

# Identifying biomarkers of cardiotoxicity with a genome-scale model of metabolism constrained with 'omics data

---

A

Dissertation

Presented to

the faculty of the School of Engineering and Applied Science  
University of Virginia

---

in partial fulfillment  
of the requirements for the degree

Doctor of Philosophy

by

Bonnie Victoria Dougherty

December 2020

# APPROVAL SHEET

This  
Dissertation  
is submitted in partial fulfillment of the requirements  
for the degree of  
Doctor of Philosophy

Author: **Bonnie Victoria Dougherty**

This Dissertation has been read and approved by the examining committee:

Advisor: **Jason Papin**

Advisor:

Committee Member: **Jeffrey Saucerman**

Committee Member: **Shayn Peirce-Cottler**

Committee Member: **Kevin Janes**

Committee Member: **Zhen Yan**

Committee Member:

Committee Member:

Accepted for the School of Engineering and Applied Science:



Craig H. Benson, School of Engineering and Applied Science

**December 2020**

# Abstract

With recent improvements in the detection and treatment of cancer, the adverse side effects of chemotherapeutics, particularly cardiotoxicity, have become more apparent. Many chemotherapeutics are now associated with adverse cardiovascular events, however, there are no clinical measures to detect, limit, or prevent cardiotoxicity. Although research work has demonstrated that metabolites are good biomarkers of early cardiotoxicity, further work is needed. Here, we utilize genome-scale metabolic network reconstructions (GENREs) to identify new biomarkers of cardiotoxicity. First, we built a heart-specific GENRE and used the model with a novel approach, the Tasks Inferred from Differential Expression (TIDEs) approach, to identify shifts in metabolic functions in heart failure. Next, we collected paired transcriptomics and metabolomics data in primary rat neonatal cardiomyocytes exposed to three compounds (5-fluorouracil, acetaminophen, and doxorubicin) to characterize *in vitro* cardiotoxicity. Finally, we integrated our collected data with a model of heart metabolism to identify shifts in metabolic functions, unique metabolic reactions, and shifts in metabolic reactions that are unique to cardiotoxicity. For each compound, we identified unique shifts in metabolism, confirming mechanisms of toxicity for doxorubicin and proposing new hypotheses for mechanisms of toxicity for 5-fluorouracil and acetaminophen. Given that our experiments are done in rats, future work is needed to address translatability in humans. To this end, finally, we highlight the utility of data-driven and mechanistic modeling approaches in making cross-species comparisons.

# Acknowledgements

My dissertation work would not have been possible without the guidance, love, and support of a number of people. First, I would like to thank my entire family for being a constant source of support and encouragement throughout my entire life; I would not be the person that I am today without you. To my mom, thank you for inspiring my love of engineering early in life and for all of the sacrifices that you have made to enable me to pursue science. To my grandparents, thank you for your unconditional love and support through every aspect of my life, especially for all of the sacrifices that you made ensure my happiness, for that I am eternally grateful. You have clearly demonstrated the love of family and I'm so happy to share this accomplishment with you. To both my mother and my grandmother, I owe my love of math and science. To my dad, thank you for your constant love and support for anything that I've wanted to pursue. You have demonstrated to me true perseverance during adversity and taught me to always appreciate the blessings that I have in life, especially the importance of family. To my brothers, I know that I was a bit of a perfectionist growing up, so I thank you for both your patience and putting me in my place when I needed it. I love how our relationship has grown since we were kids and I look forward to the future. To my extended family, my aunts, uncles, cousins, and the great-grandkids, thank you for being a constant in my life, for continuing to encourage me and support me as I pursue my dreams. I got through grad school knowing that each member of my family was just a phone call away for support. Finally, to my late grandmother, I know that you would have been so proud to see me come to this moment, from seeing me learn to count to 10 so early because I spent so much time in time out to defending my PhD, I am grateful for your love and support and I know you are looking down with love on all of us from above.

I have been immensely blessed to not only have my family as a support system but also a phenomenal group of friends. I'd like to thank my best friend from high school, April, for being a steadfast friend over the past I don't want to admit how many years. Thank for you countless

adventures (Ireland, camping, hiking, and more), laughs, and high stakes games of nertz. To Becca, even though we don't talk on the phone everyday anymore, I wouldn't have made it this far in life without your love and constant support. I especially wouldn't have made it through the first two years of grad school without our daily chats. To Kerianne, thank you for being the best roommate and fellow wanderluster; Italy, Paris, and southeast Asia were the highlights of my grad school travels and I'm excited for our next adventure. Oh and thanks to Penny too. To Laura (and Connor and Tessa), who could have predicted this is where we'd end up after meeting during third year of undergrad. Thanks for being my oldest friend in Charlottesville and getting a really cute puppy during quarantine that I could play with. To Lauren (and Garrett, but mostly Lauren), I'm so thankful that our paths crossed during Charlottesville kickball and I can't imagine my life without you in it, especially the late-night wine chats, I can't live without those. To Kris and Adrienne, you both bring a palatable joy and radiance to my life and I most certainly wouldn't have made it through grad school without your continued encouragement and support. Thank you for always being my biggest cheerleaders and calling me out when I'm being ridiculous. To pizza group, my first real friends out of undergrad, thanks for bringing a consistent laugh to my face and for some really tasty pizza. Melissa, again, many thanks to Charlottesville kickball for bringing you into my life. Thank you for your constant enthusiasm, fresh perspective, and genuine kindness. Oh and that stellar New Year's trip to Death Valley, until our next campervan adventure! Brittany, thanks to fate for bringing us together during grad school, and thanks to you for being a source of constant encouragement and genuine friendship. My life won't be the same without our weekly lunches. Bryan and Mukti, thanks for the late-night lifting workouts, the nourishing tears, the much-needed ice cream, and the constant source of laughs. I wouldn't have made it through as sane or as strong without you both as gym buddies and friends. To Hannah and Helen, thanks for the memory of that fateful day in the atrium that brought us together, thanks for the many laughs. But more, thank you for being equal parts crazy, adventurous, and quirky, just like me. Finally, thank

you to my entire bible study. Bible study has been a constant source of affirmation, prayer, grounding, and friendship over the past almost 4 years and I certainly wouldn't have made it through grad school without you lovely ladies to support me physically, emotionally, and in prayer.

Next, I would like to thank my colleagues and lab mates. First, my writing group through the Graduate Writing Lab. Thank you for your continued feedback and helping me to improve both my writing and critiquing. Second, to my lab mates. I consider myself lucky to have you as both colleagues and friends. To Glynis, thank you for your consistent mentorship, both in science and in life. Thank you for always being willing to take the time to teach me a new technique, talk through my experimental design, and talk through my life struggles. To Kris, grad school definitely would not have been the same without you. Thank you for being a constant resource and faithful friend. I wouldn't have succeeded in grad school without your mentorship and help. To Laura, thanks for the constant sass, genuine laughter, and being a whiz at figure design. I'm thankful to have gone through the stress of dissertation writing with you. To Matt, thanks for always answering my barrage of random questions and always being willing to commiserate about science. Finally, thanks to all the other members of the lab, past and present, Jennie, Edik, Anna, Phil, Greg, Maureen, Tom, Dawson, Lillian, and Deb for demonstrating genuine collegiality and giving me honest feedback on my work.

Finally, thank you to my advisor, Jason Papin for being an amazing mentor and teacher. Thank you for allowing me the opportunity to work, learn, and grow in the lab. You supported me in pursuing my varied interests both in and out of the lab that have shaped me into the independent researcher that I am today. Thank you for supporting me through the varied struggles that I encountered during grad school, from problems with experiments to personal issues, you provided a listening ear and support. Thank you for believing in me even when I didn't believe in myself.

# Table of Contents

Abstract.....	i
Acknowledgements.....	iv
List of Figures .....	x
List of Supplemental Tables, Figures, and Data .....	xi
Chapter 1 : Background and Significance .....	1
1.1    Cardiotoxicity is a serious clinical problem that lacks appropriate diagnostic measures .....	1
1.2    Genome-scale metabolic network reconstructions provide an opportunity to better understand metabolic shifts in cardiotoxicity .....	2
1.3    Building a new, heart-specific model of metabolism to integrate collected data to identify metabolic shifts <i>in vitro</i> cardiotoxicity.....	3
1.4    Paired models of human and rat heart metabolism can facilitate cross-species comparisons to identify shared markers of toxicity .....	5
1.5    References .....	7
Chapter 2: Identifying functional metabolic shifts in heart failure with the integration of omics data and a heart-specific, genome-scale model.....	11
2.1    Summary.....	12
2.2    Introduction .....	13
2.3    Results.....	15
2.3.1    Building and validating <i>iCardio</i> using metabolic tasks.....	15
2.3.2    Identifying Tasks Inferred from Differential Expression (TIDEs) using <i>iCardio</i> for heart failure gene expression data.....	18
2.3.3    Comparison of TIDEs with GSEA by KEGG pathway.....	22
2.4    Discussion.....	24
2.5    Methods.....	28
2.5.1    Developing heart-relevant metabolic tasks.....	28
2.5.2    Generating and curating a heart-specific metabolic model .....	29
2.5.3    Validating <i>iCardio</i> using reaction coverage and ATP yields .....	30
2.5.4    Analyzing transcriptomics data from heart failure patients .....	31
2.5.5    Identifying Tasks Inferred from Differential Expression (TIDEs) using <i>iCardio</i> with expression data.....	31
2.5.6    Gene set enrichment analysis.....	32
2.6    Acknowledgements.....	33
2.7    Author contributions.....	33
2.8    Conflicts of Interest.....	33
2.9    Data and Code Availability .....	34
2.10   Supplemental Tables, Figures, and Data.....	34

2.11	References .....	38
Chapter 3: Identifying biomarkers of chemotherapy-induced cardiotoxicity using paired transcriptomics and metabolomics data integrated with a model of heart metabolism..... 43		
3.1	Abstract.....	44
3.2	Introduction .....	45
3.3	Results.....	46
3.3.1	Optimizing concentrations of compounds to characterize <i>in vitro</i> cardiotoxicity .....	46
3.3.2	Unsupervised machine learning of transcriptomics and metabolomics data highlights underlying drug-induced shifts in cellular activity.....	48
3.3.3	Gene enrichment and metabolomics data identify common signatures of toxicity but cannot readily identify mechanisms of cardiotoxicity .....	50
3.3.4	Reconstruction of a rat-specific heart GENRE from an existing human-specific heart GENRE.....	52
3.3.5	Integrating a rat-specific, heart GENRE and transcriptomics data predicts novel metabolic tasks altered in <i>in vitro</i> cardiotoxicity .....	53
3.3.6	Combined omics datasets predict novel biomarkers of toxicity through integrated network-based analysis .....	56
3.4	Discussion.....	59
3.5	Methods.....	62
3.5.1	<i>In vitro</i> culture conditions.....	62
3.5.2	RNA isolation, sequencing, and analysis.....	64
3.5.3	Collecting and analyzing metabolomics data.....	65
3.5.4	Building a rat-specific heart model from the human-specific heart model, <i>iCardio</i> ....	<b>Error!</b>
	<b>Bookmark not defined.</b>	
3.5.5	Computationally predicting biomarkers of toxicity using TIMBR and RIPTiDe.....	67
3.6	Acknowledgements:.....	68
3.7	Author Contributions: .....	69
3.8	Supplemental Figures, Tables, and Data.....	69
3.9	References .....	74
Chapter 4: Systems biology approaches help to facilitate interpretation of cross-species comparisons.. 79		
4.1	Abstract.....	80
4.2	Introduction .....	81
4.3	Data-driven models.....	82
4.3.1	Machine learning .....	82
4.3.2	Rationally choosing model organisms .....	84
4.4	Mechanism-driven models .....	84
4.4.1	Pharmacokinetic (PKPB) models.....	85
4.4.2	Genome scale metabolic network reconstructions (GENRES) .....	86
4.4.3	Pathway-based models.....	87
4.5	Conclusions .....	89
4.6	Acknowledgements.....	89

4.7	References .....	90
Chapter 5:	Discussion and Future Work .....	94
5.1	Discussion.....	94
5.1.1	Identifying metabolic shifts in the context of heart failure and <i>in vitro</i> cardiotoxicity using metabolic models.....	94
5.1.2	General utility of the Tasks Inferred from Differential Expression (TIDEs) approach in identifying shifted metabolic functions from transcriptomics data .....	96
5.1.3	Difficulty in predicting biomarkers using transcriptomics data integrated with metabolic models.....	98
5.1.4	RIPTiDe models highlight the utility of integrating both transcriptomics and metabolomics data.....	101
5.2	Future work.....	102
5.2.1	Experimentally confirming identified shifts in toxicity .....	102
5.2.2	Developing a more comprehensive, heart-specific objective function to probe changes in metabolism .....	103
5.2.3	Integrating human-specific data to identify shared biomarkers of doxorubicin-induced cardiotoxicity.....	104
5.2.4	Developing pathway tracing algorithms .....	105
5.2.5	Implementing a transcript-based TIDEs approach.....	105
5.2.6	Future manual curation of the human and rat models using collected paired data .....	106
5.3	Conclusion.....	106
5.4	References .....	107

# List of Figures

Figure 2.1 Building a heart metabolic model ( <i>iCardio</i> ) by integrating protein data and curating with metabolic tasks. ....	15
Figure 2.2 Validating <i>iCardio</i> (A) descriptively using the number of reactions covered by metabolic tasks and (B) quantitatively using ATP yields for common carbon substrates. ....	17
Figure 2.3 Identifying functional metabolic changes from gene expression data using metabolic tasks. .	18
Figure 2.4 Distributions for model-predicted task scores from gene expression data for ischemic heart failure for each of the 307 tested tasks. ....	19
Figure 2.5 TIDEs analysis identifies functional metabolic changes across different datasets (A) compared to results from a traditional GSEA analysis (B). ....	21
Figure 2.6 TIDEs analysis identifies differences in datasets for genes which catalyze the conversion of arginine to nitric oxide. ....	24
Figure 3.1 Choosing cardiotoxic drug concentrations that elicited a change in metabolism without a significant increase in cell death. ....	47
Figure 3.2 PCA of transcript counts (A) and scaled metabolite abundances (B) for the three compounds at 6 and 24 hours. ....	48
Figure 3.3 Identifying biomarkers of toxicity from the transcriptomics and metabolomics data sets. ....	50
Figure 3.4 TIDEs analysis reveals distinct changes in metabolic function across compounds. ....	54
Figure 3.5 Condition-specific models integrating the metabolomics and transcriptomics 24-hour data identify unique reaction biomarkers of cardiotoxicity. ....	57
Figure 4.1 Strengths and weaknesses of data- vs. mechanism-driven models. ....	81
Figure 4.2 The iterative model building process. ....	85
Figure 5.1 TIMBR predictions compared to measured changes in metabolites. ....	98
Figure 5.2 Integrating both metabolomics and transcriptomics data reveals unique metabolism compared to either alone. ....	102

# List of Supplemental Tables, Figures, and Data

Supplemental Table 2.1 Changes to the <i>iHsa</i> network reconstruction resulting from metabolic task curation. ....	34
Supplemental Table 2.2 Evaluating integration algorithms for construction of draft <i>iCardio</i> models.....	35
Supplemental Table 2.3 Summary of DEGs for individual microarray studies. ....	35
Supplemental Figure 2.1 GSEA for all KEGG pathways. ....	36
Supplemental Figure 2.2 Subsystem-level analysis using <i>iCardio</i> with gene expression data. ....	36
Supplemental Figure 3.1 Confirming changes in measures of cellular respiration for the chosen Dox concentration using the MST assay. ....	69
Supplemental Figure 3.2 Cell reducing potential and cell death measures following 6 hours of exposure to compounds. ....	70
Supplemental Figure 3.3 Additional PCA plots demonstrate (A) separation of treatment vs control groups and (B) the top 10 metabolites separating the PCA of the scaled metabolite abundances.....	70
Supplemental Figure 3.4 Metabolomics data shows differential production and consumption of metabolites across compounds and timepoints.....	71
Supplemental Figure 3.5 The rat-specific heart model captures changes in DEGs and metabolomics. ....	72
Supplemental Figure 3.6 Distribution of random task scores for the metabolic tasks for arginine to nitric oxide and ROS detoxification demonstrate the underlying distribution of DEGs. ....	72

# Chapter 1 : Background and Significance

## **1.1 Cardiotoxicity is a serious clinical problem that lacks appropriate diagnostic measures**

Recent advances in the detection and treatment of cancer have increased survival rates. However, with increased survival, the side effects of chemotherapeutics have become more evident, particularly the higher incidence of adverse cardiovascular events [1]. It is now well established that some chemotherapeutics are associated with several types of cardiovascular damage, ranging from acute myocarditis to left ventricular dysfunction and subsequent heart failure [1]. The most well-studied chemotherapeutic that is associated with an increased risk for cardiovascular damage is doxorubicin. Doxorubicin-induced cardiotoxicity often presents as left ventricular dysfunction that eventually progresses to heart failure [2]. The current clinical standard of care for preventing cardiotoxicity is limiting the overall drug dose and tracking left ventricular ejection fraction (LVEF) using echocardiography [2]. However, changes in LVEF indicate damage that has significantly decreased heart function and, in the case of cardiotoxicity, this damage is usually irreversible, highlighting the need for earlier biomarkers of changes in cardiac function.

Biomarkers, such as B-type natriuretic peptide (BNP) and cardiac troponins, are used as non-invasive, early measures of cardiac function [3] and have been explored as markers of cardiotoxicity [4–6]. In diagnosing myocardial infarction, BNP and cardiac troponins increase before a change in LVEF. However, since the release of BNP and cardiac troponins is associated with myocardial necrosis [7], increases in BNP or cardiac troponin still only indicate irreversible damage. More recently, changes in the uptake of metabolites have been used to identify subclinical dysfunction. For example, an increase in glucose uptake was measured before a decrease in LVEF in spontaneously hypertensive rats [8]. Similarly, an increase in glucose uptake has been measured both in mouse studies and in patients treated with doxorubicin [9,10].

However, it is still unclear if this increase in glucose uptake predicts the development of cardiotoxicity. These changes in the uptake of metabolites suggest that metabolites and changes in metabolism are an opportunity for early markers of subclinical dysfunction in the heart.

Typical metabolomics studies to identify new metabolic biomarkers involve profiling patient serum or media from *in vitro* studies to identify significantly changed metabolites associated with a condition or disease. While this untargeted approach identifies potential biomarkers, it does not include a potential mechanism for the production of the metabolite. Additionally, an untargeted approach only measures metabolites that are defined in a panel *a priori*, and may therefore miss key biomarkers which either weren't defined or cannot be measured using current metabolomics approaches.

## **1.2 Genome-scale metabolic network reconstructions provide an opportunity to better understand metabolic shifts in cardiotoxicity**

Mechanistic models, such as genome-scale metabolic network reconstruction (GENREs) provide a computational approach for the systematic integration of experimental data to identify potential mechanisms driving a disease or condition of interest. GENREs represent: (a) the known metabolic reactions that an organism undergoes, (b) the stoichiometry of the metabolites that are converted in these reactions, and (c) the genes encoding for the proteins which catalyze these reactions [11]. GENREs have been used to identify metabolic markers of specific diseases or conditions [12–16], identify new drug targets [17,18], and predict pathways associated with drug side effects [19–21].

Previous work from our lab has demonstrated the value of utilizing GENREs to integrate experimental omics datasets to predict metabolic biomarkers [22]. The Transcriptionally Inferred Metabolic Biomarker Response (TIMBR) algorithm has been used to integrate transcriptomics data to explore biomarkers of hepatotoxicity [22–24] and nephrotoxicity [25] but has not yet been

extended to cardiotoxicity. In the case of both hepatotoxicity and nephrotoxicity, paired metabolomics data was collected and used to validate predictions of *in vitro* biomarkers of toxicity, providing both a validated biomarker and a mechanism for production. Further, while the TIMBR algorithm explored integrating transcriptomics data, no studies have explored the potential of integrating paired transcriptomics and metabolomics data with GENREs to predict potential biomarkers and mechanisms of toxicity.

Here, we explore the utility of GENREs to integrate experimental 'omics data and identify new potential mechanisms and biomarkers of cardiotoxicity. Numerous models of human metabolism have been published [22,26–28] and serve as resources for all reactions that are catalyzed based on the human genome. However, not all reactions are catalyzed in every tissue. Therefore, we first need to identify an appropriate heart-specific model that contains only the reactions known to be catalyzed in the heart. Previous GENREs of heart metabolism are available but either only cover mitochondrial metabolism [29] and thus are limited in scope, or were built from older models of human metabolism [30,31] and do not adequately capture the current cumulative knowledge of the field, necessitating the generation of a new, heart-specific reconstruction.

### **1.3 Building a new, heart-specific model of metabolism to integrate collected data to identify metabolic shifts *in vitro* cardiotoxicity**

Given that rats are the organism of choice for toxicity studies, we have chosen to build a new heart-specific model from the recently published, paired models of general human and rat metabolism [22] (Chapter 2). Tissue-specific models are often built using publicly available data, such as transcriptomics or proteomics data. Extensive immunohistochemistry data for tissue-specific protein expression is available for humans [v18.proteinstlas.org; [32]] but is not available for rats. Therefore, building from these paired models will facilitate both the integration of tissue-

specific proteomics data that is available for humans and *in vitro* experimental cardiotoxicity data in rats.

Tissue-specific models of metabolism have been developed for a range of tissues to explore a variety of questions [15,33–35]. A common theme amongst these studies is optimizing an objective function, which represents a hypothesis for the overall function of the tissue or cell-type of interest. For many models, such as with the liver [22] or cancer [33,35], an objective function is developed that represents overall cell growth and production of key intermediate metabolites, such as bile acids in the liver. However, a healthy heart does not regenerate or synthesize key intermediate metabolites, making it difficult to identify one comprehensive objective function. Previous work has utilized the concept of metabolic tasks, which describe general metabolic functions, e.g. the generation of ATP from glucose, to validate metabolic models [22,28,36]. Here, we present a new method that utilizes metabolic tasks to identify metabolic functions that are significantly changed in a transcriptomics data set. First, we demonstrate the utility of the method by identifying common metabolic shifts in heart failure using publicly available data (Chapter 2), then we identify metabolic shifts in cardiotoxicity using our own collected data (Chapter 3).

For the case of heart failure, there are many publicly available microarray datasets collected from patients that are undergoing heart transplants. For cardiotoxicity, there is no publicly available data that thoroughly characterizes *in vitro* toxicity to allow for the systematic identification of new biomarkers. Therefore, we have collected paired transcriptomics and metabolomics data to characterize *in vitro* cardiotoxicity in primary rat neonatal cardiomyocytes exposed to three compounds: 5-fluorouracil, doxorubicin, and acetaminophen (Chapter 3). We chose these compounds based on the established cardiotoxicity for doxorubicin [2] and 5-fluorouracil [37]. Acetaminophen, a compound with established hepatotoxicity [38] and

nephrotoxicity [39], served as a common compound with other studies that explored using transcriptomics data with GENREs to predict biomarkers [23–25].

Using our new heart-specific model of metabolism (Chapter 2), we integrate our experimentally paired transcriptomics and metabolomics data to yield insight into cardiotoxicity (Chapter 3). First, we identify metabolic functions that are significantly associated with differentially expressed genes (DEGs) for each of our three tested compounds, confirming known therapeutic mechanisms and identifying potential markers of cardiotoxicity. Second, we demonstrate the utility of integrating both transcriptomics and metabolomics data to identify new potential biomarkers of cardiotoxicity. For doxorubicin, we confirm known mechanisms of cardiotoxicity and propose new mechanisms for 5-fluorouracil and acetaminophen cardiotoxicity.

#### **1.4 Paired models of human and rat heart metabolism can facilitate cross-species comparisons to identify shared markers of toxicity**

Here, we have utilized data collected from primary rat neonatal cardiomyocytes to yield insight into metabolic changes in *in vitro* cardiotoxicity. However, even though rats are often used as surrogates for human in toxicity studies, there are limitations to translating between both *in vitro* and *in vivo* models as well as between rats and humans [40,41]. In Chapter 4, we discuss some opportunities for utilizing large data sets and mechanistic models to facilitate cross-species comparisons [42]. Although large, comprehensive datasets are available for both *in vitro* and *in vivo* toxicity in the liver and kidney [40], comprehensive data for cardiotoxicity is not available.

Our recently published paired models of rat and human metabolism [22] capture key differences in species-specific metabolism. Further, these models can facilitate cross-species comparisons to yield insight into shared mechanisms of toxicity between species. In Chapter 5, we discuss the possibilities for future integration of human-specific, *in vitro* cardiotoxicity data to identify shared mechanisms of toxicity between rats and humans for doxorubicin-induced

cardiotoxicity. Data is publicly available for doxorubicin treated human induced pluripotent stem cell-derived cardiomyocytes (iPSC-CMs) [43], but not for the other compounds studied here, necessitating future work to determine if the metabolic shifts identified in rats translate to humans. The observed identified metabolic shifts and proposed mechanisms of toxicity in this work provide an opportunity for identifying the role of metabolites to diagnose, or metabolic interventions to prevent or mitigate cardiotoxicity.

## 1.5 References

1. Albini A, Pennesi G, Donatelli F, Cammarota R, De Flora S, Noonan DM. Cardiotoxicity of anticancer drugs: the need for cardio-oncology and cardio-oncological prevention. *J Natl Cancer Inst.* 2010;102: 14–25. doi:10.1093/jnci/djp440
2. Volkova M, Russell R. Anthracycline Cardiotoxicity: Prevalence, Pathogenesis and Treatment. *Curr Cardiol Rev.* 2011;7: 214–220. doi:10.2174/157340311799960645
3. Magnussen C, Blankenberg S. Biomarkers for heart failure: small molecules with high clinical relevance. *J Intern Med.* 2018;283: 530–543. doi:10.1111/joim.12756
4. Cardinale D, Sandri MT, Martinoni A, Tricca A, Civelli M, Lamantia G, et al. Left ventricular dysfunction predicted by early troponin I release after high-dose chemotherapy. *J Am Coll Cardiol.* 2000;36: 517–522. doi:10.1016/s0735-1097(00)00748-8
5. Cardinale D, Sandri MT, Martinoni A, Borghini E, Civelli M, Lamantia G, et al. Myocardial injury revealed by plasma troponin I in breast cancer treated with high-dose chemotherapy. *Ann Oncol Off J Eur Soc Med Oncol.* 2002;13: 710–715. doi:10.1093/annonc/mdf170
6. Tan L-L, Lyon AR. Role of Biomarkers in Prediction of Cardiotoxicity During Cancer Treatment. *Curr Treat Options Cardiovasc Med.* 2018;20. doi:10.1007/s11936-018-0641-z
7. Park KC, Gaze DC, Collinson PO, Marber MS. Cardiac troponins: from myocardial infarction to chronic disease. *Cardiovasc Res.* 2017;113: 1708–1718. doi:10.1093/cvr/cvx183
8. Li J, Kemp BA, Howell NL, Massey J, Mińczuk K, Huang Q, et al. Metabolic Changes in Spontaneously Hypertensive Rat Hearts Precede Cardiac Dysfunction and Left Ventricular Hypertrophy. *J Am Heart Assoc.* 2019;8: e010926. doi:10.1161/JAHA.118.010926
9. Bauckneht M, Ferrarazzo G, Fiz F, Morbelli S, Sarocchi M, Pastorino F, et al. Doxorubicin Effect on Myocardial Metabolism as a Prerequisite for Subsequent Development of Cardiac Toxicity: A Translational 18F-FDG PET/CT Observation. *J Nucl Med.* 2017;58: 1638–1645. doi:10.2967/jnumed.117.191122
10. Borde C, Kand P, Basu S. Enhanced myocardial fluorodeoxyglucose uptake following Adriamycin-based therapy: Evidence of early chemotherapeutic cardiotoxicity? *World J Radiol.* 2012;4: 220–223. doi:10.4329/wjr.v4.i5.220
11. Rawls KD, Dougherty BV, Blais EM, Stancliffe E, Kolling GL, Vinnakota K, et al. A simplified metabolic network reconstruction to promote understanding and development of flux balance analysis tools. *Comput Biol Med.* 2019;105: 64–71. doi:10.1016/j.combiomed.2018.12.010

12. Granata I, Troiano E, Sangiovanni M, Guarracino MR. Integration of transcriptomic data in a genome-scale metabolic model to investigate the link between obesity and breast cancer. *BMC Bioinformatics*. 2019;20: 162. doi:10.1186/s12859-019-2685-9
13. Graudenzi A, Maspero D, Di Filippo M, Gnugnoli M, Isella C, Mauri G, et al. Integration of transcriptomic data and metabolic networks in cancer samples reveals highly significant prognostic power. *J Biomed Inform*. 2018;87: 37–49. doi:10.1016/j.jbi.2018.09.010
14. Hiemer S, Jatav S, Jussif J, Alley J, Lathwal S, Piotrowski M, et al. Integrated Metabolomic and Transcriptomic Profiling Reveals Novel Activation-Induced Metabolic Networks in Human T cells. *bioRxiv*. 2019; 635789. doi:10.1101/635789
15. Stempler S, Yizhak K, Ruppin E. Integrating transcriptomics with metabolic modeling predicts biomarkers and drug targets for Alzheimer’s disease. *PloS One*. 2014;9: e105383. doi:10.1371/journal.pone.0105383
16. Våremo L, Scheele C, Broholm C, Mardinoglu A, Kampf C, Asplund A, et al. Proteome- and Transcriptome-Driven Reconstruction of the Human Myocyte Metabolic Network and Its Use for Identification of Markers for Diabetes. *Cell Rep*. 2015;11: 921–933. doi:10.1016/j.celrep.2015.04.010
17. Dougherty BV, Moutinho TJ, Papin J. Accelerating the Drug Development Pipeline with Genome-Scale Metabolic Network Reconstructions. *Systems Biology*. John Wiley & Sons, Ltd; 2017. pp. 139–162. doi:10.1002/9783527696130.ch5
18. Rawls K, Dougherty BV, Papin J. Metabolic Network Reconstructions to Predict Drug Targets and Off-Target Effects. In: Nagrath D, editor. *Metabolic Flux Analysis in Eukaryotic Cells: Methods and Protocols*. New York, NY: Springer US; 2020. pp. 315–330. doi:10.1007/978-1-0716-0159-4\_14
19. Guebila MB, Thiele I. Predicting gastrointestinal drug effects using contextualized metabolic models. *PLOS Comput Biol*. 2019;15: e1007100. doi:10.1371/journal.pcbi.1007100
20. Shaked I, Oberhardt MA, Atias N, Sharan R, Ruppin E. Metabolic Network Prediction of Drug Side Effects. *Cell Syst*. 2016;2: 209–213. doi:10.1016/j.cels.2016.03.001
21. Zielinski DC, Filipp FV, Bordbar A, Jensen K, Smith JW, Herrgard MJ, et al. Pharmacogenomic and clinical data link non-pharmacokinetic metabolic dysregulation to drug side effect pathogenesis. *Nat Commun*. 2015;6: 7101. doi:10.1038/ncomms8101
22. Blais EM, Rawls KD, Dougherty BV, Li ZI, Kolling GL, Ye P, et al. Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nat Commun*. 2017;8: 14250. doi:10.1038/ncomms14250

