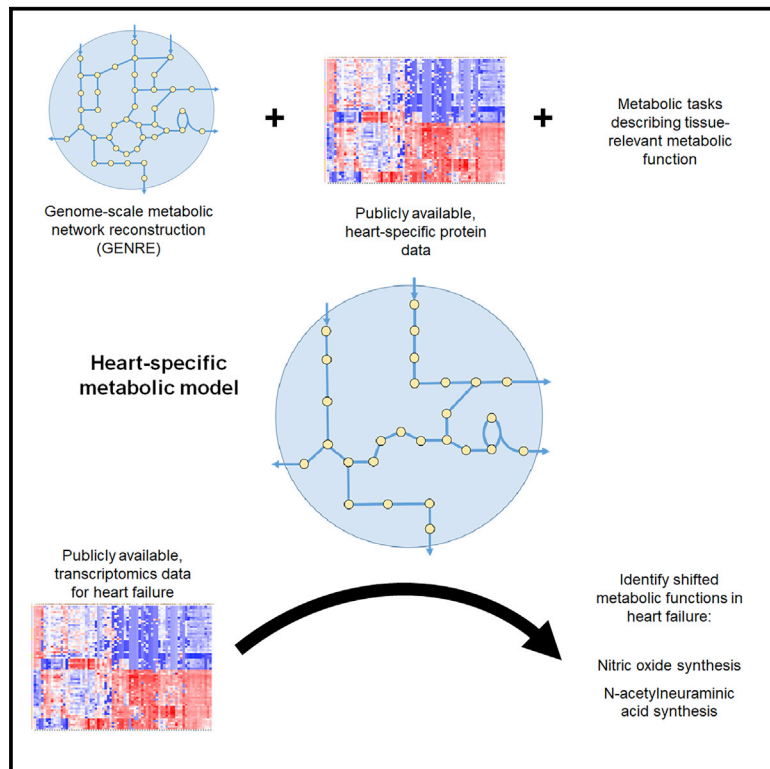


Identifying functional metabolic shifts in heart failure with the integration of omics data and a heart-specific, genome-scale model

Graphical Abstract



Authors

Bonnie V. Dougherty,
 Kristopher D. Rawls, Glynis L. Kolling,
 Kalyan C. Vinnakota, Anders Wallqvist,
 Jason A. Papin

Correspondence

papin@virginia.edu

In brief

Dougherty et al. build a heart-specific metabolic model by using tissue-specific data and broad metabolic functions in the heart. Using that model, they integrate publicly available transcriptomics data for human heart failure to identify conserved transcriptional changes in the synthesis of nitric oxide and *N*-acetylneuraminic acid across studies.

Highlights

- Development of a heart-specific metabolic model
- Interpretation of gene expression data using metabolic functions and metabolic models
- Identification of common changes in metabolic functions in late-stage heart failure



Article

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

Bonnie V. Dougherty,¹ Kristopher D. Rawls,¹ Glynis L. Kolling,^{1,2} Kalyan C. Vinnakota,^{3,4} Anders Wallqvist,³ and Jason A. Papin^{1,2,5,6,*}

¹Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22908, USA

²Department of Medicine, Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, VA 22908, USA

³Department 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, MD 21702, USA

⁴The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, MD 20817, USA

⁵Department of Biochemistry & Molecular Genetics, University of Virginia, Charlottesville, VA 22908, USA

⁶Lead contact

*Correspondence: papin@virginia.edu

<https://doi.org/10.1016/j.celrep.2021.108836>

SUMMARY

In diseased states, the heart can shift to use different carbon substrates, measured through changes in uptake of metabolites by imaging methods or blood metabolomics. However, it is not known whether these measured changes are a result of transcriptional changes or external factors. Here, we explore transcriptional changes in late-stage heart failure using publicly available data integrated with a model of heart metabolism. First, we present a heart-specific genome-scale metabolic network reconstruction (GENRE), *iCardio*. Next, we demonstrate the utility of *iCardio* in interpreting heart failure gene expression data by identifying tasks inferred from differential expression (TIDEs), which represent metabolic functions associated with changes in gene expression. We identify decreased gene expression for nitric oxide (NO) and *N*-acetylneuraminic acid (Neu5Ac) synthesis as common metabolic markers of heart failure. The methods presented here for constructing a tissue-specific model and identifying TIDEs can be extended to multiple tissues and diseases of interest.

INTRODUCTION

The heart maintains its functions by using oxygen and various nutrients through metabolic pathways to synthesize contractile proteins, generate necessary lipid species, and produce ATP as a fuel for muscle contraction. Metabolic changes, such as changes in substrate utilization, have been measured in many diseased states, such as left ventricular hypertrophy (Kundu et al., 2015) and cardiotoxicity (Bauckneht et al., 2017; Borde et al., 2012). In some cases, changes in substrate utilization occur before functional and/or structural changes to the heart, suggesting that metabolism has a key role in the downstream development of disease or could be a target to prevent disease (Kundu et al., 2015; Li et al., 2019). However, it is not clear whether measured changes in substrate utilization are driven by changes in the transcriptome or other factors. Further, consistent transcriptional changes in metabolic pathways could point toward potential contributors to, or biomarkers of, heart failure. Therefore, there is a need for a comprehensive, descriptive model of the metabolic function of the heart to interrogate the relationships among substrates, gene expression, and metabolic functions.

A common tool to interrogate relationships among metabolic substrates, gene expression, and metabolic functions is a genome-scale metabolic network reconstruction (GENRE). GENREs are mathematical representations of metabolism, which 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. Further, the complex GPR rules associated with each reaction allow for the systematic integration of different types of data, such as transcriptomics or proteomics data.

However, not all genes are expressed in every tissue, necessitating the construction of tissue-specific models of metabolism. GENREs of human metabolism (Blais et al., 2017; Brunk et al., 2018; Duarte et al., 2007; Ma et al., 2007; Mardinoglu et al., 2013; Swainston et al., 2016; Thiele et al., 2013) have been used to generate tissue-, disease-, or cell-specific models for various analyses, such as predicting drug targets (Agren et al.,



2014; Folger et al., 2011; Rawls et al., 2020), identifying biomarkers of disease (Zhang et al., 2013; Zhao and Huang, 2011), and understanding drug toxicity or side effects (Blais et al., 2017; Shaked et al., 2016; Zielinski et al., 2015). Tissue-specific models are built by integrating tissue-specific omics data, usually transcriptomic or proteomic data, with a human GENRE using various integration algorithms (some of which are summarized in Blazier and Papin (2012) (Robaina Estévez and Nikoloski, 2014) to obtain a tissue- or cell-type-specific model. To date, there are two existing heart models (Karlstädt et al., 2012; Zhao and Huang, 2011), both of which were built from Recon1, the first human GENRE (Duarte et al., 2007). These models were used to examine the relationship between substrate use and the efficiency of the heart (Karlstädt et al., 2012) and to predict epistatic interactions in the heart and biomarkers of heart disease (Zhao and Huang, 2011). However, human models have been expanded since Recon1 to more comprehensively describe human metabolism. A recently published human GENRE, *iHsa* (Blais et al., 2017), is more comprehensive than Recon1 (8,336 reactions versus 3,311 reactions) and, therefore, offers the potential to generate a more comprehensive heart-specific model. An expanded, heart-specific model can be used as a tool to interpret large datasets, such as transcriptomic data, to provide functional insight into how metabolism might change in a diseased state.

Here, we present a validated, heart-specific metabolic model, *iCardio*, which was built using *iHsa* (Blais et al., 2017) with tissue-specific protein data from the Human Protein Atlas (HPA) (version 18; <https://www.proteinatlas.org>; Uhlén et al., 2015). 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 (Blais et al., 2017) and curated, heart-relevant tasks. We validated the model qualitatively by the number of reactions covered by the metabolic tasks and quantitatively by ATP predictions for common carbon sources.

