INTERPRETABLE MECHANISTIC AND MACHINE LEARNING MODELS FOR PREDICTING CARDIAC REMODELING FROM BIOCHEMICAL AND BIOMECHANICAL FEATURES

A Dissertation
Presented to
The Graduate School of
Clemson University

In Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy
Biomedical Data Science and Informatics

---

by
Anamul Haque
December 2023

Accepted by:
Nina C. Hubig, PhD, Committee Chair
William J. Richardson, PhD
Brian C. Dean, PhD
Bethany J. Wolf, PhD
Awe J. Schoepf, MD

Abstract

Biochemical and biomechanical signals drive cardiac remodeling, resulting in altered heart physiology and the precursor for several cardiac diseases, the leading cause of death for most racial groups in the USA. Reversing cardiac remodeling requires medication and device-assisted treatment such as Cardiac Resynchronization Therapy (CRT), but current interventions produce highly variable responses from patient to patient. Mechanistic modeling and Machine learning (ML) approaches have the functionality to aid diagnosis and therapy selection using various input features. Moreover, 'Interpretable' machine learning methods have helped make machine learning models fairer and more suited for clinical application. The overarching objective of this doctoral work is to develop computational models that combine an extensive array of clinically measured biochemical and biomechanical variables to enable more accurate identification of heart failure patients prone to respond positively to therapeutic interventions. In the first aim, we built an ensemble ML classification algorithm using previously acquired data from the SMART-AV CRT clinical trial. Our classification algorithm incorporated 26 patient demographic and medical history variables, 12 biomarker variables, and 18 LV functional variables, yielding correct CRT response prediction in 71% of patients. In the second aim, we employed a machine learning-based method to infer the fibrosis-related gene regulatory network from RNA-seq data from the MAGNet cohort of heart failure patients. This network identified significant interactions between transcription factors and cell synthesis outputs related to cardiac fibrosis - a critical driver of heart failure. Novel filtering methods helped us prioritize the most critical regulatory interactions of mechanistic forward simulations. In the third aim, we developed a logic-based model for the mechanistic network of cardiac fibrosis, integrating the gene regulatory network derived from aim two into a previously constructed cardiac

fibrosis signaling network model. This integrated model implemented biochemical and biomechanical reactions as ordinary differential equations based on normalized Hill functions. The model elucidated the semi-quantitative behavior of cardiac fibrosis signaling complexity by capturing multi-pathway crosstalk and feedback loops. Perturbation analysis predicted the most critical nodes in the mechanistic model. Patient-specific simulations helped identify which biochemical species highly correlate with clinical measures of patient cardiac function.

Dedication

I want to dedicate this dissertation work to my father, who died from a sudden cardiac arrest on the first day of graduate school. His passion for education has brought me to this country for higher education and helped me find my passion for cardiovascular disease research. I also want to dedicate this to my mother, whose whole life has passed raising our six siblings. My dedication also includes my siblings, especially my elder brother, for his endless contributions to shaping our family's bright future.

My Ph.D. journey would not be possible without the support and encouragement of the love of my life, my wife, Nazneen Sultana. She believed me and supported my decision to move to Biomedical Data Science. Without her, this journey would be unthinkable. Finally, I want to dedicate this to my son, Ibrazul Haque. His presence around me was uplifting and helped me to stay spirited.

Acknowledgement

I sincerely thank my adviser, Dr. William J. Richardson, for allowing me to work in his system mechanobiology research group for my dissertation. His profound knowledge and expertise in cardiac mechanobiology helped me to ignite my passion for cardiovascular disease research. His nurturing guidance, patience, and compassion helped me become a proficient researcher and a compassionate individual. Dr. Richrdson introduced me to the field of interpretable white box mechanistic models alongside machine learning techniques. His expert advice propelled me through this academic journey. Next, I would like to thank my committee chair Dr. Nina C. Hubig, for her encouragement to incorporate interpretable machine-learning approaches into my dissertation. I am also thankful to Dr. Brian C. Dean for wholeheartedly accepting me into the Biomedical Data Science and Informatics (BDSI) Ph.D. program. His confidence in my capabilities is a consistent source of motivation. I am also indebted to Dr. Bethany J. Wolf for her invaluable contributions as a committee member and for offering insightful feedback on my research proposal. Finally, I am grateful to Dr. Awe J. Schoepf for sparing his precious time from a demanding schedule to serve on my committee.

I am also grateful to the members of the System Mechanobiology Lab: Brendyn Miller, Jonathan Heywood, and Drs. Jesse Rogers, Amir Yeganegi, Kelsey Watts, Sam Coeyman, and Jake Potter. Special thanks to Dr. Jesse Rogers and Dr. Kelsey Watts for their well-written codes that I also used in my projects. I am also thankful to my BDSI cohort, especially Drs. MinJae Woo and Paritra Mandal for their constant encouragement and lengthy discussions. Finally, I express my heartfelt gratitude to my family members, especially my wife, Nazneen Sultana, for providing me with unfailing support and continuous

encouragement throughout my research and writing process. This milestone would not have been possible without them.

List of Figures

List of Tables

Table of Contents

Chapter 1

Literature Review

1.1 Hypertension, cardiac remodeling, risk factors, and guidelines

Hypertension is the marked increase in systolic and diastolic blood pressure from the normal level. We can classify this increase into different categories based on the measurements of systolic and diastolic pressure, which is applicable for all age groups. Hypertension for a prolonged time results in future health complications and vital organ failures such as heart, brain, kidney, and eyes [1]. According to World Health Organization (WHO), an estimated 1.28 billion adults aged 30-79 years have hypertension worldwide [2]. An estimated 46% of adults with hypertension are unaware of the presence of hypertension. Only 1 in 5 adults with hypertension have it under control. WHO targeted hypertension as the leading non-communicable disease to eradicate by 2030. Nearly half of the US adults (47%) have hypertension. Only 1 in 4 adults with hypertension have their condition under control [2]. About 34 million adults in the USA take hypertension medication. Hypertension is the contributing factor to nearly half a million death per year in the USA [3].

Several factors increase the susceptibility to develop hypertension [4]. The risk of hypertension increases in elderly, male, and black individuals more prone to die from hypertension [5]. Individuals' family history also has a role in hypertension susceptibility. The presence of co-morbidity in parents risks the offspring developing hypertension. Higher BMI, physical inactivity, and a sedentary lifestyle also play a contributing factor in hypertension. Dietary habits such as high sodium, sugar, and alcohol intake are also positively correlated with hypertension. Stress and certain kidney diseases also increase the

probability of dying from hypertension. According to the European Society of Hypertension guideline (ESC/ESH guidelines on arterial hypertension (management of), hypertension evaluation consists of four categories based on their systolic and diastolic blood pressure range: Normal, Elevated, Stage 1, and Stage 2 hypertension. Hypertension evaluation also considers the secondary analysis of biochemical panels such as fasting glucose, lipid panel, thyroid-stimulating hormone. Several other biomechanical assays used for the assessment include eGFR, electrocardiogram, urinalysis [6].

Cardiac remodeling is the biochemical and biomechanical manifestation of the heart that results in the heart's change in size, mass, geometry, and function [7]. Cardiac remodeling primarily focuses on the heart's left ventricle because this is responsible for pumping oxygenated blood throughout the body. Cardiac function follows the simple Laplace formula ($T = Pr/2h$), where T is LV wall stress, P is pressure, r is the radius of the ventricle, and h is the LV wall thickness [8]. Increased blood pressure increases the stress on the LV wall so LV wall thickens to offset this pressure and normalize wall stress. This thickening results in the development of hypertrophy. There are two types of LV hypertrophy: concentric and eccentric. When elevated blood pressure increases LV wall stress, LV walls thicken (h) to offset this and normalize the wall stress. This is known as concentric hypertrophy. Eccentric hypertrophy develops when high blood volume increases the radius (r) of the heart chamber (V). The structure of the left ventricle is not only affected by pressure or volume overloads. Several other factors such as ethnicity, gender, neurohormonal characteristics, environmental and genetic factors are also responsible and may have additive effects in the overload of the cardiac chambers. Underlying medical conditions such as diabetes and chronic kidney diseases are the two confounders of cardiac remodeling [9]. Cardiac remodeling is reversible by doing regular exercise and lifestyle

modification [10]. Some studies have shown a negative correlation between physical activity and cardiac hypertrophy development.

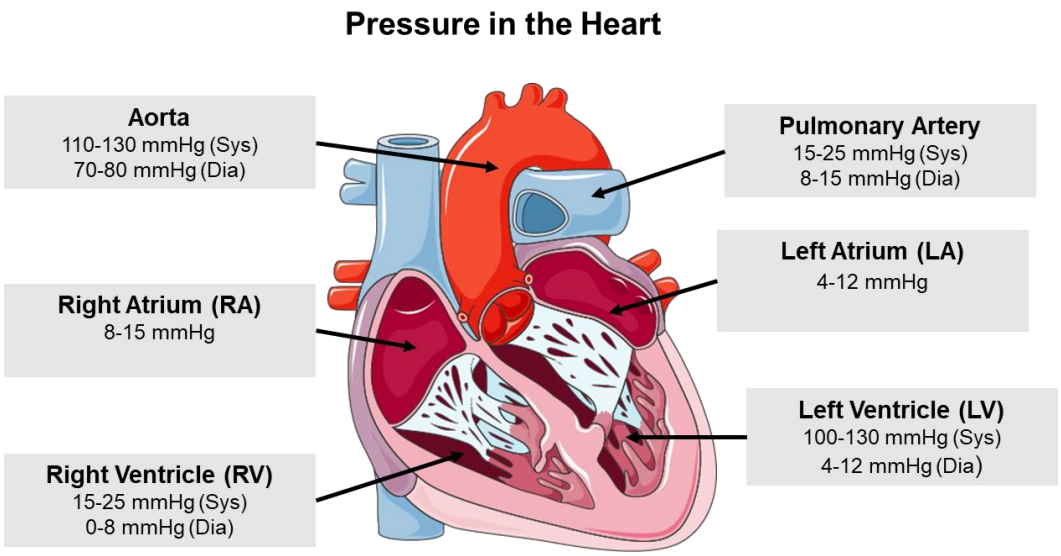


Figure 1.1. Pressure on a non-hypertensive heart. Illustration modified from Servier medical art (https://smart.servier.com/) and conceptualized from Elira Maksuti [11]

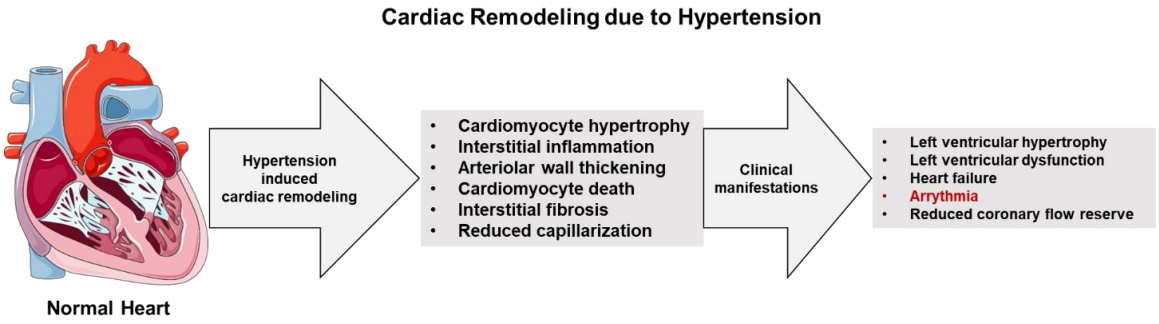


Figure 1.2. Cardiac remodeling due to hypertension and its clinical manifestation. Illustration modified from Servier medical art (https://smart.servier.com/) [9].

1.2 Biochemical markers for predicting cardiac remodeling

There are several vital biomarkers available for myocardial organ damage in hypertension [11]. Cardiac troponin is the first group of biomarkers that is part of the contractile apparatus of cardiac myocytes. Myocardial injury and cardiac remodeling increase the concentration of cardiac troponin [12]. The most common biomarkers for cardiac damage are C-reactive proteins (CRP) [13]. Many studies have shown that CRP concentration is associated with the increased risks of myocardial infarction and stroke. Mid-regional pro adrenomedullin (MR-proADM) is a peptide hormone generated by multiple tissues [14]. This hormone has a natriuretic, vasodilatory, and hypertensive effect. Soluble ST2 (sST2), a receptor for Interleukin 33 (IL-33), is also a prognostic marker for cardiac remodeling [15]. Several studies have shown that sST2 is associated with cardiac remodeling due to volume and pressure overload. The most common biomarker of cardiac damage is the low-density lipoprotein (LDL) [16]. Numerous studies have shown the positive correlation of LDL with increased atherosclerosis and cardiac remodeling [16]. Natriuretic peptides such as brain natriuretic peptide (BNP) and N-terminal proBNP (NT-proBNP) are also diagnostic biomarkers for cardiac dysfunction and remodeling [17]. Increased concentration of these peptides indicates higher ventricular wall stress, glomerular filtration rate, sodium and water extraction, and vasodilation. Another biomarker that also works as an indicator of cardiac and kidney damage is creatinine [18]. Copeptin is another biomarker responsible for the regulation of vascular tone and free water reabsorption [19]. Its lower level relates to cardiac remodeling with co-morbidity like diabetes mellitus.

In addition to circulating biomarkers for heart failure there are groups of biomarkers that actively participate in the cardiac remodeling process. Those molecules can be

separated into multiple groups such as C-terminal Propeptide of Procollagen Type I(PICP), Procollagen Type I N-terminal Propeptide (PINP); Procollagen Type III Amino-terminal Propeptide (PIINP), C-terminal Telopeptide of Collagen Type I (CITP), Matrix metalloproteinases (MMPs), Tissue inhibitors of metalloproteinase (TIMPs), Transforming growth factor-β (TGF-β), connective tissue growth factor (CTGF), endothelial to mesenchymal transition (EndoMT); Galectin-3 protein (Gal-3), tumor necrosis factor α (TNFα) [54]. These biomarkers are described in the later chapters in detail.

1.3 Predictive modeling for cardiac remodeling

Electrocardiogram (ECG) is the most common diagnostic tool to determine the biophysical profile of the heart. ECG is notorious for its low accuracy and sensitivity in diagnosing left ventricular hypertrophy [20]. Computational modeling and machine learning can help increase the predictive accuracy for left ventricular remodeling. Several studies have used machine learning techniques to improve the diagnostic capability to detect left ventricular remodeling. Fernando *et al*. used the C5.0 algorithm to improve the prediction of LVH [21]. Their resultant five-level binary decision tree used only six predictive variables and had an accuracy of 71.4% (95%CI, 65.5–80.2), a sensitivity of 79.6%, specificity of 53%. Another group compared different algorithms for predicting the LVH. They found the area under curves (AUC) for Logistic Regression (0.81), GLMNet (0.87), Random Forest (0.82), Gradient Boosting (0.80). Another interesting study has detected LVH using ECG signals based on machine learning techniques [22]. They used pathological attributes such as R wave, S wave, inversion of QRS complex, changes in ST-segment noticed in the ECG signal as their variables. After feature transformation, they have used the Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Ensemble of Bagged Tree, AdaBoost

classifiers and compared them with four neural network classifiers including Multilayer Perceptron (MLP), Scaled Conjugate Gradient Backpropagation Neural Network (SCG NN), Levenberg–Marquardt Neural Network (LMNN) and Resilient Backpropagation Neural network (RPROP). They have found accuracy in detecting LVH is 86.6%, 84.4%, 93.3%,75.6%, 95.6%, 97.8%, 97.8%, 88.9% accordingly.

All the above and other models are not interpretable and therefore need improved interpretability. Moreover, none of these models were targeted toward the prediction of cardiac remodeling in hypertension. In addition, many research studies adopted interpretable machine learning models for hypertension [23]. They rely on easily explainable models such as linear and decision tree models [24]. The decision tree is easier to use for two reasons: Its simple representation of the complex model and readily identifiable features from the tree's top. Decomposing neural network into decision trees is another approach widely used by clinicians [25]. Some researchers also use a variant of traditional interpretable techniques to interpret their models. One such technique is Anchors [26], which is an extension of the common interpretable technique local interpretable model-agnostic explanations (LIME). Some rule extraction techniques are also used (MofN algorithm) [27], which tries to extract rules that explain single neurons by clustering the least significant neurons. Interpretation of the black box models is also used via visualization. Also, these visualization tools deal with only specific types of data (image, text, ECG data). However, this research used various types of data (ECG, Anthropomorphic, or Biophysical Data), very little information is available for their integration with biochemical data such as inflammatory mediators and RAAS fingerprinting [28]. This type of data should be a part of an interpretable machine model for cardiac remodeling.

1.4 Cardiac Resynchronization Therapy (CRT) and its benefit

The four chambers of our heart (two upper atrium and two lower ventricles) form two atrioventricular pairs consisting of one atrium and ventricle. Oxygen-depleted blood from the different parts of the body enters the heart by the right atrium and is pumped out by the right ventricle to the pulmonary artery to the lung and mixed with the oxygen. This oxygen-rich blood enters the heart again in the left atria, then to the left ventricle. Aorta then pumps out this blood to the body. This synchronous activity among four heart chambers results in the efficient pumping of oxygen-rich blood. When ventricles do not pump out the blood efficiently, it results in a dyssynchronous rhythm in the heart. As a result, arrhythmia symptoms start to appear. This Ventricular dyssynchrony is of three categories: Atrioventricular dyssynchrony, Interventricular dyssynchrony, and Intraventricular dyssynchrony [29]. Therapy aims to bring synchronization among heart chambers to manage heart failure signs. There are several methods of treating cardiac resynchronization such as medication, surgery, device-assisted ventricular functionality, heart transplant, lifestyle modification, and lifestyle management.

The most common procedure for treating the rhythm is cardiac resynchronization therapy (CRT). CRT works are performed by implanting a small device below the collar bone [30]. This device works as a pacemaker by monitoring the heart rate to detect abnormal heart rhythms and sending a small electric pulse to correct the pumping activity of the ventricles. As a result, the lower ventricles have better pumping ability and leak a lower amount of blood in the mitral valve. The muscle of the lower ventricles pumps better as a result of this coordination. CRT provides some advantages in patients with left ventricular dysfunction and prolonged QRS duration [31]. It reduces heart failure and hospitalization, and

prolonged survival compared with medical treatment such as medication. CRT is also valuable for patients to improve the symptoms and quality of life (QOL), increase exercise tolerance, and reduce the mechanical change in the left ventricle. Even though CRT provides many benefits, CRT fails for around 10-15% of the patients [32].

Patients who undergo CRT have two choices for devices [30]. They can opt for a Cardiac Resynchronization Therapy Pacemaker (CRT-P) or a Cardiac Resynchronization Therapy Defibrillator (CRT-D). The choice of treatment depends on the heart condition. If the heart is miserably failing, then the CRT-D is the only chance for cure [33]. In addition to pacemaker activity, CRT-D also serves as a defibrillator. This functionality is significant for patients who are at the advanced stage of arrhythmia.

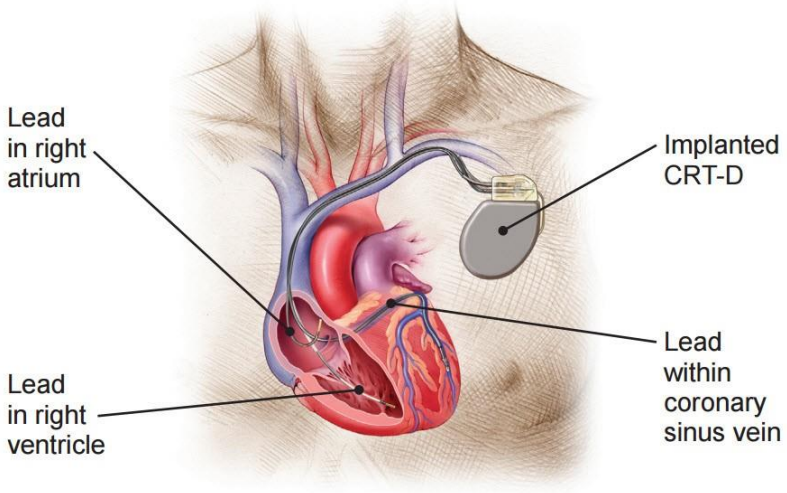


Figure 1.3. CRT-D Device. In addition to providing pacemaker activity, CRT-D provides defibrillation. It is taken from Boston scientific (https://www.bostonscientific.com/)

1.4.1 Role of biomarkers in the CRT response and cardiac remodeling

The American Heart Association and American College of Cardiology (AHA/ACC) 2008 and European Society of Cardiology (ESC) 2007 guidelines recommend some criteria such as patients with sinus rhythm, left ventricular ejection fraction ≤ 35%, QRS > 120ms, NYHA class III/IV [34]. Even after following all proper guidelines, almost one-third of the patients respond unfavorably to the treatment [35]. Biomarkers can help identify the patients who will better respond to the treatment [36]. Spinale and colleagues recently showed that specific serum protein biomarkers hold predictive power for CRT response. Expressly, elevated levels of a soluble suppressor of tumorgenicity-2 (sST-2), soluble tumor necrosis factor receptor-II (sTNFr-II), matrix metalloproteinases-2 (MMP-2), and C-reactive protein (CRP) indicated a reduced likelihood of benefit across 752 patients from the SMART-AV CRT trial. However, many recent clinical studies have tested the utility of advanced machine learning algorithms for predicting the response to CRT using various patient data, including electronic health records, clinical imaging, and others [36, 37-40]. These studies never showed the importance of using biomarkers in their models. Therefore, a biomarker scoring system is helpful in building predictive modeling for CRT response and cardiac remodeling. Thus, combining the biomarker-based matrices with various features from the SMART-AV trail clinical patient data is helpful in predicting response to CRT and cardiac remodeling.

1.4.2 Predictive Modeling for CRT Patient Selection

Machine learning can identify risks associated with a CRT implant and identify significant risks related to the treatment by finding essential variables. There are several studies have been conducted for algorithm-based CRT patient selection. The most recent

study included the use of electronic health record data to predict the outcome methods. They used gradient boosting algorithms to expect the response but ended up with a low recall value. Also, their study has lower accuracy (0.65) compared to other models to date [39]. The second study conducted by the Cleveland Clinic and John Hopkins University has shown that 690 and 253 patients have found AUC 0.72 for the Naïve Bayes Classifier. They have also seen a good result for logistic regression. They also used Adaptive Boosting but did not find satisfactory results like the first study [38]. The third study attempted to cluster the patients rather than predict the outcome of the treatment. They have used Multiple Kernel Learning for pressing the patients rather than classifying them [37]. Kalscheur et al. used the result of the COMPANION trial to predict the outcome of the treatment. They have found the best result using the Random Forest Algorithm (AUC 0.76) [40].

1.4.3 AV delay based clinical trials: SMART-AV clinical trial

The Atrioventricular (AV) node is situated at the center of the heart's electrical system and is responsible for transmitting the heart's electrical impulse from the atria to the ventricle. First, the Sinus (S) node generates an electrical signal that travels through the atria, resulting in the contraction/beat. Then these signals are collected by the AV node. After a brief delay, this signal passes through the ventricles. For normal heart function, there should be coordination between these two signals. For a healthy heart to work, we need a brief delay between these two signals. Without any coordination, ventricles may prematurely fill with blood from the atria and result in premature pumping. Disorder in the AV node results in the disruption of cardiac electrical signaling resulting in the fast and

slow pacing of the heart. These result in several symptoms due to the failure of the heart's electrical system.

The success of CRT treatment depends on the optimal atrioventricular (AV) delay. Therefore, CRT treatment requires import optimization of the AV node. Without proper optimization, CRT treatment can see 10-15% less efficiency [32]. The SMART-AV clinical trial uses three different algorithms to adjust the AV delay to synchronize the heart's electrical system. The outcome from the three different techniques did not vary significantly. The researcher hypothesized that systemic AV delay optimization with echocardiography and the SD algorithm is superior to a fixed nominal AV delay. This study showed that LV geometry improves after six months regardless of the AV delay method.

1.5 Interpretable Machine Learning

Machine learning interpretability entails the transparency and explainability of a machine learning model. Most machine learning models are not easily interpretable due to the complex internal mechanism of these models [41]. Application of machine Learning in large healthcare datasets requires understanding the variable interaction and how they interact with the labels or target. The lack of domain knowledge forces the modeler to emphasize more on accuracy than interpretability [42]. Complex black box models such as Support Vector Machines (SVM), Random Forest (RF), and Artificial Neural Networks (ANN) provide more accuracy but are not easy to interpret. Therefore, they are limited in the use of clinical decision making. General Data Protection Regulation (GDPR) by European Union (EU) is already limiting the use of Machine Learning in clinical decision-making. Model Explainability will be very important for popularizing their use in clinical interventions [43].

Table 1.1. Variable Measured in the SMART-AV Trial [31].

Domain	Individual Feature
Demographic (2)	Sex, Age
Physical Characteristics (5)	Height, Weight, Body Mass Index (BMI), Systolic blood pressure (BPSYS), Diastolic Blood Pressure (BPDIA), Pulse, Heart Rate at Rest (HRREST)
Co-Morbid Situation (6)	Paroxysmal Atrial Fibrillation (PAF), Renal Disease, Chronic Obstructive Pulmonary Disease (COPD), History of Left Bundle Branch Block (LBBB), History of Right Bundle Branch Block (RBBB), Ischemic Cardiomyopathy
Heart Failure (2)	QOL Score, 6 Minute Walk Distance
Surgical Interventions (3)	Sinoatrial (SA) Node Surgery, Coronary Artery Bypass Grafting (CABG), Pre-Cutaneous Coronary Intervention (PCI)
Medication (4)	Beta-Blocker (BB), Diuretics, Ace inhibitors or ARBs (ACE-ARB), Digoxin
Circulating Biomarkers (4)	Soluble Suppressor of Tumorgenicity-2(sST-2), Soluble Tumor Necrosis Factor Receptor-II (sTNFr-II), Matrix Metalloproteinases-2 (MMP-2), and C-Reactive Protein (CRP)
LV Assessment (7)	Left Ventricular End Diastolic Volume (LVEDV), Left Ventricular End Systolic Volume (LVESV), Left Ventricular Ejection Fraction (LVEF), Stroke Volume (SV), EDV/ESV, Cube Root
ECG (3)	AV Interval with Atrial Pacing, PR Interval with Atrial Pacing, QRS Width

1.6 Techniques used for interpreting machine learning models

Interpretable machine learning techniques are of two types: model-specific and model-agnostic. Model-specific specific interpretable techniques change the interpretability with the change of the model [44]. Model agnostic interpretability is not particular to any specific models. Therefore, these Explainability does not change with models. We can also explain the models on a global or local scale. Global model interpretability techniques show the response of individual features to the overall model outcome. Local model interpretation only deals with the small part of the model. For example, if we model the prediction of treatment, individual features have a specific weight toward the overall predictive outcome of the treatment. The importance of a particular variable varies for each of the patients. In the second case, the local interpretation is the way to explain the model. Local model interpretation is crucial for the precision medicine perspective. Table 4 represents some of the examples of interpretable techniques.

Table 1.2. Approaches of the interpretability models with examples [45].

	Global	**Local**
Model Specific	• Decision trees • Regression models • Naive Bayes classifier	• Set of rules (for a specific individual) • Decision trees (by tree -decomposition) • Most visual analytics-based • approaches • k-nearest neighbors

Model Agnostic	• Different variants of model compression/knowledge distillation/global surrogate models • Partial Dependence Plots (PDP), • Individual Conditional Expectation (ICE) plots • Black Box Explanations through Transparent approximations (BETA) (Lakkaraju) • Model Understanding through Subspace Explanations (MUSE) [46]	• Local interpretable model agnostic explanations (LIME) • Shapley additive explanations (SHAP) • Anchors • Attention map visualization, • Model Understanding through Subspace Explanations (MUSE)

All interpretable techniques are not great for their application in the field of clinical data science. The following table shows the advantage and disadvantages of using them in data science.

Table 1.3. Advantages and disadvantages of some common Explainability techniques [23]

Technique	**Advantage**	**Disadvantage**
Feature Importance	A highly compressed interpretation that considers the interaction between features	Unclear of usage in training or testing dataset

Partial Dependence Plot (PDP)	Clear interpretation	Assumption of independence between features
Individual Conditional Expectation (ICE)	Clear interpretation	Plots are overcrowded to understand
Feature Interaction	Detect all interactions been features	Computationally expensive
Global Surrogate Model	Easy to measure the goodness of the surrogate model using R^2	No clear-cut cutoff for R^2 makes trust issue of the surrogate model
Local Surrogate Model	A comprehensive explanation for different data types	Instable and close points have completely different explanations
Shapely Value Explanation	Based on Game Theory Theorem	Computationally very expensive

1.7 Logic-Based ODE Models in System Biology

Signaling pathways allow the cells to sense the sudden change in their surrounding environment. Cells respond by changing the transcriptional activity, metabolism, and other cellular activity inside the cell [47]. Signaling pathways are the interaction of several linear biochemical reactions. The combination of all these linear reactions makes signaling networks very complicated. Researchers have developed various cellular interaction network models. The two kinds of cellular interaction networks are abstract and dynamic models. Abstract or conceptual models only capture some static part of the cellular interaction [48].

Dynamic models discover the dynamics of the interaction of different reactions and pathways in the cell. Modeling biochemical interaction of the cell helps us to integrate the hypothetical and experimental results into a complete system. These systems are helpful to understand, support, or falsify the underlying mechanism of the cellular interaction network and understand the biological interaction at the system level.

Most of the cell signaling networks use standard cellular components. The most common element in the cellular signaling network is the receptor, which receives the extracellular stimuli from the outside environment. Some of these receptors are common across different eukaryotic cellular pathways. Therefore, the same receptor can be seen in various cellular networks. These interactions in different pathways inside of the cell finally end in producing or degrading specific proteins. The enzymes that catalyze these interactions are also produced by the intermediate proteins products in the cellular networks. For cardiac remodeling, most of the interacting proteins are either produced or degraded by specific gene-protein interaction networks. These receptors can switch from their active state to inactive states resulting is a biochemical reaction cascade inside the cell. The field of molecular biology consists of a vast amount of interaction data consisting of genomics, proteomics, metabolic, and immunological data. The main challenge of working with these data is finding interactions in these discrete data types [49]. Knowing the exact nature of these interactions is very important for the point of translation biology. Suppose anyone wants to design a drug targeting a specific component of the cell signaling network. While the drug might work on a particular target protein, they must know how this target protein interacts with all the parts of the cell signaling network. This knowledge of detailed interaction networks reduces off-target drug interaction. Therefore, abstract and dynamic models are both essential for capturing valuable information in the cell

signaling network. Structural or static networks correlate several pathways into an extensive network. Even though these networks are vast, they lack the directionality. Some of these static network only provides snapshot of the system at a specific point. An example of such a network is the gene co-expression network which only captures the gene expression pattern under a particular condition. They are not good at capturing the dynamic nature of any biological system [50].

Ordinary Differential Equations (ODE) are very good at capturing the dynamical characteristics of the biological system. Cellular interactions in such systems are a collection of chemical reactions and follow chemical rate laws. These laws represent better by the ODE. The only problem with the ODE model is that they have a vast parameter space. This parameter space is problematic for a large-scale network like biological systems. Therefore, a combination of such models is the logic-based ODE models. We do not need vast knowledge about the parameters for the logic-based ODE models. The only information we need here is the directionality. Therefore they are relatively easy to construct while they can easily catch the system wide response to perturbation [51].

The Hill equation is one of the most commonly used mathematical methods for studying enzyme reaction kinetics [52]. In biochemical interaction analysis, its primary use is for cooperative binding. The Hill equation describes the fraction of macromolecules saturated by ligand as a function of ligand concentration mathematically presented as-

$$\theta = \frac{[l]^n}{k_d + [l]^n} = \frac{[l]^n}{(k_a)^n + [l]^n} = \frac{1}{\left(\frac{k_a}{[l]}\right)^n + 1}$$

In this equation, θ is the fraction of occupied sites where the ligand can bind to the active site of the receptor protein. *[l]* is the free(unbound) ligand concentration, k_d is the apparent dissociation constant derived from the law of mass action, k_a is the ligand concentration producing half occupation, and n is the Hill Coefficient. When n>1, the cooperative binding is positive. If n<1, the cooperative binding is negative, and n=1 indicates no cooperative binding. For any biochemical species $x \in [0, 1]$, activation of *x*, *F(x)* modeled as normalized Hill function of the following form:

$F(x) = \frac{Bx^n}{k^n + x^n}$ with $B = \frac{EC_{50}^n - 1}{2EC_{50}^n - 1}$ and $k = (B-1)^{1/n}$

In this equation, *n* is the Hill Coefficient related to the curve's steepness, *and* EC_{50} is the enzyme concentration at which half-maximal activation occurs. Standard logic gate functions capture reaction interactions:

X AND Y = F(X)F(Y),

X OR Y = F(X) + F(Y) – F(X)F(Y),

X AND NOT Y = F(X)(1-F(Y))

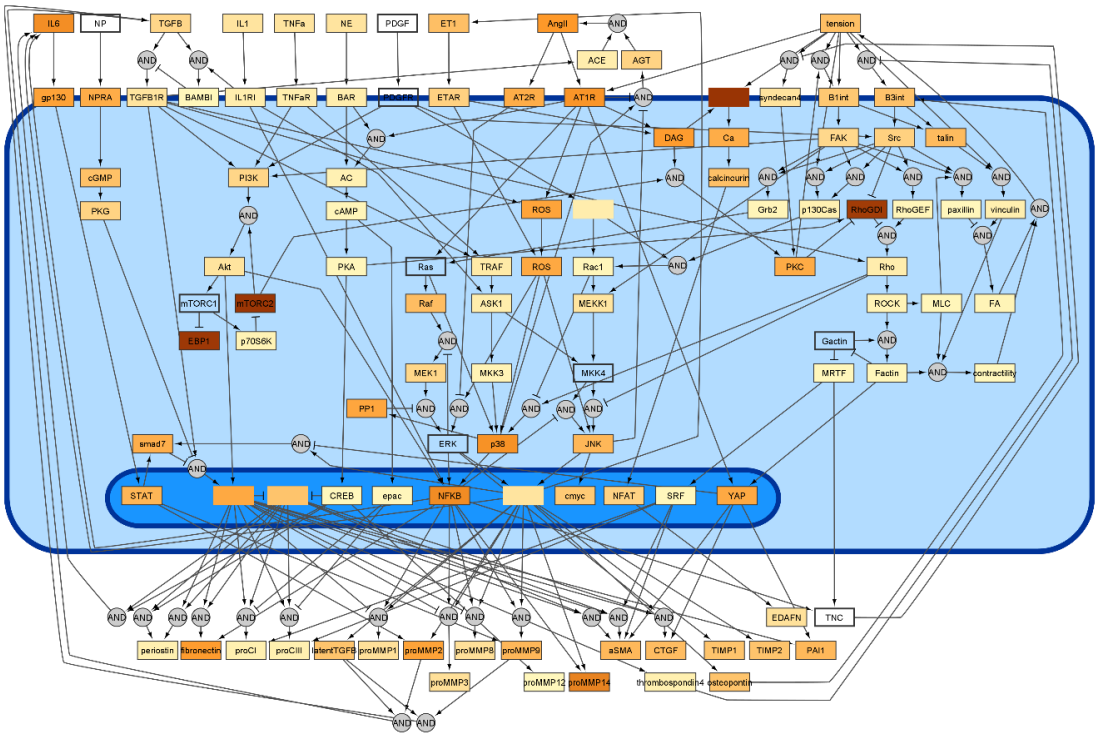


Figure 1.4. Fibroblast mechanotransduction network constructed using the logic-based ODE models.

Our lab has expertise in building cardiac cell signaling networks by manually curating biochemical transduction pathways from existing literature searches and their experimental validation (Figure 1.4). A study by Rogers et al. [53] has expanded a fibroblast mechano-chemo signal transduction capable of accurately predicting fibrosis-related protein expression in response to biochemical and biophysical force. They have shown the biochemical dose-response behavior under varying levels of mechanical stimulation. Their comprehensive model simulations of fibroblast responses to drugs in low- or high-tension contexts and identified several drug combinations that adapted fibrotic activity to the local mechanical state.

In this dissertation I will conduct the following research: In Aim 1, we will build an interpretable ML model for CRT response using biomarker and biomechanical features. We will build interpretable ML models using blood derived biochemical markers plus echocardiography-derived biomechanical features. These features will help us identify the patients who respond positively to CRT treatment before implanting the CRT device. Also, our model will show patient-specific features and how these features contribute to CRT. In aim 2, we will infer a gene regulatory network using the novel inference algorithm and filter them in the context of fibrosis to build a dilated cardiomyopathy specific network. We will use RNAseq data from Myocardial Applied Genomic Network (MGNet) study for this. In aim 3, we will developed a logic-based model for the mechanistic network of cardiac fibrosis integrating the gene regulatory network derived from aim 2 into a previously constructed cardiac fibrosis signaling network model. This integrated model implemented biochemical and biomechanical reactions as ordinary differential equations based on normalized Hill functions. The model elucidated the semi-quantitative behavior of cardiac fibrosis signaling complexity by capturing multi-pathway crosstalk and feedback loops. Perturbation analysis predicted the most important nodes in the mechanistic model and patient-specific simulations helped identify which molecules are most highly correlated with clinical measures of patient cardiac function.