23. Pannala VR, Vinnakota KC, Rawls KD, Estes SK, O'Brien TP, Printz RL, et al. Mechanistic identification of biofluid metabolite changes as markers of acetaminophen-induced liver toxicity in rats. *Toxicol Appl Pharmacol.* 2019;372: 19–32. doi:10.1016/j.taap.2019.04.001
24. Rawls KD, Blais EM, Dougherty BV, Vinnakota KC, Pannala VR, Wallqvist A, et al. Genome-Scale Characterization of Toxicity-Induced Metabolic Alterations in Primary Hepatocytes. *Toxicol Sci Off J Soc Toxicol.* 2019;172: 279–291. doi:10.1093/toxsci/kfz197
25. Rawls KD, Dougherty BV, Vinnakota KC, Pannala VR, Wallqvist A, Kolling GL, et al. Genome-Scale Metabolic Model Predicts Changes in Renal Metabolism from Chemical Exposure. In prep.
26. Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol.* 2018;36: 272–281. doi:10.1038/nbt.4072
27. Pornputtapong N, Nookaew I, Nielsen J. Human metabolic atlas: an online resource for human metabolism. *Database J Biol Databases Curation.* 2015;2015. doi:10.1093/database/bav068
28. Swainston N, Smallbone K, Hefzi H, Dobson PD, Brewer J, Hanscho M, et al. Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics.* 2016;12: 109. doi:10.1007/s11306-016-1051-4
29. Smith AC, Eyassu F, Mazat J-P, Robinson AJ. MitoCore: a curated constraint-based model for simulating human central metabolism. *BMC Syst Biol.* 2017;11. doi:10.1186/s12918-017-0500-7
30. Karlstädt A, Fliegner D, Kararigas G, Ruderisch HS, Regitz-Zagrosek V, Holzhütter H-G. CardioNet: A human metabolic network suited for the study of cardiomyocyte metabolism. *BMC Syst Biol.* 2012;6: 114. doi:10.1186/1752-0509-6-114
31. Zhao Y, Huang J. Reconstruction and analysis of human heart-specific metabolic network based on transcriptome and proteome data. *Biochem Biophys Res Commun.* 2011;415: 450–454. doi:10.1016/j.bbrc.2011.10.090
32. Uhlen M, Hallstrom BM, Lindskog C, Mardinoglu A, Ponten F, Nielsen J. Transcriptomics resources of human tissues and organs. *Mol Syst Biol.* 2016;12: 862–862. doi:10.15252/msb.20155865
33. Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol Syst Biol.* 2014;10. doi:10.1002/msb.145122

34. Chang RL, Xie L, Xie L, Bourne PE, Palsson BØ. Drug Off-Target Effects Predicted Using Structural Analysis in the Context of a Metabolic Network Model. *PLoS Comput Biol.* 2010;6: e1000938. doi:10.1371/journal.pcbi.1000938
35. Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. Predicting selective drug targets in cancer through metabolic networks. *Mol Syst Biol.* 2011;7: 501–501. doi:10.1038/msb.2011.35
36. Richelle A, Chiang AWT, Kuo C-C, Lewis NE. Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. *PLoS Comput Biol.* 2019;15: e1006867. doi:10.1371/journal.pcbi.1006867
37. Sara JD, Kaur J, Khodadadi R, Rehman M, Lobo R, Chakrabarti S, et al. 5-fluorouracil and cardiotoxicity: a review. *Ther Adv Med Oncol.* 2018;10. doi:10.1177/1758835918780140
38. Yoon E, Babar A, Choudhary M, Kutner M, Pyrsopoulos N. Acetaminophen-Induced Hepatotoxicity: a Comprehensive Update. *J Clin Transl Hepatol.* 2016;4: 131–142. doi:10.14218/JCTH.2015.00052
39. Mazer M, Perrone J. Acetaminophen-induced nephrotoxicity: pathophysiology, clinical manifestations, and management. *J Med Toxicol Off J Am Coll Med Toxicol.* 2008;4: 2–6. doi:10.1007/bf03160941
40. Igarashi Y, Nakatsu N, Yamashita T, Ono A, Ohno Y, Urushidani T, et al. Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res.* 2015;43: D921–D927. doi:10.1093/nar/gku955
41. Waters MD, Fostel JM. Toxicogenomics and systems toxicology: aims and prospects. *Nat Rev Genet.* 2004;5: 936–948. doi:10.1038/nrg1493
42. Dougherty BV, Papin JA. Systems biology approaches help to facilitate interpretation of cross-species comparisons | Elsevier Enhanced Reader. [cited 5 Oct 2020]. doi:10.1016/j.cotox.2020.06.002
43. Burridge PW, Li YF, Matsa E, Wu H, Ong S, Sharma A, et al. Human Induced Pluripotent Stem Cell–Derived Cardiomyocytes Recapitulate the Predilection of Breast Cancer Patients to Doxorubicin–Induced Cardiotoxicity. *Nat Med.* 2016;22: 547–556. doi:10.1038/nm.4087

## Chapter 2: Identifying functional metabolic shifts in heart failure with the integration of omics data and a heart-specific, genome-scale model

Bonnie V. Dougherty<sup>1</sup>, Kristopher D. Rawls<sup>1</sup>, Glynis L. Kolling<sup>1,2</sup>, Kalyan C. Vinnakota<sup>3,4</sup>, Anders Wallqvist<sup>3</sup>, Jason A. Papin<sup>\*1,2,5,6</sup>

<sup>1</sup>Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, 22908, USA.

<sup>2</sup>Department of Medicine, Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, VA, 22908, USA.

<sup>3</sup>Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Development Command, Fort Detrick, Maryland, 21702 USA.

<sup>4</sup>The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, Maryland, 20817, USA.

<sup>5</sup>Department of Biochemistry & Molecular Genetics, University of Virginia, Charlottesville, VA, 22908, USA.

<sup>6</sup>Lead contact

\*Corresponding author. Tel: +1 434 924 8195; Email: papin@virginia.edu

The text included in this chapter has been published here:

Dougherty BV, Rawls KD, Kolling GL, Vinnakota KC, Wallqvist A, Papin JA. Identifying functional metabolic shifts in heart failure with the integration of omics data and a cardiomyocyte-specific, genome-scale model. bioRxiv. 2020; 2020.07.20.212274. doi:10.1101/2020.07.20.212274

As of the writing of this dissertation, the manuscript has been submitted to Cell Reports.

## 2.1 Summary

The heart is a metabolic omnivore, known to consume many different carbon substrates in order to maintain function. In diseased states, the heart's metabolism can shift between different carbon substrates; however, there is some disagreement in the field as to the metabolic shifts seen in end-stage heart failure and whether all heart failure converges to a common metabolic phenotype. Here, we present a new, validated heart-specific GEnome-scale metabolic Network REconstruction (GENRE), *iCardio*, and use the model to identify common shifts in metabolic functions across heart failure omics datasets. We demonstrate the utility of *iCardio* in interpreting heart failure gene expression data by identifying Tasks Inferrred from Differential Expression (TIDEs) which represent metabolic functions associated with changes in gene expression. We identify decreased NO and Neu5Ac synthesis as common metabolic markers of heart failure across datasets. Further, we highlight the differences in metabolic functions seen across studies, further highlighting the complexity of heart failure. The methods presented for constructing a tissue-specific model and identifying TIDEs can be extended to multiple tissue and diseases of interest.

## 2.2 Introduction

In order for the heart to maintain its function, it uses oxygen and various nutrients through metabolic pathways to maintain contractile proteins, synthesize necessary lipid species, and produce ATP as a fuel for muscle contraction. Metabolic changes, such as changes in substrate utilization, have been noted in many diseased states, such as left ventricular hypertrophy [1] and cardiotoxicity [2,3]. In some cases, changes in substrate utilization occur before functional and/or structural changes to the heart, suggesting that metabolism plays a key role in the downstream development of disease or could be a target to prevent disease [1,4]. However, it is not understood if changes in heart metabolism are the result of, a contributor to, or the cause of disease. Therefore, there is a need for a comprehensive, descriptive model of the metabolic function of the heart to interrogate the relationships between substrates and function.

A common tool to interrogate relationships between metabolic substrates and metabolic function is a GENome-scale metabolic Network REconstruction (GENRE). GENREs are mathematical representations of metabolism that use the enzymes encoded in an organism's genome to define the biochemical metabolic reactions and associated metabolites that comprise that organism's metabolism. Each metabolic reaction is associated with a gene-protein-reaction (GPR) rule relating genes to the proteins they encode and proteins to the reactions they catalyze. Human GENREs account for the function of the biochemical reactions that humans catalyze according to annotation of the human genome.

However, not all genes are expressed in every tissue, necessitating the construction of tissue-specific models of metabolism. GENREs of human metabolism [5–11] have been used to generate tissue-, disease-, or cell-specific models for various analyses such as predicting drug targets [12–14], identifying biomarkers of disease [15,16], and understanding drug toxicity or side effects [5,17,18]. Tissue-specific models are built by integrating tissue-specific omics data, usually transcriptomic or proteomic data, with a human GENRE to obtain a tissue- or cell-type specific

model using various integration algorithms (some of which are summarized in [19,20]). To date, there are two existing heart models [16,21], both of which were built from Recon1, the first human GENRE [7]. These models were used to examine the relationship between substrate utilization and efficiency of the heart [21] and to predict epistatic interactions in the heart and biomarkers of heart disease [16]. However, human models have been expanded since Recon1 to more comprehensively describe human metabolism. Our recently published human GENRE, *iHsa* [5], is more comprehensive than Recon1 (8336 reactions vs. 3311 reactions) and therefore offers the potential to generate a more comprehensive heart-specific model. A new heart-specific model can be used as a tool to interpret large data sets, such as transcriptomic data, to provide functional insight into how metabolism might change in a diseased state.

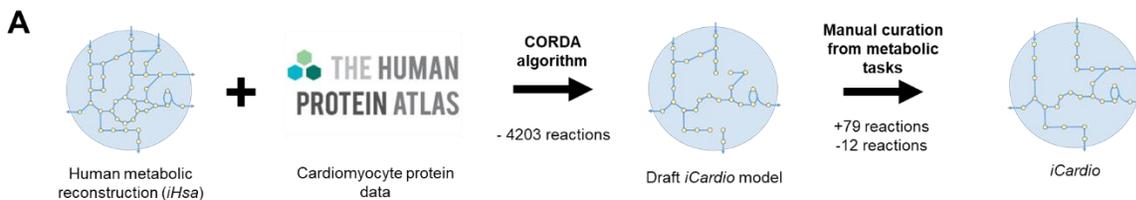
Here, we present a new, validated, heart-specific metabolic model, *iCardio*, which was built using *iHsa* [5] with data from the Human Protein Atlas (HPA) [22]. The draft model was curated using metabolic tasks, which are mathematical descriptions of metabolic functions that a model should be able to perform. The metabolic tasks represent both previously published tasks [5] as well as newly developed, heart-relevant tasks. We validated the model qualitatively by the number of reactions covered by the metabolic tasks and quantitative ATP predictions for common carbon sources. Finally, we demonstrate the utility of metabolic models in identifying functional changes in metabolism using a new method that identifies metabolic tasks associated with significant changes in gene expression, called Tasks Inferred from Differential Expression (TIDEs). TIDEs are a unique type of gene set enrichment analysis (GSEA) that take into account both the stoichiometric balance of reactions necessary to achieve metabolic functions by including transporter reactions between metabolic compartments as well as the complex relationship between genes, proteins, and the reactions they catalyze. Therefore, TIDEs offer biological insight into metabolic functions affected in a disease state that are not readily apparent from gene

expression data or gene set enrichment analyses alone. We use heart failure as a case study because of the critical role that metabolism plays in the progression and diagnosis of disease.

## 2.3 Results

### 2.3.1 Building and validating *iCardio* using metabolic tasks

The CORDA algorithm was used to build the draft *iCardio* model from *iHsa*; CORDA removed 4203 reactions from *iHsa* to build a draft *iCardio* model that had a heart-specific task accuracy of 89% (Figure 1B-C, Supplemental File 2). The draft model was then curated using metabolic tasks to obtain the final *iCardio* model (Figure 2.1A). To achieve 100% task accuracy, 79 reactions were added to and 12 removed from the draft *iCardio* model (Figure 2.1A) based on literature and manual curation (Supplemental File 2). The 79 reactions that were added to *iCardio* are reactions that were removed from *iHsa* to build the *iCardio* draft model were added back to the *iCardio* model to ensure functionality. We can map the protein data from HPA using the associated GPR rules. Of the 79 reactions added back to *iCardio*, 73% (58) are associated with either high (1), medium (8), or low (16) protein evidence, no GPR rule (28), or no data (5); the remaining 21 reactions were associated with no protein evidence. As an example, the synthesis of nitric oxide was a metabolic task that originally failed but was curated to pass in the final *iCardio*



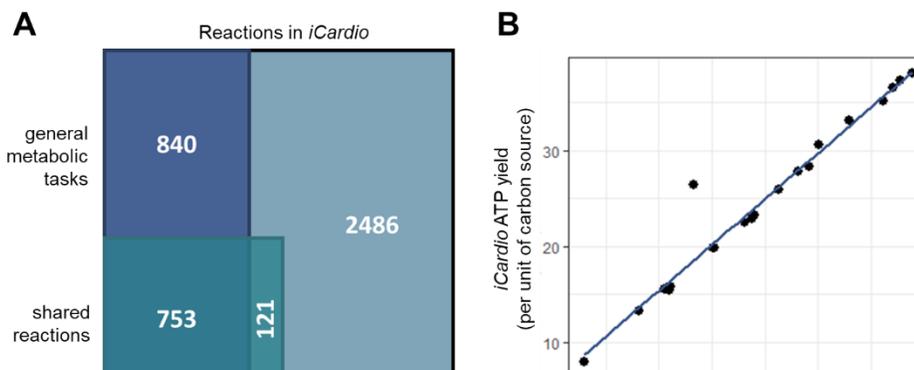
**Figure 2.1 Building a heart metabolic model (*iCardio*) by integrating protein data and curating with metabolic tasks.**

(A) The draft *iCardio* model was built by integrating protein data from the human protein atlas (HPA) with a human metabolic reconstruction (*iHsa*) using the CORDA algorithm. The draft *iCardio* model was curated using pre-defined metabolic tasks, resulting in the final model, *iCardio*. (B) The CORDA algorithm removed 4203 from *iHsa* to produce a draft *iCardio* model that had a heart-specific task accuracy of 89%. (C) The superset of 421 metabolic tasks represented metabolic tasks from a previous publication (Blais et al, 2017) with additional 112 metabolic tasks that were curated from the literature based on cardiac metabolism. These tasks were used to test if the model could perform functions known to be present in the heart while removing metabolic functions not present in the heart. After curation, the final *iCardio* model has a 100% heart-specific task accuracy.

model. For this task to pass, three reactions were added back to the model, two reaction that were associated with proteins that were not detected and one that had no GPR rule (Supplemental File 2.2). All three isoforms of nitric oxide synthase (NOS), which catalyze the two reactions, were not detected in v18 of the HPA tissue data but two of the NOS isoforms, NOS1 and NOS2, are associated with low expression in v19, while NOS3 has significantly higher transcript counts in heart muscle in the RNA consensus tissue HPA data [22]. Of note, after curation, the final *iCardio* model contains approximately 50% of the reactions that are present in the *iHsa* network reconstruction. Using the complex GPR rules that are associated with each reaction, we can also identify the number of genes that are associated with a reaction in *iHsa* but are not associated with a reaction in the *iCardio* model; only 583 genes are associated with a reaction in *iHsa* but not included in *iCardio*, highlighting both the promiscuity of enzymes and the complex relationship between genes and metabolic functions (Figure 2.1B). Since metabolic tasks have been used as a metric for building *iCardio*, the number of reactions covered by this set of metabolic tasks provides a qualitative validation of the model. As has been done with other models [23], pFBA was used to determine the reactions that each passing metabolic task utilized in *iCardio*. The 216 previously published heart-relevant tasks [5] covered 38% (1593/4200) of reactions in *iCardio* and the 93 passing new heart-relevant tasks covered 21% (874/4200) of reactions in *iCardio* (Figure 2A). There is overlap between these two sets of reactions; in total, the two sets of tasks covered 41% (1714/4200) of reactions in *iCardio*. It is important to note that although the tasks may cover the same reactions, they cover different combinations of reactions and pathways for each task. The two sets of tasks together may cover 1,700 reactions but in total over 20,000 reactions are used in different combinations to complete the tasks, indicating that a number of reactions are repeated between tasks. This result is to be expected, especially for reactions that involve central carbon metabolism and ATP production, such as ATP synthase. The maximum number of reactions covered by a given task was 788 reactions (1 task), the metabolic task that describes

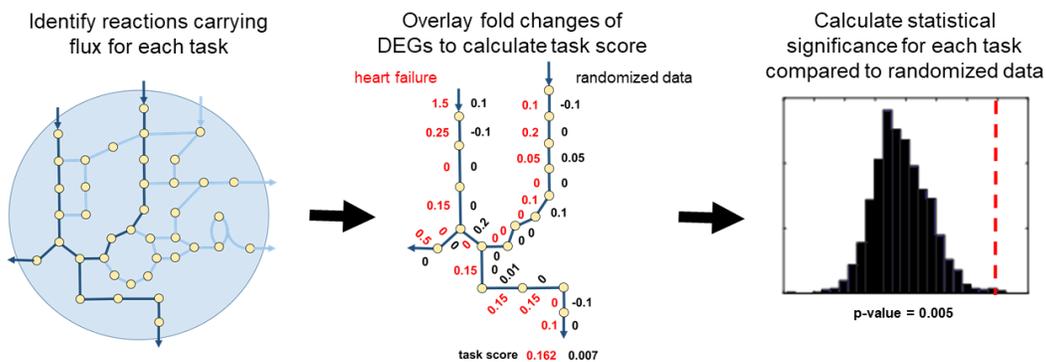
the de novo synthesis of lipids from glucose and essential fatty acids, and the minimum number of reactions covered by a task was one reaction (24 tasks), metabolic tasks that describe transport. Overall, the reaction coverage of tasks demonstrates a qualitative validation of *iCardio*.

To provide a quantitative validation of *iCardio*, ATP yields were predicted for a number of carbon substrates and amino acids and compared to another metabolic model, *MitoCore* [24]. The *MitoCore* model was chosen for comparison because of its focus on mitochondrial metabolism. For almost all carbon sources (other than methionine), *iCardio* predicted ATP yields within 10% of the values calculated with *MitoCore* [24] (Figure 2.2B). For methionine, the ATP prediction from *iCardio* matches the methionine prediction from Recon 2.2 [10]. It is important to note the difference in scope between the two models: *MitoCore* contains 342 reactions whereas *iCardio* contains 4200 reactions while still maintaining accurate ATP yields. This result highlights that, even with the increased size of *iCardio*, there are not infeasible energy-generating loops, which would artificially inflate ATP yield predictions and influence the reactions necessary for different metabolic tasks. Together, the qualitative and quantitative validation of *iCardio* demonstrate the ability of the model to accurately and more comprehensively represent heart metabolism.



**Figure 2.2 Validating *iCardio* (A) descriptively using the number of reactions covered by metabolic tasks and (B) quantitatively using ATP yields for common carbon substrates.**

(A) A descriptive validation of *iCardio* using the number of reactions covered using the different metabolic tasks. Together, the tasks account for almost half of the reactions in *iCardio*. The remaining reactions represent areas for future improvement of metabolic tasks. (B) A qualitative validation of *iCardio* using ATP yields for a variety of common carbon sources. The ATP yields for *iCardio* (y-axis) are compared to another recently published, but smaller, metabolic model, *MitoCore* (Smith et al., 2017), which contains 324 reactions. The agreement between the models demonstrates the lack of energy generating cycles and infinite loops in *iCardio*.



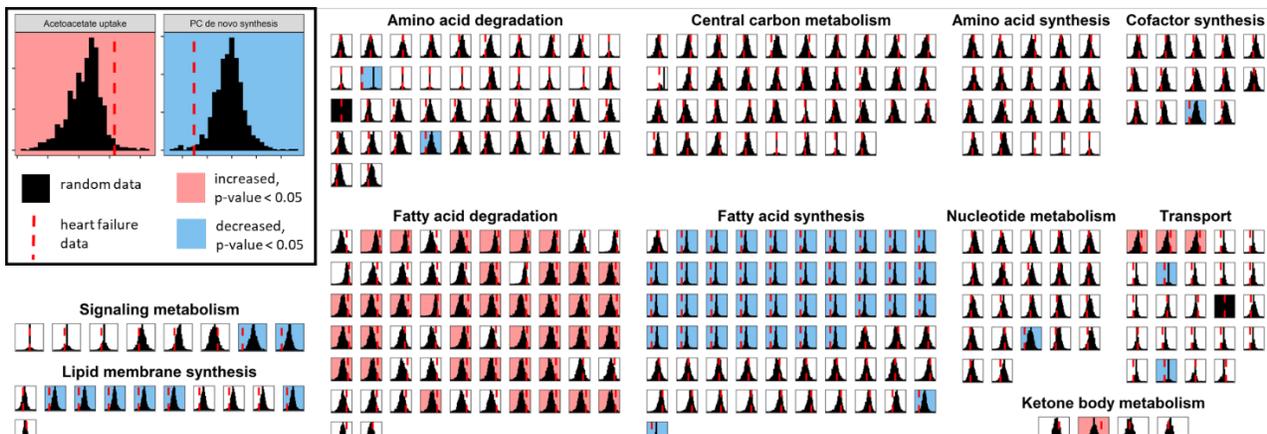
**Figure 2.3 Identifying functional metabolic changes from gene expression data using metabolic tasks.** *Metabolic tasks quantitatively describe metabolic functions that a tissue or organism is known to catalyze. iCardio was used to identify the reactions that are utilized in order to perform each of the previously defined metabolic tasks. Next, each reaction is assigned a reaction weight based on gene expression fold changes. A task score is calculated based on the weights for each reaction and represents an average gene expression value over reactions utilized in that metabolic task. In order to determine the statistical significance of this task score, the original gene expression data is shuffled over all genes with data and task scores are recalculated to give a distribution of task scores. A p-value is assigned based on where the original task score falls in the distribution of randomized data.*

### 2.3.2 Identifying Tasks Inferred from Differential Expression (TIDEs) using iCardio for heart failure gene expression data

Metabolic models provide an alternative approach for interpreting changes in gene expression to yield insight into metabolic shifts that may be contributing to a diseased state. Here, we use *iCardio* to identify metabolic tasks that were significantly associated with differentially expressed genes, called Tasks Inferred from Differential Expression (TIDEs), which represent metabolic functions that are shifted in an experimental versus control state (Figure 2.3). Heart failure is a complex disease, both in etiology and presentation, but heart failure is associated with a shift in metabolism [25]. Several different metabolic shifts have been noted in heart failure, such as decreased fatty acid utilization [25,26], increased ketone body utilization [27], and decreased utilization of branched chain amino acids [28]. However, there are still disagreements in the field. For example, one study noted an increase in fatty acid uptake in heart failure [29], while other studies reveal decreased fatty acid utilization [30]. *iCardio* provides an opportunity to contextualize gene expression data from end-stage heart failure patients to identify changes in

metabolic functions, such as those listed above, that are associated with significant changes in gene expression.

To do this, we analyzed gene expression data from patients undergoing heart transplants for either advanced ischemic or idiopathic heart failure compared to healthy hearts. We integrated these differentially expressed genes (DEGs) and their associated log fold changes with *iCardio* to determine TIDEs, representing metabolic functions associated with a significant change in gene expression. When compared to healthy hearts, the ischemic hearts from GSE5406 had 2678 DEGs; 392 of these DEGs were present in *iCardio* (Supplemental Table 2.3). After integrating these DEGs using *iCardio* and the TIDEs pipeline, 94 of the 307 metabolic tasks were designated as TIDEs (Figure 2.4), representing metabolic functions associated with significant changes in gene expression in these heart failure conditions. In the randomized data used to determine the statistical significance of each task, only 5 out of the 1000 random iterations had more than 94 tasks identified as significantly changed, indicating that the identified TIDEs cannot be attributed to random changes in gene expression, but rather distinct and coordinated shifts in gene expression resulting in changes in specific metabolic functions. Across the 307 metabolic tasks,



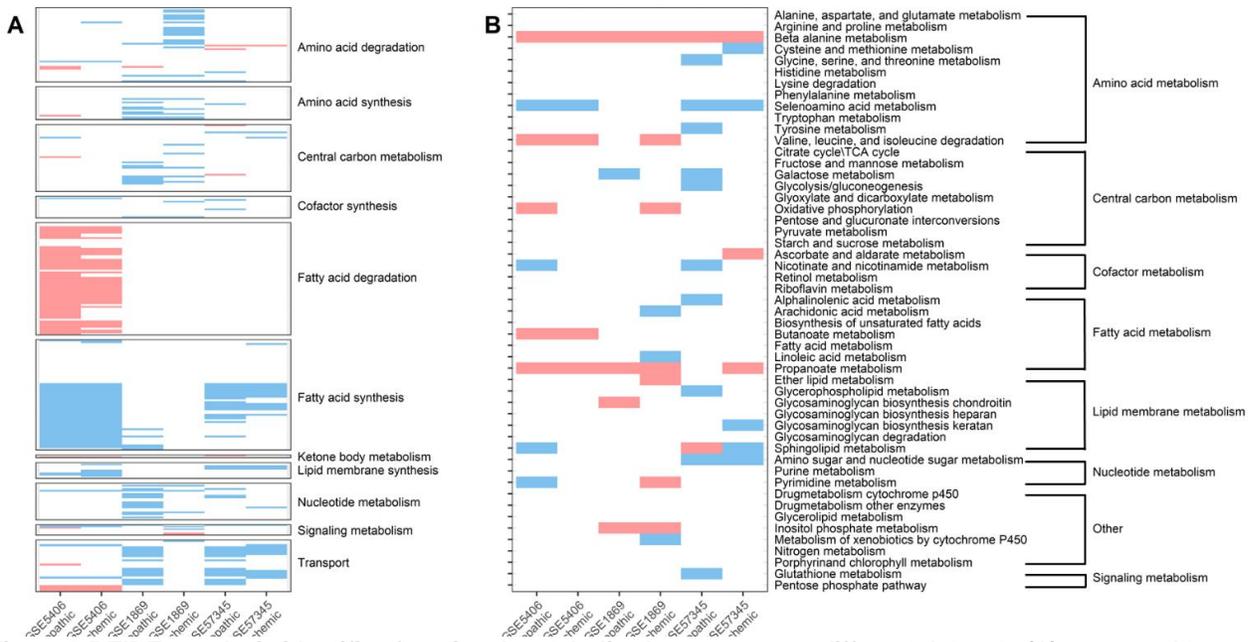
**Figure 2.4** Distributions for model-predicted task scores from gene expression data for ischemic heart failure for each of the 307 tested tasks.

Gene expression data for ischemic heart failure vs. healthy hearts from GSE5406 was integrated with *iCardio* to identify shifts in metabolic functions. The red dashed lines indicate the task score for the heart failure dataset and the black histograms are the distribution of randomized task scores. A red line to the right indicates a metabolic function associated with increased gene expression whereas a red line to the left indicates a metabolic function associated with decreased gene expression. The color of the background indicates if the metabolic function had a statistically significant ( $p$ -value < 0.05) increased (red) or decreased (blue) task score.

there are varied distributions in the underlying randomized task scores (Figure 2.4), representing the complex relationships between gene expression and metabolic function that are captured using *iCardio* and the TIDEs reaction-centric approach.

Some general trends appear from the TIDEs identified from the ischemic heart failure data (Figure 4). First, metabolic tasks related to fatty acid degradation (41) were increased in the heart failure dataset compared to randomized data while metabolic tasks related to fatty acid synthesis (38) were decreased. Within tasks related to fatty acid synthesis, significant decreases were observed for saturated fatty acids but not for unsaturated fatty acids. Finally, metabolic tasks related to lipid membrane synthesis were decreased (6). Taken together, these metabolic shifts seem to support the common hypothesis of heart failure as a state of increased demand for ATP [31,32], seen through an increased degradation of fatty acids, while limiting other uses of ATP and carbon sources for other functions, such as synthesis of fatty acids and components of the lipid membrane. However, other previously reported metabolic signatures of heart failure, such as increased ketone body degradation [27], and decreased breakdown of branched chain amino acids [28], were not identified using this dataset and approach.

Next, we expanded the TIDE analysis to the two remaining studies (Supplemental Table 2.3), including both idiopathic and ischemic heart failure. Here, there is no consistent TIDE across all 6 datasets (Figure 2.5A). However, there were a few frequently observed changes in TIDEs across the datasets. First, nitric oxide synthesis from arginine is decreased in 5 of the 6 datasets (excluding the ischemic heart failure data from GSE1869). Synthesis of long-chain, unsaturated fatty acids was decreased in 4 of the 6 datasets (excluding both ischemic and idiopathic heart failure data from GSE1869). De novo synthesis of Neu5Ac was decreased in 4 of the 6 datasets (excluding the ischemic and idiopathic heart failure data from GSE57345). The breakdown of valine to succinyl-CoA was increased in 2 of the 6 datasets (the idiopathic heart failure data from GSE1869 and GSE5406) and the breakdown of threonine to a Krebs cycle intermediate was increased in 2 of the 6 datasets (both ischemic and idiopathic heart failure in GSE57345). Finally, a metabolic task related to ketone body metabolism was increased in 3 of the 6 datasets, both



**Figure 2.5 TIDEs analysis identifies functional metabolic changes across different datasets (A) compared to results from a traditional GSEA analysis (B).**

(A) Three datasets that contained samples for ischemic, idiopathic, and healthy hearts were downloaded, analyzed, and integrated with iCardio using the TIDEs pipeline to identify shifts in metabolic functions. For each of the three datasets, both ischemic and idiopathic heart failure are shown. TIDEs cluster by study rather than type of heart failure. (B) The same three datasets were used with a traditional GSEA analysis with gene sets defined by KEGG pathways. The KEGG metabolic pathways are shown here and grouped similar to the TIDE groupings shown in (A).

ischemic and idiopathic heart failure data from GSE5406 and idiopathic heart failure from GSE57345. Across the different datasets, the previously reported metabolic signatures of heart failure appear. However, they tend to appear within a study rather than across all studies or types of heart failure. Finally, rather than clustering by type of heart failure (ischemic vs idiopathic), the TIDEs results cluster by the study.