Finally, we demonstrate the utility of metabolic models in identifying functional changes in metabolism using an approach that identifies metabolic tasks associated with significant changes in gene expression, called tasks inferred from differential expression (TIDEs). The TIDEs approach represents a reaction-centric view in contrast to standard gene-centric views, which are taken with approaches such as standard gene enrichment or gene set enrichment analysis (GSEA). This reaction-centric view accounts for both the stoichiometric balance of reactions necessary to achieve metabolic functions as well as the complex relationships among genes, proteins, and the reactions they catalyze. Therefore, the TIDEs approach offers biological insight into metabolic functions affected in a disease state that are not readily apparent from gene expression data or GSEA alone. Here, we demonstrate the utility of the TIDEs approach using heart failure as a case study because of the critical role that metabolism has in the progression and diagnosis of disease. In contrast to previous studies which measure changes in the uptake of metabolic substrates in the heart in diseased states (Bauckneht et al., 2017; Borde et al., 2012; Kundu et al., 2015; Murashige et al., 2020; Taylor et al., 2001), here, we identify metabolic functions that are associated with significant changes in transcription across

previously published transcriptomics heart-failure datasets. Using publicly available data for ischemic and idiopathic heart failure, we demonstrate that both gene expression and metabolic functions cluster based on studies rather than origin of heart failure. Across the different studies and types of heart failure, we identified two common metabolic functions associated with changes in gene expression, the synthesis of nitric oxide (NO) and the synthesis of *N*-acetylneuraminic acid (Neu5Ac), a sialic acid. Further, we demonstrate that both NO and Neu5Ac synthesis are consistently associated with decreased gene expression across other independent, previously published datasets for general heart failure, including both microarray and RNA sequencing (RNA-seq). Together, this suggests that both NO and Neu5Ac synthesis are robust metabolic functions associated with changes in gene expression in heart failure across studies and can serve as either biomarkers of heart failure or potential targets for the treatment of heart failure.

RESULTS

Building and validating *iCardio* using metabolic tasks

Before generating a heart-specific GENRE, we turned to the literature to identify metabolic functions that are known to be important for heart metabolism. These identified metabolic functions were formulated as metabolic tasks and added to the previously published metabolic task list (Blais et al., 2017), which, together, served as a benchmark for model function. Although multiple algorithms exist for integrating tissue-specific data with metabolic models, we selected the cost-optimization reaction dependency assessment (CORDA) algorithm because it generated the best-performing draft model for heart-specific task accuracy (method details).

The CORDA algorithm was used to build the draft *iCardio* model from *iHsa* by integrating tissue-specific protein data from the HPA (method details); CORDA removed 4,203 reactions from *iHsa* to build a draft *iCardio* model that had a heart-specific task accuracy of 89% (Figures 1B and 1C; Data S2). The draft model was then curated using metabolic tasks to obtain the final *iCardio* model (Figure 1). To achieve 100% task accuracy, 79 reactions were added to, and 12 removed from, the draft *iCardio* model (Figure 1A), based on literature and manual curation (Data S2). The 79 reactions that were added to *iCardio* are reactions that were removed from *iHsa* to build the *iCardio* draft model and 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 NO was a metabolic task that originally failed but was curated to pass in the final *iCardio* model. For that task to pass, three reactions were added back to the model, including two reactions that were associated with proteins that were not detected and one that had no GPR rule (Data S2). All three isoforms of NO synthase (NOS), which catalyze the two reactions with GPR rules, were not detected in version 18 of the HPA tissue data, but two of the NOS isoforms, NOS1 and NOS2, are associated with low protein evidence in version 19,

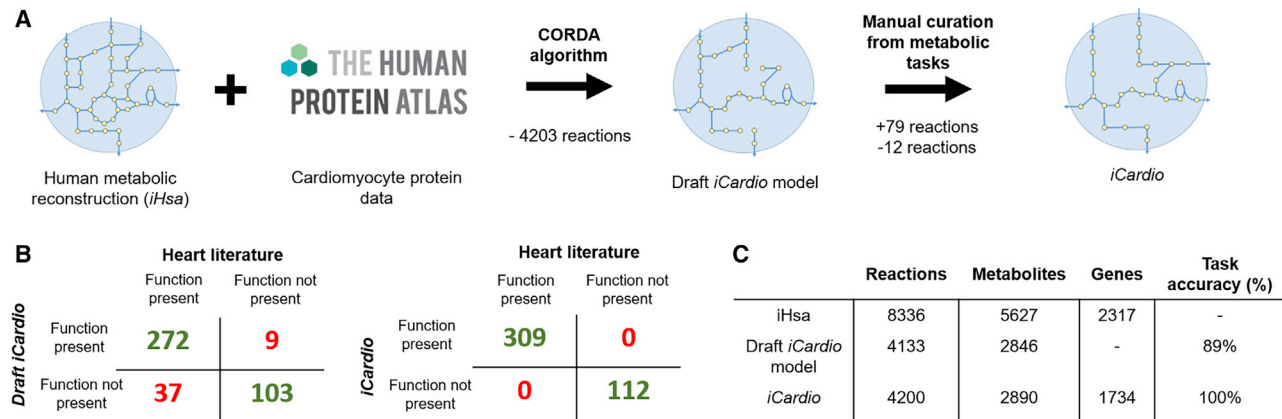


Figure 1. Building a cardiomyocyte metabolic model (*iCardio*) by integrating protein data and curating with metabolic tasks

(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 superset of 421 metabolic tasks represented metabolic tasks from a previous publication (Blais et al., 2017) with additional metabolic tasks that were curated from the literature based on heart metabolism. These tasks were used to test whether the model could perform functions known to be present in the heart and remove metabolic functions not present in the heart.

(C) The CORDA algorithm removed 4,203 reactions from *iHsa* to produce a draft *iCardio* model that had a cardiomyocyte-specific task accuracy of 89%. Further curation, based on the previously defined list of metabolic tasks, resulted in a final model with 100% heart-specific task accuracy.

whereas NOS3 has significantly higher transcript counts in heart muscle in the RNA consensus tissue HPA data (Uhlén et al., 2015).

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 1C).

Because 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. Similar to what has been done with other models (Richelle et al., 2019), parsimonious flux balance analysis (pFBA) was used to determine the reactions used in *iCardio* for each passing metabolic task. The 216 previously published heart-relevant tasks (Blais et al., 2017) covered 38% (1,593/4,200) of reactions in *iCardio* and the 93 passing heart-relevant tasks covered 21% (874/4,200) of reactions in *iCardio* (Figure 2A). There is overlap between these two sets of reactions; in total, the two sets of tasks covered 41% (1,714/4,200) 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, more than 20,000 reactions are used in different combinations to complete the metabolic tasks, indicating that a number of reactions are repeated among 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 for 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 reactions. 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* (Smith et al., 2017). 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* (Smith et al., 2017) (Figure 2B). For methionine, the ATP prediction from *iCardio* matches the methionine prediction from Recon 2.2 (Swainston et al., 2016). It is important to note the difference in scope between the two models: *MitoCore* contains 342 reactions, whereas *iCardio* contains 4,200 reactions and still maintains 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 validations of *iCardio* demonstrate the ability of the model to accurately and more comprehensively represent heart metabolism.

Identifying 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 (DEGs), called TIDEs, which represent metabolic functions that are associated

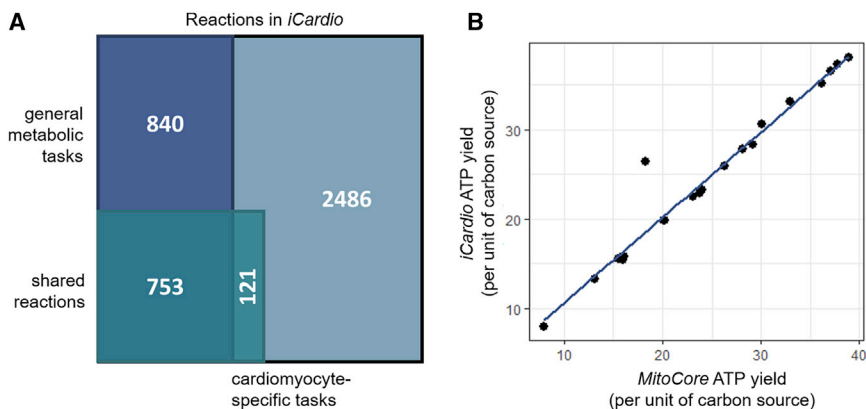


Figure 2. Validating *iCardio* descriptively using the number of reactions covered by metabolic tasks and quantitatively using ATP yields for common carbon substrates

(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 with 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*.

with significant shifts in gene expression in an experimental versus control state (Figure 3). Heart failure is a complex disease, both in etiology and presentation, but heart failure is associated with a shift in metabolism (Wende et al., 2017). Several different metabolic shifts have been measured in heart failure, such as decreased fatty acid utilization (Lopaschuk, 2017; Wende et al., 2017), increased ketone body utilization (Janardhan et al., 2011), and decreased utilization of branched-chain amino acids (Sun et al., 2016). However, there has been limited systematic analysis of the transcriptional metabolic changes that could be influencing these measured changes in metabolic uptake in heart failure. *iCardio* provides an opportunity to contextualize gene expression data from patients with end-stage heart failure to identify transcriptional changes in metabolic functions, such as those listed above, and to identify metabolic functions associated with changes in transcription across studies.