1.8 References

1. Mensah, G. A., Croft, J. B., & Giles, W. H. (2002). The heart, kidney, and brain as target organs in hypertension. Cardiology Clinics, 20(2), 225-247. doi:10.1016/s0733-8651(02)00004-8

2. World Health Organization. (2020). Hypertension fact sheet. Retrieved from https://www.who.int/news-room/fact-sheets/detail/hypertension.

3. CDC. (2021). CDC facts about hypertension. Retrieved from https://www.cdc.gov/bloodpressure/facts.htm.

4. Young, J. H., Chang, Y. C., Kim, J. D., Chretien, J., Klag, M. J., Levine, M. A., . . . Chakravarti, A. (2005). Differential susceptibility to hypertension is due to selection during the out-of-Africa expansion. PLoS Genetics, 1(6) doi:10.1371/journal.pgen.0010082

5. Fuchs, F. D. (2011). Why do black americans have higher prevalence of hypertension? Hypertension, 57(3), 379-380. doi:10.1161/HYPERTENSIONAHA.110.163196

6. McNaughton, C. D., Brown, N. J., Rothman, R. L., Liu, D., Kabagambe, E. K., Levy, P. D., . . . Roumie, C. L. (2017). Systolic blood pressure and biochemical assessment of adherence. Hypertension, 70(2), 307-314. doi:10.1161/HYPERTENSIONAHA.117.09659

7. Azevedo, P. S., Polegato, B. F., Minicucci, M. F., Paiva, S. A. R., & Zornoff, L. A. M. (2016). Cardiac remodeling: Concepts, clinical impact, pathophysiological mechanisms and pharmacologic treatment. Arquivos Brasileiros De Cardiologia, 106(1), 62-69. doi:10.5935/abc.20160005

8. Nadruz, W. (2015). Myocardial remodeling in hypertension. Journal of Human Hypertension, 29(1), 1-6. doi:10.1038/jhh.2014.36

9. González, A., Ravassa, S., López, B., Moreno, M. U., Beaumont, J., San José, G., . . . Díez, J. (2018). Myocardial remodeling in hypertension. Hypertension (Dallas, Tex.: 1979), 72(3), 549-558. doi:10.1161/HYPERTENSIONAHA.118.11125

10. Fulghum, K., & Hill, B. G. (2018). Metabolic mechanisms of exercise-induced cardiac remodeling. Frontiers in Cardiovascular Medicine, 5, 127. doi:10.3389/fcvm.2018.00127

11. Aimo, A., Gaggin, H. K., Barison, A., Emdin, M., & Januzzi, J. L. (2019). Imaging, biomarker, and clinical predictors of cardiac remodeling in Heart Failure with Reduced Ejection Fraction. JACC. Heart Failure, 7(9), 782-794. doi:10.1016/j.jchf.2019.06.004

12. Ohtsuki, I., & Morimoto, S. (2013). Troponin. In W. J. Lennarz, & M. D. Lane (Eds.), Encyclopedia of biological chemistry (second edition) (pp. 445-449). Waltham: Academic Press. Retrieved from https://www.sciencedirect.com/science/article/pii/B978012378630200195X

13. Swiatkiewicz, I., Kozinski, M., Magielski, P., Fabiszak, T., Sukiennik, A., Navarese, E. P., . . . Kubica, J. (2012). Value of C-reactive protein in predicting left ventricular remodelling in patients with a first ST-segment elevation myocardial infarction. Mediators of Inflammation, 2012 doi:10.1155/2012/250867

14. Morbach, C., Marx, A., Kaspar, M., Güder, G., Brenner, S., Feldmann, C., . . . Angermann, C. E. (2017). Prognostic potential of midregional pro-adrenomedullin following

decompensation for systolic heart failure: Comparison with cardiac natriuretic peptides. European Journal of Heart Failure, 19(9), 1166-1175. doi:10.1002/ejhf.859

15. Zhang, T., Xu, C., Zhao, R., & Cao, Z. (2021). Diagnostic value of sST2 in cardiovascular diseases: A systematic review and meta-analysis. Frontiers in Cardiovascular Medicine, 8 doi:10.3389/fcvm.2021.697837

16. Bostan, M., Stătescu, C., Anghel, L., Șerban, I., Cojocaru, E., & Sascău, R. (2020). Post-myocardial infarction ventricular remodeling Biomarkers—The key link between pathophysiology and clinic. Biomolecules, 10(11) doi:10.3390/biom10111587

17. Katsi, V., Kallistratos, M. S., Kontoangelos, K., Sakkas, P., Souliotis, K., Tsioufis, C., . . . Tousoulis, D. (2017). Arterial hypertension and health-related quality of life. Frontiers in Psychiatry, 0 doi:10.3389/fpsyt.2017.00270

18. Bansal, N., Zelnick, L., Go, A., Anderson, A., Christenson, R., Deo, R., . . . Rao, P. S. (2019). Cardiac biomarkers and risk of incident heart failure in chronic kidney disease: The CRIC (chronic renal insufficiency cohort) study. Journal of the American Heart Association, 8(21), e012336. doi:10.1161/JAHA.119.012336

19. Berezin, A. E., & Berezin, A. A. (2020). Adverse cardiac remodelling after acute myocardial infarction: Old and new biomarkers. Disease Markers, 2020, 1-21. doi:10.1155/2020/1215802

20. Siontis, K. C., Noseworthy, P. A., Attia, Z. I., & Friedman, P. A. (2021). Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nature Reviews Cardiology, 18(7), 465-478. doi:10.1038/s41569-020-00503-2

21. Garza-Salazar, F., Romero-ibarguengoitia, M., Rodriguez-Diaz, E., Azpiri-López, J., & Gonzalez, A. (2020). Improvement of electrocardiographic diagnostic accuracy of left ventricular hypertrophy using a machine learning approach. Plos One, 15, e0232657. doi:10.1371/journal.pone.0232657

22. MANIKANDAN, R., Jothiramalingam, R., Jude, A., Patan, R., D, J., & Gandomi, A. (2021). Machine learning-based left ventricular hypertrophy detection using multi-lead ECG signal. Neural Computing and Applications, doi:10.1007/s00521-020-05238-2

23. Elshawi, R., Al-Mallah, M. H., & Sakr, S. (2019). On the interpretability of machine learning-based model for predicting hypertension. BMC Medical Informatics and Decision Making, 19(1), 146. doi:10.1186/s12911-019-0874-0

24. Weng, W. (2020). Machine learning for clinical predictive analytics. (pp. 199-217)

25. Cohen, S., Rokach, L., & Maimon, O. (2007). Decision-tree instance-space decomposition with grouped gain-ratio. Information Sciences, 177, 3592-3612. doi:10.1016/j.ins.2007.01.016

26. Ribeiro, M. T., Singh, S., & Guestrin, C. (2018). Anchors: High-precision model-agnostic explanations. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1) Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/11491

27. Setiono, R. (2000). Extracting M-of-N rules from trained neural networks. IEEE Trans. Neural Networks Learn. Syst., doi:10.1109/72.839020

28. Pacurari, M., Kafoury, R., Tchounwou, P. B., & Ndebele, K. (2014). The renin-angiotensin-aldosterone system in vascular inflammation and remodeling. International Journal of Inflammation, 2014 doi:10.1155/2014/689360

29. Bleeker, G. B., Bax, J. J., Steendijk, P., Schalij, M. J., & van der Wall, Ernst E. (2006). Left ventricular dyssynchrony in patients with heart failure: Pathophysiology, diagnosis and treatment. Nature Clinical Practice. Cardiovascular Medicine, 3(4), 213-219. doi:10.1038/ncpcardio0505

30. Gassis, S., & León, A. R. (2005). Cardiac resynchronization therapy: Strategies for device programming, troubleshooting and follow-up. Journal of Interventional Cardiac Electrophysiology: An International Journal of Arrhythmias and Pacing, 13(3), 209-222. doi:10.1007/s10840-005-3247-9

31. Ellenbogen, K. A., Gold, M. R., Meyer, T. E., Fernndez Lozano, I., Mittal, S., Waggoner, A. D., . . . Stein, K. M. (2010). Primary results from the SmartDelay determined AV optimization: A comparison to other AV delay methods used in cardiac resynchronization therapy (SMART-AV) trial. Circulation, 122(25), 2660-2668. doi:10.1161/CIRCULATIONAHA.110.992552

32. Fatemi, M., Etienne, Y., Castellant, P., & Blanc, J. (2008). Primary failure of cardiac resynchronization therapy: What are the causes and is it worth considering a second attempt? A single-centre experience. Europace: European Pacing, Arrhythmias, and Cardiac Electrophysiology: Journal of the Working Groups on Cardiac Pacing, Arrhythmias, and Cardiac Cellular Electrophysiology of the European Society of Cardiology, 10(11), 1308-1312. doi:10.1093/europace/eun245

33. Shakibfar, S., Yazdchi, M., & Aliakbaryhosseinabadi, S. (2019). Effectiveness of CRT-D versus ICD on prevention of electrical storm: Big data from the USA. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2019, 302-304. doi:10.1109/EMBC.2019.8857530

34. Hawkins, N. M., Petrie, M. C., MacDonald, M. R., Hogg, K. J., & McMurray, J. J. V. (2006). Selecting patients for cardiac resynchronization therapy: Electrical or mechanical dyssynchrony? European Heart Journal, 27(11), 1270-1281. doi:10.1093/eurheartj/ehi826

35. Daubert, C., Behar, N., Martins, R. P., Mabo, P., & Leclercq, C. (2017). Avoiding non-responders to cardiac resynchronization therapy: A practical guide. European Heart Journal, 38(19), 1463-1472. doi:10.1093/eurheartj/ehw270

36. Spinale, F. G., Meyer, T. E., Stolen, C. M., Van Eyk, J. E., Gold, M. R., Mittal, S., . . . Ellenbogen, K. A. (2019). Development of a biomarker panel to predict cardiac resynchronization therapy response: Results from the SMART-AV trial. Heart Rhythm, 16(5), 743-753. doi:10.1016/j.hrthm.2018.11.026

37. Cikes, M., Sanchez-Martinez, S., Claggett, B., Duchateau, N., Piella, G., Butakoff, C., . . . Bijnens, B. (2019). Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. European Journal of Heart Failure, 21(1), 74-85. doi:10.1002/ejhf.1333

38. Feeny, A. K., Rickard, J., Patel, D., Toro, S., Trulock, K. M., Park, C. J., . . . Chung, M. K. (2019). Machine learning prediction of response to cardiac resynchronization therapy:

Improvement versus current guidelines. Circulation. Arrhythmia and Electrophysiology, 12(7), e007316. doi:10.1161/CIRCEP.119.007316

39. Hu, S., Santus, E., Forsyth, A. W., Malhotra, D., Haimson, J., Chatterjee, N. A., . . . Lindvall, C. (2019). Can machine learning improve patient selection for cardiac resynchronization therapy? Plos One, 14(10), e0222397. doi:10.1371/journal.pone.0222397

40. Kalscheur, M. M., Kipp, R. T., Tattersall, M. C., Mei, C., Buhr, K. A., DeMets, D. L., . . . Page, C. D. (2018). Machine learning algorithm predicts cardiac resynchronization therapy outcomes: Lessons from the COMPANION trial. Circulation. Arrhythmia and Electrophysiology, 11(1), e005499. doi:10.1161/CIRCEP.117.005499

41. Che, Z., Purushotham, S., Khemani, R., & Liu, Y. (2016). Interpretable deep models for ICU outcome prediction. AMIA ... Annual Symposium Proceedings, 2016, 371-380. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28269832

42. Petch, J., Di, S., & Nelson, W. (2021). Opening the black box: The promise and limitations of explainable machine learning in cardiology. Canadian Journal of Cardiology, doi:10.1016/j.cjca.2021.09.004

43. Mourby, M., Ó Cathaoir, K., & Collin, C. B. (2021). Transparency of machine-learning in healthcare: The GDPR & european health law. Computer Law & Security Review, 43, 105611. doi:10.1016/j.clsr.2021.105611

44. Ozaydin, B., Berner, E. S., & Cimino, J. J. (2021). Appropriate use of machine learning in healthcare. Intelligence-Based Medicine, 5, 100041. doi:10.1016/j.ibmed.2021.100041

45. Štiglic, G., Kocbek, P., Fijačko, N., Zitnik, M., Verbert, K., & Cilar, L. (2020). Interpretability of machine learning-based prediction models in healthcare. Wiley Interdiscip. Rev. Data Min. Knowl. Discov., doi:10.1002/widm.1379

46. Lakkaraju, H., Kamar, E., Caruana, R., & Leskovec, J. (January 27, 2019). (January 27, 2019). Faithful and customizable explanations of black box models. Paper presented at the 131-138. doi:10.1145/3306618.3314229

47. Wynn, M. L., Consul, N., Merajver, S. D., & Schnell, S. (2012). Logic-based models in systems biology: A predictive and parameter-free network analysis method. Integrative Biology : Quantitative Biosciences from Nano to Macro, 4(11) doi:10.1039/c2ib20193c

48. Bhan, A., & Mjolsness, E. (2006). Static and dynamic models of biological networks: Research articles. Complexity, 11, 57-63. doi:10.1002/cplx.20140

49. Urbanski, A. H., Araujo, J. D., Creighton, R., & Nakaya, H. I. (2019). Integrative biology approaches applied to human diseases. Exon Publications, , 19-36. doi:10.15586/computationalbiology.2019.ch2

50. Zhang, L., Feng, X. K., Ng, Y. K., & Li, S. C. (2016). Reconstructing directed gene regulatory network by only gene expression data. BMC Genomics, 17 Suppl 4(1), 430. doi:10.1186/s12864-016-2791-2

51. Zeigler, A. C., Richardson, W. J., Holmes, J. W., & Saucerman, J. J. (2016). A computational model of cardiac fibroblast signaling predicts context-dependent drivers of myofibroblast differentiation. Journal of Molecular and Cellular Cardiology, 94, 72-81. doi:10.1016/j.yjmcc.2016.03.008

52. Chow, C. C., Ong, K. M., Dougherty, E. J., & Simons, S. S. (2011). Inferring mechanisms from dose-response curves. Methods in Enzymology, 487, 465-483. doi:10.1016/B978-0-12-381270-4.00016-0

53. Rogers JD, Richardson WJ. Fibroblast mechanotransduction network predicts targets for mechano-adaptive infarct therapies. Elife. 2022;11:e62856. Published 2022 Feb 9. doi:10.7554/eLife.62856

54. Ding Y, Wang Y, Zhang W, et al. Roles of Biomarkers in Myocardial Fibrosis. Aging Dis. 2020;11(5):1157-1174. Published 2020 Oct 1. doi:10.14336/AD.2020.0604

Chapter 2

Interpretable machine learning predicts cardiac resynchronization therapy responses from personalized biochemical and biomechanical features

2.1 Introduction

Cardiac resynchronization therapy (CRT) is the preferred treatment method for patients with ventricular desynchrony accompanied by reduced ejection fraction and bundle branch block [1]. CRT reduces the risk of sudden heart failure due to the weakening of the heart muscle and canhelp alleviate the symptoms of heart failure and improve the quality of life [2]. The American Heart Association and American College of Cardiology (AHA/ACC) 2008 and European Society of Cardiology (ESC) 2007 guidelines recommend the following criteria for selecting patients for CRT: patients with sinus rhythm, left ventricular ejection fraction ≤ 35%, QRS > 120ms, NYHA class III/IV [3]. Unfortunately, roughly one-third of CRT recipients do not respond favorably to the treatment [4]. Given its expense and surgical risks, the ability to accurately predict individual patientbenefits from this treatment could hold great clinical value [5].

Spinale and colleagues recently showed that specific serum protein biomarkers hold predictive power for CRT response [6]. Notably, elevated levels of the soluble suppressor of tumorgenicity-2 (sST-2), soluble tumor necrosis factor receptor-II (sTNFr-II), matrix metalloproteinases-2 (MMP-2), and C-reactive protein (CRP) indicated a reduced likelihood of benefit across ~800 patients from the SMART-AV CRT trial. Other recent

studies have tested theutility of advanced machine learning algorithms for predicting the response to CRT using variouspatient data, including electronic health records, clinical imaging, and others [7–11]. These studies have shown modest predictive capabilities but have primarily utilized data types that largely ignored biochemical features such as serum biomarkers.

Most machine learning models are not easily interpretable due to their complex black-box nature for clinical decision-making [12-13]. A fair application of machine learning models in clinical datasets requires understanding the interaction of features with each other and target variables. Complex black box models such as Support Vector Machines (SVM), Random Forest (RF), and Artificial Neural Networks (ANN) though provide more accuracy but are not easy to interpret [14]. Therefore, they are limited in the use of clinical decision making. General Data Protection Regulation (GDPR) by European Union (EU) is already restricting machine learning in clinical decision-making [15]. Model interpretability is very important in clinical data science and should be part of clinical decision-making. Interpretable machine learning techniques are of two types: model-specific and model-agnostic [16]. Model-specific specific interpretable methods change the interpretability with the change of the model. Model agnostic interpretability is not particular to any specific models [17]. Therefore, these explainabilities do not change with a new model. We can also interpret machine learning models on a global or local scale. Global model interpretability techniques show the response of individual features to the overall model outcome. Local model interpretation only deals with a small part of the model. We have used model agnostic Global method SHapley Additive exPlanations (SHAP) [19] and Partial Dependence Plots (PDP) [20] for this study. We also used Local Interpretable Model agnostic Explanations (LIME) methods [21] for exploring the model locally.

In this study, we sought to computationally predict and interpret the patient response to CRT using a combination of demographics, physical characteristics, co-morbidities, medication history, circulating biomarker levels, and echo-based LV assessment. Building upon the previous work ofSpinale et al., we combined their biomarker-based metric with various features from the SMART-AV clinical patient data [23]. We assessed the performance of our resulting ensemble machine learning classification model using receiver-operating curve analysis for a hold-out patient dataset and comparisons of 6-month cardiac measures between model-predicted responder and non-responder groups. We also performed SHapley Additive exPlanations (SHAP) analysis to help interpret the global importance of all features included in the model. We also demonstrated how the top three features affect the overall model outcome. Finally, we have shown the top 10 features for two CRT recipients and how these ten features determine their response to CRT.

2.2 Methods

2.2.1 Study Population and Data Preparation

The data source for our model training and testing was the SMART-AV trial published previously [22]. In that study, 794 patients with NYHA class II and IV, LVEF ≤ 35%, and QRS duration ≥ 120 milliseconds were randomly assigned to different defibrillation protocols and evaluated at 0, 3, and 6 months with echocardiography and serum biomarker panels. The complete list of recorded features is organized in Table 1. A positive CRT response was defined as a decrease in ESV of at least 15 mL between 0 to 6 months post-surgery [23], and the patient cohort held a nearly equal split of responders (n = 396) and non-responders (n=398).

Patients with missing data were imputed using two different methods for our study. Surgical intervention features, PCI and CABG, were imputed to match the most frequent value for each of those features. Categorical data were transformed using one-hot encoding. Non-categorical data/continuous data were mean imputed, followed by the scaling using the StandardScaler methods [23]. The patients were split into training and testing datasets, with 80%in the training dataset (n=635) and 20% in the testing dataset (n=159).

In addition to feature imputation and scaling, continuous variables were compared between CRT responders vs. non-responders using the t-test. The mean and the standard deviation are reported with respective p-values. Categorical variables were compared using the chi-square test. The result from the statistical analysis is presented in Table 2.

2.2.2 Machine Learning Model Development

Using Python 3.6.4 and scikit-learn 0.23.2, we tested a wide variety of supervised classification machine learning algorithms, including K-Nearest Neighbors, Support Vector Classifier,Decision Tree Classifier, Random Forest, Adaptive Boosting, Gradient Boosted Classifier, Gaussian Naive Bayes classifier, Linear Discriminant Analysis, XGBoost, Catboost, Logistic regression, and Multi-Layer Perceptron Neural Network [24]. In addition, we tested Stacked and Voting ensembles with the training dataset [25-27]. Each model was tuned using a cross-validated grid search across hyperparameters with parameters selected to maximize the area under the receiver-operating characteristic curve (AUC) for binary classification of patients in thetraining set. Notably, the algorithm only used 0-month (pre-surgery) feature data to predict the 6-month post-surgery response vs. non-response outcome. Model performance was evaluated using 5-fold cross-validation, and the final model was selected based on the highest mean AUC. We have only presented the

performance of the top 6 models in Table 3. We have also demonstrated the overall model-predicted response rate for CRT responders vs. non-responders and stratified them based on the model-predicted responses probability score. We have also shown how the changes in secondary variables are related to the model-predicted response.

Feature selection was performed using a backward stepwise methodology, eliminating features that did not improve the model training score. A guiding hypothesis for this work wasthat combining the previously identified serum biomarkers with demographic and echo-based features would improve predictive capability. To evaluate this hypothesis, we trained and tested our ensembled model using three different sets of features, including all features listed in Table 1, and (1) no biomarker values, (2) all 12 biomarker values, or (3) a biomarker score based on previousanalysis by Spinale et al. [6]. The biomarker score is calculated for each patient by counting how many of the four critical biomarker analytes exceed a risk threshold (MMP-2, sST-2, CRP, sTNFR-II). All these results are presented in Table 4.

2.2.3 Model Interpretation

We have used several interpretable techniques to describe our model. To help interpret global feature importance, we performed a SHapley Additive exPlanations (SHAP) analysis using the Python tool SHAP 0.37.0 KernelExplainer and KernelSHAP [19]. We have also shown the partial dependence plot [20] to present the effect of the top three features on the overall model. Finally, we picked two examples of CRT recipients to show how the model behaves locally for responders and non-responders. To show that local behavior, we have used the LIME tool to create the LIME graph [21].

2.3 Results

2.3.1 Model Predictive Performance

Across all the algorithms tested, a majority-voting ensemble classification model demonstrated the best performance. The ensemble consisted of nine equally weighted models, each voting with their respective probability of surgical success: a Linear Discriminant Analysis classifier, a Catboost Classifier, a Gradient Boosted classifier, a Random Forest classifier, an XGBoost classifier, a Support Vector Classifier, a 3-layer Multilevel Perceptron Neural Network, a Logistic Regression Classifier, and an Adaboost classifier. Without using biomarker data, our algorithm approach demonstrated modest predictive performance with an AUC of 0.63 in the training patient set (Table 3). The addition of biomarker data substantially improved model performance, with an AUC reaching 0.75 in the training patient set and almost 0.784 in the test patient set using either all 12 biomarkers or the simplified biomarker composite score (Table 2.1, Figure 1A). Using the biomarker score reached the highest AUC in both the training and test patient set, so we proceeded with the biomarker score model for the remaining analyses.

Our binary classification model correctly predicted 71% of patient responses in the test set, with 61/88 classified responders and 52/71 classified non-responders matching the trial result (Figure 2.1B). To analyze more detailed patient stratifications, we separated patients into five groups according to the model-predicted probability of response (i.e., p = 1-0.8, 0.8-0.6, 0.6-0.4, 0.4-0.2, or 0.2-0). Across the stratified patients, the model correctly identified 96% of patients in the highest and lowest response groups, with 14/15 patient responders in the high probability score group and 8/8 non-responders in the low probability score group (Figure 2.1B).

In addition to response rate (which is judged by a strict over/under -15mL threshold for ESV change over six months), we also explored quantitative changes in left ventricle remodeling metrics across the model classification groups (Figure 2.2). Over six months after the procedure, patients classified by the model as responders showed significant reductions in both ESV and EDV, while patients classified as non-responders showed no change in ESV and a slight increasein EDV over six months. Both responders and non-responders showed increased stroke volumes and ejection fractions, but the model-predicted responders showed a statistically more significant improvement in ejection fraction (~40% compared to ~20%). These discrepancies between groups were amplified further across the 5-group patient stratification using the model probability score (Figure 2.2B). In the most extreme case, the high response probability group exhibited almost a 75% improvement in ejection fraction, while the low response probability groupexhibited no change in ejection fraction over the 6 months after surgery.

2.3.2 Model Interpretability

To improve the interpretability of our ensemble classification algorithm, we performed a SHAP analysis and corresponding visualization of feature importance (Figure 2.3A). Briefly, this technique calculates a collective, global average of how much each feature value contributed to each patient's classification in order to indicate both magnitude and direction that each feature contributes to the overall probability of falling on either side of the binary classifier (i.e., respondersvs. non-responders). SHAP analysis indicated that, in general, lower 1D stretch, lower biomarkerscore, absence of ischemic cardiomyopathy, lower QOL score, and higher age were strong contributors within the algorithm for identifying responders. In Figure 2.3B, we have shown the partial dependence between

CRT response and the top three features from our fitted Majority Voting model: 1D stretch, Biomarker Score, and Ischemic Cardiomyopathy. From the topmost plot, we could see that the chance of responding to CRT significantly decreases with increasing 1D Stretch between 1.03 to 1.39. The biomarker score follows the same trend. Increased biomarker score results in lower response to CRT. Finally, the history of Ischemic Cardiomyopathy also decreases the chance of CRT response.

We also implemented our model locally using the LIME method. We have presented two local interpretations of our model. These interpretations demonstrated how patient-specific biomarkers vary for response vs. non-response to CRT. Figure 2.4A showed a patient with a higher probability of responding to CRT (0.81) and the top ten features contributing to that. For this patient, 1D Stretch of less than or equal to 1.08 and no history of RBBB, Atrial Flutter, Ischemic Cardiomyopathy, AT-PSVT, PAF, and SA are helping to move the patient to the response regimen. But No history of VT-SVT, Non-stained VT, and Biomarker Score of more than 2 contribute to non-responsiveness for this patient. Figure 2.4B showed a patient with a higher probability of non-responding to CRT (0.75) and the top ten features contributing to that. For this patient, a 1D Stretch of greater than 1.14, a history of Ischemic Cardiomyopathy, and no history of variables VT-SVT, Afib, and Nonsustained VT are helping to move the patient to the non-response group. But No history of RBBB, Atrial Flutter, SA surgery, PAF, and Biomarker Score of zero is responsible for a small probability of response in this patient.

2.4 Discussion

Machine learning has a trust issue in clinical decision-making due to the black-box nature of the machine learning algorithms [28]. Most sophisticated black-box algorithms

provide a higher accuracy but are impossible to interpret in terms of patient variables. On the other hand, the Glassbox machine learning model, such as linear regression, though very easy to interpret but comes with a risk of low model accuracy. We need to balance accuracy and interpretability to adopt machine learning for therapeutic decision-making like CRT [29]. Our majority voting algorithm has higher accuracy and covers a higher area under the curve (Figure 2.1, Table 2.3/2.4). Also, in this study, we used a biomarker scoring system to improve the predictive capability of our model. The only study that used biomarkers to predict CRT response was the COVERT-HF study [30]. Though they used a wide variety of biomarkers, our study showed that a biomarker scoring consisting of four biomarkers: MMP-2, sST-2, CRP, and sTNFR-II, is enough to build an interpretable model with high accuracy and AUC than using twelve biomarkers (Table 2.4).

While CRT offers significant clinical benefits for many heart failure patients, a large proportion of the population does not respond positively to treatment [4]. This high patient-to-patientvariability presents a need for predictive methods to help identify which patients will or will not benefit from CRT based on information obtained prior to the procedure. Several studies havedeveloped computational algorithms spanning black-box machine learning and sophisticated electromechanical finite element modeling to help predict CRT response based on varying information, including electronic health records, clinical imaging, demographic data, and more [7-11]. Some of the studies have had moderate predictive success, but most studies have generally focused their analyses on only one type of data source, and very little attention has been given to the predictive capability of circulating biomarker panels. We hypothesized thatintegrating multiple data sources and including biochemical levels from serum panels would significantly improve the predictive ability of machine learning algorithms. Using previously obtained patient data in the SMART-AV

trial, we built a novel algorithm that integrates demographic data, physical characteristics, medical history, circulating biomarkerlevels, and echocardiography data to improve the prediction of CRT response prior to surgical intervention. In a previous study, Spinale and colleagues showed significant predictive power foridentifying CRT response using pre-surgical levels of specific serum protein biomarkers (sST-2, sTNFr-II, MMP-2, and CRP)[6]. Given the crucial roles of inflammation and extracellular matrixturnover for regulating cardiac remodeling related to CRT, it should be no surprise that circulatingproteins are associated with CRT response either as upstream regulators or downstreamcorrelates. We combined the Spinale et al. patient biomarker score with 40 other input features spanning echo-based LV metrics, medical history, demographic information, and basic clinical assessments. Using these features enabled our ensemble machine learning classifier to correctlyidentify 71% of patient response outcomes, achieving an AUC of 0.784 – a substantial improvement over the previous study using the biomarker score alone.

A significant limitation of many machine learning approaches is their ‘black-box’ nature of predictions, or in other words, their un-explainability. Future adoption of artificial intelligence into the clinical decision-making process will undoubtedly be affected by an ability to explain (to somedegree at least) why models predict what they predict and to identify the driving variables within the algorithms, especially for high-risk and costly decisions like CRT treatment. In order to improve such interpretability, a growing emphasis is being put on ‘glass-box’ or ‘white-box’ techniques . We employed SHAP analysis to elucidate the relative contribution of each feature to the patient response probability output of our model (Figure 2.3). This analysis revealed that important featurescame from diverse data sources, with the top five features including echo-based data (1D stretch),serum protein data (biomarker score), co-morbidity data (ischemic cardiomyopathy), clinical

evaluation data (QOL score), and demographic data (patient age). In addition, LIME revealed features responsible for personalized prediction and showed diverse feature sets responsible for individual response to treatment. All of these interpretable techniques only provide a clear picture of the mechanistic insight of the model, not the causality [31]. Of course, we must emphasizethat the power of these features to predict CRT response is indicative of their correlation to cardiacremodeling and not necessarily indicative of their mechanistic causation of cardiac remodeling. Additional notable limitations include a relatively short follow-up time of 6 months and a relativelysmall patient sample size (compared to thousands of patients' data used in electronic health record-based algorithms).

Current clinical guidelines define specific eligibility criteria for physicians to base their CRT recommendations [32]. The increasing accuracy of computational predictions suggests that incorporating personalized model-based probabilities could benefit such recommendation criteria.Encouragingly, our patient stratification demonstrated 96% accuracy in the highest and lowest response subgroups with significant differences in volume changes and functional changes over six months post-CRT. Our algorithm was built and tested using only baseline, pre-CRT measurements, demonstrating that it is feasible for machine learning algorithms to harness a composite set of data from the demographic, functional, and biomarker domains obtained at the time of patient evaluation for CRT and provide predictive value on the ultimate CRT response. As future model developments are likely to further improve prediction accuracy across a broader number of patients, future clinical and ethical discussions will prove vital to appropriately leveragethis predictive information into CRT decisions.

2.5 Conclusion

In this study, we have shown that integrating multiple types of data and biomarkers scoring improves the predictive capability of machine learning algorithms to identify CRT responders and non-responders. Our ensemble model combining all data types has better predictive power than only using cardiac functional markers. We have also shown the Global & Local Explainability of our ML model in terms of overall and personalized prediction. These explainabilities will be helpful to understand the response to the treatment in depth. They will help understand the nature of response for individual CRT recipients. Therefore, this research perfectly aligns with the goal of personalized precision medicine in cardiovascular diseases.

2.6 Acknowledgments

We gratefully acknowledge Doug Stubbs, Dr. Nina Hubig, and Dr. Frank Spinale (USC School of Medicine) for their collaboration on this chapter.

2.7 References

1. Prinzen FW, Vernooy K, Auricchio A. Cardiac resynchronization therapy: state-of-the-art of current applications, guidelines, ongoing trials, and areas of controversy. Circulation. 2013;128(22):2407-2418. doi:10.1161/CIRCULATIONAHA.112.000112

2. McAlister FA, Ezekowitz J, Hooton N, et al. Cardiac resynchronization therapy for patients with left ventricular systolic dysfunction: a systematic review. JAMA. 2007;297(22):2502-2514. doi:10.1001/jama.297.22.2502

3. Hawkins NM, Petrie MC, MacDonald MR, Hogg KJ, McMurray JJ. Selecting patients

for cardiac resynchronization therapy: electrical or mechanical dyssynchrony?. Eur Heart J. 2006;27(11):1270-1281. doi:10.1093/eurheartj/ehi826

4. Daubert C, Behar N, Martins RP, Mabo P, Leclercq C. Avoiding non-responders to cardiac resynchronization therapy: a practical guide. Eur Heart J. 2017;38(19):1463-1472. doi:10.1093/eurheartj/ehw270

5. Achilli A, Peraldo C, Sassara M, et al. Prediction of response to cardiac resynchronization therapy: the selection of candidates for CRT (SCART) study. Pacing Clin Electrophysiol. 2006;29 Suppl 2:S11-S19. doi:10.1111/j.1540-8159.2006.00486.x

6. Spinale FG, Meyer TE, Stolen CM, et al. Development of a biomarker panel to predict cardiac resynchronization therapy response: Results from the SMART-AV trial. Heart Rhythm. 2019;16(5):743-753. doi:10.1016/j.hrthm.2018.11.026

7. Hu SY, Santus E, Forsyth AW, et al. Can machine learning improve patient selection for cardiac resynchronization therapy?. PLoS One. 2019;14(10):e0222397. Published 2019 Oct 3. doi:10.1371/journal.pone.0222397

8. Feeny AK, Rickard J, Patel D, et al. Machine Learning Prediction of Response to Cardiac Resynchronization Therapy: Improvement Versus Current Guidelines. Circ Arrhythm Electrophysiol. 2019;12(7):e007316. doi:10.1161/CIRCEP.119.007316

9. Cikes M, Sanchez-Martinez S, Claggett B, et al. Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. *Eur J Heart Fail*. 2019;21(1):74-85. doi:10.1002/ejhf.1333

10. Kalscheur MM, Kipp RT, Tattersall MC, et al. Machine Learning Algorithm Predicts

Cardiac Resynchronization Therapy Outcomes: Lessons From the COMPANION Trial. Circ Arrhythm Electrophysiol. 2018;11(1):e005499. doi:10.1161/CIR-CEP.117.005499

11. Howell SJ, Stivland T, Stein K, Ellenbogen KA, Tereshchenko LG. Using Machine-Learning for Prediction of the Response to Cardiac Resynchronization Therapy: The SMART-AV Study. JACC Clin Electrophysiol. 2021;7(12):1505-1515. doi:10.1016/j.jacep.2021.06.009

12. Zihni E, Madai VI, Livne M, et al. Opening the black box of artificial intelligence for clinical decision support: A study predicting stroke outcome. PLoS One. 2020;15(4):e0231166. Published 2020 Apr 6. doi:10.1371/journal.pone.0231166

13. Petch J, Di S, Nelson W. Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology. Can J Cardiol. 2022;38(2):204-213. doi:10.1016/j.cjca.2021.09.004

14. Amann J, Blasimme A, Vayena E, Frey D, Madai VI; Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310. Published 2020 Nov 30. doi:10.1186/s12911-020-01332-6

15. Chico V. The impact of the General Data Protection Regulation on health research. Br Med Bull. 2018;128(1):109-118. doi:10.1093/bmb/ldy038

16. Elshawi R, Al-Mallah MH, Sakr S. On the interpretability of machine learning-based model for predicting hypertension. BMC Med Inform Decis Mak. 2019;19(1):146.

Published 2019 Jul 29. doi:10.1186/s12911-019-0874-0

17. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy (Basel). 2020;23(1):18. Published 2020 Dec 25. doi:10.3390/e23010018

18. Lundberg SM, Erion G, Chen H, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell. 2020;2(1):56-67. doi:10.1038/s42256-019-0138-9

19. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in neural information processing systems. 2017;30. doi:10.1145/2939672.2939778

20. Friedman JH. Greedy function approximation: A Gradient boosting machine. The Annals of Statistics. 2001;29(5):1189-1232. doi:10.1214/aos/1013203451

21. Ribeiro MT, Singh S, Guestrin C. Why should I trust you?. Proceedings of the 22nd ACM SIGKDD International Conference on knowledge discovery and data mining. 2016:1135-1144. doi:10.1145/2939672.2939778

22. Ellenbogen K, Gold M, Meyer T, et al. Primary results from the SmartDelay determined AV optimization: a comparison to other AV delay methods used in cardiac resynchronization therapy (SMART-AV) trial: a randomized trial comparing empirical, echocardiography-guided, and algorithmic atrioventri. *Circulation*. 2010;122(25):2660- 2668. doi:10.1161/CIRCULATIONAHA.110.992552

23. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res*. 2011;12(85):2825-2830.

http://jmlr.org/papers/v12/pedregosa11a.html. Accessed August 10, 2021.

24. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak. 2019;19(1):281. Published 2019 Dec 21. doi:10.1186/s12911-019-1004-8

25. Shen Z, Wu Q, Wang Z, Chen G, Lin B. Diabetic Retinopathy Prediction by Ensemble Learning Based on Biochemical and Physical Data. Sensors (Basel). 2021;21(11):3663. Published 2021 May 25. doi:10.3390/s21113663

26. Ali S, Hussain A, Aich S, et al. A Soft Voting Ensemble-Based Model for the Early Prediction of Idiopathic Pulmonary Fibrosis (IPF) Disease Severity in Lungs Disease Patients. Life (Basel). 2021;11(10):1092. Published 2021 Oct 15. doi:10.3390/life11101092

27. Ali S, Hussain A, Aich S, et al. A Soft Voting Ensemble-Based Model for the Early Prediction of Idiopathic Pulmonary Fibrosis (IPF) Disease Severity in Lungs Disease Patients. Life (Basel). 2021;11(10):1092. Published 2021 Oct 15. doi:10.3390/life11101092

28. Asan O, Bayrak AE, Choudhury A. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. J Med Internet Res. 2020;22(6):e15154. Published 2020 Jun 19. doi:10.2196/15154

29. Luo Y, Tseng HH, Cui S, Wei L, Ten Haken RK, El Naqa I. Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling. BJR Open. 2019;1(1):20190021. Published 2019 Jul 4.

doi:10.1259/bjro.20190021

30. McAloon CJ, Barwari T, Hu J, et al. Characterisation of circulating biomarkers before and after cardiac resynchronisation therapy and their role in predicting CRT response: the COVERT-HF study. Open Heart. 2018;5(2):e000899. Published 2018 Oct 18. doi:10.1136/openhrt-2018-000899

31. Alber M, Buganza Tepole A, Cannon WR, et al. Integrating machine learning and multiscale modeling-perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. NPJ Digit Med. 2019;2:115. Published 2019 Nov 25. doi:10.1038/s41746-019-0193-y

32. Osmanska J, Hawkins NM, Toma M, Ignaszewski A, Virani SA. Eligibility for cardiac resynchronization therapy in patients hospitalized with heart failure. ESC Heart Fail. 2018;5(4):668-674. doi:10.1002/ehf2.12297

2.8 Figures Legends

Figure 2.1. Overall performance of the machine learning model.

The receiver-Operating Characteristic curve for the supervised, binary classification ensemble model demonstrates high predictive capability with an area-under-the-curve of 0.784 for the majority voting classifier. (B) Model-predictedresponders exhibited a 69% response rate (61/88), while model-predicted non-responders exhibited only a 27% response rate (19/71). Further stratification based on the model-predicted responses probability score demonstrated a greater predictive accuracy.

Figure 2.2. Cardiac remodeling across patient stratifications.

Model-predicted responders showed statistically significant differences in left ventricle remodelingmetrics compared to the model-predicted non-responders. In particular, binary classification (A) identified a responder group with substantially greater improvements in ESV, EDV and EF from 0-6 months after CRT intervention. (B) More detailed patient stratification further amplified the remodeling differences across groups.

Figure 2.3. SHAP plot and PDP plot showing the feature importance and marginal effect of one feature at a time in the prediction model. (A) SHAP plot showing the feature importance in our model. 1D stretch, biomarker score, ischemic cardiomyopathy, QOL score, and age were indicated as the top 5 most important features for determining patient response probability. The scatter width and separation indicate the feature importance, and the color indicates which direction of that feature value is predictive of high

vs. low patient response. (B) An increase in 1D Stretch from (1.03 to 1.39), Biomarker Score, and History of Ischemic Cardiomyopathy increases the risk of not responding to CRT.

Figure 2.4. LIME plot showing local approximation and interpreting the model locally for two patients; responder vs non-responder. LIME plot is also explaining how top 10 features contributing for individualized response to CRT. (A) 1D Stretch of ≤ 1.08, and no history of RBBB, Atrial Flutter, Ischemic Cardiomyopathy, AT_PSVT, PAF, and SA surgery increased probability of responding favorably to CRT treatment. (B) 1D Stretch of > 1.14 and no history of VT-SVT, Afib, and Non-sustained VT increased the probability of not responding to treatment. On the other hand, history of Ischemic Cardiomyopathy also greatly affecting the non-response to CRT.

2.9 Tables

Table 2.1. Variables acquired from the SMART-AV clinical trial.

Domain	Individual Feature
General Characteristics (9)	Sex, Age, Height, Weight, Systolic bloodpressure (BPsys), Diastolic Blood Pressure (BPdia), Heart Rate at Rest (HRrest), QOL Score, 6-Minute Walk Distance
Co-Morbidities (10)	Atrial Fibrillation (Afib), Paroxysmal Atrial Fibrillation (PAF), Atrial Flutter, Renal Disease,Chronic Obstructive Pulmonary Disease (COPD), Premature Ventricular Contractions (PVC), Atrial Tachycardia Paroxysmal Supraventricular Tachycardia (AT-PSVT), History of Left Bundle Branch Block (LBBB), History of Right Bundle Branch Block (RBBB), Ischemic Cardiomyopathy
Surgical History (3)	Sinoatrial (SA) Node Surgery, Coronary Artery Bypass Graft (CABG), Pre-Cutaneous Coronary Intervention (PCI)
Medications (3)	Diuretics, Ace inhibitors, or ARBs (ACE-ARB), Digoxin
Echo-based Assessment (7)	Left Ventricular End Diastolic Volume (LVEDV), Left Ventricular End Systolic Volume (LVESV), Left Ventricular Ejection Fraction (LVEF), Stroke Volume (SV), EDV/ESV, 1-Dimensional Stretch (cube root of EDV/ESV), Center size
ECG (12)	AV Interval without Atrial Pacing, PR Interval without Atrial Pacing, QRSWidth, VT None, VT Non-sustained, VT Supraventricular

	Tachycardia (VT-SVT), Sick Sinus, Paced AV Delay, Echo Optimized AV Delay, Iterative AVDelay, Fixed AV Delay, Sensed AV Delay
Biomarker	Matrix Metalloproteinase 2 (MMP-2), Matrix Metalloproteinase 9 (MMP9), Suppression of Tumorigenicity 2 (sST-2), C-ReactiveProtein (CRP), N-terminal pro B-type Natriuretic Peptide (NT- proBNP), Tissue Inhibitors of Metalloproteinase 1 (TIMP1), Tissue Inhibitors of Metallo-proteinase 2 (TIMP2), Tissue Inhibitors of Metalloproteinase 4 (TIMP4), Soluble Glycoprotein 130 (sGP130), Soluble Interleukin 2 Receptor Al-pha (sIL2Ra), Tumor Necrosis Factor Receptor II (sTNFR-II), Interferon Gamma (IFNg)

*** All the Biomarkers were used for creating the initial model. MMP-2, sST-2, CRP, and sTNFR-II are used for Biomarker Scoring.

Table 2.2. Baseline characteristics of CRT Responders and Non-Responders

Feature Name	All (n=794)	CRT Responder (n=396)	CRT Non-Responder (n=398)	p-value
Continuous Variables, unit				
Age, year	65.8±10.8	65.6±10.7	66.1±10.9	0.54
Height, cm	171.4±10.4	171.6±10.0	171.3±10.7	0.67
Weight, kg	87.4±20.8	88.2±20.6	86.6±20.9	0.27
BPSys, mm Hg	123.9±20.3	123.2±19.4	124.6±21.0	0.32
BPDia, mm Hg	71.4±13.4	71.3±13.5	71.5±13.3	0.84
HRrest, bpm	71.1±12.3	70.9±12.7	71.3±11.0	0.68
QOL Score	46.6±24.9	50.0±25.8	43.3±23.5	<0.001
6MW, m	273.4±124.6	262.3±133.1	284.5±114.6	0.012
LVEDV, mL	176.7±72.1	162.9±66.9	190.5±74.5	<0.001
LVESV, mL	131.6±65.6	118.0±60.0	145.1±68.2	<0.001
LVEF, %	27.7±8.8	29.7±9.3	25.6±7.8	<0.001
SV, mL	45.1±14.1	44.9±14.6	45.3±13.6	0.65
EDV/ESV Ratio	1.4±0.2	1.5±0.2	1.4±0.2	<0.001
1D Stretch/Cube Root of EDV/ESV	1.1±0.0	1.1±0.1	1.1±0.0	<0.001
AV Interval (Without Atrial Pacing), ms	252.5±69.1	252.7±70.6	252.2±67.6	0.92

PR Interval (Without Atrail Pacing), ms	197.2±49.8	200.2±51.3	194.2±48.3	0.09
QRS Width (Without Atrial Pacing), ms	153.6±27.3	150.6±26.7	156.7±27.6	0.002
Iterative AV Delay (Recommended), ms	127.7±38.2	131±36.3	123.4±39.6	0.002
Paced AV Delay (Recommended), ms	174.9±39	181.2±41.2	168.6±37.7	<0.001
Sensed AV Delay (Recommended), ms	127.1±37.3	132.4±39	121.8±34.9	<0.001
Biomarker CRT Score (0,1,2,3,4)	1.7±1.2	2±1.1	1.4±1.1	<0.001
Binary Categorical Variables, n (%)				
Sex (Male)	565 (67.4%)	281 (71%)	254 (63.8%)	0.04
Atrial Fibrillation (Afib)	99 (12.5%)	57 (14.4%)	42 (10.6%)	0.13
PAF	97 (12.2%)	56 (14.1%)	41 (10.3%)	0.12
Atrial Flutter	10 (1.3%)	9 (2.3%)	1 (0.3%)	0.03
Renal Disease	117 (14.7%)	63 (15.9%)	54 (13.6%)	0.41
COPD	115 (14.5%)	64 (16.2%))	51 (12.8%)	0.22
PVC	13 (1.6%)	8 (19.7%)	5 (1.3%)	0.57
AT-PSVT	12 (1.5%)	8 (2%)	4 (1%)	0.38
LBBB	611 (77%)	272 (68.7%)	339 (85.2%)	<0.001
RBBB	103 (13%)	78 (19.7%)	25 (6.3%)	<0.001

Ischemic Cardiomyo-pathy	445 (56%)	271 (68.4%)	174 (43.7%)	<0.001
SA Surgery	93 (11.7%)	58 (14.6%)	35 (8.8%)	0.01
CABG	257 (32.4%)	163 (41.2%)	94 (23.6%)	<0.001
PCI	248 (31.2%)	155 (39.1%)	93 (23.4%)	<0.001
Diuretic	637 (80.2%)	316 (79.8%)	321 (80.7%)	0.83
ACE-ARB	674 (84.9%)	325 (82.1%)	349 (87.7%)	0.03
Digoxin	176 (22.2%)	96 (24.2%)	80 (20.1%)	0.19
Small Centersize	176 (22.2%)	86 (21.7%	90 (22.6%)	0.83
Sick Sinus	53 (6.7%)	24 (6.1%0	29 (7.3%)	0.58
VT None	654 (82.4%)	311 (78.5%)	343 (86.2%)	0.006
VT Non Sustained	95 (12%)	51 (12.9%)	44 (11.1%)	0.5
VT SVT	3 (0.4%)	2 (0.5%)	1 (0.3%)	0.99
Echo Optimized AV Delay Group	261 (32.9%)	128 (32.3%)	133 (33.4%)	0.8
Fixed AV Delay Group	262 (33.0%)	139 (35.1%)	123 (30.9%)	0.24

Table 2.3. Comparison of the performance of the top 6 models in our study using Biomarker Scoring

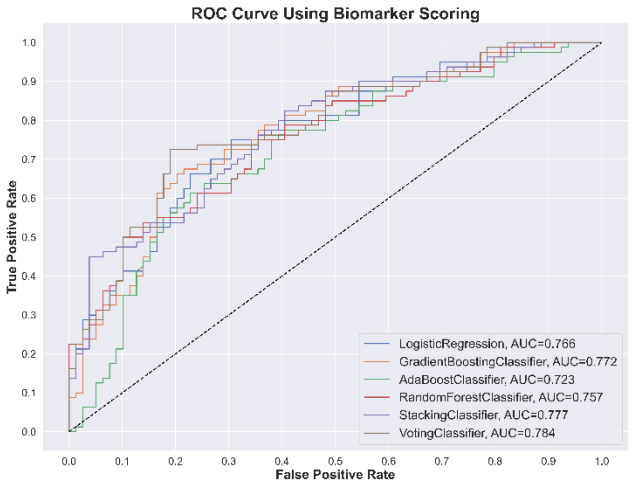
Model Name	Accuracy Train	Accuracy Test	Recall Train	Recall Test	ROC-AUC Test	F1 Test	MCC Test
Voting Classifier	1.000	0.730	0.997	0.713	0.784	0.726	0.460
Stacking Classifier	0.855	0.723	0.915	0.800	0.772	0.744	0.451
Gradient Boosting Classifier	1.000	0.730	1.000	0.625	0.775	0.699	0.470
Logistic Regression	0.706	0.692	0.733	0.763	0.766	0.713	0.387
Random Forest Classifier	0.935	0.679	0.940	0.750	0.757	0.702	0.361
Adaptive Boosting Classifier	0.751	0.667	0.802	0.775	0.723	0.701	0.340

Table 2.4. Area-Under-the-Curves (AUC) for the ML models with or without the biomarker data.

Feature Set	Biomarker Feature Used	Train AUC (n=635)	Test AUC (n=159)
No Biomarkers	None	0.63	0.74
All Biomarkers	MMP-2, MMP9, sST-2, CRP, NT- proBNP, TIMP1, TIMP2, TIMP4, sGP130, sIL2Ra, sTNFR-II, IFNg	0.75	0.77
Biomarker Score (0,1,2,3,4)	MMP-2 (≥982,000 pg/mL), sST-2(≥23,721 pg/mL), CRP (≥7381 ng/mL), sTNFR-II (≥7,090 pg/mL)	0.75	0.78

2.10 Figures

A)



B)

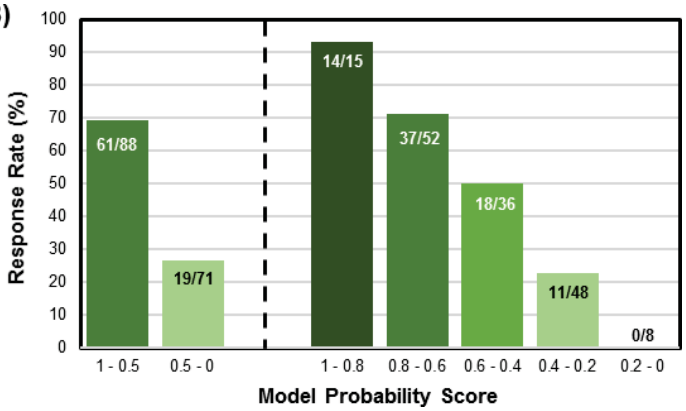


Figure 2.1.

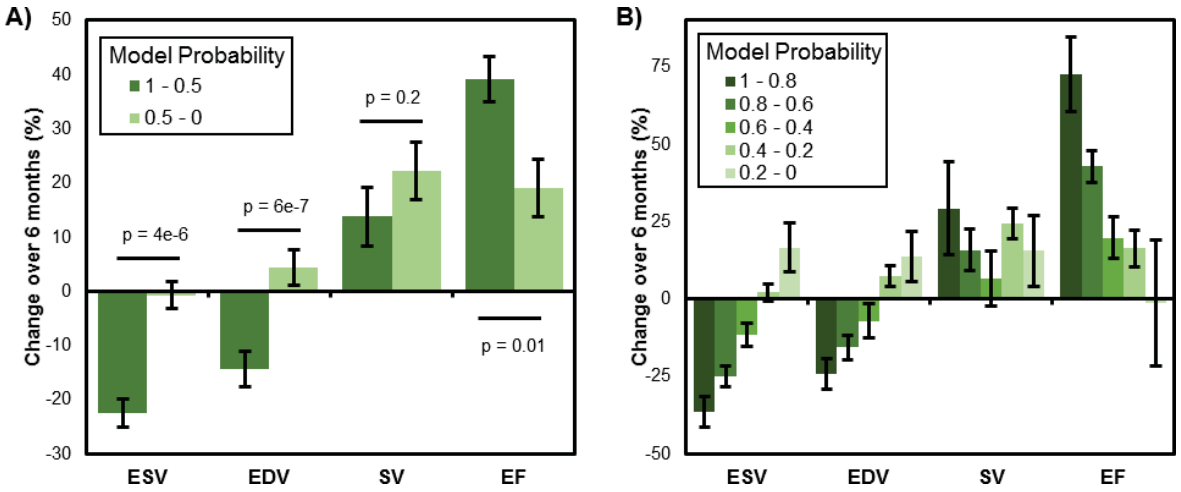

A)
Model Probability
1 - 0.5
0.5 - 0
Change over 6 months (%)
50
40
30
20
10
0
-10
-20
-30
p = 4e-6
p = 6e-7
p = 0.2
p = 0.01
ESV
EDV
SV
EF
B)
Model Probability
1 - 0.8
0.8 - 0.6
0.6 - 0.4
0.4 - 0.2
0.2 - 0
Change over 6 months (%)
75
50
25
0
-25
-50
ESV
EDV
SV
EF

Figure 2.2.

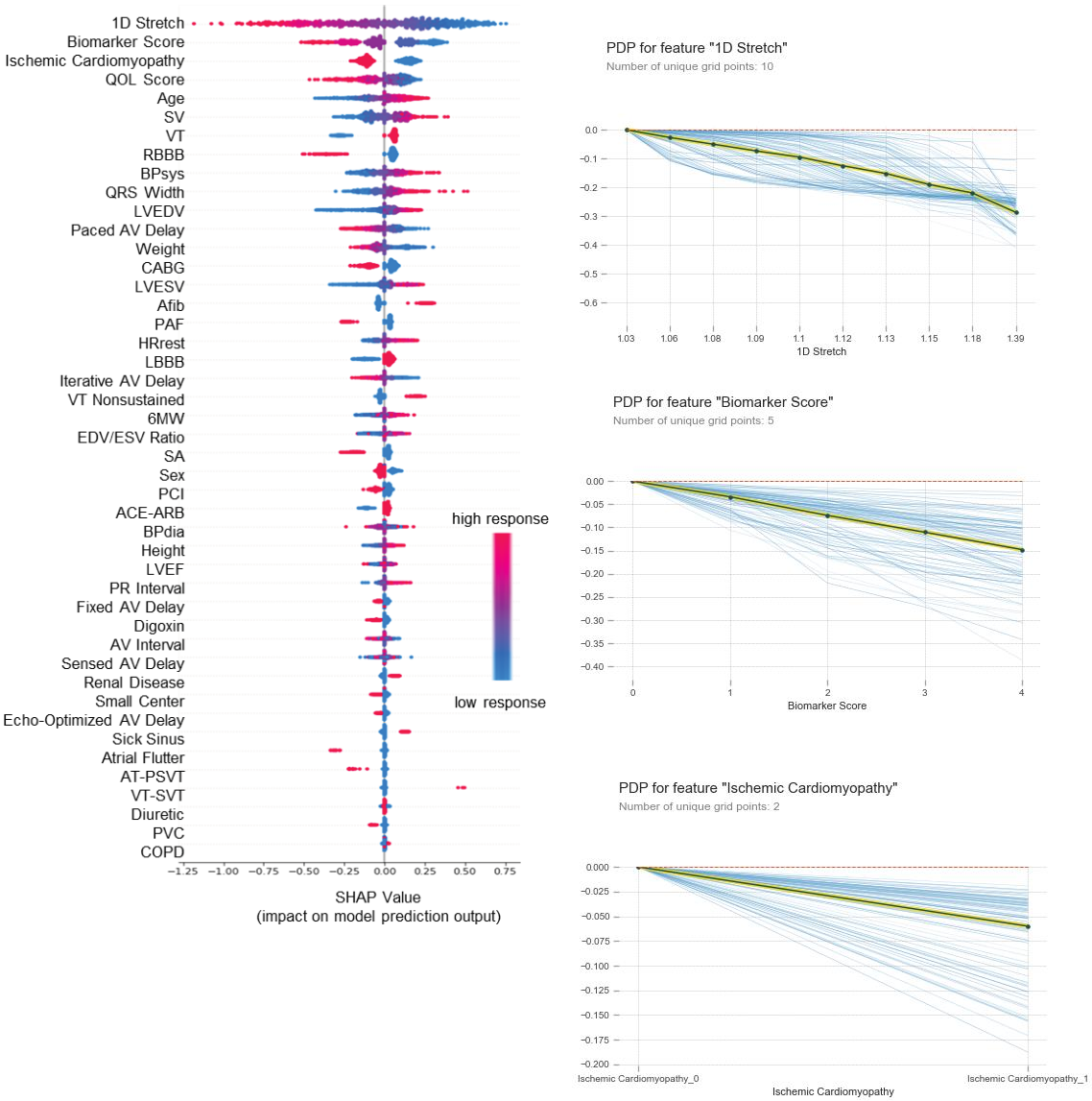

1D Stretch
Biomarker Score
Ischemic Cardiomyopathy
QOL Score
Age
SV
VT
RBBB
BPsys
QRS Width
LVEDV
Paced AV Delay
Weight
CABG
LVESV
Afib
PAF
HRrest
LBBB
Iterative AV Delay
VT Nonsustained
6MW
EDV/ESV Ratio
SA
Sex
PCI
ACE-ARB
BPdia
Height
LVEF
PR Interval
Fixed AV Delay
Digoxin
AV Interval
Sensed AV Delay
Renal Disease
Small Center
Echo-Optimized AV Delay
Sick Sinus
Atrial Flutter
AT-PSVT
VT-SVT
Diuretic
PVC
COPD
high response
low response
SHAP Value
(impact on model prediction output)
PDP for feature "1D Stretch"
Number of unique grid points: 10
1D Stretch
PDP for feature "Biomarker Score"
Number of unique grid points: 5
Biomarker Score
PDP for feature "Ischemic Cardiomyopathy"
Number of unique grid points: 2
Ischemic Cardiomyopathy_0
Ischemic Cardiomyopathy_1
Ischemic Cardiomyopathy

Figure 2.3.

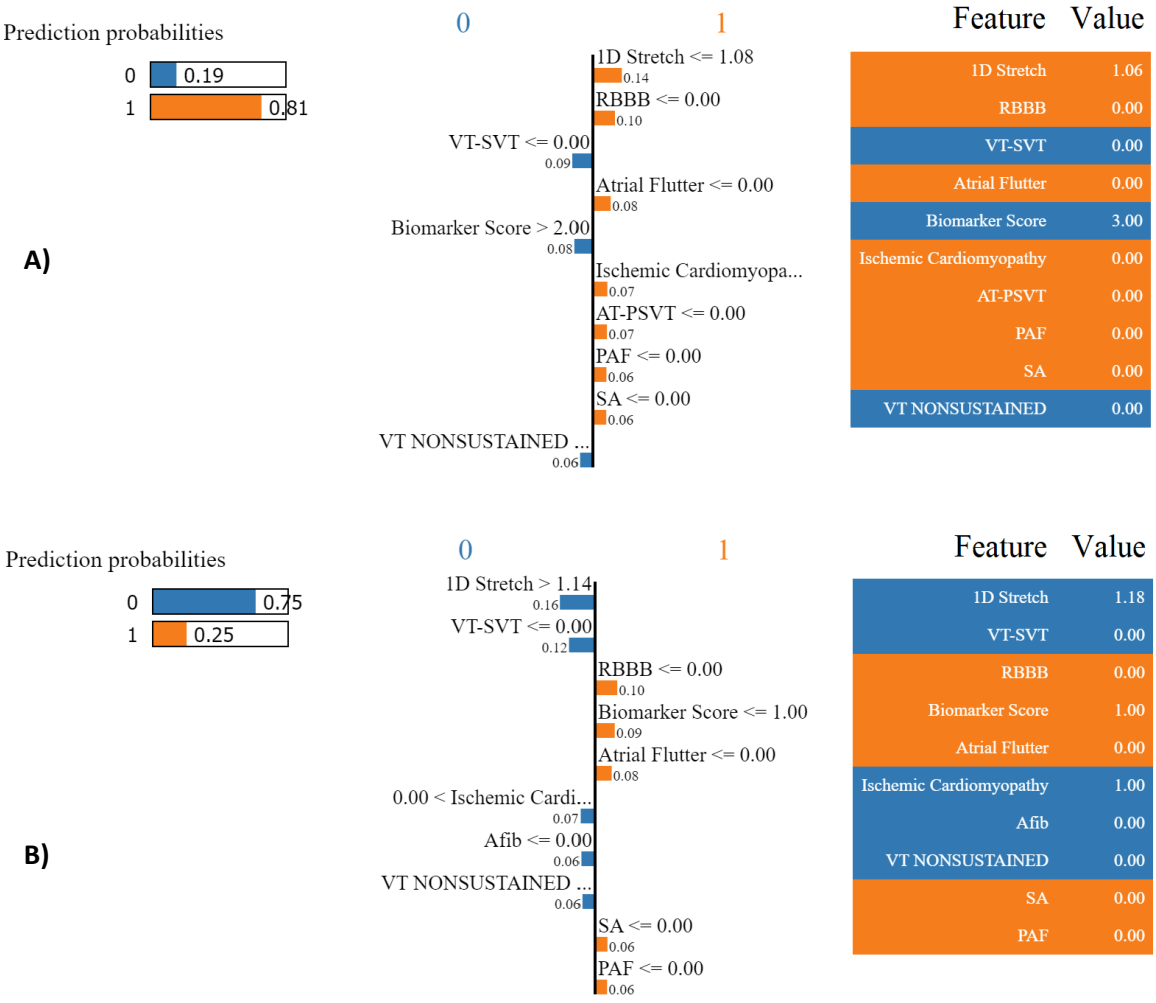

Prediction probabilities
0
1
0.19
0.81
A)
0
1
1D Stretch <= 1.08
0.14
RBBB <= 0.00
0.10
VT-SVT <= 0.00
0.09
Atrial Flutter <= 0.00
0.08
Biomarker Score > 2.00
0.08
Ischemic Cardiomyopa...
0.07
AT-PSVT <= 0.00
0.07
PAF <= 0.00
0.06
SA <= 0.00
0.06
VT NONSUSTAINED ...
0.06
Feature
Value
1D Stretch 1.06
RBBB 0.00
VT-SVT 0.00
Atrial Flutter 0.00
Biomarker Score 3.00
Ischemic Cardiomyopathy 0.00
AT-PSVT 0.00
PAF 0.00
SA 0.00
VT NONSUSTAINED 0.00
Prediction probabilities
0
1
0.75
0.25
B)
0
1
1D Stretch > 1.14
0.16
VT-SVT <= 0.00
0.12
RBBB <= 0.00
0.10
Biomarker Score <= 1.00
0.09
Atrial Flutter <= 0.00
0.08
0.00 < Ischemic Cardi...
0.07
Afib <= 0.00
0.06
VT NONSUSTAINED ...
0.06
SA <= 0.00
0.06
PAF <= 0.00
0.06
Feature
Value
1D Stretch 1.18
VT-SVT 0.00
RBBB 0.00
Biomarker Score 1.00
Atrial Flutter 0.00
Ischemic Cardiomyopathy 1.00
Afib 0.00
VT NONSUSTAINED 0.00
SA 0.00
PAF 0.00

Figure 2.4.

Chapter 3

Building a Fibrosis Related Gene Regulatory Network in Dilated Cardiomyopathy Patients

3.1 Introduction

Cardiomyopathies are pathologies of cardiac muscle that result in the change of the mechanical and electrical disruption of the heart [1]. Cardiomyopathies have several phenotypes, such as Hypertrophic (HCM), Dilated (DCM), Peripartum (PPCM), and Restrictive Cardiomyopathy [2]. Some types of cardiomyopathies are common in young adults (DCM), while some types of cardiomyopathies may result from physical change due to age (restrictive) or pregnancy (PPCM). Dilated cardiomyopathy is non-ischemic cardiomyopathy that may result in ventricular dysfunction even without risk factors such as coronary artery disease, valvular disease, or congenital disorders. According to American Heart Association definitions, DCM can be hereditary and nonhereditary. Genetic cardiomyopathy involves the mutation of genes vital for cardiac functions. Nonhereditary dilated cardiomyopathy results from bacterial infection, protozoal infection, viral infections, autoimmune reactions, toxin exposure, and adverse neuromuscular changes. Dilated cardiomyopathy sometimes results in no symptoms, especially in the early stage of the disease. As the symptoms progress, cardiomyopathy patients feel fatigued, have shortness of breath during physical activity or while lying down, reduced ability to exercise, swelling (edema) in the legs, feet, and abdomen, discomfort in the chest, and rapid palpitations. Globally the incidence of cardiomyopathy has not drastically decreased even with

improved treatment and disease management [3]. According to a CDC study, 1 in 500 adults have cardiomyopathy-related conditions [4].

In the USA, several studies have attempted to identify risk factors for cardiovascular diseases. Some of these studies have been incredibly successful and have contributed to our predictive knowledge of cardiovascular disease [5]. One study, the Myocardial Applied Genomic Network (MAGNet) [7], collected extensive data from the left ventricle of DCM patient. The data of this study has been widely used to investigate the population genetics of DCM. Our study is unique in terms of its goal. While other studies aim to investigate the differential gene expression in dilated cardiomyopathy compared to non-failing patients, our goal is to identify the fibrotic factors involved in dilated cardiomyopathy.

A large part of dilated cardiomyopathy involves hereditary changes [8]. Therefore, it is important to understand the genetic landscape of cardiac myopathy. Several genes are involved in dilated cardiomyopathy phenotypes. The main genetic factors of dilated cardiomyopathy are Titin (TTN), Lamin A/C (LMNA), Myosin heavy chain (MYH7), Myosin binding protein C (MYBPC3), Myopalladin (MYPN), Sodium Channel alpha unit (SCN5A), BaCl2-associated Athano Gene 3 (BAG3), and Phospholamban (PLN). While they are important biomarkers for DCM, little efforts have been made to integrate these markers in a multipath way network of interactions to mechanistically understand their role in cardiac remodeling.

Myocardial fibrosis, which is defined as the deposition of extracellular matrix in the heart, is an important remodeling process and is intrinsically involved with cardiomyopathy phenotypes [9]. Therefore, it is important to study cardiomyopathy in the context of fibrosis. The very first step is to understand the genomic signature/hallmark signature of cardiac

fibrosis to find a link between the genetic signature of DCM and fibrosis. Gene regulatory networks are collections of different molecules within cells that work together to activate specific target molecules, leading to specific phenotypes. These molecules include DNA, RNA, and proteins [10]. Transcription factors (TFs) and their target relationships are important components of gene regulatory networks. Differential gene regulatory networks activate different cellular responses by activating different genes and transcription factors in the gene regulatory networks. Therefore, specific disease phenotypes have their own network of transcription factors (TFs). In this study, we are studying such networks in the context of cardiac fibrosis in dilated cardiomyopathy patients.

This chapter has these purposes: identifying important genetic modules that differ between non-failure controls (NF) and DCM patients, identifying pathways that are involved in this differential expression and finally identifying the transcription factors that paly major role in cardiac fibrosis in DCM patients. The purpose of the study is depicted in Figure 3.1. The first step involves differential expression analysis, followed by gene set enrichment analysis and weighted gene co-expression analysis. This is followed by inferring a large subnetwork in the upregulated and downregulated genes in the differentially expressed gene analysis. For the second part, we have built a transcription factor and target network and subsequently filtered it to identify the TF factors and targets responsible for causing fibrosis in DCM patients. These filtered transcription factors will be used subsequently to build the composite network in aim 3.

3.2 Materials and Methods

3.2.1 Data Source

The study was conducted using the Myocardial Applied Genomic Network Study (MAGNet) data source. MAGNet aimed to collect human left ventricle tissues for research purposes. The left ventricular free-wall tissue was collected from cardiomyopathy patients during cardiac surgery for heart transplant. For non-failing controls, samples were collected from unused donor hearts with normal functions [4]. RNA-Sequencing libraries were prepared against the hg19/hGRC37 using the STAR aligner. The RNA-seq data from this study has been stored in the NCBI GEO database (GSE141910). Two sets of data were utilized for this study. Raw read counts from the study were used for differential expression analysis, and these counts were obtained from the study's GitHub page (https://github.com/mpmorley/MAGNet). To identify important transcription factors, Surrogate Variable Analysis (SVA) normalized data from the same GitHub repository was retrieved [13]. The dataset consisted of 166 non-failing samples and 166 DCM samples.

3.2.2 Identification of the Differentially Expressed Genes

For the differential expression analysis, functional enrichment, and GO analysis, we utilized the integrated Differential Expression and Pathway analysis (iDEP) web application [11]. This web application comprises 63 R/Bioconductor packages, 2 web services, and is connected to several data repositories containing 220 plant and animal species. Since the raw data included numerous zero Counts Per Million Values (CPM) rows, we filtered the dataset for a minimum CPM of 1 and presence in at least 2 libraries. For clustering and Principal Component Analysis (PCA), we transformed the data using the EdgeR log2(CPM+c) transformation [12]. Missing values were imputed using the median method

to replace the missing CPM values. To cluster the gene expression data, we employed the elbow method to determine the optimal number of clusters. For the differential expression analysis, we utilized the limma-voom transformation method to analyze the data. We only considered genes whose log2 fold change was at least 2-fold. Due to the large sample size, we could not use the DESeq2 method. Furthermore, we performed enrichment analysis on the differentially expressed genes, both upregulated and downregulated, using the GO biological process for pathway enrichment.

3.2.3 Functional Enrichment Analysis

We utilized the Generally Applicable Gene-set/Pathway Analysis (GAGE) for pathway enrichment analysis [14]. A pathway significance cutoff (FDR) of 0.4 was applied, and the top pathways were identified. Additionally, we identified the significant KEGG pathways that were upregulated in the differentially expressed genes.

3.2.4 Weighted Gene CO-expression analysis

We also used the identified the weighted gene expression network WGCNA package [15]. We have only included the genes that are differentially expressed and shared among pathways between the Non Failing and DCM patients.

3.2.5 Target Gene-TF regulatory network analysis

3.2.5.1 Gene Regulatory Network Inference

We applied the method described by Rogers et al. [16] to derive the TF-target gene regulatory network using the dataset derived from the MAGNet study. To maintain the reproducibility of the study, we used the dataset from the MAGNet study to infer the gene regulatory network. To normalize the expression levels of the dataset, we employed the

counts per million (CPM) method and then transformed the data using the VOOM procedure in the LIMMA R package. Surrogate Variables (SV), which account for latent sources of variation such as batch effects, were calculated using the svaseq function from the R SVA package [13]. Differential gene expression analysis between races was performed using the LIMMA R package. We used a linear model with the following form: $Y = \beta0 + \beta1 \times race + \beta2 \times sex + \beta3 \times Age + \beta4\text{-}14 \times SVA1{:}SVA11$. In this model, Y represents the log2-transformed gene expression, and race is either African Americans or European Americans. We adjusted for gender, age, and 11 other surrogate variables, including etiology, weight, height, heart weight (hw), size of left ventricular mass, atrial fibrillation, presence of ventricular fibrillation/ventricular flutter, diabetes, hypertension, Left Ventricular Ejection Fraction (LVEF), RNA integrity number (RIN), and Transcript Integrity Number (TIN).

For the gene regulatory network analysis, we implemented the GRNBoost2 algorithm using the Arboretto library in Python [17]. The GRNBoost2 method is based on the GENIE3 algorithm [18], which predicts regulatory links between input and target genes using a decision tree ensemble method. The expression of a group of input genes is used to predict the expression of a target gene through the decision tree ensemble method. We chose the GENIE3 algorithm for several reasons [19]: it is widely used as a standard gene regulatory network inference algorithm, requires minimal assumptions about the network topology, provides directed interactions, and can infer nonlinear and combinatorial regulation compared to other regression-based methods.

Overall, we used the method by Rogers et al. to construct the TF-target gene regulatory network using the MAGNet study dataset. We performed preprocessing steps, conducted differential gene expression analysis, and employed the GRNBoost2 algorithm

based on the GENIE3 algorithm for network inference, taking advantage of its ability to infer regulatory interactions in gene networks.

3.2.5.2 Network Pruning

The initial network consisted of multiple interactions that may or may not be relevant to the context of cardiac fibrosis. To fit the network to our specific context and dataset, we applied three filtering steps, similar to Rogers et al.'s approach. In the first filtering step, we only selected interactions that are experimentally validated and involve transcription factors. This ensures that our network includes only reliable interactions supported by experimental evidence. In the second step, we focused on transcription factors that have evidence-based connections with mediators of cardiac fibrosis. This step helps refine the network by prioritizing transcription factors directly associated with the pathways involved in cardiac fibrosis. In the third filtering step, we retained only the interactions that were ranked highly by the GRNBoost2 algorithm. GRNBoost2 is used to predict regulatory links, and by selecting high-ranking interactions, we aim to prioritize the most influential and significant connections in the network.

