

## **$\Delta$ FBA: Predicting metabolic flux alterations using genome-scale metabolic models and differential transcriptomic data**

Sudharshan Ravi<sup>a,b</sup>, Rudiyanto Gunawan<sup>a</sup>

<sup>a</sup> Department of Chemical and Biological Engineering, University, Buffalo, NY 14260, USA

<sup>b</sup> Institute for Chemical and Bioengineering, ETH Zurich, 8093 Zurich, Switzerland

## Abstract

Genome-scale metabolic models (GEMs) provide a powerful framework for simulating the entire set of biochemical reactions occurring in a cell. Constraint-based modeling tools like flux balance analysis (FBA) developed for the purposes of predicting metabolic flux distribution using GEMs face considerable difficulties in estimating metabolic flux alterations between experimental conditions. Particularly, the most appropriate metabolic objective for FBA is not always obvious, likely context-specific, and not necessarily the same between conditions. Here, we propose a new method, called  $\Delta$ FBA (deltaFBA), that employs constraint-based modeling, in combination with differential gene expression data, to evaluate changes in the intracellular flux distribution between two conditions. Notably,  $\Delta$ FBA does not require specifying the cellular objective to produce the flux change predictions. We showcased the performance of  $\Delta$ FBA through several case studies involving the prediction of metabolic alterations caused by genetic and environmental perturbations in *Escherichia coli* and caused by Type-2 diabetes in human muscle.

## Introduction

In the post-genomic era, there has been intense efforts directed toward the reconstruction of genome-scale models of cellular networks. An important portion of these efforts focuses on metabolic networks due to the significance of cellular metabolism for understanding diseases such as cancer <sup>1–4</sup> as well as for metabolic engineering applications in biomanufacturing <sup>5</sup>. Recent advances in high-throughput sequencing technologies, gene functional annotation, and metabolic pathway databases, and developments of algorithms for mapping gene-protein-reaction (GPR) associations and identifying missing metabolic reactions systematically (gap-filling), have enabled the reconstruction of thousands of genome-scale metabolic models (GEMs), from single cell organisms to human <sup>6,7</sup>. A GEM provides the gene-protein-reaction associations that encompass the set of metabolites and metabolic reactions in an organism as prescribed by its genome. Concurrent with these developments are the creation of efficacious algorithms that use GEMs to predict intracellular metabolic fluxes – i.e. the rates of metabolic reactions – and how these fluxes vary under different environmental, genetic, and disease conditions <sup>8–10</sup>.

A prominent class of algorithms based on a constrained-based modeling technique, called flux balance analysis (FBA), have flourished due to its ease of formulation and flexibility, using the stoichiometric coefficients of the metabolic reactions in a GEM, an assumed cellular objective such as maximization of biomass production, and various experimental data on metabolic capabilities and constraints of the cells, to predict metabolic fluxes <sup>11</sup>. Although FBA is effective in handling large networks and predicting cell behavior in many metabolic engineering studies <sup>12–15</sup>, considerable uncertainty still remains about the appropriate choice of cellular objective for different conditions and cell types – a choice that requires expert knowledge of the cells and their phenotype in a given condition. Such an issue is particularly prominent for complex organisms such as human. Moreover, the flux solutions produced by applying FBA have degeneracy – that is, multiple solutions exist that give the same cellular objective value <sup>16</sup>. Not to mention, standard FBA often produces biologically unrealistic flux solutions <sup>17,18</sup>.

Driven by the increasing ease and availability of whole-genome omics profiling data, a multitude of FBA-based algorithms have been developed to incorporate additional datasets to create context-specific metabolic networks and to improve flux prediction accuracy<sup>19–26</sup>. Several of these methods, such as GIMME (Gene Inactivity Moderated by Metabolism and Expression)<sup>19</sup>, iMAT (integrative Metabolic Analysis Tools)<sup>21</sup>, and MADE (Metabolic Adjustment by Differential Expression)<sup>25</sup>, are based on maximizing the consistency between the predicted flux distribution and the mRNA transcript abundance of metabolic genes, where the higher the transcript level of an enzyme, the larger should the flux of the corresponding reactions. Other methods, such as E-Flux<sup>23</sup>, use data of transcript abundance for setting the bounds on reaction fluxes. More recent methods, such as ME-model<sup>27</sup> and GECKO<sup>28</sup>, combine FBA with an explicit modeling of enzyme / protein expression and thus, are able to directly account for protein abundance. Thermodynamics constraints have also been integrated with the FBA to eliminate thermodynamically infeasible fluxes, and at the same time enable the integration of metabolite concentration data, as done in recent methods such as ETFL<sup>29</sup>.

