A Quantitative Pipeline For The Identification of Combinations of Targets for Claudin-Low Triple Negative Breast Cancer Reversion

Madeleine S. Gastonguay

B.S. Degree Candidate in Applied Mathematics

Thesis Advisor: Paola Vera-Licona, PhD

Honors Advisor: Vasileios Chousionis, PhD

University of Connecticut College of Liberal Arts and Sciences, Storrs CT

UConn Health Center For Quantitative Medicine, Farmington CT

May 1st, 2020

Contents

List of Figures

List of Tables

Abstract

Claudin-Low Triple Negative Breast Cancer (CL TNBC) has high relapse and low survival rates. Due to the tumors' decreased response to cytotoxic drugs and hormone therapy, alternative therapeutic strategies should be explored. One such strategy is tumor reversion, the biological process by which tumor cells lose a significant fraction of their malignant phenotype. Tumor reversion has been observed for over a century and has been achieved *in vitro*, *in vivo*, and *ex vivo*. In particular, tumor reversion has been achieved *in vitro* with the CL cell line MDA-MB-231, and *ex vivo* in mice xenografted with MDA-MB-231 cells. This project takes a dynamical systems approach to identify *in silico* combinations of therapeutic targets for CL TNBC reversion. An intracellular signaling network was reconstructed with multi-omics profile data for MDA-MB-231. Then a structure-based attractor-based control method for nonlinear dynamic systems was applied to the network to identify driver nodes of the system. Topological signal flow analysis was applied to the network for virtual screenings of driver node perturbations to predict their effect on the system. Combinations of nodes whose concerted perturbation resulted in the system shifting from the tumorigenic basin of attraction to the normal-like basin of attraction were deemed putative concerted reversion targets. Through this methodology, several potential combinations of targets that may shift the cell from a tumorigenic to a normal-like phenotype have been identified to be further validated in future work.

Acknowledgements

First and foremost, I am extremely thankful to my mentor, Dr. Paola Vera-Licona, for taking me under her wing the past two years. I have learned so many valuable lessons under her tutelage; her guidance and feedback have helped shape me into the scientist I am today. I am constantly inspired by her passion for her research and so fortunate to have had the opportunity to work with the Computational Systems Medicine Team.

I am also incredibly grateful for the technical help and moral support of "my" graduate student, Lauren Marazzi. I have learned so much from watching her work and discussing all things computational biology. This experience certainly would not have been the same without her. I will greatly miss our Rebel Dog coffee runs and Journal Club field trips, and wish her the best of luck finishing her dissertation work.

I would like to thank the Trimble family for funding the Summer Undergraduate Research Fund Award I received to complete this work. It has afforded me a formative experience from which I've learned far more than I ever expected. The relationships and connections I've made through it will serve me well as my career as a scientist unfolds.

I also owe a debt of gratitude to everyone who guided me through the 2017 Holster Scholar Program. Many thanks to Robert and Carlotta Holster for their financial support, to Dr. Vin Moscardelli for his wisdom and guidance, and to Dr. Rachel O'Neill for stimulating my interest in the field of genetics. The program was an incredible foundation for my research experience at UConn and I'm so lucky to have been involved.

Last but not least, I am so thankful for the endless love and support from my family. Without their encouragement (and home cooked meals), I would not have been able to accomplish this work. Their selflessness inspires me daily and I am truly fortunate to be surrounded by them.

Chapter 1

Introduction

1.1 Claudin-Low Triple Negative Breast Cancer

Triple Negative Breast Cancer (TNBC) is an aggressive, heterogeneous subtype of breast cancer (BC) accounting for 15-20% of BC cases [1]. Unlike other subtypes of BC which are characterized by the presence of estrogen receptors, progesterone receptors, or HER2 amplification, TNBC is characterized by the lack of all three markers. Hence, it is not responsive to hormone therapy [2]. TNBC is characterized by high rates of relapse and shorter overall survival compared with other BC subtypes [3]. Although the current standard of care is chemotherapy, it shows low efficacy in treating TNBC [2, 4]. A targeted antibody–drug conjugate for metastatic TNBC, Sacituzumab Govitecan-hziy, has recently been approved for use on refractory tumors [2, 5]. Adverse effects of this drug include neutropenia (abnormally low white blood cells) and severe diarrhea [5]. It is currently only approved for use in patients who have previously received at least two therapies for the disease [2]. While this is a tremendous improvement for patients with metastatic TNBC, there is still a need for an efficacious treatment preventing relapse or the development of refractory tumors. Combinations of immunotherapy, targeted therapy, and/or chemotherapy may increase efficacy in treating TNBC. Currently, almost 80% of clinical trials studying TNBC therapeutics are

exploring combinatorial approaches [6].

Prior to 2007, BC was classified into four intrinsic subtypes : Luminal A, Luminal B, HER2-enriched, and basal-like. As gene expression technologies evolved, a new subtype accounting for 7-14% of invasive breast cancer was classified and termed Claudin-Low (CL) [7, 8]. This subtype is characterized by low to absent expression of tight junction proteins Claudin 3, 4, and 7, as well as the cell adhesion molecule E-Cadherin [9]. CL tumors are also known to have increased expression of immune response and mesenchymal genes [7, 10]. These molecular characteristics are associated with the epithelial-to-mesenchymal transition (EMT) and the development of stem cell features [8, 11]. EMT is the process by which epithelial cells gain migratory and invasive properties like that of mesenchymal stem cells [12, 13]. EMT and an enriched expression of immune response genes lead to the formation of stem cell characteristics (stemness), which is associated with tumor growth, therapeutic resistance, and metastasis [8, 14]. Furthermore, CL tumors exhibit low expression of proliferation-associated genes. This indicates they may be slower cycling tumors, resulting in poor success rates when treated with traditional cytotoxic chemotherapies that depend on rapid cell division for selectivity [8]. Due to its decreased response to targeted and cytotoxic therapies, this aggressive disease has a poor prognosis and low survival rates [8]. New therapeutic strategies are thus required for the treatment of CL TNBC.

1.2 Tumor Reversion

Conventional cancer therapies are designed to induce apoptosis in cancerous cells by targeting rapidly dividing cells [15]. Immunotherapies activate the body's natural immune response to kill cancer cells by ramping up the immune system or preventing the ability of a cancer cell to "hide" from the immune system by inducing the expression of antigens [16, 17]. The limitations to these approaches lie in that it is difficult to selectively choose which cells to kill, and ultimately normal cells are damaged in the process. Furthermore, acquired

resistance is a challenge for both treatment options [15, 16]. An alternative therapeutic approach for cancer therapy is tumor reversion. Tumor reversion is the biological process by which tumor cells lose a significant fraction of their malignant phenotype [15]. It is unlike conventional therapies in that instead of inducing apoptosis of cancerous cells, tumor reversion is aimed at reprogramming them to normal-like cells. In doing so, this approach reduces the deleterious impact on normal cells [18].

Tumor reversion is not merely the reversal of the development of a malignant cell. Instead, it is its own biological process involving cell reprogramming through epigenetic and genetic tools that supersede the tumorigenic characteristics of a cell [15]. Regulatory elements of a cell such as transcription factors can be targeted to induce the reversion of tumorigenic cells by triggering a cascading effect to alter the cell's transcriptional profile. Reversion targets differ from tumor suppressor targets because only reversion targets can trigger the shift from a malignant to a normal-like phenotype, rather than solely suppressing the malignant phenotype [15]. The systematic identification of these putative reversion targets is a problem well suited to investigation via systems biology methods.

Tumor reversion has been observed for over a century. It was first observed by Askanazy in 1907 when ovarian teratoma cells spontaneously differentiated into a normal somatic cell lineage [19]. Since then, it has been achieved *in vitro*, *in vivo*, and *ex vivo* [15]. In particular, reversion of the EMT phenotype of the well characterized CL TNBC cell line MDA-MB-231 has been observed *in vitro* through activation of GATA3. Its over expression leads to elevated expression of E-cadherin, ER, and increased disease-free survival [20]. The same observation was made *ex vivo* when mice xenografted with MDA-MB-231 cells overexpressing GATA3 lead to smaller primary tumors without metastasis compared to mice xenografted with control MDA-MB-231 cells [20]. GATA3 overexpression has also been observed to restore sensitivity to the tumor suppressor properties of TGFβ in normal epithelial cells [21, 22]. The EMT phenotype has also been reversed in MDA-MB-231 and mice xenografted with MDA-MB-231 cells through pan- Histone Deacetylace Inhibitors [23]. Due to previous

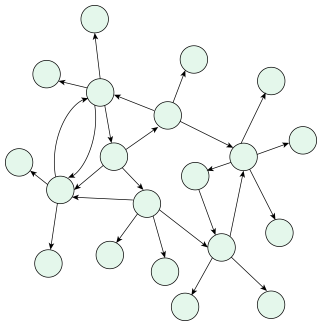
success with reversion of EMT both *in vitro* and *ex vivo*, data from MDA-MB-231 was used for this study to identify combinations of putative reversion targets.

1.3 Cells as Dynamical Systems

A cell can be viewed as a dynamical system - one that evolves over time. At a systems level, this evolution can be modeled using a dynamical model that combines information about the topology of regulatory interactions within the cell and their dynamics. Examples of dynamical models include systems of Ordinary Differential Equations (ODEs) and Boolean models, which represent system dynamics through ODEs with kinetic parameters or logic formalisms, respectively. In a dynamical model, the topology of intracellular interactions can be represented as a network where *nodes* represent molecular components of the cell and *edges* represent regulatory interactions between them (**Figure 1.1a**). Each state of the network corresponds to a specific pattern of expression or activity of genes, proteins, and other molecular components in the cell. The collection of all possible network states makes up a high-dimensional space known as the state space of the system [24]. As a cell develops, its molecular profile changes according to the dynamics of its intracellular processes until it reaches a fully differentiated stable state. This change in molecular profile corresponds to the movement of the network state to a new point in the state space. Such movement occurs along a trajectory specified by the dynamics of the system until it reaches a steady state, or *attractor* (**Figure 1.1b**).

In non-chaotic dynamical models, an attractor state corresponds to a steady state associated with a cell phenotype [25]. Attractors are characterized as "low potential states" in the state space that attract unstable states in their *basin of attraction*. Networks containing positive feedback loops, as is characteristic of biological networks, can exhibit *multistability*. This refers to the coexistence of multiple attractor states within the state space - which attractor the system ends in depends on its initial state [24]. Although the high-dimensional state

(a) Example of Network Topology. Nodes indicate molecular components of a cell and edges are regulatory interactions.



(b) A Portion of the State Space of a Dynamical System. Yellow indicates decreased activity and green denotes increased activity. The activity pattern of the network nodes determines the state of the system (s(t)). The system starts at an initial state $s(t_1)$ and changes over time according to its dynamics until it reaches a stable attractor state ($s(t_n)$) at some time t_n.

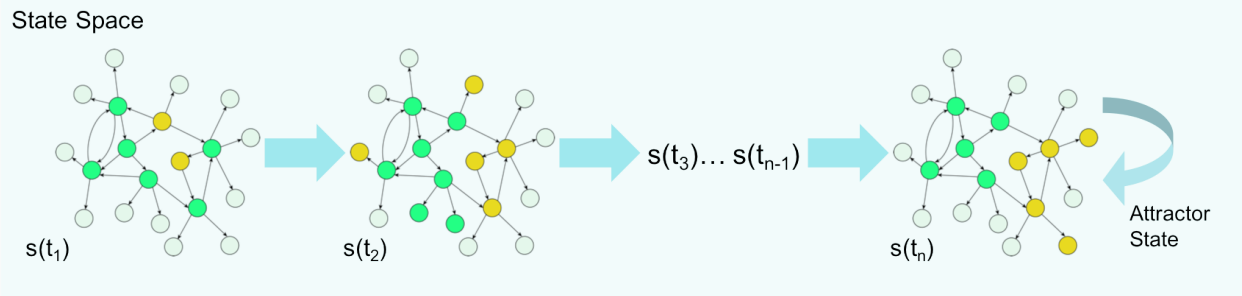


Figure 1.1: A dynamical model of the cell consists of a) network topology of intracellular interactions and b) dynamics of the interactions that govern how the system changes over time. Figure courtesy of Vera-Licona and Marazzi.

space is impossible to visualize, the idea of multistability and developmental trajectories of dynamical models representing cells can be conceptualized through the use of Waddington's Epigenetic landscape.

1.3.1 Waddington's Epigenetic Landscape

Waddington's Epigenetic landscape was first developed in the 1950s as a metaphor for cell differentiation [26]. It is a concept from developmental biology in which a differentiating cell is likened to a ball rolling down a hill. Just as a ball rolls down a hill until it reaches a valley, an undifferentiated cell starts at the top of the landscape and "rolls" down a developmental trajectory until it reaches a valley as fully differentiated cell. In this metaphor, the valleys of Waddington's epigenetic landscape each represent a cell type (**Figure 1.2**). This metaphor

has been extended to explain how cells can escape a mature cell type as described in section 1.4.2.

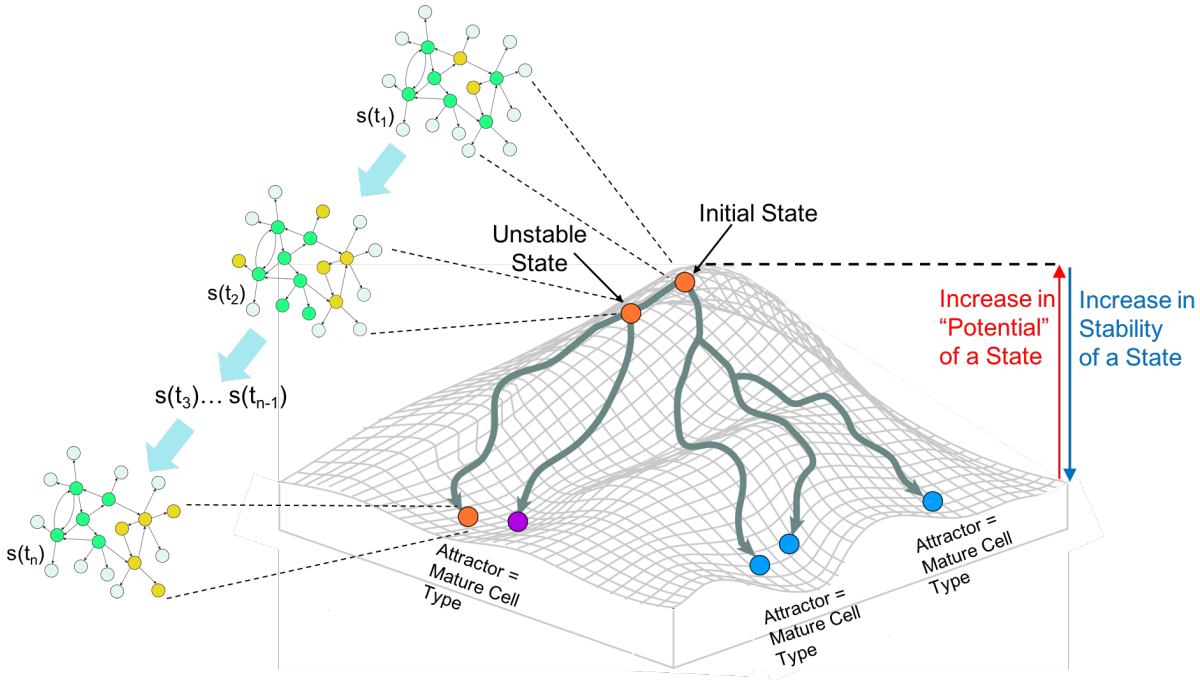


Figure 1.2: Waddington's Epigenetic Landscape as a Quasi-Potential Landscape of the High-Dimensional State Space of a Dynamical System. The trajectory from the top to the bottom of the landscape represents cell development. Network states in valleys of the landscape are attractors of the system, corresponding to cell types. States at the bottom of the landscape are more stable while states at the top of the landscape have more "potential". Figure courtesy of Vera-Licona and Marazzi.