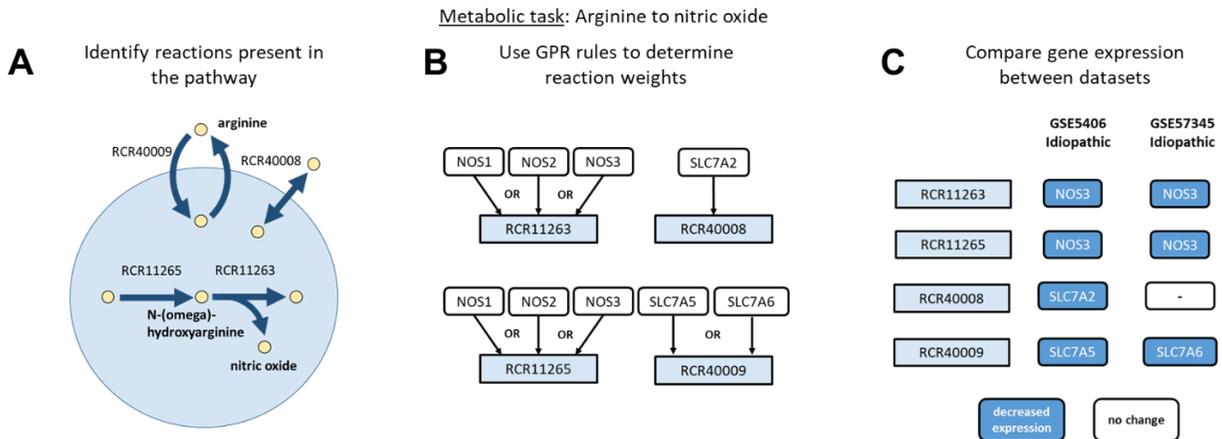
### 2.3.3 *Comparison of TIDEs with GSEA by KEGG pathway*

Next, we wanted to compare the TIDEs analysis to a common gene-centric approach, gene set enrichment analysis (GSEA), using KEGG pathways to define the gene sets. Rather than taking a reaction-centric approach, GSEA takes a gene-centric approach and defines a group of genes for specific functions [33]. While KEGG includes many other pathways other than metabolism (Supplemental Figure 1), we have chosen to focus on a comparison of metabolic pathways. Using GSEA with the data from the studies cited above (Figure 2.5B), the most commonly changed metabolic pathways across the datasets were increased beta alanine metabolism in all datasets, increased propanoate metabolism in 5 of the 6 datasets, decreased selenoamino acid metabolism in 4 of the 6 datasets and increased valine, isoleucine, and leucine degradation in 3 of the 6 datasets. For the commonly cited metabolic shifts in heart failure, the GSEA analysis shows no change in fatty acid metabolism, one dataset with a decrease in genes associated with glycolysis, two studies that show increased oxidative phosphorylation, and 3 datasets with increased valine, isoleucine, and leucine metabolism. For comparison with the metabolic tasks, the breakdown of propanoate and synthesis of beta-alanine were included as metabolic tasks task but showed no change across datasets. Selenoamino acids are included in the model but are not included as a metabolic task. Valine, isoleucine, and leucine metabolism cover six metabolic tasks in the model covering uptake and breakdown of each amino acid. Valine, isoleucine, and leucine were associated with decreased uptake in idiopathic heart failure for

GSE1869 and GSE57345. Breakdown of valine to succinyl-CoA was increased in idiopathic heart failure for GSE1869 and GSE5406 while the breakdown of isoleucine to acetyl-CoA and succinyl-CoA was increased in idiopathic heart failure for GSE5406.

The difference in identified changes in metabolism highlight the differences between a reaction-centric approach, such as TIDEs, versus a gene-centric approach. GSEA KEGG metabolic pathways can cover more than one metabolic function, such as fatty acid metabolism, while the TIDEs approach allows for a distinction between the synthesis and degradation of different fatty acids. We can use *iCardio* for a similar approach by using the metabolic subsystems assigned to each reaction to identify a set of reactions. Using TIDEs, we can determine if the reactions in a given subsystem are associated with increased or decreased gene expression (Supplemental Figure 2.2). While this approach also reveals some interesting trends, such as a decrease in expression for reactions in the arachidonic acid metabolic subsystem, it fails to highlight changes in fatty acid metabolism or other smaller pathways, such as nitric oxide synthesis or ketone body degradation.

Second, a GSEA approach includes every gene that can be present in the pathway even if the function is redundant, such as with isozymes. For example, take the metabolic task of synthesis of nitric oxide from arginine (Figure 2.6). Using the TIDEs approach with *iCardio*, this metabolic task covers 5 reactions, 4 of which are associated with GPR rules and therefore included in downstream analysis (Figure 2.6A). Two of these reactions can be catalyzed by one of three enzymes, NOS1, NOS2, or NOS3, which are known to have tissue-specific expression. The TIDEs reaction-centric approach uses the log-fold change of one gene to determine the reaction weight, which could differ between studies and across tissues. In this study, the two reaction weights for the transporters responsible for the import of arginine to and the export of citrulline from the cytosol are determined by different genes between the idiopathic heart failure



**Figure 2.6 TIDEs analysis identifies differences in datasets for genes which catalyze the conversion of arginine to nitric oxide.**

(A) For the metabolic task of arginine to nitric oxide, the *iCardio* model uses four reactions, which cover the transport of arginine and citrulline into the cell and the conversion of arginine to the intermediate *N*-( $\omega$ )-hydroxyarginine and finally conversion to nitric oxide. (B) The model uses the GPR rules associated with each of these reactions to determine the reaction weight based off gene expression data, thereby assigning the expression of one gene to govern the reaction. (C) The gene whose gene expression was used in the analysis can therefore differ between different datasets while still producing the same result, as seen here with both datasets still showing a statistically significant decrease the metabolic function of conversion of arginine to nitric oxide.

data in GSE5406 and GSE57345 (Figure 2.6C). However, the metabolic task of nitric oxide production from arginine was associated with decreased gene expression in both datasets. A similar GSEA analysis would have included all 6 genes rather than the 2-3 used in the TIDEs analysis.

## 2.4 Discussion

Here, we present a new, validated model of heart metabolism, *iCardio*, with a method for analyzing changes in metabolic functions based on gene expression data. We present a new approach for building tissue-specific models using tissue-specific protein data from the HPA integrated with a general human reconstruction, *iHsa*, followed by manual curation with metabolic tasks to ensure general metabolic functionality. Using a task-driven approach for model curation demonstrated that the CORDA method produced a draft model that was more accurate with respect to performance of heart-specific metabolic tasks than other integration methods (Supplemental Table 2.2). It is interesting to note that the CORDA algorithm produced one of the

smaller draft models while still maintaining the highest task accuracy. As noted, although the CORDA algorithm produced one of the smaller draft models and resulted in the final *iCardio* model that contained 50% of the reactions in *iHsa*, 75% of the genes in *iHsa* are still represented in the *iCardio* model. This result highlights the ability of the complex GPR rules to capture tissue-specific function. For example, the conversion of arginine to nitric oxide (NO) is an established mechanism for intracellular and extracellular signaling and the production of NO is mediated through tissue-specific expression of different NOS isoforms (Figure 2.6). *iCardio* captures the complex relationship between gene expression and function through the complex GPR rules even though all of the isoforms were not detected in the original HPA dataset.

The metabolic tasks used for benchmarking the integration algorithms and further manual curation cover 41% of the network while also identifying areas for future development of both general and tissue-specific metabolic tasks. Here, metabolic tasks were formulated agnostic to the gene expression data that was integrated with the resulting model. Additional tasks could serve as hypotheses for changes in metabolic functions of the heart that could then be tested with specific gene expression data sets. Finally, although not every reaction is covered by a metabolic task, these reactions represent metabolic functions with either evidence of protein expression in the HPA or pathway connectivity from the CORDA integration algorithm. These reactions and associated pathways further support the use of metabolic network reconstructions for generating hypotheses for important, tissue-specific metabolic functions.

The pFBA approach assumes that the pathway for each metabolic task remains the same, independent of the data. This reaction-centric TIDE approach for determining significantly changed metabolic tasks emphasizes that (a) metabolic functions require multiple complex, stoichiometrically balanced reactions in order to proceed, and (b) some genes may disproportionately influence the completion of metabolic functions. For example, two metabolic tasks cover the synthesis of nitric oxide from arginine (Supplemental File 1). The first task

comprehensively covers the entire pathway, providing only arginine and other ions extracellularly as inputs while requiring the production of nitric oxide and the release of only metabolites that have transport reactions from the cytosol to the extracellular space. The second task only accounts for the central reactions for nitric oxide production, providing arginine and NADPH and require the production of nitric oxide. While both tasks cover the synthesis of nitric oxide from arginine, the first metabolic task uses 66 reactions and covers the entire pathway, including transport of metabolites between compartments, which are necessary to maintain the stoichiometric balance of the pathway. However, because the second metabolic task covers mainly reactions in the cytosol, it requires only 5 reactions. While the first task was present in the original *iHsa* task, the second task was added because of the importance of the synthesis of nitric oxide in the heart and therefore covers the core reactions for nitric oxide synthesis. Together, these two tasks highlight the potential for both breadth and specificity in identifying shifts in metabolic functions.

Second, reactions are associated with complex GPR rules, allowing for multiple genes representing different isozymes to catalyze the same reaction. By accounting for these complex relationships in the calculation of task scores, the expression values of individual genes can affect multiple functions. For example, fatty acid synthase, which catalyzes multiple reactions, can disproportionately influence a final task score. The associated fold change for fatty acid synthase may determine the reaction weight for multiple reactions in a metabolic task. In contrast, previous gene-centric approaches would have evenly weighted all genes in a specific pathway. The TIDEs analysis represents a reaction-centric approach that focuses on metabolic functions that can provide insight into broad metabolic changes that may not be immediately apparent using gene-centric approaches.

The presented TIDEs pipeline offers an alternative, reaction-centric approach to interpret complex changes in gene expression data to identify non-obvious changes in metabolic functions.

In the case of heart failure, the TIDEs pipeline was able to identify some of the common shifts noted in heart failure in at least one of the three studies, such as increased fatty acid utilization, increased utilization of branched chain amino acids, and increased utilization of ketone bodies. The TIDEs analysis also revealed two changes that were common across more than one dataset: decreased synthesis of nitric oxide from arginine and decreased de novo synthesis of Neu5Ac. With an evaluation of all six data sets, decreased expression of NOS3, also referred to as eNOS, was driving the weight of two of the reactions in the metabolic task (Figure 2.6). Previous work has highlighted the important role of NO in cardiac function [34,35] and more recent work has suggested a role for increasing NO synthesis for the increasing efficacy of beta-blockers [36]. Neu5Ac is the most common sialic acid associated with multiple functional roles in the body. Studies have shown that increased Neu5Ac is associated with the development of atherosclerosis and synthesis of Neu5Ac has been suggested as a therapeutic target [37]. Using TIDEs, we can identify genes that are driving the change in the metabolic task, both genes that are shared across and are different between studies which can serve as a starting point for future work. However, more work is needed to identify the role that decreased synthesis of Neu5Ac may play in end-stage heart failure. Both NO and Neu5Ac synthesis represent potential biomarkers of and targets for treating heart failure.

Although there were some common trends across datasets, no TIDEs could discriminate between ischemic and idiopathic heart failure across studies. This characteristic was also true for the KEGG GSEA results. Together, this result highlights the complexity of heart failure, both in etiology and presentation, suggesting that classifications such as ischemic and idiopathic may be insufficient to capture distinct metabolic changes. In addition, the datasets cluster within each study for both the TIDEs and GSEA KEGG metabolic pathway analysis, suggesting, again, that there is a large amount of heterogeneity in changes in gene expression for heart failure. However, it is important to note that changes in gene expression are not the only drivers of changes in

metabolic function. Other studies have noted the role of changing metabolic milieu in the blood as a driver of changes in the metabolic functions of the heart during heart failure [32]. Future work can integrate clinical measures, such as LVEF or FDG-PET glucose uptake measures, which could help to separate clusters of patients and more clearly identify the influence of metabolism in the progression of heart failure.

Here, we provide a validated approach for constructing a tissue-specific metabolic model and demonstrate the utility of metabolic models to interpret changes in metabolic functions based on gene expression data. The model-building process can be extended to other cell- or tissue-type specific models. The metabolic tasks provide a two-fold role for model validation and concrete metabolic functions to identify metabolic shifts in gene expression data. TIDEs represent a new, reaction-centric approach to identifying changes in metabolic functions and testing new hypotheses for changes in gene expression for metabolic functions. These new hypotheses can be formulated as metabolic tasks based on the current literature, based on reactions present in a metabolic model but not covered in the current list of metabolic tasks, or results from other gene-centric approaches, such as GSEA. The method is not limited to use with *iCardio* but can be used with any published metabolic model that includes GPR rules. Further, the method can be used with either microarray or RNA-seq data. We demonstrate that a heart-specific model, *iCardio*, with the TIDEs pipeline was able to identify decreased NO synthesis and decreased Neu5Ac synthesis across different heart failure datasets that were not identified using conventional gene-centric approaches, such as gene set enrichment analysis.

## 2.5 Methods

### 2.5.1 Developing heart-relevant metabolic tasks

Metabolic tasks describe metabolic functions that a tissue or organism is known to be able to catalyze. These metabolic tasks are represented as mathematical constraints on input and

output metabolites to the model, where a task is considered to “pass” if there is a feasible flux distribution through the model with the specified constraints. Metabolic tasks have been published with metabolic network reconstructions to demonstrate general metabolic function [5,11]. For example, *iHsa* was published with 327 metabolic tasks which describe both general metabolism, *i.e.* the generation of ATP from glucose, as well as liver specific metabolism, *i.e.* bile acid synthesis. To expand upon the general metabolic functions in the *iHsa* task list, we curated 94 new, heart-relevant tasks, including 25 that quantitatively describe ATP generation from various carbon sources. Testing these new heart-relevant tasks with *iHsa* resulted in five changes to the network reconstruction (Supplementary Table 2.1), generating an updated human model which served as the starting point for building the heart-specific model, *iCardio*. Two notable changes were (a) the addition of reactive oxygen species (ROS) formation as 0.1% of the flux through Complex I of the electron transport chain (ETC) as has been done with another model [38] and (b) the change from 4 protons to 2.7 protons to generate one ATP molecule to be consistent with recently published data [39]. The heart-relevant and general metabolic tasks together represent 421 metabolic tasks (Supplemental File 2.1) that cover a wide range of metabolic functions that both do and do not occur in the heart and therefore serve as a resource for curation of the draft *iCardio* model to ensure model functionality (Figure 2.1).

### 2.5.2 *Generating and curating a heart-specific metabolic model*

The general human model, *iHsa*, was able to pass all metabolic tasks successfully, but achieved a heart-specific task accuracy of 78% since *iHsa* was able to pass the metabolic tasks known to not occur in the heart. Since *iHsa* covers all human metabolism, it was necessary to prune reactions from *iHsa* that do not have evidence for presence in the heart. To do this, we integrated protein data from the HPA [22] with *iHsa* [5]. Various algorithms have been published to generate tissue-specific models from tissue transcriptomics or proteomics data. Here, we

implemented 5 of these algorithms [40–44] to generate draft heart-specific models. For GIMME, ATP production was used as the objective function. Data was integrated from the HPA, which contains tissue-specific protein expression where each protein is assigned either a high, medium, low, or no expression based on data from antibody-based immunohistochemistry [22]. Code for implementing each algorithm is available at <https://github.com/csbl/iCardio>. The CORDA algorithm was chosen from among the algorithms given its accuracy for the pre-defined list of heart-specific metabolic tasks (Supplemental File 2.1, Supplemental Table 2.2).

The CORDA algorithm takes as an input user-defined high, medium, and negative confidence reactions to produce a model that is (1) consistent (*i.e.* all reactions can carry flux) and (2) maximizes high and medium confidence reactions while minimizing the number of negative confidence reactions. Proteins that corresponded with high, medium, or low/no expression in the heart as indicated in the HPA were included as high ( $n = 1005$ ), medium ( $n = 2168$ ), or no confidence reactions ( $n = 5163$ ) respectively based on the model's GPR rules. Reactions without GPR rules (~2300 reactions) or reactions associated with no data were included in the negative confidence reactions.

### 2.5.3 Validating *iCardio* using reaction coverage and ATP yields

*iCardio* was validated qualitatively by determining the number of reactions covered by each metabolic task and quantitatively by comparing ATP yields for common carbon sources between *iCardio* and another metabolic model, *MitoCore* [24]. Parsimonious flux balance analysis (pFBA) [45] determines the lowest sum of fluxes, and therefore reactions, necessary to complete an objective. Here, pFBA was used to identify the reactions necessary for each metabolic task. Previous work has shown that, in most cases, pFBA produces more physiologically relevant flux distributions compared to flux balance analysis (FBA) or flux-based algorithms which incorporate data [46]. As a final, quantitative validation step, we calculated ATP yields for a variety of carbon

sources *in silico* as the maximum flux through the ATP synthase reaction for one unit of each carbon source and compared the results to a smaller heart-specific model of mitochondrial metabolism [24].

#### 2.5.4 Analyzing transcriptomics data from heart failure patients

Microarray data (Supplemental Table 2.3) [47–49] from patients undergoing heart transplants for advanced heart failure were downloaded from the Gene Expression Omnibus (GEO) database. Datasets were selected that (a) contained samples for both ischemic and idiopathic heart failure and (b) resulted in at least 50 differentially expressed genes for each type of heart failure. Since all the datasets had been background corrected using RMA, the limma package in R [50] was used to determine genes that were differentially expressed between healthy hearts and ischemic or idiopathic hearts. Genes corresponding to expression values with an FDR < 0.1 were considered to be differentially expressed and their corresponding fold change was used in subsequent analysis.

#### 2.5.5 Identifying Tasks Inferred from Differential Expression (TIDEs) using *i*Cardio with expression data

Metabolic tasks and their associated reactions, as identified using *i*Cardio with pFBA, were used to identify metabolic functions that are significantly associated with differentially expressed genes in a condition of interest. This method is referred to as Tasks Inferred from Differential Expression (TIDEs) (Figure 2.3). A total of 307 metabolic tasks was used for this analysis, representing the tasks functionally present in *i*Cardio from the original task list (Supplemental File 2.1) that also contained at least one reaction with an associated GPR rule (Supplemental File 2.3). Reactions that carry flux for each task are identified using a pFBA assumption without previous knowledge of the gene expression data (Figure 2.3a), as has been done with a related approach [51]. Gene expression log fold changes are overlaid onto reactions in the network using

the GPR rules to give each reaction a weight. GPR rules represent the proteins necessary to catalyze a specific reaction through AND or OR relationships. The AND relationships represent a protein complex where different genes encode unique protein subunits necessary for enzyme function while OR relationships represent isozymes. Reactions with complex GPR rules are assigned the maximum absolute fold change for OR relationships and the minimum fold change for AND relationships. For OR relationships where there is a disagreement in the direction of change, *i.e.* where there are genes associated with both a positive and negative fold change, the positive fold change is taken as the reaction weight. The assigned reaction weight values across all reactions in a task are averaged to calculate the task score (Figure 2.3b). To assign statistical significance to these task scores, the gene expression fold changes are randomized 1000 times amongst the genes measured in each dataset and task scores are recalculated based on the randomized data to create a distribution of task scores. The p-value for each task score that corresponded to the original data is calculated as the number of random task scores greater/less than the original data, depending on how the task score falls relative to the mean randomized task score (Figure 2.3c). TIDEs are identified as tasks with a p-value < 0.025. A p-value of 0.025 was chosen over a p-value of 0.05 based on the use of a two-sided t-test for calculating significance. Finally, *iCardio* was also used to identify reactions that belonged to each metabolic subsystem in the model. These sets of reactions were also used with the TIDEs method to identify reaction subsystems that were significantly associated with changes in gene expression data. Code is available for re-producing this analysis in MATLAB <https://github.com/csbl/iCardio>.

### 2.5.6 Gene set enrichment analysis

The same gene expression datasets (GSE1869, GSE5406, GSE57345) were also analyzed using gene set enrichment analysis (GSEA) [33] using the pre-defined gene sets by KEGG pathways. To more closely replicate the TIDE analysis, genes were shuffled within a

sample rather than shuffling across samples within each dataset. Pathway enrichment scores with a nominal p-value < 0.05 were defined as statistically significant.

## **2.6 Acknowledgements**

The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or of the U.S. Department of Defense, or The Henry M. Jackson Foundation for Advancement of Military Medicine, Inc. This paper has been approved for public release with unlimited distribution. The authors were supported by the U.S. Army Medical Research and Development Command, Ft. Detrick, MD, as part of the U.S. Army's Network Science Initiative. Support for this project was provided by the United States Department of Defense (W81XWH-14-C-0054 to JP) and the National Science Foundation Graduate Research Fellowship Program (awarded to BD).

## **2.7 Author contributions**

BD, AW, and JP conceived the study. BD performed the computational modeling and data analysis. BD wrote the initial draft of the manuscript. BD, KR, GK, KV, AW, and JP edited and wrote the final manuscript.

## **2.8 Conflicts of Interest**

The authors declare no competing interests.

## 2.9 Data and Code Availability

The code generated in this study is available at <https://github.com/csbl/iCardio>.

## 2.10 Supplemental Tables, Figures, and Data

<u>Description</u>	<u>Change</u>	<u>Reactions affected</u>	<u>Citation</u>
Directionality of DHAP reaction	reverse reaction direction	RCR21050	
Remove superoxide transport from the model	Change reaction bounds to [0,0]	RCR40428	
Include catalase reactions in the model	Change reaction bounds to [0,1000]	RCR10165, RCR10607, RCR11007, RCR11029, RCR14124	
Include superoxide production in Complex I	Change reaction to consume 0.04 units of O <sub>2</sub> and produce 0.04 units of O <sub>2</sub> <sup>-</sup>	RCR21048	Smith et al, 2011
Change the number of protons necessary to generate one unit of ATP	Change reaction to 2.7 protons moved from the cytosol to the mitochondrial to produce 1 unit of ATP	RCR20085	Watt et al, 2010

**Supplemental Table 2.1 Changes to the *iHsa* network reconstruction resulting from metabolic task curation.**

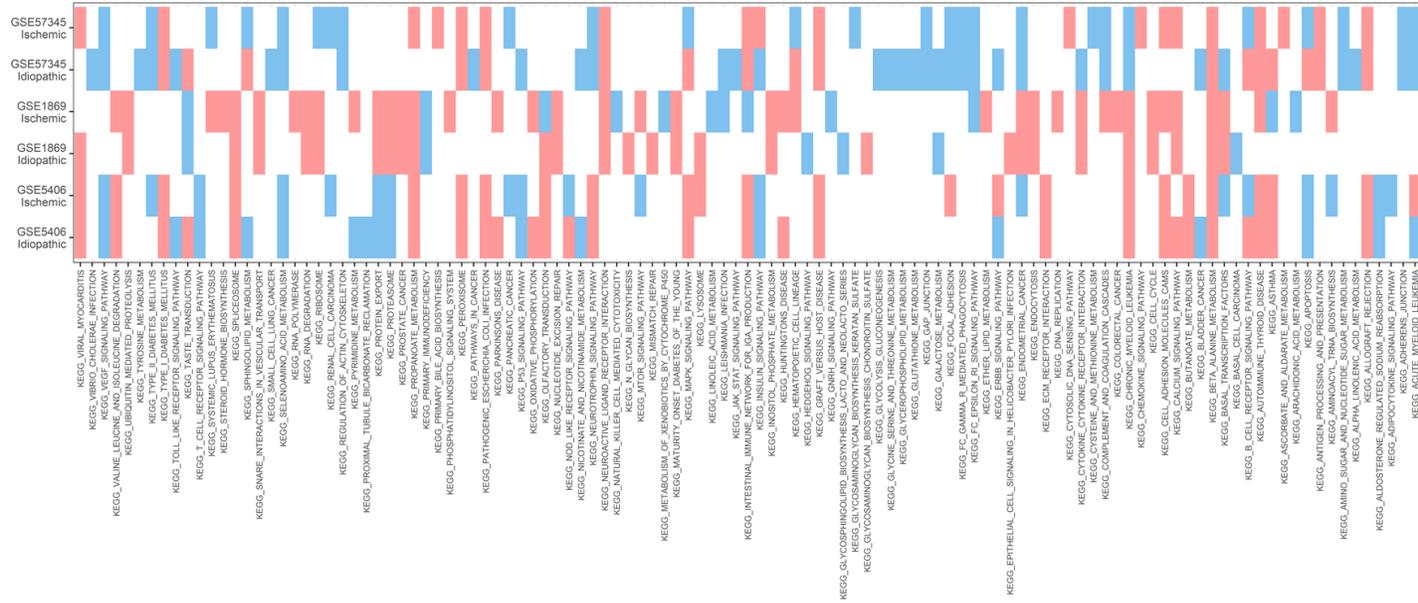
		Reactions	Metabolites	Heart-specific task accuracy (%)
<b>Original model</b>	iHsa	8336	5627	-
<b>Integration algorithms</b>	<b>CORDA</b>	<b>4133</b>	<b>2846</b>	<b>89%</b>
	iMAT	4434	3074	82%
	FastCore (high, medium, low)	4624	3215	55%
	GIMME	5526	5013	37%
	MBA (high, medium, low)	3709	2630	37%
	FastCore (only high)	2671	2187	36%
	MBA (only high)	2510	2119	34%

**Supplemental Table 2.2 Evaluating integration algorithms for construction of draft iCardio models.**

*Five different algorithms were used to integrate data from the Human Protein Atlas (HPA) with iHsa to produce draft iCardio models. Unless otherwise indicated, high and medium proteins, as indicated in the HPA, were used as an input to each algorithm. Heart-specific accuracy was evaluated based on completion of a list of metabolic tasks that covered both general and heart-relevant metabolism (Supplemental File 1).*

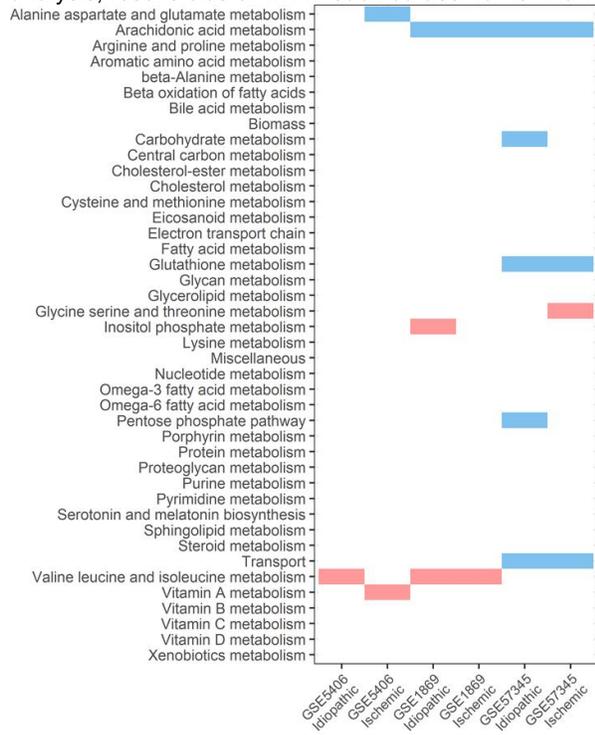
dataset	type of HF	# diseased	# controls	# DEGs	# model DEGs	total genes measured
GSE1869	Ischemic	3	6	2162	278	8426
GSE1869	Idiopathic	13	6	6974	896	8426
GSE57345	Ischemic	95	136	5938	808	8839
GSE57345	Idiopathic	82	136	5879	778	8839
GSE5406	Ischemic	86	16	2678	392	8242
GSE5406	Idiopathic	108	16	2571	398	8242

**Supplemental Table 2.3 Summary of DEGs for individual microarray studies.**



### Supplemental Figure 2.1 GSEA for all KEGG pathways.

Results from the GSEA for all KEGG pathways, including pathways that were not metabolic. Pathways with that were not statistically different across any of the studies are not shown. As with the TIDEs and KEGG metabolic pathway analysis, results cluster within each dataset rather than by type of heart failure.



### Supplemental Figure 2.2 Subsystem-level analysis using iCardio with gene expression data.

Using the reaction category assignments already in iCardio, reactions associated with each subsystem were determined and processed with the TIDEs pipeline. Subsystems that were significantly increased or decreased based on the gene expression data are shown here.

Supplemental File 2.1. Metabolic tasks used to evaluate draft *iCardio* models. Includes previously developed *iHsa* tasks and *iCardio* tasks indicating if the task should pass or fail.

Supplemental File 2.2. Reactions added and removed from the model during manual curation.

Supplemental File 2.3. Metabolic tasks used for TIDE analysis with results for each GSE study.