A number of methods focus on using differential expression data between two conditions (e.g., perturbation vs. control) to predict metabolic alterations – a particular focus of our study. The method Relative Expression and Metabolic Integration (REMI)<sup>30</sup> used differential expression of transcriptome and metabolome to estimate metabolic flux profiles in *Escherichia coli* under varying dilution and genetic perturbations. The method relies on maximizing the agreement between the fold-changes of metabolic fluxes and the fold-changes of enzyme expressions between two conditions. The metabolome data, if available, are used to determine the flux directionality using reaction thermodynamics. Among the alternative flux solutions, the L1-norm minimal solution is adopted to give a representative flux distribution. Another method by Zhu *et al.*<sup>31</sup> employed a softer definition when assessing consistency between the metabolic fluxes and enzyme differential expressions, where only the sign of the differences needs to agree. The method provides a qualitative determination of metabolic flux changes by determining the maximum and minimum flux through each reaction in the GEM. Both of the above methods generate metabolic flux predictions for each condition in comparison, and like standard FBA, both methods require an assumption on the cell's metabolic objective. Generally, model prediction inaccuracy is

amplified when evaluating the differences between two model predictions. Another related method MOOMIN<sup>32</sup> uses a Bayesian approach to integrate differential gene expression profiles with GEMs to predict the qualitative change in the metabolic fluxes—increased, decreased or no change.

In this work, we developed  $\Delta$ FBA (deltaFBA) for predicting the metabolic flux difference given a GEM and differential transcriptomic data between two conditions.  $\Delta$ FBA relies on a constrained-based model that governs the balance of flux difference in the GEM, while maximizing the consistency and minimizing inconsistency between the flux alterations and the gene expression changes.  $\Delta$ FBA is created in MATLAB to work with the CONstraint-Based Reconstruction and Analysis (COBRA) toolbox<sup>33</sup>. We applied the  $\Delta$ FBA to analyze the metabolic changes of *Escherichia coli* in response to environmental and genetic perturbations using data from the studies of Ishii *et al.*<sup>34</sup> and Gerosa *et al.*<sup>35</sup>. We compared the performance of  $\Delta$ FBA to REMI for evaluating flux differences between conditions. We also demonstrated the application of  $\Delta$ FBA to human GEM, specifically evaluating the metabolic alterations associated with type-2-diabetes in skeletal muscle using myocyte-specific GEM<sup>36</sup>.

## Materials and Methods

### Method Formulation

$\Delta$ FBA generates a prediction for metabolic flux changes between a pair of conditions, such as treated vs. untreated or mutant vs. wild-type strains. In the following, we use the superscript C to denote the control (reference) condition and P to denote the perturbed condition.  $\Delta$ FBA enforces that the flux changes  $\Delta \mathbf{v} = (\mathbf{v}^P - \mathbf{v}^C)$  for each metabolite are balanced, as follows:

$$\mathbf{S}\Delta \mathbf{v} = \mathbf{S}(\mathbf{v}^P - \mathbf{v}^C) = \mathbf{S}\mathbf{v}^P - \mathbf{S}\mathbf{v}^C = \mathbf{0}$$

where  $\mathbf{S}$  denotes the  $m \times n$  stoichiometric matrix for  $m$  metabolites that are involved in  $n$  metabolic reactions in the GEM, and  $\mathbf{v}$  and  $\Delta \mathbf{v}$  denote the vector of metabolic fluxes and

flux changes, respectively. The flux balance above is a direct consequence of the steady-state flux balance in each condition, i.e.  $\mathbf{S} = \mathbf{0}$

$$\max_{\Delta \mathbf{v}, \mathbf{z}^U, \mathbf{z}^D} \Phi = \sum_{i \in \mathbf{R}^U} w_i (z_i^U - z_i^D) + \sum_{j \in \mathbf{R}^D} w_j (z_j^D - z_j^U) \quad (1)$$

subject to:

$$\mathbf{S} \Delta \mathbf{v} = \mathbf{0} \quad (1)$$

$$\Delta \mathbf{v}_{min} \leq \Delta \mathbf{v} \leq \Delta \mathbf{v}_{max} \quad (2)$$

$$\Delta \mathbf{v} - m \mathbf{z}^U \geq \boldsymbol{\mu} - m \mathbf{1} \quad (3)$$

$$\Delta \mathbf{v} - m \mathbf{z}^U \leq \boldsymbol{\mu} \quad (4)$$

$$\Delta \mathbf{v} + m \mathbf{z}^D \leq -\boldsymbol{\eta} + m \mathbf{1} \quad (5)$$

$$\Delta \mathbf{v} + m \mathbf{z}^D \geq -\boldsymbol{\eta} \quad (6)$$

$$\Delta \mathbf{v} + m \mathbf{z}^0 \leq m \mathbf{1} \quad (7)$$

$$\Delta \mathbf{v} - m \mathbf{z}^0 \geq -m \mathbf{1} \quad (8)$$

$$z_k^0 + z_k^U \geq +z_{k'}^D \quad (9)$$

$$z_k^0 + z_k^D \geq +z_{k'}^U \quad (11)$$

The weighting coefficients  $w_i$  allows users to prioritize certain reactions (e.g., reaction(s) corresponding to gene knock-out(s)) for (in)consistency. The set of upregulated reactions  $\mathbf{R}^U$  and the set of downregulated reactions  $\mathbf{R}^D$  are user-defined inputs. Typically, the sets  $\mathbf{R}^U$  and  $\mathbf{R}^D$  include reactions with statistically significant increase and decrease in gene expression between the perturbed condition and the control, respectively. Equation (2) specifies the upper and lower bounds for the flux change, which can be set based on experimental data (e.g. the change in the biomass production or growth rates), or based on the corresponding bounds from the perturbed ( $\mathbf{v}_{min}^P, \mathbf{v}_{max}^P$ ) and the control condition ( $\mathbf{v}_{min}^U, \mathbf{v}_{max}^U$ ) as follows:

$$\Delta v_{min,i} = (v_{min,i}^P - v_{max,i}^C) \quad (12)$$

$$\Delta v_{max,i} = (v_{max,i}^P - v_{min,i}^C) \quad (13)$$

The MILP above will generate flux change prediction  $\Delta v$  whose signs (positive and negative) agree with the direction of the gene expression changes. Specifically, the binary decision variables  $z^U$ ,  $z^D$ , and  $z^0$  in the MILP formulation set the sign (directionality) of the flux changes  $\Delta v$ . When  $z_i^U = 1$ ,  $\Delta v_i$  takes a positive value above a threshold  $\mu_i$ , as specified by Equations (3)-(4). Vice versa, when  $z_i^D = 1$ ,  $\Delta v_i$  takes a negative value below a threshold  $\eta_i$ , as specified by Equations (5)-(6). The thresholds  $\mu_i$  and  $\eta_i$  for the positive and negative flux changes, respectively, are user-defined parameters. For example, the thresholds can be set to the same constant value  $\varepsilon$ , or to a value that scales with the fold-change reaction expression  $e_i^{P/C}$  (see Supplementary Materials), as follows:

$$\mu_i = \varepsilon e_i^{P/C} \quad (14)$$

$$\eta_i = \varepsilon / e_i^{P/C} \quad (15)$$

Clearly,  $z_i^U$  and  $z_i^D$  cannot simultaneously be equal to 1 since no feasible  $\Delta v_i$  exists that simultaneously satisfy Equations (3)-(6). The decision variable  $z^0$  is used to force select reactions to have zero flux change value, as specified by Equations (7)-(8). Together with Equation (9)-(11), the forward and reverse halves of a reversible reaction,  $\Delta v_k$  and  $\Delta v_{k'}$ , are prevented to simultaneously have non-zero values, so as to reduce degeneracy of the flux change solution  $\Delta v$ .

Given the degrees of freedom in GEMs for  $\Delta v$ , many equivalent optimal solutions often exist that give the same objective function value  $\Phi^*$  as specified in Equation **Error! Reference source not found.**. By assuming parsimony in  $\Delta v$  – that is,  $\Delta v$  is minimal between the perturbed and control condition, a two-step optimization procedure is implemented in  $\Delta$ FBA. The first step is to maximize consistency with gene expression changes as prescribed in Equations **Error! Reference source not found.**-(11) with the maximum objective function value denoted by  $\Phi^*$ . The second step is to produce an L2 norm minimal solution for  $\Delta v$ , as follows:

$$\min_{\Delta v, z^U, z^D} \|\Delta v\|_2^2 \quad (16)$$

subject to the same constraints in Equations (1)-(11) while achieving the same level of consistency  $\Phi^*$ , implemented by the additional constraint:

$$\sum_{i \in \mathbf{R}^U} w_i (z_i^U - z_i^D) + \sum_{j \in \mathbf{R}^D} w_j (z_j^D - z_j^U) = \Phi^* \quad (17)$$

The L2 minimization is based on the premise that the perturbed metabolic fluxes should not deviate far from the control condition, similar to the method called Minimization of Metabolic Adjustment (MOMA) <sup>37</sup>. An alternative to L2-norm minimization is L1-norm minimization, which is analogous to maximizing sparsity of  $\Delta \mathbf{v}$ . The L1-norm minimization was previously used in the parsimonious FBA (pFBA) method, but such an approach often still leads to multiple degenerate solutions. On the other hand, the L2-norm minimization will produce a unique solution, but the mixed integer quadratic optimization that is required to find the solution may have high computational requirement.

$\Delta$ FBA is available as MATLAB scripts and are compatible with the COBRA toolbox.  $\Delta$ FBA requires Gurobi optimizer (<http://www.gurobi.com>) as a pre-requisite.  $\Delta$ FBA has been tested on a Windows PC using a 6-core Intel Xeon (2146G) Processor with 16 GB RAM.

### Case studies: Data and Implementation

The first case study involved the response of *E. coli*'s metabolism to genetic (single-gene deletions) and environmental perturbations (dilution rates) performed by Ishii *et al.* <sup>34</sup>. The study provided <sup>13</sup>C-based flux data and RT-PCR mRNA abundances for the central carbon metabolism, pentose phosphate pathway (PPP), and the tricarboxylic acid (TCA) cycle for wild-type K12 *E. coli* culture in chemostat under different dilution rates (0.1, 0.2, 0.4, 0.5, and 0.7 hours<sup>-1</sup>) and for 24 single-gene perturbations along the glycolysis and PPP <sup>34</sup>. The global transcriptional response was only captured for 5 of the 24 single-gene deletions (*pgm*, *pgi*, *gapC*, *zwf* and *rpe*) and two of the 4 dilution conditions (0.5 and 0.7 hours<sup>-1</sup>). The differential (fold-change) gene expression levels were computed with respect to the control condition, set to be wild-type K12 *E. coli* cultured at a dilution rate of 0.2 h<sup>-1</sup>. The differential (fold-change) reaction expressions were subsequently evaluated based on the fold-change gene expression using the GPR Max/Min rule available in the COBRA toolbox (MATLAB) <sup>38</sup>.