Waddington's Epigenetic Landscape has since been formally explained as a generalized quasi-potential landscape for the state space of a dynamical model, shaped by the model's topology and dynamics [24, 25]. The height of a state in Waddington's epigenetic landscape corresponds to how much "potential" the state has, or how unstable it is (**Figure 1.2**). Here, potential does not refer to potential described in physics, but rather an abstract quasi-potential. Hence, the valleys of Waddington's landscape hold low-potential states - the attractors of the system - and hills correspond to unstable, transient states separating attractors. It is important to note that the term "epigenetic" in Waddington's epigenetic landscape does not refer to the molecular biology term used to describe chromatin modifications. Instead, it is more closely related to the notion of an "epigenetic state" from physics – a state that arises from genetic interactions [25]. In this framework, a switch between two

cell phenotypes corresponds to the transition of a cell from one valley to another.

1.4 A Dynamical Systems Perspective of Cancer

So often is cancer reduced to its genetic component that the larger picture of tumor cells in the context of cell development is lost. Normal cells, such as blood cells and epithelial cells, differentiate into distinct cell types despite having identical genomes. This paradigm can be extended to the development of cancer in that genetic mutations do not dictate how cancer cells develop. Instead, cancer cells are simply an abnormal cell type resulting from errors in the complex cell developmental processes that guides the differentiation of normal cells [25]. Thus, the development of cancer can be viewed as a dynamical systems process of abnormal cell development, driven by a regulatory network with dysregulated interactions.

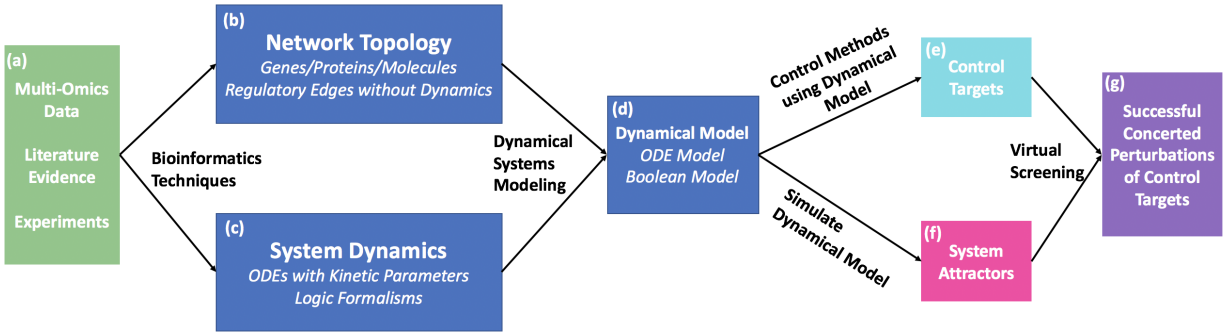


Figure 1.3: One typical dynamical systems approach to model a cancer cell and identify putative reversion targets. In this workflow, multi-omics data, literature evidence, and experiments **(a)** are used to determine the topology of interactions within the cell **(b)** and their dynamics **(c)**, which are combined to construct a dynamical model of the system **(d)**. Then, optimal control methods using the dynamics of the network are applied to identify control targets of the system **(e)**. The model is simulated to identify the attractors of the system **(f)**, and virtual screenings of concerted perturbations applied to control targets are done to determine which perturbations trigger a shift towards the normal-like attractor **(g)**.

Modeling cancer cells as dynamical systems allows for a dynamical systems approach to identify putative reversion targets. One typical approach that requires knowledge of the system dynamics is described in **Figure 1.3**. A dynamical model of the system is built from Multi-Omics Data, Literature Evidence, and Experiments. Then, the dynamics of the model are used to identify control targets and system attractors. Lastly, successful

perturbations that can trigger a shift toward the normal-like attractor are identified through virtual screening of concerted perturbations of control targets.

1.4.1 Cancer Attractors

Since attractors of the system correspond to cell types, and cancer is an abnormal cell type, one might infer that there are attractors of the system representing cancer cells. Kauffman first described the cancer attractor in 1971 as a preexisting attractor state of the network that cannot be accessed by normal cells [27]. This concept stems from the idea that a dynamical model of a normal cell can have many attractors that are not associated with normal physiological cell types, but are a byproduct of the systems' complex dynamics. Many of these attractors may correspond to non-viable gene expression profiles, but some may be associated with a viable, malignant phenotype, representing cancer. Normal developmental trajectories steer cells away from these attractors, but when those processes are dysregulated, as they are in the development of cancer, cells can fall into the basin of these cancer attractors [25].

The concept of the cancer attractor allows for the consideration of genetic mutations in the context of cell development. In a dynamical model of a normal cell, cancer attractors are not accessible. Genetic mutations in cancer cells alter the topology of the network by removing a node or edge in the case of loss-of-function mutations, and strengthening or adding additional edges in the case of gain-of-function mutations. These changes in the network structure alter the topography of the epigenetic landscape by lowering barriers to reach cancer attractors, or in the case of a bifurcation event, creating a de novo cancer attractor from a previously unstable state [25]. These mutations increase the probability that a developing cells falls into a cancer attractor. Other factors, such as changes in the tumor micro environment, can have a similar effect on the topography of the landscape and increase the probability of the development of cancer [25].

1.4.2 Tumor Reversion on Waddington's Epigenetic Landscape

While Waddington's epigenetic landscape was originally developed to explain cell differentiation, it has been extended to explain how cells respond to changes in environmental conditions causing perturbations to the system [18, 24]. Perturbations of a few regulatory molecule in a cell can propagate gene expression changes throughout the cell at a genome-scale until the cell returns to a steady state [24]. Such transient non-genetic perturbations to the network state move it from one site in the landscape to another, potentially changing which basin of attraction the system falls into (**Figure 1.4**) [18]. Alternatively, constitutively controlling target nodes can alter the dynamics of the system, and therefore the topography of the epigenetic landscape, possibly creating a path for the system to move out of the cancer attractor and into the basin of the normal-like attractor [18, 25]. The landscape is robust to small perturbations, but concerted perturbations of multiple network

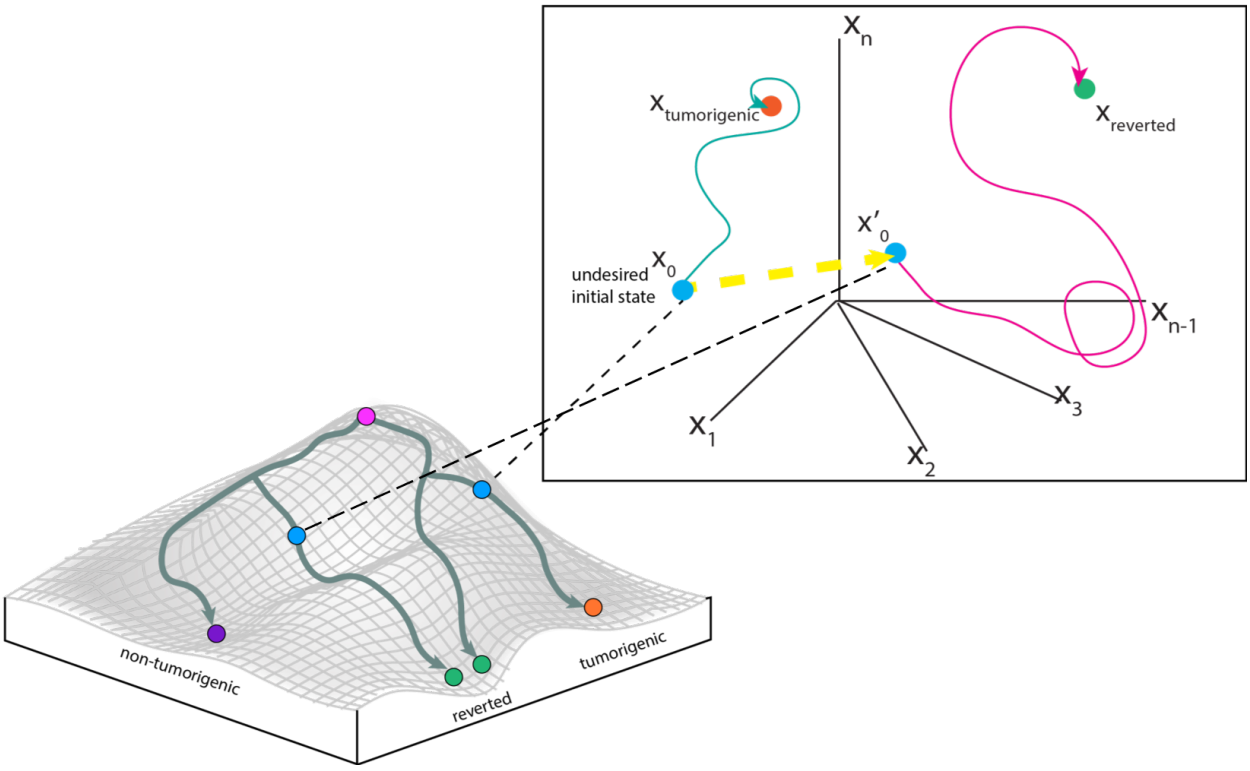


Figure 1.4: Tumor Reversion of Waddington's Epigenetic Landscape. Altering the initial condition of a cell presenting an undesired phenotype can send it down a trajectory towards a different attractor. The goal of tumor reversion is to move the system away from the cancer attractor and towards a normal-like attractor. Figure courtesy of Vera-Licona and Marazzi.

nodes may be sufficient to alter the trajectory of the system [24].

Thus, tumor reversion can be viewed as an optimal control problem in dynamical systems theory to identify combinations of targets that can trigger the shift out of the basin of a cancer attractor and towards that of a normal-like attractor.

1.5 Optimal Control Of Dynamical Systems

Control theory refers to influencing the behavior of a system to achieve a desired goal. Several methods exist for control of linear and nonlinear dynamical systems such as [28–33]. The objective of many of these control methods is full control of the system - the ability to steer the system from any initial state to any desired target state. As full control is not always necessary, attractor-based control methods have been developed to control the system by restricting the target states to naturally occurring attractors of the system.

Structure-based control methods distinguish themselves from the aforementioned control methods due to their ability to identify control targets of the system without knowledge of its dynamics [34]. Many of these methods require linear dynamics or linearized non-linear dynamics [34]. Such approaches are not appropriate for applications to biological networks, which are governed by nonlinear dynamics. Recently, a structure-based attractor-based control method for systems with nonlinear dynamics has been proposed. Feedback Vertex Set Control (FC) focuses on the controllability of the system based on its structure by restricting the target states to attractors [34, 35].

The Feedback Vertex Set (FVS) of a network is a subset of nodes whose removal leaves the network without directed cycles. Mochizuki et al. mathematically proved that, for a network governed by non-linear dynamics, controlling the dynamics of the FVS is sufficient to drive the dynamics of the whole system to converge to any dynamical attractor. Overriding the state variables of the FVS nodes into the trajectory specified by a given dynamical attractor ensures that the network will approach that desired dynamical attractor (**Figure 1.4**) [35].

Therefore, the smallest number of nodes needed to control the network is the minimal FVS (mFVS).

Zañudo further expanded this framework to include source nodes - nodes with no incoming edges - as necessary control targets [34]. Mochizuki and colleagues assume that source nodes converge to a unique trajectory and do not need independent control, but Zañudo asserts that source nodes can denote external stimuli a cell responds to, which could result in different attractors for source node states. The combination of the FVS and source nodes constitutes the FC set; in Zañudo's framework, control of these nodes guarantees that the system can be driven from any initial state to any dynamical attractor [34]. Thus, FC nodes whose concerted override steers the network away from the cancer attractor and towards the normal-like attractor may be putative reversion targets.

1.6 Intracellular Signaling Networks

The introduction of structure-based control methods for non-linear systems provides a way to identify control targets without the construction of a dynamical model. Instead, the system can be represented as an intracellular signaling network of interactions within a cell. Represented as a graph, the nodes of an intracellular signaling network represent molecular components of a cell and edges are the regulatory interactions in between them. Unlike dynamical models, this topological representation of the cell contains no information about the dynamics of interactions between molecules. **Figure 1.5** outlines the approach this study takes to identifying putative reversion targets using the intracellular signaling network without knowledge of the system dynamics.

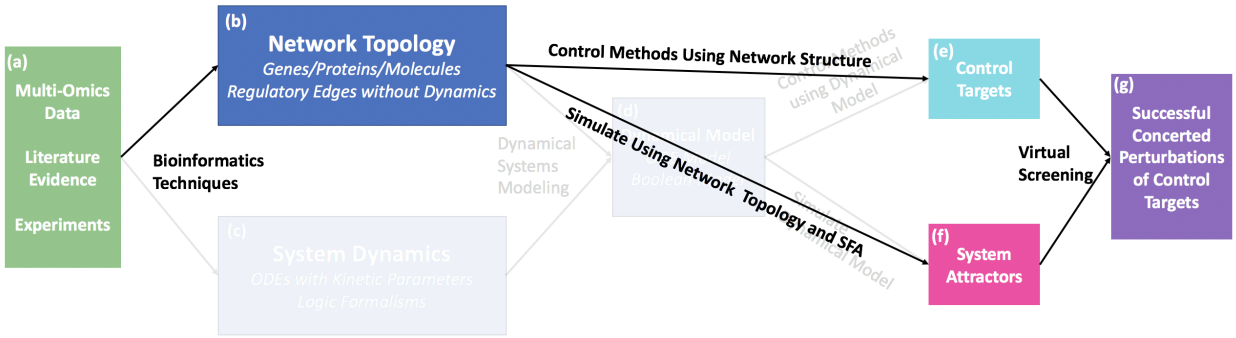


Figure 1.5: An alternative approach to identify putative reversion targets in which an intracellular signaling network is used to describe the system. Multi-omics data describing the cell **(a)** is used to construct a network of intracellular signaling pathways **(b)**. Then, structure-based control methods can be used to identify control targets if the system **(e)**, and system attractors can be estimated via methods developed to estimate system dynamics from the topology of the network **(f)**. Lastly, virtual screenings can be done to identify which concerted perturbations to control targets set the system on a trajectory towards the normal-like attractor **(g)**.

Advancements in high throughput measurement technologies have enabled experiments to produce various genomic, transcriptomic, proteomic, and epigenomic profiles to describe a system. The combination of these multi-omics profiles can be used to infer the underlying signaling processes of a cell. Taking a data-driven approach to constructing an intracellular signaling network reduces bias and allows for a genome-wide representation of the cell [36]. This representation places individual signaling pathways in the context of the cell by including their interactions with other molecular components. Such a view is necessary to identify key drivers of tumorigenesis and putative targets for cell fate reprogramming [36].

The proposed intracellular signaling network consists of an input layer, a target layer, and an intermediate layer canalizing intracellular signals between the two [37]. The target layer consists of functionally related differentially expressed genes (FunDEGs). These genes may not all be highly differentially expressed, but form a functional core whose expression pattern determines the phenotype of the system [38, 39]. The next layer consists of transcription factors (TFs) that transcribe the genes in the target layer to capture how small changes in regulatory elements alter the phenotype [39]. The last layer consists of genes, proteins, complexes, or other molecules considered to be Master Regulators (MRs) of the phenotype [40]. These elements are connected to the TF layer via intracellular signaling pathways that represents the functional flow of information from the MRs to the target layer [39].