## 2.11 References

1. Kundu BK, Zhong M, Sen S, Davogustto G, Keller SR, Taegtmeier H. Remodeling of Glucose Metabolism Precedes Pressure Overload-Induced Left Ventricular Hypertrophy: Review of a Hypothesis. *Cardiology*. 2015;130: 211–220. doi:10.1159/000369782
2. Bauckneht M, Ferrarazzo G, Fiz F, Morbelli S, Sarocchi M, Pastorino F, et al. Doxorubicin Effect on Myocardial Metabolism as a Prerequisite for Subsequent Development of Cardiac Toxicity: A Translational 18F-FDG PET/CT Observation. *J Nucl Med*. 2017;58: 1638–1645. doi:10.2967/jnumed.117.191122
3. Borde C, Kand P, Basu S. Enhanced myocardial fluorodeoxyglucose uptake following Adriamycin-based therapy: Evidence of early chemotherapeutic cardiotoxicity? *World J Radiol*. 2012;4: 220–223. doi:10.4329/wjr.v4.i5.220
4. Li R, He H, Fang S, Hua Y, Yang X, Yuan Y, et al. Time series characteristics of serum branched-chain amino acids for early diagnosis of chronic heart failure. *J Proteome Res*. 2019 [cited 26 Mar 2019]. doi:10.1021/acs.jproteome.9b00002
5. Blais EM, Rawls KD, Dougherty BV, Li ZI, Kolling GL, Ye P, et al. Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nat Commun*. 2017;8: 14250. doi:10.1038/ncomms14250
6. Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol*. 2018;36: 272–281. doi:10.1038/nbt.4072
7. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci*. 2007;104: 1777–1782. doi:10.1073/pnas.0610772104
8. Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, et al. The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol*. 2007;3: 135. doi:10.1038/msb4100177
9. Mardinoglu A, Gatto F, Nielsen J. Genome-scale modeling of human metabolism – a systems biology approach. *Biotechnol J*. 8: 985–996. doi:10.1002/biot.201200275
10. Swainston N, Smallbone K, Hefzi H, Dobson PD, Brewer J, Hanscho M, et al. Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics*. 2016;12: 109. doi:10.1007/s11306-016-1051-4
11. Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, et al. A community-driven global reconstruction of human metabolism. *Nat Biotechnol*. 2013;31: 419–425. doi:10.1038/nbt.2488

12. Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol Syst Biol.* 2014;10. doi:10.1002/msb.145122
13. Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. Predicting selective drug targets in cancer through metabolic networks. *Mol Syst Biol.* 2011;7: 501–501. doi:10.1038/msb.2011.35
14. Rawls K, Dougherty BV, Papin J. Metabolic Network Reconstructions to Predict Drug Targets and Off-Target Effects. *Methods Mol Biol Clifton NJ.* 2020;2088: 315–330. doi:10.1007/978-1-0716-0159-4\_14
15. Zhang A-D, Dai S-X, Huang J-F. Reconstruction and Analysis of Human Kidney-Specific Metabolic Network Based on Omics Data. In: *BioMed Research International [Internet].* 2013 [cited 7 Mar 2019]. doi:10.1155/2013/187509
16. Zhao Y, Huang J. Reconstruction and analysis of human heart-specific metabolic network based on transcriptome and proteome data. *Biochem Biophys Res Commun.* 2011;415: 450–454. doi:10.1016/j.bbrc.2011.10.090
17. Shaked I, Oberhardt MA, Atias N, Sharan R, Ruppin E. Metabolic Network Prediction of Drug Side Effects. *Cell Syst.* 2016;2: 209–213. doi:10.1016/j.cels.2016.03.001
18. Zielinski DC, Filipp FV, Bordbar A, Jensen K, Smith JW, Herrgard MJ, et al. Pharmacogenomic and clinical data link non-pharmacokinetic metabolic dysregulation to drug side effect pathogenesis. *Nat Commun.* 2015;6: 7101. doi:10.1038/ncomms8101
19. Blazier AS, Papin JA. Integration of expression data in genome-scale metabolic network reconstructions. *Front Physiol.* 2012;3: 299. doi:10.3389/fphys.2012.00299
20. Robaina Estévez S, Nikoloski Z. Generalized framework for context-specific metabolic model extraction methods. *Front Plant Sci.* 2014;5. doi:10.3389/fpls.2014.00491
21. Karlstädt A, Fliegner D, Kararigas G, Ruderisch HS, Regitz-Zagrosek V, Holzhütter H-G. CardioNet: A human metabolic network suited for the study of cardiomyocyte metabolism. *BMC Syst Biol.* 2012;6: 114. doi:10.1186/1752-0509-6-114
22. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. *Science.* 2015;347: 1260419. doi:10.1126/science.1260419
23. Richelle A, Chiang AWT, Kuo C-C, Lewis NE. Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. *PLoS Comput Biol.* 2019;15: e1006867. doi:10.1371/journal.pcbi.1006867

24. Smith AC, Eyassu F, Mazat J-P, Robinson AJ. MitoCore: a curated constraint-based model for simulating human central metabolism. *BMC Syst Biol.* 2017;11. doi:10.1186/s12918-017-0500-7
25. Wende AR, Brahma MK, McGinnis GR, Young ME. Metabolic Origins of Heart Failure. *JACC Basic Transl Sci.* 2017;2: 297–310. doi:10.1016/j.jacbts.2016.11.009
26. Lopaschuk GD. Metabolic Modulators in Heart Disease: Past, Present, and Future. *Can J Cardiol.* 2017;33: 838–849. doi:10.1016/j.cjca.2016.12.013
27. Janardhan A, Chen J, Crawford PA. Altered Systemic Ketone Body Metabolism in Advanced Heart Failure. *Tex Heart Inst J.* 2011;38: 533–538.
28. Sun H, Olson KC, Gao C, Prosdocimo DA, Zhou M, Wang Z, et al. Catabolic Defect of Branched-Chain Amino Acids Promotes Heart Failure. *Circulation.* 2016;133: 2038–2049. doi:10.1161/CIRCULATIONAHA.115.020226
29. Taylor M, Wallhaus TR, Degrado TR, Russell DC, Stanko P, Nickles RJ, et al. An evaluation of myocardial fatty acid and glucose uptake using PET with [18F]fluoro-6-thiaheptadecanoic acid and [18F]FDG in Patients with Congestive Heart Failure. *J Nucl Med Off Publ Soc Nucl Med.* 2001;42: 55–62.
30. Doenst T, Nguyen TD, Abel ED. Cardiac Metabolism in Heart Failure - Implications beyond ATP production. *Circ Res.* 2013;113: 709–724. doi:10.1161/CIRCRESAHA.113.300376
31. Ingwall JS. Transgenesis and cardiac energetics: new insights into cardiac metabolism. *J Mol Cell Cardiol.* 2004;37: 613–623. doi:10.1016/j.yjmcc.2004.05.020
32. Neubauer S. The Failing Heart — An Engine Out of Fuel. *N Engl J Med.* 2007;356: 1140–1151. doi:10.1056/NEJMra063052
33. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102: 15545–15550. doi:10.1073/pnas.0506580102
34. Li M, Parker BL, Pearson E, Hunter B, Cao J, Koay YC, et al. Core functional nodes and sex-specific pathways in human ischaemic and dilated cardiomyopathy. *Nat Commun.* 2020;11: 2843. doi:10.1038/s41467-020-16584-z
35. Massion PB, Feron O, Dessy C, Balligand J-L. Nitric oxide and cardiac function: ten years after, and continuing. *Circ Res.* 2003;93: 388–398. doi:10.1161/01.RES.0000088351.58510.21
36. Hayashi H, Hess DT, Zhang R, Sugi K, Gao H, Tan BL, et al. S-Nitrosylation of  $\beta$ -Arrestins Biases Receptor Signaling and Confers Ligand Independence. *Mol Cell.* 2018;70: 473-487.e6. doi:10.1016/j.molcel.2018.03.034

37. Zhang C, Chen J, Liu Y, Xu D. Sialic acid metabolism as a potential therapeutic target of atherosclerosis. *Lipids Health Dis.* 2019;18: 173. doi:10.1186/s12944-019-1113-5
38. Smith AC, Robinson AJ. A metabolic model of the mitochondrion and its use in modelling diseases of the tricarboxylic acid cycle. *BMC Syst Biol.* 2011;5: 102. doi:10.1186/1752-0509-5-102
39. Watt IN, Montgomery MG, Runswick MJ, Leslie AGW, Walker JE. Bioenergetic cost of making an adenosine triphosphate molecule in animal mitochondria. *Proc Natl Acad Sci.* 2010;107: 16823–16827. doi:10.1073/pnas.1011099107
40. Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol.* 2008;4: e1000082. doi:10.1371/journal.pcbi.1000082
41. Jerby L, Shlomi T, Ruppin E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol Syst Biol.* 2010;6: 401. doi:10.1038/msb.2010.56
42. Schultz A, Qutub AA. Reconstruction of Tissue-Specific Metabolic Networks Using CORDA. *PLOS Comput Biol.* 2016;12: e1004808. doi:10.1371/journal.pcbi.1004808
43. Vlassis N, Pacheco MP, Sauter T. Fast Reconstruction of Compact Context-Specific Metabolic Network Models. *PLOS Comput Biol.* 2014;10: e1003424. doi:10.1371/journal.pcbi.1003424
44. Zur H, Ruppin E, Shlomi T. iMAT: an integrative metabolic analysis tool. *Bioinformatics.* 2010;26: 3140–3142. doi:10.1093/bioinformatics/btq602
45. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, et al. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol.* 2010;6: 390. doi:10.1038/msb.2010.47
46. Machado D, Herrgård M. Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism. *PLOS Comput Biol.* 2014;10: e1003580. doi:10.1371/journal.pcbi.1003580
47. Hannenhalli S, Putt ME, Gilmore JM, Wang J, Parmacek MS, Epstein JA, et al. Transcriptional genomics associates FOX transcription factors with human heart failure. *Circulation.* 2006;114: 1269–1276. doi:10.1161/CIRCULATIONAHA.106.632430
48. Kittleson MM, Minhas KM, Irizarry RA, Ye SQ, Edness G, Breton E, et al. Gene expression analysis of ischemic and nonischemic cardiomyopathy: shared and distinct genes in the development of heart failure. *Physiol Genomics.* 2005;21: 299–307. doi:10.1152/physiolgenomics.00255.2004
49. Liu H, Lv L, Yang K. Chemotherapy targeting cancer stem cells. *Am J Cancer Res.* 2015;5: 880–893.

50. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43: e47–e47. doi:10.1093/nar/gkv007
51. Jerby L, Ruppin E. Predicting Drug Targets and Biomarkers of Cancer via Genome-Scale Metabolic Modeling. *Clin Cancer Res.* 2012;18: 5572–5584. doi:10.1158/1078-0432.CCR-12-1856

# Chapter 3: Identifying biomarkers of chemotherapy-induced cardiotoxicity using paired transcriptomics and metabolomics data integrated with a model of heart metabolism

Bonnie V. Dougherty<sup>1</sup>, Kristopher D. Rawls<sup>1</sup>, Matthew L. Jenior<sup>1</sup>, Bryan Chun<sup>1</sup>, Sarbajeet Nagdas<sup>2</sup>, Jeffery J. Saucerman<sup>1</sup>, Glynis L. Kolling<sup>1,3</sup>, Anders Wallqvist<sup>4,5</sup>, Jason A. Papin<sup>1,3,6,7</sup>

<sup>1</sup>Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, 22908, USA.

<sup>2</sup>Department of Microbiology, Immunology, and Cancer Biology, University of Virginia Health System, Charlottesville, VA 22908, USA.

<sup>3</sup>Department of Medicine, Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, VA, 22908, USA.

<sup>4</sup>Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Development Command, Fort Detrick, Maryland, 21702 USA.

<sup>5</sup>The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, Maryland, 20817, USA.

<sup>6</sup>Department of Biochemistry & Molecular Genetics, University of Virginia, Charlottesville, VA, 22908, USA.

<sup>7</sup>Lead contact

\*Corresponding author. Tel: +1 434 924 8195; Email: papin@virginia.edu

### 3.1 Abstract

Improvements in the diagnosis and treatment of cancer has revealed the long-term side effects of chemotherapeutics, particularly cardiotoxicity. Current clinical measures to track cardiotoxicity are insufficient to diagnose damage before it has been done, necessitating new, early biomarkers of cardiotoxicity. Here, we present paired transcriptomics and metabolomics data characterizing in vitro cardiotoxicity to three compounds: 5-fluorouracil, acetaminophen, and doxorubicin. Standard gene enrichment and metabolomics approaches identify some commonly affected pathways and metabolites but are not able to readily identify mechanisms of cardiotoxicity. Here, we integrate this paired data with a genome-scale metabolic network reconstruction (GENRE) of the heart to identify shifted metabolic functions, unique metabolic reactions, and changes in flux in metabolic reactions in response to these compounds. Using this approach, we are able to confirm known mechanisms of doxorubicin-induced cardiotoxicity and provide hypotheses for mechanisms of cardiotoxicity for 5-fluorouracil and acetaminophen.

## 3.2 Introduction

As the treatment of cancer improves, multiple chemotherapeutics have been identified as increasing the incidence of cardiovascular events in cancer patients [1]. This observation has prompted the exploration of potential mechanisms of chemotherapy-induced cardiotoxicity. It is now well-established that multiple chemotherapeutics are associated with adverse cardiovascular events, such as left ventricular dysfunction and chronic heart failure [1]. Current approaches to limit the development of chemotherapy-induced cardiotoxicity include limiting the dose of the chemotherapeutic, potentially decreasing the efficacy of treatment [2,3], and using clinical measures, such as left ventricular ejection fraction (LVEF), to monitor heart function [4,5]. However, LVEF only demonstrates a change after irreversible cardiac damage, highlighting the need for early and more sensitive markers of cardiotoxicity. Along these lines, changes in glucose uptake have recently been noted to precede clinical measures of heart dysfunction in both spontaneously hypertensive rats [6] and in doxorubicin models of cardiotoxicity [7,8], suggesting metabolites and changes in metabolism in the heart are an opportunity for early biomarkers of drug-induced cardiotoxicity.

Typical studies to predict new metabolic biomarkers rely on untargeted metabolomics. However, these methods lack mapping to a potential mechanism for the change in a metabolite and may miss key biomarkers since the panel of screened metabolites must be defined *a priori*. Genome-scale metabolic network reconstructions (GENREs) provide an opportunity to mechanistically connect changes in metabolomics with changes in transcriptomics, identifying potential mechanisms for biomarker production. GENREs provide a mechanistic representation of cellular metabolism, including the stoichiometric coefficients for metabolic reactions and the connectivity between genes and the individual reactions they govern. Previous studies have used transcriptomics data with metabolic network reconstructions to predict biomarkers of hepatotoxicity [9–12] and nephrotoxicity [13,14].

In the current study, we extend that work to integrate paired transcriptomics and metabolomics data with a heart-specific genome-scale metabolic network (GENRE) [15] to predict biomarkers of cardiotoxicity. We collect paired transcriptomics and metabolomics data for primary rat neonatal cardiomyocytes exposed to three compounds: 5-fluorouracil (5FU), acetaminophen (Ace), and doxorubicin (Dox). Both Dox and 5FU were selected based on their established cardiotoxicity [4,16] while Ace was chosen based on previous studies exploring hepatotoxicity and nephrotoxicity [9,13]. Furthermore, Dox has multiple hypothesized mechanisms of toxicity [4] whereas 5FU does not have established hypotheses for mechanisms of cardiotoxicity [16]. We demonstrate the utility of integrating multiple forms of omics data with functional models of human metabolism to yield unique insight into potential biomarkers of cardiotoxicity and their associated mechanisms. Through this integrated approach, we confirm known mechanisms of Dox cardiotoxicity and propose new potential mechanisms for 5FU toxicity through increased metabolic stress from its primary chemotherapeutic mechanism of action and Ace cardiotoxicity through changes in phospholipid metabolism and the pentose phosphate pathway.

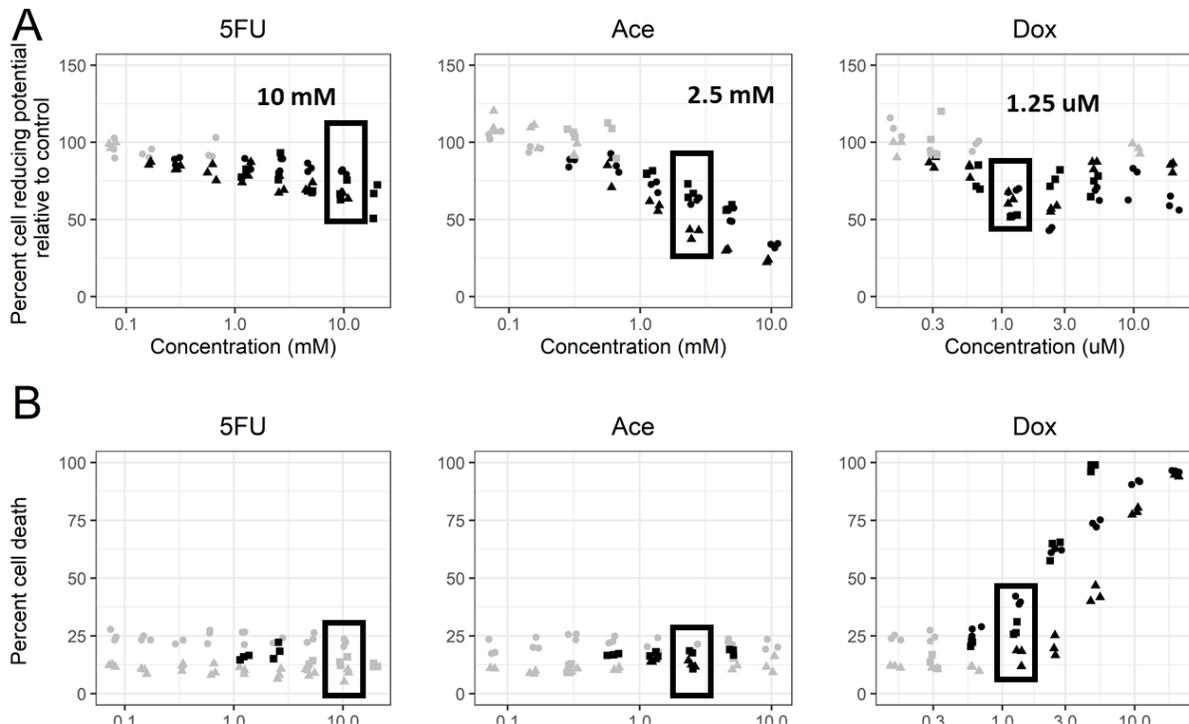
### **3.3 Results**

#### *3.3.1 Optimizing concentrations of compounds to characterize in vitro cardiotoxicity*

Given that most toxicity studies have limited rationale for chosen concentrations, we aimed to deliberately choose cardiotoxic concentrations that both elicited a measured change in cell metabolism and enable comparison across compounds. We determined cardiotoxic concentrations for our *in vitro* studies as concentrations that elicited a significant decrease in cell reducing potential without a concordant significant increase in cell death compared to controls at 24 hours ( ). 10 mM 5FU and 2.5 mM Ace elicited a significant decrease in cell reducing potential at 24 hours without a significant increase in cell death. 1.25  $\mu$ M Dox induced a significant decrease in cell reducing potential but also a significant increase in cell death (Figure 3.1). To confirm that

we had elicited a change in metabolism and not simply an increase in cell death, we performed a Seahorse mitochondrial stress test for which the measured OCR was normalized to live cell counts (Supplemental Figure 3.1). There was a significant increase in OCR for ATP production for 10 mM 5FU and 1.25  $\mu$ M Dox and a significant increase in basal respiration for Dox. This confirms a metabolic stress on a per cell level, seen as an increase in the flux of oxygen being used for the electron transport chain (ATP production) and increase in the basal rate of oxygen consumption (basal respiration).

As with previous studies [9,13], we profiled cell responses at both 6 (Supplemental Figure 3.2) and 24 hours (Figure 3.1) after drug exposure to capture early and late toxicity. There was no significant increase in cell death for any compounds for our chosen concentrations at 6 hours. For cell reducing potential, we see a consistent decrease at our chosen concentration for Ace,



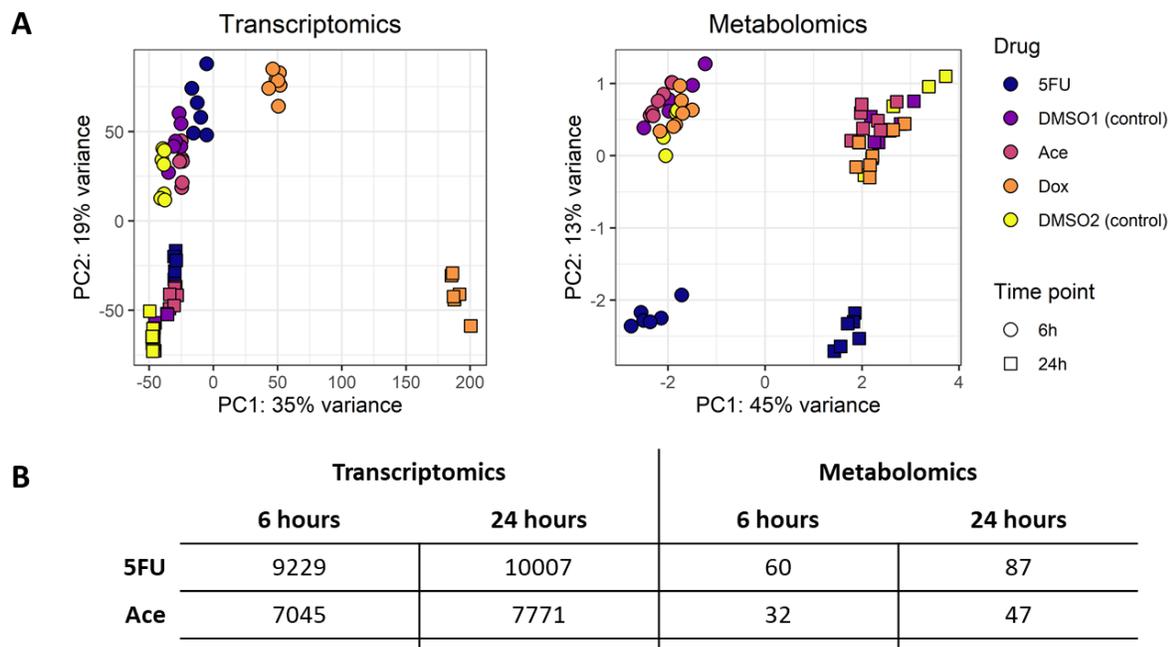
**Figure 3.1 Choosing cardiotoxic drug concentrations that elicited a change in metabolism without a significant increase in cell death.**

(A) Percent cell reducing potential with respect to controls was measured for a range of concentrations for the chosen compounds. The shapes represent different biological replicates. Black dots indicate a statistically significant change from the control condition, determined using Dunnett's test with a  $p$ -value  $< 0.05$ . Boxes indicate the concentrations chosen for the experimental studies. (B) Percent cell death calculated using a PI/Hoescht stain. The shapes represent different biological replicates. The black dots indicate a statistically significant change from the control condition, calculated using Dunnett's test with  $p$ -value  $< 0.05$ .

indicating that Ace has an early but sustained, negative effect on cell metabolism in cardiomyocytes (Supplemental Figure 3.2).

### 3.3.2 Unsupervised machine learning of transcriptomics and metabolomics data highlights underlying drug-induced shifts in cellular activity

In order to identify the largest sources of variability between our conditions, individual samples were clustered in an unsupervised fashion using Principal Component Analysis (PCA) (Figure 3.2). For reference, the 5FU condition is paired with the DMSO1 control and the Ace and Dox conditions are paired with the DMSO2 control (Methods). Analysis of normalized transcript counts for each gene shows separation in the first principal component by the Dox treatment and in the second principal component by time. For the other compounds, within each time point, the control conditions (DMSO1 and DMSO2) separate from the treatment groups (5FU and Ace)



**Figure 3.2 PCA of transcript counts (A) and scaled metabolite abundances (B) for the three compounds at 6 and 24 hours.**

(A) PCA of transcript counts separate by time and treatment condition, specifically the Dox treatment. (B) PCA of scaled metabolite abundances separate by time and treatment condition, specifically the 5FU treatment. (C) Quantification of the number of differentially expressed genes and differentially changed metabolites for each condition. DEGs were genes with an FDR < 0.01 and differential metabolites were metabolites with an FDR < 0.1.

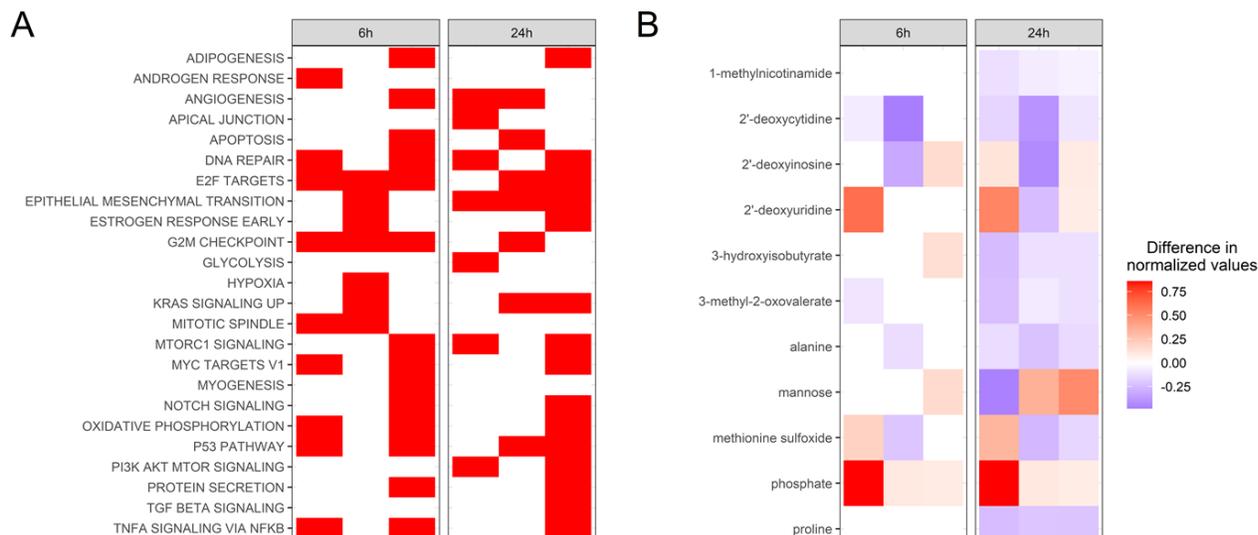
(perMANOVA for treatment vs control,  $p$ -value  $< 0.001$ ). This separation is clearer at the 24-hour time point (perMANOVA for treatment vs control,  $p$ -value  $< 0.001$ ) (Supplemental Figure 3.3A). The second principal component separates between the 6-hour and 24-hour samples, indicating a large change in the underlying cellular transcriptomics and metabolomic response. This separation may be a byproduct of the cells adapting to *in vitro* culture conditions. Given that two of the proposed chemotherapeutic mechanisms of action of Dox are intercalation with DNA and the inhibition of topoisomerase II [17], we would expect clear separation in the transcriptome for Dox-treated cells compared to the other compounds. A gene enrichment analysis for Hallmark pathways from the Molecular Signatures Database [18,19] using the top 100 genes in the first principal component identifies the p53 pathway as the only significantly enriched pathway, suggesting that Dox induces a unique DNA repair response compared to the other compounds. In contrast, a gene enrichment analysis for the top 100 genes in the second principal component identifies Myc targets as uniquely enriched, suggesting a transition between fatty acid oxidation and glucose utilization [20].

The PCA of log-scaled and median-centered metabolite abundances separates by time and, in the second principal component, by the 5FU treatment (Figure 3.2B). In agreement with the transcriptomics data, the clear separation by time point suggests a potential adaption of the primary cells to *in vitro* culture conditions. The bi-plot (Supplemental Figure 3.3B) of the top 10 metabolites responsible for separation identifies erythritol, a derivative of glucose metabolism [21] and ethylmalonate, a branched chain fatty acid, for separation in the first principal component, again suggesting a change in glucose and fatty acid metabolism over time. However, in contrast to the transcriptomics results, there is no clear separation among any conditions, except for 5FU treatment. This lack of separation could result from the number and type of metabolites that were profiled or could suggest a more nuanced change between different conditions. Given that 5FU acts as an analogue for uracil and interferes with RNA synthesis [22], we would expect clear

separation in the metabolomics data for cells treated with 5FU. The bi-plot (Supplemental Figure 3.3B) of the top 10 metabolites responsible for separation identifies uracil and 2-deoxyuridine, indicating changes in uracil synthesis, and phosphate, indicating a metabolic effect, as separating 5FU from the other conditions. Finally, we can quantify the number of differentially expressed genes (DEGs) and changing metabolites for each treatment and time point (Figure 3.2C). As we would expect from the PCA, the Dox condition has the largest number of DEGs at both 6 and 24-hours. The 5FU condition has the largest number of differentially changed metabolites at the 6 and 24-hour conditions.

### 3.3.3 Gene enrichment and metabolomics data identify common signatures of toxicity but cannot readily identify mechanisms of cardiotoxicity

The large number of DEGs for each condition necessitates an enrichment approach to identify changed pathways across conditions and time points. We performed enrichment analysis using the Hallmark gene sets defined in the Molecular Signatures Database (MSigDB) [18,19] (Figure 3.3A). Given the large number of DEGs, it was necessary to use both a lower FDR cutoff



**Figure 3.3 Identifying biomarkers of toxicity from the transcriptomics and metabolomics data sets.** (A) Enrichment analysis for genes with an FDR < 0.01 using the Hallmark gene sets from the Molecular Signatures database. A red box indicates enrichment with a p-value < 0.1. (B) Metabolites that were identified to be significantly changed in production between the treatment and control (Mann-Whitney, FDR < 0.1). The color represents the mean difference in normalized metabolites between the treatment and control, where a negative value indicates decreased production and a positive value indicates increased production.

and a higher p-value cutoff for enriched gene sets; genes with an FDR < 0.01 and gene sets with an adjusted p-value < 0.1 were identified as significantly changed. Consistent with their known mechanisms of chemotherapeutic efficacy, we see that the 5FU and Dox conditions are enriched for genes related to DNA repair at both the 6 and 24-hour timepoint. As with the PCA data, we see enrichment for genes for the p53 pathway for both Dox and 5FU. Finally, we see enrichment for genes related to oxidative phosphorylation, particularly at 6 hours, for both 5FU and Dox, confirming changes in cellular metabolism.