For the first pruning step, we specifically chose transcription factors that have been validated in the CHIPX and TRANSFAC databases [20]. Transcription factors not found in these databases are not considered regulatory elements for our network. In the second step, we selected transcription factors that are connected to proteins involved in the extracellular matrix (ECM) turnover, ECM organization, and acute immune responses during cardiac fibrosis. This additional criterion ensures that the network is centered with transcription factors relevant to cardiac fibrosis. In addition, we added top genes that are differentially expressed in NF vs DCM patients (MYH6, COL22A1, COL10A1)

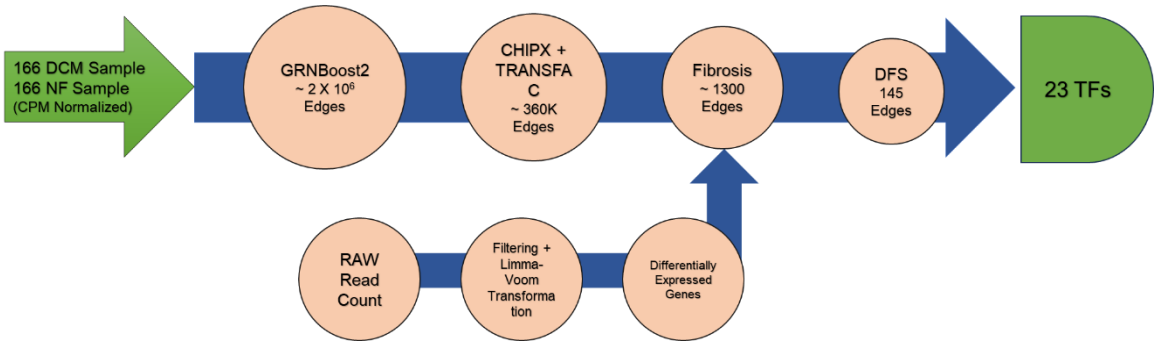


Figure 3.1. Differential gene expression analysis and integration of the DEGs into Gene Regulatory Network to build DCM related network.

3.3 Results

3.3.1 Differentially expressed genes in Normal vs Dilated Cardiomyopathy Patients

After preprocessing and filtering the data to ensure a minimum of 1 Count Per Million (CPM) read in at least 2 libraries, we obtained a total of 20,853 genes that passed the filter. To handle missing values, we employed the median-based imputation method. Following the imputation, we identified 870 genes that were significantly upregulated in the DCM tissue samples compared to the non-failing (NF) heart tissues, with a P-value < 0.05. In contrast, 358 genes were found to be downregulated in the DCM patients compared to NF samples.

We visualized the log2 fold change of all the genes in the volcano plot, where the grey area represents genes with no significant change in expression (neither up nor downregulated). The left blue region corresponds to downregulated genes, while the right red region indicates upregulated genes. Additionally, we presented a scatter plot illustrating the expression levels of the genes across all the samples.

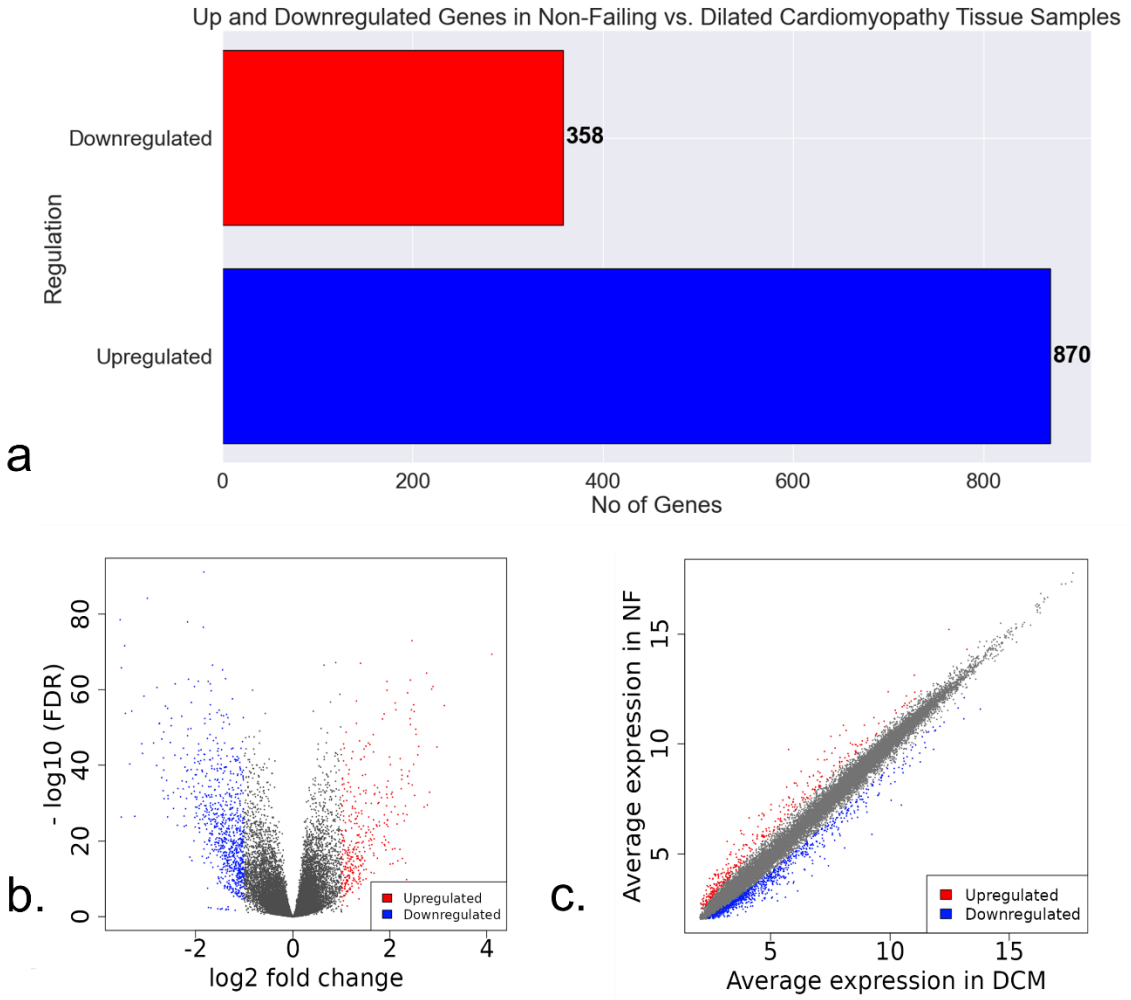


Figure 3.2. In the differential expression analysis between Non-Failing (NF) and Dilated Cardiomyopathy (DCM) tissue samples, we calculated the total number of differentially expressed genes using the limma-voom method. a. The total number of differentially expressed genes calculated using the limma-voom method. b. The Volcano plot shows the distribution of genes across different expression levels between NF and DCM samples. c. The Scatter plot demonstrates the total number of genes upregulated vs downregulated in Non-Failing and Dilated heart samples.

3.3.2 Gene Enrichment Pathways in Differentially Expressed Genes

Enrichment analysis using the Gene Ontology (GO) process has revealed that significantly downregulated genes are prominently enriched (indicated by low adjusted p-values) in pathways associated with extracellular matrix and structural organization, external encapsulating structure organization, immune response, and cell exterior organization. Conversely, upregulated genes are predominantly enriched in pathways related to immune response activation. Furthermore, we have individually identified the top upregulated and downregulated genes. Additionally, we observed that certain cell signaling pathway modules were upregulated in our analysis.

Table 3.1. Enriched pathways and the number of significantly up and down regulate genes in those pathways.

Direction	Adjusted P-values	No of the Genes	Pathways that have these genes
Down regulated	8.8e-17	58	Extracellular matrix organization
	8.8e-17	58	Extracellular structure organization
	8.8e-17	58	External encapsulating structure organization
	2.8e-13	147	Immune response
	5.2e-12	118	Regulation of immune system process
	6.9e-11	175	Cell surface receptor signaling pathway
	4.2e-10	175	Immune system process
	6.1e-10	104	Biological adhesion
	6.3e-10	111	Defense response

	8.7e-10	103	Cell adhesion
	1.1e-09	48	Leukocyte migration
	1.1e-09	160	Response to external stimulus
	1.3e-09	23	Collagen fibril organization
	3.1e-09	155	Regulation of multicellular organismal process
	1.3e-08	48	T cell activation
Up regulated	7.2e-06	9	Acute-phase response
	7.2e-06	35	Inflammatory response
	2.7e-04	11	Acute inflammatory response
	3.6e-03	46	Defense response
	4.4e-03	5	Leukocyte migration involved in inflammatory response
	6.6e-03	19	Regulation of inflammatory response
	8.6e-03	23	Chemotaxis
	8.6e-03	64	Response to external stimulus
	8.6e-03	23	Response to wounding
	8.6e-03	12	Negative regulation of peptidase activity
	8.6e-03	3	Negative regulation of plasminogen activation
	8.6e-03	23	Taxis
	8.6e-03	40	Secretion
	8.6e-03	3	Negative regulation of fibrinolysis
	8.6e-03	19	Cellular divalent inorganic cation homeostasis

In addition to identifying the pathways enriched by the differentially expressed genes, we have listed the top genes in Table 3.2. Most of the genes presented in the table are protein-coding genes. Additionally, there are several long non-coding RNAs (lncRNAs) among the pool of differentially expressed genes. Both genes and lncRNAs are listed based on their corresponding p-values, with those having the largest p-values shown.

Table 3.2. Top genes that are differentially expressed in non-failing vs the dilated cardio-myopathy tissue samples.

Ensemble ID	log2 Fold Change	Adjusted P-value	Gene Symbol	Type
ENSG00000115602	4.10	4.19e-70	IL1RL1	Protein Coding
ENSG00000233485	-3.56	3.19e-79	FHAD1-AS1	lncRNA
ENSG00000188536	-3.53	4.24e-27	HBA2	Protein Coding
ENSG00000100095	-3.53	1.62e-66	SEZ6L	Protein Coding
ENSG00000106483	-3.46	2.47e-72	SFRP4	Protein Coding
ENSG00000169436	-3.45	2.02e-54	COL22A1	Protein Coding
ENSG00000244734	-3.36	4.84e-41	HBB	Protein Coding
ENSG00000181195	-3.32	5.04e-55	PENK	Protein Coding
ENSG00000206172	-3.26	2.65e-27	HBA1	Protein Coding
ENSG00000169385	3.13	1.66e-56	RNASE2	Protein Coding
ENSG00000105205	-3.11	2.21e-46	CLC	Protein Coding
ENSG00000100079	-3.10	7.97e-44	LGALS2	Protein Coding

ENSG00000198768	-3.06	5.42e-59	APCDD1L	Protein Coding
ENSG00000164694	-3.00	6.27e-85	FNDC1	Protein Coding
ENSG00000187922	2.97	1.60e-45	LCN10	Protein Coding
ENSG00000152086	2.89	1.52e-61	TUBA3E	Protein Coding
ENSG00000162383	-2.87	1.36e-46	SLC1A7	Protein Coding
ENSG00000075886	2.86	7.35e-61	TUBA3D	Protein Coding
ENSG00000179593	2.82	1.10e-33	ALOX15B	Protein Coding
ENSG00000145681	-2.80	1.29e-40	HAPLN1	Protein Coding
ENSG00000143768	-2.80	2.79e-61	LEFTY2	Protein Coding
ENSG00000137558	2.77	2.55e-30	PI15	Protein Coding
ENSG00000197616	2.76	4.33e-65	MYH6	Protein Coding
ENSG00000254636	-2.75	9.91e-52	ARMS2	Protein Coding
ENSG00000179639	-2.74	9.07e-52	FCER1A	Protein Coding
ENSG00000248187	2.71	9.33e-30	N/A	lncRNA
ENSG00000150551	-2.70	1.60e-37	LYPD1	Protein Coding
ENSG00000100450	-2.70	1.27e-47	GZMH	Protein Coding
ENSG00000077274	-2.67	4.46e-34	CAPN6	Protein Coding
ENSG00000144406	-2.66	1.19e-42	UNC80	Protein Coding
ENSG00000080644	-2.66	2.07e-39	CHRNA3	Protein Coding
ENSG00000166428	-2.66	1.71e-49	PLD4	Protein Coding
ENSG00000169245	-2.65	5.69e-32	CXCL10	Protein Coding
ENSG00000101825	-2.62	2.31e-56	MXRA5	Protein Coding
ENSG00000165623	-2.60	8.05e-37	UCMA	Protein Coding

ENSG00000198033	2.60	1.13e-45	TUBA3C	Protein Coding
ENSG00000225526	-2.58	3.85e-27	MKRN2OS	Protein Coding
ENSG00000125869	-2.58	9.32e-56	LAMP5	Protein Coding
ENSG00000124713	2.58	1.95e-43	GNMT	Protein Coding
ENSG00000123500	-2.57	8.37e-33	COL10A1	Protein Coding
ENSG00000000005	-2.57	5.79e-32	TNMD	Protein Coding
ENSG00000267206	2.54	6.86e-49	LCN6	Protein Coding
ENSG00000142973	2.53	2.95e-50	CYP4B1	Protein Coding
ENSG00000121005	-2.52	1.01e-46	CRISPLD1	Protein Coding
ENSG00000261116	2.51	7.99e-33	N/A	lncRNA
ENSG00000233682	-2.51	3.65e-51	N/A	lncRNA
ENSG00000138356	2.50	6.27e-55	AOX1	Protein Coding
ENSG00000242258	-2.50	4.11e-53	LINC00996	lncRNA
ENSG00000106236	2.49	1.04e-56	NPTX2	Protein Coding
ENSG00000267653	2.49	6.11e-41	N/A	lncRNA

3.3.3 Overlapping relationships among enriched gene-sets

We presented a network that illustrates the overlapping number of genes in enriched pathways. This network reveals a significant number of genes shared by different pathways in DCM samples.

3.3.4 Identification of Transcription factors that are import for fibrosis

We used an extremely stringent filtering process to identify the 23 transcription factors (TFs) for the context of our study. Using the GRnBoost2 algorithm we first ended up getting almost 2 million interactions. Initial filtering with CHIPX and TRANSFAC reduces the total no TFs- target interactions (Edges) to ~ 360,000. Fibrosis related filtering then reduces the number of edges to around 1300. In our final step, we used an important score >3 and only selected TFs that are only present in at least 8 out of 10 grids of network training. List of the TFs are presented in table 3.3 below.

Figure 3.3. The relationship between two pathways using WGCNA analysis. Here we used a 10% edge cutoff to allow connection between two pathways when there is 10% similarity.

Table 3.3. List of transcription factors derived from MAGNet Study. These transcription factors are connected to 178 edges with an importance score of greater than 3 and present in at least 8 out of 10 training datasets.

Transcription Factors	BCL6, CACYBP, CEBPD, CUX1, EGR1, ETS1, ETS2, HIF1A, LEF1, IKZF1, KLF4, MITF, NR5A2, PPARA, RARG, RUNX1, RELA, TEAD4, RUNX2, TCF4, TFCP2L1, WT1, ZNF281

3.4 Discussion

Cardiac fibrosis is a major pathophysiological manifestation of various heart diseases and can be simply defined as the excessive accumulation of the heart's extracellular matrix. As a pathological manifestation, patients may experience changes in the electrophysiological conduction of the heart. As defined, this process is very complex and involves several pathways and biochemical markers, such as the Matrix Metalloproteases (MMPs) and their tissue Inhibitors (TIMPS), renin-angiotensin-aldosterone system (RAAS), Transformation Growth Factor Beta (TGFβ) pathway, inflammatory mediators (IL1, IL6, and TNFα), and a multitude of long noncoding RNAs (lncRNAs) [21, 22]. Moreover, recent studies have also shown that male-female sex differences result in varying manifestations of cardiac fibrosis [23]. Overall, the process of cardiac fibrosis, though it may seem well-defined, requires further study to validate new target biomarkers for treatment.

Dilated cardiomyopathy is the most prevalent form of cardiomyopathy among the elderly. Surprisingly, there are no large-scale studies that have investigated the pathogenicity, disease manifestation, clinical course of disease, survival analysis, and drug

target discovery using genomic data for dilated cardiomyopathy. The purpose of the MAG-Net study is to shed light on these aspects of cardiomyopathy. In the current study, we performed bioinformatics analysis to identify the differentially expressed genes and enriched pathways that play an important role in dilated cardiomyopathies. Though further experimental validation is required, this study can serve as a primer for further validation of biomarker discovery.

After analyzing 332 patient tissue samples, the bioinformatics analysis revealed 359 downregulated and 870 upregulated proteins in DCM tissue samples compared to the NF patients. Subsequently, we conducted enrichment analysis and employed Whole Genome Co-expression network analysis to study potentially enriched pathways and build a network among them. From this study, we identified 23 unique transcription factors, which we validated using existing literature. This network was used for building the network for aim 3. Our differential expression analysis included both protein-coding genes and numerous long non-coding RNAs (lncRNAs). While lncRNAs play essential roles in post-translational gene modification and self-coding [22], our focus in this study solely centered on the cell signaling pathways of proteins, and therefore, we did not include the long non-coding RNAs in this analysis.

Among the differentially expressed genes, several genes seem very important in the context of fibrosis. The most differentially expressed gene was the Interleukin receptor type 1 gene (IL1RL1), responsible for binding and releasing IL-1α from damaged cardiomyocytes in the early stage of cardiac remodeling [24]. Downregulated in DCM patients were the Hemoglobin Subunit Alpha (HBA1 & HBA2) genes, which has a significant role in hypertension [25]. The gene SFRP4 plays a major role in the progression of myocardial ischemia [26], while Proenkephalin (PENK) along with IL1RL1 [27], is also differentially

expressed in Heart Failure patients. The MYH6 gene is strongly associated with hypertrophic cardiomyopathy according to some studies. Additionally, we identified two differentially expressed collagen genes (COL22A1 & COL10A1) [28, 29]. The COL22A1 gene encodes a protein found in the myotendinous junction of the heart, skin, and tendons, and its knockdown causes muscular dystrophy in animal models. On the other hand, COL10A1 serves as a biomarker for tumors in prostate cancer. Tubulin alpha genes (TUBA3A, TUBA3D, TUBA3E) are also upregulated in dilated cardiomyopathy patients [30]. Regarding the pathways most affected in dilated cardiomyopathy, ECM organization pathways were the most impacted, followed by immune response, biological adhesion, and matrix migration pathways. Conversely, acute phase inflammatory responses were the most upregulated responses in cardiac fibrosis.

Among the 23 derived transcription factors, some of them have unique implications in cardiac fibrosis. For instance, Bcl6 has been identified as an important transcription factor that suppresses cardiac fibroblast activation and function by directly binding to Smad4 [31]. EGR1, on the other hand, is involved in regulating cell growth, differentiation, and the response to stress and injury, making it crucial to understand its role in fibrosis [32]. Additionally, HIF1A, a master regulator of transcription, plays an important role in cellular responses to low oxygen levels [33]. Finally, ZNF281, another major player in fibrosis [34] has been discovered to be a significant transcription factor in our dataset.

3.5 Conclusion

Though this differential expression and functional enrichment analysis shed light on the crucial players of DCM, further validation is necessary to confirm them as potential biomarkers. Cell culture and animal model studies are needed to establish their role as

viable biomarkers for diagnosis or drug targets. Nevertheless, this study has identified several upregulated and downregulated genes that can serve as a starting point for further investigation. In addition, we identified important transcription factors that can be used in our aim 3 to build a GRN specific to Cardiac fibrosis. Utilizing these biomarkers also entails developing blood tests that can correlate with their expression levels, as obtaining left ventricle tissue samples poses ethical challenges.

3.6. References

1. Schaufelberger M. Cardiomyopathy and pregnancy. Heart. 2019;105(20):1543-1551. doi:10.1136/heartjnl-2018-313476

2. Precone V, Krasi G, Guerri G, et al. Cardiomyopathies. Acta Biomed. 2019;90(10-S):32-43. Published 2019 Sep 30. doi:10.23750/abm.v90i10-S.8755

3. Ba H, Zhang D, Guan S, Zheng J. Global burden of myocarditis and cardiomyopathy in children and prediction for 2035 based on the global burden of disease study 2019. Front Cardiovasc Med. 2023;10:1173015. Published 2023 May 2. doi:10.3389/fcvm.2023.1173015

4. https://www.cdc.gov/heartdisease/cardiomyopathy.htm

5. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet. 2014;383(9921):999-1008. doi:10.1016/S0140-6736(13)61752-3

6. Bellazzi R, Larizza C, Gabetta M, et al. Information technology solutions to support translational research on inherited cardiomyopathies. Stud Health Technol Inform. 2011;169:907-911.

7. www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001539.v1.p1

8. Dellefave L, McNally EM. The genetics of dilated cardiomyopathy. Curr Opin Cardiol. 2010;25(3):198-204. doi:10.1097/HCO.0b013e328337ba52

9. Eijgenraam TR, Silljé HHW, de Boer RA. Current understanding of fibrosis in genetic cardiomyopathies. Trends Cardiovasc Med. 2020;30(6):353-361. doi:10.1016/j.tcm.2019.09.003

10. Emmert-Streib F, Dehmer M, Haibe-Kains B. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. Front Cell Dev Biol. 2014;2:38. Published 2014 Aug 19. doi:10.3389/fcell.2014.00038

11. Ge SX, Son EW, Yao R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. BMC Bioinformatics. 2018;19(1):534. Published 2018 Dec 19. doi:10.1186/s12859-018-2486-6

12. Liu S, Wang Z, Zhu R, Wang F, Cheng Y, Liu Y. Three Differential Expression Analysis Methods for RNA Sequencing: limma, EdgeR, DESeq2. J Vis Exp. 2021;(175):10.3791/62528. Published 2021 Sep 18. doi:10.3791/62528

13. Leek JT. svaseq: removing batch effects and other unwanted noise from sequencing data. Nucleic Acids Res. 2014;42(21):e161. doi:10.1093/nar/gku864

14. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics. 2009;10:161. Published 2009 May 27. doi:10.1186/1471-2105-10-161

15. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. Published 2008 Dec 29. doi:10.1186/1471-2105-9-559

16. Rogers JD, Aguado BA, Watts KM, Anseth KS, Richardson WJ. Network modeling predicts personalized gene expression and drug responses in valve myofibroblasts cultured with patient sera. Proc Natl Acad Sci U S A. 2022;119(8):e2117323119. doi:10.1073/pnas.2117323119

17. Moerman T, Aibar Santos S, Bravo González-Blas C, et al. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. Bioinformatics. 2019;35(12):2159-2161. doi:10.1093/bioinformatics/bty916

18. Kumar N, Mishra B, Athar M, Mukhtar S. Inference of Gene Regulatory Network from Single-Cell Transcriptomic Data Using pySCENIC [published correction appears in Methods Mol Biol. 2021;2328:C1]. Methods Mol Biol. 2021;2328:171-182. doi:10.1007/978-1-0716-1534-8_10

19. Kang Y, Thieffry D, Cantini L. Evaluating the Reproducibility of Single-Cell Gene Regulatory Network Inference Algorithms. Front Genet. 2021;12:617282. Published 2021 Mar 22. doi:10.3389/fgene.2021.617282

20. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments

21. Cabral-Pacheco GA, Garza-Veloz I, Castruita-De la Rosa C, et al. The Roles of Matrix Metalloproteinases and Their Inhibitors in Human Diseases. Int J Mol Sci. 2020;21(24):9739. Published 2020 Dec 20. doi:10.3390/ijms21249739

22. Zhou J, Tian G, Quan Y, et al. The long noncoding RNA THBS1-AS1 promotes cardiac fibroblast activation in cardiac fibrosis by regulating TGFBR1. JCI Insight. 2023;8(6):e160745. Published 2023 Mar 22. doi:10.1172/jci.insight.160745

23. Watts K, Richardson WJ. Effects of Sex and 17 β-Estradiol on Cardiac Fibroblast Morphology and Signaling Activities In Vitro. Cells. 2021;10(10):2564. Published 2021 Sep 28. doi:10.3390/cells10102564

24. Bageghni SA, Hemmings KE, Yuldasheva NY, et al. Fibroblast-specific deletion of interleukin-1 receptor-1 reduces adverse cardiac remodeling following myocardial infarction. JCI Insight. 2019;5(17):e125074. Published 2019 Aug 8. doi:10.1172/jci.insight.125074

25. Filser M, Gardie B, Wemeau M, Aguilar-Martinez P, Giansily-Blaizot M, Girodon F. Importance of Sequencing HBA1, HBA2 and HBB Genes to Confirm the Diagnosis of High Oxygen Affinity Hemoglobin. Genes (Basel). 2022;13(1):132. Published 2022 Jan 12. doi:10.3390/genes13010132

26. Huang A, Huang Y. Role of Sfrps in cardiovascular disease. Ther Adv Chronic Dis. 2020;11:2040622320901990. Published 2020 Jan 28. doi:10.1177/2040622320901990

27. Emmens JE, Ter Maaten JM, Damman K, et al. Proenkephalin, an Opioid System Surrogate, as a Novel Comprehensive Renal Marker in Heart Failure. Circ Heart Fail. 2019;12(5):e005544. doi:10.1161/CIRCHEARTFAILURE.118.005544

28. Shorter JR, Huang W, Beak JY, et al. Quantitative trait mapping in Diversity Outbred mice identifies two genomic regions associated with heart size [published correction appears in Mamm Genome. 2018 Dec 4;:]. Mamm Genome. 2018;29(1-2):80-89. doi:10.1007/s00335-017-9730-7

29. Song C, Wei S, Fan Y, Jiang S. Bioinformatic-based Identification of Genes Associated with Aortic Valve Stenosis. Heart Surg Forum. 2022;25(1):E069-E078. Published 2022 Jan 24. doi:10.1532/hsf.4263

30. Childers RC, Sunyecz I, West TA, Cismowski MJ, Lucchesi PA, Gooch KJ. Role of the cytoskeleton in the development of a hypofibrotic cardiac fibroblast phenotype in volume overload heart failure. Am J Physiol Heart Circ Physiol. 2019;316(3):H596-H608. doi:10.1152/ajpheart.00095.2018

31. Ni J, Wu QQ, Liao HH, Fan D, Tang QZ. Bcl6 Suppresses Cardiac Fibroblast Activation and Function via Directly Binding to Smad4. Curr Med Sci. 2019;39(4):534-540. doi:10.1007/s11596-019-2070-y

32. Li G, Qin Y, Cheng Z, et al. Gpx3 and Egr1 Are Involved in Regulating the Differentiation Fate of Cardiac Fibroblasts under Pressure Overload. Oxid Med Cell Longev. 2022;2022:3235250. Published 2022 Jun 28. doi:10.1155/2022/3235250

33. Legendre C, Mooij MJ, Adams C, O'Gara F. Impaired expression of hypoxia-inducible factor-1α in cystic fibrosis airway epithelial cells - a role for HIF-1 in the pathophysiology of CF?. J Cyst Fibros. 2011;10(4):286-290. doi:10.1016/j.jcf.2011.02.005

34. Zhou H, Morales MG, Hashimoto H, et al. ZNF281 enhances cardiac reprogramming by modulating cardiac and inflammatory gene expression. Genes Dev. 2017;31(17):1770-1783. doi:10.1101/gad.305482.117

Chapter 4

Building a Composite Gene Regulatory Fibroblast Network Model

4.1 Introduction

Cardiac fibrosis is a common manifestation of many cardiovascular diseases and can be defined as the excessive accumulation of extracellular matrix by cardiac fibroblasts [1]. Cardiac fibroblasts are similar to cells that produce connective tissue but are dissimilar to bones and tendons. In the heart, they build an extracellular matrix (ECM) that is dense, irregular, and composed of collagen, proteoglycans, and glycoproteins [2]. Cardiac fibrosis is the inherent response to any injury to the heart [3]. Cardiac fibroblasts produce a response to inflammation, proliferation of nonmyocytes, and scar mutation as a first line of defense. Later this response produces excess collagens and other extracellular matrix proteins. While the primary purpose of this protective mechanism is to maintain the integrity of the heart, long-time exposure results in the loss of the heart's contractile power [4]. Therefore, it is crucial to know the mechanistic regulation of cardiac fibrosis [6].

Cardiac fibrosis event consists of different pathways that include the neurohumoral pathway [7,8] such as the RAAS system, GPCR/Adrenergic signaling pathway, Endothelin -1 pathway, Fibrogenic Growth Pathway, TGFβ signaling pathway, Platelet Derived Growth Factor, Inflammatory Pathway such as the TNFα and IL6 pathways. Several biochemical molecule species from these pathways play an essential role in cardiac fibrosis. These molecules are from different pathways and are involved in the intricate cycle of cardiac fibrosis. These pathways crosslink with each other and crosstalk through cell signaling pathways. There are several modeling approaches for cardiac fibrosis primarily

falling into two broad groups: machine learning vs.mechanistic approaches. Each of these approaches has its strengths and limitations. Machine learning models are the most convenient approach for any domain nowadays due to the availability of many frameworks. However, they are limited in capturing biological processes, such as the input-output relation in cell signaling pathways. [9, 10]

Mechanistic models have been used for decades to understand the mechanism of disease and complex biological networks. These models are primarily dynamic and can be used to understand disease networks, drug response, and complex disease network analysis. However, the recent advancement of interpretable machine learning methods also seems promising for understanding the mechanistic approach of disease modeling. Mechanistic models have another significant limitation compared to machine learning approaches. It is tough to incorporate data from different sources in mechanistic models. However, mechanistic models are most usable when we have a small dataset and a single data type. It is sometimes tough to obtain large datasets for machine learning for studies like MAGNet [11] and SMART-AV clinical trials [12]. In those cases, mechanistic models are more accurate in qualitative prediction of the model output because of their deterministic nature.

Understanding the dynamics of large network models requires a complete understanding of the network and its parameters [13, 14]. As more parameters get involved, the model becomes complex and requires more experimentally measured parameter data to validate those models. Some of these chemical reactions are very prompt, and measurement of individual reactions is not feasible. Measurement of such parameters is problematic if multiple crosstalk is involved in the network. Parameter optimization and parameter exploration are easy for small biochemical networks but difficult for more extensive

networks. For larger systems without knowledge about the parameters, logic-based systems can provide significant mechanistic insight into any model. The advantage of these models is that they can keep the network's magnitude but provide qualitative insight into the model by setting a logical relation among the network species. Reactions in which one species activates/inhibits another species, along with another species, can be defined using logical AND, OR, and NOT gates [15]. Since biological reactions are limited to a few basic types of reactions (translation, transcription, and replications), we can set basic parameters related to concentration and time and validate models only using those parameters.

The Myocardial Applied Genomic Network (MAGNet) is a large-scale clinical study to collect cardiac tissues [18]. Dilated cardiomyopathy (DCM) tissue samples were collected from patients undergoing heart transplants. On the other hand, healthy heart donor tissues are Non-Failing (NF) tissue samples collected from the left ventricle-free walls and processed for the RNAseq data for different downstream analyses. MAGNet is one of the most significant studies to compare to study the genomic landscape of the DCM.

In this chapter, we used the gene regulatory network inferred from Chapter 3 to build a composite gene regulatory network. We integrated the previously built cardiac fibrosis network into our study. We simulated the model to predict the qualitative change in the pro-fibrotic molecules. In addition, we will do the perturbation analysis to study the most critical nodes in our model. Finally, we will conduct a patient-specific simulation to identify the most relevant clinical variables in the MAGNet.

4.2 Methods

4.2.1 Building a composite gene regulatory network for NF and DCM patients

The previously published signaling network model for cardiac fibrosis relied on a manual literature search [19, 20]. They used more than 300 articles to build the cardiac fibrosis network. This network has expanded using a gene regulatory network derived from different GRN inference algorithms. The Rogers et al. study [21] presented a way to incorporate the gene regulatory network into the network. In our study, we have followed a similar approach for DCM patients. We build the gene regulatory network using the GRNBoost2 [22] algorithm by the approach used by the authors.GRNBoost2 algorithm is a regression-based method to predict the regulatory link between the input and the target gene. The famous Gene Network Inference with Ensemble of Trees (GENIE3) algorithm works in the base GRNBoost2 method. The GENIE3 algorithm performs well in all benchmarking analyses and is consistently a top performer in inferring gene regulatory interactions. In GENEI3, ensemble trees predict the expression of a given target gene from the expression of all other genes. This algorithm is better than other gene regulatory network inference algorithms. For example, it does not require understanding network topology and can infer directed interactions by comparing the correlation and probability-based methods. Moreover, this algorithm can infer nonlinear regulations as well. For our study, we used a different approach to pruning the inferred GRN. The initially inferred network has thousands of non-specific interactions unrelated to fibrosis. Multi-step network pruning removes these non-specific interactions. First, we only allowed the literature-supported transcription factor target interactions. We used the Chromatin Immune Precipitation (CHIP-X) and Transcription Factor (TRANSFAC) database to cover the maximum number of transcription factors with experimental evidence. Upon filtering for the transcription

factors, we again filtered out the network that is only related to the output of the fibroblast mechanotransduction network. We also added genes (MYH6, COL22A1, COL10A1) significantly different between nonfailing vs. dilated cardiomyopathy tissue samples. Finally, we used a depth-first search (DFS) algorithm to prune the network edges. This algorithm helps us to identify possible pathways between each primary transcription factor and target and check whether each interaction meets the required importance score. Only highly ranked interactions are allowed here. Interactions are only allowed if the importance score of each edge is higher than the 90th percentile. These filtering steps help to reduce the noises by non-specific interactions. Since our dataset has a lot of replicates, we, trained the network model by dividing the dataset into ten folds. We only selected the edges derived from network inference , pruning, and, present in at least 8/10 fold of data. Finally, we filtered the gene regulatory network using literature/articles that validate the presence of those TF in at least one article.

4.2.2 Building the Logic Based ODE model.

We integrated this transcription factor and target network into the cardiac fibroblast mechanotransduction network, a logic-based ODE model. This model is logic-based because the approach allows logical interaction between the biochemical species/model components. For any cell signaling species x, the activation and inhibition can be modeled as a Normalized Hill Equation [23] as –

$$fact(x) = \frac{Bx^n}{k^n + x^n}$$

$$finhib(x) = 1 - \frac{Bx^n}{k^n + x^n}$$

Here n is the Hill coefficient related to the gradient of the activation or inhibition dose-response. The constant B and K are additional constants characterizing the dose-response curves and can be expressed as -

$$B = \frac{EC_{50}^{n} - 1}{2EC_{50}^{n} - 1}$$

$$k = (B - 1)^{\frac{1}{n}}$$

Where EC_{50} is the value of x when half maximum activation occurs. The primary logics here are 'AND', 'OR', 'ANDNOT' involved in two types of reaction (activation/inhibition). If two biochemical species involved in an activation/inhibition reaction, the equations can be written as -

$$xANDy = fact(x) * fact(y)$$

$$xORy = fact(x) + fact(y) - fact(x) * fact(y)$$

$$xANDNOTy = fact(x)(1 - fact(y))$$

Complex logical reactions can be used for complex reactions involving 2 or 3 species. As we can see these reactions are only dependent on the reactants not the product. In addition to these n and EC_{50}, which are specific to any specific biochemical reactions, another parameter Reaction Weight (w) can be used to better for the quantitative experiments. Besides these three reaction parameters, there are three different species/node specific parameters. Each node has a decay timescale tau (τ) and a maximal activity level, ymax $\in$ [0,1]. For any reaction an ymax value of 1 is used by default. We can use a different value if we want to lower the maximum activation level of any node (between 0 and 1) or knockdown of any node (ymax = 0). In our model we have set specific value for these

parameters. We set the initial activation level y0 = 0 for all species, τ is set to 1, 0.1, 10 based on the type of reaction. For translation it is set to 0.1, 1, 10 based on the type of translation reactions. On the other hand, τ for all GRN transcription reactions are set to 1. We used normalized RNASeq data between 0 and 1, and set that as ymax values. For our model we used ymax value for all the input reactions as w, all internal and output reaction weights are set to 1, n is set to 1.4, EC_{50} set to 0.6. These logics are implemented into a dynamical model where change of each species is integrated into the differential equations presented in the supplemental table S4.1 & S4.2. The implementation of these models is done in MATLAB package Netflux [23]. The model was created using Cytoscape and PowerPoint.

4.2.3 Model Validation

In all previous studies, the network validation steps consist of literature validation. However, for this current study, we have the RNASeq dataset to compare the input-output relation in the dataset. We normalized the RNAseq data to compare the NF and DCM tissue samples. To do that validation, we have normalized the SVA samples. We exclude the outliers in each gene. To do that, we excluded the data that spread between 10-90% of the data. We have excluded the dataset above these spreads and replaced them with the top 90 or 10% percentile data points. Then each gene expression data was normalized between 0-1. Since we removed the data's outliers, we included a change in gene expressions at a minimum of 1%. Since this is our logical model, we compared the qualitative outcome of the model rather than the numerical aspect of the model. We compare the model output in response to the input values as RNASeq. The output of the model compared with the bulk RNAseq data. These differences have three categories increase, decrease, and no change. Instead of predicting like machine learning algorithms, this model

represents the mechanical aspects of cardiac fibrosis. The matching and mismatching model predictions are presented as a matrix in Figure 4.2. All the simulation data were generated in MATLAB, and all the graphs were prepared using the Python package Matplotlib.

4.2.4 Network Perturbation Analysis

Network perturbation analysis was conducted to identify the influential signaling node in the DCM tissue samples. First, the model was simulated using the baseline parameter values. To do that we first set the ymax for each node from normalized RNAseq data then set the ymax of all the nodes to 10% of the set ymax. We then also set all the input reaction weights (w) equivalent to the baseline ymax value. We ran the simulation using this baseline values and saved as. Then, we simulated the model using the ymax and input weight value derived from RNASeq. We recorded the change of value as Δ(Activity). The changes are plotted in a heatmap in Figure 4.3.

4.2.5 Correlation between Model Predicted output clinical variables.

We also used the patient's clinical and demographic variables, such as age, weight, height, heart weight, left ventricular mass, and left ventricular ejection fraction (LVEF), to find a correlation with the model-predicted outputs. To do that, we first simulated the model for all patients. We measured the output of the model. We then filtered the data that has all clinical variables present. We then used Pearson correlation coefficient for this. This correlation matrix is presented in Figure 4.4.