1.7 Estimation of Network Dynamics

Though signaling networks do not describe the temporal behavior of a system, dynamical systems based approaches for network analysis have recently been developed. One such approach is the estimation of system dynamics based on network structure. Lee and Cho developed an algorithm that estimates signal flow between network nodes to predict attractors from user provided initial conditions based on the network topology [41]. In this signal propagation algorithm, the activity of a node i at time $t+1$ is determined by equation (1.1):

$$a_i(t+1) = \left(\prod_j a_j(t)^{W_{ij}} \right)^{\alpha} a_b(i)^{1-\alpha} \tag{1.1}$$

Where $a_j(t)$ is the activity of node j regulating node i at the previous time step, W_{ij} is the weight of the interaction between the two nodes, $a_b(i)$ is the basal activity of node i, and α is a hyper parameter between 0 and 1. The algorithm assumes that input stimulation does not change according to time. Taking the logarithm and using matrix notation produces equation (1.2):

$$x(t+1) = \alpha W x(t) + (1-\alpha)b \tag{1.2}$$

Where x is $log(a) \in P^N$, b is $log(a_b) \in P^N$, and $W \in P^{N \times N}$ is a weight matrix. This equation can be solved iteratively until it converges at some tolerance level, or the steady state solution of the equation can be solved with equation (1.3).

$$x_s = (1-\alpha)(I - \alpha W)^{-1} b \tag{1.3}$$

The link weight matrix, W, is normalized so that all links become a decay link. This ensures that the value of signal flow between two molecules will be smaller than the activity of the source of the signal. W is defined as:

$$W = D_{in}^{-1/2} A D_{out}^{-1/2} \quad , \quad (D_{in})_{ii} = \sum_j A_{ij} \quad , \quad (D_{out})_{jj} = \sum_j A_{ij}$$

Where A is the adjacency matrix of the network, and D_{in} and D_{out} are diagonal matrices consisting of the in-degree and out-degree for each node, respectively.

To compare the effects of several perturbations, the difference between log steady state attractor values can be calculated as

$$x^{fold} = x^{c_1} - x^{c_2}$$

Where $x^{c_i} \in P^N$ is a vector of the log steady state activity of each node for each condition. Thus, x^{fold} is essentially a log fold change. The signal flow algorithm is designed to predict the direction of activity change (DAC) (i.e. up or down regulated) rather than the accurate amount of change.

Applying signal flow analysis (SFA) to the network allows for the estimation of system attractors from predefined initial conditions based on the topology of the network. This will aid in generating hypotheses for experimental validation regarding which concerted perturbations of the FC nodes steer the system towards a normal-like attractor.

1.8 Objectives

This project takes a dynamical systems approach to systematically identify combinations of putative reversion targets for CL TNBC *in silico*. Multi-omics data from the normal-like breast cell line MCF10A and the CL TNBC cell line MDA-MB-231 were used to construct an intracellular signaling network of CL TNBC. Then, control targets were identified via FC, and SFA was applied to determine which concerted perturbations set the system on a trajectory towards the normal-like attractor. Ultimately, this pipeline was used to generate experimentally testable hypotheses for combinations of CL TNBC reversion interventions.

Chapter 2

Methods

2.1 Reconstruction of the Tumorigenic Intracellular Signaling Network With Multi-Omics Data

The intracellular signaling network of the CL TNBC cell line MDA-MB-231 is comprised of three layers: Functionally related differentially expressed genes, Transcription Factors, and Master Regulators. It was reconstructed using transcriptomics, proteomics, methylation, and mutational profiles of MDA-MB-231 to capture relevant tumorigenic signaling processes (**Figure 2.1**).

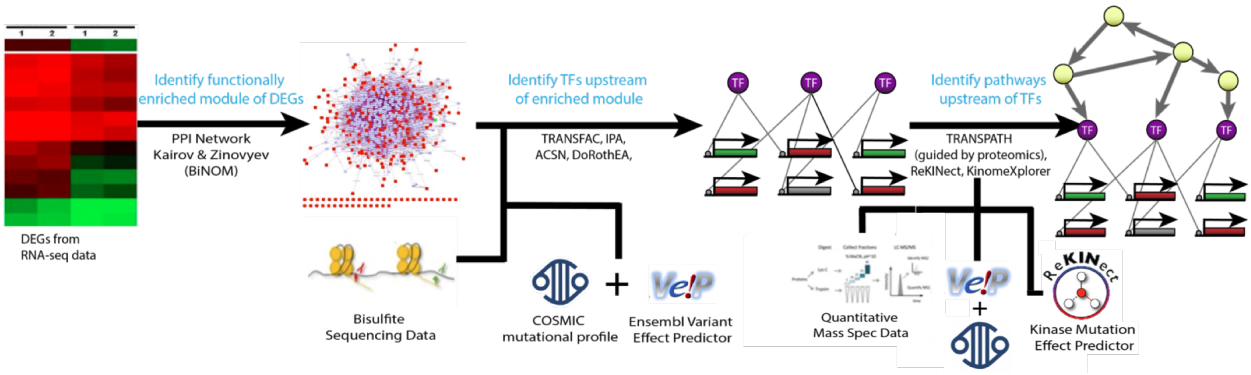


Figure 2.1: Pipeline for the reconstruction of the CL TNBC intracellular signaling network using multi-omics data. Differential expression of RNA-seq data is used to identify a module of FunDEGs. Then, TFs of the FunDEGs are identified, taking into consideration methylation and mutational profiles. Lastly, signaling pathways upstream of the TFs are identified using proteomics and mutational data. Figure courtesy of Vera-Licona and Marazzi.

2.1.1 Data Collection

Due to data availability, the network was constructed with data from cell-lines instead of patient data. The normal-like P53 wild-type epithelial cell line MCF10A was selected as a control as it is the standard choice in breast cancer research [42]. Pair-wise RNA-seq data for MDA-MB-231 and MCF10A was taken from [43], deposited at Gene Expression Omnibus (GEO) under GSE96860. Information regarding how these cell lines were grown can be found in Appendix A. Each cell line has two biological and two technical replicates. Bisulfite data for MDA-MB-231 was obtained from [44] (GEO accession number GSE42944). Single nucleotide variant (SNV) and copy number variant (CNV) profiles were obtained for MDA-MB-231 from Catalogue of Somatic Mutations in Cancer (COSMIC) Cell Lines Project under sample ID COSS905960 [45]. Label free quantitative mass spectrometry for MDA-MB-231 was obtained from [46], deposited in proteomicsDB under the accession number PRDB004167.

2.1.2 RNA-seq Data Processing and Differential Expression

The raw RNA-seq files were sent to Psomagen (formerly known as Macrogen) for a differential expression analysis. The Trimmomatic program was used to drop adapter sequences and bases with base quality lower than three from the ends, and the sliding window method was used to remove bases that do not qualify for window size 4 and mean quality 15. Reads with length shorter than 36bp were also dropped to produce trimmed data. The trimmed reads were mapped to hg19 using HISAT2, a program that handles spliced reads through Bowtie2 aligner, and known transcripts were assembled with StringTie. If more than one read count for the eight samples was zero, the gene was excluded from the analysis. This includes the removal of all genes with copy number loss as denoted by COSMIC CNV data. Reads were normalized with Relative Log Expression normalization and differential expression was calculated using the Wald test with a negative binomial distribution in the DESeq2 R package [47].

2.1.3 Ranking Differentially Expressed Genes

Differentially expressed genes (DEGs) were filtered for a False Discovery Rate (FDR) < 0.05, and a p-value < 0.001. By DESeq2 default, the FDR is equivalent to the Benjamini-Hochberg adjusted p-value [47]. DEGs were ranked using two methods. First, they were ranked by ascending FDR. The second method, adapted from Rathi et al., involved computing a pseudo-z-score for each gene and ranking by its descending absolute value [48]. The pseudo-z-score of the expression of node i is:

$$Z_i = \frac{e_{i,N} - \mu_{i,N}}{\sigma_{i,N}}$$

Where $e_{i,N}$ is the average expression of gene i in the cancerous cell line, $\mu_{i,N}$ is the average expression of gene i in the normal cell line, and $\sigma_{i,N}$ is the standard deviation of the 4 replicates of the expression of gene i from the normal cell line. This score reflects the cancerous expression level of a gene compared to the distribution of expression in the normal samples, and is related to the probability that the expression level in the tumor is from the same distribution as is in the normal tissue.

2.1.4 Identification of Functionally Related Differentially Expressed Genes

The first step in the pipeline is to identify a module of functionally related DEGs (FunDEGs). Optimally Functionally Enriched Network (OFTEN) analysis, developed by Kairov et al., considers functional relationships between molecules based on protein-protein interactions (PPI) as opposed to a statistical cutoff to identify a functionally related core of genes [38]. Thus, more biologically relevant interactions may be captured, even if some of the genes involved are not highly deferentially expressed. The method stems from percolation in graph theory, where one can estimate the expected size of the largest connected component given a graph and k randomly selected nodes. For some k_{crit} number of ranked nodes, most

of them are included in the largest connected component (LCC). Thus, if the first k ranked nodes ($k << k_{crit}$) form a largest connected component larger than expected by random, their connections are non-random, and they form a functionally related core.

This method is implemented in the BiNOM Cytoscape 2.8.3 plugin [49, 50]. Using the Human Protein Reference Database (HPRD) as the underlying PPI, the percolation scores for multiple sets of k top-ranked DEGs (k=50-1000 by increments of 50) were computed [51, 52]. The percolation score S is defined as:

$$S = \frac{C(k') - Mean(R(k'))}{k'}$$

Where k' is the number of ranked DEGs found in the HPRD PPI ($k' \leq k$). $C(k')$ is the size of the largest connected component (LCC) formed by these k' genes. $Mean(R(k'))$ is the mean LCC size of 10,000 networks of k' genes, each randomly sampled from the HPRD while preserving the connectivity distribution of the k' ranked DEGs.

The optimal k (k_{opt}) is chosen as the smallest k such that S decreases as k increases for $k > k_{opt}$. This analysis was run separately on the DEGs ranked by FDR and the DEGs ranked by pseudo z-score. Housekeeping genes, except for VIM because it is a marker of EMT, were filtered from the functionally related core and the resultant LCC was selected as the set of First Order FunDEGs [53].

After identifying First Order FunDEGs, BiNOM was used to identify Second Order FunDEGs. This step identifies nodes that share an edge between two First Order FunDEGs in the HPRD. Thus, the genes in the Second Order FunDEGs need not be differentially expressed, but could play important roles in tumorigenesis. Housekeeping genes and genes that are not expressed in MDA-MB-231 were removed from this set of genes and the LCC was selected as the set of Second Order FunDEGs.

To determine which set of genes to use for network construction (The first k ranked DEGs, First Order FunDEGs, or Second Order FunDEGs), a weighted sums metric was employed. This compares the contents of each gene set to Claudin-Low related genes, Breast Diseases

Ontology genes from DOLite in the GeneAnswers R package, and genes associated with the EMT and innate immune response hallmarks of cancer from the Atlas of Cancer Signaling Network (ACSN) [7, 36, 54–56]. The ACSN is a manually curated map of frequently deregulated molecular interactions within cancer cells and tumor microenvironments [56]. The fraction of a gene set that intersects with each these three lists was summed, each with a weight of one third. The gene set with the highest weighted sum was selected as the first layer of the network.

2.1.5 Identification of Transcription Factors

The next component of the intracellular signaling network is transcription factors (TFs) that regulate the FunDEGs. These were identified with the Analyze Promoters workflow using TRANSFAC and MATCH^{TM} in GeneXplain [57–59]. TRANSFAC is a manually curated database of eukaryotic transcription factors, their genomic binding sites and DNA-binding profiles [57]. It stores DNA binding patterns as positional weight matrices (PWMs) that can be used to identify putative transcription factor binding sites (TFBS) with the MATCH^{TM} algorithm [58]. The algorithm takes a DNA sequence as input and uses two scores- the matrix similarity score (MSS) and the core similarity score (CSS) - to measure the quality of the match between the sequence and the PWM. Matches that have a MSS and CSS higher than pre-defined cutoff values are reported as putative TFBSs. The Analyze Promoters workflow compares putative TFBSs within a specified promoter window size between a "Yes" set and a "No" set of genes. The yes/no ratio is calculated for each TF to indicate the proportion of targets the TF regulates that are in the "Yes" set compared to the "No" set.

For this analysis, PWMs for human TFBS with a p-value < 0.001 were used. The "Yes" set was the list of FunDEGs and the "No" set was the list of expressed housekeeping genes. The workflow was run with five different promoter window start sights: 500bp, 1000bp, 1500bp, 2000bp, and 2500bp upstream of the transcription start sight. The results were filtered for TFs with a yes/no ratio > 1.5 and a p-value < 0.05. Additionally, Ingenuity

Pathway Analysis (IPA) upstream regulator analysis was used to identify TF – FunDEG interactions that have been experimentally validated in any tissue or breast cancer sample. IPA contains a manually curated database of experimentally validated cause-effect relationships between genes, proteins, chemical compounds, and microRNA [60]. The upstream regulator analysis algorithm takes a statistical approach to identify regulators whose network connections to FunDEGs are unlikely to occur in a random model [60]. These results were filtered for interactions with a p-value < 0.001. The intersection of the TRANSFAC and IPA results at each promoter window size were filtered against gene and protein expression data for MDA-MB-231 and unmethylated CpG regions determined by bisulfite data. A weighted sums metric for the lists of TFs from each window size was calculated using the previously described criteria, and coverage of FunDEGs was checked. The TF list with the highest weighted sum such that all FunDEGs had a putative interaction with a TF was selected.

2.1.6 TF - FunDEG Interactions

Although TRANSFAC provides putative TFBS, it does not provide information regarding the sign (activation or inhibition) of the interaction between a TF and FunDEG. Those interactions that were not included in the IPA output were mapped into OmniPath using Cytoscape [50]. Omnipath is a network of high quality, manually curated signaling pathways, including TF-target interactions from curated databases, ChIP-seq binding data, TFBS predictions on human promoters, and inferences from normal tissue gene expression profiles from the Genotype-Tissue Expression (GTEx) project [61]. Only TF-target interactions with likely to high confidence (level A and B) were considered. Interactions with confidence level A are supported by at least two curated databases, a signed database and any of the other resources in Omnipath, or by all four types of resources. Interactions with confidence level B are supported by a curated resource and ChIP-seq data; a curated resource, TFBS, and GTEx inferences; or by ChIP-seq data, TFBS, and GTEx inferences [62]. The sign of the

relationship for any remaining TF – FunDEG interactions that were not in OmniPath was inferred by Pearson Correlation using the ExpressionCorrelation application in Cytoscape [63].

Mutations in regulatory regions can influence TF binding affinity. To identify mutations in regulatory regions of the FunDEGs, the COSMIC SNV data for MDA-MB-231 was filtered for SNVs in regulatory regions as determined by the Ensembl Variant Effect Predictor (VEP) [64]. The impact of mutations in regulatory regions on TF binding affinity was predicted as described in [65]. In short, each SNV in a regulatory region is analyzed for its ability to modulate the sequence to PWM match score. SNVs that increase the score compared to the wildtype sequence are potential gain-of-binding site events, while SNVs that decrease the score compared to the wildtype sequence are potential loss-of-binding site events. This is done in GeneXplain using the Mutation Effect on Sites Analysis workflow. Necessary modifications to TF-FunDEG interactions are modified manually according to an increase or decrease in binding affinity.