We identified DEGs from microarray data for samples collected from non-infarcted regions of the left ventricle for patients undergoing heart transplants for either ischemic or idiopathic heart failure compared with samples from healthy hearts. We integrated these DEGs and their associated log-fold changes with *iCardio* to determine TIDEs, representing metabolic functions associated with a significant change in gene expression. For example, when compared with healthy hearts, the ischemic hearts from GEO: GSE5406 (Hannenhalli et al., 2006) had 2,678 DEGs; 392 of these DEGs were represented in *iCardio* (Table S3). After integrating these DEGs using *iCardio* and the TIDEs pipeline, 85 of the 307 metabolic tasks were designated as TIDEs (Figure 4; Data S3), representing metabolic functions associated with significant changes in gene expression for this ischemic heart-failure dataset. In the randomized data used to determine the statistical significance for that dataset, only 13 of the 1,000 random iterations had more than 85 tasks identified as significantly changed, indicating that the identified TIDEs cannot be attributed to random changes in gene expression, but rather, to distinct and coordinated shifts in gene expression resulting in changes in specific metabolic functions. Across the 307 metabolic tasks, there are varied distributions in the underlying randomized task scores (Figure 4), representing the complex relationships between

gene expression and metabolic function that are captured with *iCardio* and the TIDEs reaction-centric approach.

Some general trends appear from the TIDEs identified from this ischemic heart-failure dataset (Figure 4). First, metabolic tasks related to fatty acid synthesis (39 tasks) were decreased. Within tasks related to fatty acid synthesis, significant decreases were observed for synthesis of saturated fatty acids but not for synthesis of unsaturated fatty acids. Metabolic tasks related to lipid synthesis were decreased (6 tasks). Finally, metabolic tasks related to signaling metabolism, the synthesis of NO and Neu5Ac, were associated with decreased gene expression. Taken together, these TIDEs support a potential role for decreased gene expression for fatty acid synthesis and lipid synthesis as well as decreased synthesis of NO and Neu5Ac for signaling in heart failure. However, other, previously reported, metabolic signatures of heart failure, such as increased ketone body degradation (Janardhan et al., 2011) and decreased breakdown of branched-chain amino acids (Sun et al., 2016), were not associated with significant changes in transcription for this dataset.

To determine common shifted metabolic functions across datasets and types of heart failure (ischemic versus idiopathic), we expanded the TIDE analysis to two additional studies (Table S3), which included samples for idiopathic and ischemic heart failure. Although we examined more than these three datasets, these were the only datasets that had publicly available data and at least 50 DEGs (Table S4). First, we see clustering of both the DEGs (Figure S1A) and the TIDEs (Figure S1B; Data S4) within each dataset, rather than across the types of heart failure, highlighting the complexity and heterogeneity of the disease. Although we see no common TIDE appears across datasets (Figure S2A), there are some TIDEs that appear in at least four of the six datasets (Figure 5A).

First, *de novo* synthesis of Neu5Ac was associated with decreased gene expression in five of the six datasets (excluding GEO: GSE1869 ischemic). Second, both the synthesis of NO and the synthesis of NAD from tryptophan were associated with decreased gene expression in four of the six datasets (excluding ischemic and idiopathic GEO: GSE1869). Decreased synthesis of NO, NAD, and Neu5Ac represent potential biomarkers of heart

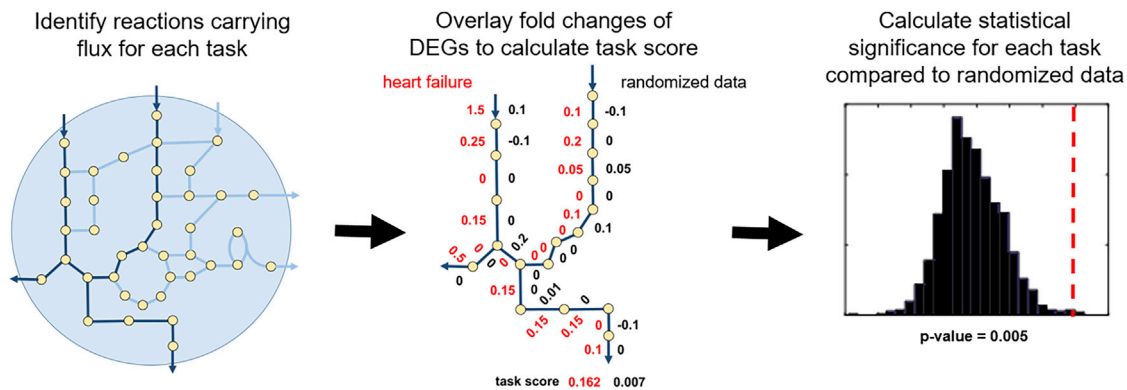


Figure 3. Identifying functional metabolic changes from gene expression data using metabolic tasks

Metabolic tasks quantitatively describe metabolic functions that a tissue or organism is known to catalyze. *iCardio* was used to identify the reactions that are used 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 used in that metabolic task. 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.

failure that are shared across types of heart failure (ischemic versus idiopathic) and datasets. To confirm the role of these metabolic functions in heart failure, four additional datasets (Greco et al., 2012; Ren et al., 2020; Schiano et al., 2017) (Table S5) for general heart failure were analyzed using the TIDEs pipeline (Figure S3; Data S5). Both the synthesis of NO and Neu5Ac are significantly associated with decreased gene expression changes in three of the four heart failure datasets, whereas the synthesis of NAD from tryptophan is significantly associated with decreased gene expression in two of the four datasets.

As mentioned previously, several metabolic shifts are associated with heart failure: decreased fatty acid utilization (Lopaschuk, 2017; Murashige et al., 2020; Wende et al., 2017), increased ketone body utilization (Janardhan et al., 2011; Murashige et al., 2020), and decreased utilization of branched-chain amino acids (Sun et al., 2016). However, it is unclear whether these changes are driven by external factors (i.e., changes in hormones or circulating levels of metabolites) or transcriptional changes. Here, we see decreased synthesis of fatty acids in four of the six datasets (Figure 5A), suggesting downregulation of fatty acid synthesis, rather than consistent increased gene expression for fatty acid oxidation (Figure 5B). For ketone body metabolism, branched-chain amino acid metabolism, and fatty acid oxidation, we see no consistent changes in gene expression for these metabolic functions across the datasets. Finally, for glucose metabolism, we see decreased gene expression for the breakdown of glucose in the presence of oxygen for three of six datasets, suggesting decreased use of glucose oxidation (Figure 5B). The lack of transcriptional changes for ketone body metabolism, branched-chain amino acid metabolism, and fatty acid metabolism are confirmed in the general heart-failure dataset (Figure S3; Data S5). Additionally, decreased gene expression for the metabolic function of the breakdown of glucose in the presence of oxygen was seen in three of the four heart-failure datasets, confirming the role of decreased gene expression for glucose oxidation in heart failure (Figure S3B).