4.3. Results

4.3.1 Integration of gene regulatory network to the fibroblast signaling network

After inferring the gene regulatory network using the GRNBoost2 algorithm, we found around 1300 unique edges. These edges are further filtered for the nodes that are connected to either the list of transcription factors or connected to the output of the mechanistic pathway of the cardiac fibrosis model. All these final edges were again validated with the literature supporting the interaction. Moreover, we had to add connection reactions to the intermediates to the network. This results in total addition of 142 more edges.

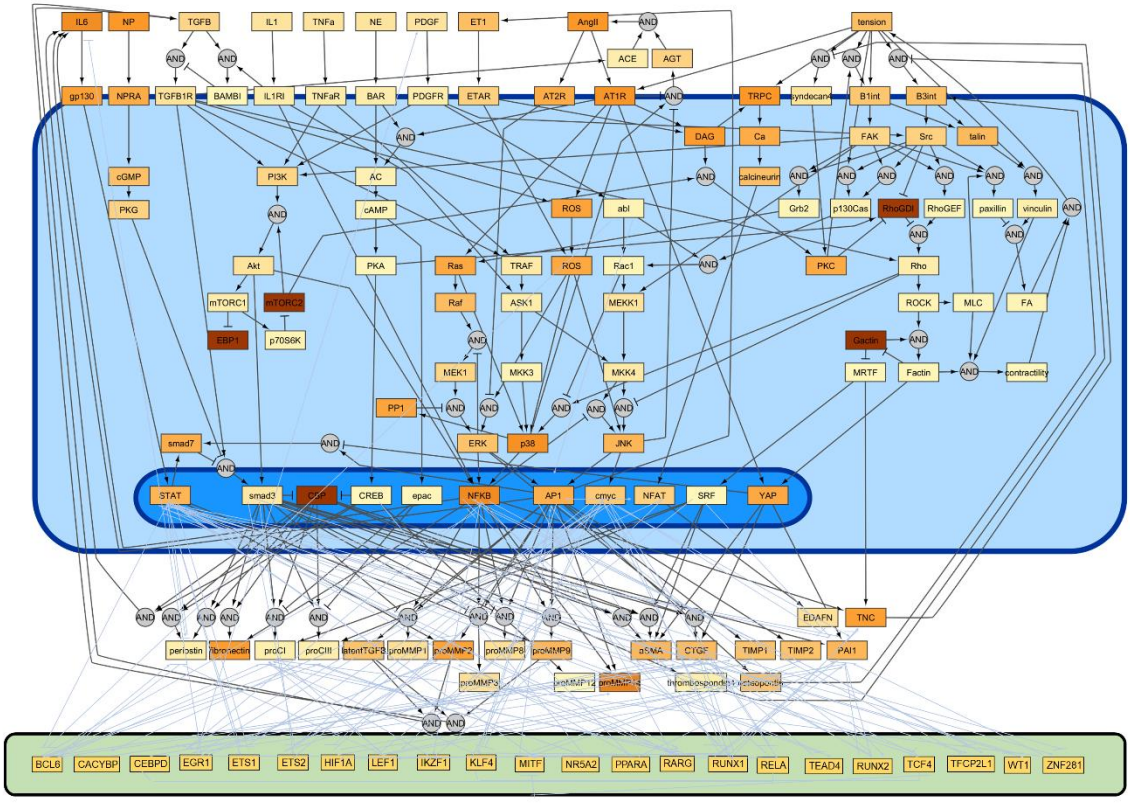


Figure 4.1. The combined gene regulatory network. The model consists of two parts. The

top one is adopted from the previous fibroblast mechanotransduction network. The bottom green part is inferred from the MAGNet gene regulatory network.

4.3.2 Model Accuracy after integration of the gene regulatory network

To validate whether the model still shows the same performance as the original fibroblast mechanotransduction model [20]. We have added 142 more edges and made the model show a similar level of accuracy. The final accuracy of the model was relatively the same after adding those edges to the model (Figure 4.2). The simulations ran for only 100 sec to validate the expression change in the input, intermediates, and output of the model.

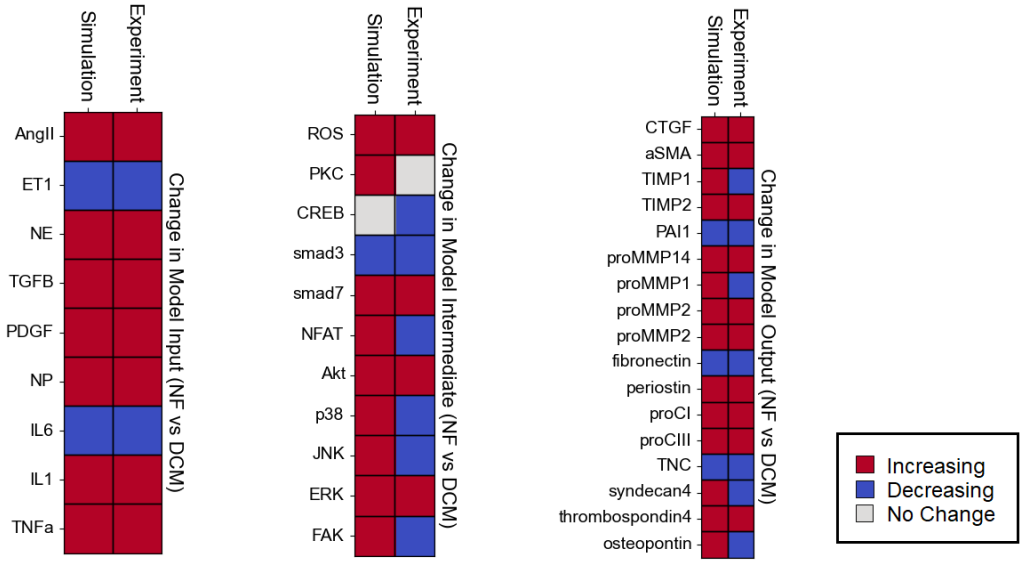


Figure 4.2. Validation of the composite model. Addition of the Gene Regulatory Network did not change the accuracy of the model compared to the previously reported models. The model showed better performance in predicting the qualitative changes in Non Failing to Dilated Cardiomyopathy tissue samples.

4.3.3 Perturbation analysis identifies important network drivers

Perturbation analysis showed the most important nodes in our model. Tension is the most important factor here. Its knockdown showed the single most effect in the network. The second most effected node is the activator protein (AP1) which regulates gene expression in the presence of various stimuli, such as stress/mechanical tension, cytokines, growth factors, and infections [24]. The third most perturbed node is the Nuclear factor kappa-light-chain-enhancer of activated B cells (NFKB) is the third most perturbed node. During cardiac remodeling, NFKB exerts cytotoxic effect by prolonging inflammatory response, which is a hallmark of cardiac fibrosis. Early growth response 1 (EGR1) is the only transcription factor that can be included in the top perturbed node [25, 26]. This transcription factor is activated with the onset of cardiac hypertrophy in the left ventricle of the heart. Among the receptors, Angiotensin II type 1 receptor (AT1R) is the most perturbed. Angiotensin II activates mechanotransduction through the receptor AT1R, increases the intracellular Ca^{2+} concentration and promotes hypertrophic cardiac remodeling [27]. Nuclear Factor of Activated T-cells (NFAT) is also perturbed to the highest degree [28].

4.3.4 Correlation among clinical variables and model outputs

Correlation analysis showed the relationship between cardiac function variables Left Ventricular Mass (LV_Mass), Heart Weight, and Left Ventricular Ejection Fraction (LVEF). There is no strong correlation found in this correlation analysis except for the Tissue Inhibitors for Metalloprotease 2 (TIMP2).

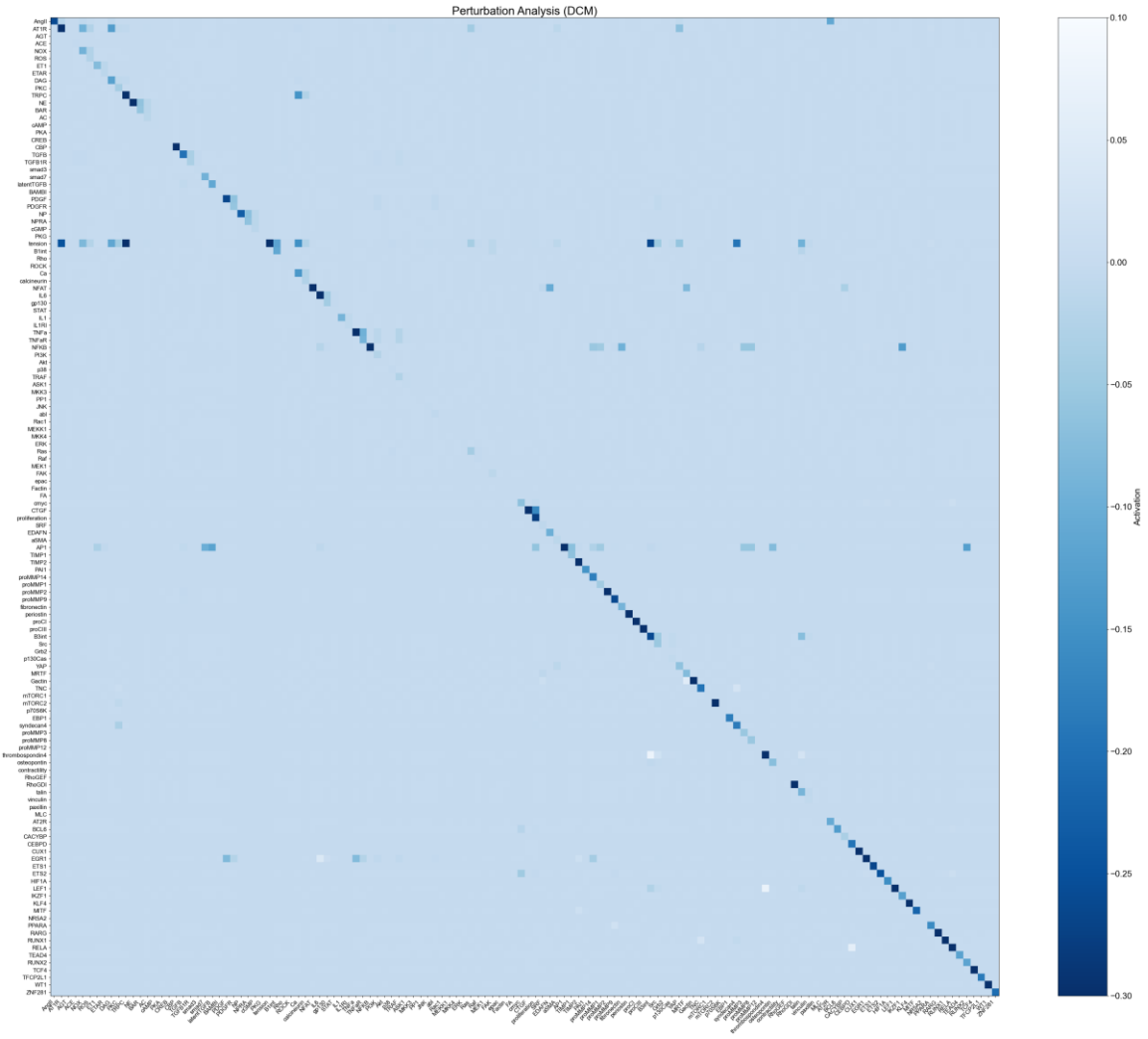


Figure 4.3. Perturbation analysis showed the most important nodes in the network. We simulated the model with baseline ymax values (0.2) for each of the nodes.

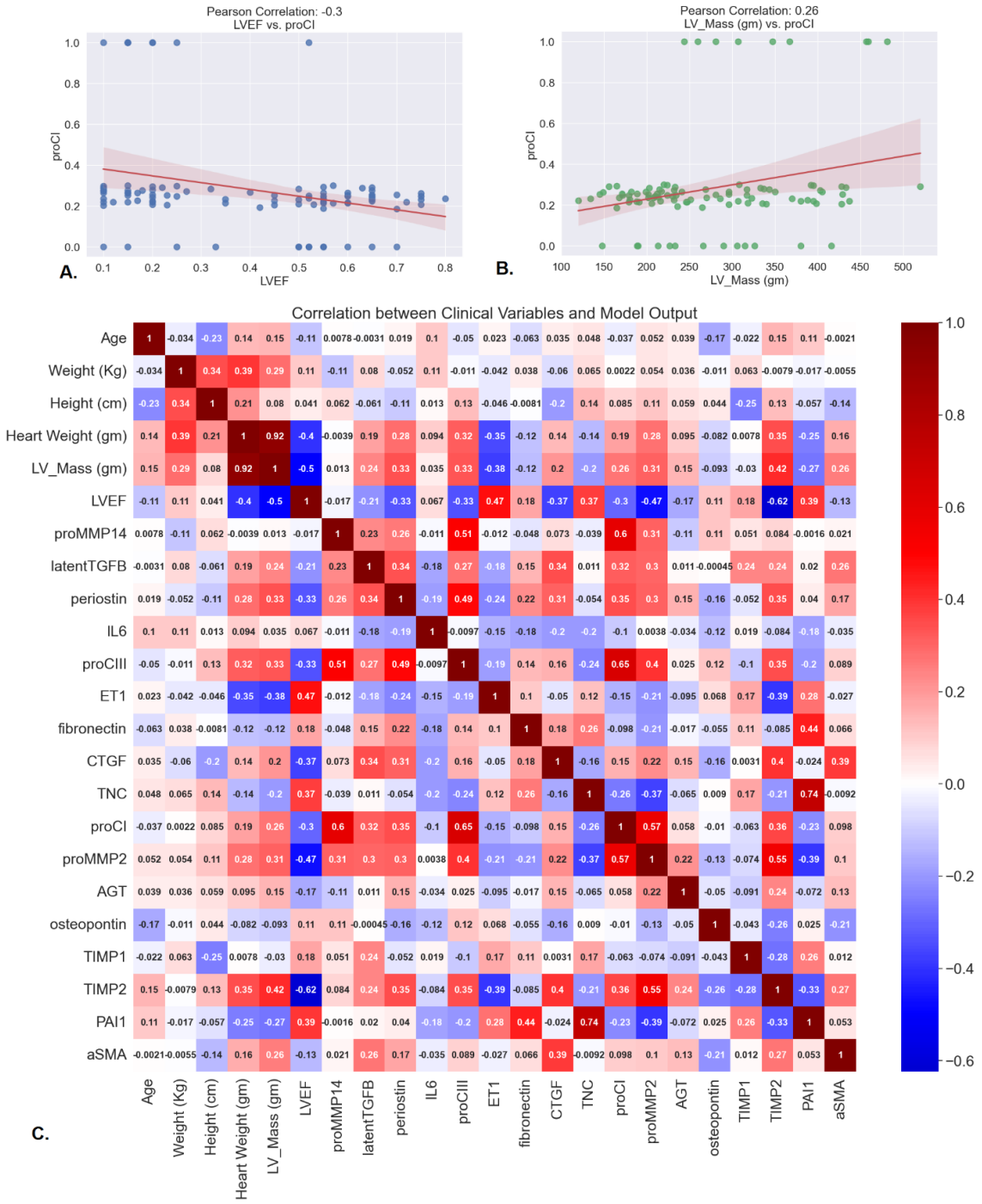


Figure 4.5. Pearson correlation among the model predicted output and clinical variables. We ran the simulation for each patients who has complete clinical variables. A. Correlation between simulated proCI and Left Ventricular Ejection Fraction (LVEF). B. Correlation

between simulated proCI and Left Ventricular Mass. C. Correlation among the outputs and numerical measurements. All of these simulated data have some outliers values due to reaching into steady state or not expressing the fibrotic outputs.

4.4 Discussion

Cardiac fibrosis is a complex disease that involves crosslinking multiple pathways from cell development and growth to immune responses. Developing a proper therapeutic target requires a complete understanding of the complex network. Gene regulatory networks, though insufficient to elucidate the intracellular dynamics, play an essential role in how cellular effectors interact. Therefore disease-specific gene regulatory networks are essential to infer and integrate into the mechanistic model of cardiac fibrosis. With the availability and ease of machine learning algorithms, modelers are more prone to machine learning models. However, with a limited amount of patient-specific data and a large number of variables, it is more rational to use mechanistic models instead of machine learning to understand diseases. Mechanistic models are more accurate and applicable to patient-specific modeling [29].

In this chapter, we integrated fibrosis related gene regulatory networks into our model. Upon comparing them with the experimental data, we found the most similarities in output nodes, followed by the input and intermediate nodes. The accuracy in predicting the qualitative change in output is very encouraging (77%) which is almost similar to the original fibroblast models.This is interesting since we did not fit it to experimental data, rather used them as input. This also validate that gene expression does not equate protein expression in intracellular signalling models. Therefore, further proteomics study needs to validate the qualitative output change of the model [30].

The perturbation study analysis uncovered tension as the most influential node in the DCM patient. Tension is the mechanical stimulant for left ventricular hypertrophy. The heart must work hard to counteract tension and maintain its pumping capability. This results in the thickening of the heart , which is responsible for pumping oxygen-rich blood to the organs. Therefore, tension is the most critical node in our model. In cardiac diseases, AP1, NFKB, and EGR1 are important intermediates crosslinking different pathways, resulting in left ventricular hypertrophy development. Due to mechanical stress, several kinase pathways get activated and activate NFKB, AP1, and EGR1. Activation of these crosslinking molecules, a cascade of transcriptional programming starts and induce the transcription of Platelet-Derived Growth Factors (PDGFs), Transforming Growth Factors (TGFs), tissue factors, matrix metalloproteinases (MMPs), and collagens (such as COL1, COL3) [25-27,31]. These start a cascade of physiological changes such as atherosclerosis, angiogenesis, ischemic disease, and cardiac hypertrophy. All of these are potential triggers for cardiac remodeling. The Angiotensin II receptor ATR1 plays a vital role in cardiac hypertrophy. Its activation results in the activation of the TGFβ1 pathway through its receptor. The activation of TGFβ1 activates Extracellular Signal-Regulated Kinase 1 and 2 (ERK1 and 2) pathways. Phosphorylated ERK1/ERK2 stimulate overexpression of growth factors and neurohormonal mediators. All these lead to left ventricular hypertrophy [32].

Though we did patient-specific simulations to show the relation between clinical variables and model outputs, we did not find a strong correlation. This is somewhat expected because of the lack of different data types. In our study, we have used RNASeq data as model input as well as to validate the model. But the correlation between mRNA/RNASeq and protein is not straightly correlated. This is due to the different half-life

for different transcript and protein, post transcriptional modifications of the mRNA transcripts [30]. Therefore, we need a proteomic data set to fit and train the model for more accurate qualitative prediction of the change in model intermediates and outputs. Another factor that is strongly contributing to this low correlation is the outlier data derived from the simulation. This is because we have a small panel of complete patient data. After imputing, we ended up getting only 99/332 patients who has complete set of clinical data. Extensive studies should be conducted on such relations to find relations among pro-fibrotic outputs and cardiac function variables.

4.5 Conclusion

In this chapter we integrated the gene regulatory network derived from the previous chapter into previously built fibroblast mechanotransduction network. These combined mechanistic models showed very similar qualitative prediction of the mediators of fibrosis, especially in terms of predicting the qualitative change in input. Further analysis of the clinical variables has shown how they correlate with the model outputs. Due to ethical issues, it is very hard to obtain real tissue samples from actively pumping hearts left ventricles. MAGNet study has given us the opportunity to compare the dynamics of transcriptomics between the Non Failing donor heart and Dilated Cardiomyopathy patients. We leveraged this opportunity to build a logic based model. Future direction of this study will be building a most robust model by integrating an efficient parameter estimation method and adding a drug perturbation study.

4.6 References

1. Jiang W, Xiong Y, Li X, Yang Y. Cardiac Fibrosis: Cellular Effectors, Molecular Pathways, and Exosomal Roles. Front Cardiovasc Med. 2021;8:715258. Published 2021 Aug 16. doi:10.3389/fcvm.2021.715258

2. Ivey MJ, Tallquist MD. Defining the Cardiac Fibroblast. Circ J. 2016;80(11):2269-2276. doi:10.1253/circj.CJ-16-1003

3. Kania G, Blyszczuk P, Eriksson U. Mechanisms of cardiac fibrosis in inflammatory heart disease. Trends Cardiovasc Med. 2009;19(8):247-252. doi:10.1016/j.tcm.2010.02.005

4. Burchfield JS, Xie M, Hill JA. Pathological ventricular remodeling: mechanisms: part 1 of 2. Circulation. 2013;128(4):388-400. doi:10.1161/CIRCULATIONAHA.113.001878

5. Travers JG, Tharp CA, Rubino M, McKinsey TA. Therapeutic targets for cardiac fibrosis: from old school to next-gen. J Clin Invest. 2022;132(5):e148554. doi:10.1172/JCI148554

6. Ferrario CM, Mullick AE. Renin angiotensin aldosterone inhibition in the treatment of cardiovascular disease. Pharmacol Res. 2017;125(Pt A):57-71. doi:10.1016/j.phrs.2017.05.020

7. He X, Du T, Long T, Liao X, Dong Y, Huang ZP. Signaling cascades in the failing heart and emerging therapeutic strategies. Signal Transduct Target Ther. 2022;7(1):134. Published 2022 Apr 23. doi:10.1038/s41392-022-00972-6

8. Kong P, Christia P, Frangogiannis NG. The pathogenesis of cardiac fibrosis. Cell Mol Life Sci. 2014;71(4):549-574. doi:10.1007/s00018-013-1349-6

9. Baker RE, Peña JM, Jayamohan J, Jérusalem A. Mechanistic models versus machine learning, a fight worth fighting for the biological community?. Biol Lett. 2018;14(5):20170660. doi:10.1098/rsbl.2017.0660

10. Procopio A, Cesarelli G, Donisi L, Merola A, Amato F, Cosentino C. Combined mechanistic modeling and machine-learning approaches in systems biology - A systematic literature review [published online ahead of print, 2023 Jun 17]. Comput Methods Programs Biomed. 2023;240:107681. doi:10.1016/j.cmpb.2023.107681

11. Arking DE, Pulit SL, Crotti L, et al. Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. Nat Genet. 2014;46(8):826-836. doi:10.1038/ng.3014

12. Spinale FG, Meyer TE, Stolen CM, et al. Development of a biomarker panel to predict cardiac resynchronization therapy response: Results from the SMART-AV trial. Heart Rhythm. 2019;16(5):743-753. doi:10.1016/j.hrthm.2018.11.026

13. Prasse B, Van Mieghem P. Predicting network dynamics without requiring the knowledge of the interaction graph. Proc Natl Acad Sci U S A. 2022;119(44):e2205517119. doi:10.1073/pnas.2205517119

14. Schwab JD, Kühlwein SD, Ikonomi N, Kühl M, Kestler HA. Concepts in Boolean network modeling: What do they all mean?. Comput Struct Biotechnol J. 2020;18:571-582. Published 2020 Mar 10. doi:10.1016/j.csbj.2020.03.001

15. Irons L, Humphrey JD. Cell signaling model for arterial mechanobiology. PLoS Comput Biol. 2020;16(8):e1008161. Published 2020 Aug 24. doi:10.1371/journal.pcbi.1008161

16. Berman MN, Tupper C, Bhardwaj A. Physiology, Left Ventricular Function. [Updated 2022 Sep 19]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK541098/

17. Reichart D, Magnussen C, Zeller T, Blankenberg S. Dilated cardiomyopathy: from epidemiologic to genetic phenotypes: A translational review of current literature. J Intern Med. 2019;286(4):362-372. doi:10.1111/joim.12944

18. Lin H, Dolmatova EV, Morley MP, et al. Gene expression and genetic variation in human atria. Heart Rhythm. 2014;11(2):266-271. doi:10.1016/j.hrthm.2013.10.051

19. Zeigler AC, Richardson WJ, Holmes JW, Saucerman JJ. Computational modeling of cardiac fibroblasts and fibrosis. J Mol Cell Cardiol. 2016;93:73-83. doi:10.1016/j.yjmcc.2015.11.020

20. Rogers JD, Richardson WJ. Fibroblast mechanotransduction network predicts targets for mechano-adaptive infarct therapies. Elife. 2022;11:e62856. Published 2022 Feb 9. doi:10.7554/eLife.62856

21. Rogers JD, Aguado BA, Watts KM, Anseth KS, Richardson WJ. Network modeling predicts personalized gene expression and drug responses in valve myofibroblasts cultured with patient sera. Proc Natl Acad Sci U S A. 2022;119(8):e2117323119. doi:10.1073/pnas.2117323119

22. Moerman T, Aibar Santos S, Bravo González-Blas C, et al. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. Bioinformatics. 2019;35(12):2159-2161. doi:10.1093/bioinformatics/bty916

23. Kraeutler MJ, Soltis AR, Saucerman JJ. Modeling cardiac β-adrenergic signaling with normalized-Hill differential equations: comparison with a biochemical model. BMC Syst Biol. 2010;4:157. Published 2010 Nov 18. doi:10.1186/1752-0509-4-157

24. Windak R, Müller J, Felley A, et al. The AP-1 transcription factor c-Jun prevents stress-imposed maladaptive remodeling of the heart. PLoS One. 2013;8(9):e73294. Published 2013 Sep 10. doi:10.1371/journal.pone.0073294

25. Gaspar-Pereira S, Fullard N, Townsend PA, et al. The NF-κB subunit c-Rel stimulates cardiac hypertrophy and fibrosis. Am J Pathol. 2012;180(3):929-939. doi:10.1016/j.ajpath.2011.11.007

26. Ramadas N, Rajaraman B, Kuppuswamy AA, Vedantham S. Early growth response-1 (EGR-1) - a key player in myocardial cell injury. Cardiovasc Hematol Agents Med Chem. 2014;12(2):66-71. doi:10.2174/1871525713666150123152131

27. Schnee JM, Hsueh WA. Angiotensin II, adhesion, and cardiac fibrosis. Cardiovasc Res. 2000;46(2):264-268. doi:10.1016/s0008-6363(00)00044-4

28. Wilkins BJ, Dai YS, Bueno OF, et al. Calcineurin/NFAT coupling participates in pathological, but not physiological, cardiac hypertrophy. Circ Res. 2004;94(1):110-118. doi:10.1161/01.RES.0000109415.17511.18

29. Gherman IM, Abdallah ZS, Pang W, Gorochowski TE, Grierson CS, Marucci L. Bridging the gap between mechanistic biological models and machine learning surrogates. PLoS Comput Biol. 2023;19(4):e1010988. Published 2023 Apr 20. doi:10.1371/journal.pcbi.1010988

30. Haider S, Pal R. Integrated analysis of transcriptomic and proteomic data. Curr Genomics. 2013;14(2):91-110. doi:10.2174/1389202911314020003

31. Khachigian LM. Early growth response-1 in cardiovascular pathobiology. Circ Res. 2006;98(2):186-191. doi:10.1161/01.RES.0000200177.53882.c3

32. Duangrat R, Parichatikanond W, Morales NP, Pinthong D, Mangmool S. Sustained AT1R stimulation induces upregulation of growth factors in human cardiac fibroblasts via Gαq/TGF-β/ERK signaling that influences myocyte hypertrophy. Eur J Pharmacol. 2022;937:175384. doi:10.1016/j.ejphar.2022.175384

Chapter 5

Conclusion, Limitation, and Future Direction

5.1 Summary of Findings

In this dissertation, we utilized interpretable mechanistic and machine learning models to predict cardiac remodeling based on biochemical and biomechanical features. Aim 1 demonstrated the effectiveness of integrating multiple data types, such as demographics, comorbidities, therapy history, circulating biomarker levels, and echo-based left ventricular function data. This integration significantly improved the predictive capability of our machine learning algorithms in identifying responders and non-responders to cardiac resynchronization therapy (CRT). By leveraging this diverse set of features, we achieved an identification rate of 71% for patient response, with an area under the curve (AUC) of 0.784. In Aim 2, we focused on identifying differentially expressed genes upregulated in the non-failing (NF) versus dilated cardiomyopathy (DCM) left ventricle tissue samples obtained from donors and recipients. We employed a gene regulatory network inference algorithm called GRNBoost2 to achieve this and incorporated these identified genes into the analysis. We identified 23 unique transcription factors connected to 17 profibrotic biochemical species and other cellular intermediates through a novel multistep filtering algorithm. These connections represent 158 activation/inhibition reactions (edges). Aim 3 integrated the 158 GRN edges into an existing logic-based ordinary differential equation (ODE) model to predict the qualitative changes in model input, intermediate, and output components. Our model demonstrated comparable predictive capability to the

previously established fibroblast mechanistic network. Furthermore, we conducted patient-specific simulations to illustrate the relevance of clinical variables in our findings.

Overall, our research contributes to a better understanding of cardiac remodeling prediction by applying both interpretable mechanistic models and machine learning algorithms. By combining various data types and integrating gene regulatory networks, we have gained valuable insights into identifying responders to CRT and analyzing key transcriptional factors involved in cardiac remodeling. These findings have significant implications for advancing personalized cardiac disease treatment and management approaches.

5.2 Study limitations

Our aims in this study have certain limitations stemming from the nature of the research and the data availability. Firstly, the SMART-AV clinical trial data used in Aim 1 had a relatively short follow-up period of only six months for the patients. This limited timeframe prevented us from conducting a comprehensive survival analysis of the patients. It would have been valuable to have access to long-term follow-up studies [1] as well, as they could have aided in identifying patient-specific variables that contribute to long-term survival analysis. In Aim 2, we employed the GRNBoost2 algorithm to identify crucial interactions between transcription factors and biochemical species. However, to further refine the gene regulatory network, conducting a perturbation analysis to identify upstream regulatory signaling pathways from downstream gene expression (GEX) would have been beneficial [2].

Additionally, exploring patient-specific gene regulatory networks could have provided more profound insights into individual variations. For the final aim, we utilized the GRN network obtained from Aim 2 and used RNASeq data to fit both the input and output of the mechanistic model. Since our model comprises multiple parameters (y0, ymax, tau, w, n, and EC50), employing a parameter optimization method could have significantly enhanced the accuracy of model predictions. Some studies have successfully employed parameter optimization methods, such as nonlinear programming [3], genetic algorithms [4], and uncertainty quantification [5], to refine their models. Furthermore, it would have been valuable to modify the original fibroblast mechanistic model by adding or removing nodes not pertinent to dilated cardiomyopathy. This refinement could have resulted in more patient-specific and disease-specific mechanistic models.

In conclusion, while our research has contributed valuable insights into cardiac remodeling, it is essential to acknowledge these limitations. The short follow-up period of the clinical trial data, the need for further analyses in the gene regulatory network, and the potential for parameter optimization in the final model could be addressed in future studies to enhance our findings' robustness and applicability.

5.3 Future Direction

The limitations encountered in our aims have provided valuable opportunities for future research. In the first aim, further biochemical exploration of CRT patients can complement our machine learning models. By incorporating mechanistic models alongside machine learning, we could gain deeper insights into the input-output relationships through an extended study. Such an exploration would offer mechanistic insights into the effects of CRT treatment on patients, such as changes in pressure and volume in the left ventricle.

Understanding these aspects could enhance our comprehension of CRT treatment outcomes. We can extend the study for the second and third aims by incorporating additional data types beyond RNASeq data. Integrating proteomic studies would add another layer of information and potentially lead to more comprehensive findings. Moreover, implementing an efficient parameter estimation method would refine our models and improve the accuracy of predictions.

Expanding the scope of this study to include different types of cardiomyopathies, such as Dilated, Hypertrophic, and Peripartum, in patients would be worthwhile. Before extending to such studies, constructing a mechanistic network specific to cardiomyopathies would be paramount. The overarching goal of this study is to build interpretable models for cardiac remodeling using either machine learning or mechanistic modeling, both of which have their strengths and limitations. By combining these approaches, we have the potential to uncover new insights and enrich our current knowledge base. This integration may lead to more accurate and comprehensive models, ultimately advancing our understanding of cardiac remodeling and its associated complexities.

In conclusion, the limitations encountered in our study present exciting possibilities for future research. By exploring additional biochemical factors, employing mechanistic modeling alongside machine learning, integrating diverse data types, and extending the study to include different cardiomyopathies, we can advance our understanding of cardiac remodeling and pave the way for improved therapeutic approaches in cardiac mechanobiology.

5.4 References

1. Hadwiger M, Dagres N, Haug J, et al. Survival of patients undergoing cardiac resynchronization therapy with or without defibrillator: the RESET-CRT project. Eur Heart J. 2022;43(27):2591-2599. doi:10.1093/eurheartj/ehac053

2. Liu A, Trairatphisan P, Gjerga E, Didangelos A, Barratt J, Saez-Rodriguez J. From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL. NPJ Syst Biol Appl. 2019;5:40. Published 2019 Nov 11. doi:10.1038/s41540-019-0118-z

3. Khalilimeybodi A, Paap AM, Christiansen SLM, Saucerman JJ. Context-specific network modeling identifies new crosstalk in β-adrenergic cardiac hypertrophy. PLoS Comput Biol. 2020;16(12):e1008490. Published 2020 Dec 18. doi:10.1371/journal.pcbi.1008490

4. Rogers JD, Aguado BA, Watts KM, Anseth KS, Richardson WJ. Network modeling predicts personalized gene expression and drug responses in valve myofibroblasts cultured with patient sera. Proc Natl Acad Sci U S A. 2022;119(8):e2117323119. doi:10.1073/pnas.2117323119

5. Wang A, Cao S, Aboelkassem Y, Valdez-Jasso D. Quantification of uncertainty in a new network model of pulmonary arterial adventitial fibroblast pro-fibrotic signalling. Philos Trans A Math Phys Eng Sci. 2020;378(2173):20190338. doi:10.1098/rsta.2019.0338

Chapter 6

Supplementary Materials

6.1 Supplementary Tables

Table S4.1 Biochemical reactions and their parameter values used in Chapter 4.