2.1.7 Identification of Master Regulators

The final layer of the network consists of master regulators (MRs) – genes, proteins, or complexes that regulate the TFs in the network. Pathways in which the TFs participate were identified using the Master Regulators with Context Genes with Weights pipeline in GeneXplain and the manually curated TRANSPATH database [66]. This workflow uses a specified radius and Dijkstra's shortest path algorithm on the TRANSPATH database to identify upstream molecules regulating the TFs. Proteomics data for MDA-MB-231 was used as context to guide the shortest path algorithm towards nodes with protein expression. The workflow weighs the context genes based off of the average protein intensity for each gene by decreasing the cost of the edges around genes with higher levels of protein expression, thereby attracting the algorithm to those molecules [67]. The analysis was run with radii ranging from 5 to 10, indicating how far upstream from the TFs to search for MRs. The

smallest radius such that all the TFs have an upstream MR was chosen. Resultant master regulators were filtered for a z-score > 1.5, an FDR < 0.05, and to include only those that have protein expression. Nodes in MR pathways without gene expression data were removed from the network so long as their removal did not create a source node.

SNVs have the capability to inactivate, activate, or modify the specificity of a protein's kinase domain which may cause profound signaling changes within a cell. To determine the impact of kinase mutations in the network, the Ensembl VEP ProteinSeqs plugin was used to obtain reference (wild-type) and mutated protein transcripts. Then, the ReKINect software was used to predict and classify kinase modifying mutations in the SNV data for MDA-MB-231 [68]. ReKINect employs a database of all known human kinase domains, 111 SH2 domains, and 149,838 phosphorylation sites and will predict the functional impact of the SNV. For mutations affecting protein kinase function, the downstream targets of the mutated kinases were determined with KINSpect, using the ReKINect database [69]. KinomeXplorer was used to investigate upstream kinase specificity changes for mutated proteins [70]. The identified mutated pathways were added to or removed from the network manually, as they are not included in signaling databases.

2.1.8 Addition/Removal of Mutated Pathways

After considering the effect of mutations in TF binding regions and kinase signaling pathways, the effect of additional mutated network nodes must be considered manually. Ensembl Variant Effect Predictor (VEP) was used to provide information regarding the location and consequence of each SNV [64]. Mutations that lie in protein coding regions were filtered to include only those that modify the protein amino acid sequence (missense, splice-site variation, frameshift, deletion, stop gained). A literature curation of the gene variants that cause protein transcript alterations was done to explore their impact on TNBC tumorigenesis. If a gene variant causes gain-of-function (GOF) activities that affect a specific TNBC pathway, the interaction and pathway were added to the network manually as appropriate, since they

are not included in signaling databases. Additional interactions between the added pathways and other network nodes were determined by the shortest paths algorithm in GeneXplain.

2.2 Identification of Putative Reversion Targets

2.2.1 Characterising Normal and Cancer Associated Attractors

To identify attractor states of the intracellular signaling network that are associated with the normal and cancerous phenotype, attractors resulting from the experimental data for the normal and cancerous cell lines were estimated. The network was initialized with the normalized RNA-seq expression data for each of the eight experimental replicates (four from the normal cell line and four from the cancerous cell line). Then, SFA was applied to the unperturbed network with these initializations and the resultant attractors were estimated. Non-metric multi-dimensional scaling (MDS) using sammon mapping from the MASS R package and unsupervised k-means clustering using Scikit Learn in python were applied to these attractors [71, 72]. The optimal number of clusters (k_{opt}) for k-means was identified by creating a Silhouette plot of the sum of squared distances between samples and their closest cluster center for k (number of clusters) ranging from 2 to 8 [73]. k_{opt} was selected such that for all $k > k_{opt}$, the sum of squared distances did not significantly decrease and attractors from normal initial conditions clustered separately from the attractors from cancerous conditions [73]. The direction of activity change (DAC) between the estimated attractors and the median of the cancer attractors was calculated to compare attractors associated with the normal and cancerous phenotype. MDS and k-means were applied to the DAC of the attractors as well as the concatenation of the attractors and their DAC. To reduce emphasis on the magnitude of the SFA output, the SFA output and DAC values were grouped into three values: 1 for positive values, -1 for negative values, and 0 for zeros. MDS and k-means were also applied to these data sets to determine if manipulating the SFA output impacted the association of attractors to phenotypes.

2.2.2 Identification of the Minimal Feedback Vertex Set

The OCSANA+ Cytoscape app was used to identify the minimal FVSes (mFVSes) of the intracellular signaling network [74]. Identifying the mFVS is a well-known NP-hard problem. As such, several algorithms have been developed to identify near-minimum FVSes. OCSANA+ employs a simulated annealing local search approach, SA-FVSP, originally described by Galinier et al. to do so [74, 75].

2.2.3 Virtual Screenings

Virtual screenings were run using SFA to predict the effect of concerted perturbations of the FVS nodes on the long-term behavior of the system. FVS nodes were either knocked-out, knocked-in, or left unchanged. Knock-out perturbations were simulated by setting the value of the FVS node to 0, and fixing it at 0 until the system reached an attractor state. Similarly, knock-ins were simulated by fixing the value of FVS nodes 2-fold higher than the mean normalized RNA-seq expression of the FVS node across all four experimental replicates for MDA-MB-231.

The network was initialized with normalized RNA-seq expression data for the tumorigenic cell. 100,000 random perturbations of the FVS nodes were applied to this initial state, and the resultant attractors were estimated with SFA. Steady states were reached by running SFA iteratively instead of solving the steady state equation in order to maintain the fixed basal perturbation values for the FVS nodes. This process was repeated for each of the four experimental replicates of MDA-MB-231.

2.2.4 Classification of Attractors from Virtual Screenings

The attractors produced from virtual screenings of the FVS nodes perturbations need to be associated to a phenotype to identify perturbations that result in normal-like attractors. To do so, k-nearest neighbors (knn), a supervised machine learning algorithm for classification

implemented in python Scikit Learn, was used [72]. The training set was composed of the estimated attractors from the experimental data and their associated clusters from k-means analysis. A perturbation was deemed successful if it led to an attractor classified as part of a normal cluster when applied to all four cancerous initializations. The DAC of the attractors from virtual screenings was computed against the median of the cancerous attractors from the training set, and the data sets were grouped into ones, negative ones, and zeros as previously described. Knn was applied to each of these data sets and their concatenations, first using euclidean distance for all datasets, and then using hamming distance for the discrete datasets. The resultant classifications were compared to ensure selectivity of successful perturbations.

2.2.5 Prioritization of Reversion Targets

After identifying successful perturbations, they were prioritized for future validation. First, perturbations were prioritized by the number of nodes whose state override is required to trigger a shift to the normal-like cluster. This is an important concern for experimental validation, where it is not feasible to knock-out or knock-in a large number of nodes. Additionally, the dependency of cell survival on each FVS node was assessed using the Cancer Dependency Map (DepMap). DepMap includes data from Project Achilles, which studies the dependencies of cancer cell lines *in vitro* via CRIPSR/Cas9 loss-of-function screens [76]. Gene effect scores are calculated for each cell line as a reflection of the probability that the cell line is dependent on it for survival. A score of zero indicates that it is not essential and a score of -1 corresponds to the median of all common essential genes. Thus, a gene scoring less than -1 for MDA-MB-231 indicates that the cell line is dependent on it. As the objective is not to kill the cell, perturbations requiring the knock-out of a gene upon which MDA-MB-231 is dependent were excluded from further consideration. Lastly, perturbations were compared to previous findings in literature as additional validation.

Chapter 3

Results

3.1 Proposed CL TNBC Intracellular Signaling Network

The proposed intracellular signaling network for MDA-MB-231 includes 227 nodes and 572 edges (**Figure 3.1**). 90 nodes are associated with the EMT, Stemness, and immune response hallmarks of cancer; 27 are associated with breast disease ontology; and 5 are markers of CL TNBC [7, 36, 54]. Additionally, 142 of the network nodes are in the Atlas of Cancer Signaling Network, a manually curated map of frequently deregulated molecular interactions within cancer cells and tumor microenvironments [56].

3.1.1 Identification of Network Components

Functionally Related Differentially Expressed Gene Module

To construct the first layer of the intracellular signaling network, the most representative module of FunDEGs was identified using BiNOM. After filtering DEGs for a p-value < 0.001 and FDR < 0.05, 4652 DEGs remained. None of these genes had complete copy number loss. **Table 3.1** shows that the optimal k (k_{opt}) for DEGs ranked by z-score produced a

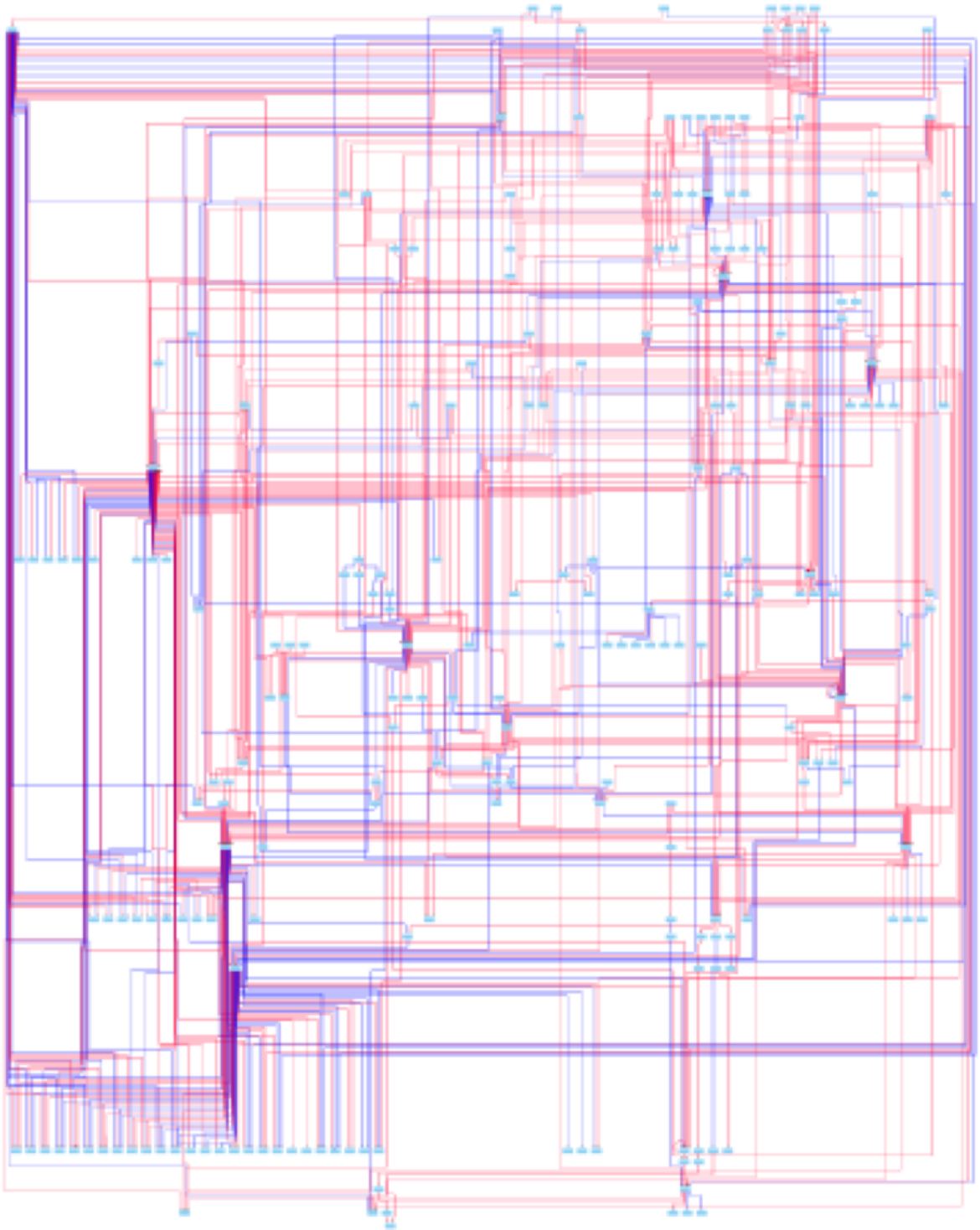


Figure 3.1: Intracellular Signaling Network for CL TNBC. The network consists of 227 nodes and 572 edges. Edges in red are activating, while edges in blue are inhibitory. The network contains 90 nodes associated with EMT, Stemness, and immune response; 27 associated with breast disease ontology; and 5 makers of CL TNBC. It also includes 142 nodes involved in pathways frequently deregulated in cancer cells and tumor microenvironments, according to the ACSN.

larger connected component and a larger percolation score than the results for DEGs ranked by FDR. The removal of housekeeping genes resulted in a set of 80 first order FunDEGs. BiNOM identified 378 second order FunDEGs, which were reduced to a LCC of 316 genes after removing housekeeping and unexpressed genes.

Table 3.1: Comparison of BiNOM results for the identification of First Order FunDEGs from DEGs ranked by FDR and pseudo z-score.

	Ranking Method	
Parameter	**FDR**	**Pseudo Z-score**
k_{opt}	700	650
k'	383	342
LCC size	64	83
Percolation Score (S)	0.080	0.106

Weighted sums of the 4652 DEGs, the 80 First Order FunDEGS, and the 316 Second Order FunDEGs show that the list of Second Order FunDEGs contains the highest proportion of CL TNBC relevant genes (**Table 3.2**).

Table 3.2: Number of genes from each potential list of FunDEGs that are found in the EMT and Innate Immune Hallmarks of Cancer (HOC), Breast Diseases Ontology (BDO), and CL Related Genes (CLG) and the corresponding weighted sum.

FunDEG Set	**Total**	**HOC**	**BDO**	**CLG**	**Weighted Sum**
DEGs	4652	275	187	20	0.0345
First Order	80	24	10	2	0.15
Second Order	316	126	35	7	0.177

Transcription Factors

The MATCH^{TM} algorithm and IPA upstream regulator search were applied to the First Order FunDEGs and Second Order FunDEGs to identify transcription factors. After identifying transcription factors of the Second Order FunDEGs and pulling interactions from available databases, there were more than 70 Second Order FunDEGs without known interactions or significant expression correlations with the network TFs. Including additional

TFs from IPA did not significantly increase the number of Second Order FunDEGs with experimentally validated interactions with the TFs. Thus, the list of First Order FunDEGs was selected as the first layer of the network.

A comparison of weighted sums for First Order FunDEG TFs at each promoter window size can be found in **Table 3.3**. The results from the first four promoter window sizes have the same weighted sum, so the list with the largest number of transcription factors was selected because it holds the largest number of CL TNBC relevant genes. The nine TFs in the intracellular signaling network are TCF3, FOXM1, ZBTB17, SNAI1, CTNNB1, STAT3, FOS, RELA, and JUN.

Table 3.3: Number of genes from each potential list of TFs that are found in the EMT and Innate Immune Hallmarks of Cancer (HOC), Breast Diseases Ontology (BDO), and CL Related Genes (CLG), and the corresponding weighted sum.

TF Promoter Window Size	Total	HOC	BDO	CLG	Weighted Sum
500bp	9	7	2	0	$0.3\bar{3}$
1000bp	7	6	1	0	$0.3\bar{3}$
1500bp	5	4	1	0	$0.3\bar{3}$
2000bp	7	6	1	0	$0.3\bar{3}$
2500bp	6	5	0	0	0.278

Master Regulators

When searching for master regulators (MRs) of the TFs, a radius of 9 was the smallest choice such that all transcription factors were regulated by an upstream pathway. Filtering against proteomic data for MDA-MB-231 resulted in 50 MRs.

3.1.2 Assembly of the Intracellular Signaling Network

The interactions between TFs and FunDEGs were identified as follows: 47 from IPA, 10 from Omnipath, 141 with pearson correlations greater than 0.90, and 2 with Pearson correlations greater than 0.8. After the inclusion of these edges, there were three FunDEGs lacking interactions with the network TFs. Their removal did not disrupt the LCC of the FunDEGs

and they are not related to TNBC, so they were removed from the network. According to VEP, there were no mutations in the regulatory regions of the First Order FunDEGs to take into consideration.