Next, we identified metabolites that were significantly changed across all conditions within a timepoint (Figure 3.3B). We see increased production of phosphate for all three compounds, confirming a significant change in metabolism for the chosen concentrations. Further, we see a consistent increase in production of 2'-deoxyinosine for 5FU and Dox at the 24-hour timepoint, consistent with a response to reactive oxygen species (ROS) stress [23] as well as increased production of 2'-deoxyuridine in the 5FU condition, consistent as a by-product of DNA damage and uracil metabolism [24]. Further, we see differences in metabolites that are consumed between conditions (Supplemental Figure 3.4A) or metabolites that are ambiguously produced or consumed differentially between conditions (Supplemental Figure 3.4B). Any of these metabolites can serve as potential biomarkers of *in vitro* cardiotoxicity; however, a specific mechanism for the change in production or consumption is difficult to assess. Further, there are no available methods for connecting the list of DEGs and metabolites to confirm or identify mechanisms of chemotherapeutic toxicity or mechanisms of cardiotoxicity. Metabolic network reconstructions provide an opportunity to connect the measured changes in the transcriptome with the measured changes in the metabolome to identify potential mechanisms of toxicity.

### 3.3.4 *Reconstruction of a rat-specific heart GENRE from an existing human-specific heart GENRE*

Before integrating our collected data, we needed to construct a rat-specific heart metabolic model. We started from the previously published human-specific heart model [15], which was built from a general human metabolic network developed in parallel with a general rat metabolic network [25]. These paired models are able to capture species-specific functions and species-specific gene-protein-reaction (GPR) rules [25]. Given that the two original general models were generated simultaneously, we can directly map reactions included in the human-specific heart model to a rat-specific heart model using the common reaction IDs. After including these reactions, we added 13 rat-specific reactions from the general rat model that had literature evidence for expression in the heart (Supplemental File 3.1). Next, we utilized the metabolomics data to identify metabolites that were both measured in our dataset and included in the metabolic model. 121 metabolites mapped between the metabolomics dataset and metabolites in the model. 75 of these metabolites had associated exchange reactions in the general rat model, indicating that the metabolite could be either consumed or produced in the model. From these exchange reactions, 37 were added to the rat-specific heart model from the general model to ensure that constraints could be placed for either production or consumption of that metabolite. Manual curation is often a time-intensive process and here we demonstrate the value of paired transcriptomics and metabolomics data in identifying new potential reactions that should be added to tissue-specific models of metabolism.

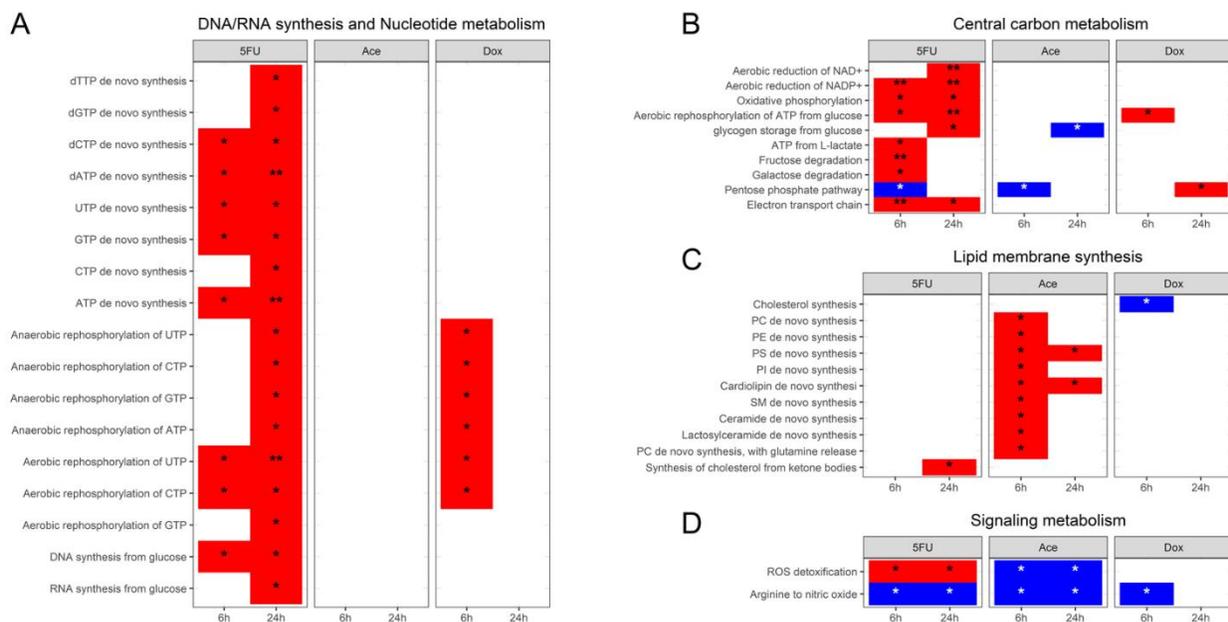
Before integrating the transcriptomics and metabolomics data with the model, we first identified the number of DEGs and differentially changed metabolites that the model captures (Supplemental Figure 3.5A). Here, we see that the model captures ~10% of the DEGs across all treatments and time points and between 40-65% of the differentially changed metabolites. Next, in order to confirm that the metabolic genes captured in the model are still capturing the underlying

variability in the data (Figure 3. 2A), we again used PCA to identify the largest sources of variability in the metabolic genes (Supplemental Figure 3.5B). As with the previous PCA (Figure 3.2), we see a similar distribution of data. Using a similar enrichment approach as above, both the glycolysis and hypoxia Hallmark pathways, among others, were enriched for the first principal component. For the second principal component, the Hallmark pathways of cholesterol homeostasis, glycolysis, oxidative phosphorylation, and fatty acid metabolism, among others, were enriched. Again, this result confirms a significant metabolic component in response to the drug treatment. However, as with the above analysis, general pathway level changes do not yield insight into potential mechanisms of cardiotoxicity.

### *3.3.5 Integrating a rat-specific, heart GENRE and transcriptomics data predicts novel metabolic tasks altered in in vitro cardiotoxicity*

Gene enrichment analysis can be helpful in identifying broad changes in a data set. However, as noted above, it can be difficult to identify concrete mechanisms that may be related to a drug's mechanism of action or mechanism of cardiotoxicity. Here, we apply the TIDEs approach [15] to identify metabolic functions that are associated with a significant change in gene expression using a rat-specific metabolic network reconstruction of the heart (Figure 3.4, Supplemental File 3.1). The TIDEs approach uses the complex GPR rules in a metabolic model to assign weights to individual reactions necessary to complete a metabolic task, such as production of ATP from glucose. Using this approach, we can identify metabolic tasks that are either associated with increased or decreased gene expression in response to each drug treatment. Statistical significance is calculated by randomizing gene expression values to create a distribution of task scores.

We implemented this approach by overlaying metabolic DEGs with an FDR < 0.01 onto the rat-specific heart model developed for this study (Methods, Supplemental File 3.1). Metabolic



**Figure 3.4 TIDEs analysis reveals distinct changes in metabolic function across compounds.**

A red box indicates significantly higher gene expression and a blue box indicates significantly lower gene expression. Statistical significance is indicated for a  $p$ -value  $< 0.1$  (\*) and  $p$ -value  $< 0.01$  (\*\*). (A) Metabolic tasks related to DNA and RNA synthesis or nucleotide metabolism. (B) Metabolic tasks related to central carbon metabolism. (C) Metabolic tasks related to lipid membrane synthesis. (D) Metabolic tasks related to signaling metabolism.

tasks with a  $p$ -value  $< 0.1$  were identified as differentially changed. Through this approach, nucleoside triphosphates metabolism (NTPs) is increased in both the 5FU and Dox condition, whereas dNTP metabolism, precursors to DNA, are uniquely increased in the 5FU condition (Figure 3.4A). This result suggests that although the chemotherapeutic mechanism of action of Dox interferes with DNA synthesis, the metabolic stress induced by the drug mainly effects the metabolism of precursors to RNA metabolism and the general metabolic state of the cell. In the case of 5FU, we see changes in both NTP and dNTP metabolism, suggesting that 5FU has an effect on both DNA and RNA synthesis, which is confirmed in the metabolic tasks for DNA and RNA synthesis. In contrast to the gene enrichment analysis, the TIDEs approach uniquely identified UTP metabolism as altered at both 6 and 24 hours, and a highly significant increase in anaerobic rephosphorylation of UTP at 24 hours ( $p$ -value  $< 0.01$ ) (Figure 3.4A), consistent with the known chemotherapeutic mechanism of action of 5FU.

Next, given that the Hallmark pathway of oxidative phosphorylation was enriched from the Hallmark gene set (Figure 3.3A), we examined changes in metabolic tasks related to central carbon metabolism (Figure 3.4B). Here, we see a significant number of tasks changed for the 5FU condition at both timepoints, indicating an overall strong metabolic shift in response to 5FU. Of note, both the 5FU and Ace conditions are associated with decreased gene expression for the pentose phosphate pathway as 6 hours. In contrast, Dox shows increased gene expression for the pentose phosphate pathway, confirming the role of ROS production in cardiotoxicity.

Given that Ace is not a known cardiotoxic compound, we next identified that metabolic tasks related to lipid membrane synthesis were uniquely increased in the Ace condition (Figure 3.4C). The TIDEs approach revealed early upregulation of metabolic tasks for a variety of phospholipids, specific to Ace compared to the other treatments. These metabolic tasks suggested a unique potential mechanism of action of Ace cardiotoxicity. Of particular interest, the metabolic task for cardiolipin synthesis, a key component of the mitochondrial metabolism, is significantly increased. One mechanism of Ace hepatotoxicity is associated significant lipid peroxidation in response to increased oxidative stress [26], suggesting a potential shared mechanism between hepatotoxicity and cardiotoxicity.

Finally, we identified metabolic tasks that were shared in at least four of the six conditions (Figure 3.4D). Here, we found a decrease in the synthesis of nitric oxide from arginine for five of the six conditions as well as differences in ROS detoxification between 5FU and Ace. Given the known role of Ace ROS production in Ace hepatotoxicity [26], it is interesting to see decreased gene expression for the metabolic task of ROS detoxification and the pentose phosphate pathway. Decreased synthesis of NO has been noted in heart failure [15,27,28], suggesting a shared metabolic marker of heart dysfunction. It is interesting to note that, overall, the Dox treatment condition has few significant metabolic tasks. For the metabolic task of arginine to nitric oxide and ROS detoxification, we can examine the underlying distribution of randomized task

scores used to determine statistical significance (Supplemental Figure 3.5). From these data, it is clear that Dox has a higher overall task score, indicating a higher overall average gene expression across reactions, but also has a wider spread for the underlying distribution of randomized task scores as a result of the large number of DEGs.

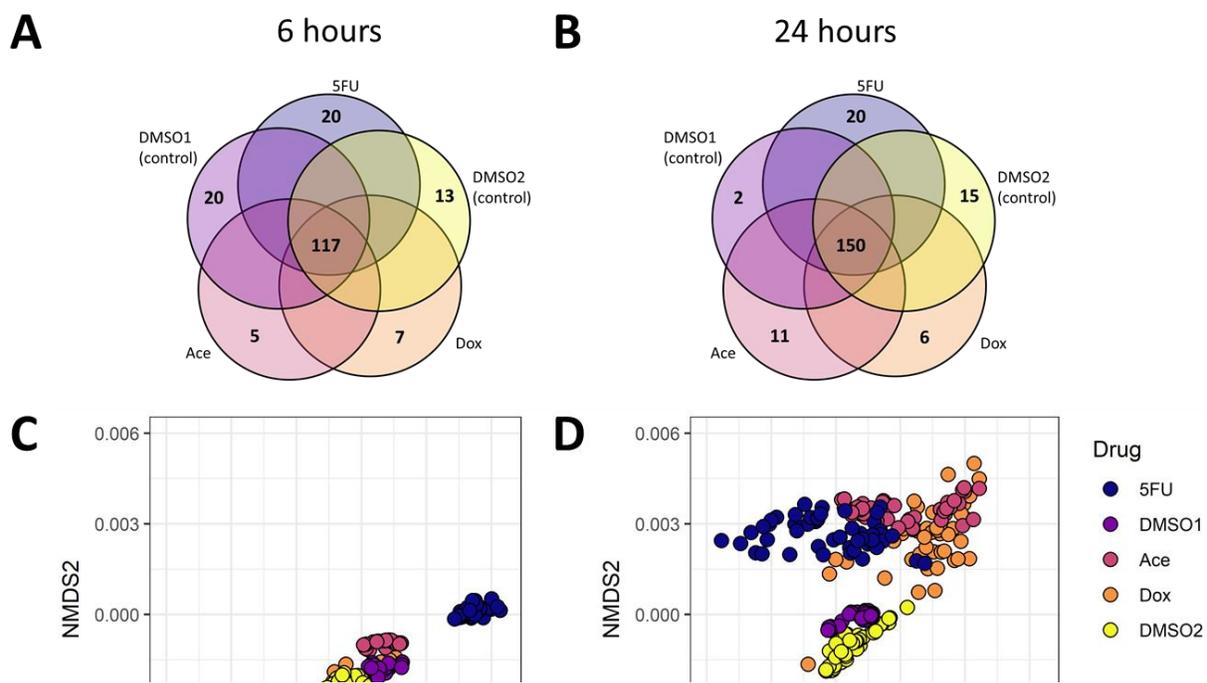
### *3.3.6 Combined omics datasets predict novel biomarkers of toxicity through integrated network-based analysis*

Pathway-level analysis provides one point of view for interpreting changes in gene expression. However, pathways do not work independently, but rather in a coordinated effort to maintain cell function. GENREs are able to capture this relationship by determining flux through reactions in a model while meeting an objective function, which represents a hypothesis for cell function. Here, we utilized our transcriptomics data integrated with the rat-specific heart model using the RIPTiDe algorithm [29] to determine the reactions and fluxes that met the constraints provided by our metabolomics data and objective function. Objective functions for non-proliferative cells are often hard to define so we used the objective function of ATP hydrolysis, representing the ATP generated for cardiomyocyte contraction, as well as requiring minimal synthesis of DNA and RNA (Methods). Using the condition-specific transcriptomics and metabolomics data, we generated condition-specific models for each condition, both treatment and control groups. The condition-specific constraints are contained in Supplemental Table 3.1 while details for the condition-specific models are in Supplemental Table 3.2.

Each condition-specific model, with the exception of the 5FU 6-hour condition, had a significant correlation ( $p$ -value < 0.001) between the transcript data and the flux samples (Supplemental Table 3.2). This result indicates that the context-specific patterns of metabolism predicted with RIPTiDe are significantly correlated with the experimentally measured omics data, further supporting the validity of our predictions for the underlying biology. First, we identified

reactions that are unique to each condition (Figure 3.5A-B). We see expansion of the number of reactions that are included in all of the 24-hour condition models compared to the 6-hour condition models, suggesting an overall expansion in core metabolism. Given the separation observed in the PCA of the transcriptomics and metabolomics data (Figure 3.2A), it is not surprising to see this expansion in core metabolism. This expansion of reactions included reactions involved in central carbon metabolism, the pentose phosphate pathway, and metabolite transport or exchange.

For the 5FU condition, at 6 hours we see unique reactions for pyrimidine synthesis that are included in all context-specific models at 24 hours. Across all the conditions, only the Dox condition has reactions that are unique at both 6 and 24 hours; these reactions utilize cysteine to oxidize glutathione, representing a potential mechanism of Dox cardiotoxicity. Most reactions that appear as unique in the 6-hour condition are included in all of the models in the 24-hour condition,



**Figure 3.5 Condition-specific models integrating the metabolomics and transcriptomics 24-hour data identify unique reaction biomarkers of cardiotoxicity.**

(A) VennDiagram of reactions that are shared and distinct between different condition-specific models. (B) NMDS of the 50 flux samples for each condition. For each flux sample, fluxes were only taken for the 112 reactions that are shared between all conditions.

with the exception of one reaction that appears in the 6-hour 5FU condition and all treated conditions (5FU, Ace, and Dox) in the 24-hour condition. This reaction is related to pyrimidine metabolism for DNA synthesis, highlighting a potential unique role of changes in pathways involved in DNA synthesis as a mechanism of shared toxicity. Finally, the unique reactions at the 24-hour condition (Figure 3.5B) indicate potential mechanisms for late toxicity. For 5FU, we see unique reactions related to beta oxidation of fatty acids, suggesting a unique energy source, and for Ace, unique reactions for carbohydrate metabolism.

A non-metric multidimensional scaling (NMDS) was used to visually display the 50 flux samples for each condition in an unsupervised fashion (Figure 3.5C-D). For this approach, we can only plot fluxes for the 112 reactions that are shared amongst all conditions. At 6-hours, the 5FU condition clearly separates from the other conditions. While at 24-hours, all treatment groups clearly separate from the control groups. Supervised machine learning with random forest feature selection was used to identify the reactions that most distinguish the active metabolism of treatment from the respective control conditions. This analysis highlighted that every treatment and time point contained at least one reaction involved in central carbon metabolism as highly distinguishing, suggesting unique divergent pathways for ATP production. For the 6-hour Dox condition, the reaction with the highest mean decrease in accuracy was involved in reducing thioredoxin, again confirming the role of ROS detoxification in the Dox condition. However, this reaction was not identified as separating in the 24-hour condition, suggesting an early, but no sustained increase in ROS detoxification. Surprisingly, for the 5FU condition, the highly ranked reactions were involved in central carbon metabolism, uracil exchange (24-hour time point), and purine metabolism, suggesting that although 5FU interferes with uracil metabolism, the same pathways are used as in the control condition. Finally, for Ace, reactions related to nucleotide metabolism and pyrimidine metabolism were highly ranked at 6 hours while reactions for the pentose phosphate pathway were highly ranked at 24 hours, suggesting a role for diverting flux

through the pentose phosphate pathway for either nucleotide metabolism or detoxification of intracellular ROS.

### 3.4 Discussion

Here, we present paired transcriptomics and metabolomics data characterizing *in vitro* cardiotoxicity for three compounds: 5-fluorouracil, acetaminophen, and doxorubicin. In contrast to previous studies, we present rationally identified concentrations that induce a change in metabolism without a significant change in cell death for all three compounds. The desired metabolic disruption at the chosen concentrations for 5FU and Dox was indicated with results from a Seahorse assay to measure changes in oxygen consumption along the mitochondrial respiratory chain. The data provides the first, to our knowledge, paired transcriptomic and metabolic characterization of *in vitro* cardiotoxicity. We confirm our treatment using enrichment analysis, demonstrating significant enrichment for DNA repair in both 5FU and Dox, both of whose chemotherapeutic mechanisms of action target DNA or RNA synthesis. Additionally, we demonstrate metabolic changes through enrichment in oxidative phosphorylation and increased release of phosphate measured in the metabolomics data across all conditions. However, the cardiotoxic mechanisms of these three compounds are not obvious from the transcriptomics and metabolomics data alone. Further, we cannot directly connect what changes in the transcriptome may be influencing changes seen in the metabolomics data.

The chemotherapeutic mechanisms of action for 5FU and Dox are well understood [4,16]. However, the mechanisms by which these two compounds induce cardiotoxicity are not as well studied. There are multiple, well-established, hypothesized mechanisms of doxorubicin-induced cardiotoxicity, including increased ROS production and increased eNOS production [30,31]. 5-fluorouracil has no established mechanisms for cardiotoxicity. Standard enrichment analyses and measured changes in metabolites were insufficient to hypothesize mechanisms responsible for

the measured changes in metabolism and metabolites. Alternatively, by integrating transcriptomics and metabolomics data with a metabolic model of the rat heart, we provide multiple novel hypotheses for mechanisms behind the measured changes.

Consistent with the enrichment analysis, integration of DEGs with the TIDEs approach specifically identified changes in metabolic tasks related to DNA/RNA synthesis but also specifically identifying the role of all nucleotides in both Dox and 5FU, particularly metabolism of UTP in 5FU. In the case of Ace, the TIDEs pipeline identified increased synthesis of multiple phospholipids as potential markers of cardiotoxicity. Lipid peroxidation of membrane phospholipids is one proposed mechanism of acetaminophen-induced hepatotoxicity [26], but has not been proposed in cardiomyocytes. Nitric oxide synthesis and ROS detoxification were identified as common metabolic functions altered across conditions. Nitric oxide synthesis is a proposed mechanism of doxorubicin-induced cardiotoxicity [32] but has not yet been identified for 5FU cardiotoxicity. Here, we demonstrate that decreased NO synthesis and changes in ROS detoxification can serve as early and shared markers of *in vitro* cardiotoxicity. Finally, it is important to note the difference in number of TIDEs identified between conditions, where the Dox 24-hour condition has the lowest number of TIDEs. Given that there are the largest number of DEGs in the Dox condition, most genes that were included in the TIDEs analysis for the Dox 24-hour condition were differentially expressed, meaning that a metabolic function had to have a large gene expression signature in order to be identified as different (Supplemental Figure 3.5).

Although enrichment analyses are helpful in determining broad metabolic changes, it is important to view these changes in the context of the entire metabolic network. By incorporating constraints from the metabolomics data with a pre-defined objective, RIPTiDe was able to identify both the reactions necessary and the fluxes through those reactions that satisfied the given metabolomics and transcriptomics constraints. We were able to identify a reaction unique to the Dox condition that is related to ROS detoxification, again identifying a suggested mechanism of

cardiotoxicity. It is important to note that increased ROS production was not included for the Dox condition but can be explored in the future for a better understanding of how metabolism may adapt to an increased in ROS production. In addition, we identified a reaction common to all treatment conditions at the 24-hour time point which was related to pyrimidine metabolism for DNA synthesis, suggesting a unique mechanism for diverting flux in cardiotoxicity.

Finally, random forest variable selection identified reactions that were shared among individual conditions but whose flux separated between conditions. These reactions represent changes in flux that are necessary either for ATP production, metabolite production, or DNA/RNA synthesis that differ between condition, indicating divergent pathways of flux. Across all three treatment groups, we see reactions for central carbon metabolism, indicating a baseline divergent flux for ATP production. Next, not surprisingly, the Dox condition again identified a reaction involved in ROS detoxification. For 5FU, we see most pathways related to central carbon metabolism, suggesting similar pathways for DNA and RNA synthesis as in the control condition. However, given the TIDEs results indicate increased gene expression for DNA/RNA synthesis, the increased flux through these pathways and their accompanying metabolic stress are potential mechanisms of cardiotoxicity. Finally, in the Ace treatment, we identified reactions related to the pentose phosphate pathway and cholesterol metabolism, and with the TIDEs analysis, suggest divergent pathways for nucleotide metabolism, cholesterol metabolism, and phospholipid metabolism as potential mechanisms of toxicity.

In summary, we present an approach that identifies shifted metabolic functions and specific metabolic reactions that together point toward potential mechanisms of toxicity. Together, the paired transcriptomics and metabolomics data integrated with the rat-specific model of metabolism provide insight that was not clear from either set of data on its own. For Dox, we identified shifts in metabolic tasks related to nucleotide metabolism, ROS detoxication, and NO synthesis and reactions related to ROS detoxification as potential markers of toxicity, consistent

with previously published hypotheses for mechanisms of toxicity [33,34]. For 5FU, we identified shifts in metabolic tasks related to nucleotide metabolism, ROS detoxification, and NO synthesis and reactions related to central carbon metabolism, suggesting a mechanism of toxicity related to increased metabolic stress from the chemotherapeutic mechanism of action of 5FU. Finally, for Ace, we identified shifts in metabolic tasks related to lipid synthesis and reactions related to the pentose phosphate pathway as markers of toxicity.

Future work is necessary to trace pathway fluxes to determine how fluxes through these individual reactions influence other parts of metabolism. In addition, the provided paired transcriptomics and metabolomics data provide a starting point for improvements to the present metabolic network reconstructions. A number of metabolites were measured as produced but could not be produced with the individual condition-specific models, either because of missing internal reactions or missing constraints. Future work can explore the use of additional objective functions that replicate proposed mechanisms of toxicity, such as increased ROS production or synthesis of key cellular proteins, which may provide further explanation for the measured changes in the metabolomics data. Finally, given the paired nature of the human model that has previously been developed [15], the presented approach and previous model can be used to integrate data from human-induced pluripotent stem cell cardiomyocytes exposed to doxorubicin [35] to confirm that the mechanisms identified here also translate to humans.

## **3.5 Methods**

### *3.5.1 In vitro culture conditions*

Primary neonatal rat cardiomyocytes were isolated and cultured according to previously published protocols [36]. After the initial plating, cells were maintained for ~36 hours in plating media containing low glucose DMEM and M199 supplemented with L-glutamine, Penicillin-Streptomycin, 10% horse serum and 5% FBS. Cells were serum starved overnight (~12 hours)

before running experiments in serum free, ITSS-supplemented plating media. Cells were observed to beat spontaneously within 24-48 hours after isolation, confirming metabolic activity and functionality.

For experiments to determine optimal drug concentrations, cells were seeded in 96 well plates at a density of 100k cell/well. The initial range of concentrations used for 5FU, Dox, and Ace (Tocris) were selected based off previous studies for Dox [35,37,38], 5-FU [39,40], and Ace [41]. Drug stocks were prepared according to manufacturer's instructions using sterile DMSO and were diluted in plating media before treatment. Concentrations that induced cardiotoxicity were determined using parallel measures of cell death and cell reducing potential (10 mM 5FU, 2.5 mM Ace, 1.25  $\mu$ M Dox). Cell death was determined as the number of propidium iodide (PI) positive cells divided by the total number of Hoescht positive cells. Fluorescence data from treated cells was background corrected using blank wells before using CellProfiler [42] to segment nuclei and measure fluorescence intensity for both PI and Hoescht. In the case of doxorubicin, which is fluorescent at overlapping wavelengths with PI, we used wells containing the respective concentrations of doxorubicin for background subtraction. Measures were aggregated from four fields of view for each drug concentration. Cell viability, which measures cell reducing potential and thus cell metabolism, was measured using the RealTime-Glo MT Cell Viability kit (Promega, Catalog #G9711). Both measures were repeated for three separate wells, representing three technical replicates for each condition, as well as on three separate days using different primary cell isolations, representing three biological replicates for each condition. Statistical significance was calculated using Dunnet's t-test [43] which accounts for the dose-dependent nature of the data. A p-value < 0.05 was considered statistically significant.

Oxygen Consumption Rate (OCR) for Mitochondrial Stress Test (MST) assay was measured using a Seahorse XF24 Extracellular Flux Analyzer [44,45]. Primary rat neonatal

cardiomyocytes were plated on Seahorse assay plates and cultured according to the protocol described above. MST media was an unbuffered, phenol-red free, serum-free media from above that was adjusted to a pH of 7.4 and filter-sterilized before use. For the MST assay, oligomycin, BAM15 (Cayman), and Rotenone and Antimycin A (Sigma) were injected to final concentrations of 1  $\mu$ M, 10  $\mu$ M, 1  $\mu$ M and 2  $\mu$ M respectively. The OCR for each measure was taken as the first measurement after injection with the inhibitor or the first measurement in the case of baseline. ATP production was defined as the OCR at baseline minus the OCR after the oligomycin injection. OCR for each well was normalized to well cell numbers which were collected using a PI/Hoescht stain prior to the assay.

### 3.5.2 *RNA isolation, sequencing, and analysis*

For the paired transcriptomics and metabolomics data, hearts were harvested in parallel from three litters of rats on the same day. After parallel digestion, cells were mixed from all isolations before plating in 12-well plates at a density of 1.2 million cells/well. For reference, two separate DMSO controls were run (DMSO1 at 1% DMSO for the 5FU condition; DMSO2 at 0.25% DMSO for the Ace and Dox conditions). Primary rat neonatal cardiomyocytes were exposed to the chemicals mentioned above at the chosen concentrations for either 6 or 24 hours. After exposure, as has been done in previous studies [9,13], the cells were lysed with Trizol to begin RNA extraction. Cell lysates were mixed with chloroform and spun in phase-lock gel tubes inside a cold room and the upper phase was then decanted into new tubes. Isopropanol and glycogen were added to the mixture, incubated overnight at -20C and spun again resulting in an RNA pellet, which was washed with 75% ethanol twice. DNA was removed using the TURBO DNA-free kit (Invitrogen, #AM1907) and then RNA was quantified using the QuBit RNA Broad Range detection kit (Invitrogen, #Q10210). RNA was sent to GeneWiz (<https://www.genewiz.com/en>) for PolyA selection, library construction, and sequencing. RNA was sequenced using 2x150bp paired-end

(PE) readings and fastq files were generated. Kallisto v 0.46.0 [46] was used to pseudo-align raw fastq files under default settings to the *Rattus norvegicus* Ensembl v96 transcriptome. Transcript abundances were then aggregated to the Entrez gene level in R v. 3.6.3 with the package tximport [47]. Genes with consistently low counts (< 10) across all samples were removed. Differentially expressed genes (DEGs) were determined using DESeq2 [48] with a significance threshold of  $FDR < 0.1$ .

Principal component analysis (PCA) was performed using the variance stabilized gene counts [49] with the prcomp function in R. Statistical significance for the separation between treatment and control groups was calculated using adonis2 function in the Vegan package in R [50].

Following identification of DEGs, two approaches were used to identify pathways significantly changed in the data: enrichment using Hallmark gene sets from the Molecular Signatures Database [18,19] and Tasks Inferred from Differential Expression (TIDEs) for identifying differentially changed metabolic functions [15]. For the enrichment analysis, due to the large number of DEGs, enrichment was determined using genes with an  $FDR < 0.01$  and pathways were defined as statistically significant with a  $p$ -value  $< 0.1$  following Benjamini-Hochberg (BH) correction [51]. For the TIDEs analysis, the subset of genes that mapped to the rat model and that had an  $FDR < 0.01$  were used; pathways with a  $p$ -value  $< 0.1$  when compared to randomly shuffled DEGs were defined as statistically significant.

### 3.5.3 *Collecting and analyzing metabolomics data*

As described above, primary rat neonatal cardiomyocytes were exposed to the compounds at the chosen concentrations. Before lysing the cells with Trizol, the cell supernatant was removed and sent to Metabolon for analysis (<https://www.metabolon.com/>). Raw area counts

were obtained for 181 named metabolites. Metabolites that had greater than 60% missing values (13 named metabolites) were removed from the data set. Next, missing values were imputed as half of the minimum raw area count within a metabolite. Values were then log-scaled and mean-centered within a metabolite. The Mann-Whitney U-test was used to determine if a metabolite was produced or consumed relative to the blank media samples ( $n = 3$ ) as well as differences between treatment and control conditions. Metabolites were considered to be significantly changed if the p-value  $< 0.1$  following Benjamini-Hochberg correction [51].