Reaction Information							
module	ID	Rule	Weight	n	EC50	source	notes
input	i1	=> AngII	0.45	1.4	0.6	neonatal rat cardiac fibroblasts	increased via RAS in hypertension and heart failure
input	i2	=> TGFB	0.37	1.4	0.6		increased in response to injury
input	i3	=> tension	0.10	1.4	0.6		increased with integrin stimulation
input	i4	=> IL6	0.42	1.4	0.6		increased in hypertension
input	i5	=> IL1	0.27	1.4	0.6		
input	i6	=> TNFa	0.25	1.4	0.6		
input	i7	=> NE	0.25	1.4	0.6		most likely NE signaling
input	i8	=> PDGF	0.36	1.4	0.6		increased post-MI
input	i9	=> ET1	0.43	1.4	0.6		increased from stretch of vascular endothelial cells
input	i10	=> NP	0.29	1.4	0.6		increased in pressure
fback	r1	proMMP9 & latentTGFB => TGFB	1.00	1.4	0.6	in vitro	release of latent protein
fback	r2	proMMP2 & latentTGFB => TGFB	1.00	1.4	0.6	in vitro	release of latent protein
fback	r3	ACE & AGT => AngII	1.00	1.4	0.6	neonatal cardiac fibroblasts	enzymatic modification
output/fback	r4	CREB & CBP => IL6	1.00	1.4	0.6	neonatal rat cardiac fibroblasts	txn
output/fback	r5	NFKB => IL6	1.00	1.4	0.6	neonatal rat cardiac fibroblasts	txn
output/fback	r6	AP1 => IL6	1.00	1.4	0.6	neonatal cardiac fibroblasts	txn
output/fback	r7	AP1 => ET1	1.00	1.4	0.6	neonatal rat cardiac fibroblasts	txn
middle	r8	AngII => AT1R	1.00	1.4	0.6	neonatal cardiac fibroblasts	receptor binding

middle	r9	AT1R => NOX	1.00	1.4	0.6	adult rat cardiac fibroblast	-
middle	r10	NOX => ROS	1.00	1.4	0.6	adult rat cardiac fibroblast	enzymatic production
middle	r11	IL6 => gp130	1.00	1.4	0.6	neonatal rat cardiac fibroblasts	receptor binding
middle	r12	ROS => p38	1.00	1.4	0.6	neonatal rat cardiac fibroblasts	activation
middle	r13	ROS => JNK	1.00	1.4	0.6	neonatal rat cardiac fibroblasts	activation
middle	r14	IL1RI => NFKB	1.00	1.4	0.6	neonatal rat cardiac fibroblasts	release of blocking and increased abundance
middle	r15	gp130 => STAT	1.00	1.4	0.6	neonatal mouse fibroblasts	activation (via JAK)
middle	r16	TNFaR => PI3K	1.00	1.4	0.6	human cardiac fibroblasts	activation
middle	r17	!AT1R & !JNK & p38 => AGT	1.00	1.4	0.6	neonatal rat cardiac fibroblasts	txn
middle	r18	TGFB1R & !PKG & !smad7 => smad3	1.00	1.4	0.6	adult rat cardiac fibroblast	activation
output	r19	smad3 & CBP & ERK => CTGF	1.00	1.4	0.6	neonatal rat cardiac fibroblasts	txn
output	r20	STAT => proMMP2	1.00	1.4	0.6	neonatal rat cardiac fibroblasts	txn
output	r21	STAT => proMMP9	1.00	1.4	0.6	mouse cardiac fibroblasts	txn
output	r22	smad3 & CBP => periostin	1.00	1.4	0.6	adult rat cardiac fibroblasts	txn
output	r23	CREB & CBP => periostin	1.00	1.4	0.6	adult rat cardiac fibroblasts	txn
middle	r24	ERK => NFKB	1.00	1.4	0.6	human cardiac fibroblast	activation
middle	r25	p38 => NFKB	1.00	1.4	0.6	human cardiac fibroblast	activation
output	r26	NFKB & AP1 & !smad3 => proMMP1	1.00	1.4	0.6	human cardiac fibroblast	txn
middle	r27	ETAR => ROS	1.00	1.4	0.6	neonatal rat cardiac fibroblasts	activation
middle	r28	ERK => AP1	1.00	1.4	0.6	neonatal rat cardiac fibroblasts	activation

output	r29	AP1 => proMMP2	1.00	1.4	0.6	human cardiac fibroblasts	txn
output	r30	AP1 & NFKB => proMMP9	1.00	1.4	0.6	human cardiac fibroblasts	txn
output	r31	AP1 => TIMP1	1.00	1.4	0.6	human cardiac fibroblasts	txn
output	r32	AP1 => TIMP2	1.00	1.4	0.6	human cardiac fibroblast	txn
middle	r33	PKC & tension => B1int	1.00	1.4	0.6	adult rat cardiac fibroblasts	activation
middle	r34	cAMP => PKA	1.00	1.4	0.6	adult rat cardiac fibroblasts	activation
output	r35	smad3 & CBP => fibronectin	1.00	1.4	0.6	human lung fibroblast	txn
middle	r36	!smad3 => CBP	1.00	1.4	0.6	adult rat cardiac fibroblasts	depletion of txn factor binding partner
middle	r37	!CREB => CBP	1.00	1.4	0.6	adult rat cardiac fibroblasts	depletion of txn factor binding partner
middle	r38	tension => B1int	1.00	1.4	0.6	neonatal rat cardiac fibroblasts	activation
output	r39	NFAT => EDAFN	1.00	1.4	0.6	neonatal mice cardiac fibroblast	txn activation
middle	r40	TGFB1R => ACE	1.00	1.4	0.6	rat cardiac fibroblasts	increased txn
middle	r41	TGFB & !BAMBI => TGFB1R	1.00	1.4	0.6	mice cardiac fibroblast	binding to receptor
middle	r42	AP1 => proliferation	1.00	1.4	0.6	adult rat cardiac fibroblasts	via activation of Kca3.1 channels
middle	r43	PKA => CREB	1.00	1.4	0.6	rat cardiac fibroblasts	activation
middle	r44	CREB => proliferation	1.00	1.4	0.6	rat cardiac fibroblasts	
middle	r45	NE => BAR	1.00	1.4	0.6	rat cardiac fibroblasts	receptor binding
middle	r46	ET1 => ETAR	1.00	1.4	0.6	neonatal rat cardiac fibroblasts	receptor binding
middle	r47	CTGF => proliferation	1.00	1.4	0.6	human cardiac fibroblast	
middle	r48	IL1 => IL1RI	1.00	1.4	0.6	mouse cell line	receptor binding
middle	r49	PKC => proliferation	1.00	1.4	0.6	adult rat cardiac fibroblasts	activation
output	r50	smad3 & CBP & !epac => proCI	1.00	1.4	0.6	adult rat cardiac fibroblasts	txn

output	r51	smad3 & CBP & !epac => proCIII	1.00	1.4	0.6	adult rat cardiac fibroblasts	txn
output	r52	AP1 => proMMP14	1.00	1.4	0.6	mouse cardiac fibroblasts	correlated increase with cFOS
middle	r53	PDGF => PDGFR	1.00	1.4	0.6	adult rat cardiac fibroblasts	receptor binding
middle	r54	BAR => AC	1.00	1.4	0.6	adult rat cardiac fibroblasts	activation
middle	r55	BAR & AT1R => AC	1.00	1.4	0.6	adult rat cardiac fibroblasts	activation with potentiation
middle	r56	AC => cAMP	1.00	1.4	0.6	adult rat cardiac fibroblasts	activation
middle	r57	FAK =>MEKK1	1.00	1.4	0.6	mouse embryonic fibroblasts	activation
output	r58	AP1 => latentTGFB	1.00	1.4	0.6	mouse lung fibroblasts	txn activation
middle	r59	cAMP => epac	1.00	1.4	0.6	adult rat cardiac fibroblasts	activation
middle	r60	Rho => ROCK	1.00	1.4	0.6	rat embryonic fibroblasts	activation
middle	r61	TNFa => TNFaR	1.00	1.4	0.6	human cardiac fibroblast	receptor binding
middle	r62	NP => NPRA	1.00	1.4	0.6	human cardiac fibroblast	receptor binding
middle	r63	NPRA => cGMP	1.00	1.4	0.6	adult rat cardiac fibroblast	activation
middle	r64	cGMP => PKG	1.00	1.4	0.6	adult rat cardiac fibroblast	
middle	r65	Ras => Raf	1.00	1.4	0.6	neonatal rat cardiac fibroblast	possibly via recruitment and Src phosphorylation
middle	r66	Raf & !ERK => MEK1	1.00	1.4	0.6	adult rat cardiac fibroblast	
middle	r67	MEK1 & !PP1=> ERK	1.00	1.4	0.6	adult rat cardiac fibroblast	
middle	r68	p38 => PP1	1.00	1.4	0.6	3T3 cells, adult and neonatal human dermal fibroblast	via activation
middle	r69	MKK3 => p38	1.00	1.4	0.6	3T3 cells, adult and neonatal human dermal fibroblast	activation
middle	r70	TGFB1R => TRAF	1.00	1.4	0.6	adult mouse cardiac fibroblast	activation
middle	r71	Rac1 => MEKK1	1.00	1.4	0.6	NIH-3T3, HeLa	activation

middle	r72	MEKK1 => MKK4	1.00	1.4	0.6	NIH-3T3, HeLa	activation
middle	r73	MKK4 & !NFKB => JNK	1.00	1.4	0.6	NIH-3T3, HeLa	activation
middle	r74	PDGFR => abl	1.00	1.4	0.6	3T3	activation
middle	r75	abl => Rac1	1.00	1.4	0.6	3T3	activation
middle	r76	JNK => cmyc	1.00	1.4	0.6	3T3	activation
middle	r77	cmyc => proliferation	1.00	1.4	0.6	3T3	activation
middle	r78	TNFaR => TRAF	1.00	1.4	0.6	293 cells	activation
middle	r79	TRAF => ASK1	1.00	1.4	0.6	293 cells	activation - most likely binding allows the receptor to eventually activate ASK1
middle	r80	ASK1 => MKK3	1.00	1.4	0.6	COS7 cells	activation
middle	r81	ASK1 => MKK4	1.00	1.4	0.6	COS7 cells	activation
middle	r82	IL1RI => ASK1	1.00	1.4	0.6	fibroblast-like synoviocytes	assumed activation
middle	r83	smad3 => PAI1	1.00	1.4	0.6	adult mouse cardiac fibroblast	transcription
output	r84	NFKB => proMMP14	1.00	1.4	0.6	human dermal fibroblast	transcription
middle	r85	Ras => p38	1.00	1.4	0.6	adult rat cardiac fibroblast	unknown
middle	r86	TGFB1R => PI3K	1.00	1.4	0.6	adult rat cardiac fibroblast	activation
middle	r87	PDGFR => PI3K	1.00	1.4	0.6	3T3	activation
middle	r88	FAK => PI3K	1.00	1.4	0.6	human lung fibroblast	activation
middle	r89	TGFB1R => NOX	1.00	1.4	0.6	human cardiac fibroblast	activation
middle	r90	Akt => NFKB	1.00	1.4	0.6	human cardiac fibroblast	activation by removal of IKK
output	r91	NFKB => fibronectin	1.00	1.4	0.6	human cardiac fibroblast	transcription
middle	r92	JNK => AP1	1.00	1.4	0.6	human periodontal ligament fibroblast	activation
middle	r93	IL1RI & TGFB => BAMBI	1.00	1.4	0.6	mice cardiac fibroblast	increased transcription (unsure of transcription factor)
middle	r94	STAT => smad7	1.00	1.4	0.6	UA4 cell line	STAT necessary for smad expression
output	r95	SRF => proCI	1.00	1.4	0.6	10t1/2 cells, cardiac fibroblasts	MRTF directly activates the expression of COL1
middle	r96	Rho & !Rac1 => p38	1.00	1.4	0.6	neonatal rat cardiac fibroblast	
middle	r97	MKK4 & !Rho => JNK	1.00	1.4	0.6	neonatal rat cardiac fibroblast	

output	r98	SRF => proCIII	1.00	1.4	0.6	mouse cardiac fibroblasts	
middle	r99	calcineurin => NFAT	1.00	1.4	0.6	mouse cardiac fibroblasts	activation/nuclear translocation
middle	r100	AT1R => Ras	1.00	1.4	0.6	neonatal cardiac fibroblasts	
output	r101	smad3 & CBP => aSMA	1.00	1.4	0.6	human cardiac fibroblast	txn activation
output	r102	SRF => aSMA	1.00	1.4	0.6	rat cardiac fibroblasts	transcription
middle	r103	ETAR => DAG	1.00	1.4	0.6	rat embryonic fibroblasts	production
middle	r104	AT1R => DAG	1.00	1.4	0.6	CHO cells	production
middle	r105	DAG => TRPC	1.00	1.4	0.6	human cardiac fibroblast	activation
middle	r106	TRPC => Ca	1.00	1.4	0.6	human cardiac fibroblast	channel opening
middle	r107	Ca => calcineurin	1.00	1.4	0.6	adult rat cardiac fibroblast	activation
middle	r108	TGFB1R => Rho	1.00	1.4	0.6	human gingivial fibroblasts	
middle	r109	B3int => Src	1.00	1.4	0.6	human lung fibroblasts	dephosphorylation: Y530, autophosphorylation: Y419
middle	r110	B1int => FAK	1.00	1.4	0.6	mouse embryonic fibroblasts, COS7	autophosphorylation: Y397
middle	r111	FAK & Src => Grb2	1.00	1.4	0.6	mouse embryonic fibroblasts	activation via Src
middle	r112	Grb2 => Ras	1.00	1.4	0.6	rat cardiac fibroblasts	activation via SOS
middle	r113	FAK & Src => RhoGEF	1.00	1.4	0.6	mouse embryonic fibroblasts	activation
middle	r114	!Src => RhoGDI	1.00	1.4	0.6	mouse embryonic fibroblasts, HeLa	phorphorylation: decreases binding Rho binding affinity
middle	r115	FAK & Src => p130Cas	1.00	1.4	0.6	mouse embryonic fibroblasts	activation via Src
middle	r116	PDGFR => Src	1.00	1.4	0.6	mouse embryonic fibroblasts	activation
middle	r117	tension & Src => p130Cas	1.00	1.4	0.6	mouse embryonic fibroblasts	activation
middle	r118	p130Cas & abl => Rac1	1.00	1.4	0.6	HEK293	activation
middle	r119	Factin => YAP	1.00	1.4	0.6	mouse embryonic fibroblasts	dephosphorylation + translocation

middle	r120	PKA => RhoGDI	1.00	1.4	0.6	rat cardiac fibroblasts	phosphorylation
middle	r121	RhoGEF & !RhoGDI & !PKG => Rho	1.00	1.4	0.6	mouse embryonic fibroblasts	activation
output	r122	YAP => CTGF	1.00	1.4	0.6	mouse embryonic fibroblasts	txn via TEAD
middle	r123	syndecan4 => PKC	1.00	1.4	0.6	rat embryonic fibroblasts	activation
middle	r124	!PKC => RhoGDI	1.00	1.4	0.6	rat embryonic cardiomyocytes	phosphorylation
middle	r125	NFAT & !Gactin => MRTF	1.00	1.4	0.6	mouse cardiac fibroblasts	translocation
middle	r126	ROCK & Gactin => Factin	1.00	1.4	0.6	rat cardiac fibroblasts	polymerization
middle	r127	!Factin => Gactin	1.00	1.4	0.6	mouse embryonic fibroblasts	polymerization
middle	r128	MRTF => SRF	1.00	1.4	0.6	human lung fibroblasts	
fback	r129	!TNC & tension => syndecan4	1.00	1.4	0.6	mouse cardiac fibroblasts	dephosphorylation
middle	r130	Akt => mTORC1	1.00	1.4	0.6	HEK293	activation via TSC1/2, PRAS40 inhibition
middle	r131	mTORC1 => p70S6K	1.00	1.4	0.6	mouse embryonic fibroblasts	activation
middle	r132	!mTORC1 => EBP1	1.00	1.4	0.6	mouse embryonic fibroblasts	phosphorylation
middle	r133	!EBP1 & p70S6K => proliferation	1.00	1.4	0.6	rat cardiac fibroblasts	mRNA translation
middle	r134	Akt => smad3	1.00	1.4	0.6	mouse cardiac fibroblasts	activation via GSK3B inhibition
output	r135	NFKB => TNC	1.00	1.4	0.6	human cardiac fibroblasts	txn
output	r136	MRTF => TNC	1.00	1.4	0.6	mouse embryonic fibroblasts	txn
middle	r137	!p70S6K => mTORC2	1.00	1.4	0.6	mouse embryonic fibroblasts	phosphorylation via Rictor
middle	r138	mTORC2 & PI3K => Akt	1.00	1.4	0.6	mouse embryonic fibroblasts	activation
middle	r139	mTORC2 & DAG => PKC	1.00	1.4	0.6	HEK293	activation
output	r140	YAP => PAI1	1.00	1.4	0.6	human lung fibroblasts	txn

middle	r141	smad3 => thrombospondin4	1.00	1.4	0.6	human dermal fibroblasts	txn
fback	r142	!thrombospondin4 & tension => B3int	0.80	1.4	0.6	mouse cardiac fibroblasts	receptor binding
output	r143	NFKB & AP1 & !smad3 => proMMP8	1.00	1.4	0.6	mouse cardiac fibroblasts	txn
output	r144	NFKB & AP1 & !smad3 => proMMP3	1.00	1.4	0.6	mouse cardiac fibroblasts	txn
output	r145	AP1 => osteopontin	1.00	1.4	0.6	rat cardiac fibroblasts	txn
fback	r146	osteopontin => B3int	0.80	1.4	0.6	rat cardiac fibroblasts	receptor binding
output	r147	CREB => proMMP12	1.00	1.4	0.6	human dermal fibroblasts	txn
middle	r148	AP1 & !YAP => smad7	1.00	1.4	0.6	human dermal fibroblasts	txn: YAP/TAZ knockdown required for smad7 expression
middle	r149	FAK & Src & MLC => paxillin	1.00	1.4	0.6	human foreskin fibroblasts	activation
middle	r150	vinculin & !paxillin => FA	1.00	1.4	0.6	human foreskin fibroblasts	stabilization: paxillin increases FA turnover for increased migration
middle	r151	B1int => talin	1.00	1.4	0.6	mouse embryonic fibroblasts	activation
middle	r152	B3int => talin	1.00	1.4	0.6	mouse embryonic fibroblasts	activation
middle	r153	talin & tension => vinculin	1.00	1.4	0.6	mouse embryonic fibroblasts	activation
middle	r154	Factin & MLC & vinculin => contractility	1.00	1.4	0.6	mouse embryonic fibroblasts	binding via vinculin tail region
fback	r155	contractility & FA => tension	0.80	1.4	0.6	human foreskin fibroblasts	force generation via molecular clutch theory
middle	r156	ROCK => MLC	1.00	1.4	0.6	mouse embryonic fibroblasts	activation via MLCK activation, MBS inhibition
middle	r157	tension => TRPC	1.00	1.4	0.6	mouse embryonic fibroblasts	activation
middle	r158	tension => AT1R	1.00	1.4	0.6	rat cardiac fibroblasts	activation
middle	r159	AT1R => YAP	1.00	1.4	0.6	rat cardiac fibroblasts	dephosphorylation + translocation
middle	r160	YAP => aSMA	1.00	1.4	0.6	rat cardiac fibroblasts	txn
middle	r161	AngII => AT2R	1.00	1.4	0.6	human cardiac fibroblasts	activation

middle	r162	ROS & !AT2R => ERK	1.00	1.4	0.6	neonatal rat cardiac fibro-blasts	activation
middle	r163	!abl => BCL6	1.00	1.4	0.6	GRN	inhibition
output	r164	!BCL6 => proCI	1.00	1.4	0.6	GRN	inhibition
output	r165	!BCL6 => proCIII	1.00	1.4	0.6	GRN	inhibition
output	r166	!BCL6 => CTGF	1.00	1.4	0.6	GRN	inhibition
middle	r167	!BCL6 => NFKB	1.00	1.4	0.6	GRN	inhibition
output	r168	!BCL6 => peri-ostin	1.00	1.4	0.6	GRN	inhibition
output	r169	!CEBPD => CTGF	1.00	1.4	0.6	GRN	inhibition
output	r170	!EGR1 => IL6	1.00	1.4	0.6	GRN	inhibition
output	r171	!EGR1 => proMMP2	1.00	1.4	0.6	GRN	inhibition
output	r172	!EGR1 => TIMP2	1.00	1.4	0.6	GRN	inhibition
output	r173	!ETS2 => proMMP2	1.00	1.4	0.6	GRN	inhibition
output	r174	!HIF1A => PAI1	1.00	1.4	0.6	GRN	inhibition
middle	r175	!AP1 => KLF4	1.00	1.4	0.6	GRN	inhibition
middle	r176	!AP1 => RARG	1.00	1.4	0.6	GRN	inhibition
output	r177	!KLF4 => proCI	1.00	1.4	0.6	GRN	inhibition
output	r178	!LEF1 => proCI	1.00	1.4	0.6	GRN	inhibition
output	r179	!LEF1 => peri-ostin	1.00	1.4	0.6	GRN	inhibition
output	r180	!LEF1 => throm-bospondin4	1.00	1.4	0.6	GRN	inhibition
output	r181	!MITF => TIMP2	1.00	1.4	0.6	GRN	inhibition
middle	r182	!cmyc => EGR1	1.00	1.4	0.6	GRN	inhibition
middle	r183	!cmyc => ETS2	1.00	1.4	0.6	GRN	inhibition
middle	r184	!cmyc => AP1	1.00	1.4	0.6	GRN	inhibition
middle	r185	!cmyc => HIF1A	1.00	1.4	0.6	GRN	inhibition
middle	r186	!cmyc => KLF4	1.00	1.4	0.6	GRN	inhibition
middle	r187	!cmyc => NFAT	1.00	1.4	0.6	GRN	inhibition
middle	r188	!cmyc => RELA	1.00	1.4	0.6	GRN	inhibition
middle	r189	!cmyc => RUNX1	1.00	1.4	0.6	GRN	inhibition

output	r190	!cmyc => PAI1	1.00	1.4	0.6	GRN	inhibition
middle	r191	!NFKB => EGR1	1.00	1.4	0.6	GRN	inhibition
middle	r192	!NFKB => RUNX1	1.00	1.4	0.6	GRN	inhibition
middle	r193	!NFKB => TFCP2L1	1.00	1.4	0.6	GRN	inhibition
output	r194	!NR5A2 => proCIII	1.00	1.4	0.6	GRN	inhibition
output	r195	!PPARA => proMMP9	1.00	1.4	0.6	GRN	inhibition
output	r196	!RARG => proCI	1.00	1.4	0.6	GRN	inhibition
output	r197	!RARG => proMMP2	1.00	1.4	0.6	GRN	inhibition
middle	r198	!RELA => CE-BPD	1.00	1.4	0.6	GRN	inhibition
output	r199	!RUNX1 => CTGF	1.00	1.4	0.6	GRN	inhibition
middle	r200	!RUNX1 => KLF4	1.00	1.4	0.6	GRN	inhibition
middle	r201	!RUNX1 => LEF1	1.00	1.4	0.6	GRN	inhibition
output	r202	!RUNX1 => PAI1	1.00	1.4	0.6	GRN	inhibition
output	r203	!RUNX1 => TNC	1.00	1.4	0.6	GRN	inhibition
output	r204	!RUNX2 => CTGF	1.00	1.4	0.6	GRN	inhibition
output	r205	!RUNX2 => PAI1	1.00	1.4	0.6	GRN	inhibition
output	r206	!smad3 => CTGF	1.00	1.4	0.6	GRN	inhibition
middle	r207	!smad3 => ETS2	1.00	1.4	0.6	GRN	inhibition
middle	r208	!SRF => LEF1	1.00	1.4	0.6	GRN	inhibition
output	r209	!SRF => PAI1	1.00	1.4	0.6	GRN	inhibition
output	r210	!STAT => CTGF	1.00	1.4	0.6	GRN	inhibition
middle	r211	!STAT => ETS1	1.00	1.4	0.6	GRN	inhibition
middle	r212	!STAT => MITF	1.00	1.4	0.6	GRN	inhibition
output	r213	!STAT => proMMP2	1.00	1.4	0.6	GRN	inhibition
middle	r214	!STAT => NFAT	1.00	1.4	0.6	GRN	inhibition
output	r215	!STAT => peri-ostin	1.00	1.4	0.6	GRN	inhibition
middle	r216	!STAT => RARG	1.00	1.4	0.6	GRN	inhibition
output	r217	!STAT => PAI1	1.00	1.4	0.6	GRN	inhibition

middle	r218	!STAT => TCF4	1.00	1.4	0.6	GRN	inhibition
middle	r219	!STAT => TEAD4	1.00	1.4	0.6	GRN	inhibition
middle	r220	!STAT => TFCP2L1	1.00	1.4	0.6	GRN	inhibition
middle	r221	!STAT => WT1	1.00	1.4	0.6	GRN	inhibition
middle	r222	!STAT => ZNF281	1.00	1.4	0.6	GRN	inhibition
middle	r223	!TCF4 => CTGF	1.00	1.4	0.6	GRN	inhibition
middle	r224	!TCF4 => KLF4	1.00	1.4	0.6	GRN	inhibition
middle	r225	!TCF4 => MITF	1.00	1.4	0.6	GRN	inhibition
output	r226	!TEAD4 => CTGF	1.00	1.4	0.6	GRN	inhibition
middle	r227	!TEAD4 => AP1	1.00	1.4	0.6	GRN	inhibition
middle	r228	!TEAD4 => NFAT	1.00	1.4	0.6	GRN	inhibition
output	r229	!TFCP2L1 => PAI1	1.00	1.4	0.6	GRN	inhibition
middle	r230	!YAP => NFKB	1.00	1.4	0.6	GRN	inhibition
middle	r231	!YAP => PPARA	1.00	1.4	0.6	GRN	inhibition
output	r232	!ZNF281 => proCI	1.00	1.4	0.6	GRN	inhibition
output	r233	!ZNF281 => PAI1	1.00	1.4	0.6	GRN	inhibition
middle	r234	abl => AP1	1.00	1.4	0.6	GRN	activation
output	r235	BCL6 => proCI	1.00	1.4	0.6	GRN	activation
output	r236	BCL6 => proCIII	1.00	1.4	0.6	GRN	activation
output	r237	BCL6 => CTGF	1.00	1.4	0.6	GRN	activation
middle	r238	BCL6 => cmyc	1.00	1.4	0.6	GRN	activation
output	r239	BCL6 => perios-tin	1.00	1.4	0.6	GRN	activation
output	r240	CACYBP => TIMP2	1.00	1.4	0.6	GRN	activation
output	r241	CEBPD => CTGF	1.00	1.4	0.6	GRN	activation
middle	r242	CREB => AP1	1.00	1.4	0.6	GRN	activation
middle	r243	CREB => NR5A2	1.00	1.4	0.6	GRN	activation
output	r244	EGR1 => proMMP14	1.00	1.4	0.6	GRN	activation
output	r245	EGR1 => proMMP2	1.00	1.4	0.6	GRN	activation

output	r246	EGR1 => PDGF	1.00	1.4	0.6	GRN	activation
output	r247	EGR1 => TIMP2	1.00	1.4	0.6	GRN	activation
output	r248	EGR1 => TNFa	1.00	1.4	0.6	GRN	activation
output	r249	ETS2 => proMMP2	1.00	1.4	0.6	GRN	activation
output	r250	ETS2 => proMMP9	1.00	1.4	0.6	GRN	activation
middle	r251	ETS2 => cmyc	1.00	1.4	0.6	GRN	activation
output	r252	HIF1A => PAI1	1.00	1.4	0.6	GRN	activation
output	r253	IKZF1 => proMMP9	1.00	1.4	0.6	GRN	activation
middle	r254	AP1 => KLF4	1.00	1.4	0.6	GRN	activation
middle	r255	AP1 => RUNX2	1.00	1.4	0.6	GRN	activation
output	r256	KLF4 => proCI	1.00	1.4	0.6	GRN	activation
output	r257	LEF1 => proCI	1.00	1.4	0.6	GRN	activation
output	r258	LEF1 => perios-tin	1.00	1.4	0.6	GRN	activation
output	r259	LEF1 => throm-bospondin4	1.00	1.4	0.6	GRN	activation
middle	r260	MITF => AP1	1.00	1.4	0.6	GRN	activation
middle	r261	MITF => HIF1A	1.00	1.4	0.6	GRN	activation
output	r262	MITF => TIMP2	1.00	1.4	0.6	GRN	activation
middle	r263	cmyc => EGR1	1.00	1.4	0.6	GRN	activation
middle	r264	cmyc => ETS2	1.00	1.4	0.6	GRN	activation
middle	r265	cmyc => AP1	1.00	1.4	0.6	GRN	activation
middle	r266	cmyc => HIF1A	1.00	1.4	0.6	GRN	activation
middle	r267	cmyc => KLF4	1.00	1.4	0.6	GRN	activation
middle	r268	cmyc => NFAT	1.00	1.4	0.6	GRN	activation
middle	r269	cmyc => RELA	1.00	1.4	0.6	GRN	activation
middle	r270	cmyc => RUNX1	1.00	1.4	0.6	GRN	activation
output	r271	cmyc => PAI1	1.00	1.4	0.6	GRN	activation
middle	r272	NFAT => CA-CYBP	1.00	1.4	0.6	GRN	activation
middle	r273	NFAT => LEF1	1.00	1.4	0.6	GRN	activation

middle	r274	NFKB => IKZF1	1.00	1.4	0.6	GRN	activation
middle	r275	NFKB => RUNX1	1.00	1.4	0.6	GRN	activation
middle	r276	NFKB => TFCP2L1	1.00	1.4	0.6	GRN	activation
output	r277	NR5A2 => proCIII	1.00	1.4	0.6	GRN	activation
output	r278	PPARA => PAI1	1.00	1.4	0.6	GRN	activation
output	r279	RARG => proCI	1.00	1.4	0.6	GRN	activation
output	r280	RARG => proMMP2	1.00	1.4	0.6	GRN	activation
middle	r281	RELA => CE-BPD	1.00	1.4	0.6	GRN	activation
middle	r282	RELA => NFKB	1.00	1.4	0.6	GRN	activation
middle	r283	RUNX1 => AP1	1.00	1.4	0.6	GRN	activation
output	r284	RUNX1 => CTGF	1.00	1.4	0.6	GRN	activation
middle	r285	RUNX1 => KLF4	1.00	1.4	0.6	GRN	activation
middle	r286	RUNX1 => LEF1	1.00	1.4	0.6	GRN	activation
output	r287	RUNX1 => PAI1	1.00	1.4	0.6	GRN	activation
output	r288	RUNX1 => TNC	1.00	1.4	0.6	GRN	activation
output	r289	RUNX2 => CTGF	1.00	1.4	0.6	GRN	activation
output	r290	RUNX2 => PAI1	1.00	1.4	0.6	GRN	activation
output	r291	smad3 => CTGF	1.00	1.4	0.6	GRN	activation
middle	r292	smad3 => ETS2	1.00	1.4	0.6	GRN	activation
middle	r293	SRF => AP1	1.00	1.4	0.6	GRN	activation
middle	r294	STAT => AP1	1.00	1.4	0.6	GRN	activation
middle	r295	STAT => HIF1A	1.00	1.4	0.6	GRN	activation
middle	r296	STAT => PPARA	1.00	1.4	0.6	GRN	activation
middle	r297	STAT => BCL6	1.00	1.4	0.6	GRN	activation
middle	r298	STAT => ETS1	1.00	1.4	0.6	GRN	activation
middle	r299	STAT => MITF	1.00	1.4	0.6	GRN	activation
middle	r300	STAT => NFAT	1.00	1.4	0.6	GRN	activation
output	r301	STAT => perios-tin	1.00	1.4	0.6	GRN	activation

middle	r302	STAT => RARG	1.00	1.4	0.6	GRN	activation
output	r303	STAT => PAI1	1.00	1.4	0.6	GRN	activation
middle	r304	STAT => TCF4	1.00	1.4	0.6	GRN	activation
middle	r305	STAT => TEAD4	1.00	1.4	0.6	GRN	activation
middle	r306	STAT => TFCP2L1	1.00	1.4	0.6	GRN	activation
middle	r307	STAT => WT1	1.00	1.4	0.6	GRN	activation
middle	r308	STAT => ZNF281	1.00	1.4	0.6	GRN	activation
output	r309	TCF4 => CTGF	1.00	1.4	0.6	GRN	activation
output	r310	TEAD4 => CTGF	1.00	1.4	0.6	GRN	activation
middle	r311	TEAD4 => AP1	1.00	1.4	0.6	GRN	activation
middle	r312	TEAD4 => NFAT	1.00	1.4	0.6	GRN	activation
output	r313	TFCP2L1 => PAI1	1.00	1.4	0.6	GRN	activation
middle	r314	WT1 => AP1	1.00	1.4	0.6	GRN	activation
middle	r315	YAP => NFKB	1.00	1.4	0.6	GRN	activation
middle	r316	YAP => PPARA	1.00	1.4	0.6	GRN	activation
output	r317	ZNF281 => proCI	1.00	1.4	0.6	GRN	activation
output	r318	ZNF281 => PAI1	1.00	1.4	0.6	GRN	activation
output	r319	!CUX1 => CTGF	1.00	1.4	0.6	GRN	inhibition
middle	r320	!STAT => CUX1	1.00	1.4	0.6	GRN	inhibition
middle	r321	!TEAD2 => CUX1	1.00	1.4	0.6	GRN	inhibition
output	r322	CUX1 => CTGF	1.00	1.4	0.6	GRN	activation
middle	r323	STAT => CUX1	1.00	1.4	0.6	GRN	activation
middle	r324	TEAD2 => CUX1	1.00	1.4	0.6	GRN	activation

Table S4.2 Species and Species parameters used in Chapter 4.

Species information							
module	ID	name	Yinit	Ymax	tau	type	gene name

g-coupled	AngII	angiotensin II	0	0.45	1	protein	AGT
g-coupled	AT1R	angiotensin II receptor type 1	0	0.49	0.1	protein	AGTR1
g-coupled	AGT	angiotensinogen	0	0.45	10	protein	AGT
g-coupled	ACE	angiotensin converting enzyme	0	0.32	0.1	protein	ACE; ACE2
g-coupled	NOX	NAD(P)H oxidase	0	0.34	0.1	protein	NOX4; NOX5
g-coupled	ROS	reactive oxygen species	0	0.50	0.1	protein	
g-coupled	ET1	endothelin 1	0	0.43	1	protein	EDN1
g-coupled	ETAR	endothelin 1 receptor A	0	0.27	0.1	protein	EDNRA
g-coupled	DAG	diacyl-glycerol	0	0.50	0.1	small	
g-coupled	PKC	protein kinase C	0	0.41	0.1	protein	PRKCA; PRKCE;
pressure/stretch	TRPC	transient receptor potential canonical	0	0.41	0.1	protein	TRPC6;TRPC3
g-coupled	NE	norepinephrine	0	0.50	1	small	
g-coupled	BAR	beta adrenergic receptor 1 or 2	0	0.45	0.1	protein	ADRB1; ADRB2
g-coupled	AC	adenylate cyclase	0	0.42	0.1	protein	ADCY6
g-coupled	cAMP	cyclic adenosine monophosphate	0	0.50	0.1	small	
g-coupled	PKA	protein kinase A	0	0.46	0.1	protein	PRKACA
g-coupled	CREB	cAMP response-element binding protein	0	0.48	0.1	protein	CREB1; CREB3
g-coupled	CBP	CREB - binding protein	0	0.42	0.1	protein	CREBBP
growth factor	TGFB	transforming growth factor beta 1	0	0.37	1	protein	TGFB1
growth factor	TGFB1R	TGFB receptor	0	0.47	0.1	protein	TGFBR1; TGFBR2
growth factor	smad3	small mothers against decapentaplegic 2 and 3	0	0.45	0.1	protein	SMAD2; SMAD3
growth factor	smad7	small mothers against decapentaplegic 7	0	0.40	10	protein	SMAD7
growth factor	latentTGFB	TGFB1 with latent protein complex	0	0.37	10	protein	TGFB1
growth factor	BAMBI	BMP and activin bound inhibitor	0	0.32	0.1	protein	BAMBI
growth factor	PDGF	platelet derived growth factor	0	0.36	1	protein	PDGFA; PDGFB; PDGFD
growth factor	PDGFR	platelet derived growth factor receptor	0	0.50	0.1	protein	PDGFRA; PDGFRB
g-coupled	NP	natriuretic peptide	0	0.29	1	protein	NPPA; NPPB

g-coupled	NPRA	natriuretic peptide receptor	0	0.32	0.1	protein	NPR1; NPR2; NPR3
g-coupled	cGMP	cyclic guanosine monophosphate	0	0.60	0.1	small	
g-coupled	PKG	protein kinase G	0	0.49	0.1	protein	PRKG1
pressure/stretch	tension	stretch	0	0.25	1	process	
pressure/stretch	B1int	beta 1 integrin	0	0.42	0.1	protein	ITGB1
pressure/stretch	Rho	a Rho-dependent GTPase	0	0.50	0.1	protein	RHOA
pressure/stretch	ROCK	rho associated protein kinase	0	0.50	0.1	protein	ROCK1
pressure/stretch	Ca	calcium	0	0.50	0.1	small	
pressure/stretch	calcineurin	calcineurin	0	0.55	0.1	protein	PPP3CA; PPP3CB
pressure/stretch	NFAT	nuclear factor of activated T-cells	0	0.33	0.1	protein	NFATC1
cytokine	IL6	interleukin-6	0	0.42	1	protein	IL6
cytokine	gp130	IL-6 receptor complexed to gp130 for signal transduction	0	0.47	0.1	protein	IL6ST; IL6R
cytokine	STAT	signal transducers and activators of transcription 1 and 3	0	0.37	0.1	protein	STAT1; STAT3
cytokine	IL1	interleukin-1 alpha and beta	0	0.27	1	protein	IL1B; IL1A
cytokine	IL1RI	IL1 receptor type I	0	0.51	0.1	protein	IL1R1
cytokine	TNFa	tissue necrosis factor alpha	0	0.50	1	protein	TNF
cytokine	TNFaR	TNF alpha receptor	0	0.45	0.1	protein	TNFRSF1A ;TNFRSF1B
cytokine	NFKB	nuclear factor kappa-light-chain-enhancer of activated B cells	0	0.45	0.1	protein	NFKB1
cytokine	PI3K	phosphoinositide 3-kinase	0	0.47	0.1	protein	PIK3CA
cytokine	Akt	protein kinase B	0	0.44	0.1	protein	AKT1; AKT2; AKT3
MAPK	p38	a MAP kinase	0	0.52	0.1	protein	MAPK14
MAPK	TRAF	tnf receptor associated factor either 2/6	0	0.53	0.1	protein	TRAF6
MAPK	ASK1	apoptosis signal related kinase 1	0	0.49	0.1	protein	MAP3K5
MAPK	MKK3	mitogen activated protein kinase kinase	0	0.38	0.1	protein	MAP2K3