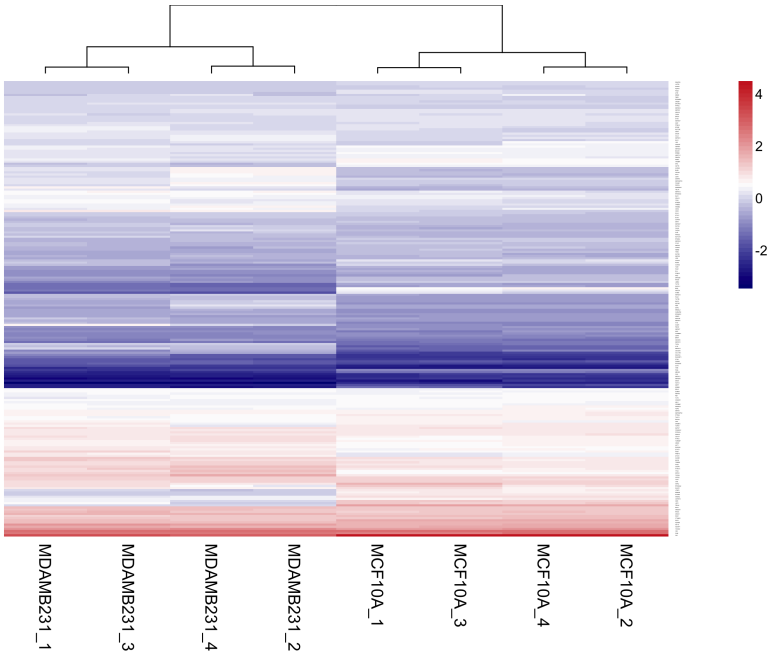
The reKINect output indicated that there was one mutated kinase in the network. The BRAF p.G464V mutation hits an essential residue on the kinase domain and has been shown to increases MAPK growth pathway signaling through MEK and ERK [77–79]. These interactions were already included in the network so no alterations were made. In addition to the kinase mutations, there was a nonsense p.Q249* TAB2 mutation and a gain-of-function p.R280K TP53 mutation in the network [80]. TAB2 was removed from the network without creating any source nodes. Pathways non-canonically activated by R280k p53 were added to the network. The R280K mutation also causes a loss of normal TP53 activity, thus the outgoing edges from TP53 determined by the master regulator search were removed [80]. The exception to this is NOTCH1, due to literature evidence that mutant TP53 binds to and activates NOTCH1 [81].

3.2 Identification of Reversion Targets

3.2.1 Estimating Cancer and Normal-Like Attractors

Attractors estimated from the MCF10A and MDA-MB-231 normalized RNA-seq expression exhibit similar expression patterns and hierarchical clustering results to the experimental data (**Figure 3.2**). Hierarchical clustering, k-means clustering, and MDS on the attractors identified four distinct clusters - two with cancer attractors and two with normal-like attractors (**Figure 3.3b**). These results are consistent with the output of clustering and MDS applied to normalized RNA-seq data for the eight experimental replicates (**Figure 3.3a**). In both datasets, the biological replicates of each cell line (ending in "_1" or "_3") cluster separately and their technical replicate (ending in "_2" or "_4") clusters with them. Thus, two normal-like clusters and two cancerous clusters were used in the training set for knn.

(a) A heatmap of experimental RNA-seq expression values of the network nodes for the normal-like and cancerous cell lines. The color gradient indicates the z-score resulting from centering and standardizing the expression profile of each sample.



(b) A heatmap of the expression values of the estimated attractors for the normal-like and cancerous cell lines. The color gradient indicates the z-score resulting from centering and standardizing the expression profile of each sample.

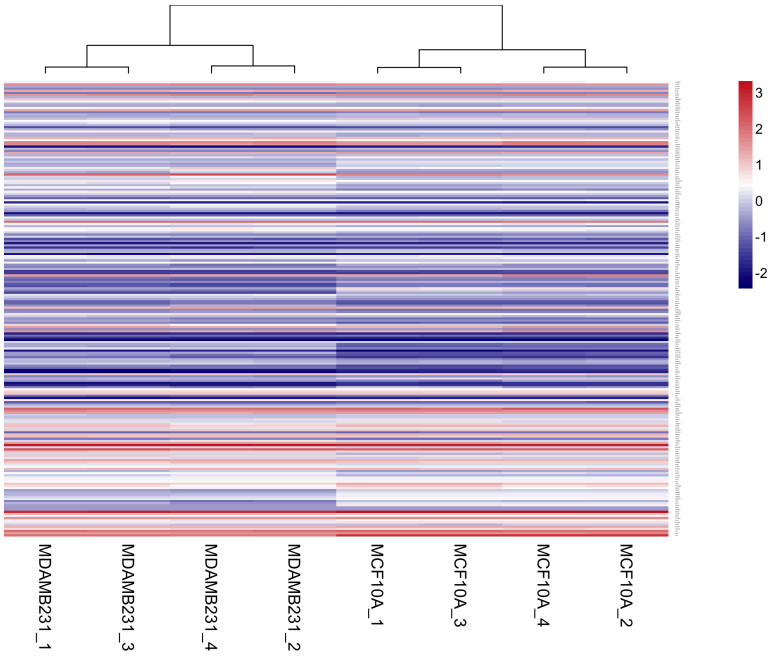
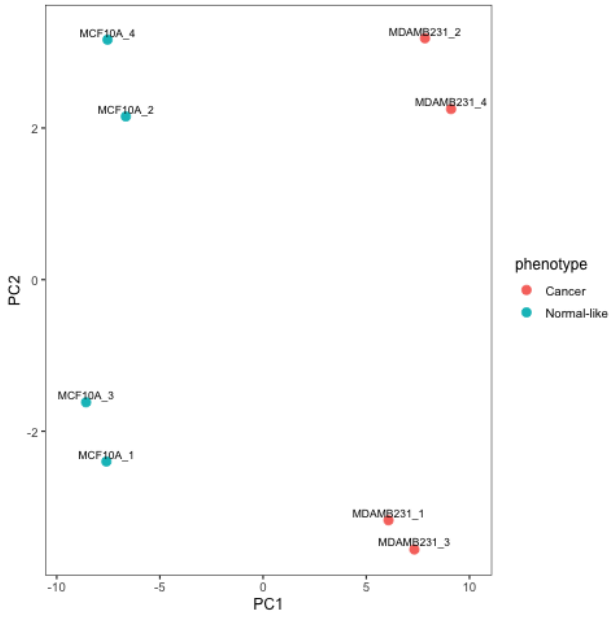


Figure 3.2: Heatmaps of a) experimental expression values and b) estimated attractor values of the two cell lines exhibit similar expression patterns and identical hierarchical clustering.

(a) MDS on Normalized RNA-seq Expression of Experimental Replicates of the Normal-Like and Cancerous Cell Lines.



(b) MDS on Estimated Attractors of the Normal-Like and Cancerous Cell Lines.

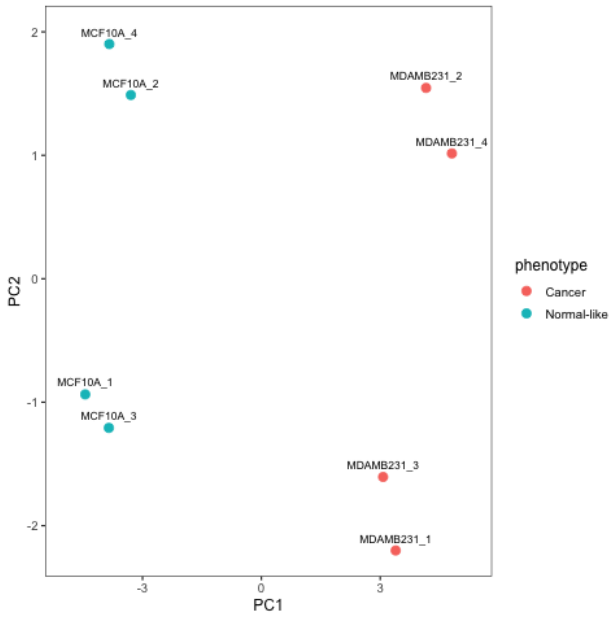


Figure 3.3: Non-metric MDS with Sammon mapping on a) Normalized RNA-seq expression data for the experimental replicates of MCF10A and MDA-MB-231 and b) Attractors of experimental conditions predicted by SFA. MDS plots show similar grouping of the replicates and their estimated attractors.

3.2.2 Virtual Screenings

FVS control applied to the network produced two mFVSes, each of 12 nodes. Eleven nodes were identical in the two sets: AKT1, AURKA, CTNNB1, FOXM1, GSK3B, HDAC3, JUN, MAPK1, RELA, STAT3, and TCF3. One of the mFVSes contained SUMO1 while the other contained PIAS1. Analyses were done using the set containing PIAS1 as it plays a crucial role in the regulation of several pathways involved in tumorigenesis [82–84]. The FC set consists of these 12 nodes and 25 source nodes.

Analyses were focused on the 12 FVS nodes to minimize the number of control targets. There are $531,441$ possible concerted perturbations of these nodes. Due to computational limitations, a subset of 100,000 perturbations were randomly generated. It was confirmed that these perturbations capture a representative sample of all possible perturbations (**Table 3.4**).

Table 3.4: Frequency of knock-ins, knock-outs, and no changes for each FVS node in the 100,000 random perturbations.

	Perturbation Orientation		
FVS Node	**Knocked-in (%)**	**Knocked-out (%)**	**No Change (%)**
AKT1	33.426	33.206	33.368
AURKA	33.083	33.264	33.653
CTNNB1	33.249	33.406	33.345
FOXM1	33.389	33.305	33.306
GSK3B	33.268	33.473	33.259
HDAC3	33.455	33.335	33.21
JUN	33.306	33.416	33.278
MAPK1	33.411	33.332	33.257
PIAS1	33.28	33.103	33.617
RELA	33.318	33.372	33.31
STAT3	33.488	33.275	33.237
TCF3	33.382	33.355	33.263

To classify the attractors simulated from virtual screenings, knn was applied to six datasets in total: the log steady state attractors values, the DAC of each attractor against the median cancer attractor values, the concatenation of these two datasets, and all three of

these datasets grouped into discrete values of ones, negative ones, and zeros. The training set consisted of two clusters of normal-like attractors and two clusters of cancer attractors. Classifications based on the discrete datasets did not differ when using euclidean or hamming distance as a distance metric. Classifications from the concatenated sets of values were used because they did not differ greatly from classifications by the other two datasets, and incorporated information from both the attractor state and DAC to classify the the virtual screening results. Three nearest neighbors were used to classify these attractors.

Perturbations resulting in an attractor classified as part of a normal-like cluster were compared between all four experimental initial conditions for MDA-MB-231. Classifications using the discrete dataset did not result in any perturbations triggering a shift to the normal-like cluster in more than two of the initial conditions. On the contrary, **Table 3.5** shows that using the floating point numbers to classify the attractors identified perturbations that shift the network to the normal-like attractor from multiple initial conditions. There were 157 perturbations that triggered a shift to the normal-like cluster from all four cancerous initial conditions based on these classifications. The distribution of the orientation of each FVS node in these 157 perturbations can be found in **Table 3.6**.

Table 3.5: Size of the intersections between perturbations that result in an attractor in the normal-like cluster from all four initial conditions when classified using floating point values.

	MDAMB231_1	**MDAMB231_2**	**MDAMB231_3**	**MDAMB231_4**
MDAMB231_1	270	173	256	159
MDAMB231_2	173	213	185	191
MDAMB231_3	256	185	277	169
MDAMB231_4	159	191	169	191

The fewest number of FVS nodes perturbed to trigger the shift to the normal-like cluster was 6 nodes. **Figure 3.4** shows the distribution of the number of nodes perturbed in the 157 successful perturbations. **Table 3.6** includes the dependency of MDA-MB-231 on each FVS node as indicated by the Gene Effect score. The only FVS node that MDA-MB-231 shows a high dependency on is HDAC3. Successful perturbations requiring its knock-out will not be considered for future validation.

Table 3.6: Frequency of knock-ins, knock-outs, and no changes for each FVS node in the 157 successful perturbations and the dependency of MDA-MB-231 on each node.

	Perturbation Orientation			
FVS Node	**Knocked-in (%)**	**Knocked-out (%)**	**No Change (%)**	**Gene Effect**
AKT1	12.102	58.599	29.299	0.145
AURKA	0	68.153	31.847	-0.894
CTNNB1	82.166	7.006	10.828	-0.169
FOXM1	0	64.968	35.032	-0.145
GSK3B	89.172	1.274	9.554	-0.0388
HDAC3	82.166	6.369	11.465	-1.31
JUN	36.943	42.675	20.382	-0.299
MAPK1	29.936	31.847	38.217	-0.523
PIAS1	68.79	7.006	24.204	-0.323
RELA	0	68.79	31.21	-0.287
STAT3	0.637	57.962	41.401	0.246
TCF3	2.548	52.229	45.223	0.149

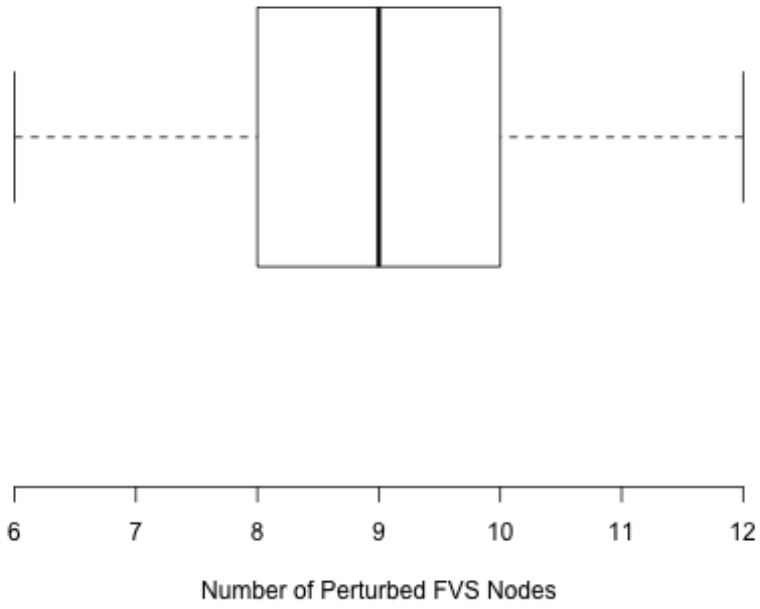


Figure 3.4: Distribution of the Number of Perturbed Nodes in the 157 Successful Perturbations. The number of nodes that must be perturbed to trigger a shift to the normal-like cluster exhibits a symmetric distribution between 6 and 12, with a median of 9.

Chapter 4

Discussion

In this study, a dynamical systems approach was used to identify putative concerted reversion targets for Claudin-Low Triple Negative Breast Cancer. This approach provides a framework to systematically uncover drivers of tumorigenesis. Optimal control theory was used to identify control targets of the system based on its structure, and a topological estimation of signal flow was applied to determine the effect of their concerted perturbation. This pipeline has resulted in the identification of putative reversion interventions for CL TNBC which will be validated in future studies.

4.1 Construction of the Proposed Intracellular Signaling Network

Advancements in high-throughput measurement technologies make it possible to characterize intracellular signaling processes via multi-omics data. The data-driven construction of the intracellular signaling network reduces bias and provides a genome-wide representation of the cell. This representation includes many signaling pathways and their interactions so that those who are integral to tumorigenesis can be captured and the impact of therapeutic interventions on the entire cell can be examined. The proposed intracellular signaling

network for CL TNBC includes nodes that are markers of, or involved in the regulation of EMT and Stemness, two characteristic phenotypes of the disease. Additionally, over 60% of network nodes are found in the Atlas of Cancer Signaling Network, demonstrating that the developed pipeline for network construction captures signaling pathways relevant to CL TNBC [56].