Comparison of TIDEs with GSEA by KEGG pathway

Next, we wanted to compare the TIDEs analysis to a common gene-centric approach, GSEA, using the Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways to define the gene sets. Although KEGG includes pathways other than metabolism (Figure S4), we have chosen to focus on a comparison of metabolic pathways. Using GSEA with the data from the studies cited above, the most commonly changed metabolic pathways across the datasets were increased beta alanine metabolism in all datasets, increased propanoate metabolism in five of the six datasets, decreased selenoamino acid metabolism in four of the six datasets and increased valine, isoleucine, and leucine degradation in three of the six datasets (Figure S2B). For the commonly cited changes in uptake of metabolites 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 three 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 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, and only one task, breakdown of valine, was significantly changed in one dataset.

The difference in identified changes in metabolism highlights 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, whereas the TIDEs approach allows for a distinction between the synthesis and degradation of different fatty acids. We can use *iCardio* for a similar general metabolic approach by using the metabolic subsystems assigned to each reaction to identify sets of reactions. Using TIDEs, we can determine whether the reactions in a given subsystem are

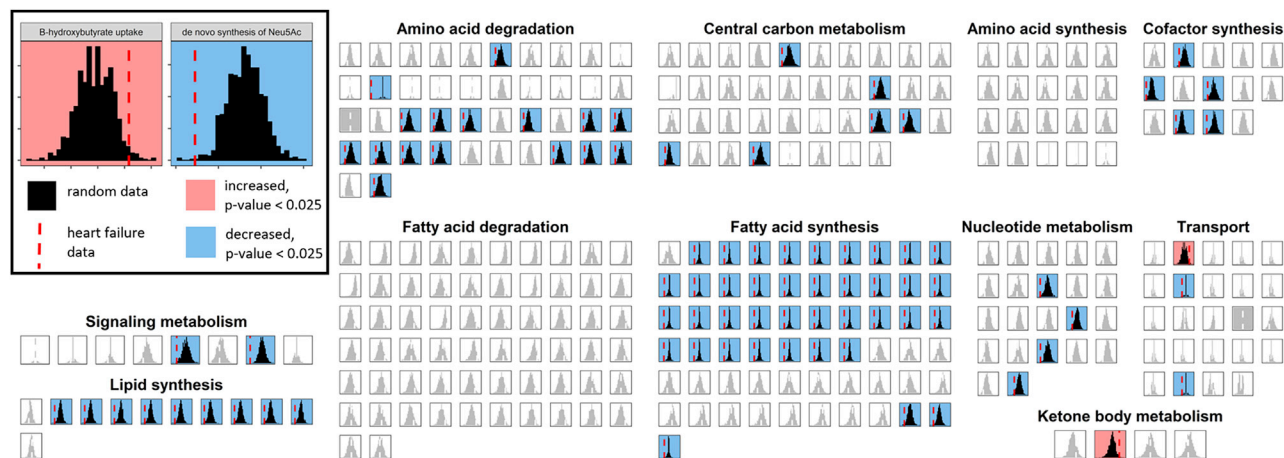


Figure 4. Distributions for model-predicted task scores from gene expression data for ischemic heart failure for each of the 307 tested tasks
Gene expression data for ischemic heart failure versus healthy hearts from GEO: GSE5406 was integrated with *iCardio* to identify shifts in metabolic functions. 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 whether the metabolic function had a statistically significant ($p < 0.025$) increased (red) or decreased (blue) task score.

associated with increased or decreased gene expression (Figure S5). Although 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 pathways, such as NO synthesis or Neu5Ac synthesis.

Second, a GSEA approach includes every gene that can be present in the pathway, even if the gene function is redundant, such as with isozymes. The TIDEs approach allows for the fold-change of one gene to determine the weight of each reaction. For example, one of the metabolic tasks related to synthesis of NO from arginine (Figure 6), a subset of the reactions necessary for that metabolic task is displayed in Figure 6. Two of those reactions can be catalyzed by one of three enzymes, NOS1, NOS2, or NOS3, which are known to have tissue-specific expression (Mattila and Thomas, 2014). The TIDEs reaction-centric approach uses the log-fold-change of one gene to determine the reaction weight, allowing a different gene to determine a reaction's weight among studies and across tissues. In this study, the reaction weight for the transporter responsible for the import of arginine to, and the export of citrulline from, the cytosol was determined by different genes among the idiopathic heart-failure data in GEO: GSE5406 and GSE57345 (Figure 6C). However, the metabolic task of NO production from arginine was statistically associated with decreased gene expression in both datasets (Figure 5A). A similar GSEA analysis would have included all five genes rather than the two to three used in the TIDEs analysis.

DISCUSSION

Here, we present a validated model of heart metabolism, *iCardio*, with an approach for analyzing changes in metabolic functions based on gene expression data. We present an 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 algorithm (Schultz and Qutub, 2016) produced a draft tissue-specific model that was more accurate with respect to performance of heart-relevant metabolic tasks than other integration methods (Table S2). It is interesting to note that the CORDA algorithm produced one of the smaller draft models and still maintained 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*, approximately 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 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 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 and also identifies 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 can serve as hypotheses for changes in metabolic functions of the heart or other tissues that could then be tested using gene expression datasets. Finally, although not every reaction is covered by a metabolic task, these reactions represent metabolic functionality with either evidence of protein expression in the HPA or pathway

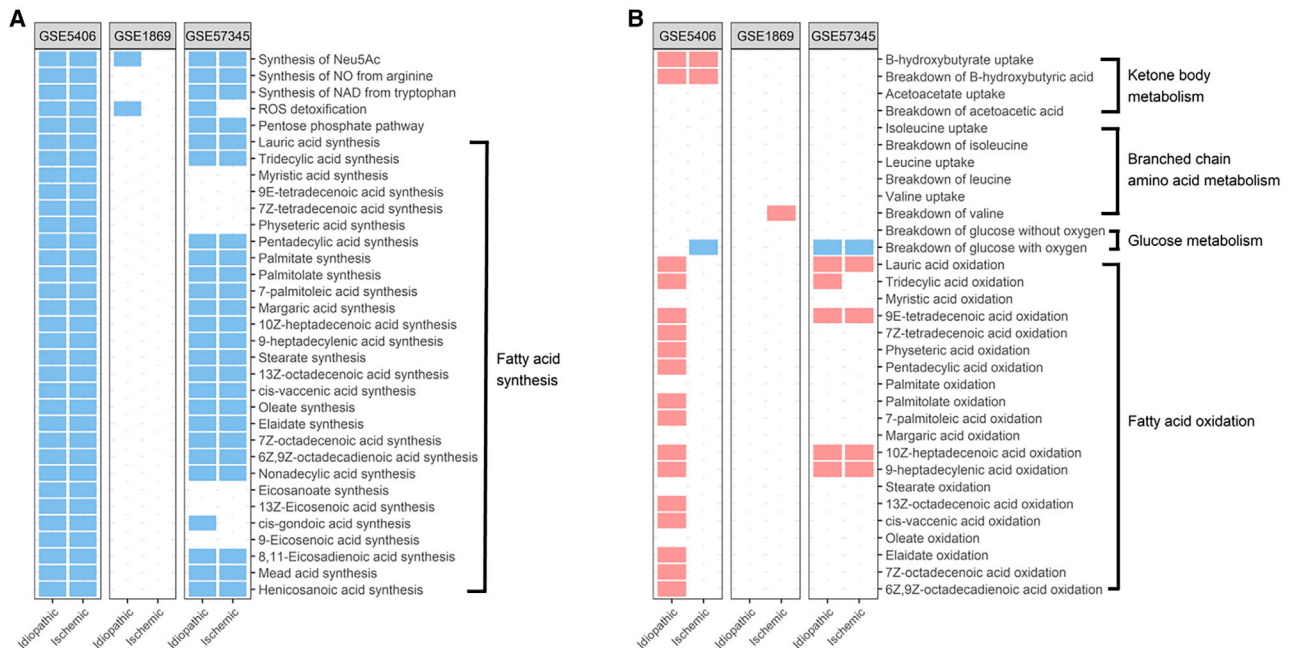


Figure 5. TIDEs analysis identifies functional metabolic changes shared across different datasets, whereas no changes in transcription were seen for common metabolic shifts in heart failure

A red box indicates a significant positive association (p value < 0.025), and a blue box indicates a significant negative association (p value < 0.025). Three datasets that contained samples for ischemic, idiopathic, and healthy hearts were downloaded, analyzed, and integrated with *iCardio* using the TIDEs approach to identify shifts in metabolic functions. (A) Here, we display a subset of those metabolic functions that were identified to be either changing consistently transcriptionally across datasets or had been previously reported to be associated with a change in uptake rates in heart failure (B).

connectivity. These reactions and associated pathways can serve as starting points for future metabolic task curation and further support the use of metabolic network reconstructions for generating hypotheses for important, tissue-specific metabolic functions.