For samples with only RT-PCR mRNA abundance data, the set of up- and downregulated reactions,  $\mathbf{R}^U$  and  $\mathbf{R}^D$ , included all reactions with fold-change reaction expressions higher than 1 and those with fold-change lower than 1, respectively. In the additional analyses for samples with whole-genome transcriptome data, the set of up- and downregulated reactions,  $\mathbf{R}^U$  and  $\mathbf{R}^D$ , respectively, were taken from the top and bottom 5<sup>th</sup> percentile of the differential reaction expressions. The differences of the measured cell specific glucose uptake rates between perturbed and control experiments were used as constraints.  $\Delta$ FBA was applied using the two-step optimization with the L2 norm minimization, as described above.

The second case study came from a study of *E. coli* growth on 8 different carbon sources performed by Gerosa *et al.* <sup>35</sup>. Unprocessed global transcriptomic data were obtained from ArrayExpress (E-MTAB-3392) <sup>39</sup>, and differential expression analyses between every pair of carbon sources were evaluated using the *Limma* package in R <sup>40</sup>. As before, the fold-change reaction expressions were computed based on fold-change in the global gene expression using the Max-Min GPR rule using COBRA toolbox <sup>38</sup>. The up- and downregulated set of reactions were taken from the top and bottom 5<sup>th</sup> percentile of the differential reaction expressions. In addition, cell culture data on specific growth rates were used to compute the flux change bounds for biomass production rate. The uptake rates of the carbon source changes were also incorporated as constraints. We implemented the two-step optimization of  $\Delta$ FBA using L2 norm minimization.

The third case study came from two studies of skeletal muscle tissue metabolism in type-2 diabetes (T2D) patients by van Tienen *et al.* <sup>41</sup> and Jin *et al.* <sup>42</sup>. The microarray gene expression datasets were obtained from GEO (GSE19420 <sup>41</sup> and GSE25462 <sup>42,43</sup>) and the differential (fold-change) expression of genes for each dataset were computed using the *Limma* package in R <sup>40</sup>. The fold-change reaction expressions were computed based on the differential gene expression using the Max/Min GPR rule. In the absence of additional constraints in the form of exchange fluxes or growth characteristics, we set the up- and downregulated reactions from the top and bottom 25<sup>th</sup>-percentile in differential reaction expressions, rather than the 5<sup>th</sup> percentile threshold used in *E. coli* case studies above, so as to incorporate more differentially expressed transcripts. We implemented an L1-norm

minimization in the second step of  $\Delta$ FBA to reduce computational complexity (time) due to the large number of constraints associated with the differential reaction expressions.

### Implementation of REMI

The method Relative Expression and Metabolomic Integrations (REMI) was developed for predicting individual flux distributions of a pair of conditions ( $v^P$  and  $v^C$ ) using multi-omics dataset. The toolbox was downloaded from <https://github.com/EP-LCSB/remi>. The differential gene expressions in each case study were obtained as described above. The mapping from differential gene expression to the corresponding reaction expressions were done using the procedure detailed in REMI<sup>30</sup>. Briefly, the authors followed the implementation of Fang *et al.*<sup>33</sup> to translate gene expression ratios to obtain reaction expression ratios. When several enzyme subunits are required for a reaction, a geometric mean of expression ratios is chosen to represent the reaction ratio. In the case where multiple isozymes catalyze a reaction, the arithmetic mean of the individual expression ratios of the isozymes is used for the reaction ratio. The set of up- and down-regulated reactions  $R^U$  and  $R^D$  were taken from the computed differential reaction expressions as in  $\Delta$ FBA implementation. Unlike  $\Delta$ FBA, REMI produces solutions for the metabolic fluxes of perturbed  $v^P$  and control condition  $v^C$ . For comparison, we evaluated the flux change predicted by REMI by taking the difference:  $\Delta v = v^P - v^C$ .

### Performance evaluation

The quantitative agreement between the measured flux difference  $\Delta v^M$  and the predicted flux changes  $\Delta v^*$  was assessed by using two accuracy metrics: uncentered Pearson correlation coefficient and normalized root mean square error (NRMSE). The uncentered Pearson correlation coefficient  $\rho$  was computed as follows

$$\rho = \frac{\Delta v^M \cdot \Delta v^*}{\|\Delta v^M\|_2 \|\Delta v^*\|_2} \quad (10)$$

Meanwhile, the NRMSE was according to the following equation – using *tdStats* package in R:

$$NRMSE = \frac{1}{\Delta v_{max}^M - \Delta v_{min}^M} \sqrt{\frac{\|\Delta \mathbf{v}^M - \Delta \mathbf{v}^*\|_2^2}{m}} \quad (11)$$

Besides the quantitative agreement in flux changes, we also evaluated the qualitative agreement by comparing the signs of the flux changes between experimental measurements and predictions. To this end, we discretized the measured and predicted flux changes into +1, 0, and −1, to describe upregulated, no change, and downregulated reactions, respectively. The agreement in the direction of the flux changes was evaluated as the number of correct sign predictions divided by the total number of fluxes.