Traditional methods of interpreting differential expression data use a statistical cut-off to identify the most differentially expressed genes [85]. This poses a problem because highly differentially expressed genes tend to be downstream of TFs and other regulators that may not exhibit large changes in activity, especially when compared to changes in downstream gene expression. By considering only highly differentially expressed genes, upstream elements that play an important role in their regulation may be excluded from the analysis [85]. To prevent this, several methods have been developed to incorporate biological information when analyzing high-throughput expression data such as Over-representation Analysis, Gene Set Enrichment Analysis, Global Test with generalized linear models and kernel machines, and Principle Component Analysis [86–89]. These methods focus on determining if pre-defined sets of biologically related genes are differentially expressed as a whole instead of focusing on individual gene expression. Some limitations of these methods include the requirement of a statistical cut-off for differential expression of a gene set and the assumption that each gene in the gene set is differentially expressed in the same direction. Kairov et al.'s method implemented in BiNOM has a similar objective, but does not require a statistical cut-off or a specification of gene sets a priori [38]. Using the HPRD as prior biological knowledge, it identifies functionally related genes within the DEGs and includes genes that may not be highly differentially expressed, but interact with those that are. When a weighted sums metric is applied the FunDEG sets, it is clear that this method captures more CL TNBC relevant genes than a statistical cutoff of differential expression. This reinforces that consideration of biological information is paramount when analyzing differential expression analyses.

Applying OFTEN analysis to different ranking methods of DEGs reveals that the pseudo z-score ranks functionally related genes higher than ranking by the FDR. The FDR is a reflection of the probability that a log2 fold change for a gene is a false positive; it is used to determine which genes are significantly deferentially expressed [47]. On the other hand, the pseudo z-score is an indicator of the probability that the expression level of a gene in the cancerous cells comes from the distribution of expression values in the normal cells [48]. Larger z-scores indicate a larger difference in expression between the two conditions. Hence, the z-score includes information regarding *how* differentially expressed two genes are as opposed to the confidence that the difference in expression is not a false positive. This distinction could explain why ranking by pseudo z-score identified a larger module of functionally related genes.

Results from the MATCH^{TM} algorithm for the identification of TFBS indicate putative interactions between FunDEGs and TFs. Thus, even if a TF can bind to the promoter of a FunDEG, it does not necessarily transcribe said FunDEG. This was observed with the set of Second Order FunDEGs - many interactions predicted by MATCH^{TM} were not experimentally observed. Although the decision to use the set of First Order FunDEGs resulted in a loss of CL TNBC genes in the network, there is experimental evidence of their regulation by TFs of MDA-MB-231, which is not the case for many Second Order FunDEGs. Further work should be done to determine if interactions between TFs and Second Order FunDEGs are missing because they do not exist or because they are not included in the resources used for network construction.

Genetic mutations play an integral role in shaping the topology of the intracellular signaling network of a cancer cell. The mutational profile of MDA-MB-231 revealed several gain-of and loss-of function mutations altering the network topology. Several of these mutations were found in kinases, which can have profound effects on signaling pathways regulated through phosphorylation and de-phosphorylation. Creixell et al.'s ReKINect software provides an automated way to identify the upstream and downstream effect of kinase mu-

tations on elements of the network [68]. In addition to kinase mutations, MDA-MB-231 has an R280K TP53 mutation whose non-canonical activation of several signaling pathways is well characterized [80]. One such pathway is the mevalonate pathway, whose increased activity promotes tumor cell survival [90]. Mutant p53 also activates the TGFβ pathway by binding to and inhibiting p63, which typically inhibits the pathway, leading to increased cell migration [90, 91]. Furthermore, mutant p53 interacts functionally and physically with the vitamin D receptor to increase transcription of vitamin D response elements, transforming vitamin D into an antiapoptotic agent [92]. Increased vitamin D and TGFβ pathway signaling by mutant p53 causes an increase of the non-canonical WNT5A pathway, leading to increased cell migration [93]. As all of these pathways were identified and added to the network manually, an automated process to do so would be ideal, especially for tumors with a large number of GOF mutations.

4.2 Identification of Putative Reversion Interventions

Clustering of the estimated attractors for the experimental conditions of MDA-MB-231 and MCF10A demonstrates that applying SFA to the network can replicate experimental observations. Interestingly, the biological replicates of the cell lines cluster separately from each other, while their technical replicates cluster with them. Examination of steady state expression levels of the FVS nodes did not reveal large variation in their expression level among attractors of the same phenotype, indicating that other nodes in the attractors were causing this separation. Thus, an increase in the number of experimental replicates of the cancerous and normal-like cell line would help to better characterize the cancer attractors and normal-like attractors of the network.

The putative reversion targets identified by FVS control (AKT1, AURKA, CTNNB1, FOXM1, GSK3B, HDAC3, JUN, MAPK1, RELA, STAT3, TCF3, and PIAS1) are known regulators of EMT, stemness, tumor growth, and cell migration [3, 83, 94–110]. In fact,

several of them have been studied as therapeutic targets in TNBC [111, 112]. The caveat of these results is that network components are limited to information found in databases used for network construction. There is ample data regarding well-known oncogenes and drivers of tumorigenesis, but less common or unknown mechanisms of tumorigenesis are not as well characterized. Thus, the ability of this pipeline to identify novel reversion targets is restricted.

Estimating the dynamics of the network using SFA allows for the identification of perturbations of the FVS nodes that may shift the system towards a normal-like phenotype. It is important to consider that the output of SFA is a prediction of the system's steady state activity based on the topology of the network. Comparisons between the effect of a perturbation on a system predicted by simulating with SFA and by using a dynamical model of the system demonstrate that SFA replicates the DAC of each node observed from the dynamical model with 60-80% accuracy [41]. Therefore, the accuracy of virtual screening results is limited to the accuracy of SFA. Originally, the discrete datasets were used to reduce the emphasis on the magnitude of the SFA output as SFA was developed to predict the sign of the direction of activity change opposed to the accurate amount. However, the floating point values of the attractors and their DACs identified perturbations that resulted in the normal-like attractor when applied to all four cancerous initial conditions, indicating that they encode information about the attractor states that is necessary to associate them with the normal-like phenotype.

The 157 successful perturbations demonstrated clear patterns in the FVS node orientations. The exception to this is MAPK1, which is set in each orientation at a roughly equal frequency. Not all of the orientation patterns agree with literature. For example, Histone Deacetylace (HDAC) inhibitors have been used as anticancer agents and shown to partially reverse EMT, but HDAC3 knocked-in in 82% of the successful perturbations. However, the results from DepMap indicate that MDA-MB-231 is dependent on HDAC3, so its inhibition may induce cell death as opposed to reversion in MDA-MB-231. Thus, perturbations where

HDAC3 is left unchanged or knocked-in are given a higher priority. Perturbations where HDAC3 is left unchanged require a knock-in of JUN, which promotes stemness in TNBC, to trigger a shift to the normal-like cluster [113]. Therefore, only perturbations in which HDAC3 is knocked-in remain as potentially successful perturbations.

Additionally, Glycogen Synthase Kinase-3 Beta (GSK3B) and Beta-Catenin (CTNNNB1) are knocked-in in the majority of successful perturbations. These findings are contrary to prior evidence that GSK3B inhibition decreases EMT and Stemness and CTNNB1 inhibition has reversed EMT [100, 102, 111]. One possible explanation for these results is that in the cancerous and normal experimental data, GSK3B and CTNNB1 have higher normalized expression in the normal-like cell line than in the cancerous cell line (MDAMB231/MCF10A log2 fold change = -1.85 and -1.82, respectively). This observation was compared to normalized RNA-seq expression of MCF10A and MDA-MB-231 from GSE75168 [114]. The log2 fold change values from this study showed the same trend for GSK3B (logfc = -1.13), but the opposite trend for CTNNB1 (logfc = 1.56). These trends were further corroborated by RPKM data from the Library of Integrated Network-based Cellular Signatures [115]. Cell growth condition for these two studies can be found in Appendix A. Thus, lower expression of GSK3B in MDA-MB-231 than in MCF10A appears to be a trend, while the difference in expression for CTNNB1 is a characteristic of the data used in this study. If the normalized expression values of CTNNB1 and GSK3B for MDA-MB-231 and MCF10A are swapped, the results show that CTNNB1 should be knocked-out and GSK3B should be left unchanged to trigger a shift to the normal-like cluster. Therefore, the results of classification of attractors from virtual screening are sensitive to the initial conditions used to estimate attractors for the training set.

Of the 157 successful perturbations, 2 of them did not require the activation of GSK3B and CTNNB1. One of them requires the override of 10 FVS nodes while the other only requires 7. The combined knock-out of AKT1, FOXM1, JUN, RELA, and TCF3, and knock-in of HDAC3 and PIAS1 may trigger a shift to a normal-like phenotype in MDA-

MB-231. Aside from HDAC3, this result is supported by literature. PIAS1 is a regulator of EMT, metastasis and stemness. Its inhibition has increased metastasis in TNBC, although there is little information regarding its activation. FOXM1 inhibition leads to the reversal of EMT and has been shown to suppress MDA-MB-231 cell tumorigenesis [45]. Inhibition of AKT1, JUN, RELA, and TCF3 have been shown to reduce stemness and tumor formation in BC [106, 108, 109, 113]. The concerted perturbation of these targets has not yet been studied.

4.3 Limitations and Future Directions

This study analyzed a subset of all possible perturbations of the FVS nodes. An exhaustive search for successful perturbations of the system should be done by examining the remaining concerted FVS node perturbations and including source nodes as control targets. Although the introduction of source nodes as control targets will increase the number of potential nodes to control, some of them may already be set in the correct trajectory to reach the normal-like attractor. Thus, new perturbation combinations including source nodes could identify alternative putative reversion targets without requiring the activation of GSK3B and CTNNB1. Furthermore, this pipeline used normal-like attractors as the goal for classifying successful perturbations. However, a perturbation need not result in the attractor itself to trigger reversion. Moving the network state into the normal-like basin of attraction would be sufficient. Therefore, estimation of the attractor landscape - the attractors of the network and all the states in their basins - using SFA could aid in identifying perturbations that trigger a shift into the basin of the normal-like attractor. Preliminary analyses have demonstrated that this is possible, but refinement of the application of SFA for these purposes is necessary.

In cases where the majority of successful perturbations agree with literature, further prioritization is necessary to delimit perturbations to be experimentally validated. This can

be done by examining synthetic lethality among the targets. DepMap indicated that of the FVS nodes, MDA-MB-231 is dependent only on HDAC3, however concerted inhibition of multiple control targets could lead to cell death even if their individual inhibition does not [116, 117]. Hence, perturbations requiring the simultaneous inhibition of two such genes should be excluded from consideration as possible reversion strategies. Additional prioritization can be done by selecting perturbations that do not require the override of an FVS node involved in negative feedback loops. Recent work has shown that control of positive feedback loops is more important than control of negative feedback loops for controlling dynamical systems [118]. Thus, perturbations to nodes involved in positive feedback loops should be prioritized over those in negative feedback lops. Lastly, perturbations can be classified by the druggability of control nodes. Perturbations of nodes that are easier to control *in vivo* should be prioritized over more difficult interventions.

In using one CL cell line, the results can be interpreted as hypothesis-generating. Should the developed pipeline be applied to data of more CL TNBC cell lines and patient data, hypotheses of generalized targets for the reversion of CL TNBC tumors as a subtype can be made. This pipeline can be adapted and applied to other cell types and other therapeutic goals. For example, it can be used to identify targets for reversion of resistance or to trigger the expression of a phenotype that can be treated with current therapies.

Future work includes the validation of concerted perturbations triggering a shift to a normal-like cluster through the construction of a dynamical model. This is most easily done with a Boolean model, which represents system dynamics in the form of logic formalisms. A Boolean model can be obtained by training the signaling network with phosphoproteomic perturbation data to infer the system dynamics. Several software have been developed to identify candidate Boolean models from the signaling network using either a heuristic approach [119, 120] or via answer set programming [121]. Ultimately, concerted perturbations validated through dynamical modeling should be experimentally validated *in vitro.*

Chapter 5

Conclusions

The developed pipeline for the construction of an intracellular signaling network captures signaling pathways relevant to the cell it represents. The application of structure-based control to the signaling network successfully identifies nodes that can alter the long-term behavior of the system. This work shows that estimation of network dynamics by SFA can be used to predict the effect of perturbations on the long-term behavior of the system and identify putative concerted perturbations to trigger tumor reversion. Lastly, the concerted perturbation of AKT1, FOXM1, HDAC3, JUN, PIAS1, RELA, and TCF3 may trigger tumor reversion in the CL TNBC cell line MDA-MB-231. These findings are based on a literature review and should be further validated through experiments.

Appendix A

Cell Line Growth Conditions

Raw RNA-seq data for MDA-MB-231 and MCF10A was taken from Gene Expression Omnibus series GSE96860 for the construction of the network [43]. This study provided two biological and two technical replicates for each cell line. Both cell lines were grown at 37°C with 6.5% CO_2. MCF10A was grown in D-MEM with equal parts of alpha-MEM and Ham's F12 base media with the following additions: 0.1 M HEPES, 2 mM L-glutamine, 1% FBS, 0.035 mg/ml of Bovine Pituitary Extract, 0.01 mM Ascorbic Acid, 2 nM β-estradiol, 2.5 ng/ml Sodium Selenite, 10 nM Triiodothryonine, Ethanolamine, 1 µg/ml Insulin, 1 ng/ml Hydrocortisone, 0.1 mM Phosphoethanolamine, 0.01 mg/ml Transferrin, 12.5 ng/ml Epidermal Growth Factor, and 1% Penn/Strep. MDA-MB-231 was grown in α-MEM base media with the following additions: 0.1 M HEPES, 10% Fetal Calf Serum, 1% non-essential amino acids, 2 mM L-glutamine, 1% Sodium Pyruvate, 1 µg/ml insulin, 1 ng/ml Hydrocortisone, 12.5 ng/ml Epidermal Growth Factor, 1% Penn/Strep.

Additional RNA-seq data for the two cell lines was collected from GSE75168 to assesses expression trends in the dataset used in this study [114]. This data provided three biological replicates for both cell lines. In this study, MCF10A cells were grown in D-MEM with F12 (Hyclone-SH30271), 5% (v/v) horse serum (Gibco 16050 lot 1075876) + 10µg/ml human insulin (Sigma I-1882)+ 20ng/ml recombinant hEGF (Peprotech AF-100-15) + 100ng/ml

Cholera toxin (Sigma C-8052) + 0.5 µg/ml Hydrocortisone (Sigma H-0888) Pen/Strep (Life Technologies) and Glutamine (Life Technologies). MDA-MB-231 cells were grown in D-MEM with F12 (Hyclone-SH30271) + 10% (v/v) FBS (Atlanta Biologicals) Pen/Strep (Life Technologies) Glutamine (Life Technologies).

RPKM data for MDA-MB-231 and MCF10A from a third study was collected to further verify expression trends [115]. MDA-MB-231 was grown at 37°C in the absence of CO_2 in D-MEM + 10% FBS + 1% P/S. MCF10A was grown at 37°C with 5% CO_2 in D-MEM/F12 (1:1) + 5% HS + 1% P/S, 20ng/ml EGF, 0.5mg/ml HC, 10 µg/ml IN, and 100ng/ml CT.