When identifying reactions necessary for each metabolic task, the pFBA approach assumes that the pathway for each task remains the same, independent of the data. This reaction-centric TIDEs approach for determining significantly changed metabolic functions emphasizes that (1) metabolic functions require multiple complex, stoichiometrically balanced reactions; and (2) some genes may disproportionately influence the completion of metabolic functions. For example, two metabolic tasks cover the synthesis of NO from arginine (Data S1). The first task comprehensively covers the entire pathway, providing only arginine and other ions extracellularly as inputs while requiring the production of NO 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 NO production, providing arginine and dihydronicotinamide adenine dinucleotide phosphate (NADPH), and requires the production of NO. Although both tasks cover the synthesis of NO from arginine, the first metabolic task uses 53 reactions and covers the entire pathway, including transport of metabolites between compartments, which is necessary to maintain the stoichiometric balance of the pathway. However, because the second metabolic task covers mainly reactions in the cytosol, it requires only five reactions. Although the first task was present

in the original *iHsa* task list, the second task was added because of the importance of the synthesis of NO in the heart and, therefore, covers the core reactions for NO 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 strongly influence a final task score by determining 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, multiple studies have noted changes in metabolite uptake in late-stage heart failure (Kundu et al., 2015; Paolisso et al., 1994; Taylor et al., 2001), including a recent, comprehensive metabolomics analysis (Murashige et al., 2020). However, it is unclear whether these changes are driven by changes in gene expression in the heart or other factors, such as changes in circulating levels of metabolites or

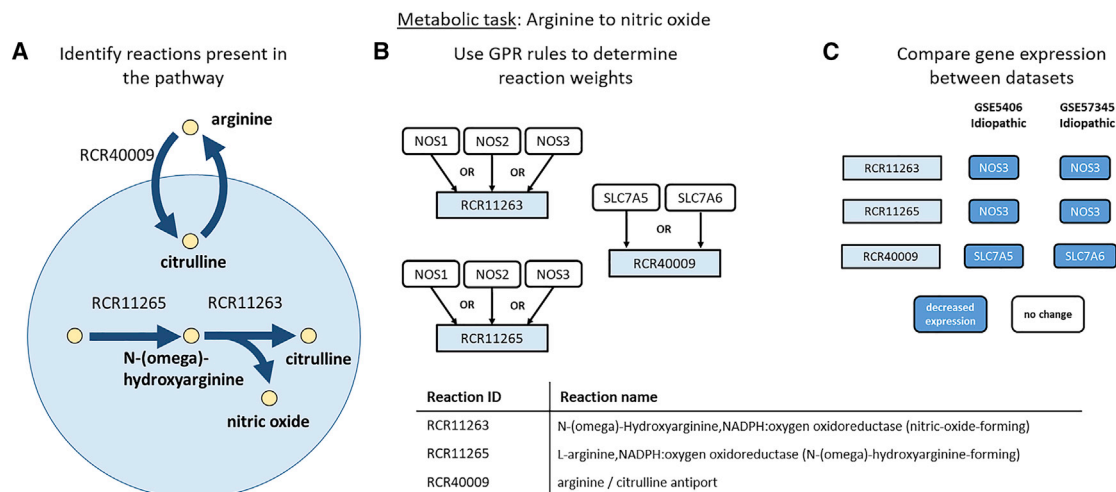


Figure 6. TIDEs analysis identifies differences in datasets for genes that catalyze the conversion of arginine to nitric oxide

(A) For the metabolic task of arginine to nitric oxide, a subset of the reactions necessary for the metabolic function 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 and still produce 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.

hormones. The TIDEs pipeline did not identify consistent, significant changes in transcription for ketone body utilization, branched-chain amino acid metabolism, or fatty acid oxidation (Figure 5). In the case of ketone body utilization, a recent metabolomics analysis suggests that changes in circulating levels of ketone bodies influence increased use in late-stage heart failure (Murashige et al., 2020). Although the role of fatty acids in late-stage heart failure remains controversial, with some studies measuring increased use (Paolisso et al., 1994; Taylor et al., 2001) whereas animal models measure no change (Zhabyeyev et al., 2013) or decreased use (Allard et al., 1994; Neubauer, 2007), the TIDEs analysis demonstrated no consistent change in gene expression for fatty acid oxidation in late-stage heart failure (Figures 5B and S3B). Together, this suggests that factors other than changes in transcription are driving the measured changes in metabolite uptake for ketone bodies, amino acids, and fatty acids in late-stage heart failure.

The TIDEs pipeline did identify three common transcriptional changes across the late-stage ischemic and idiopathic heart failure datasets: decreased gene expression for the use of glucose in the presence of oxygen, decreased gene expression for NO synthesis, and decreased gene expression for Neu5Ac synthesis. Although decreased glucose uptake has been noted in human heart failure (Paolisso et al., 1994; Taylor et al., 2001), here, we confirm the potential role for transcriptional changes as influencing changes in glucose metabolism in late-stage heart failure (Mori et al., 2012). In the ischemic and idiopathic datasets (Figure 5) and the general heart-failure datasets (Figure S3), we identified significant decreased gene expression for the use of glucose in the presence of oxygen. Recent work in animal models (Fernandez-Caggiano et al., 2020; McCommis et al., 2020; Zhang et al., 2020) demonstrated the role that the mito-

chondrial pyruvate carrier (MTC) and pyruvate metabolism have in the development of heart failure. Of note, two of these studies analyzed the expression of *MTC* in failing hearts (Fernandez-Caggiano et al., 2020; McCommis et al., 2020), noting decreased protein or normalized mRNA levels. However, we see no decreased gene expression for either *MCT-1* or *-2* in the datasets used in this current study. This result would suggest that, although decreased *MCT-1/2* expression is one potential mechanism, other potential points of regulation can influence the decreased use of glucose, which is captured in the metabolic functions presented in this study. In the case of NO synthesis, previous work has highlighted the important role of NO in cardiac function (Li et al., 2020; Massion et al., 2003), and more recent work has suggested a role for increasing NO synthesis for the increasing efficacy of beta-blockers (Hayashi et al., 2018).

Neu5Ac is the most common sialic acid associated with multiple functional roles in the body. Increased Neu5Ac levels in blood plasma have been measured for coronary artery disease (Zhang et al., 2018), atrial fibrillation (Hu et al., 2020), and chronic heart failure (Rajendiran et al., 2014), and Neu5Ac is generally associated with increased risk for cardiovascular diseases (Gopaul and Crook, 2006). However, increased Neu5Ac in blood plasma has mainly been attributed to changes in glycosylation in circulating lipoproteins and has been correlated with increased circulating triglyceride levels (Israr et al., 2018). Here, we identified a significant decrease in gene expression for the production of Neu5Ac in the heart: first in the idiopathic and ischemic heart-failure datasets (Figure 5), and second, in the general heart-failure datasets (Figure S3). Previous work in a mouse model demonstrated that inhibiting protein glycosylation through deletion of *ST3Gal4*, a sialyltransferase, resulted in the development of dilated cardiomyopathy (DCM) and stress-induced heart failure

(Deng et al., 2016). Knockouts of other genes related to protein glycosylation have also been shown to alter ion signaling in cardiomyocytes (Ednie et al., 2019; Montpetit et al., 2009), resulting in cardiac abnormalities. Finally, a congenital genetic defect in Neu5Ac synthesis led to dilated cardiomyopathy, which was confirmed in a genetic knockout in zebrafish (Wen et al., 2018). Together with the results presented here, this observation suggests a role for Neu5Ac metabolism as a marker of, and potential contributing mechanism for, 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. Further, differences in classifications between different publicly available datasets (i.e., ischemic, dilated, idiopathic, etc.; Table S4) makes it increasingly difficult to draw conclusions across studies. In addition, the datasets cluster within each study for gene expression, the TIDEs approach, and the GSEA KEGG metabolic pathway analysis, suggesting, again, that there is a large amount of heterogeneity in changes in gene expression for heart failure. However, pathway-based approaches, such as the TIDEs approach, have been shown to better reconcile both species-specific differences (Brubaker et al., 2019; Normand et al., 2018) and platform differences (e.g., microarray or RNA-seq) (van der Kloet et al., 2020) when compared with using gene expression alone.

