujanić, Bengt Sandstedt, Dieter Harms, Anna Kelsey, Ivo Leuschner, Jan de Kraker, and SIOP Nephroblastoma Scientific Committee. Revised international society of paediatric oncology (siop) working classification of renal tumors of childhood. *Medical and pediatric oncology*, 38(2):79–82, 2002.

[130] Jörg Schrader, Timothy T Gordon-Walker, Rebecca L Aucott, Mariëlle van Deemter, Alexander Quaas, Shaun Walsh, Daniel Benten, Stuart J Forbes, Rebecca G Wells, and John P Iredale. Matrix stiffness modulates proliferation, chemotherapeutic response, and dormancy in hepatocellular carcinoma cells. *Hepatology*, 53(4):1192–1205, 2011.

[131] Michael W Pickup, Janna K Mouw, and Valerie M Weaver. The extracellular matrix modulates the hallmarks of cancer. *EMBO reports*, page e201439246, 2014.

[132] Daniel M Freed, Nicholas J Bessman, Anatoly Kiyatkin, Emanuel Salazar-Cavazos, Patrick O Byrne, Jason O Moore, Christopher C Valley, Kathryn M Ferguson, Daniel J Leahy, Diane S Lidke, et al. Egfr ligands differentially stabilize receptor dimers to specify signaling kinetics. *Cell*, 171(3):683–695, 2017.

[133] Bin Wu and Wei Guo. Unpublished.

[134] James N Higginbotham, Michelle Demory Beckler, Jonathan D Gephart, Jeffrey L Franklin, Galina Bogatcheva, Gert-Jan Kremers, David W Piston, Gregory D Ayers, Russell E McConnell, Matthew J Tyska, et al. Amphiregulin exosomes increase cancer cell invasion. *Current Biology*, 21(9):779–786, 2011.

[135] Richard Cerione. *Cell*, In Review.

[136] Zhongwen Chen, Dongmyung Oh, Alok Kumar Dubey, Mingxi Yao, Beverly Yang, Jay T Groves, and Michael Sheetz. Egfr family and src family kinase interactions: mechanics matters? *Current opinion in cell biology*, 51:97–102, 2018.

[137] Xiaoling Peng, Jianyong Huang, Chunyang Xiong, and Jing Fang. Cell adhesion nucleation regulated by substrate stiffness: a monte carlo study. *Journal of biomechanics*, 45(1):116–122, 2012.

[138] Mayur Saxena, Shuaimin Liu, Bo Yang, Cynthia Hajal, Rishita Changede, Junqiang Hu, Haguy Wolfenson, James Hone, and Michael P Sheetz. Egfr and her2 activate rigidity sensing only on rigid matrices. *Nature materials*, 16(7), 2017.