#### 3.5.4 Building a rat-specific heart model from the human-specific heart model, *iCardio*

The previously published human heart metabolic model, *iCardio* [15], was used to build a rat-specific heart metabolic model to contextualize changes in metabolites and DEGs. All reactions that were included in the human model were included in the rat model. The 6 updates that were made in the human model [15] were also made to the general rat model. Since each model contains species-specific reactions, we identified if each of the 23 rat-specific reactions should be included in the heart; 13 were included based on literature evidence [52] (Supplemental File 3.1).

Further curation was necessary based on metabolites that were measured to be either produced or consumed in the metabolomics data. Metabolites were mapped between the metabolomics data and the metabolic model using compound identifiers from the KEGG database. Exchange reactions, which are reactions in the model that transport a metabolite into or from the extracellular compartment, were added from the general rat model to the heart model if a metabolite was measured to be either consumed or produced relative to blank media in the metabolomics data. These reactions were necessary in order for a metabolite to be modeled as either produced or consumed. Reactions added back to the heart model from the general rat metabolic network are summarized in Supplemental File 3.1.

To identify shifted metabolic functions with the TIDEs approach, we used the previously published list of metabolic tasks [15] with the developed rat-specific heart model. Only genes that mapped to the metabolic model were included in the analysis. For each metabolic gene, a weight was assigned based on the FDR where a gene with an FDR < 0.01 was assigned its log fold change as a weight and 0 otherwise. As with the previous publication, reaction weights were determined based on the weights for individual genes in the complex gene-protein-reaction (GPR) rules. Task scores for individual tasks were calculated as the average weight across reactions in that metabolic task. To establish statistical significance, the weights for each metabolic gene were randomly shuffled 1000 times and significance was determined by comparing the original metabolic task score to random data.

### *3.5.5 Computationally predicting biomarkers of toxicity using TIMBR and RIPTiDe*

We utilized the RIPTiDe algorithm [29] to integrate the metabolomics and transcriptomics data to identify the most likely flux distributions in the rat-specific heart model network for each of the 10 conditions, including both treatment and control groups at both time points. RIPTiDe identifies possible optimal flux distributions in a metabolic network given the cellular investments indicated by the transcriptomic abundance data. The metabolomics data was used to define constraints on each condition-specific model by allowing the consumption of metabolites that were measured to be consumed and forcing production of metabolites measured to be produced. Here, we defined the cellular objective function to be ATP hydrolysis, with an upper bound of 100 units of flux, and production of 1 unit of RNA and DNA, representing general cell maintenance. The exchange reactions for consumed metabolites were given a lower bound of -10, representing a theoretical overabundance of each metabolite for the given objective, while the exchange reactions for produced metabolites were given a lower bound of 1, representing minimal production. These constraints allow for metabolites to be consumed at a relatively high flux while

only forcing production at a low level of flux. The number of constraints placed on each condition-specific model are summarized in Supplemental Table 3.1. Finally, the upper bound of internal reactions was set as 10e6 to ensure that internal fluxes were not constraining the solution space.

After applying these condition-specific constraints, condition-specific models were generated by integrating the median transcripts per million (TPM) for each gene within a condition using the RIPTiDe algorithm [29]. The RIPTiDe algorithm was run with a minimum fraction of 90% of the objective to ensure that differences in ATP flux were not the main determinants of differences between the condition-specific models. Each condition-specific model was flux-sampled 50 times to obtain a range of possible flux distributions that satisfied the pFBA assumption.

Following RIPTiDe analysis, flux samples for each condition were analyzed to identify both unique reactions for each condition and reactions whose flux separated between conditions [29]. Non-metric multidimensional scaling (NMDS) ordination of Bray-Curtis distances between flux samples, calculated using the Vegan package in R [50], was used to visualize differences for reactions that were shared between all conditions. Finally, random forest feature selection [53] was used to determine the reactions whose fluxes most separated between each treatment and control group.

Code to reproduce this analysis is available at (<https://github.com/csbl/Cardiotoxicity>). Transcript data is available on GEO. Metabolomics data is available at Metabolights.

### **3.6 Acknowledgements:**

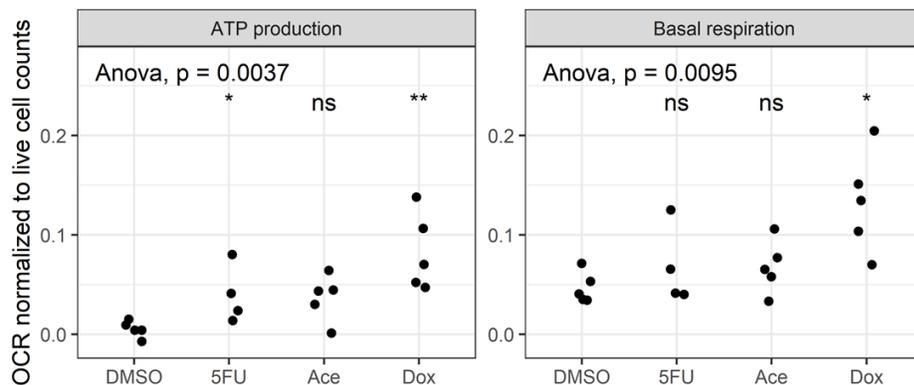
Support for this project was provided by the United States Department of Defense (W81XWH-14-C-0054 to JP). The opinions and assertions contained herein are the private views of the authors and are not to be construed as the official or as reflecting the views of the U.S.

Army, the U.S. Department of Defense, or the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc. (HJF). This manuscript has been approved for public release with unlimited distribution. We would like to thank Bethany Wissman for assisting with the isolation of the primary rat neonatal cardiomyocytes and Laura Dunphy for her feedback on the manuscript.

### 3.7 Author Contributions:

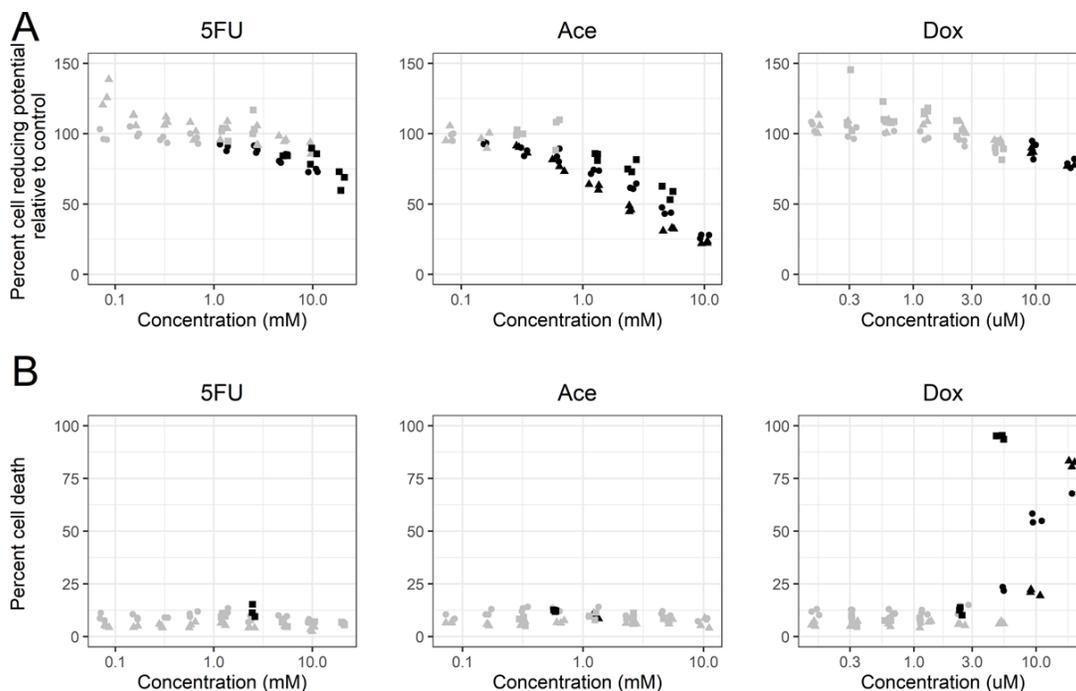
BD, GK, AW, and JP conceived the study. BD, GK, BC, and SN performed experiments. BD performed the computational modeling and data analysis. BD wrote the initial draft of the manuscript. BD, KR, MJ, BC, SN, JS, GK, AW, and JP edited and wrote the final manuscript.

### 3.8 Supplemental Figures, Tables, and Data



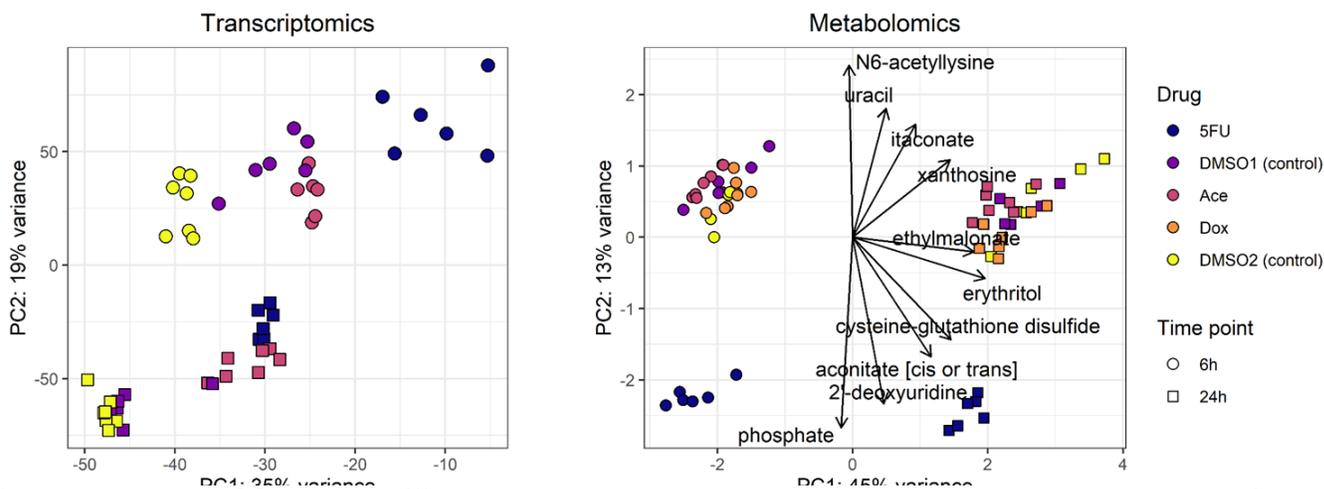
#### Supplemental Figure 3.1 Confirming changes in measures of cellular respiration for the chosen Dox concentration using the MST assay.

The MST assay was performed using the chosen concentrations for each compound. Measures of respiration where there was a statistically significant difference across groups (ANOVA,  $p$ -value  $< 0.05$ ) were followed by treatment vs control comparisons using the Wilcoxon rank-sum test for differences where ns is not significant, \* is  $P$ -value  $< 0.05$ , and \*\* is  $p$ -value  $< 0.01$ . There were 5 experimental replicates per condition, except in the case of 5FU where one well did not respond.



**Supplemental Figure 3.2 Cell reducing potential and cell death measures following 6 hours of exposure to compounds.**

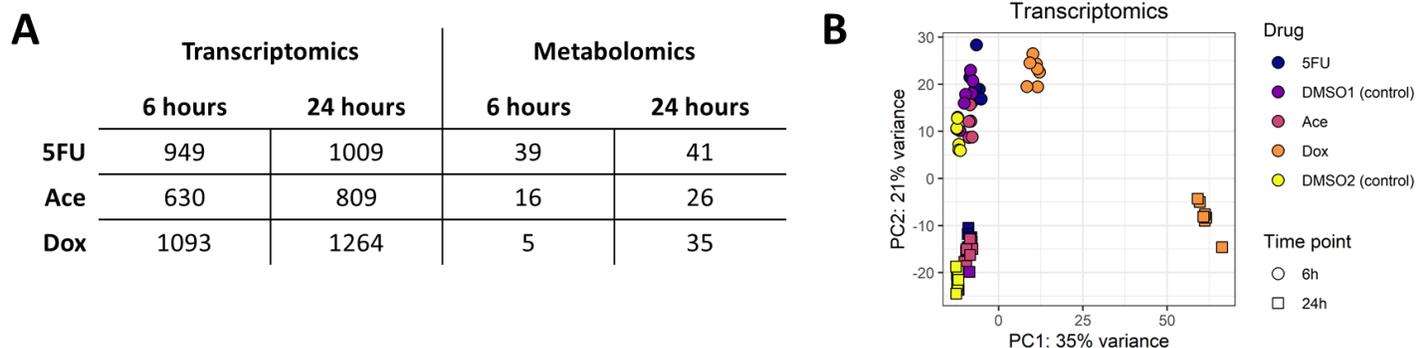
Shapes indicate different cell isolations. Black dots indicate a statistically significant change from the control condition ( $p$ -value  $< 0.05$ ) calculated using Dunnett's test. Black boxes indicate the chosen concentrations for cardiotoxicity characterization. (A) Percent cell viability for a range of concentrations of treatment following 6 hours of exposure. (B) Percent cell death measured using a Hoescht/PI stain.



**Supplemental Figure 3.3 Additional PCA plots demonstrate (A) separation of treatment vs control groups and (B) the top 10 metabolites separating the PCA of the scaled metabolite abundances.**

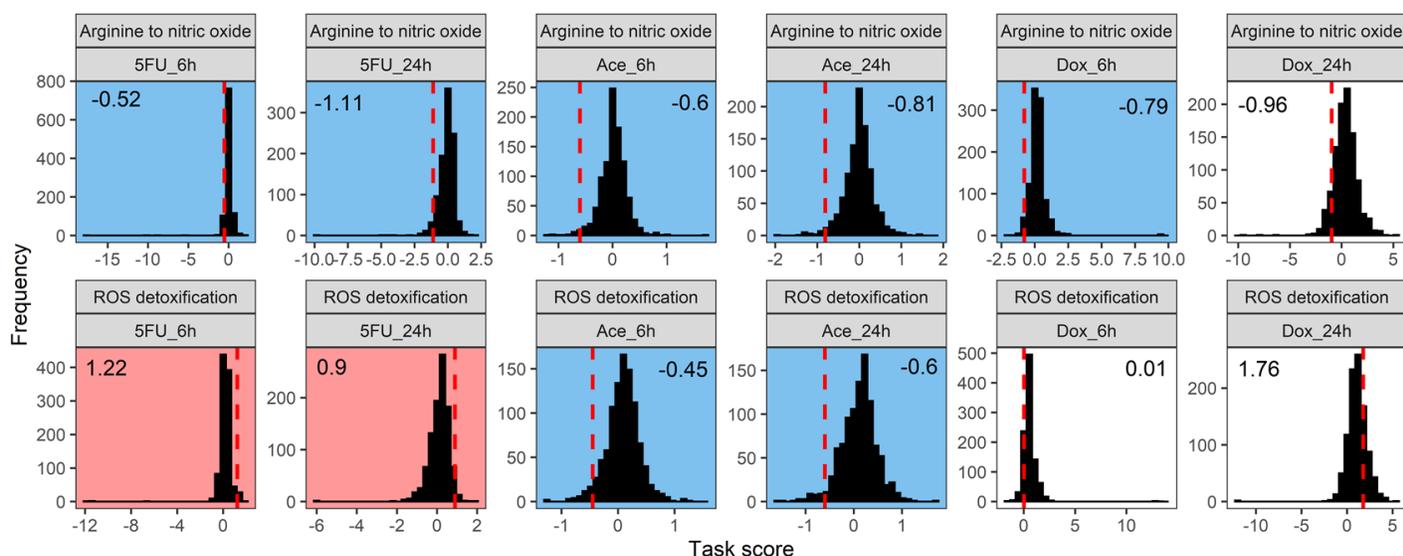
(A) PCA with the Dox samples removed demonstrates separation between treated and control samples at both 6 and 24 hours. (B) The top 10 metabolites show separation in both the first and second principal component. For the first principal component, ethylmalonate and erythritol have a strong influence, suggesting a phenotypic switch between fatty acid and glucose utilization over time, although the direction is unclear. For the second principal component, phosphate and uracil separate the 5FU condition.





**Supplemental Figure 3.5 The rat-specific heart model captures changes in DEGs and metabolomics.**

(A) The number of DEGs ( $FDR < 0.1$ ) and differentially changed metabolites ( $FDR < 0.1$ ) that map to the rat-specific metabolic mode. (B) A PCA of the normalized gene counts that map back to the rat-specific heart model demonstrate clear separation, similar to the PCA of all gene counts (Figure 2), confirming that metabolism has a large determinant in separating conditions.



**Supplemental Figure 3.6 Distribution of random task scores for the metabolic tasks for arginine to nitric oxide and ROS detoxification demonstrate the underlying distribution of DEGs.**

A red background indicates a metabolic task associated with a significant increase in gene expression and a blue background indicates a metabolic task associated with a significant decrease in gene expression ( $p$ -value  $< 0.1$ ). The red dashed line indicates the task score for the actual gene expression data whereas the black bars indicate the calculated task scores when the gene expression data is randomized. In this case, the distribution for the Dox data is significantly wider (i.e. larger range on the x-axis), indicating a larger overall absolute change in gene expression requiring a higher overall average gene expression for ROS production to be deemed significant and a lower overall average gene expression for arginine to nitric oxide to be deemed significant.

	6 hours		24 hours	
	Consumed	Produced	Consumed	Produced
<b>5FU</b>	11	12	13	14
<b>DMSO1</b>	11	13	11	16
<b>Ace</b>	11	11	15	15
<b>Dox</b>	13	13	13	15
<b>DMSO2</b>	13	13	12	16

**Supplemental Table 3.1 Constraints that were placed on individual metabolic models before RIPTIDE integration of transcript counts.**

	6 hours		24 hours	
	# reactions	p-value	# reactions	p-value
<b>5FU</b>	167	0.1741	216	0.0003
<b>DMSO1</b>	199	0.0001	207	0.0000
<b>Ace</b>	184	0.0002	206	0.0001
<b>Dox</b>	190	0.0000	192	0.0000
<b>DMSO2</b>	202	0.0000	214	0.0000

**Supplemental Table 3.2 RIPTiDe models after integration with metabolomics and transcriptomics data.**

*The number of reactions included in each model and the p-value for the Spearman correlation between reaction flux and transcript abundance. In all cases, except the 5FU 6-hour condition, there was a statistically significant correlation between transcript abundance and flux through the model.*

Supplemental File 1. Reactions added to the rat heart model during manual curation.

Supplemental File 2. All TIDEs results after integrating *in vitro* cardiotoxicity data with the rat-specific heart model.

### 3.9 References

1. Albini A, Pennesi G, Donatelli F, Cammarota R, De Flora S, Noonan DM. Cardiotoxicity of anticancer drugs: the need for cardio-oncology and cardio-oncological prevention. *J Natl Cancer Inst.* 2010;102: 14–25. doi:10.1093/jnci/djp440
2. Swain SM, Whaley FS, Ewer MS. Congestive heart failure in patients treated with doxorubicin: a retrospective analysis of three trials. *Cancer.* 2003;97: 2869–2879. doi:10.1002/cncr.11407
3. Carvalho C, Santos RX, Cardoso S, Correia S, Oliveira PJ, Santos MS, et al. Doxorubicin: the good, the bad and the ugly effect. *Curr Med Chem.* 2009;16: 3267–3285. doi:10.2174/092986709788803312
4. Volkova M, Russell R. Anthracycline Cardiotoxicity: Prevalence, Pathogenesis and Treatment. *Curr Cardiol Rev.* 2011;7: 214–220. doi:10.2174/157340311799960645
5. Avila MS, Siqueira SRR, Ferreira SMA, Bocchi EA. Prevention and Treatment of Chemotherapy-Induced Cardiotoxicity. *Methodist DeBakey Cardiovasc J.* 2019;15: 267–273. doi:10.14797/mdcj-15-4-267
6. Li J, Kemp BA, Howell NL, Massey J, Mińczuk K, Huang Q, et al. Metabolic Changes in Spontaneously Hypertensive Rat Hearts Precede Cardiac Dysfunction and Left Ventricular Hypertrophy. *J Am Heart Assoc.* 2019;8: e010926. doi:10.1161/JAHA.118.010926
7. Bauckneht M, Ferrarazzo G, Fiz F, Morbelli S, Sarocchi M, Pastorino F, et al. Doxorubicin Effect on Myocardial Metabolism as a Prerequisite for Subsequent Development of Cardiac Toxicity: A Translational 18F-FDG PET/CT Observation. *J Nucl Med.* 2017;58: 1638–1645. doi:10.2967/jnumed.117.191122
8. Borde C, Kand P, Basu S. Enhanced myocardial fluorodeoxyglucose uptake following Adriamycin-based therapy: Evidence of early chemotherapeutic cardiotoxicity? *World J Radiol.* 2012;4: 220–223. doi:10.4329/wjr.v4.i5.220
9. Rawls KD, Blais EM, Dougherty BV, Vinnakota KC, Pannala VR, Wallqvist A, et al. Genome-Scale Characterization of Toxicity-Induced Metabolic Alterations in Primary Hepatocytes. *Toxicol Sci Off J Soc Toxicol.* 2019;172: 279–291. doi:10.1093/toxsci/kfz197
10. Pannala VR, Wall ML, Estes SK, Trenary I, O'Brien TP, Printz RL, et al. Metabolic network-based predictions of toxicant-induced metabolite changes in the laboratory rat. *Sci Rep.* 2018;8: 11678. doi:10.1038/s41598-018-30149-7
11. Pannala VR, Vinnakota KC, Rawls KD, Estes SK, O'Brien TP, Printz RL, et al. Mechanistic identification of biofluid metabolite changes as markers of acetaminophen-induced liver toxicity in rats. *Toxicol Appl Pharmacol.* 2019;372: 19–32. doi:10.1016/j.taap.2019.04.001

12. Pannala VR, Estes SK, Rahim M, Trenary I, O'Brien TP, Shiota C, et al. Mechanism-based identification of plasma metabolites associated with liver toxicity. *Toxicology*. 2020;441: 152493. doi:10.1016/j.tox.2020.152493
13. Rawls KD, Dougherty BV, Vinnakota KC, Pannala VR, Wallqvist A, Kolling GL, et al. Genome-Scale Metabolic Model Predicts Changes in Renal Metabolism from Chemical Exposure. In prep.
14. Pannala VR, Vinnakota KC, Estes SK, Trenary I, O'Brien TP, Printz RL, et al. Genome-Scale Model-Based Identification of Metabolite Indicators for Early Detection of Kidney Toxicity. *Toxicol Sci Off J Soc Toxicol*. 2020;173: 293–312. doi:10.1093/toxsci/kfz228
15. Dougherty BV, Rawls KD, Kolling GL, Vinnakota KC, Wallqvist A, Papin JA. Identifying functional metabolic shifts in heart failure with the integration of omics data and a cardiomyocyte-specific, genome-scale model. *bioRxiv*. 2020; 2020.07.20.212274. doi:10.1101/2020.07.20.212274
16. Sara JD, Kaur J, Khodadadi R, Rehman M, Lobo R, Chakrabarti S, et al. 5-fluorouracil and cardiotoxicity: a review. *Ther Adv Med Oncol*. 2018;10. doi:10.1177/1758835918780140
17. Taymaz-Nikerel H, Karabekmez ME, Eraslan S, Kırdar B. Doxorubicin induces an extensive transcriptional and metabolic rewiring in yeast cells. *Sci Rep*. 2018;8: 13672. doi:10.1038/s41598-018-31939-9
18. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1: 417–425. doi:10.1016/j.cels.2015.12.004
19. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J Integr Biol*. 2012;16: 284–287. doi:10.1089/omi.2011.0118
20. Ahuja P, Zhao P, Angelis E, Ruan H, Korge P, Olson A, et al. Myc controls transcriptional regulation of cardiac metabolism and mitochondrial biogenesis in response to pathological stress in mice. *J Clin Invest*. 2010;120: 1494–1505. doi:10.1172/JCI38331
21. Schlicker L, Szebenyi DME, Ortiz SR, Heinz A, Hiller K, Field MS. Unexpected roles for ADH1 and SORD in catalyzing the final step of erythritol biosynthesis. *J Biol Chem*. 2019;294: 16095–16108. doi:10.1074/jbc.RA119.009049
22. Zhang N, Yin Y, Xu S-J, Chen W-S. 5-Fluorouracil: Mechanisms of Resistance and Reversal Strategies. *Molecules*. 2008;13: 1551–1569. doi:10.3390/molecules13081551
23. Lee C-C, Yang Y-C, Goodman SD, Chen S, Huang T-Y, Cheng W-C, et al. Deoxyinosine repair in nuclear extracts of human cells. *Cell Biosci*. 2015;5. doi:10.1186/s13578-015-0044-8

24. Evans MD, Saparbaev M, Cooke MS. DNA repair and the origins of urinary oxidized 2'-deoxyribonucleosides. *Mutagenesis*. 2010;25: 433–442. doi:10.1093/mutage/geq031
25. Blais EM, Rawls KD, Dougherty BV, Li ZI, Kolling GL, Ye P, et al. Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nat Commun*. 2017;8: 14250. doi:10.1038/ncomms14250
26. Jaeschke H, Ramachandran A. Oxidant Stress and Lipid Peroxidation in Acetaminophen Hepatotoxicity. *React Oxyg Species Apex NC*. 2018;5: 145–158.
27. Li M, Parker BL, Pearson E, Hunter B, Cao J, Koay YC, et al. Core functional nodes and sex-specific pathways in human ischaemic and dilated cardiomyopathy. *Nat Commun*. 2020;11: 2843. doi:10.1038/s41467-020-16584-z
28. Massion PB, Feron O, Dessy C, Balligand J-L. Nitric oxide and cardiac function: ten years after, and continuing. *Circ Res*. 2003;93: 388–398. doi:10.1161/01.RES.0000088351.58510.21
29. Jenior ML, Moutinho TJ, Dougherty BV, Papin JA. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. *PLoS Comput Biol*. 2020;16. doi:10.1371/journal.pcbi.1007099
30. Gorini S, De Angelis A, Berrino L, Malara N, Rosano G, Ferraro E. Chemotherapeutic Drugs and Mitochondrial Dysfunction: Focus on Doxorubicin, Trastuzumab, and Sunitinib. *Oxid Med Cell Longev*. 2018;2018: 7582730. doi:10.1155/2018/7582730
31. Tokarska-Schlattner M, Zaugg M, Zuppinger C, Wallimann T, Schlattner U. New insights into doxorubicin-induced cardiotoxicity: the critical role of cellular energetics. *J Mol Cell Cardiol*. 2006;41: 389–405. doi:10.1016/j.yjmcc.2006.06.009
32. Bahadır A, Kurucu N, Kadioğlu M, Yenilme E. The Role of Nitric Oxide in Doxorubicin-Induced Cardiotoxicity: Experimental Study. *Turk J Hematol*. 2014;31: 68–74. doi:10.4274/Tjh.2013.0013
33. Farías JG, Molina VM, Carrasco RA, Zepeda AB, Figueroa E, Letelier P, et al. Antioxidant Therapeutic Strategies for Cardiovascular Conditions Associated with Oxidative Stress. *Nutrients*. 2017;9. doi:10.3390/nu9090966
34. Deidda M, Madonna R, Mango R, Pagliaro P, Bassareo PP, Cugusi L, et al. Novel insights in pathophysiology of antineoplastic drugs-induced cardiotoxicity and cardioprotection. *J Cardiovasc Med Hagerstown Md*. 2016;17 Suppl 1 Special issue on Cardiotoxicity from Antineoplastic Drugs and Cardioprotection: e76–e83. doi:10.2459/JCM.0000000000000373
35. Burridge PW, Li YF, Matsa E, Wu H, Ong S, Sharma A, et al. Human Induced Pluripotent Stem Cell–Derived Cardiomyocytes Recapitulate the Predilection of Breast Cancer

Patients to Doxorubicin–Induced Cardiotoxicity. *Nat Med.* 2016;22: 547–556.  
doi:10.1038/nm.4087

36. Ryall KA, Bezzerides VJ, Rosenzweig A, Saucerman JJ. Phenotypic screen quantifying differential regulation of cardiac myocyte hypertrophy identifies CITED4 regulation of myocyte elongation. *J Mol Cell Cardiol.* 2014;72: 74–84. doi:10.1016/j.yjmcc.2014.02.013

37. Chen J-Y, Hu R-Y, Chou H-C. Quercetin-induced cardioprotection against doxorubicin cytotoxicity. *J Biomed Sci.* 2013;20: 95. doi:10.1186/1423-0127-20-95

38. Chao H-H, Liu J-C, Hong H-J, Lin J, Chen C-H, Cheng T-H. L-carnitine reduces doxorubicin-induced apoptosis through a prostacyclin-mediated pathway in neonatal rat cardiomyocytes. *Int J Cardiol.* 2011;146: 145–152. doi:10.1016/j.ijcard.2009.06.010

39. Lamberti M, Porto S, Marra M, Zappavigna S, Grimaldi A, Feola D, et al. 5-Fluorouracil induces apoptosis in rat cardiocytes through intracellular oxidative stress. *J Exp Clin Cancer Res CR.* 2012;31: 60. doi:10.1186/1756-9966-31-60

40. Lamberti M, Porto S, Zappavigna S, Addeo E, Marra M, Miraglia N, et al. A mechanistic study on the cardiotoxicity of 5-fluorouracil in vitro and clinical and occupational perspectives. *Toxicol Lett.* 2014;227: 151–156. doi:10.1016/j.toxlet.2014.03.018

41. Jin SM, Park K. Acetaminophen Induced Cytotoxicity and Altered Gene Expression in Cultured Cardiomyocytes of H9C2 Cells. *Environ Health Toxicol.* 2012;27.  
doi:10.5620/eh.t.2012.27.e2012011

42. McQuin C, Goodman A, Chernyshev V, Kamensky L, Cimini BA, Karhohs KW, et al. CellProfiler 3.0: Next-generation image processing for biology. *PLOS Biol.* 2018;16: e2005970. doi:10.1371/journal.pbio.2005970

43. Hothorn T, Bretz F, Westfall P. Simultaneous Inference in General Parametric Models. *Biom J.* 2008;50: 346–363. doi:10.1002/bimj.200810425

44. Kenwood BM, Weaver JL, Bajwa A, Poon IK, Byrne FL, Murrow BA, et al. Identification of a novel mitochondrial uncoupler that does not depolarize the plasma membrane. *Mol Metab.* 2014;3: 114–123. doi:10.1016/j.molmet.2013.11.005

45. Nagdas S, Kashatus JA, Nascimento A, Hussain SS, Trainor RE, Pollock SR, et al. Drp1 Promotes KRas-Driven Metabolic Changes to Drive Pancreatic Tumor Growth. *Cell Rep.* 2019;28: 1845-1859.e5. doi:10.1016/j.celrep.2019.07.031

46. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34: 525–527. doi:10.1038/nbt.3519

47. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research.* 2015;4: 1521. doi:10.12688/f1000research.7563.2

48. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15: 550. doi:10.1186/s13059-014-0550-8
49. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010;11: R106. doi:10.1186/gb-2010-11-10-r106
50. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci.* 2003;14: 927–930. doi:10.1111/j.1654-1103.2003.tb02228.x
51. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol.* 1995;57: 289–300.
52. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2018;46: D8–D13. doi:10.1093/nar/gkx1095
53. Calle ML, Urrea V, Boulesteix A-L, Malats N. AUC-RF: a new strategy for genomic profiling with random forest. *Hum Hered.* 2011;72: 121–132. doi:10.1159/000330778

## Chapter 4: Systems biology approaches help to facilitate interpretation of cross-species comparisons

Bonnie Dougherty and Jason A. Papin\*

\*Corresponding author. Tel: +1 434 924 8195; Email: [papin@virginia.edu](mailto:papin@virginia.edu); Address: Departments of Biomedical Engineering, University of Virginia, Box 800759, Health System, Charlottesville, VA 22908

Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, 22908, USA.