MAPK	PP1	protein phosphatase 1	0	0.55	0.1	protein	PPP1CA; PPP1CB; PPP1CC
MAPK	JNK	a MAP kinase	0	0.44	0.1	protein	MAPK8
MAPK	abl	abl tyrosine kinase	0	0.50	0.1	protein	ABL1; ABL2
MAPK	Rac1	a Rho-dependent GTPase	0	0.51	0.1	protein	RAC1
MAPK	MEKK1	a MAP3K associated with p38 and JNK	0	0.55	0.1	protein	MAP3K1
MAPK	MKK4	a MAP2K associated with p38 and JNK	0	0.54	0.1	protein	MAP2K4
MAPK	ERK	a MAP kinase	0	0.46	0.1	protein	MAPK1; MAPK3
MAPK	Ras	representing the family of GTPases	0	0.53	0.1	protein	KRAS
MAPK	Raf	family of raf protein ser-ine/threonine kinases	0	0.58	0.1	protein	RAF1
MAPK	MEK1	a MAP2K mainly specific to ERK	0	0.43	0.1	protein	MAP2K1
adhe-sion	FAK	focal adhesion kinase	0	0.34	0.1	protein	PTK2
g-cou-pled	epac	exchange protein activated by cAMP 1	0	0.53	0.1	protein	RAPGEF3
adhe-sion	Factin	polymerized actin	0	0.42	1		ACTG1
adhe-sion	FA	stabilization of focal adhe-sions	0	0.50	1	complex	
growth	cmyc	myc transcription factor	0	0.45	0.1	protein	MYC
ECM	CTGF	connective tissue growth factor	0	0.36	0.1	protein	CTGF
growth	prolifera-tion	proliferation	0	0.50	10	event	
adhe-sion	SRF	serum response factor	0	0.36	0.1	protein	SRF
ECM	EDAFN	extra domain A of fibronectin	0	0.33	10	protein	FN1
adhe-sion	aSMA	alpha-smooth muscle actin	0	0.40	10	protein	ACTA2
MAPK	AP1	activator protein 1	0	0.26	0.1	protein	JUN; FOS
ECM	TIMP1	tissue inhibitor of metallo-proteinase 1	0	0.41	10	protein	TIMP1
ECM	TIMP2	tissue inhibitor of metallo-proteinase 2	0	0.26	10	protein	TIMP2
ECM	PAI1	plasminogen activator inhib-itor 1	0	0.35	10	protein	SERPINE1
ECM	proMMP14	inactive MMP14	0	0.27	10	protein	MMP14
ECM	proMMP1	inactive MMP1	0	0.50	10	protein	MMP1
ECM	proMMP2	inactive MMP2	0	0.21	10	protein	MMP2
ECM	proMMP9	inactive MMP9	0	0.29	10	protein	MMP9
ECM	fibron-ectin	fibronectin	0	0.33	10	protein	FN1
ECM	periostin	periostin	0	0.27	10	protein	POSTN
ECM	proCI	procollagen I	0	0.21	10	protein	COL1A1

ECM	proCIII	procollagen III	0	0.24	10	protein	COL3A1
pressure/stretch	B3int	beta 3 integrin	0	0.50	0.1	protein	ITGB3
adhesion	Src	proto-oncogene tyrosine-protein kinase Src	0	0.35	0.1	protein	SRC
MAPK	Grb2	growth factor receptor-bound protein 2	0	0.57	0.1	protein	GRB2
adhesion	p130Cas	breast cancer anti-estrogen resistance protein 1	0	0.28	0.1	protein	BCAR1
pressure/stretch	YAP	yes-associated protein 1	0	0.38	0.1	protein	YAP1
adhesion	MRTF	myocardin-related transcription factor A	0	0.47	0.1	protein	MRTFA; MKL1
adhesion	Gactin	monomeric actin	0	0.42	1	protein	ACTG1
ECM	TNC	tenascin-c	0	0.37	10	protein	TNC
growth	mTORC1	mammalian target of rapamycin complex 1	0	0.50	0.1	complex	
growth	mTORC2	mammalian target of rapamycin complex 2	0	0.50	0.1	complex	
growth	p70S6K	p70-S6 kinase 1	0	0.54	0.1	protein	RPS6KB1
growth	EBP1	eukaryotic translation initiation factor 4E-binding protein 1	0	0.45	0.1	protein	EIF4EBP1
pressure/stretch	syndecan 4	syndecan 4	0	0.48	0.1	protein	SDC4
ECM	proMMP3	inactive MMP3	0	0.50	10	protein	MMP3
ECM	proMMP8	inactive MMP8	0	0.50	10	protein	MMP8
ECM	proMMP12	inactive MMP12	0	0.50	10	protein	MMP12
ECM	thrombospondin4	thrombospondin 4	0	0.19	10	protein	THBS4
ECM	osteopontin	osteopontin	0	0.38	10	protein	SPP1
adhesion	contractility	intracellular tension	0	0.50	10	event	
pressure/stretch	RhoGEF	a Rho guanine nucleotide exchange factor	0	0.50	0.1	protein	
pressure/stretch	RhoGDI	a Rho GDP-dissociation inhibitor	0	0.50	0.1	protein	
adhesion	talin	talin 1	0	0.59	0.1	protein	TLN1
adhesion	vinculin	vinculin	0	0.55	0.1	protein	VCL
adhesion	paxillin	paxillin	0	0.34	0.1	protein	PXN
adhesion	MLC	myosin regulatory light chain	0	0.31	0.1	protein	MYL2

g-coupled	AT2R	angiotensin II receptor type 2	0	0.51	0.1	protein	AGTR2
txn	BCL6	B-Cell Lymphoma 6 Protein	0	0.48	1	transcription factor	BCL6
txn	CACYBP	Calcyclin Binding Protein	0	0.30	1	transcription factor	CACYBP
txn	CEBPD	CCAAT Enhancer Binding Protein Delta	0	0.42	1	transcription factor	CEBPD
txn	CUX1	Cut Like 1 Homebox	0	0.50	1	transcription factor	CUX1
txn	EGR1	Early Growth Response 1	0	0.29	1	transcription factor	EGR1
txn	ETS1	ETS Proto-Oncogene 1, Transcription Factor	0	0.46	1	transcription factor	ETS1
txn	ETS2	ETS Proto-Oncogene 2, Transcription Factor	0	0.51	1	transcription factor	ETS2
txn	HIF1A	Hypoxia Inducible Factor 1 Subunit Alpha	0	0.45	1	transcription factor	HIF1A
txn	LEF1	LEF1	0	0.15	1	transcription factor	LEF1
txn	IKZF1	IKZF1	0	0.24	1	transcription factor	IKZF1
txn	KLF4	KLF4	0	0.39	1	transcription factor	KLF4
txn	MITF	Melanocyte Inducing Transcription Factor	0	0.53	1	transcription factor	MITF
txn	NR5A2	Nuclear Receptor Subfamily 5 Group A Member 2	0	0.46	1	transcription factor	NR5A2
txn	PPARA	Peroxisome Proliferator Activated Receptor Alpha	0	0.47	1	transcription factor	PPARA
txn	RARG	Retinoic Acid Receptor Gamma	0	0.23	1	transcription factor	RARG
txn	RUNX1	RUNX1	0	0.28	1	transcription factor	RUNX1
txn	RELA	RELA Proto-Oncogene, NF-KB Subunit	0	0.45	1	transcription factor	RELA
txn	TEAD4	TEA Domain Transcription Factor 4	0	0.36	1	transcription factor	TEAD4
txn	RUNX2	RUNX Family Transcription Factor 2	0	0.29	1	transcription factor	RUNX2
txn	TCF4	Transcription Factor 4	0	0.32	1	transcription factor	TCF4
txn	TFCP2L1	TFCP2L1	0	0.42	1	transcription factor	TFCP2L1
txn	WT1	WT1 Transcription Factor	0	0.21	1	transcription factor	WT1
txn	ZNF281	Zinc Finger Protein 281	0	0.46	1	transcription factor	ZNF281

6.2 Codes

Code C.3.1 Gene Regulatory Network Inference. Modified from Rogers et. al. [1] (.py)

```
"""
Scripts for inferring GRNs, including:
- Single inference
```

```
- k-fold cross validation
- Conversion to Netflux models
"""

import os
# import time
import pandas as pd
import numpy as np
import asyncio

from distributed import Client, LocalCluster
from arboreto.algo import grnboost2

from src.GRNrefinement import refineGRN
import src.GRNvalidation as gv

asyncio.set_event_loop_policy(asyncio.WindowsSelectorEventLoopPol-
icy())



def importData(filepath):
    data = pd.read_csv(filepath, index_col=0, header=1).dropna()
    return data


def importTFs(dirpath, libraryname, both=True):
    """
    Imports list of TFs from specified top-level directory
    (dirpath) and library string (libraryname). Optionally
    adds a second library to create composite list for
    wider coverage of TFs.
    """
    from src.GRNrefinement import getLibPath
    libextension = "_attribute_list_entries.txt.gz"
    libfile   =   getLibPath(dirpath,   libraryname,   filter_exten-
sion=libextension)
    tf_all = pd.read_table(libfile)
    if both:
        libname = "TRANSFACpredicted"
        libfile   =   getLibPath(dirpath,   libname,   filter_exten-
sion=libextension)
        tf_2 = pd.read_table(libfile)
        tf_all = pd.concat([tf_all, tf_2], axis=0).drop_duplicates()
    return tf_all


def processData(data, cutoff=1, num_samples=2):
    """Given a pandas dataframe, returns a transposed numpy array
    of values and associated gene names for input into grnboost2.
```

```
    Additionally applies a threshold to data as used for EdgeR.

    Note: assumes data is CPM data for thresholding purposes"""
    data_t = data.transpose()
    # apply threshold via EdgeR method
    data_threshold = data_t >= cutoff
    data_keep = data_t.loc[:, (data_threshold.sum(axis=0) >= num_sam-
ples).values]
    data_array = data_keep.values
    data_genes = data_keep.columns.values
    return data_array, data_genes


def saveGRN(grn, savedir, fold=None,
            trainingset=None, testingset=None,
            suffix=None):
    if suffix is not None:
        ext_csv = "_"+suffix+".csv"
        ext_txt = "_"+suffix+".txt"
    else:
        ext_csv = ".csv"
        ext_txt = ".txt"
    if fold is not None:
        filename_grn = "GRN_CV_fold"+str(fold)+ext_csv
        if trainingset and testingset is not None:
            filename_sets = "datasets_CV_fold"+str(fold)+ext_txt
            with open(savedir+filename_sets, "w") as output:
                output.write("_Training_\n")
                for train in trainingset[fold]:
                    output.write(train + "\n")
                output.write("_Testing_\n")
                for test in testingset[fold]:
                    output.write(test + "\n")
    else:
        filename_grn = "GRN_single"+ext_csv
    grn.to_csv(savedir+filename_grn)
    return



# runtime scripts
def inferGRN(filename,
            libpath, libname, lib_both=True,
            savedir=None, suffix=None, seed=None):
    """
    Top-level script for inferring gene regulatory network
    from a given dataset using the Arboreto GRNboost2 algorithm.
    :filename:  path to CSV file containing gene expression data.
    :libpath:   path to directory containing sub-folders for TF-target
                libraries.
    :libname:   string of TF-target library used for inference.
```

```
    :lib_both:  (optional) Boolean operator determining use of addi-
tional
                library (TRANSFACpredicted) for wider TF coverage
    :savedir:   (optional) path to directory for saving final CSV.
    :seed:      (optional) integer for inference algorithm seed
    """

    # import cpm + library data
    cpm = importData(filename)
    cpm_array, cpm_genes = processData(cpm)

    tf_all = importTFs(libpath, libname, lib_both)
    tf_names = tf_all["GeneSym"].to_list()


    # setup Dask cluster
    client = Client(LocalCluster())
    print(client.dashboard_link)



    # infer + refine GRN
    grn = grnboost2(expression_data=cpm_array,
                    gene_names=cpm_genes,
                    tf_names=tf_names,
                    client_or_address=client,
                    seed=seed)
    grn_refined = refineGRN(grn, libname, dir_path=libpath)

    if savedir is not None:
        saveGRN(grn_refined, savedir, suffix=suffix)

    client.shutdown()
    return grn_refined


def crossvalidateGRN(filename,
                    libpath, libname, k, lib_both=True,
                    savedir=None, suffix=None, seed=None):
    """
    Top-level script for k-fold cross validation of gene regulatory
    network inference using the Arboreto GRNboost2 algorithm.
    :filename:  path to CSV file containing gene expression data.
    :libpath:   path to directory containing sub-folders for TF-target
                libraries.
    :libname:   string of TF-target library used for inference.
    :k:         integer specifying number of folds for CV
    :lib_both:  (optional) Boolean operator determining use of addi-
tional
                library (TRANSFACpredicted) for wider TF coverage
    :savedir:   (optional) path to directory for saving final CSV.
```

```
    :seed:      (optional) integer for inference algorithm seed
    """

    # import cpm + library data
    cpm = importData(filename)

    tf_all = importTFs(libpath, libname, lib_both)
    tf_names = tf_all["GeneSym"].to_list()

    # create and assign CV folds
    folds = gv.makeFolds(cpm, k)
    training, testing = gv.assignFolds(folds)

    # setup Dask cluster
    client = Client(LocalCluster())
    print(client.dashboard_link)

    # infer + refine GRN for each fold
    fold = 0
    while fold < k:
        cpm_fold = cpm.loc[:, training[fold]]
        cpm_array, cpm_genes = processData(cpm_fold)

        grn = grnboost2(expression_data=cpm_array,
                        gene_names=cpm_genes,
                        tf_names=tf_names,
                        client_or_address=client,
                        seed=seed)
        grn_refined = refineGRN(grn, libname, dir_path=libpath)

        if savedir is not None:
            saveGRN(grn_refined, savedir, fold=fold, suffix=suffix,
                    trainingset=training, testingset=testing)

        # store all refined GRNs
        grn_refined["fold"] = fold
        if fold == 0:
            grn_all = grn_refined
        else:
            grn_all = grn_all.append(grn_refined)

        fold = fold + 1

    client.shutdown()
    return grn_all
```

Code C.3.2 Gene Regulatory Network Refinement. Modified from Rogers et. al. [1](.py)

```
"""
Functions for refining GRN as inferred from Arboreto
```

```
"""
import os
import pandas as pd
import numpy as np
# Library filtering functions
def    getLibPath(start_directory,    sub_directory,    filter_exten-
sion=None):
    """
    Using top-level and subdirectory, returns
    string of filename matching extension
    """
    for root, _, files in os.walk(start_directory+sub_directory):
        for file in files:
            if filter_extension is None or file.lower().endswith(fil-
ter_extension):
                return os.path.join(root, file)
def importLibraries(dirpath, libraryname, both=True):
    """
    Using filename of GRN, imports libraries of TF-target interac-
tions.
    :dirpath:   name of top-level directory containing library fold-
ers.
    :filename:  name of library to use
                    Library options: CHEA, TRANSFAC, ENCODE
    """
    libextension = "_gene_attribute_edges.txt.gz"
    libfile   =   getLibPath(dirpath,   libraryname,   filter_exten-
sion=libextension)
    library = pd.read_csv(libfile,
                          index_col=None, header=0,
                          low_memory=False, sep=r"\s+")
    if both:
        libname = "TRANSFACpredicted"
        libfile   =   getLibPath(dirpath,   libname,   filter_exten-
sion=libextension)
        library2 = pd.read_csv(libfile,
                          index_col=None, header=0,
                          low_memory=False, sep=r"\s+")
        library  =  pd.concat([library,library2],  axis=0).drop_dupli-
cates()
    return library.drop(index=0)
def filterWithLibrary(grn, library):
    """given a grn df and library df, finds edges matching
    the library and returns filtered grn with matching edges"""
    grn_pairs = grn["TF"]+"-"+grn["target"]
    lib_pairs = library["target"]+"-"+library["source"]
    inLibrary = grn_pairs.isin(lib_pairs)
    grn_filtered = grn.loc[inLibrary, :]
    return grn_filtered
# ========================
# Input-output filtering functions
```

```
def findOutputs(grn, *regs, indivregs=None):
    """Given regex strings for output gene families,
    finds outputs included in full GRN"""
    if indivregs is not None:
        reg_list = "|".join((regs) + indivregs)
    else:
        reg_list = "|".join((regs))
    outputs   =   grn.loc[grn["target"].str.contains(reg_list,   re-
gex=True).values,"target"].unique()
    return outputs
def findInputsOrOutputs(grn,values,column):
    """
    Given a list of gene names,
    finds names located in either TF or target columns in GRN
    """
    vals_any = []
    for val in values:
        val_any = grn[column].isin([val]).any()
        vals_any.append(val_any)
    vals_dict = dict(zip(values,vals_any))
    # filter dictionary for true items
    keys = []
    for item in vals_dict.items():
        if item[1] == True:
            keys.append(item[0])
    return keys
def filterInputsOrOutputs(grn,keys,column):
    """Given a list of gene names and the corresponding
    column, returns a filtered GRN containing only
    elements in the list."""
    for key in keys:
        grn_key = grn.loc[grn[column]==key,:]
        # remove 'index' column if necessary
        # (in order to avoid issues with concatenation)
        if grn.columns.isin(["index"]).any():
            grn_key = grn_key.drop(columns=["index"])
        if key == keys[0]:
            grn_filt = grn_key
        else:
            grn_filt = pd.concat([grn_filt, grn_key], axis=0)
    return grn_filt
def filterInOutNetwork(grn_in, grn_out, grn_targets, include_interme-
diates=False):
    """Given GRNs filtered for inputs TFs only, outputs targets only,
    and those included in libraries, function returns a subnetwork
    containing edges leading from inputs to outputs via intermediate
    TFs, with optional inclusion of indermediate (TF-TF) edges"""
    tfs = grn_targets["TF"].unique()
    in_tf = grn_in.loc[grn_in["target"].isin(tfs), :]
    tf_out = grn_out        # assumed b/c of TF column
    if include_intermediates:
```

```
        # find tf-tf edges connected to "input-tf" or "tf-output"
edges
        grn_tfs = grn_targets.loc[grn_targets["target"].isin(grn_tar-
gets["TF"].unique()), :]
        inclInput = grn_tfs["TF"].isin(in_tf["target"])
        inclOutput = grn_tfs["target"].isin(tf_out["TF"])
        tftf_in = grn_tfs.loc[inclInput, :]
        tftf_out = grn_tfs.loc[inclOutput, :]
        tftf = tftf_in.loc[tftf_in.index.isin(tftf_out.index), :]
        grn_filtered = pd.concat([in_tf, tftf, tf_out], axis=0)
    else:
        grn_filtered = pd.concat([in_tf, tf_out], axis=0)
    return grn_filtered
# ========================
# Simultaneous top-down + bottom-up DFS scripts
def testPathInv(path,grn,rules):
    """Test that a found path meets bottom-up search requirements"""
    meetsRules = []
    for item in range(len(path)):
        if item < len(path)-1:
            target = path[item]
            TF = path[item+1]
            toTest      =      grn.loc[(grn["target"]==target)      &
(grn["TF"]==TF),"importance"]
            neighbors_all = grn.loc[grn["target"]==target,:]
            quant = neighbors_all["importance"].quantile(q=rules[1])
            if   ((toTest   >=   rules[0]).bool())   |   ((toTest   >=
quant).bool()):
                meetsRules.extend([True])
            else:
                meetsRules.extend([False])
    return all(meetsRules)
def findPathsBoth(grn,start,output_keys,rules=[1,0.5]):
    """
    Given a network of pairwise TF-target interactions, a
    starting TF, and a set of output genes, 1) uses a top-down
    search algorithm to find paths between the input and outputs,
    and 2) checks that found paths meets the same rules for a
    bottom-up search.
    """
    stack = [(start,[start])]
    while stack:
        (vertex,path) = stack.pop()
        neighbors_all = grn.loc[grn["TF"]==vertex,:]
        quant = neighbors_all["importance"].quantile(q=rules[1])
        toKeep  =  ((neighbors_all["importance"]>rules[0])  |  (neigh-
bors_all["importance"]>quant))
        neighbors = neighbors_all.loc[toKeep,"target"]
        for neigh in neighbors:
            if neigh not in path:
                if neigh in output_keys:
```

```
                    meetsInvRules = testPathInv(list(reversed(path +
[neigh])),grn,rules)
                    if meetsInvRules:
                        yield path + [neigh]
                else:
                    stack.append((neigh, path + [neigh]))
def findPathImps(grn,pathlist):
    """

    """
    tot_imp = []
    vars_imp = []
    for paths in pathlist:
        path_imp = []
        for item in range(len(paths)):
            if item < len(paths)-1:
                TF = paths[item]
                target = paths[item+1]
                criteria = ((grn["TF"] == TF) & (grn["target"] == tar-
get))
                imp = grn.loc[criteria, "importance"]
                if item == 0:
                    path_imp = [float(imp)]
                else:
                    path_imp = path_imp + [float(imp)]
        tot_imp.extend([np.sum(path_imp)])
        vars_imp.extend([np.std(path_imp)])
    paths_imp = pd.DataFrame(data=[pathlist, tot_imp, vars_imp],
                             index=["path", "importance_total", "im-
portance_sd"]).transpose()

    # add additional metrics + metadata
    paths_imp["importance_mean"]  =  paths_imp["importance_total"]  /
paths_imp["path"].str.len()
    paths_imp["importance_cv"]    =    paths_imp["importance_sd"]    /
paths_imp["importance_mean"]
    paths_str = []
    for _, series in paths_imp.iterrows():
        paths_str.extend(['->'.join('->'+str(i)    for    i    in    se-
ries["path"])[2:]])
    paths_imp["path_string"] = paths_str
    paths_imp["input"] = paths_imp["path"].str[0]
    paths_imp["output"] = paths_imp["path"].str[-1]
    paths_imp["TF"] = paths_imp["path"].str[-2]
    return paths_imp
def findPathRows(grn,pathlist):
    """Extract interaction pairs from paths"""
    idxs = []
    for paths in pathlist:
        for item in range(len(paths)):
            if item < len(paths)-1:
```

```
                TF = paths[item]
                target = paths[item+1]
                criteria = ((grn["TF"]==TF) & (grn["target"]==tar-
get))
                idx = grn.loc[criteria,:].index
                idxs.extend(idx)
    paths_idx = grn.loc[idxs,:].sort_values("importance", ascend-
ing=False).drop_duplicates()
    return paths_idx
# ========================
# Runtime function
def refineGRN(grn,
            libraryname,
            dir_path="D:\\Research\\Aim3\\data_TFdatabases\\",
            lib_both=True,
            output_regex=False):
    """
    Top-level function for GRN refinement.
    :grn:           n x 3 pandas dataframe containing "TF", "target"
and "importance"
    :libraryname:    string containing TF-target database used for
pruning
                        Current options:      "CHEA", "TRANSFACpre-
dicted",
                                              "TRANSFACcurated",  "EN-
CODE"
    :dir_path:        string containing top-level directory for all
databases,
                      libraries should be located in folders matching
libraryname
    :lib_both:       Boolean determining 'both' argument in importLi-
braries fcn
    :output_regex:  Boolean determining whether regex should be used
in
                      choosing gene outputs (i.e. gene families)
    :grn_final:     n x 3 pandas datafram containing refined edges
    """
    # import df
    # db_path = "CHEA\\"
    # file_name = "CHEA_both_GRN_09012020_EdgeR.csv"
    #  grn = pd.read_csv(dir_path+db_path+file_name, header=0, in-
dex_col=0)
    # grn = grn.reset_index()
    # print(grn.shape)
    # filter for edges contained in librar(ies)
    library = importLibraries(dir_path, libraryname, both=lib_both)
    grn_targets = filterWithLibrary(grn, library)
    print(grn_targets.shape)
    # filter for edges connected to desired inputs/outputs
    if output_regex:
        reg_col = r"^COL\d{1,2}A\d$"
```

```
        reg_mmp = r"^MMP\d"
        reg_timp = r"TIMP\d"
        reg_cts = r"CTS+[A-Z]"
        reg_tgfb = r"TGFB\d$"
        reg_thbs = r"THBS\d"
        reg_lox = r"^LOX"
        reg_others = ("SPP1","POSTN","^FN1","SPARC$",
                      "TNC","CTGF","SERPINE1","ACTA2",
                        "HBA1","HBA2","HBB","SFRP4","PENK","COL22A1",
                        "LGALS2","FNDC1","MYH6")
        outputs = findOutputs(grn,
                              reg_col, reg_mmp, reg_timp, reg_cts,
                              reg_tgfb, reg_thbs, reg_lox,
                              indivregs=reg_others)
    else:
        outputs     =     ["CTGF","FN1","ACTA2","TIMP1","TIMP2","SER-
PINE1","MMP12",

"MMP14","MMP1","MMP2","MMP3","MMP8","MMP9","POSTN","COL1A1",
                "COL1A2","COL3A1","TNC","THBS4","SPP1"]

    inputs                                                          =
["STAT1","STAT3","JUN","FOS","NFKB1","RELA","CREB1","CREBBP",

"SMAD3","MYC","NFATC1","NFATC3","SRF","TEAD2","TEAD4","YAP1","WWTR1"
]

    input_keys = findInputsOrOutputs(grn_targets, inputs, "TF")
    output_keys = findInputsOrOutputs(grn_targets, outputs, "target")
    grn_outputs  =  filterInputsOrOutputs(grn_targets,  output_keys,
"target")
    grn_inputs = filterInputsOrOutputs(grn_targets, input_keys, "TF")
    paths_all_inout  =  filterInOutNetwork(grn_inputs,  grn_outputs,
grn_targets)
    print(paths_all_inout.shape)
    # find input-output paths via modified DFS algorithm
    paths_found_bothsearch = []
    for inp in input_keys:
        paths_found_bothsearch.extend(list(find-
PathsBoth(paths_all_inout,
                                                        inp,    out-
put_keys,

rules=[1,0.90])))
    paths_found_bothsearch_imp = findPathImps(paths_all_inout,
                                              paths_found_bothsearch)
    grn_final = findPathRows(paths_all_inout.reset_index(),
                             paths_found_bothsearch_imp["path"])
    print(grn_final.shape)
    return grn_final
```

Code C.3.3 Gene Regulatory Network Validation. Modified from Rogers et. al. [1] (.py)

```
"""
Scripts for generating cross validation folds and datasets
"""

import random

def makeFolds(data, k):
    """
    Given a dataframe of cpm values, randomly
    creates k folds and returns nested list of
    column names.
    """
    # randomize columns
    order = data.columns.tolist()
    random.shuffle(order)
    # split into folds (specified by k)
    folds = []
    fold = 0
    dist = len(order) / k
    while fold < k:
        start = int(round(fold * dist))
        end = int(round(start + dist))
        folds.append(order[start:end])
        fold = fold + 1
    return folds

def assignFolds(folds):
    """
    Given a nested list of strings from makeFolds,
    assigns testing sets from each list(fold) and
    assigns rest to training set for each fold. K
    is determined from length of nested list as the
    number of folds.
    """
    training = []
    testing = []
    k = len(folds)
    fold=0
    while fold < k:
        testing.append(folds.pop(fold))
        training.append([y for x in folds for y in x])
        folds.insert(fold, testing[fold])
        fold = fold+1
    return training, testing
```

Code C.3.4 Gene Regulatory Network Execution in Dask (.py)

```
from src.GRNinference import inferGRN, crossvalidateGRN
```

```
# Specify data and library arguments:
# - `path_to_data`: address of CSV file containing gene expression
data (formatted as genes x samples)
# - `lib_dir`: address of directory containing sub-directories for all
TF-target databases (included in repository as `data`)
# - `lib_name`: string specifying the desired library to use for
inference and refinement
#   - Here, the [CHEA](https://pubmed.ncbi.nlm.nih.gov/20709693/) da-
tabase of transcription factor targets is used
path_to_data    =    "data\\expression\\CPMS_SVA_corrected_geneid_re-
paired_06022023.csv"
lib_dir = "data\\"
lib_name = "CHEA"
k = 10
path_to_save = "data\\networks\\"
grn_all  =  crossvalidateGRN(path_to_data,  lib_dir,  lib_name,  k,
savedir=path_to_save)
grn_all.to_csv('grn_all_MAGNet_DE.csv', index=False)
```