### Metabolic subsystem enrichment analysis

The flux differences obtained from applying ΔFBA were first filtered according to the directionality of their change. The significantly altered fluxes ( $|\Delta v_i| > \varepsilon$ ) were grouped based on the subsystem to which the fluxes belong. A Fisher exact test (*fisher.test* function in the R-package) was used in determining over-represented subsystems in upregulated (positive change) and downregulated (negative change) fluxes. The statistical significance *p*-values were corrected for multiple hypothesis testing using the *p.adjust* function in R.

## **Results**

### *Escherichia coli* response to genetic and environmental variations

Ishii *et al.*<sup>34</sup> studied the robustness of *E. coli* K12 metabolism in chemostat in response to changes in dilution rates and to gene deletions. The study generated multi-omics data, including transcriptomic, proteomic, metabolomic, and <sup>13</sup>C metabolic fluxes, and demonstrated the remarkable ability of *E. coli* to reroute its metabolic fluxes to maintain metabolic homeostasis in response to environmental and genetic perturbations. Only a small fraction of variation in the measured flux ratios can be explained by the fold-change in reaction expressions, as indicated by the low coefficient of determinations  $R^2$ , regardless of the GPR mapping procedures in ΔFBA and REMI (see **Figure 1A**). The weak agreement between reaction expressions and metabolic fluxes motivates using a system-oriented approach that consider the global network changes based on GEMs<sup>44</sup>. We applied ΔFBA using *E. coli*'s *iJO1366* GEM to predict the metabolic flux shifts from the control condition

(wild-type K12 at 0.2 hour<sup>-1</sup> dilution rate), caused by alterations in dilution rates (0.1, 0.4, 0.5, and 0.7 hours<sup>-1</sup>) and by 24 single-gene deletions (*galM*, *glk*, *pgm*, *pgi*, *pfkA*, *pfkB*, *fbp*, *fbaB*, *gapC*, *gpmA*, *gpmB*, *pykA*, *pykF*, *ppsA*, *zwf*, *pgl*, *gnd*, *rpe*, *rpiA*, *rpiB*, *tktA*, *tktB*, *talA*, and *talB*). We compared the predicted flux changes using  $\Delta$ FBA with the differences of 46 measured metabolic fluxes along the central carbon metabolism by incorporating the enzyme expression obtained from RT-PCR. Uncentered Pearson correlations, NRMSE, and sign accuracy of the flux change predictions are depicted in **Figure 1B** and **C**, indicating a general agreement between the predicted flux changes and the differences in the measured fluxes. We compared the accuracy of flux change predictions by  $\Delta$ FBA and REMI<sup>30</sup>. As illustrated in **Figure 1B** and **C**,  $\Delta$ FBA generally outperforms REMI in predicting the flux changes by having lower NRMSE, higher Pearson correlations, and higher sign accuracy. The results using the whole-genome gene expression profiles for a subset of perturbation experiments are comparable with those using RT-PCR data above (see also **Supplementary Figure S2** and **Supplementary Figure S3**).

Another study, carried out by Gerosa *et al.*<sup>35</sup>, looked at how *E. coli*'s central carbon metabolism adapts to 8 different carbon sources: acetate, fructose, galactose, glucose, glycerol, gluconate, pyruvate and succinate. The study generated <sup>13</sup>C metabolic flux, metabolite concentration and microarray gene expression data from exponentially growing *E. coli* under each carbon source. The study found that only a small subset of the numerous transcriptome changes translates to notable shifts in the corresponding metabolic fluxes, indicating non-trivial relationships between transcriptional regulations and metabolic fluxes. We applied  $\Delta$ FBA to predict flux changes between every pair of the carbon sources, treating one as the perturbation and another as the control condition. **Figure 2** describes the good agreement between the flux change predictions by  $\Delta$ FBA with the measured differences of 34 metabolic fluxes between any two carbon sources, specifically in terms of uncentered Pearson correlation (mean: 0.61), NRMSE (mean: 0.15), and sign accuracy (mean: 0.66). The findings from Ishii *et al.* and Gerosa *et al.* highlight the ability of  $\Delta$ FBA in accurately predicting metabolic flux alterations using transcriptomic data for both environmental (e.g., dilution rates, carbon sources) and genetic perturbations.