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 (Murashige et al., 2020; Neubauer, 2007). Future work can integrate clinical measures, such as left ventricular ejection fraction (LVEF) or fluorodeoxyglucose-positron emission tomography (FDG-PET) 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 an 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 2-fold role for model validation and concrete metabolic functions to identify metabolic shifts in gene expression data. TIDEs represent a reaction-centric approach to identifying changes in metabolic functions and testing hypotheses for changes in gene expression for metabolic functions. These 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.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- METHOD DETAILS
 - Developing heart-relevant metabolic tasks
 - Curating a heart-specific metabolic model
 - Validating *iCardio*
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Analyzing transcriptomics data
 - Identifying TIDEs
 - Gene set enrichment analysis

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2021.108836>.

ACKNOWLEDGMENTS

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, USA) as part of the U.S. Army's Network Science Initiative. Support for this project was provided by the U.S. Department of Defense (W81XWH-14-C-0054 to J.A.P.) and the National Science Foundation graduate research fellowship program (awarded to B.V.D.).

AUTHOR CONTRIBUTIONS

B.V.D., A.W., and J.A.P. conceived the study. B.V.D. performed the computational modeling and data analysis. B.V.D. wrote the initial draft of the manuscript. B.V.D., K.D.R., G.L.K., K.C.V., A.W., and J.A.P. edited and wrote the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 22, 2020
Revised: January 7, 2021
Accepted: February 17, 2021
Published: March 9, 2021

REFERENCES

Agren, R., Mardinoglu, A., Asplund, A., Kampf, C., Uhlen, M., and Nielsen, J. (2014). Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol. Syst. Biol.* 10, 721.

- Allard, M.F., Schönekeß, B.O., Henning, S.L., English, D.R., and Lopaschuk, G.D. (1994). Contribution of oxidative metabolism and glycolysis to ATP production in hypertrophied hearts. *Am. J. Physiol.* *267*, H742–H750.
- Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* *41*, D991–D995.
- Bauckneht, M., Ferrarazzo, G., Fiz, F., Morbelli, S., Sarocchi, M., Pastorino, F., Ghidella, A., Pomposelli, E., Miglino, M., Ameri, P., et al. (2017). Doxorubicin effect on myocardial metabolism as a prerequisite for subsequent development of cardiac toxicity: a translational ¹⁸F-FDG PET/CT observation. *J. Nucl. Med.* *58*, 1638–1645.
- Becker, S.A., and Palsson, B.O. (2008). Context-specific metabolic networks are consistent with experiments. *PLoS Comput. Biol.* *4*, e1000082.
- Blais, E.M., Rawls, K.D., Dougherty, B.V., Li, Z.I., Kolling, G.L., Ye, P., Wallqvist, A., and Papin, J.A. (2017). Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nat. Commun.* *8*, 14250.
- Blazier, A.S., and Papin, J.A. (2012). Integration of expression data in genome-scale metabolic network reconstructions. *Front. Physiol.* *3*, 299.
- Borde, C., Kand, P., and Basu, S. (2012). Enhanced myocardial fluorodeoxyglucose uptake following Adriamycin-based therapy: Evidence of early chemotherapeutic cardiotoxicity? *World J. Radiol.* *4*, 220–223.
- Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* *34*, 525–527.
- Brubaker, D.K., Proctor, E.A., Haigis, K.M., and Lauffenburger, D.A. (2019). Computational translation of genomic responses from experimental model systems to humans. *PLoS Comput. Biol.* *15*, e1006286.
- Brunk, E., Sahoo, S., Zielinski, D.C., Altunkaya, A., Dräger, A., Mih, N., Gatto, F., Nilsson, A., Preciat Gonzalez, G.A., Aurich, M.K., et al. (2018). Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.* *36*, 272–281.
- Deng, W., Ednie, A.R., Qi, J., and Bennett, E.S. (2016). Aberrant sialylation causes dilated cardiomyopathy and stress-induced heart failure. *Basic Res. Cardiol.* *111*, 57.
- Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srivas, R., and Palsson, B.O. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA* *104*, 1777–1782.
- Ednie, A.R., Deng, W., Yip, K.-P., and Bennett, E.S. (2019). Reduced myocyte complex N-glycosylation causes dilated cardiomyopathy. *FASEB J.* *33*, 1248–1261.
- Fernandez-Caggiano, M., Kamynina, A., Francois, A.A., Pryszyzna, O., Eykyn, T.R., Krasemann, S., Crespo-Leiro, M.G., Vieites, M.G., Bianchi, K., Morales, V., et al. (2020). Mitochondrial pyruvate carrier abundance mediates pathological cardiac hypertrophy. *Nat. Metab.* *2*, 1223–1231.
- Folger, O., Jerby, L., Frezza, C., Gottlieb, E., Ruppín, E., and Shlomi, T. (2011). Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* *7*, 501.
- Gopaul, K.P., and Crook, M.A. (2006). Sialic acid: a novel marker of cardiovascular disease? *Clin. Biochem.* *39*, 667–681.
- Greco, S., Fasanaro, P., Castelvécchio, S., D’Alessandra, Y., Arcelli, D., Di Donato, M., Malavazos, A., Capogrossi, M.C., Menicanti, L., and Martelli, F. (2012). MicroRNA dysregulation in diabetic ischemic heart failure patients. *Diabetes* *61*, 1633–1641.
- Hannenhalli, S., Putt, M.E., Gilmore, J.M., Wang, J., Parmacek, M.S., Epstein, J.A., Morrisey, E.E., Margulies, K.B., and Cappola, T.P. (2006). Transcriptional genomics associates FOX transcription factors with human heart failure. *Circulation* *114*, 1269–1276.
- Hayashi, H., Hess, D.T., Zhang, R., Sugi, K., Gao, H., Tan, B.L., Bowles, D.E., Milano, C.A., Jain, M.K., Koch, W.J., and Stamler, J.S. (2018). S-nitrosylation of β -arrestins biases receptor signaling and confers ligand independence. *Mol. Cell* *70*, 473–487.e6.
- Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S.N., Richelle, A., Heinken, A., Haraldsdóttir, H.S., Wachowiak, J., Keating, S.M., Vlasov, V., et al. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* *14*, 639–702.
- Hu, W., Xie, J., Zhu, T., Meng, G., Wang, M., Zhou, Z., Guo, F., Chen, H., Wang, Z., Wang, S., et al. (2020). Serum N-acetylneuraminic acid is associated with atrial fibrillation and left atrial enlargement. *Cardiol. Res. Pract.* *2020*, 1358098.
- Israr, M., Gul, C., Rehman, A.U., Khan, A.S., Siyar, M., Rahim, S., and Arshad, M. (2018). Relationship of lipids, C-reactive protein and sialic acid in the healthy individuals. *J. Ayub Med. Coll. Abbottabad* *30*, 213–216.
- Janardhan, A., Chen, J., and Crawford, P.A. (2011). Altered systemic ketone body metabolism in advanced heart failure. *Tex. Heart Inst. J.* *38*, 533–538.
- Jerby, L., and Ruppín, E. (2012). Predicting drug targets and biomarkers of cancer via genome-scale metabolic modeling. *Clin. Cancer Res.* *18*, 5572–5584.
- Jerby, L., Shlomi, T., and Ruppín, E. (2010). Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol. Syst. Biol.* *6*, 401.
- Karlstädt, A., Fliegner, D., Kararigas, G., Ruderisch, H.S., Regitz-Zagrosek, V., and Holzhütter, H.-G. (2012). CardioNet: a human metabolic network suited for the study of cardiomyocyte metabolism. *BMC Syst. Biol.* *6*, 114.
- Kittleson, M.M., Minhas, K.M., Irizarry, R.A., Ye, S.Q., Edness, G., Breton, E., Conte, J.V., Tomaselli, G., Garcia, J.G.N., and Hare, J.M. (2005). Gene expression analysis of ischemic and nonischemic cardiomyopathy: shared and distinct genes in the development of heart failure. *Physiol. Genomics* *21*, 299–307.
- Kundu, B.K., Zhong, M., Sen, S., Davogusto, G., Keller, S.R., and Taegtmeier, H. (2015). Remodeling of glucose metabolism precedes pressure overload-induced left ventricular hypertrophy: review of a hypothesis. *Cardiology* *130*, 211–220.
- Lewis, N.E., Hixson, K.K., Conrad, T.M., Lerman, J.A., Charusanti, P., Polpitiya, A.D., Adkins, J.N., Schramm, G., Purvine, S.O., Lopez-Ferrer, D., et al. (2010). Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* *6*, 390.
- Li, R., He, H., Fang, S., Hua, Y., Yang, X., Yuan, Y., Liang, S., Liu, P., Tian, Y., Xu, F., et al. (2019). Time series characteristics of serum branched-chain amino acids for early diagnosis of chronic heart failure. *J. Proteome Res.* *18*, 2121–2128.
- Li, M., Parker, B.L., Pearson, E., Hunter, B., Cao, J., Koay, Y.C., Guneratne, O., James, D.E., Yang, J., Lal, S., and O’Sullivan, J.F. (2020). Core functional nodes and sex-specific pathways in human ischaemic and dilated cardiomyopathy. *Nat. Commun.* *11*, 2843.
- Liu, H., Lv, L., and Yang, K. (2015). Chemotherapy targeting cancer stem cells. *Am. J. Cancer Res.* *5*, 880–893.
- Lopaschuk, G.D. (2017). Metabolic modulators in heart disease: past, present, and future. *Can. J. Cardiol.* *33*, 838–849.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*, 550.
- Ma, H., Sorokin, A., Mazein, A., Selkov, A., Selkov, E., Demin, O., and Goryain, I. (2007). The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol. Syst. Biol.* *3*, 135.
- Machado, D., and Herrgård, M. (2014). Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput. Biol.* *10*, e1003580.
- Mardinoglu, A., Gatto, F., and Nielsen, J. (2013). Genome-scale modeling of human metabolism - a systems biology approach. *Biotechnol. J.* *8*, 985–996.
- Massion, P.B., Feron, O., Dessy, C., and Balligand, J.-L. (2003). Nitric oxide and cardiac function: ten years after, and continuing. *Circ. Res.* *93*, 388–398.
- Mattila, J.T., and Thomas, A.C. (2014). Nitric oxide synthase: non-canonical expression patterns. *Front. Immunol.* *5*, 478.
- McCommis, K.S., Kovacs, A., Weinheimer, C.J., Shew, T.M., Koves, T.R., Ilkayeva, O.R., Kamm, D.R., Pyles, K.D., King, M.T., Veech, R.L., et al.