Code C.4.1 Logic based ODE Model (.m)

```
function dydt=NetfluxODE_DCM_07122023(t,y,params)
% NetfluxODE_DCM_07122023.m

% Assign names for parameters
[rpar,tau,ymax,speciesNames]=params{:};
AngII = 1;
AT1R = 2;
AGT = 3;
ACE = 4;
NOX = 5;
ROS = 6;
ET1 = 7;
ETAR = 8;
DAG = 9;
PKC = 10;
TRPC = 11;
NE = 12;
BAR = 13;
AC = 14;
cAMP = 15;
PKA = 16;
CREB = 17;
CBP = 18;
TGFB = 19;
TGFB1R = 20;
smad3 = 21;
smad7 = 22;
latentTGFB = 23;
BAMBI = 24;
```

```
PDGF = 25;
PDGFR = 26;
NP = 27;
NPRA = 28;
cGMP = 29;
PKG = 30;
tension = 31;
B1int = 32;
Rho = 33;
ROCK = 34;
Ca = 35;
calcineurin = 36;
NFAT = 37;
IL6 = 38;
gp130 = 39;
STAT = 40;
IL1 = 41;
IL1RI = 42;
TNFa = 43;
TNFaR = 44;
NFKB = 45;
PI3K = 46;
Akt = 47;
p38 = 48;
TRAF = 49;
ASK1 = 50;
MKK3 = 51;
PP1 = 52;
JNK = 53;
abl = 54;
Rac1 = 55;
MEKK1 = 56;
MKK4 = 57;
ERK = 58;
Ras = 59;
Raf = 60;
MEK1 = 61;
FAK = 62;
epac = 63;
Factin = 64;
FA = 65;
cmyc = 66;
CTGF = 67;
proliferation = 68;
SRF = 69;
EDAFN = 70;
aSMA = 71;
AP1 = 72;
TIMP1 = 73;
TIMP2 = 74;
PAI1 = 75;
```

```
proMMP14 = 76;
proMMP1 = 77;
proMMP2 = 78;
proMMP9 = 79;
fibronectin = 80;
periostin = 81;
proCI = 82;
proCIII = 83;
B3int = 84;
Src = 85;
Grb2 = 86;
p130Cas = 87;
YAP = 88;
MRTF = 89;
Gactin = 90;
TNC = 91;
mTORC1 = 92;
mTORC2 = 93;
p70S6K = 94;
EBP1 = 95;
syndecan4 = 96;
proMMP3 = 97;
proMMP8 = 98;
proMMP12 = 99;
thrombospondin4 = 100;
osteopontin = 101;
contractility = 102;
RhoGEF = 103;
RhoGDI = 104;
talin = 105;
vinculin = 106;
paxillin = 107;
MLC = 108;
AT2R = 109;
BCL6 = 110;
CACYBP = 111;
CEBPD = 112;
CUX1 = 113;
EGR1 = 114;
ETS1 = 115;
ETS2 = 116;
HIF1A = 117;
LEF1 = 118;
IKZF1 = 119;
KLF4 = 120;
MITF = 121;
NR5A2 = 122;
PPARA = 123;
RARG = 124;
RUNX1 = 125;
RELA = 126;
```

```
TEAD4 = 127;
RUNX2 = 128;
TCF4 = 129;
TFCP2L1 = 130;
WT1 = 131;
ZNF281 = 132;
dydt = zeros(132,1);
dydt(AngII)                                                              =
(OR(rpar(1,1),AND(rpar(:,13),act(y(AGT),rpar(:,13)),act(y(ACE),rpar(
:,13))))*ymax(AngII) - y(AngII))/tau(AngII);
dydt(AT1R)          =          (OR(act(y(AngII),rpar(:,18)),act(y(ten-
sion),rpar(:,168)))*ymax(AT1R) - y(AT1R))/tau(AT1R);
dydt(AGT)                      =                     (AND(rpar(:,27),in-
hib(y(AT1R),rpar(:,27)),act(y(p38),rpar(:,27)),in-
hib(y(JNK),rpar(:,27)))*ymax(AGT) - y(AGT))/tau(AGT);
dydt(ACE) = (act(y(TGFB1R),rpar(:,50))*ymax(ACE) - y(ACE))/tau(ACE);
dydt(NOX)                                                                =
(OR(act(y(AT1R),rpar(:,19)),act(y(TGFB1R),rpar(:,99)))*ymax(NOX)    -
y(NOX))/tau(NOX);
dydt(ROS)                                                                =
(OR(act(y(NOX),rpar(:,20)),act(y(ETAR),rpar(:,37)))*ymax(ROS)       -
y(ROS))/tau(ROS);
dydt(ET1)    =    (OR(rpar(1,9),act(y(AP1),rpar(:,17)))*ymax(ET1)    -
y(ET1))/tau(ET1);
dydt(ETAR) = (act(y(ET1),rpar(:,56))*ymax(ETAR) - y(ETAR))/tau(ETAR);
dydt(DAG)                                                                =
(OR(act(y(ETAR),rpar(:,113)),act(y(AT1R),rpar(:,114)))*ymax(DAG)    -
y(DAG))/tau(DAG);
dydt(PKC)                                                                =
(OR(act(y(syndecan4),rpar(:,133)),AND(rpar(:,149),act(y(DAG),rpar(:,
149)),act(y(mTORC2),rpar(:,149))))*ymax(PKC) - y(PKC))/tau(PKC);
dydt(TRPC)          =          (OR(act(y(DAG),rpar(:,115)),act(y(ten-
sion),rpar(:,167)))*ymax(TRPC) - y(TRPC))/tau(TRPC);
dydt(NE) = (rpar(1,7)*ymax(NE) - y(NE))/tau(NE);
dydt(BAR) = (act(y(NE),rpar(:,55))*ymax(BAR) - y(BAR))/tau(BAR);
dydt(AC)                                                                 =
(OR(act(y(BAR),rpar(:,64)),AND(rpar(:,65),act(y(AT1R),rpar(:,65)),ac
t(y(BAR),rpar(:,65))))*ymax(AC) - y(AC))/tau(AC);
dydt(cAMP) = (act(y(AC),rpar(:,66))*ymax(cAMP) - y(cAMP))/tau(cAMP);
dydt(PKA) = (act(y(cAMP),rpar(:,44))*ymax(PKA) - y(PKA))/tau(PKA);
dydt(CREB) = (act(y(PKA),rpar(:,53))*ymax(CREB) - y(CREB))/tau(CREB);
dydt(CBP)            =            (OR(inhib(y(smad3),rpar(:,46)),in-
hib(y(CREB),rpar(:,47)))*ymax(CBP) - y(CBP))/tau(CBP);
dydt(TGFB)                                                               =
(OR(rpar(1,2),OR(AND(rpar(:,11),act(y(latentTGFB),rpar(:,11)),act(y(
proMMP9),rpar(:,11))),AND(rpar(:,12),act(y(latentTGFB),rpar(:,12)),a
ct(y(proMMP2),rpar(:,12)))))*ymax(TGFB) - y(TGFB))/tau(TGFB);
dydt(TGFB1R)      =      (AND(rpar(:,51),act(y(TGFB),rpar(:,51)),in-
hib(y(BAMBI),rpar(:,51)))*ymax(TGFB1R) - y(TGFB1R))/tau(TGFB1R);
```

```
dydt(smad3)    =    (OR(AND(rpar(:,28),act(y(TGFB1R),rpar(:,28)),in-
hib(y(smad7),rpar(:,28)),in-
hib(y(PKG),rpar(:,28))),act(y(Akt),rpar(:,144)))*ymax(smad3)       -
y(smad3))/tau(smad3);
dydt(smad7)                                                        =
(OR(act(y(STAT),rpar(:,104)),AND(rpar(:,158),act(y(AP1),rpar(:,158))
,inhib(y(YAP),rpar(:,158))))*ymax(smad7) - y(smad7))/tau(smad7);
dydt(latentTGFB)   =    (act(y(AP1),rpar(:,68))*ymax(latentTGFB)   -
y(latentTGFB))/tau(latentTGFB);
dydt(BAMBI)                                                        =
(AND(rpar(:,103),act(y(TGFB),rpar(:,103)),act(y(IL1RI),rpar(:,103)))
*ymax(BAMBI) - y(BAMBI))/tau(BAMBI);
dydt(PDGF)  =  (OR(rpar(1,8),act(y(EGR1),rpar(:,259)))*ymax(PDGF)  -
y(PDGF))/tau(PDGF);
dydt(PDGFR)      =      (act(y(PDGF),rpar(:,63))*ymax(PDGFR)       -
y(PDGFR))/tau(PDGFR);
dydt(NP) = (rpar(1,10)*ymax(NP) - y(NP))/tau(NP);
dydt(NPRA) = (act(y(NP),rpar(:,72))*ymax(NPRA) - y(NPRA))/tau(NPRA);
dydt(cGMP)       =       (act(y(NPRA),rpar(:,73))*ymax(cGMP)       -
y(cGMP))/tau(cGMP);
dydt(PKG) = (act(y(cGMP),rpar(:,74))*ymax(PKG) - y(PKG))/tau(PKG);
dydt(tension)                                                      =
(OR(rpar(1,3),AND(rpar(:,165),act(y(FA),rpar(:,165)),act(y(contrac-
tility),rpar(:,165))))*ymax(tension) - y(tension))/tau(tension);
dydt(B1int)  =  (OR(AND(rpar(:,43),act(y(PKC),rpar(:,43)),act(y(ten-
sion),rpar(:,43))),act(y(tension),rpar(:,48)))*ymax(B1int)         -
y(B1int))/tau(B1int);
dydt(Rho)    =    (OR(act(y(TGFB1R),rpar(:,118)),AND(rpar(:,131),in-
hib(y(PKG),rpar(:,131)),act(y(RhoGEF),rpar(:,131)),in-
hib(y(RhoGDI),rpar(:,131))))*ymax(Rho) - y(Rho))/tau(Rho);
dydt(ROCK) = (act(y(Rho),rpar(:,70))*ymax(ROCK) - y(ROCK))/tau(ROCK);
dydt(Ca) = (act(y(TRPC),rpar(:,116))*ymax(Ca) - y(Ca))/tau(Ca);
dydt(calcineurin)  =  (act(y(Ca),rpar(:,117))*ymax(calcineurin)   -
y(calcineurin))/tau(calcineurin);
dydt(NFAT)       =       (OR(act(y(calcineurin),rpar(:,109)),OR(in-
hib(y(cmyc),rpar(:,198)),OR(inhib(y(STAT),rpar(:,226)),OR(in-
hib(y(TEAD4),rpar(:,240)),OR(act(y(cmyc),rpar(:,281)),OR(act(y(STAT)
,rpar(:,314)),act(y(TEAD4),rpar(:,326)))))))*ymax(NFAT)            -
y(NFAT))/tau(NFAT);
dydt(IL6)                                                          =
(OR(rpar(1,4),OR(AND(rpar(:,14),act(y(CREB),rpar(:,14)),act(y(CBP),r
par(:,14))),OR(act(y(NFKB),rpar(:,15)),OR(act(y(AP1),rpar(:,16)),in-
hib(y(EGR1),rpar(:,181))))))*ymax(IL6) - y(IL6))/tau(IL6);
dydt(gp130)      =      (act(y(IL6),rpar(:,21))*ymax(gp130)       -
y(gp130))/tau(gp130);
dydt(STAT)      =      (act(y(gp130),rpar(:,25))*ymax(STAT)       -
y(STAT))/tau(STAT);
dydt(IL1) = (rpar(1,5)*ymax(IL1) - y(IL1))/tau(IL1);
dydt(IL1RI)      =      (act(y(IL1),rpar(:,58))*ymax(IL1RI)       -
y(IL1RI))/tau(IL1RI);
```

```
dydt(TNFa)   =   (OR(rpar(1,6),act(y(EGR1),rpar(:,261)))*ymax(TNFa)   -
y(TNFa))/tau(TNFa);
dydt(TNFaR)       =       (act(y(TNFa),rpar(:,71))*ymax(TNFaR)       -
y(TNFaR))/tau(TNFaR);
dydt(NFKB)                                                            =
(OR(act(y(IL1RI),rpar(:,24)),OR(act(y(ERK),rpar(:,34)),OR(act(y(p38)
,rpar(:,35)),OR(act(y(Akt),rpar(:,100)),OR(in-
hib(y(BCL6),rpar(:,177)),OR(in-
hib(y(YAP),rpar(:,242)),OR(act(y(RELA),rpar(:,295)),act(y(YAP),rpar(
:,329)))))))))*ymax(NFKB) - y(NFKB))/tau(NFKB);
dydt(PI3K)                                                            =
(OR(act(y(TNFaR),rpar(:,26)),OR(act(y(TGFB1R),rpar(:,96)),OR(act(y(P
DGFR),rpar(:,97)),act(y(FAK),rpar(:,98)))))*ymax(PI3K)                -
y(PI3K))/tau(PI3K);
dydt(Akt)                                                             =
(AND(rpar(:,148),act(y(PI3K),rpar(:,148)),act(y(mTORC2),rpar(:,148))
)*ymax(Akt) - y(Akt))/tau(Akt);
dydt(p38)                                                             =
(OR(act(y(ROS),rpar(:,22)),OR(act(y(MKK3),rpar(:,79)),OR(act(y(Ras),
rpar(:,95)),AND(rpar(:,106),act(y(Rho),rpar(:,106)),in-
hib(y(Rac1),rpar(:,106))))))*ymax(p38) - y(p38))/tau(p38);
dydt(TRAF)                                                            =
(OR(act(y(TGFB1R),rpar(:,80)),act(y(TNFaR),rpar(:,88)))*ymax(TRAF)  -
y(TRAF))/tau(TRAF);
dydt(ASK1)                                                            =
(OR(act(y(TRAF),rpar(:,89)),act(y(IL1RI),rpar(:,92)))*ymax(ASK1)    -
y(ASK1))/tau(ASK1);
dydt(MKK3)       =       (act(y(ASK1),rpar(:,90))*ymax(MKK3)        -
y(MKK3))/tau(MKK3);
dydt(PP1) = (act(y(p38),rpar(:,78))*ymax(PP1) - y(PP1))/tau(PP1);
dydt(JNK)     =     (OR(act(y(ROS),rpar(:,23)),OR(AND(rpar(:,83),in-
hib(y(NFKB),rpar(:,83)),act(y(MKK4),rpar(:,83))),AND(rpar(:,107),in-
hib(y(Rho),rpar(:,107)),act(y(MKK4),rpar(:,107)))))*ymax(JNK)       -
y(JNK))/tau(JNK);
dydt(abl) = (act(y(PDGFR),rpar(:,84))*ymax(abl) - y(abl))/tau(abl);
dydt(Rac1)                                                            =
(OR(act(y(abl),rpar(:,85)),AND(rpar(:,128),act(y(abl),rpar(:,128)),a
ct(y(p130Cas),rpar(:,128))))*ymax(Rac1) - y(Rac1))/tau(Rac1);
dydt(MEKK1)                                                           =
(OR(act(y(FAK),rpar(:,67)),act(y(Rac1),rpar(:,81)))*ymax(MEKK1)     -
y(MEKK1))/tau(MEKK1);
dydt(MKK4)                                                            =
(OR(act(y(MEKK1),rpar(:,82)),act(y(ASK1),rpar(:,91)))*ymax(MKK4)    -
y(MKK4))/tau(MKK4);
dydt(ERK)                    =                    (OR(AND(rpar(:,77),in-
hib(y(PP1),rpar(:,77)),act(y(MEK1),rpar(:,77))),AND(rpar(:,172),act(
y(ROS),rpar(:,172)),inhib(y(AT2R),rpar(:,172))))*ymax(ERK)           -
y(ERK))/tau(ERK);
dydt(Ras)                                                             =
(OR(act(y(AT1R),rpar(:,110)),act(y(Grb2),rpar(:,122)))*ymax(Ras)    -
y(Ras))/tau(Ras);
```

```
dydt(Raf) = (act(y(Ras),rpar(:,75))*ymax(Raf) - y(Raf))/tau(Raf);
dydt(MEK1)                    =                    (AND(rpar(:,76),in-
hib(y(ERK),rpar(:,76)),act(y(Raf),rpar(:,76)))*ymax(MEK1)           -
y(MEK1))/tau(MEK1);
dydt(FAK) = (act(y(B1int),rpar(:,120))*ymax(FAK) - y(FAK))/tau(FAK);
dydt(epac)       =       (act(y(cAMP),rpar(:,69))*ymax(epac)       -
y(epac))/tau(epac);
dydt(Factin)  =  (AND(rpar(:,136),act(y(ROCK),rpar(:,136)),act(y(Gac-
tin),rpar(:,136)))*ymax(Factin) - y(Factin))/tau(Factin);
dydt(FA) = (AND(rpar(:,160),act(y(vinculin),rpar(:,160)),inhib(y(pax-
illin),rpar(:,160)))*ymax(FA) - y(FA))/tau(FA);
dydt(cmyc)                                                          =
(OR(act(y(JNK),rpar(:,86)),OR(act(y(BCL6),rpar(:,250)),act(y(ETS2),r
par(:,264))))*ymax(cmyc) - y(cmyc))/tau(cmyc);
dydt(CTGF)                                                          =
(OR(AND(rpar(:,29),act(y(CBP),rpar(:,29)),act(y(smad3),rpar(:,29)),a
ct(y(ERK),rpar(:,29))),OR(act(y(YAP),rpar(:,132)),OR(in-
hib(y(BCL6),rpar(:,176)),OR(inhib(y(CEBPD),rpar(:,179)),OR(in-
hib(y(CUX1),rpar(:,180)),OR(inhib(y(RUNX1),rpar(:,210)),OR(in-
hib(y(RUNX2),rpar(:,215)),OR(inhib(y(smad3),rpar(:,217)),OR(in-
hib(y(STAT),rpar(:,221)),OR(inhib(y(TCF4),rpar(:,235)),OR(in-
hib(y(TEAD4),rpar(:,238)),OR(act(y(BCL6),rpar(:,249)),OR(act(y(CE-
BPD),rpar(:,253)),OR(act(y(CUX1),rpar(:,256)),OR(act(y(RUNX1),rpar(:
,297)),OR(act(y(RUNX2),rpar(:,302)),OR(act(y(smad3),rpar(:,304)),OR(
act(y(TCF4),rpar(:,323)),act(y(TEAD4),rpar(:,324))))))))))))))))))))
*ymax(CTGF) - y(CTGF))/tau(CTGF);
dydt(proliferation)                                                 =
(OR(act(y(AP1),rpar(:,52)),OR(act(y(CREB),rpar(:,54)),OR(act(y(CTGF)
,rpar(:,57)),OR(act(y(PKC),rpar(:,59)),OR(act(y(cmyc),rpar(:,87)),AN
D(rpar(:,143),act(y(p70S6K),rpar(:,143)),in-
hib(y(EBP1),rpar(:,143)))))))))*ymax(proliferation)   -   y(prolifera-
tion))/tau(proliferation);
dydt(SRF) = (act(y(MRTF),rpar(:,138))*ymax(SRF) - y(SRF))/tau(SRF);
dydt(EDAFN)       =       (act(y(NFAT),rpar(:,49))*ymax(EDAFN)       -
y(EDAFN))/tau(EDAFN);
dydt(aSMA)                                                          =
(OR(AND(rpar(:,111),act(y(CBP),rpar(:,111)),act(y(smad3),rpar(:,111)
)),OR(act(y(SRF),rpar(:,112)),act(y(YAP),rpar(:,170))))*ymax(aSMA)  -
y(aSMA))/tau(aSMA);
dydt(AP1)                                                           =
(OR(act(y(ERK),rpar(:,38)),OR(act(y(JNK),rpar(:,102)),OR(in-
hib(y(cmyc),rpar(:,195)),OR(in-
hib(y(TEAD4),rpar(:,239)),OR(act(y(abl),rpar(:,246)),OR(act(y(CREB),
rpar(:,254)),OR(act(y(MITF),rpar(:,273)),OR(act(y(cmyc),rpar(:,278))
,OR(act(y(RUNX1),rpar(:,296)),OR(act(y(SRF),rpar(:,306)),OR(act(y(ST
AT),rpar(:,307)),OR(act(y(TEAD4),rpar(:,325)),act(y(WT1),rpar(:,328)
))))))))))))))*ymax(AP1) - y(AP1))/tau(AP1);
dydt(TIMP1)       =       (act(y(AP1),rpar(:,41))*ymax(TIMP1)       -
y(TIMP1))/tau(TIMP1);
```

```
dydt(TIMP2)            =            (OR(act(y(AP1),rpar(:,42)),OR(in-
hib(y(EGR1),rpar(:,183)),OR(inhib(y(MITF),rpar(:,192)),OR(act(y(CA-
CYBP),rpar(:,252)),OR(act(y(EGR1),rpar(:,260)),act(y(MITF),rpar(:,27
5)))))))*ymax(TIMP2) - y(TIMP2))/tau(TIMP2);
dydt(PAI1)                                                          =
(OR(act(y(smad3),rpar(:,93)),OR(act(y(YAP),rpar(:,150)),OR(in-
hib(y(HIF1A),rpar(:,185)),OR(inhib(y(cmyc),rpar(:,201)),OR(in-
hib(y(RUNX1),rpar(:,213)),OR(inhib(y(RUNX2),rpar(:,216)),OR(in-
hib(y(SRF),rpar(:,220)),OR(inhib(y(STAT),rpar(:,229)),OR(in-
hib(y(TFCP2L1),rpar(:,241)),OR(in-
hib(y(ZNF281),rpar(:,245)),OR(act(y(HIF1A),rpar(:,265)),OR(act(y(cmy
c),rpar(:,284)),OR(act(y(PPARA),rpar(:,291)),OR(act(y(RUNX1),rpar(:,
300)),OR(act(y(RUNX2),rpar(:,303)),OR(act(y(STAT),rpar(:,317)),OR(ac
t(y(TFCP2L1),rpar(:,327)),act(y(ZNF281),rpar(:,332)))))))))))))))))
)*ymax(PAI1) - y(PAI1))/tau(PAI1);
dydt(proMMP14)                                                      =
(OR(act(y(AP1),rpar(:,62)),OR(act(y(NFKB),rpar(:,94)),act(y(EGR1),rp
ar(:,257))))*ymax(proMMP14) - y(proMMP14))/tau(proMMP14);
dydt(proMMP1)                  =                  (AND(rpar(:,36),in-
hib(y(smad3),rpar(:,36)),act(y(NFKB),rpar(:,36)),act(y(AP1),rpar(:,3
6)))*ymax(proMMP1) - y(proMMP1))/tau(proMMP1);
dydt(proMMP2)                                                       =
(OR(act(y(STAT),rpar(:,30)),OR(act(y(AP1),rpar(:,39)),OR(in-
hib(y(EGR1),rpar(:,182)),OR(inhib(y(ETS2),rpar(:,184)),OR(in-
hib(y(RARG),rpar(:,208)),OR(in-
hib(y(STAT),rpar(:,225)),OR(act(y(EGR1),rpar(:,258)),OR(act(y(ETS2),
rpar(:,262)),act(y(RARG),rpar(:,293)))))))))*ymax(proMMP2)        -
y(proMMP2))/tau(proMMP2);
dydt(proMMP9)                                                       =
(OR(act(y(STAT),rpar(:,31)),OR(AND(rpar(:,40),act(y(NFKB),rpar(:,40)
),act(y(AP1),rpar(:,40))),OR(in-
hib(y(PPARA),rpar(:,206)),OR(act(y(ETS2),rpar(:,263)),act(y(IKZF1),r
par(:,266))))))*ymax(proMMP9) - y(proMMP9))/tau(proMMP9);
dydt(fibronectin)                                                   =
(OR(AND(rpar(:,45),act(y(CBP),rpar(:,45)),act(y(smad3),rpar(:,45))),
act(y(NFKB),rpar(:,101)))*ymax(fibronectin) - y(fibronectin))/tau(fi-
bronectin);
dydt(periostin)                                                     =
(OR(AND(rpar(:,32),act(y(CBP),rpar(:,32)),act(y(smad3),rpar(:,32))),
OR(AND(rpar(:,33),act(y(CREB),rpar(:,33)),act(y(CBP),rpar(:,33))),OR
(inhib(y(BCL6),rpar(:,178)),OR(inhib(y(LEF1),rpar(:,190)),OR(in-
hib(y(STAT),rpar(:,227)),OR(act(y(BCL6),rpar(:,251)),OR(act(y(LEF1),
rpar(:,271)),act(y(STAT),rpar(:,315)))))))))*ymax(periostin) - y(per-
iostin))/tau(periostin);
dydt(proCI)                                                         =
(OR(AND(rpar(:,60),act(y(CBP),rpar(:,60)),act(y(smad3),rpar(:,60)),i
nhib(y(epac),rpar(:,60))),OR(act(y(SRF),rpar(:,105)),OR(in-
hib(y(BCL6),rpar(:,174)),OR(inhib(y(KLF4),rpar(:,188)),OR(in-
hib(y(LEF1),rpar(:,189)),OR(inhib(y(RARG),rpar(:,207)),OR(in-
hib(y(ZNF281),rpar(:,244)),OR(act(y(BCL6),rpar(:,247)),OR(act(y(KLF4
),rpar(:,269)),OR(act(y(LEF1),rpar(:,270)),OR(act(y(RARG),rpar(:,292
```

```
)),act(y(ZNF281),rpar(:,331))))))))))))))*ymax(proCI)                -
y(proCI))/tau(proCI);
dydt(proCIII)                                                        =
(OR(AND(rpar(:,61),act(y(CBP),rpar(:,61)),act(y(smad3),rpar(:,61)),i
nhib(y(epac),rpar(:,61))),OR(act(y(SRF),rpar(:,108)),OR(in-
hib(y(BCL6),rpar(:,175)),OR(in-
hib(y(NR5A2),rpar(:,205)),OR(act(y(BCL6),rpar(:,248)),act(y(NR5A2),r
par(:,290)))))))*ymax(proCIII) - y(proCIII))/tau(proCIII);
dydt(B3int)   =   (OR(AND(rpar(:,152),act(y(tension),rpar(:,152)),in-
hib(y(thrombospondin4),rpar(:,152))),act(y(osteopon-
tin),rpar(:,156)))*ymax(B3int) - y(B3int))/tau(B3int);
dydt(Src)                                                            =
(OR(act(y(B3int),rpar(:,119)),act(y(PDGFR),rpar(:,126)))*ymax(Src)  -
y(Src))/tau(Src);
dydt(Grb2)                                                           =
(AND(rpar(:,121),act(y(FAK),rpar(:,121)),act(y(Src),rpar(:,121)))*ym
ax(Grb2) - y(Grb2))/tau(Grb2);
dydt(p130Cas)                                                        =
(OR(AND(rpar(:,125),act(y(FAK),rpar(:,125)),act(y(Src),rpar(:,125)))
,AND(rpar(:,127),act(y(ten-
sion),rpar(:,127)),act(y(Src),rpar(:,127))))*ymax(p130Cas)          -
y(p130Cas))/tau(p130Cas);
dydt(YAP)                        =                        (OR(act(y(Fac-
tin),rpar(:,129)),act(y(AT1R),rpar(:,169)))*ymax(YAP)               -
y(YAP))/tau(YAP);
dydt(MRTF)  =  (AND(rpar(:,135),act(y(NFAT),rpar(:,135)),inhib(y(Gac-
tin),rpar(:,135)))*ymax(MRTF) - y(MRTF))/tau(MRTF);
dydt(Gactin)  =  (inhib(y(Factin),rpar(:,137))*ymax(Gactin)  -  y(Gac-
tin))/tau(Gactin);
dydt(TNC)                                                            =
(OR(act(y(NFKB),rpar(:,145)),OR(act(y(MRTF),rpar(:,146)),OR(in-
hib(y(RUNX1),rpar(:,214)),act(y(RUNX1),rpar(:,301)))))*ymax(TNC)    -
y(TNC))/tau(TNC);
dydt(mTORC1)      =      (act(y(Akt),rpar(:,140))*ymax(mTORC1)       -
y(mTORC1))/tau(mTORC1);
dydt(mTORC2)    =    (inhib(y(p70S6K),rpar(:,147))*ymax(mTORC2)     -
y(mTORC2))/tau(mTORC2);
dydt(p70S6K)     =     (act(y(mTORC1),rpar(:,141))*ymax(p70S6K)     -
y(p70S6K))/tau(p70S6K);
dydt(EBP1)      =      (inhib(y(mTORC1),rpar(:,142))*ymax(EBP1)     -
y(EBP1))/tau(EBP1);
dydt(syndecan4)  =  (AND(rpar(:,139),act(y(tension),rpar(:,139)),in-
hib(y(TNC),rpar(:,139)))*ymax(syndecan4)                             -
y(syndecan4))/tau(syndecan4);
dydt(proMMP3)                    =                    (AND(rpar(:,154),in-
hib(y(smad3),rpar(:,154)),act(y(NFKB),rpar(:,154)),act(y(AP1),rpar(:
,154)))*ymax(proMMP3) - y(proMMP3))/tau(proMMP3);
dydt(proMMP8)                    =                    (AND(rpar(:,153),in-
hib(y(smad3),rpar(:,153)),act(y(NFKB),rpar(:,153)),act(y(AP1),rpar(:
,153)))*ymax(proMMP8) - y(proMMP8))/tau(proMMP8);
```

```
dydt(proMMP12)    =    (act(y(CREB),rpar(:,157))*ymax(proMMP12)    -
y(proMMP12))/tau(proMMP12);
dydt(thrombospondin4)     =     (OR(act(y(smad3),rpar(:,151)),OR(in-
hib(y(LEF1),rpar(:,191)),act(y(LEF1),rpar(:,272))))*ymax(thrombos-
pondin4) - y(thrombospondin4))/tau(thrombospondin4);
dydt(osteopontin)   =   (act(y(AP1),rpar(:,155))*ymax(osteopontin)   -
y(osteopontin))/tau(osteopontin);
dydt(contractility)          =          (AND(rpar(:,164),act(y(Fac-
tin),rpar(:,164)),act(y(vincu-
lin),rpar(:,164)),act(y(MLC),rpar(:,164)))*ymax(contractility)     -
y(contractility))/tau(contractility);
dydt(RhoGEF)                                                        =
(AND(rpar(:,123),act(y(FAK),rpar(:,123)),act(y(Src),rpar(:,123)))*ym
ax(RhoGEF) - y(RhoGEF))/tau(RhoGEF);
dydt(RhoGDI)                          =                          (OR(in-
hib(y(Src),rpar(:,124)),OR(act(y(PKA),rpar(:,130)),in-
hib(y(PKC),rpar(:,134))))*ymax(RhoGDI) - y(RhoGDI))/tau(RhoGDI);
dydt(talin)                                                         =
(OR(act(y(B1int),rpar(:,161)),act(y(B3int),rpar(:,162)))*ymax(talin)
- y(talin))/tau(talin);
dydt(vinculin)             =             (AND(rpar(:,163),act(y(ten-
sion),rpar(:,163)),act(y(talin),rpar(:,163)))*ymax(vinculin) - y(vin-
culin))/tau(vinculin);
dydt(paxillin)                                                      =
(AND(rpar(:,159),act(y(FAK),rpar(:,159)),act(y(Src),rpar(:,159)),act
(y(MLC),rpar(:,159)))*ymax(paxillin) - y(paxillin))/tau(paxillin);
dydt(MLC) = (act(y(ROCK),rpar(:,166))*ymax(MLC) - y(MLC))/tau(MLC);
dydt(AT2R)      =      (act(y(AngII),rpar(:,171))*ymax(AT2R)      -
y(AT2R))/tau(AT2R);
dydt(BCL6)                            =                           (OR(in-
hib(y(abl),rpar(:,173)),act(y(STAT),rpar(:,310)))*ymax(BCL6)       -
y(BCL6))/tau(BCL6);
dydt(CACYBP)   =   (act(y(NFAT),rpar(:,285))*ymax(CACYBP)   -   y(CA-
CYBP))/tau(CACYBP);
dydt(CEBPD)                            =                           (OR(in-
hib(y(RELA),rpar(:,209)),act(y(RELA),rpar(:,294)))*ymax(CEBPD)     -
y(CEBPD))/tau(CEBPD);
dydt(CUX1)                            =                           (OR(in-
hib(y(STAT),rpar(:,222)),act(y(STAT),rpar(:,311)))*ymax(CUX1)      -
y(CUX1))/tau(CUX1);
dydt(EGR1)          =          (OR(inhib(y(cmyc),rpar(:,193)),OR(in-
hib(y(NFKB),rpar(:,202)),act(y(cmyc),rpar(:,276))))*ymax(EGR1)     -
y(EGR1))/tau(EGR1);
dydt(ETS1)                            =                           (OR(in-
hib(y(STAT),rpar(:,223)),act(y(STAT),rpar(:,312)))*ymax(ETS1)      -
y(ETS1))/tau(ETS1);
dydt(ETS2)          =          (OR(inhib(y(cmyc),rpar(:,194)),OR(in-
hib(y(smad3),rpar(:,218)),OR(act(y(cmyc),rpar(:,277)),act(y(smad3),r
par(:,305)))))*ymax(ETS2) - y(ETS2))/tau(ETS2);
dydt(HIF1A)                           =                           (OR(in-
hib(y(cmyc),rpar(:,196)),OR(act(y(MITF),rpar(:,274)),OR(act(y(cmyc),
```

```
rpar(:,279)),act(y(STAT),rpar(:,308)))))*ymax(HIF1A)                -
y(HIF1A))/tau(HIF1A);
dydt(LEF1)          =          (OR(inhib(y(RUNX1),rpar(:,212)),OR(in-
hib(y(SRF),rpar(:,219)),OR(act(y(NFAT),rpar(:,286)),act(y(RUNX1),rpa
r(:,299)))))*ymax(LEF1) - y(LEF1))/tau(LEF1);
dydt(IKZF1)       =       (act(y(NFKB),rpar(:,287))*ymax(IKZF1)       -
y(IKZF1))/tau(IKZF1);
dydt(KLF4)           =           (OR(inhib(y(AP1),rpar(:,186)),OR(in-
hib(y(cmyc),rpar(:,197)),OR(inhib(y(RUNX1),rpar(:,211)),OR(in-
hib(y(TCF4),rpar(:,236)),OR(act(y(AP1),rpar(:,267)),OR(act(y(cmyc),r
par(:,280)),act(y(RUNX1),rpar(:,298)))))))*ymax(KLF4)               -
y(KLF4))/tau(KLF4);
dydt(MITF)           =           (OR(inhib(y(STAT),rpar(:,224)),OR(in-
hib(y(TCF4),rpar(:,237)),act(y(STAT),rpar(:,313))))*ymax(MITF)      -
y(MITF))/tau(MITF);
dydt(NR5A2)       =       (act(y(CREB),rpar(:,255))*ymax(NR5A2)       -
y(NR5A2))/tau(NR5A2);
dydt(PPARA)                             =                           (OR(in-
hib(y(YAP),rpar(:,243)),OR(act(y(STAT),rpar(:,309)),act(y(YAP),rpar(
:,330))))*ymax(PPARA) - y(PPARA))/tau(PPARA);
dydt(RARG)           =           (OR(inhib(y(AP1),rpar(:,187)),OR(in-
hib(y(STAT),rpar(:,228)),act(y(STAT),rpar(:,316))))*ymax(RARG)      -
y(RARG))/tau(RARG);
dydt(RUNX1)          =          (OR(inhib(y(cmyc),rpar(:,200)),OR(in-
hib(y(NFKB),rpar(:,203)),OR(act(y(cmyc),rpar(:,283)),act(y(NFKB),rpa
r(:,288)))))*ymax(RUNX1) - y(RUNX1))/tau(RUNX1);
dydt(RELA)                              =                           (OR(in-
hib(y(cmyc),rpar(:,199)),act(y(cmyc),rpar(:,282)))*ymax(RELA)       -
y(RELA))/tau(RELA);
dydt(TEAD4)                              =                          (OR(in-
hib(y(STAT),rpar(:,231)),act(y(STAT),rpar(:,319)))*ymax(TEAD4)      -
y(TEAD4))/tau(TEAD4);
dydt(RUNX2)       =       (act(y(AP1),rpar(:,268))*ymax(RUNX2)       -
y(RUNX2))/tau(RUNX2);
dydt(TCF4)                              =                           (OR(in-
hib(y(STAT),rpar(:,230)),act(y(STAT),rpar(:,318)))*ymax(TCF4)       -
y(TCF4))/tau(TCF4);
dydt(TFCP2L1)         =         (OR(inhib(y(NFKB),rpar(:,204)),OR(in-
hib(y(STAT),rpar(:,232)),OR(act(y(NFKB),rpar(:,289)),act(y(STAT),rpa
r(:,320)))))*ymax(TFCP2L1) - y(TFCP2L1))/tau(TFCP2L1);
dydt(WT1)                               =                           (OR(in-
hib(y(STAT),rpar(:,233)),act(y(STAT),rpar(:,321)))*ymax(WT1)        -
y(WT1))/tau(WT1);
dydt(ZNF281)                             =                          (OR(in-
hib(y(STAT),rpar(:,234)),act(y(STAT),rpar(:,322)))*ymax(ZNF281)     -
y(ZNF281))/tau(ZNF281);

% utility functions
function fact = act(x,rpar)
% hill activation function with parameters w (weight), n (Hill coeff),
EC50
```

```
    w = rpar(1);
    n = rpar(2);
    EC50 = rpar(3);
    beta = (EC50.^n - 1)./(2*EC50.^n - 1);
    K = (beta - 1).^(1./n);
    fact = w.*(beta.*x.^n)./(K.^n + x.^n);
    if fact>w,                  % cap fact(x)<= 1
        fact = w;
    end

function finhib = inhib(x,rpar)
% inverse hill function with parameters w (weight), n (Hill coeff),
EC50
    finhib = rpar(1) - act(x,rpar);

function z = OR(x,y)
% OR logic gate
    z = x + y - x*y;

function z = AND(rpar,varargin)
% AND logic gate, multiplying all of the reactants together
    w = rpar(1);
    if w == 0,
        z = 0;
    else
        v = cell2mat(varargin);
        z = prod(v)/w^(nargin-2);
    end
```

Code C.4.2 Logic based ODE Model Parameters (.m) [2]

```
function [params,y0] = NetfluxODE_DCM_07122023_loadParams()
% Automatically generated by Netflux on 11-Jul-2023

% species parameters
speciesNames                          =                          {'An-
gII','AT1R','AGT','ACE','NOX','ROS','ET1','ETAR','DAG','PKC','TRPC',
'NE','BAR','AC','cAMP','PKA','CREB','CBP','TGFB','TGFB1R','smad3','s
mad7','latentTGFB','BAMBI','PDGF','PDG-
FR','NP','NPRA','cGMP','PKG','ten-
sion','B1int','Rho','ROCK','Ca','calcineu-
rin','NFAT','IL6','gp130','STAT','IL1','IL1RI','TNFa','TNFaR','NFKB'
,'PI3K','Akt','p38','TRAF','ASK1','MKK3','PP1','JNK','abl','Rac1','M
EKK1','MKK4','ERK','Ras','Raf','MEK1','FAK','epac','Fac-
tin','FA','cmyc','CTGF','proliferation','SRF','EDAFN','aS-
MA','AP1','TIMP1','TIMP2','PAI1','proMMP14','proMMP1','proMMP2','pro
MMP9','fibronectin','perios-
tin','proCI','proCIII','B3int','Src','Grb2','p130Cas','YAP','MRTF','
Gactin','TNC','mTORC1','mTORC2','p70S6K','EBP1','syndecan4','proMMP3
```

```
','proMMP8','proMMP12','thrombospondin4','osteopontin','contractili-
ty','RhoGEF','RhoGDI','talin','vinculin','paxil-
lin','MLC','AT2R','BCL6','CACYBP','CE-
BPD','CUX1','EGR1','ETS1','ETS2','HIF1A','LEF1','IKZF1','KLF4','MITF
','NR5A2','PPA-
RA','RARG','RUNX1','RELA','TEAD4','RUNX2','TCF4','TFCP2L1','WT1','ZN
F281',};
tau = [1, 1.000000e-01, 10, 1.000000e-01, 1.000000e-01, 1.000000e-01,
1, 1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01, 1,
1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01,
1.000000e-01, 1, 1.000000e-01, 1.000000e-01, 10, 10, 1.000000e-01, 1,
1.000000e-01, 1, 1.000000e-01, 1.000000e-01, 1.000000e-01, 1,
1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01,
1.000000e-01, 1, 1.000000e-01, 1.000000e-01, 1, 1.000000e-01, 1,
1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01,
1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01,
1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01,
1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01,
1, 1, 1.000000e-01, 1.000000e-01, 10, 1.000000e-01, 10, 10, 1.000000e-
01, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 1.000000e-01,
1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01,
1, 10, 1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01,
1.000000e-01, 10, 10, 10, 10, 10, 10, 1.000000e-01, 1.000000e-01,
1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
];
ymax = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ];

% reaction parameters
w = [1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01,
1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01, 1.000000e-01,
1.000000e-01, 8.000000e-01, 8.000000e-01, 8.000000e-01, 8.000000e-01,
8.000000e-01, 8.000000e-01, 8.000000e-01, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 8.000000e-
01, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 8.000000e-01, 1, 1, 1,
8.000000e-01, 1, 1, 1, 1, 1, 1, 1, 1, 8.000000e-01, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
```

```
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
];
n  =  [1.400000e+00,  1.400000e+00,  1.400000e+00,  1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
```

```
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00, 1.400000e+00,
1.400000e+00, 1.400000e+00, 1.400000e+00, ];
EC50  =  [6.000000e-01,  6.000000e-01,  6.000000e-01,  6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
```

```
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01, 6.000000e-01,
6.000000e-01, 6.000000e-01, 6.000000e-01, ];
rpar = [w;n;EC50];

params = {rpar,tau,ymax,speciesNames};

y0 = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ];
```

Code C.4.3 Running Logic Based ODE Model (.m) [2]

```
% NetfluxODE_DCM_07122023_run.m
 % Load parameters
 [params,y0] = NetfluxODE_DCM_07122023_loadParams();
```

```
% Run single simulation
tspan = [0 10];
options = [];
[t,y] = ode23(@NetfluxODE_DCM_07122023,tspan,y0,options,params);
```

Code C.4.4 Sensitivity Analysis (.m) [3, 4]

```
% Loading all default parameters
[params, y0] = MAGNet_DCM_ODE_loadParams();

% Save baseline simulation in a table
baselineTable = table('Size', [1, numel(params{4})+1], 'Variable-
Types', [{'string'}, repmat({'double'}, 1, numel(params{4}))], 'Var-
iableNames', [{'Species Name'}, params{4}]);
baselineTable(:, 1) = params{4}';

% Running baseline simulation
tspan_baseline = [0 10];
options_baseline = [];
[t_baseline, y_baseline] = ode23(@MAGNet_DCM_ODE, tspan_baseline, y0,
options_baseline, params);

baselineTable(1, 2:end) = num2cell(real(y_baseline(end, :)));

% Save simulation in another table
resultsTable = table('Size', [numel(params{3}), numel(params{3})+1],
'VariableTypes', [{'string'}, repmat({'double'}, 1, nu-
mel(params{3}))], 'VariableNames', [{'Species Name'}, params{4}]);
resultsTable(:, 1) = params{4}';

% Simulate with ymax = 0.1
for i = 1:numel(params{3})
    % Setting ymax to 0.1 for current species
    ymax_modified = params{3};
    ymax_modified(i) = 0.1;

    % Updating parameters with modified ymax
    params_modified = params;
    params_modified{3} = ymax_modified;

    % Running single simulation
    tspan = [0 100];
    options = [];
    [t, y] = ode23(@MAGNet_DCM_ODE, tspan, y0, options, params_modi-
fied);

    % Saving real part of the simulation result (extracting real part)
    resultsTable(i, 2:end) = num2cell(real(y(end, :)));
end
```

```
% Subtracting baseline result
for i = 1:size(resultsTable, 1)
    resultsTable(i,   2:end)   =   num2cell(cell2mat(resultsTable(i,
2:end)) - cell2mat(baselineTable(1, 2:end)));
end

% Saving the data as a CSV file
writetable(resultsTable, 'perturbation.csv');
```

Code C.4.4 Sensitivity Analysis plotting (.py) [3, 4]

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from matplotlib import font_manager
data = pd.read_csv('perturbation.csv')
species_names = data['Species Name']
data = data.drop(columns='Species Name')

# High Resolution image
fig, ax = plt.subplots(figsize=(36, 30))

# using mild color (Blues)
cmap = plt.cm.get_cmap('Blues')

# Heatmap
heatmap = ax.imshow(data, cmap=cmap, aspect='auto')

# y-axis tick labels as the species names
ax.set_yticks(range(len(species_names)))
ax.set_yticklabels(species_names)

# x-axis tick labels
ax.set_xticks(range(data.shape[1]))
ax.set_xticklabels(data.columns, rotation=45, ha='right')

# heatmap colorbar
cbar = plt.colorbar(heatmap)
cbar.set_label('Activation', fontname='Arial', fontsize=18)

# Plot title setup
ax.set_title('Perturbation Analysis (DCM)', fontname='Arial', font-
size=24)

# Setting font size and properties
font_properties    =    font_manager.FontProperties(family='Arial',
size=12)
ax.tick_params(axis='both', which='major', labelsize=12)
ax.tick_params(axis='both', which='minor', labelsize=12)
```

```
ax.set_yticklabels(species_names, fontproperties=font_properties)
ax.set_xticklabels(data.columns, rotation=45, ha='right', fontproper-
ties=font_properties)

font_properties_cbar  =  font_manager.FontProperties(family='Arial',
size=18)
cbar.ax.tick_params(labelsize=18)
cbar.ax.set_xticklabels(cbar.ax.get_xticklabels(),       fontproper-
ties=font_properties_cbar)
cbar.ax.set_yticklabels(cbar.ax.get_yticklabels(),       fontproper-
ties=font_properties_cbar)

# Setting tight layout
plt.tight_layout()

# Showing the heatmap
plt.show()
```

Code C.4.5 Pearson correlation plotting (.py) [3, 4]

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Group 2 and Group 3 (excluding 'proMMP1')are numerical data
group2_columns = ['Age', 'Weight (Kg)', 'Height (cm)', 'Heart Weight
(gm)', 'LV_Mass (gm)', 'LVEF']
group3_columns  =  ['proMMP14',  'latentTGFB',  'periostin',  'IL6',
'proCIII', 'ET1', 'fibronectin', 'CTGF',
                   'TNC',  'proCI',  'proMMP2',  'AGT',  'osteopontin',
'TIMP1', 'TIMP2', 'PAI1', 'aSMA']

# Correlation matrix for Group 2 and Group 3
correlation_matrix   =   final_df[group2_columns   +   group3_col-
umns].corr(method='pearson')

sns.set(font="Arial", font_scale=2)

# Big image
plt.figure(figsize=(24, 20))
annot_font_size = 14
annot_kws = {'fontsize': annot_font_size, 'fontweight': 'bold'}

# Correlation heatmap in seaborn
sns.heatmap(correlation_matrix, annot=True, cmap='seismic', center=0,
annot_kws=annot_kws)
plt.title('Correlation between Clinical Variables and Model Output')
plt.show()
```

6.3 References

1. Rogers JD, Aguado BA, Watts KM, Anseth KS, Richardson WJ. Network modeling predicts personalized gene expression and drug responses in valve myofibroblasts cultured with patient sera. Proc Natl Acad Sci U S A. 2022;119(8):e2117323119. doi:10.1073/pnas.2117323119

2. https://github.com/saucermanlab/Netflux

3. https://stackoverflow.com/

4. https://chat.openai.com/