## Dysregulation of Skeletal Muscle Metabolism in Type – 2 – Diabetes

In this case study, we looked at metabolic alterations of human muscle using the myocyte GEM *iMyocyte2419*<sup>36</sup> and gene expression datasets from two type-2 diabetes (T2D) studies, one by van Tienen *et al.*<sup>41</sup> and another by Jin *et al.*<sup>42</sup>. In long term T2D patients compared to age-matched cohort, van Tienen *et al.*<sup>41</sup> reported the downregulation of gene expression related to substrate transport into mitochondria, conversion of pyruvate into acetyl-CoA, aspartate-malate shuttle in mitochondria, glycolysis, TCA cycle, and electron transport chain. Similarly, Jin *et al.*<sup>42</sup> reported a significant enrichment of pathways involved the oxidative phosphorylation among the downregulated genes in their T2D cohort compared to control. Jin *et al.*<sup>42</sup> further identified the transcription factor SRF and its cofactor MKL1 among the top-ranking enriched gene sets with increased expression. But, the correlation between the differential gene expressions in the two studies is only modest.<sup>36</sup>

We applied  $\Delta$ FBA to predict the flux changes based on the differential gene expressions in each of the two studies above (see Methods). We grouped the reactions based on whether the predicted flux changes are positive or negative, denoted by up- and down-reactions, respectively. We performed metabolic subsystem enrichment analysis using the subsystems defined in myocyte specific GEM *iMyocyte2419*<sup>36</sup> to identify over-represented metabolic subsystems among the up- and down-reactions (see Methods). As summarized in **Figure 3**, the enrichment analysis of metabolic changes in the van Tienen *et al.* study shows a significant over-representation of  $\beta$ -oxidation and BCAA (branched-chain amino acids) metabolism among the down-reactions, and of extracellular transport and lipid metabolism among the up-reactions. The enrichment analysis of flux changes in the Jin *et al.* study also indicates an over-representation of lipid metabolism among the up-reactions in T2D patients, as well as an over-representation of  $\beta$ -oxidation pathway among the down-reactions (see **Figure 3**).

Furthermore, we evaluated the change in the flux throughput for every metabolite irrespective of its compartmental location, more specifically by computing the change in the total production flux of each metabolite. Metabolites with a large change in the flux throughput are of particular interest for disease biomarkers. In the following, we focus on metabolites that have a flux throughput change above a threshold ( $|\Delta v_i| > \varepsilon$ ;  $\varepsilon = 1$ ), and capture intermediary metabolites that participate in linear reaction sequences. **Figure 4**

shows the flux throughput changes predicted by  $\Delta$ FBA for various metabolites. Among the metabolites with a large drop in the flux throughput in both studies are Coenzyme A (CoA), Acetyl-CoA and AMP (Adenosine monophosphate), all of which have been previously identified as metabolite reporters of diabetes<sup>36,45</sup>. Other metabolic biomarkers that have been previously proposed for T2D, such as repression of FAD (Flavin adenine dinucleotide), FADH<sub>2</sub> and NADH by van Tienen *et al.* study<sup>41</sup> and increased glycerol by Jin *et al.* study<sup>42</sup>, are confirmed by  $\Delta$ FBA (see **Figure 4**). Våremo *et al.*<sup>36</sup> had also identified the metabolites above as T2D reporters via meta-analysis of numerous T2D datasets—including the two studies used here—and topology-based analysis using the GEM *iMyocyte2419*. Besides these confirmatory observations, our  $\Delta$ FBA results further suggest that arachidonate and palmitate are candidate metabolic reporters for T2D, both of which have a significant positive flux throughput change in the two T2D studies (see **Figure 4**). The roles of arachidonate and palmitate in the progression and cause of T2D have also been previously investigated<sup>46–48</sup>. The results above showcase the ability of  $\Delta$ FBA in elucidating metabolic flux alterations in a complex human GEM and identifying key metabolites of interest in human diseases.

## Discussions

GEMs and constraint-based modeling using FBA and the myriad FBA variants have proven to be important enabling tools for establishing genotype-phenotype relationship<sup>10,49,50</sup>. The increasing availability of omics data have driving the development of FBA-based strategies that are able to use such data to improve the accuracy of predictions of intracellular metabolic fluxes. In this work, we present a new FBA-based method, called  $\Delta$ FBA, built for the purpose of analyzing the metabolic alterations between two conditions given data on differential gene expression.  $\Delta$ FBA does not require the specification of the metabolic objective, and thus, eliminates any potential pitfalls that are associated with an incorrect selection of this objective. Note that  $\Delta$ FBA does not generate the flux prediction for a given condition; rather, the method produces differences of metabolic fluxes between two conditions. Differential flux predictions are indispensable in formulating hypothesis and in understanding the physiological response of cells to changes in the environment.  $\Delta$ FBA can be easily integrated and have been tested to work with the widely popular COBRA toolbox.

We showed the applicability and performance of  $\Delta$ FBA for predicting metabolic flux changes in an array of experimental perturbations and in both simple prokaryotic *E. coli* and complex multicellular human muscle cells. In comparison to a related method REMI<sup>30</sup>,  $\Delta$ FBA show a markedly better accuracy in prediction the magnitude and direction of metabolic flux changes in *E. coli*. Further, the application of  $\Delta$ FBA to two T2D studies shed light on the rewiring of muscle metabolism associated with type-2 diabetes that leads to the repression of  $\beta$ -oxidation and activation of glycerolphospholipids, pointing to increased lipid metabolism in the T2D patients. Interestingly, serum metabolic profiling of T2D patients showed increased glycerophospholipids when compared to healthy controls<sup>51</sup>. Besides, clinical and experimental studies have demonstrated the association between phospholipids and insulin resistance<sup>52</sup>. Furthermore, by looking at the changes in the flux throughput of metabolites, the results of  $\Delta$ FBA suggest two fatty acids, arachidonate and palmitate, for candidate biomarkers of T2D.