The text included in this chapter has been published here:

Dougherty BV, Papin JA. Systems biology approaches help to facilitate interpretation of cross-species comparisons | Elsevier Enhanced Reader. [cited 5 Oct 2020].  
doi:10.1016/j.cotox.2020.06.002

## 4.1 Abstract

Translation of biological knowledge from animal models to humans is an important step in the development of therapeutics but there remain limitations for effective translation. Systems biology offers approaches to understand the limitations for translation between species through data-driven models, such as methods that rely on learning patterns from data, and mechanism-driven models of biological processes, such as pharmacokinetic models. Here, we describe recent advances in both data-driven and mechanism-driven systems biology approaches to better understand limitations to translation from animal models to humans. Both approaches to modeling have their strengths and weaknesses but still provide key biological insight for translating between model systems and humans (Figure 4.1). The presented methods not only identify differences between different model organisms but also provide opportunities to identify shared biomarkers and unique biological insight.

## 4.2 Introduction

Animal models are commonly used to understand the basic biology of human disease, identify new drug targets, and predict the toxicity of new drugs in humans. However, animal models have been shown to be both good [1] and poor [2–4] models of human biology, depending on the disease or model system. In order to facilitate better translation between model systems and humans, we need a better understanding of both the strengths and limitations of animal models. Systems biology is an approach that could help to articulate the limitations and opportunities for cross-species comparisons. For example, evidence of protein homology between model organisms and humans does not ensure direct translation due to inherent differences, such as levels of transcription and network structure. Consequently, systems biology, which enables the interpretation of large datasets such as transcriptomics, proteomics, or metabolomics, either with or without knowledge of the underlying biological mechanisms, can help to understand the drivers of these complex species-specific differences between genotype and phenotype to facilitate better translation to humans.

Systems biology can involve both models driven primarily by relationships in the data, rather than integration of underlying biological knowledge, as well as models that utilize *a priori*

	<b>Data-driven models</b>	<b>Mechanism-driven models</b>
<b>Strengths</b>	<ul style="list-style-type: none"> <li>• No need for underlying details of biological mechanism</li> <li>• Can identify important biological features of data</li> </ul>	<ul style="list-style-type: none"> <li>• Provides mechanistic understanding of predictions</li> </ul>
<b>Weaknesses</b>	<ul style="list-style-type: none"> <li>• Can require large amounts of data</li> <li>• No mechanistic understanding supporting predictions</li> </ul>	<ul style="list-style-type: none"> <li>• Need underlying network structure and interactions</li> <li>• Can require extensive parameter values</li> </ul>

**Figure 4.1 Strengths and weaknesses of data- vs. mechanism-driven models.**

mechanistic biological knowledge to gain insight into species-specific differences. In this review, we first discuss recent efforts using data-driven models, such as machine learning, to guide choosing appropriate model systems and predicting pathway and gene levels in humans based on model organism data. Second, we discuss recent efforts using mechanism-driven models of biological processes, such as pharmacokinetic models and genome-scale metabolic network reconstructions, to predict drug dosing in humans from animal studies and identify shared biomarkers of toxicity. Together, these methods highlight how big data and systems biology can facilitate translation between model organisms and humans while yielding key biological insights.

### **4.3 Data-driven models**

One common form of data-driven modeling in systems biology is machine learning, which encompasses a broad range of computational methods that learn patterns from data to make new predictions. In order to learn these patterns, a large amount of data is needed. Recent efforts to centralize data storage in both general [5,6] and field-specific [7,8] databases have facilitated efforts to use systems biology to yield insights into species-specific differences. Here, we discuss how machine learning approaches have been used to (a) predict gene expression data in humans from model organism data and (b) rationally choose model systems to improve translation to humans.

#### *4.3.1 Machine learning*

Recent studies have utilized data-driven approaches to identify human differentially expressed genes from rat epithelial cells [9] and mouse models of human diseases [10], sepsis [11\*], and immune responses [12]. These studies highlighted that data-driven approaches, such as machine learning methods, were better at predicting human differentially expressed genes than directly translating human gene expression from mouse data using only protein homology.

In addition to predicting human differentially expressed genes, Normand et al, 2019 identified new pathways relevant to human diseases by using predicted differentially expressed genes whose homologs were not differentially expressed in the mouse data. As results from animal models of sepsis have led to poor translation in the clinic [13,14], Brubaker et al, 2019 used different machine learning approaches to predict both human differentially expressed genes and pathway-level changes. The authors were able to identify pathways that had a similar response in both mouse models and humans and could therefore serve as biomarkers for successful therapeutics for sepsis. Another targeted study, the IMPROVER toxicology challenge, identified common biomarkers of smoking between rat epithelial cells and humans using machine learning approaches [15\*]. Although both studies were able to use machine learning to improve predictions of human gene expression through identification of new disease-relevant pathways and shared biomarkers, there were still significant gaps between predictions and human data. For example, although predictions of human differentially expressed genes were better than different homology-based translation, the machine learning models had low F-scores, representing low accuracy for the models [11\*]. Both studies acknowledged that more heterogeneous data that covers a wider range of mouse strains, mouse models, and human data may help to improve predictions.

In addition to predicting human differentially expressed genes, machine learning has been used to identify the most appropriate model system for a specific disease in order to improve translation to humans. For example, a recent study utilized gene expression data from the Japanese Toxicogenomics Database [7] to determine what model system, either human or rat hepatocytes *in vitro* or the rat liver or kidney *in vivo*, best captured human adverse events reported in the literature [16]. The authors demonstrated that each model system captured different adverse events and that no adverse event was captured by all model systems. The results from this study could be used to rationally choose a model system depending on the adverse event that is being tested for [17].

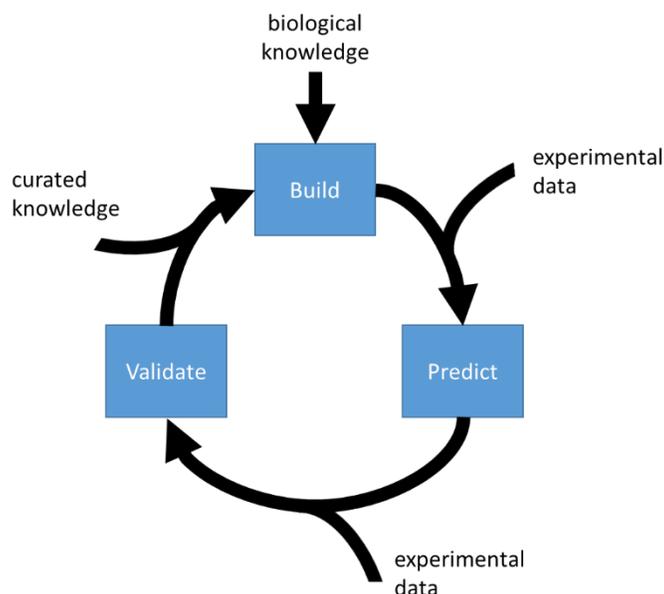
### 4.3.2 *Rationally choosing model organisms*

Similarly, McGary et al [18] described the use of “phenologs”, or model organisms with a specific phenotype whose genotype is similar to a human disease, as a way to rationally choose model organisms for human diseases. These phenologs do not share a common phenotype but a common genotype, through common orthologous genes, suggesting that the underlying biology driving the different phenotypes may be similar. These phenologs were more recently expanded upon using an approach that represents homologous proteins in a network to identify functional similarity [19]. The functional homology network was used to identify phenologs between model organisms and human diseases. Finally, another study used gene expression across various tissues and species to identify genes that are dominated by variance across tissues rather than species, suggesting that model organisms are appropriate for representing human diseases associated with these genes [20]. Together, these methods demonstrate approaches that utilize large data to make more informed decisions regarding model systems to facilitate translation from model organisms to humans.

## 4.4 **Mechanism-driven models**

Mechanism-driven models, another systems biology approach, are models that incorporate biological knowledge, such as interactions between proteins or biochemical reactions that a cell catalyzes, into a mathematical framework to interpret data. Given the biological knowledge incorporated into these models, the predictions generated can also lead to identification of mechanisms driving a specific phenotype of interest or driving differences between mouse models and humans. These models are usually built using an iterative process (Figure 4.2), incorporating biological knowledge, experimental data, and new predictions to continuously refine the model. The model building process itself can yield insight into mechanistic

differences between model systems and humans as well as parameters that inform differences between model systems and humans. Here, we discuss four groups of mechanistic computational models that have recently yielded insights into differences between model organisms and



**Figure 4.2 The iterative model building process.**

humans: pharmacokinetic models, genome-scale metabolic network reconstructions, signaling network models, and protein-protein interaction networks.

#### *4.4.1 Pharmacokinetic (PKPB) models*

Pharmacokinetic (PKPD) models are often ordinary differential equation (ODE) based models used to model the absorption, metabolism, and secretion of compounds. Thiel et al [21] used a PKPD model to explore how different parameterizations of the model affect the accuracy of translation from preclinical animal models to humans. The authors highlighted that even though the presence of orthologous genes in the model were highly correlated between mice and humans, the parameters for transporters and enzymes were not, highlighting the importance of species-relevant parameters. Other studies have also used PKPD models to explore the effects

of dose amount and scheduling on side effects with cancer therapeutics [22–24]. These models helped to identify appropriate dosing in humans from model organism data to minimize toxic events. While useful in this case, PKPD models require significant mechanistic information about the side effects of chemotherapeutics, which in the case of other compounds, is not readily available. Similarly, a mathematical model of drug-induced liver injury (DILIsym) has been developed (<https://www.simulations-plus.com/software/dilisyml/>) that models common mechanisms of liver injury and has been used to better predict differences in human and rat responses to compounds [25,26]. This model, along with data from rats and humans, was able to identify why one drug failed in pre-clinical models but was successful in the clinic [25] and why one drug failed in clinic after being successful in pre-clinical models [26]. By using a mechanistic model, the authors were able to identify the mechanisms that potentially caused the differences in toxicity, highlighting species-specific differences that are relevant for future toxicity profiling.

#### 4.4.2 Genome scale metabolic network reconstructions (GENREs)

Genome-scale metabolic network reconstructions (GENREs) are mathematical representations of the genes and proteins, biochemical reactions that these proteins catalyze, and the metabolites converted in those biochemical reactions for an organism of interest [27]. These models allow for the systematic integration of large data sets to interpret metabolic shifts and predict downstream changes in metabolism. A model of paired rat and human metabolism was recently published to facilitate cross-species comparisons [28\*]. The models were built using protein homology to translate proteins present in a human model of metabolism to a rat model. The model building process revealed key differences in the metabolic structure of humans and rats both at a reaction and gene-protein-reaction level. Further, the model was used to integrate transcriptomics data for both rat and human cells treated *in vitro* with a variety of compounds to predict biomarkers of chemical exposure. The authors demonstrated that rats, a model organism

used for human toxicity studies, can respond differently to compounds compared to humans. The model and analytical framework allow for the identification of biological differences, either in network structure or transcriptomics data, which drive differences in responses between the two organisms. The models have been used with both *in vitro* [29] and *in vivo* [30,31] rat models of hepatotoxicity to identify new biomarkers of toxicity as well as potential mechanisms for these biomarkers. Other models exist for model organisms frequently used in toxicity studies, such as the mouse [32], but need further curation to allow for comparisons with updated models of human metabolism. Finally, a database of species-specific essential reactions was recently published to allow for cross-species comparisons [33], which can be integrated with these network models to identify functional metabolic similarities and differences between species.

#### 4.4.3 Pathway-based models

Pathway-based models, such as signaling network models, are often ODE representations of the dynamics of key signaling pathways. A recent paper [34\*] used a signaling network model of the EGF, Wnt, and Notch pathways to explore how two *Caenorhabditis* species respond differently to perturbations in the signaling network. The authors showed that, even with a similar overall network structure, perturbations on 13 parameters in the network resulted in the differences seen in development in the two species. Together, this work highlights that even though two species may have similar overall network structure, responses can differ based on individual nodes in the network. These signaling networks require detailed knowledge of the individual signaling pathways which are often available for both model organisms and humans. However, the models also require either a large number of measured parameters or experimental data to inform model parameters.

Protein-protein interaction (PPI) network models represent the interactions between proteins within the context of a cell to allow for identification of key network modules as well as

general network structure. These protein-protein relationships are often published in large databases and include interactions across species that can be utilized to build these models [35–37]. A recent paper [38] explored differences in the overall network structure for cytochrome P450 (CYP) enzymes, which have a major role in metabolizing drugs. Overall, the authors demonstrated that there are network-level differences in PPI networks between rats, humans, and mice. This study was a first step in demonstrating key structural differences. Further work is needed to understand how changes at the network and structural level impact the differences between model organisms and the human response to perturbations, particularly in response to therapeutics and potential toxicity.

Finally, an alternative method that uses the strengths of mechanistic-modeling, such as the incorporation of biological knowledge, and the strengths of data-driven modeling is using pathway-centric approaches. Previous studies have shown that pathway level activity is more translatable between model organisms and humans than homology-based translation of gene expression [10,11\*,39]. A recent study expanded on this concept by using pathway databases to map from differentially expressed genes to transcription factors responsible for the observed changes. The authors used both mouse and human data to demonstrate that changes in transcription factor levels are indirectly conserved in species-specific responses [40\*], suggesting that condensing species-specific gene expression changes to the pathway level may help to reconcile differences [41\*]. Further, recently published pathway databases can help to consolidate links between molecular initiating events, such as a change in a specific transcription factor or pathway, and adverse toxicological outcomes to help predict if a chemical will elicit a specific adverse outcome [42,43]. These pathway-level approaches offer an alternative for reconciling known differences between model systems and humans, utilizing existing databases of biological knowledge, to more effectively translate between preclinical studies and the clinic.

## 4.5 Conclusions

As described in this review, systems biology approaches can have significant value in delineating the strengths and weaknesses of cross-species comparisons for toxicology applications. The methods presented here have advantages and disadvantages (Figure 4.1). Machine learning applied to large biological datasets facilitated insights into species-specific differences through translation from model organisms to human gene expression and the rational selection of model systems. However, limitations of such approaches include: (1) the dependence on large amounts of data for both model organisms and humans, and (2) lack of a connection between the observed phenotypes and associated mechanisms.

Mechanistic computational models offer a potential solution to interpret large data sets and that can map phenotypes to potential mechanisms. Such models have made progress in identifying key mechanisms that govern differences between model organisms and humans. However, these methods often require deep biological knowledge that may not be available for every disease or compound of interest. In addition, the connection between genotypic differences, network structure, and gene expression and the resulting differences in phenotypes is often quite complex and much work remains to sufficiently delineate the needed parameters and network structures. Together, the approaches presented here offer a myriad of ways to learn from the wealth of data and knowledge already published in the literature to gain biological insight on species-specific differences of relevance to toxicology.

## 4.6 Acknowledgements

Support for this project was provided by the United States Department of Defense (W81XWH-14-C-0054 to JP) and the National Science Foundation Graduate Research Fellowship Program (awarded to BD).

## 4.7 References

1. Takao K, Miyakawa T. Genomic responses in mouse models greatly mimic human inflammatory diseases. *Proc Natl Acad Sci U S A*. 2015;112: 1167–1172. doi:10.1073/pnas.1401965111
2. Monticello TM, Jones TW, Dambach DM, Potter DM, Bolt MW, Liu M, et al. Current nonclinical testing paradigm enables safe entry to First-In-Human clinical trials: The IQ consortium nonclinical to clinical translational database. *Toxicol Appl Pharmacol*. 2017;334: 100–109. doi:10.1016/j.taap.2017.09.006
3. Tamaki C, Nagayama T, Hashiba M, Fujiyoshi M, Hizue M, Kodaira H, et al. Potentials and limitations of nonclinical safety assessment for predicting clinical adverse drug reactions: correlation analysis of 142 approved drugs in Japan. *J Toxicol Sci*. 2013;38: 581–598. doi:10.2131/jts.38.581
4. Clark M, Steger-Hartmann T. A big data approach to the concordance of the toxicity of pharmaceuticals in animals and humans. *Regul Toxicol Pharmacol RTP*. 2018;96: 94–105. doi:10.1016/j.yrtph.2018.04.018
5. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41: D991–D995. doi:10.1093/nar/gks1193
6. Athar A, Füllgrabe A, George N, Iqbal H, Huerta L, Ali A, et al. ArrayExpress update – from bulk to single-cell expression data. *Nucleic Acids Res*. 2019;47: D711–D715. doi:10.1093/nar/gky964
7. The Japanese toxicogenomics project: Application of toxicogenomics - Uehara - 2010 - Molecular Nutrition & Food Research - Wiley Online Library. [cited 28 Jan 2020]. Available: <https://onlinelibrary.wiley.com/doi/full/10.1002/mnfr.200900169>
8. Darde TA, Gaudriault P, Beranger R, Lancien C, Caillarec-Joly A, Sallou O, et al. TOXslgN: a cross-species repository for toxicogenomic signatures. *Bioinformatics*. 2018;34: 2116–2122. doi:10.1093/bioinformatics/bty040
9. Rhrissorrakrai K, Belcastro V, Bilal E, Norel R, Poussin C, Mathis C, et al. Understanding the limits of animal models as predictors of human biology: lessons learned from the sbv IMPROVER Species Translation Challenge. *Bioinforma Oxf Engl*. 2015;31: 471–483. doi:10.1093/bioinformatics/btu611
10. Normand R, Du W, Briller M, Gaujoux R, Starosvetsky E, Ziv-Kenet A, et al. Found In Translation: a machine learning model for mouse-to-human inference. *Nat Methods*. 2018;15: 1067–1073. doi:10.1038/s41592-018-0214-9

11. Brubaker DK, Proctor EA, Haigis KM, Lauffenburger DA. Computational translation of genomic responses from experimental model systems to humans. *PLOS Comput Biol*. 2019;15: e1006286. doi:10.1371/journal.pcbi.1006286
12. Seok J. Evidence-Based Translation for the Genomic Responses of Murine Models for the Study of Human Immunity. *PLoS ONE*. 2015;10. doi:10.1371/journal.pone.0118017
13. Seok J, Warren HS, Cuenca AG, Mindrinos MN, Baker HV, Xu W, et al. Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proc Natl Acad Sci U S A*. 2013;110: 3507–3512.
14. Leist M, Hartung T. Inflammatory findings on species extrapolations: humans are definitely no 70-kg mice. *Arch Toxicol*. 2013;87: 563–567. doi:10.1007/s00204-013-1038-0
15. Belcastro V, Poussin C, Xiang Y, Giordano M, Tripathi KP, Boda A, et al. The sbv IMPROVER Systems Toxicology computational challenge: Identification of human and species-independent blood response markers as predictors of smoking exposure and cessation status. *Comput Toxicol*. 2018;5: 38–51. doi:10.1016/j.comtox.2017.07.004
16. Taškova K, Fontaine J-F, Mrowka R, Andrade-Navarro MA. Evaluation of in vivo and in vitro models of toxicity by comparison of toxicogenomics data with the literature. *Methods*. 2018;132: 57–65. doi:10.1016/j.ymeth.2017.07.010
17. Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, et al. Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul Toxicol Pharmacol RTP*. 2000;32: 56–67. doi:10.1006/rtph.2000.1399
18. McGary KL, Park TJ, Woods JO, Cha HJ, Wallingford JB, Marcotte EM. Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc Natl Acad Sci*. 2010;107: 6544–6549. doi:10.1073/pnas.0910200107
19. Fan J, Cannistra A, Fried I, Lim T, Schaffner T, Crovella M, et al. Functional protein representations from biological networks enable diverse cross-species inference. *Nucleic Acids Res*. 2019;47: e51–e51. doi:10.1093/nar/gkz132
20. Breschi A, Djebali S, Gillis J, Pervouchine DD, Dobin A, Davis CA, et al. Gene-specific patterns of expression variation across organs and species. *Genome Biol*. 2016;17: 151. doi:10.1186/s13059-016-1008-y
21. Thiel C, Schneckener S, Krauss M, Ghallab A, Hofmann U, Kanacher T, et al. A Systematic Evaluation of the Use of Physiologically Based Pharmacokinetic Modeling for Cross-Species Extrapolation. *J Pharm Sci*. 2015;104: 191–206. doi:10.1002/jps.24214
22. Zhu AZ. Quantitative translational modeling to facilitate preclinical to clinical efficacy & toxicity translation in oncology. *Future Sci OA*. 2018;4: FSO306. doi:10.4155/foa-2017-0152

23. Shankaran H, Cronin A, Barnes J, Sharma P, Tolsma J, Jasper P, et al. Systems Pharmacology Model of Gastrointestinal Damage Predicts Species Differences and Optimizes Clinical Dosing Schedules. *CPT Pharmacomet Syst Pharmacol*. 2018;7: 26–33. doi:10.1002/psp4.12255
24. Chen Y, Zhao K, Liu F, Xie Q, Zhong Z, Miao M, et al. Prediction of Deoxy podophyllotoxin Disposition in Mouse, Rat, Monkey, and Dog by Physiologically Based Pharmacokinetic Model and the Extrapolation to Human. *Front Pharmacol*. 2016;7. doi:10.3389/fphar.2016.00488
25. Battista C, Yang K, Stahl SH, Mettetal JT, Watkins PB, Siler SQ, et al. Using Quantitative Systems Toxicology to Investigate Observed Species Differences in CKA-Mediated Hepatotoxicity. *Toxicol Sci Off J Soc Toxicol*. 2018;166: 123–130. doi:10.1093/toxsci/kfy191
26. Generaux G, Lakhani VV, Yang Y, Nadanaciva S, Qiu L, Riccardi K, et al. Quantitative systems toxicology (QST) reproduces species differences in PF-04895162 liver safety due to combined mitochondrial and bile acid toxicity. *Pharmacol Res Perspect*. 2019;7: e00523. doi:10.1002/prp2.523
27. Rawls KD, Dougherty BV, Blais EM, Stancliffe E, Kolling GL, Vinnakota K, et al. A simplified metabolic network reconstruction to promote understanding and development of flux balance analysis tools. *Comput Biol Med*. 2019;105: 64–71. doi:10.1016/j.combiomed.2018.12.010
28. Blais EM, Rawls KD, Dougherty BV, Li ZI, Kolling GL, Ye P, et al. Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nat Commun*. 2017;8: 14250. doi:10.1038/ncomms14250
29. Rawls KD, Blais EM, Dougherty BV, Vinnakota KC, Pannala VR, Wallqvist A, et al. Genome-Scale Characterization of Toxicity-Induced Metabolic Alterations in Primary Hepatocytes. *Toxicol Sci Off J Soc Toxicol*. 2019;172: 279–291. doi:10.1093/toxsci/kfz197
30. Pannala VR, Wall ML, Estes SK, Trenary I, O'Brien TP, Printz RL, et al. Metabolic network-based predictions of toxicant-induced metabolite changes in the laboratory rat. *Sci Rep*. 2018;8: 11678. doi:10.1038/s41598-018-30149-7
31. Pannala VR, Vinnakota KC, Rawls KD, Estes SK, O'Brien TP, Printz RL, et al. Mechanistic identification of biofluid metabolite changes as markers of acetaminophen-induced liver toxicity in rats. *Toxicol Appl Pharmacol*. 2019;372: 19–32. doi:10.1016/j.taap.2019.04.001
32. Sigurdsson MI, Jamshidi N, Steingrimsson E, Thiele I, Palsson BØ. A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. *BMC Syst Biol*. 2010;4: 140. doi:10.1186/1752-0509-4-140
33. Labena AA, Ye Y-N, Dong C, Zhang F-Z, Guo F-B. SSER: Species specific essential reactions database. *BMC Syst Biol*. 2017;11: 50. doi:10.1186/s12918-017-0426-0

34. Dawes AT, Wu D, Mahalak KK, Zitnik EM, Kravtsova N, Su H, et al. A computational model predicts genetic nodes that allow switching between species-specific responses in a conserved signaling network. *Integr Biol.* 2017;9: 156–166. doi:10.1039/c6ib00238b
35. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43: D447–D452. doi:10.1093/nar/gku1003
36. Mellor JC, Yanai I, Clodfelter KH, Mintseris J, DeLisi C. Predictome: a database of putative functional links between proteins. *Nucleic Acids Res.* 2002;30: 306–309. doi:10.1093/nar/30.1.306
37. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* 2004;32: D449–451. doi:10.1093/nar/gkh086
38. Karthikeyan BS, Akbarsha MA, Parthasarathy S. Network analysis and cross species comparison of protein-protein interaction networks of human, mouse and rat cytochrome P450 proteins that degrade xenobiotics. *Mol Biosyst.* 2016;12: 2119–2134. doi:10.1039/c6mb00210b
39. Hafemeister C, Romero R, Bilal E, Meyer P, Norel R, Rhrissorrakrai K, et al. Inter-species pathway perturbation prediction via data-driven detection of functional homology. *Bioinforma Oxf Engl.* 2015;31: 501–508. doi:10.1093/bioinformatics/btu570
40. Holland CH, Szalai B, Saez-Rodriguez J. Transfer of regulatory knowledge from human to mouse for functional genomics analysis. *Biochim Biophys Acta BBA - Gene Regul Mech.* 2019; 194431. doi:10.1016/j.bbagr.2019.194431
41. McMullen PD, Bhattacharya S, Woods CG, Pendse SN, McBride MT, Soldatow VY, et al. Identifying qualitative differences in PPAR $\alpha$  signaling networks in human and rat hepatocytes and their significance for next generation chemical risk assessment methods. *Toxicol In Vitro.* 2019; 104463. doi:10.1016/j.tiv.2019.02.017
42. Wittwehr C, Aladjov H, Ankley G, Byrne HJ, de Knecht J, Heinzle E, et al. How Adverse Outcome Pathways Can Aid the Development and Use of Computational Prediction Models for Regulatory Toxicology. *Toxicol Sci Off J Soc Toxicol.* 2017;155: 326–336. doi:10.1093/toxsci/kfw207
43. Staal YCM, Pennings JLA, Hessel EVS, Piersma AH. Advanced Toxicological Risk Assessment by Implementation of Ontologies Operationalized in Computational Models. *Appl Vitro Toxicol.* 2017;3: 325–332. doi:10.1089/aivt.2017.0019

# Chapter 5: Discussion and Future Work

## 5.1 Discussion

In this dissertation, we provide insight into changes in heart metabolism in diseased states as well as novel approaches to integrate experimental data with genome-scale metabolic network reconstructions (GENREs). While the first GENREs were developed to study bacterial metabolism [1–3], more recent work has developed general models of human metabolism [4–7]. These models have been used to study a wide array of diseased states [8–11], identify new drug targets [12], and provide a mechanistic understanding of drug side effects [13]. In an attempt to disseminate the utility of GENREs to the general community, we have recently published a simplified metabolic network, *iSIM*, with associated methods and code for analyses [14]. Although the utility of GENREs has been demonstrated in both bacterial and human models, few studies have utilized GENREs to better understand metabolic shifts in the heart. To date, only two metabolic models of the heart have been published [15,16]. These models were used to examine the relationship between substrate utilization and efficiency of the heart [15], and to predict epistatic interactions in the heart and biomarkers of heart disease [16]. However, neither of these studies utilized updated, more-comprehensive models of human metabolism. Here, we develop new paired models of human- and rat-specific heart metabolism and demonstrate their utility in identifying metabolic shifts in both heart failure and *in vitro* cardiotoxicity.

### *5.1.1 Identifying metabolic shifts in the context of heart failure and in vitro cardiotoxicity using metabolic models*

We built paired human- and rat-specific models of heart metabolism and demonstrated their utility in predicting shifted metabolic functions in the context of late-stage human heart failure and *in vitro* cardiotoxicity of primary rat neonatal cardiomyocytes. First, we built a human heart-

specific model of metabolism using immunohistochemistry data from the Human Protein Atlas (HPA) (v18.proteinstlas.org; [17]) and metabolic tasks. Using this approach, we demonstrated the utility of the model building process in generating biological knowledge, e.g. through identification and addition of reactions related to nitric oxide (NO) synthesis (Chapter 2). Next, we used the model to interpret changes in microarray gene expression in late-stage heart failure using the novel Tasks Inferred from Differential Expression (TIDEs) approach. It is important to note the complexity of late-stage heart failure, both in etiology and progression of the disease. Here, we used publicly available datasets for ischemic and idiopathic heart failure to identify commonly shifted metabolic functions. As a result of this approach, we propose changes in NO synthesis and Neu5Ac synthesis as potential biomarkers of, or targets for, future clinical intervention for late-stage heart failure.