Bibliography

[1] P Boyle. "Triple-negative breast cancer: epidemiological considerations and recommendations". In: *Annals of Oncology* 23 (Aug. 2012), pp. vi7–vi12. ISSN: 0923-7534. DOI: 10.1093/annonc/mds187. URL: https://doi.org/10.1093/annonc/mds187.

[2] *FDA Approves Trodelvy for Metastatic Triple-Negative Breast Cancer*. URL: https://www.ajmc.com/newsroom/fda-approves-trodelvy-for-metastatic-triplenegative-breast-cancer.

[3] Eleni Andreopoulou, Catherine M. Kelly, and Hayley M. McDaid. *Therapeutic Advances and New Directions for Triple-Negative Breast Cancer*. Mar. 2017. DOI: 10.1159/000455821.

[4] Hanan Ahmed Wahba and Hend Ahmed El-Hadaad. *Current approaches in treatment of triple-negative breast cancer*. June 2015. DOI: 10.7497/j.issn.2095-3941.2015.0030.

[5] Aditya Bardia et al. "Sacituzumab Govitecan-hziy in Refractory Metastatic Triple-Negative Breast Cancer". In: *New England Journal of Medicine* 380.8 (Feb. 2019), pp. 741–751. ISSN: 0028-4793. DOI: 10.1056/NEJMoa1814213. URL: http://www.nejm.org/doi/10.1056/NEJMoa1814213.

[6] Naveen K.R. Chalakur-Ramireddy and Suresh B. Pakala. *Combined drug therapeutic strategies for the effective treatment of Triple Negative Breast Cancer*. Jan. 2018. DOI: 10.1042/BSR20171357.

[7] Kay Dias et al. "Claudin-low breast cancer; clinical & pathological characteristics". In: *PLoS ONE* 12.1 (Jan. 2017). ISSN: 19326203. DOI: 10.1371/journal.pone.0168669.

[8] Aleix Prat et al. "Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer". In: *Breast Cancer Research* 12.5 (Sept. 2010). ISSN: 14655411. DOI: 10.1186/bcr2635.

[9] Deborah L. Holliday and Valerie Speirs. *Choosing the right cell line for breast cancer research*. Aug. 2011. DOI: 10.1186/bcr2889. URL: http://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr2889.

[10] Ana C. Garrido-Castro, Nancy U. Lin, and Kornelia Polyak. *Insights into molecular classifications of triple-negative breast cancer: Improving patient selection for treatment*. 2019. DOI: 10.1158/2159-8290.CD-18-1177.

[11] Ellaine Salvador, Malgorzata Burek, and Carola Y. Förster. *Tight Junctions and the Tumor Microenvironment*. Sept. 2016. DOI: 10.1007/s40139-016-0106-6.

[12] Joëlle Roche. *The epithelial-to-mesenchymal transition in cancer*. Feb. 2018. DOI: 10.3390/cancers10020052.

[13] Kornelia Polyak and Robert A. Weinberg. *Transitions between epithelial and mesenchymal states: Acquisition of malignant and stem cell traits*. Apr. 2009. DOI: 10.1038/nrc2620.

[14] Aleix Prat and Charles M. Perou. "Deconstructing the molecular portraits of breast cancer". In: *Molecular Oncology* 5.1 (Feb. 2011), pp. 5–23. ISSN: 15747891. DOI: 10.1016/j.molonc.2010.11.003. URL: http://doi.wiley.com/10.1016/j.molonc.2010.11.003.

[15] Adam Telerman and Robert Amson. *The molecular programme of tumour reversion: The steps beyond malignant transformation*. Mar. 2009. DOI: 10.1038/nrc2589.

[16] Gabriele D'Errico, Heather L. Machado, and Bruno Sainz. "A current perspective on cancer immune therapy: step-by-step approach to constructing the magic bullet". In: *Clinical and Translational Medicine* 6.1 (Dec. 2017). ISSN: 2001-1326. DOI: 10.1186/s40169-016-0130-5.

[17] Marilina García-Aranda and Maximino Redondo. "Immunotherapy: A challenge of breast cancer treatment". In: *Cancers* 11.12 (Dec. 2019). ISSN: 20726694. DOI: 10.3390/cancers11121822.

[18] Kwang-Hyun Cho et al. "Cancer reversion, a renewed challenge in systems biology". In: 2 (2017), pp. 48–57. DOI: 10.1016/j.coisb.2017.01.005. URL: http://dx.doi.org/10.1016/j.coisb.2017.01.005.

[19] ASKANAZY and M. "Die Teratome nach ihrem Bau, ihrem Verlauf, ihrer Genese und im Vergleich zum experimentellen Teratoid". In: *Verhandlungen der Deutschen Pathologischen Gesellschaft* 11 (1907), pp. 39–82.

[20] Wei Yan et al. "GATA3 inhibits breast cancer metastasis through the reversal of epithelial-mesenchymal transition". In: *Journal of Biological Chemistry* 285.18 (Apr. 2010), pp. 14042–14051. ISSN: 00219258. DOI: 10.1074/jbc.M110.105262.

[21] Isabel M. Chu et al. "Expression of GATA3 in MDA-MB-231 Triple-negative Breast Cancer Cells Induces a Growth Inhibitory Response to TGFß". In: *PLoS ONE* 8.4 (Apr. 2013). Ed. by Elad Katz, e61125. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0061125. URL: https://dx.plos.org/10.1371/journal.pone.0061125.

[22] I. M. Chu et al. "GATA3 inhibits lysyl oxidase-mediated metastases of human basal triple-negative breast cancer cells". In: *Oncogene* 31.16 (Apr. 2012), pp. 2017–2027. ISSN: 09509232. DOI: 10.1038/onc.2011.382.

[23] "Panobinostat suppresses the mesenchymal phenotype in anovel claudin-low triple negative patient-derived breast cancermodel". In: *Oncoscience* (Apr. 2018), p. 114. DOI: 10.18632/oncoscience.412.

[24] Sui Huang and Stuart Kauffman. *How to escape the cancer attractor: Rationale and limitations of multi-target drugs*. Aug. 2013. DOI: 10.1016/j.semcancer.2013.06.003.

[25] Sui Huang, Ingemar Ernberg, and Stuart Kauffman. *Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective*. 2009. DOI: 10.1016/j.semcdb.2009.07.003.

[26] C. H. Waddington. "Canalization of development and genetic assimilation of acquired characters". In: *Nature* 183.4676 (1959), pp. 1654–1655. ISSN: 00280836. DOI: 10.1038/1831654a0.

[27] Stuart Kauffman. "Differentiation of malignant to benign cells". In: *Journal of Theoretical Biology* 31.3 (June 1971), pp. 429–451. ISSN: 10958541. DOI: 10.1016/0022-5193(71)90020-8.

[28] Jorge G.T. Zañudo and Réka Albert. "Cell Fate Reprogramming by Control of Intracellular Network Dynamics". In: *PLoS Computational Biology* 11.4 (Apr. 2015). ISSN: 15537358. DOI: 10.1371/journal.pcbi.1004193.

[29] Alexander J. Gates and Luis M. Rocha. "Control of complex networks requires both structure and dynamics". In: *Scientific Reports* 6 (Apr. 2016). ISSN: 20452322. DOI: 10.1038/srep24456.

[30] R E Kalman. *Contributions to the Theory of Optimal Control*. Tech. rep.

[31] Sean P. Cornelius and Adilson E. Motter. "NECO - A scalable algorithm for NEtwork COntrol". In: *Protocol Exchange* (July 2013). DOI: 10.1038/protex.2013.063. URL: http://arxiv.org/abs/1307.2582%20http://dx.doi.org/10.1038/protex.2013.063.

[32] Daniel K. Wells, William L. Kath, and Adilson E. Motter. "Control of stochastic and induced switching in biophysical networks". In: *Physical Review X* 5.3 (Sept. 2015), p. 031036. ISSN: 21603308. DOI: 10.1103/PhysRevX.5.031036.

[33] David Murrugarra et al. "Identification of control targets in Boolean molecular network models via computational algebra". In: *BMC Systems Biology* 10.1 (Sept. 2016), p. 94. ISSN: 17520509. DOI: 10.1186/s12918-016-0332-x. URL: http://bmcsystbiol.biomedcentral.com/articles/10.1186/s12918-016-0332-x.

[34] Jorge Gomez Tejeda Zañudo et al. "Structure-based control of complex networks with nonlinear dynamics". In: *Proceedings of the National Academy of Sciences of the United States of America* 114.28 (July 2017), pp. 7234–7239. ISSN: 10916490. DOI: 10.1073/pnas.1617387114.

[35] Atsushi Mochizuki et al. "Dynamics and control at feedback vertex sets. II: A faithful monitor to determine the diversity of molecular activities in regulatory networks". In: *Journal of Theoretical Biology* 335 (Oct. 2013), pp. 130–146. ISSN: 00225193. DOI: 10.1016/j.jtbi.2013.06.009.

[36] Juan M. Cejalvo et al. "Intrinsic subtypes and gene expression profiles in primary and metastatic breast cancer". In: *Cancer Research* 77.9 (May 2017), pp. 2213–2221. ISSN: 15387445. DOI: 10.1158/0008-5472.CAN-16-2717.

[37] Luis Sordo Vieira et al. "Computing Signal Transduction in Signaling Networks modeled as Boolean Networks, Petri Nets, and Hypergraphs". In: (2019). DOI: 10.1101/272344. URL: https://doi.org/10.1101/272344.

[38] Ulykbek Kairov et al. "Network analysis of gene lists for finding reproducible prognostic breast cancer gene signatures". In: *Bioinformation* 8.16 (Aug. 2012), pp. 773–776. ISSN: 09738894. DOI: 10.6026/97320630008773.

[39] Trey Ideker and Ruth Nussinov. "Network approaches and applications in biology". In: (2017). DOI: 10.1371/journal.pcbi.1005771. URL: https://doi.org/10.1371/journal.pcbi.1005771.

[40] Wei Keat Lim, Eugenia Lyashenko, and Andrea Califano. *MASTER REGULATORS USED AS BREAST CANCER METASTASIS CLASSIFIER**. Tech. rep.

[41] Daewon Lee and Kwang Hyun Cho. "Topological estimation of signal flow in complex signaling networks". In: *Scientific Reports* 8.1 (Dec. 2018), pp. 1–11. ISSN: 20452322. DOI: 10.1038/s41598-018-23643-5.

[42] Ying Qu et al. "Evaluation of MCF10A as a Reliable Model for Normal Human Mammary Epithelial Cells". In: *PLOS ONE* 10.7 (July 2015). Ed. by Xuefeng Liu, e0131285. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0131285. URL: https://dx.plos.org/10.1371/journal.pone.0131285.

[43] Yuanxin Xi et al. "Histone modification profiling in breast cancer cell lines highlights commonalities and differences among subtypes". In: *BMC Genomics* 19.1 (Feb. 2018). ISSN: 14712164. DOI: 10.1186/s12864-018-4533-0.

[44] Anneleen Daemen et al. "Modeling precision treatment of breast cancer". In: *Genome Biology* 14.10 (Dec. 2013). ISSN: 1474760X. DOI: 10.1186/gb-2013-14-10-r110.

[45] John G Tate et al. "COSMIC: the Catalogue Of Somatic Mutations In Cancer". eng. In: *Nucleic acids research* 47.D1 (Jan. 2019), pp. D941–D947. ISSN: 1362-4962. DOI: 10.1093/nar/gky1015. URL: https://pubmed.ncbi.nlm.nih.gov/30371878%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6323903/.

[46] Robert T. Lawrence et al. "The Proteomic Landscape of Triple-Negative Breast Cancer". In: *Cell Reports* 11.4 (Apr. 2015), pp. 630–644. ISSN: 22111247. DOI: 10.1016/j.celrep.2015.03.050.

[47] Michael I Love, Wolfgang Huber, and Simon Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome Biology* 15.12 (2014), p. 550. DOI: 10.1186/s13059-014-0550-8.

[48] Komal S. Rathi et al. "Correcting transcription factor gene sets for copy number and promoter methylation variations". In: *Drug Development Research* 75.6 (2014), pp. 343–347. ISSN: 10982299. DOI: 10.1002/ddr.21220.

[49] Eric Bonnet et al. "BiNoM 2.0, a Cytoscape plugin for accessing and analyzing pathways using standard systems biology formats". In: *BMC Systems Biology* 7 (Mar. 2013). ISSN: 17520509. DOI: 10.1186/1752-0509-7-18.

[50] Paul Shannon et al. "Cytoscape: A software Environment for integrated models of biomolecular interaction networks". In: *Genome Research* 13.11 (Nov. 2003), pp. 2498–2504. ISSN: 10889051. DOI: 10.1101/gr.1239303.

[51] Suraj Peri et al. "Development of human protein reference database as an initial platform for approaching systems biology in humans". In: *Genome Research* 13.10 (Oct. 2003), pp. 2363–2371. ISSN: 10889051. DOI: 10.1101/gr.1680803.

[52] T S Keshava Prasad et al. "Human Protein Reference Database–2009 update". eng. In: *Nucleic acids research* 37.Database issue (Jan. 2009), pp. D767–D772. ISSN: 1362-4962. DOI: 10.1093/nar/gkn892. URL: https://pubmed.ncbi.nlm.nih.gov/18988627%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2686490/.

[53] Xinwei She et al. "Definition, conservation and epigenetics of housekeeping and tissue-enriched genes". In: *BMC Genomics* 10 (June 2009). ISSN: 14712164. DOI: 10.1186/1471-2164-10-269.

[54] Pan Du et al. "From disease ontology to disease-ontology lite: statistical methods to adapt a general-purpose ontology for the test of gene-ontology associations". eng. In: *Bioinformatics (Oxford, England)* 25.12 (June 2009), pp. i63–i68. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btp193. URL: https://pubmed.ncbi.nlm.nih.gov/19478018%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2687947/.

[55] Lei Huang et al. *GeneAnswers: Integrated Interpretation of Genes*. 2019.

[56] I. Kuperstein et al. "Atlas of Cancer Signalling Network: A systems biology resource for integrative analysis of cancer data with Google Maps". In: *Oncogenesis* 4.7 (July 2015). ISSN: 21579024. DOI: 10.1038/oncsis.2015.19.

[57] E Wingender et al. "TRANSFAC: an integrated system for gene expression regulation". eng. In: *Nucleic acids research* 28.1 (Jan. 2000), pp. 316–319. ISSN: 0305-1048. DOI: 10.1093/nar/28.1.316. URL: https://pubmed.ncbi.nlm.nih.gov/10592259%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC102445/.

[58] A E Kel et al. "MATCH: A tool for searching transcription factor binding sites in DNA sequences". eng. In: *Nucleic acids research* 31.13 (July 2003), pp. 3576–3579. ISSN: 1362-4962. DOI: 10.1093/nar/gkg585. URL: https://pubmed.ncbi.nlm.nih.gov/12824369%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC169193/.

[59] F Kolpakov et al. "GeneXplain — Identification of Causal Biomarkers and Drug Targets in Personalized Cancer Pathways". eng. In: *Journal of Biomolecular Techniques : JBT* 22.Suppl (Oct. 2011), S16–S16. ISSN: 1524-0215. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3186619/%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3186619/.

[60] Andreas Krämer et al. "Causal analysis approaches in Ingenuity Pathway Analysis". eng. In: *Bioinformatics (Oxford, England)* 30.4 (Feb. 2014), pp. 523–530. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btt703. URL: https://pubmed.ncbi.nlm.nih.gov/24336805%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3928520/.

[61] Francesco Ceccarelli et al. "Bringing data from curated pathway resources to Cytoscape with OmniPath". In: *Bioinformatics* 36.8 (Dec. 2019), pp. 2632–2633. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btz968. URL: https://doi.org/10.1093/bioinformatics/btz968.