There are several limitations of  $\Delta$ FBA, the most obvious of which is that the method does not produce flux predictions for the conditions under comparison (perturbed vs. control). If the values of the metabolic fluxes are desired, one can use a FBA-based method, for example parsimonious FBA<sup>53</sup>, to evaluate metabolic fluxes for one of the conditions (perturbed or control) – preferably one that is more well characterized (e.g., more experimental data, more obvious metabolic objective) – and compute the metabolic fluxes for the other condition by combining this flux prediction and the flux changes from  $\Delta$ FBA. Also, in the formulation and the application of  $\Delta$ FBA in this work, we considered only differential gene expression data. But the method can also accommodate other omics dataset, such as proteomics, by appropriate mapping of the data to changes in reaction expressions. Metabolomics data can also be accommodated in  $\Delta$ FBA via thermodynamics constraints, as done in REMI<sup>30</sup>, in which certain reactions can only proceed in one direction.

## Conclusion

In this work, we address the challenge of studying metabolic flux alterations in organisms as a result of genetic or environmental changes. Our versatile method  $\Delta$ FBA provides a set of

functions for evaluating metabolic flux differences between two conditions using genome-scale metabolic models and differential gene expression data. The computational tool eliminates the need for assuming a metabolic objective by exploiting the fact that the flux differences also satisfy the same flux balance equation used in standard FBA. As demonstrated in several case studies,  $\Delta$ FBA provides accurate and biologically relevant predictions of metabolic alterations caused by environmental, genetic, and disease-related perturbations. With increasing research efforts directed toward the integration of omics data with biochemical network models, tools such as  $\Delta$ FBA represents an important advancement in this direction. The MATLAB implementation of  $\Delta$ FBA is freely available on <https://github.com/CABSEL/DeltaFBA>.

## Acknowledgments

The authors would like to acknowledge the Swiss National Science Foundation for funding support (grant number: 163390) and the University at Buffalo's Center for Computational Research for computational support.

## References

1. Granata, I., Troiano, E., Sangiovanni, M. & Guarracino, M. R. Integration of transcriptomic data in a genome-scale metabolic model to investigate the link between obesity and breast cancer. *BMC Bioinformatics* **20**, 162 (2019).
2. Agren, R. *et al.* Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol Syst Biol* **10**, 721 (2014).
3. Nilsson, A. & Nielsen, J. Genome scale metabolic modeling of cancer. *Metabolic Engineering* **43**, 103–112 (2017).
4. Lewis, N. E. & Abdel-Haleem, A. M. The evolution of genome-scale models of cancer metabolism. *Front. Physiol.* **4**, 237 (2013).
5. Zhang, C. & Hua, Q. Applications of Genome-Scale Metabolic Models in Biotechnology and Systems Medicine. *Front Physiol* **6**, 413 (2015).
6. Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Current status and applications of genome-scale metabolic models. *Genome Biol* **20**, 121 (2019).



7. Oberhardt, M. A., Palsson, B. & Papin, J. A. Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology* **5**, (2009).
8. Santos, F., Boele, J. & Teusink, B. A practical guide to genome-scale metabolic models and their analysis. in *Methods in Enzymology* **500**, 509–532 (Academic Press Inc., 2011).
9. Bordbar, A., Monk, J. M., King, Z. A. & Palsson, B. O. Ø. Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* **15**, 107–120 (2014).
10. Lewis, N. E., Nagarajan, H. & Palsson, B. O. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology* **10**, 291–305 (2012).
11. Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
12. Park, J. H. & Lee, S. Y. Towards systems metabolic engineering of microorganisms for amino acid production. *Current Opinion in Biotechnology* **19**, 454–460 (2008).
13. Kim, T. Y., Sohn, S. B., Kim, H. U. & Lee, S. Y. Strategies for systems-level metabolic engineering. *Biotechnology Journal* **3**, 612–623 (2008).
14. Nevoigt, E. Progress in Metabolic Engineering of *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **72**, 379–412 (2008).
15. Kim, H. U., Kim, T. Y. & Lee, S. Y. Metabolic flux analysis and metabolic engineering of microorganisms. *Mol. Biosyst.* (2008). doi:10.1039/b712395g
16. Ebrahim, A. *et al.* Do genome-scale models need exact solvers or clearer standards? *Mol. Syst. Biol.* (2015). doi:10.15252/msb.20156157
17. Richelle, A., Chiang, A. W. T., Kuo, C.-C. & Lewis, N. E. Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. *PLOS Comput. Biol.* **15**, e1006867 (2019).
18. Hyduke, D. R., Lewis, N. E. & Palsson, B. O. Analysis of omics data with genome-scale models of metabolism. *Molecular BioSystems* **9**, 167–174 (2013).
19. Becker, S. A. & Palsson, B. O. Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol* **4**, e1000082 (2008).
20. Machado, D., Herrgård, M., Herrgård, M. & Herrgård, M. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of