In the context of *in vitro* cardiotoxicity, we developed a rat-specific heart model to identify shifted metabolic functions and reactions that are differentially active in the context of *in vitro* cardiotoxicity. We used paired transcriptomics and metabolomics data to characterize the response of primary rat neonatal cardiomyocytes to three compounds: 5-fluorouracil (5FU), acetaminophen (Ace), and doxorubicin (Dox). First, using our collected transcriptomics data integrated with the rat-specific heart model, we identified the effect of these compounds on specific metabolic functions using the TIDEs approach [18]. We identified that both Dox and 5FU shifted DNA/RNA synthesis and nucleotide metabolism, consistent with their known chemotherapeutic mechanisms of action. For Ace, we identified shifts in phospholipid metabolism, suggesting a potential mechanism for cardiotoxicity consistent with toxicity in the liver and kidney. The generation of oxidized phospholipids by NAPBQI, a toxic byproduct of Ace metabolism, is an established mechanism of Ace toxicity in both the liver and the kidney [19,20]. Across all compounds, we identified shifts in nitric oxide (NO) synthesis and ROS detoxification, suggesting common mechanisms of cardiotoxicity across compounds.

Finally, using a new approach, we integrated our paired transcriptomics and metabolomics data to identify shifts in active reactions for compounds. Common across all compounds, we found shifts in central carbon metabolism, suggesting unique and divergent pathways for ATP, DNA, RNA or biomarker production. For Dox, we confirmed changes in reactions related to ROS detoxification that were identified with the Tasks Inferred from Differential Expression (TIDEs) approach. For 5FU, we identified unique reactions in fatty acid oxidation and changes in reactions involved in central carbon metabolism and beta oxidation of fatty acids. It was surprising to see that no pathways related to pyrimidine synthesis were identified as either unique to the 5FU condition or different between the 5FU and control condition. However, this would suggest that the 5FU condition model utilized pathways for DNA synthesis similar to the control condition. Taken with the increased gene expression for metabolic tasks related to DNA and RNA synthesis, this suggests that the mechanism of action for 5FU cardiotoxicity is the metabolic stress placed on cells for DNA synthesis. Finally, for Ace, we identify unique reactions for carbohydrate metabolism and changes in reactions related to the pentose phosphate pathway and nucleotide metabolism, suggesting a divergent pathway for nucleotide metabolism or ROS detoxification.

### *5.1.2 General utility of the Tasks Inferred from Differential Expression (TIDEs) approach in identifying shifted metabolic functions from transcriptomics data*

While previous studies have highlighted the utility of pathway-based approaches [21,22] and metabolic tasks [23,24] in GENREs, no studies have used metabolic tasks as a basis for interpreting differentially expressed genes (DEGs). Given the complex gene-protein-reaction (GPR) rules that govern each metabolic reaction, GENREs provide the opportunity to interpret complex changes in gene expression to yield insight into changes in metabolic function in a diseased state. Gene enrichment approaches, e.g. gene set enrichment analysis (GSEA) [25], are already commonly used to interpret broad pathway-based changes in gene expression. These

approaches utilize *a priori* defined lists of genes describing either a specific pathway or general function. In contrast, the approach presented in this dissertation, the Tasks Inferred from Differential Expression (TIDEs) approach [18], utilizes an *a priori* defined list of metabolic reactions for different metabolic functions. In this way, complex changes in gene expression are distilled down to a weight for each reaction involved in a metabolic function. This approach represents a shift from a gene-centric view to a reaction-centric view. In this way, instead of weighting genes as equally contributing to a specific pathway or function, the TIDEs approach weights reactions as equally contributing to a specific pathway or function.

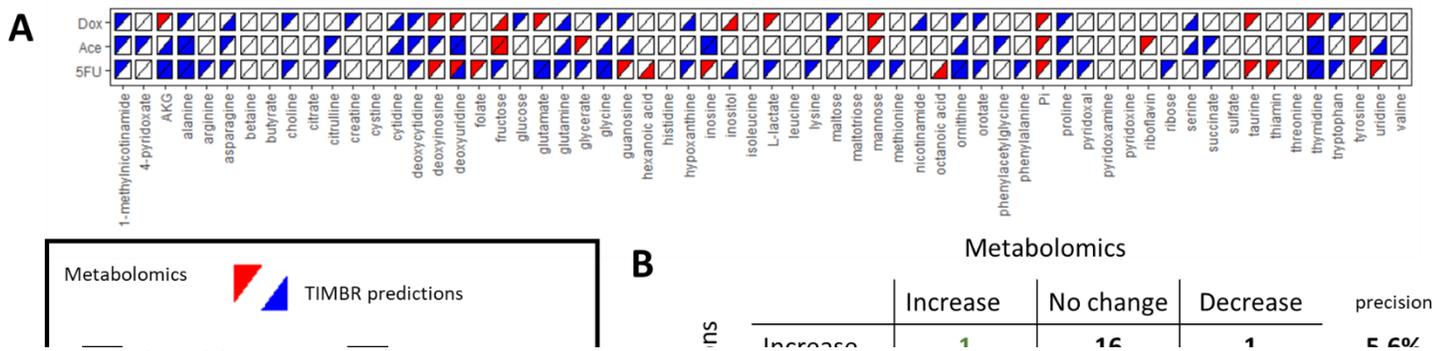
Statistical significance for over-enrichment analyses are calculated using a hypergeometric distribution [26] while GSEA uses a random-walk approach [25]. In a reaction-centric approach, such as TIDEs, neither of these statistical tests are appropriate since the log fold change of one DEG can determine the weight of more than one reaction. Therefore, rather than calculating a statistic directly from the weight of reactions, we compare the average weight across reactions in a metabolic task to a distribution of task scores calculated from randomized gene expression values (Chapter 2). In this way, we can identify metabolic functions that are significantly associated with either increased or decreased gene expression. Similar to an interpretation of gene enrichment results, the results from the TIDEs approach are interpreted as metabolic functions for which gene expression is higher/lower than would be expected given the underlying distribution of the data.

As with other approaches, before integration of DEGs for analysis, a significance threshold is specified, either based on the FDR and/or a log fold change. In the case of the *in vitro* cardiotoxicity data presented here (Chapter 3), there are a large number of DEGs across conditions, complicating both the gene enrichment and TIDEs approaches. Given the rationale that a consistent change in DEG at a lower log fold change is still relevant [25], we instead chose to change the threshold for statistical significance. In the case of gene over-enrichment, we

decreased the significance threshold to define a DEG as an FDR < 0.01 and increased the significance threshold to defined a significantly changed pathway to a BH-adjusted  $p$ -value < 0.1. For GSEA, for a number of different combinations of FDR cutoffs, we were not able to identify any pathways that were significantly changed, highlighting the complexity of using data with a large number of DEGs. For the TIDEs analysis, we also decreased the significance threshold to define a DEG to an FDR < 0.01 and increased the significance threshold to define a statistically changed pathway to a  $p$ -value < 0.1. In this way, we were able to identify both enriched pathways and TIDEs. These results highlight the limitations of current approaches and the value of rationally interpreting  $p$ -values given the data and question.

### 5.1.3 Difficulty in predicting biomarkers using transcriptomics data integrated with metabolic models

Here, we sought to use a previously developed algorithm, the TIMBR algorithm [22], to integrate transcriptomics data characterizing *in vitro* cardiotoxicity to predict new biomarkers that would be validated with the untargeted metabolomics data. The TIMBR algorithm was developed to integrate transcriptomics data in the form of the log fold changes of DEGs with GENREs to predict production of metabolites [22]. When we integrated our collected transcriptomics data with



**Figure 5.1** TIMBR predictions compared to measured changes in metabolites. (A) TIMBR predictions compared to measured changes in the metabolomic data for the 24-hour condition. The boxes in the upper left-hand corner indicate the metabolomics data and the bottom right indicate the TIMBR predictions. A red box indicates a significant change in the metabolomics data (Mann-Whitney, FDR < 0.1) or the TIMBR predictions (absolute value of the production score > 1). (B) Precision and recall for all of the measured metabolomics data across all conditions and timepoints.

the rat-specific heart model to predict changes in metabolites, we had low precision and recall for predicting metabolites that were measured to be significantly increased or decreased in the metabolomics data (Figure 5.1). However, agreement between TIMBR predictions and metabolites measured to be changed does allow for the identification of potential pathways of production. For example, in the case of thymidine production, TIMBR correctly predicted a decrease in the production of thymidine for both Ace and 5FU at 24 hours. In addition, TIMBR provides predictions for all metabolites that the model can produce given the supplied media constraints. Predicted metabolites that were similar across conditions or were associated with scores that were high within a condition are optimal metabolites for more targeted validation *in vitro*. The heterogeneous precision and recall for differentially produced metabolites indicated that another approach for integrating the transcriptomics and metabolomics data may provide more tractable results in identifying potential biomarkers of cardiotoxicity.

Prior to implementing the approach presented in this dissertation (Chapter 3), we attempted other iterations around the central idea of the TIMBR algorithm. Each of these iterations was centered around a different hypothesis for the production of biomarkers. For example, (a) implementing TIMBR while placing constraints based on measured uptake in the metabolomics data (assuming that the metabolomics data would more accurately constrain the solution space), (b) implementing an alternative algorithm, RIPTiDe [27], to predict pathway utilization and pathway weights for biomarker production (assuming that DEGs lose information and RIPTiDe with transcript abundances could more accurately predict production), and (c) implementing RIPTiDe to generate one model producing a minimal amount of ATP while optimizing production of metabolites (assuming that biomarkers are produced in response to a central cell function). None of these iterations improved predictions.

There are multiple reasons why the TIMBR algorithm, and these alternative approaches, did not capture changes in metabolite production. First, the metabolomics data captures mostly

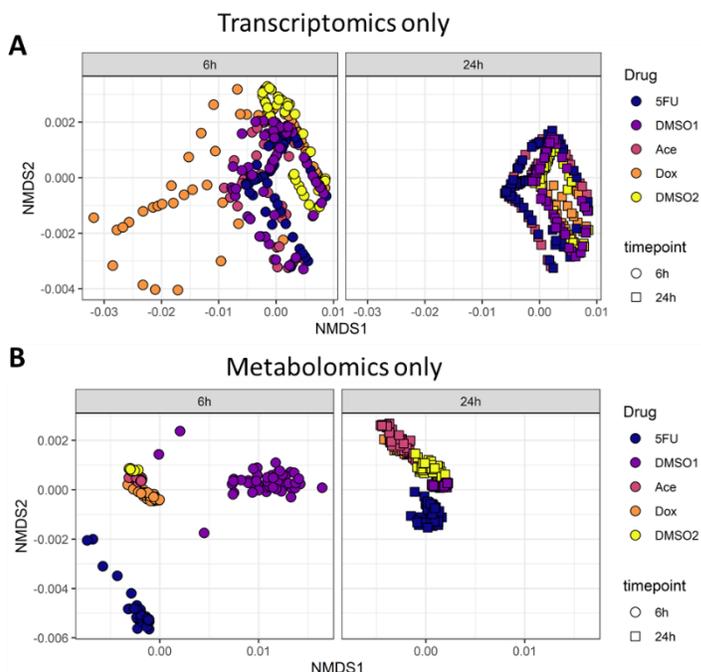
central metabolites that may (a) be subject to higher levels of regulation, both transcriptionally and metabolically, (b) participate in more reactions, or (c) be produced by more than one pathway, making accurate predictions more difficult. Second, the large number of DEGs in the data set could have influenced the accuracy of the results, as we saw with the gene enrichment and TIDEs approaches. With a large number of DEGs and a large range of fold changes, small but concerted changes in individual pathways could fail to reach statistical significance. Finally, the original TIMBR algorithm utilized a well-defined objective function for hepatocyte function to make predictions for biomarkers [22]. Here, a more well-defined objective for the heart may help to improve predictions. Nonetheless, thoroughly probing these hypotheses was outside the scope of this dissertation. Given there are no other approaches to our knowledge that predict metabolite production with GENREs and transcriptomics data, we chose to develop a novel approach that would allow for the integration of our model with our paired transcriptomics and metabolomics data.

Our novel approach was, again, based on the assumption that metabolic biomarkers are produced as a result of changes in the central function of the cell. For the heart, we chose to optimize ATP hydrolysis with a baseline production of DNA and RNA (Chapter 3). Here, we demonstrate the utility of integrating paired data in generating new predictions about reactions unique to specific condition and changes in reaction fluxes between treatment and control conditions, specifically in the case of *in vitro* cardiotoxicity. Through our results, we show the benefit in taking both a metabolic task-based approach through TIDEs and a systems-level approach by building condition-specific models using RIPTiDe. While we are able to identify changes in metabolic functions using metabolic tasks, these functions do not operate independently. The RIPTiDe approach emphasizes that these metabolic functions work together to achieve an end goal in the cell, and the approach highlights the additional insight that is gained by using this systems-level approach.

It is important to note the difficulty in integrating unitless metabolomics data with GENREs. Here, we rationally identified lower bounds for metabolites that were measured to be consumed or produced in a given condition. However, in the case of metabolites that were differentially produced between conditions, e.g. between a treatment and control, we placed no additional bounds due to the difficulty in identifying an appropriate bound for a measured statistical change in production.

#### *5.1.4 RIPTiDe models highlight the utility of integrating both transcriptomics and metabolomics data*

Here, as with previous studies [28,29], we have chosen to characterize *in vitro* cardiotoxicity using paired transcriptomics and metabolomics data. Given the size of the human model, it can be difficult to sufficiently constrain the solution space to identify concrete changes in metabolism. We see this clearly in the difficulty in predicting changes in metabolite production using transcriptomics data alone. Here, integrating the transcriptomics data alone or the metabolomics data alone with the RIPTiDe approach (Figure 5.2) demonstrates unique flux distributions. However, we see clear separation when we integrate both types of data together, highlighting the utility of both data types in constraining the solution space to capture differences in metabolism.



## Figure 5.2 Integrating both metabolomics and transcriptomics data reveals unique metabolism compared to either alone.

(A) NMDS of shared reactions fluxes for the 50 flux samples for each condition after integrating only the transcriptomics data with the defined experimental media. (B) NMDS of shared reaction fluxes for the 50 flux samples for each condition after integrating only the metabolomics data as constraints on production and consumption of metabolites for each condition. In this case, RIPTiDe determines the most parsimonious route of flux.

## 5.2 Future work

### 5.2.1 Experimentally confirming identified shifts in toxicity

Across all compounds, we identified a consistent shift in gene expression for NO synthesis and ROS detoxification, representing common biomarkers of *in vitro* cardiotoxicity. Future work will explore the use of fluorescent markers of ROS production in response to all three compounds. In the case of Ace, given the increased gene expression for phospholipid synthesis and the role of oxidized phospholipids in liver and kidney toxicity, specific markers of lipid peroxidation should be explored. Given that ROS production is an established mechanism for doxorubicin-induced cardiotoxicity, current methods for measuring ROS *in vivo* [30] can be directly translated to the clinic for both doxorubicin and 5-fluoruracil as potential early biomarkers of cardiotoxicity.

Given that the RIPTiDe approach identified divergent reactions for central carbon metabolism, predictions of differences in flux for these pathways can be validated using carbon tracing to confirm the role that central carbon metabolism plays in response to the compounds [31,32]. Further, measurements of internal flux rates can help to further constrain the model to improve predictions. Finally, as has been done with other studies [33], select external flux uptake rates can confirm the carbon sources that the model predicts were utilized but were not measured in the metabolomics data for each context-specific condition. As with carbon tracing, these flux

measures will help to validate the flux distributions predicted with the RIPTiDe models and help to further constrain the solution space for more accurate predictions.

Although we highlight the limitations of the TIMBR algorithm in this context, the metabolites that were predicted to be consistently increased across conditions are potential biomarkers that warrant further exploration. As highlighted previously, the metabolites measured in the metabolomics dataset may be particularly difficult to predict for a number of reasons. However, TIMBR provides predictions for both these central metabolites and more peripheral metabolites. Consistent changes in more peripheral metabolites across treatments are metabolites that should be investigated further. Finally, because TIMBR utilizes transcriptomics data for these predictions, we are able to identify both the pathway predicted for a change in production and the genes driving the reactions present in that pathway. Even for incorrect predictions, the TIMBR approach is a data-driven approach that is reflecting changes in the transcriptomics data that warrant future exploration.

### *5.2.2 Developing a more comprehensive, heart-specific objective function to probe changes in metabolism*

As mentioned above, here we chose to use ATP production with minimal synthesis of RNA and DNA as an objective function. However, this objective function does not capture key features of heart metabolism, such as synthesis of key proteins and maintenance of phospholipids for both the cell membrane and mitochondria. In the case of *Ace in vitro* cardiotoxicity, we identified changes in metabolic functions related to phospholipid synthesis. However, given that phospholipid synthesis was not included in the objective function, the RIPTiDe models do not capture differences in this function in response to the compounds. An improved objective function could also help to improve some of the TIMBR predictions, assuming that biomarkers are produced as a result of general cell function.

In addition, future work should explore utilizing compound-specific objective functions. In the case of Dox, given that excessive ROS production is a mechanism of cardiotoxicity, ROS should be forced into the models to study the overall effect on metabolism. In the case of 5FU, the models should be required to synthesize either more DNA or RNA, or more specifically, UTP. Integration of these compound-specific objective functions would be evaluated using the Spearman correlation between reaction fluxes and transcript abundances produced by the RIPTiDe algorithm [27], where an increased correlation would signify that the given objective better captures the underlying biology. Recent updates to the RIPTiDe package employ a brute-force approach in identifying the fraction of the objective value that best fit the data, and are therefore suited to answer these questions about the relationship between RNA or DNA synthesis and ROS production as mechanisms of toxicity.

### *5.2.3 Integrating human-specific data to identify shared biomarkers of doxorubicin-induced cardiotoxicity*

As has been highlighted previously (Chapter 4), there are limitations to translating biological discoveries between *in vitro* and *in vivo* studies as well as between model organisms and humans. The paired nature of the human and rat-specific heart models presented in this dissertation provide the opportunity to identify shared biomarkers of cardiotoxicity. Published transcriptomics data is available for human-induced pluripotent stem cell cardiomyocytes (hiPSC-CMs) treated with doxorubicin [34] or trastuzumab [35], two compounds associated with cardiotoxicity [36,37]. While these studies only collected transcriptomics data, the presented approach can still help to highlight metabolic differences that are consistent with the *in vitro* rat cardiotoxicity data presented here. Shared metabolic reactions identified using the TIDES approach or unique metabolic reactions identified using the RIPTiDe approach would indicate

specific transcripts, metabolites, or metabolic functions that warrant future study as translational biomarkers.

#### 5.2.4 *Developing pathway tracing algorithms*

Here, we demonstrate the utility of integrating paired transcriptomics and metabolomics data to identify shifts in active reactions between treated and control conditions. However, the current state of the field is limited in tracing how changes in flux through a particular reaction propagate through the network. For example, we identify changes in active flux through reactions related to central carbon metabolism for all of our treatment groups. However, we cannot, at this point, determine the source or effect of that change in flux, i.e. changes in carbon source utilization, changes in pathways for nucleotide metabolism, or changes due to biomarker production. Future work should explore developing and using pathway tracing algorithms to identify what is influencing changes in one reaction in the network. Previous work has been done in developing pathway tracing algorithms [38,39]. However, none of these approaches have publicly available code.

#### 5.2.5 *Implementing a transcript-based TIDEs approach*

The presented TIDEs approach uses an *a priori* defined list of reactions for each metabolic task [18]. For some metabolic tasks, such as synthesis of NO from arginine, there is only one pathway available in the model to complete the task. However, for others, such as synthesis of DNA, there are multiple parallel pathways available. In the current state, the TIDEs approach cannot capture these different pathways for completion of a metabolic task. However, the RIPTiDe algorithm provides the opportunity to integrate transcript abundances to identify the most feasible path of flux to meet these individual metabolic tasks. Future work should explore integrating the TIDEs approach with RIPTiDe to identify the most feasible path of flux for specific metabolic tasks. While many functions for implementing analyses with GENREs are available in both MATLAB

and Python, the functions for testing metabolic tasks are currently only available in MATLAB. Therefore, these functions will need to be extended and re-written in Python in order to utilize the TIDEs approach with the RIPTiDe algorithm. Further, integration of transcript abundances would complicate the current approach, which utilizes the average weight across reactions in a metabolic task to calculate a task score. Future work would need to explore statistical approaches for determining differences in metabolic tasks based on transcript abundances rather than log fold changes within a metabolic task.

#### 5.2.6 *Future manual curation of the human and rat models using collected paired data*

The paired transcriptomics and metabolomics data presented here adds to other paired data that has been collected to characterize *in vitro* hepatotoxicity [28] and nephrotoxicity [29]. Together, this paired data can enable extensive manual curation of the current general human reconstructions. For example, in the presented paired integration of the transcriptomics and metabolomics data (Chapter 3), a number of metabolites that were measured to be either produced or consumed in the metabolomics were either not included in the general reconstructions or could not be produced or consumed in the heart-specific model given the constraints. These discrepancies provide opportunities for future curation to both the general models of human metabolism and the heart-specific models.

### 5.3 **Conclusion**

In conclusion, here we present new paired data characterizing *in vitro* cardiotoxicity and new methods for integrating experimental data with GENREs. We present new hypotheses for

mechanisms of compound-induced cardiotoxicity based on our collected data integrated with GENREs. These model generated hypotheses demonstrate the utility of GENREs in identifying metabolic changes in the heart and provide direction for future experimental validation. The methods presented here can be extended beyond the context of the heart and applied to other tissue-specific models and diseased states in humans as well as with microbial reconstructions. Finally, the identified biomarkers can, hopefully, eventually lead to clinical biomarkers to diagnose and prevent cardiotoxicity in the clinic.

## 5.4 References

1. Edwards JS, Palsson BO. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J Biol Chem*. 1999;274: 17410–17416. doi:10.1074/jbc.274.25.17410
2. Förster J, Famili I, Fu P, Palsson BØ, Nielsen J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res*. 2003;13: 244–253. doi:10.1101/gr.234503
3. Majewski RA, Domach MM. Simple constrained-optimization view of acetate overflow in *E. coli*. *Biotechnol Bioeng*. 1990;35: 732–738. doi:10.1002/bit.260350711
4. Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, et al. A community-driven global reconstruction of human metabolism. *Nat Biotechnol*. 2013;31: 419–425. doi:10.1038/nbt.2488
5. Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M, Nielsen J. Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat Commun*. 2014;5: 3083. doi:10.1038/ncomms4083
6. Robinson JL, Kocabaş P, Wang H, Cholley P-E, Cook D, Nilsson A, et al. An atlas of human metabolism. *Sci Signal*. 2020;13. doi:10.1126/scisignal.aaz1482
7. Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol*. 2018;36: 272–281. doi:10.1038/nbt.4072

8. Yizhak K, Chaneton B, Gottlieb E, Ruppin E. Modeling cancer metabolism on a genome scale. *Mol Syst Biol.* 2015;11: 817.
9. Ghaffari P, Mardinoglu A, Asplund A, Shoaie S, Kampf C, Uhlen M, et al. Identifying anti-growth factors for human cancer cell lines through genome-scale metabolic modeling. *Sci Rep.* 2015;5: 8183. doi:10.1038/srep08183
10. Stempler S, Yizhak K, Ruppin E. Integrating transcriptomics with metabolic modeling predicts biomarkers and drug targets for Alzheimer's disease. *PLoS One.* 2014;9: e105383. doi:10.1371/journal.pone.0105383
11. Chang RL, Xie L, Xie L, Bourne PE, Palsson BØ. Drug Off-Target Effects Predicted Using Structural Analysis in the Context of a Metabolic Network Model. *PLoS Comput Biol.* 2010;6: e1000938. doi:10.1371/journal.pcbi.1000938
12. Dougherty BV, Moutinho TJ, Papin J. Accelerating the Drug Development Pipeline with Genome-Scale Metabolic Network Reconstructions. *Systems Biology.* John Wiley & Sons, Ltd; 2017. pp. 139–162. doi:10.1002/9783527696130.ch5
13. Rawls K, Dougherty BV, Papin J. Metabolic Network Reconstructions to Predict Drug Targets and Off-Target Effects. In: Nagrath D, editor. *Metabolic Flux Analysis in Eukaryotic Cells: Methods and Protocols.* New York, NY: Springer US; 2020. pp. 315–330. doi:10.1007/978-1-0716-0159-4\_14
14. Rawls KD, Dougherty BV, Blais EM, Stancliffe E, Kolling GL, Vinnakota K, et al. A simplified metabolic network reconstruction to promote understanding and development of flux balance analysis tools. *Comput Biol Med.* 2019;105: 64–71. doi:10.1016/j.combiomed.2018.12.010
15. Karlstädt A, Fliegner D, Kararigas G, Ruderisch HS, Regitz-Zagrosek V, Holzhütter H-G. CardioNet: A human metabolic network suited for the study of cardiomyocyte metabolism. *BMC Syst Biol.* 2012;6: 114. doi:10.1186/1752-0509-6-114
16. Zhao Y, Huang J. Reconstruction and analysis of human heart-specific metabolic network based on transcriptome and proteome data. *Biochem Biophys Res Commun.* 2011;415: 450–454. doi:10.1016/j.bbrc.2011.10.090
17. Uhlen M, Hallström BM, Lindskog C, Mardinoglu A, Ponten F, Nielsen J. Transcriptomics resources of human tissues and organs. *Mol Syst Biol.* 2016;12: 862–862. doi:10.15252/msb.20155865
18. Dougherty BV, Rawls KD, Kolling GL, Vinnakota KC, Wallqvist A, Papin JA. Identifying functional metabolic shifts in heart failure with the integration of omics data and a cardiomyocyte-specific, genome-scale model. *bioRxiv.* 2020; 2020.07.20.212274. doi:10.1101/2020.07.20.212274

19. Du K, Ramachandran A, Jaeschke H. Oxidative stress during acetaminophen hepatotoxicity: Sources, pathophysiological role and therapeutic potential. *Redox Biol.* 2016;10: 148–156. doi:10.1016/j.redox.2016.10.001
20. Mazer M, Perrone J. Acetaminophen-induced nephrotoxicity: pathophysiology, clinical manifestations, and management. *J Med Toxicol Off J Am Coll Med Toxicol.* 2008;4: 2–6. doi:10.1007/bf03160941
21. Zielinski DC, Filipp FV, Bordbar A, Jensen K, Smith JW, Herrgard MJ, et al. Pharmacogenomic and clinical data link non-pharmacokinetic metabolic dysregulation to drug side effect pathogenesis. *Nat Commun.* 2015;6: 7101. doi:10.1038/ncomms8101
22. Blais EM, Rawls KD, Dougherty BV, Li ZI, Kolling GL, Ye P, et al. Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nat Commun.* 2017;8: 14250. doi:10.1038/ncomms14250
23. Swainston N, Smallbone K, Hefzi H, Dobson PD, Brewer J, Hanscho M, et al. Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics.* 2016;12: 109. doi:10.1007/s11306-016-1051-4
24. Richelle A, Chiang AWT, Kuo C-C, Lewis NE. Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. *PLoS Comput Biol.* 2019;15: e1006867. doi:10.1371/journal.pcbi.1006867
25. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102: 15545–15550. doi:10.1073/pnas.0506580102
26. Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, et al. GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics.* 2004;20: 3710–3715. doi:10.1093/bioinformatics/bth456
27. Jenior ML, Moutinho TJ, Dougherty BV, Papin JA. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. *PLoS Comput Biol.* 2020;16. doi:10.1371/journal.pcbi.1007099
28. Rawls KD, Blais EM, Dougherty BV, Vinnakota KC, Pannala VR, Wallqvist A, et al. Genome-Scale Characterization of Toxicity-Induced Metabolic Alterations in Primary Hepatocytes. *Toxicol Sci Off J Soc Toxicol.* 2019;172: 279–291. doi:10.1093/toxsci/kfz197
29. Rawls KD, Dougherty BV, Vinnakota KC, Pannala VR, Wallqvist A, Kolling GL, et al. Genome-Scale Metabolic Model Predicts Changes in Renal Metabolism from Chemical Exposure. In prep.

30. Shah SA, Cui SX, Waters CD, Sano S, Wang Y, Doviak H, et al. Nitroxide-enhanced MRI of cardiovascular oxidative stress. *NMR Biomed.* 2020;33: e4359. doi:10.1002/nbm.4359
31. Pannala VR, Wall ML, Estes SK, Trenary I, O'Brien TP, Printz RL, et al. Metabolic network-based predictions of toxicant-induced metabolite changes in the laboratory rat. *Sci Rep.* 2018;8: 11678. doi:10.1038/s41598-018-30149-7
32. Pannala VR, Vinnakota KC, Rawls KD, Estes SK, O'Brien TP, Printz RL, et al. Mechanistic identification of biofluid metabolite changes as markers of acetaminophen-induced liver toxicity in rats. *Toxicol Appl Pharmacol.* 2019;372: 19–32. doi:10.1016/j.taap.2019.04.001
33. Ramirez AK, Lynes MD, Shamsi F, Xue R, Tseng Y-H, Kahn CR, et al. Integrating Extracellular Flux Measurements and Genome-Scale Modeling Reveals Differences between Brown and White Adipocytes. *Cell Rep.* 2017;21: 3040–3048. doi:10.1016/j.celrep.2017.11.065
34. Burridge PW, Li YF, Matsa E, Wu H, Ong S, Sharma A, et al. Human Induced Pluripotent Stem Cell–Derived Cardiomyocytes Recapitulate the Predilection of Breast Cancer Patients to Doxorubicin–Induced Cardiotoxicity. *Nat Med.* 2016;22: 547–556. doi:10.1038/nm.4087
35. Kitani Tomoya, Ong Sang-Ging, Lam Chi Keung, Rhee June-Wha, Zhang Joe Z., Oikonomopoulos Angelos, et al. Human-Induced Pluripotent Stem Cell Model of Trastuzumab-Induced Cardiac Dysfunction in Patients With Breast Cancer. *Circulation.* 2019;139: 2451–2465. doi:10.1161/CIRCULATIONAHA.118.037357
36. Volkova M, Russell R. Anthracycline Cardiotoxicity: Prevalence, Pathogenesis and Treatment. *Curr Cardiol Rev.* 2011;7: 214–220. doi:10.2174/157340311799960645
37. Mohan N, Jiang J, Dokmanovic M, Wu WJ. Trastuzumab-mediated cardiotoxicity: current understanding, challenges, and frontiers. *Antib Ther.* 2018;1: 13–17. doi:10.1093/abt/tby003
38. Tervo CJ, Reed JL. MapMaker and PathTracer for tracking carbon in genome-scale metabolic models. *Biotechnol J.* 2016; n/a-n/a. doi:10.1002/biot.201500267
39. Huang Y, Zhong C, Lin HX, Wang J. A Method for Finding Metabolic Pathways Using Atomic Group Tracking. *PLoS ONE.* 2017;12. doi:10.1371/journal.pone.0168725