[62] Luz Garcia-Alonso et al. "Benchmark and integration of resources for the estimation of human transcription factor activities". In: *Genome Research* 29.8 (2019), pp. 1363–1375. ISSN: 15495469. DOI: 10.1101/gr.240663.118.

[63] *Software/ExpressionCorrelation - Bader Lab @ The University of Toronto*. URL: http://baderlab.org/Software/ExpressionCorrelation.

[64] William McLaren et al. "The Ensembl Variant Effect Predictor". In: *Genome Biology* 17.1 (June 2016). ISSN: 1474760X. DOI: 10.1186/s13059-016-0974-4.

[65] Rebecca F. Lowdon and Ting Wang. "Epigenomic annotation of noncoding mutations identifies mutated pathways in primary liver cancer". In: *PLoS ONE* 12.3 (Mar. 2017). ISSN: 19326203. DOI: 10.1371/journal.pone.0174032.

[66] Mathias Krull et al. "TRANSPATH: an integrated database on signal transduction and a tool for array analysis". eng. In: *Nucleic acids research* 31.1 (Jan. 2003), pp. 97–100. ISSN: 1362-4962. DOI: 10.1093/nar/gkg089. URL: https://pubmed.ncbi.nlm.nih.gov/12519957%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC165536/.

[67] Alexander E Kel et al. "Multi-omics "upstream analysis" of regulatory genomic regions helps identifying targets against methotrexate resistance of colon cancer". eng. In: *EuPA open proteomics* 13 (Sept. 2016), pp. 1–13. ISSN: 2212-9685. DOI: 10.1016/j.euprot.2016.09.002. URL: https://pubmed.ncbi.nlm.nih.gov/29900117%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5988513/.

[68] Pau Creixell et al. "Kinome-wide Decoding of Network-Attacking Mutations Rewiring Cancer Signaling". In: *Cell* 163.1 (Sept. 2015), pp. 202–217. ISSN: 10974172. DOI: 10.1016/j.cell.2015.08.056.

[69] Pau Creixell et al. "Unmasking Determinants of Specificity in the Human Kinome". In: *Cell* 163.1 (Sept. 2015), pp. 187–201. ISSN: 10974172. DOI: 10.1016/j.cell.2015.08.057.

[70] Heiko Horn et al. *KinomeXplorer: An integrated platform for kinome biology studies*. 2014. DOI: 10.1038/nmeth.2968.

[71] W N Venables and B D Ripley. *Modern Applied Statistics with S*. Fourth. New York: Springer, 2002. URL: http://www.stats.ox.ac.uk/pub/MASS4.

[72] F Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[73] Peter J. Rousseeuw. "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". In: *Journal of Computational and Applied Mathematics* 20.C (Nov. 1987), pp. 53–65. ISSN: 03770427. DOI: 10.1016/0377-0427(87)90125-7.

[74] Lauren Marazzi, Andrew Gainer-Dewar, and Paola Vera-Licona. *OCSANA+: Optimal Control and Simulation of Signaling Networks from Network Analysis*. 2019. DOI: 10.1101/806315. URL: https://doi.org/10.1101/806315.

[75] Philippe Galinier, Eunice Lemamou, and Mohamed Wassim Bouzidi. "Applying local search to the feedback vertex set problem". In: *Journal of Heuristics* 19.5 (Oct. 2013), pp. 797–818. ISSN: 13811231. DOI: 10.1007/s10732-013-9224-z.

[76] Joshua M Dempster et al. "Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines". In: *bioRxiv* (Jan. 2019), p. 720243. DOI: 10.1101/720243. URL: http://biorxiv.org/content/early/2019/07/31/720243.abstract.

[77] Paul T.C. Wan et al. "Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF". In: *Cell* 116.6 (Mar. 2004), pp. 855–867. ISSN: 00928674. DOI: 10.1016/S0092-8674(04)00215-6.

[78] Matthew Holderfield et al. "RAF Inhibitors Activate the MAPK Pathway by Relieving Inhibitory Autophosphorylation". In: *Cancer Cell* 23.5 (May 2013), pp. 594–602. ISSN: 15356108. DOI: 10.1016/j.ccr.2013.03.033.

[79] Patrick Kwok Shing Ng et al. "Systematic Functional Annotation of Somatic Mutations in Cancer". In: *Cancer Cell* 33.3 (Mar. 2018), pp. 450–462. ISSN: 18783686. DOI: 10.1016/j.ccell.2018.01.021.

[80] Dawid Walerych et al. "The rebel angel: mutant p53 as the driving oncogene in breast cancer". eng. In: *Carcinogenesis* 33.11 (Nov. 2012), pp. 2007–2017. ISSN: 1460-2180. DOI: 10.1093/carcin/bgs232. URL: https://pubmed.ncbi.nlm.nih.gov/22822097%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3483014/.

[81] Yun Hee Bae et al. "Mutant p53-Notch1 signaling axis is involved in curcumin-induced apoptosis of breast cancer cells". In: *Korean Journal of Physiology and Pharmacology* 17.4 (Aug. 2013), pp. 291–297. ISSN: 12264512. DOI: 10.4196/kjpp.2013.17.4.291.

[82] Bin Liu et al. "PIAS1 regulates breast tumorigenesis through selective epigenetic gene silencing". In: *PLoS ONE* 9.2 (Feb. 2014). ISSN: 19326203. DOI: 10.1371/journal.pone.0089464.

[83] Shorafidinkhuja Dadakhujaev et al. "A novel role for the SUMO E3 ligase PIAS1 in cancer metastasis". In: *Oncoscience* 1.3 (2014), pp. 229–240. ISSN: 23314737. DOI: 10.18632/oncoscience.27.

[84] Andrea Rabellino, Cristina Andreani, and Pier Paolo Scaglioni. *The Role of PIAS SUMO E3-Ligases in Cancer*. 2017. DOI: 10.1158/0008-5472.CAN-16-2958.

[85] Michael C. Wu and Xihong Lin. *Prior biological knowledge-based approaches for the analysis of genome-wide expression profiles using gene sets and pathways*. 2009. DOI: 10.1177/0962280209351925.

[86] Aravind Subramanian et al. "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles". In: *Proceedings of the National Academy of Sciences of the United States of America* 102.43 (Oct. 2005), pp. 15545–15550. ISSN: 00278424. DOI: 10.1073/pnas.0506580102.

[87] Vamsi K. Mootha et al. "PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes". In: *Nature Genetics* 34.3 (July 2003), pp. 267–273. ISSN: 10614036. DOI: 10.1038/ng1180.

[88] Jelle J Goeman et al. "A global test for groups of genes: testing association with a clinical outcome". In: *Bioinformatics* 20.1 (Jan. 2004), pp. 93–99. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btg382. URL: https://doi.org/10.1093/bioinformatics/btg382.

[89] John Tomfohr, Jun Lu, and Thomas B. Kepler. "Pathway level analysis of gene expression using singular value decomposition". In: *BMC Bioinformatics* 6 (Sept. 2005). ISSN: 14712105. DOI: 10.1186/1471-2105-6-225.

[90] Patricia A.J. Muller, Karen H. Vousden, and Jim C. Norman. *p53 and its mutants in tumor cell migration and invasion*. Jan. 2011. DOI: 10.1083/jcb.201009059.

[91] Maddalena Adorno et al. "A Mutant-p53/Smad Complex Opposes p63 to Empower TGFβ-Induced Metastasis". In: *Cell* 137.1 (Apr. 2009), pp. 87–98. ISSN: 00928674. DOI: 10.1016/j.cell.2009.01.039.

[92] Perry Stambolsky et al. "Modulation of the Vitamin D3 Response by Cancer-Associated Mutant p53". In: *Cancer Cell* 17.3 (Mar. 2010), pp. 273–285. ISSN: 15356108. DOI: 10.1016/j.ccr.2009.11.025.

[93] Rosa Serra et al. *Wnt5a as an effector of TGFβ in mammary development and cancer*. June 2011. DOI: 10.1007/s10911-011-9205-5.

[94] Jennifer M Sahni et al. "BET inhibitors suppress Aurora kinases in breast cancer Bromodomain and extraterminal protein inhibition blocks growth of triple-negative breast cancers through the suppression of Aurora kinases". In: (2016). DOI: 10.1074/jbc.M116.738666. URL: http://www.jbc.org/cgi/doi/10.1074/jbc.M116.738666.

[95] Alexey A. Leontovich et al. "NOTCH3 expression is linked to breast cancer seeding and distant metastasis". In: *Breast Cancer Research* 20.1 (Sept. 2018), p. 105. ISSN: 1465542X. DOI: 10.1186/s13058-018-1020-0. URL: https://breast-cancer-research.biomedcentral.com/articles/10.1186/s13058-018-1020-0.

[96] Jie Yuan, Fei Zhang, and Ruifang Niu. "Multiple regulation pathways and pivotal biological functions of STAT3 in cancer". In: *Scientific Reports* 5.1 (Dec. 2015), pp. 1–10. ISSN: 20452322. DOI: 10.1038/srep17663.

[97] Alexander Ring et al. "CBP/B-catenin/FOXM1 is a novel therapeutic target in triple negative breast cancer". In: *Cancers* 10.12 (Dec. 2018). ISSN: 20726694. DOI: 10.3390/cancers10120525.

[98] Ruth M. O'Regan and Rita Nahta. *Targeting forkhead box M1 transcription factor in breast cancer*. Aug. 2018. DOI: 10.1016/j.bcp.2018.05.019.

[99] Yanli Tan et al. "Identification of FOXM1 as a specific marker for triple-negative breast cancer". In: *International Journal of Oncology* 54.1 (Jan. 2019), pp. 87–97. ISSN: 17912423. DOI: 10.3892/ijo.2018.4598.

[100] Hongyuan Li et al. "Luteolin suppresses the metastasis of triple-negative breast cancer by reversing epithelial-to-mesenchymal transition via downregulation of β-catenin expression". In: *Oncology Reports* 37.2 (Feb. 2017), pp. 895–902. ISSN: 17912431. DOI: 10.3892/or.2016.5311.

[101] Nan Wu et al. *Precision medicine based on tumorigenic signaling pathways for triple-negative breast cancer*. Oct. 2018. DOI: 10.3892/ol.2018.9290.

[102] Geraldine Vidhya Vijay et al. "GSK3β regulates epithelial-mesenchymal transition and cancer stem cell properties in triple-negative breast cancer". In: *Breast Cancer Research* 21.1 (Mar. 2019), pp. 1–14. ISSN: 1465542X. DOI: 10.1186/s13058-019-1125-0.

[103] Ezanee Azlina Mohamad Hanif and Shamsul Azhar Shah. *Overview on epigenetic reprogramming: A potential therapeutic intervention in triple negative breast cancers*. Dec. 2018. DOI: 10.31557/APJCP.2018.19.12.3341.

[104] Sumayah Al-Mahmood et al. *Metastatic and triple-negative breast cancer: challenges and treatment options*. Oct. 2018. DOI: 10.1007/s13346-018-0551-3.

[105] Jack J. Chan, Tira J.Y. Tan, and Rebecca A. Dent. *Novel therapeutic avenues in triple-negative breast cancer: PI3K/AKT inhibition, androgen receptor blockade, and beyond*. 2019. DOI: 10.1177/1758835919880429.

[106] Richard L. Carpenter et al. "Combined inhibition of AKT and HSF1 suppresses breast cancer stem cells and tumor growth". In: *Oncotarget* 8.43 (May 2017), pp. 73947–73963. ISSN: 19492553. DOI: 10.18632/oncotarget.18166.

[107] Abdullah Alsuliman et al. "Bidirectional crosstalk between PD-L1 expression and epithelial to mesenchymal transition: Significance in claudin-low breast cancer cells". In: *Molecular Cancer* 14.1 (Aug. 2015), pp. 1–13. ISSN: 14764598. DOI: 10.1186/s12943-015-0421-2.

[108] Michal Slyper et al. "Control of breast cancer growth and initiation by the stem cell-associated transcription factor TCF3". In: *Cancer Research* 72.21 (Nov. 2012), pp. 5613–5624. ISSN: 00085472. DOI: 10.1158/0008-5472.CAN-12-0119.

[109] Phungern Khongthong, Antonia K Roseweir, and Joanne Edwards. "The NF-KB pathway and endocrine therapy resistance in breast cancer". English. In: *Endocrine-Related Cancer* 26.6 (2019), R369–R380. DOI: 10.1530/ERC-19-0087. URL: https://erc.bioscientifica.com/view/journals/erc/26/6/ERC-19-0087.xml.

[110] Cornelia Braicu et al. *A comprehensive review on MAPK: A promising therapeutic target in cancer*. Oct. 2019. DOI: 10.3390/cancers11101618.

[111] Geraldine Vidhya Raja. "GSK3B regulates epithelial-mesenchymal transition and cancer stem cell properties and is a novel drug target for triple-negative breast cancer". In: *UT GSBS Dissertations and Theses (Open Access)* (Aug. 2017). URL: https://digitalcommons.library.tmc.edu/utgsbs_dissertations/806.

[112] Jiang-Jiang Qin et al. "STAT3 as a potential therapeutic target in triple negative breast cancer: a systematic review". In: (). DOI: 10.1186/s13046-019-1206-z. URL: https://doi.org/10.1186/s13046-019-1206-z.

[113] X. Xie et al. "C-Jun N-terminal kinase promotes stem cell phenotype in triple-negative breast cancer through upregulation of Notch1 via activation of c-Jun". In: *Oncogene* 36.18 (May 2017), pp. 2599–2608. ISSN: 14765594. DOI: 10.1038/onc.2016.417.

[114] Terri L. Messier et al. "Histone H3 lysine 4 acetylation and methylation dynamics define breast cancer subtypes". In: *Oncotarget* 7.5 (2016), pp. 5094–5109. ISSN: 19492553. DOI: 10.18632/oncotarget.6922.

[115] Caitlin Mills, Marc Hafner, and Sarah Boswell. *Breast Cancer Profiling Project, Gene Expression 1: Baseline mRNA sequencing on 35 breast cell lines, lincs, V1.* 2018. URL: http://lincsportal.ccs.miami.edu/datasets/#/view/LDS-1530.

[116] Nigel J. O'Neil, Melanie L. Bailey, and Philip Hieter. *Synthetic lethality and cancer.* Oct. 2017. DOI: 10.1038/nrg.2017.47.

[117] Sebastian M.B. Nijman. *Synthetic lethality: General principles, utility and detection using genetic screens in human cells.* Jan. 2011. DOI: 10.1016/j.febslet.2010.11.024.

[118] Jordan C. Rozum and Réka Albert. "Self-sustaining positive feedback loops in discrete and continuous systems". In: *Journal of Theoretical Biology* 459 (Dec. 2018), pp. 36–44. ISSN: 10958541. DOI: 10.1016/j.jtbi.2018.09.017.

[119] Camille Terfve et al. "CellNOptR: A flexible toolkit to train protein signaling networks to data using multiple logic formalisms". In: *BMC Systems Biology* 6 (Oct. 2012). ISSN: 17520509. DOI: 10.1186/1752-0509-6-133.

[120] Julien Dorier et al. "Boolean regulatory network reconstruction using literature based knowledge with a genetic algorithm optimization method". In: *BMC Bioinformatics* 17.1 (Oct. 2016). ISSN: 14712105. DOI: 10.1186/s12859-016-1287-z.

[121] Carito Guziolowski et al. "Exhaustively characterizing feasible logic models of a signaling network using Answer Set Programming". eng. In: *Bioinformatics (Oxford, England)* 29.18 (Sept. 2013), pp. 2320–2326. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btt393. URL: https://pubmed.ncbi.nlm.nih.gov/23853063%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3753570/.