- (2020). Nutritional modulation of heart failure in mitochondrial pyruvate carrier-deficient mice. *Nat Metab* 2, 1232–1247.
- Montpetit, M.L., Stocker, P.J., Schwetz, T.A., Harper, J.M., Norring, S.A., Schaffer, L., North, S.J., Jang-Lee, J., Gilmartin, T., Head, S.R., et al. (2009). Regulated and aberrant glycosylation modulate cardiac electrical signaling. *Proc. Natl. Acad. Sci. USA* 106, 16517–16522.
- Mori, J., Basu, R., McLean, B.A., Das, S.K., Zhang, L., Patel, V.B., Wagg, C.S., Kassiri, Z., Lopaschuk, G.D., and Oudit, G.Y. (2012). Agonist-induced hypertrophy and diastolic dysfunction are associated with selective reduction in glucose oxidation: a metabolic contribution to heart failure with normal ejection fraction. *Circ. Heart Fail.* 5, 493–503.
- Murashige, D., Jang, C., Neinast, M., Edwards, J.J., Cowan, A., Hyman, M.C., Rabinowitz, J.D., Frankel, D.S., and Arany, Z. (2020). Comprehensive quantification of fuel use by the failing and nonfailing human heart. *Science* 370, 364–368.
- Neubauer, S. (2007). The failing heart—an engine out of fuel. *N. Engl. J. Med.* 356, 1140–1151.
- Normand, R., Du, W., Briller, M., Gaujoux, R., Starosvetsky, E., Ziv-Kenet, A., Shalev-Malul, G., Tibshirani, R.J., and Shen-Orr, S.S. (2018). Found in translation: a machine learning model for mouse-to-human inference. *Nat. Methods* 15, 1067–1073.
- Paolisso, G., Gambardella, A., Galzerano, D., D’Amore, A., Rubino, P., Verza, M., Teasuro, P., Varricchio, M., and D’Onofrio, F. (1994). Total-body and myocardial substrate oxidation in congestive heart failure. *Metabolism* 43, 174–179.
- Rajendiran, K.S., Ananthanarayanan, R.H., Satheesh, S., and Rajappa, M. (2014). Elevated levels of serum sialic acid and high-sensitivity C-reactive protein: markers of systemic inflammation in patients with chronic heart failure. *Br. J. Biomed. Sci.* 71, 29–32.
- Rawls, K., Dougherty, B.V., and Papin, J. (2020). Metabolic network reconstructions to predict drug targets and off-target effects. *Methods Mol. Biol.* 2088, 315–330.
- Ren, Z., Yu, P., Li, D., Li, Z., Liao, Y., Wang, Y., Zhou, B., and Wang, L. (2020). Single-cell reconstruction of progression trajectory reveals intervention principles in pathological cardiac hypertrophy. *Circulation* 141, 1704–1719.
- Richelle, A., Chiang, A.W.T., Kuo, C.-C., and Lewis, N.E. (2019). Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. *PLoS Comput. Biol.* 15, e1006867.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47–e47.
- Robaina Estévez, S., and Nikoloski, Z. (2014). Generalized framework for context-specific metabolic model extraction methods. *Front. Plant Sci.* 5, 491.
- Schiano, C., Costa, V., Aprile, M., Grimaldi, V., Maiello, C., Esposito, R., Soricelli, A., Colantuoni, V., Donatelli, F., Ciccociola, A., and Napoli, C. (2017). Heart failure: Pilot transcriptomic analysis of cardiac tissue by RNA-sequencing. *Cardiol. J.* 24, 539–553.
- Schultz, A., and Qutub, A.A. (2016). Reconstruction of tissue-specific metabolic networks using CORDA. *PLoS Comput. Biol.* 12, e1004808.
- Shaked, I., Oberhardt, M.A., Atias, N., Sharan, R., and Ruppin, E. (2016). Metabolic network prediction of drug side effects. *Cell Syst.* 2, 209–213.
- Smith, A.C., and Robinson, A.J. (2011). A metabolic model of the mitochondrion and its use in modelling diseases of the tricarboxylic acid cycle. *BMC Syst. Biol.* 5, 102.
- Smith, A.C., Eyassu, F., Mazat, J.-P., and Robinson, A.J. (2017). MitoCore: a curated constraint-based model for simulating human central metabolism. *BMC Syst. Biol.* 11, 114.
- Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* 4, 1521.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550.
- Sun, H., Olson, K.C., Gao, C., Prosdocimo, D.A., Zhou, M., Wang, Z., Jeyaraj, D., Youn, J.-Y., Ren, S., Liu, Y., et al. (2016). Catabolic defect of branched-chain amino acids promotes heart failure. *Circulation* 133, 2038–2049.
- Swainston, N., Smallbone, K., Hefzi, H., Dobson, P.D., Brewer, J., Hanscho, M., Zielinski, D.C., Ang, K.S., Gardiner, N.J., Gutierrez, J.M., et al. (2016). Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics* 12, 109.
- Taylor, M., Wallhaus, T.R., Degrado, T.R., Russell, D.C., Stanko, P., Nickles, R.J., and Stone, C.K. (2001). An evaluation of myocardial fatty acid and glucose uptake using PET with [¹⁸F]fluoro-6-thia-heptadecanoic acid and [¹⁸F]FDG in patients with congestive heart failure. *J. Nucl. Med.* 42, 55–62.
- Thiele, I., Swainston, N., Fleming, R.M.T., Hoppe, A., Sahoo, S., Aurich, M.K., Haraldsdottir, H., Mo, M.L., Rolfsson, O., Stobbe, M.D., et al. (2013). A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* 31, 419–425.
- Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Marding, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics: tissue-based map of the human proteome. *Science* 347, 1260419.
- van der Kloet, F.M., Buurmans, J., Jonker, M.J., Smilde, A.K., and Westerhuis, J.A. (2020). Increased comparability between RNA-Seq and microarray data by utilization of gene sets. *PLoS Comput. Biol.* 16, e1008295.
- Vlassis, N., Pacheco, M.P., and Sauter, T. (2014). Fast reconstruction of compact context-specific metabolic network models. *PLoS Comput. Biol.* 10, e1003424.
- Watt, I.N., Montgomery, M.G., Runswick, M.J., Leslie, A.G.W., and Walker, J.E. (2010). Bioenergetic cost of making an adenosine triphosphate molecule in animal mitochondria. *Proc. Natl. Acad. Sci. USA* 107, 16823–16827.
- Wen, X.-Y., Tarailo-Graovac, M., Brand-Arzamendi, K., Willems, A., Rakic, B., Huijben, K., Da Silva, A., Pan, X., El-Rass, S., Ng, R., et al. (2018). Sialic acid catabolism by *N*-acetylneuraminase pyruvate lyase is essential for muscle function. *JCI Insight* 3, e122373.
- Wende, A.R., Brahma, M.K., McGinnis, G.R., and Young, M.E. (2017). Metabolic origins of heart failure. *JACC Basic Transl. Sci.* 2, 297–310.
- Zhabyeyev, P., Gandhi, M., Mori, J., Basu, R., Kassiri, Z., Clanachan, A., Lopaschuk, G.D., and Oudit, G.Y. (2013). Pressure-overload-induced heart failure induces a selective reduction in glucose oxidation at physiological afterload. *Cardiovasc. Res.* 97, 676–685.
- Zhang, A.-D., Dai, S.-X., and Huang, J.-F. (2013). Reconstruction and analysis of human kidney-specific metabolic network based on omics data. *Biomed Res. Int.* 2013, 187509.
- Zhang, L., Wei, T.-T., Li, Y., Li, J., Fan, Y., Huang, F.-Q., Cai, Y.-Y., Ma, G., Liu, J.-F., Chen, Q.-Q., et al. (2018). Functional metabolomics characterizes a key role for *N*-acetylneuraminic acid in coronary artery diseases. *Circulation* 137, 1374–1390.
- Zhang, Y., Taufalele, P.V., Cochran, J.D., Robillard-Frayne, I., Marx, J.M., Soto, J., Rauckhorst, A.J., Tayyari, F., Pawa, A.D., Gray, L.R., et al. (2020). Mitochondrial pyruvate carriers are required for myocardial stress adaptation. *Nat Metab* 2, 1248–1264.
- Zhao, Y., and Huang, J. (2011). Reconstruction and analysis of human heart-specific metabolic network based on transcriptome and proteome data. *Biochem. Biophys. Res. Commun.* 415, 450–454.
- Zielinski, D.C., Filipp, F.V., Bordbar, A., Jensen, K., Smith, J.W., Herrgard, M.J., Mo, M.L., and Palsson, B.O. (2015). Pharmacogenomic and clinical data link non-pharmacokinetic metabolic dysregulation to drug side effect pathogenesis. *Nat. Commun.* 6, 7101.
- Zur, H., Ruppin, E., and Shlomi, T. (2010). iMAT: an integrative metabolic analysis tool. *Bioinformatics* 26, 3140–3142.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Ischemic and idiopathic microarray heart failure datasets	Kittleson et al., 2005	GSE1869
	Liu et al., 2015	GSE57345
	Hannenhalli et al., 2006	GSE5406
General heart failure microarray and transcriptomics datasets	Ren et al., 2020	GSE133054
	Greco et al., 2012	GSE26887
	Schiano et al., 2017	GSE71613
	unpublished but collected by https://www.med.upenn.edu/magnet/	GSE141910
Software and algorithms		
All code used for data analysis	This paper, GitHub	https://github.com/csbl/iCardio
COBRA Toolbox	Heirendt et al., 2019	https://github.com/opencobra/cobratoolbox/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jason Papin (papin@virginia.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

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

METHOD DETAILS

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 ([Blais et al., 2017](#); [Thiele et al., 2013](#)). 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 heart-relevant tasks, including 25 that quantitatively describe ATP generation from various carbon sources. Testing these heart-relevant tasks with *iHsa* resulted in five changes to the network reconstruction ([Table S1](#)), 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 ([Smith and Robinson, 2011](#)) and (b) the change from 4 protons to 2.7 protons to generate one ATP molecule to be consistent with recently published data ([Watt et al., 2010](#)). The heart-relevant and general metabolic tasks together represent 421 metabolic tasks ([Data S1](#)) 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 1](#)).

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 (data available from version 18, <https://www.proteinatlas.org>; Uhlén et al., 2015) with *iHsa* (Blais et al., 2017). Various algorithms have been published to generate tissue-specific models from tissue transcriptomics or proteomics data. Here, we implemented 5 of these algorithms (Becker and Palsson, 2008; Jerby et al., 2010; Schultz and Qutub, 2016; Vlassis et al., 2014; Zur et al., 2010) 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 (Uhlén et al., 2015). 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 (Data S1; Table S2).

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.

Validating iCardio

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* (Smith et al., 2017). Parsimonious flux balance analysis (pFBA) (Lewis et al., 2010) 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 (Machado and Herrgård, 2014). 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 (Smith et al., 2017).

QUANTIFICATION AND STATISTICAL ANALYSIS

Analyzing transcriptomics data

Microarray data (Table S3) (Hannenhalli et al., 2006; Kittleson et al., 2005; Liu et al., 2015) from patients undergoing heart transplants for advanced ischemic or idiopathic heart failure were downloaded from the Gene Expression Omnibus (GEO) database (Barrett et al., 2013). 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 (Ritchie et al., 2015) was used to determine differentially expressed genes (DEGs) between healthy hearts and ischemic or idiopathic hearts. Genes with an FDR < 0.1 were considered to be differentially expressed and their corresponding fold change was used in subsequent analysis. Although one study contained patient metadata, this data was not included in downstream analysis.

To validate the identified metabolic functions associated with significant changes in transcription in general heart failure, four additional datasets were analyzed (Table S5) (Greco et al., 2012; Ren et al., 2020; Schiano et al., 2017), which included one microarray and three RNA-seq datasets. For the microarray dataset, DEGs were determined using the above methods. For the RNA-seq datasets, raw FASTQ files were downloaded from GEO (Barrett et al., 2013), pseudo aligned to the *Homo sapiens* transcriptome Ensembl v96 transcriptome using kallisto (Bray et al., 2016), transcript abundances were aggregated to the Entrez gene level R v. 3.6.3 with the package tximport (Soneson et al., 2015), and fold changes were calculated using DESeq2 (Love et al., 2014). Genes with an FDR < 0.1 were considered to be DEGs and their corresponding fold change was used in subsequent analysis.

Identifying TIDES

Metabolic tasks and their associated reactions, as identified using *iCardio* 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 3). A total of 307 metabolic tasks was used for this analysis, representing the tasks functionally present in *iCardio* from the original task list (Data S1) that also contained at least one reaction with an associated GPR rule (Data S4). Reactions that carry flux for each task are identified using a pFBA assumption without previous knowledge of the gene expression data (Figure 3A), as has been done with a related approach (Jerby and Rupp, 2012). 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, the fold change with the highest absolute weight 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 3B). To assign statistical significance to these task scores, the gene expression fold changes are randomized 1000 times among 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 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. For validation of selected TIDEs using general heart failure data, a p value < 0.05 was used for determining 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.

Gene set enrichment analysis

The same gene expression datasets (GSE1869, GSE5406, GSE57345) were also analyzed using gene set enrichment analysis (GSEA) (Subramanian et al., 2005) 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.