- metabolism. *PLoS Comput. Biol.* **10**, e1003580 (2014).
21. Zur, H., Rupp, E. & Shlomi, T. iMAT: an integrative metabolic analysis tool. *Bioinformatics* **26**, 3140–2 (2010).
22. Agren, R. *et al.* Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput Biol* **8**, e1002518 (2012).
23. Kim, M. K., Lane, A., Kelley, J. J. & Lun, D. S. E-Flux2 and sPOT: Validated methods for inferring intracellular metabolic flux distributions from transcriptomic data. *PLoS One* **11**, e0157101 (2016).
24. Chandrasekaran, S. & Price, N. D. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A.* (2010). doi:10.1073/pnas.1005139107
25. Jensen, P. A. & Papin, J. A. Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics* (2011). doi:10.1093/bioinformatics/btq702
26. Åkesson, M., Förster, J. & Nielsen, J. Integration of gene expression data into genome-scale metabolic models. *Metab. Eng.* (2004). doi:10.1016/j.ymben.2003.12.002
27. O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R. & Palsson, B. O. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol* **9**, 693 (2013).
28. Sánchez, B. J. *et al.* Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol. Syst. Biol.* **13**, 935 (2017).
29. Salvy, P. & Hatzimanikatis, V. The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. *Nat Commun* **11**, 30 (2020).
30. Pandey, V., Hadadi, N. & Hatzimanikatis, V. Enhanced flux prediction by integrating relative expression and relative metabolite abundance into thermodynamically consistent metabolic models. *PLoS Comput Biol* **15**, e1007036 (2019).
31. Zhu, L., Zheng, H., Hu, X. & Xu, Y. A computational method using differential gene expression to predict altered metabolism of multicellular organisms. *Mol. Biosyst.* **13**, 2418–2427 (2017).

32. Pusa, T. *et al.* MOOMIN - Mathematical explORation of 'Omics data on a Metabolic Network. *Bioinformatics* (2020). doi:10.1093/bioinformatics/btz584
33. Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* **14**, 639–702 (2019).
34. Ishii, N. *et al.* Multiple high-throughput analyses monitor the response of E. coli to perturbations. **316**, (2007).
35. Gerosa, L. *et al.* Pseudo-transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data. *Cell Syst.* (2015). doi:10.1016/j.cels.2015.09.008
36. Våremo, L. *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). doi:10.1016/j.celrep.2015.04.010
37. Segrè, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 15112–15117 (2002).
38. Opdam, S. *et al.* A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models Article A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. *Cell Syst.* **4**, 318–329 (2017).
39. Athar, A. *et al.* ArrayExpress update - From bulk to single-cell expression data. *Nucleic Acids Res.* **47**, D711–D715 (2019).
40. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* (2015). doi:10.1093/nar/gkv007
41. Van Tienen, F. H. J. *et al.* Physical activity is the key determinant of skeletal muscle mitochondrial function in type 2 diabetes. *J. Clin. Endocrinol. Metab.* (2012). doi:10.1210/jc.2011-3454
42. Jin, W. *et al.* Increased SRF transcriptional activity in human and mouse skeletal muscle is a signature of insulin resistance. *J. Clin. Invest.* (2011). doi:10.1172/JCI41940
43. Lerin, C. *et al.* Defects in muscle branched-chain amino acid oxidation contribute to impaired lipid metabolism. *Mol. Metab.* (2016). doi:10.1016/j.molmet.2016.08.001
44. Angione, C. Human Systems Biology and Metabolic Modelling: A Review—From Disease Metabolism to Precision Medicine. *Biomed Res. Int.* **2019**, (2019).
45. Misra, P. & Chakrabarti, R. The role of AMP kinase in diabetes. *Indian Journal of*

- Medical Research* **125**, 389–398 (2007).
46. Ly, L. D. *et al.* Oxidative stress and calcium dysregulation by palmitate in type 2 diabetes. *Exp. Mol. Med.* (2017). doi:10.1038/emmm.2016.157
  47. Igoillo-Esteve, M. *et al.* Palmitate induces a pro-inflammatory response in human pancreatic islets that mimics CCL2 expression by beta cells in type 2 diabetes. *Diabetologia* (2010). doi:10.1007/s00125-010-1707-y
  48. Das, U. N. Arachidonic acid in health and disease with focus on hypertension and diabetes mellitus: A review. *Journal of Advanced Research* (2018). doi:10.1016/j.jare.2018.01.002
  49. King, Z. A., Lloyd, C. J., Feist, A. M. & Palsson, B. O. Next-generation genome-scale models for metabolic engineering. *Current Opinion in Biotechnology* **35**, 23–29 (2015).
  50. O'Brien, E. J., Monk, J. M. & Palsson, B. O. Using genome-scale models to predict biological capabilities. *Cell* **161**, 971–987 (2015).
  51. Xu, F. *et al.* Metabolic signature shift in type 2 diabetes mellitus revealed by mass spectrometry-based metabolomics. *J. Clin. Endocrinol. Metab.* (2013). doi:10.1210/jc.2012-4132
  52. Chang, W., Hatch, G. M., Wang, Y., Yu, F. & Wang, M. The relationship between phospholipids and insulin resistance: From clinical to experimental studies. *Journal of Cellular and Molecular Medicine* (2019). doi:10.1111/jcmm.13984
  53. Lewis, N. E. *et al.* Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* **6**, 390 (2010).

## Figures

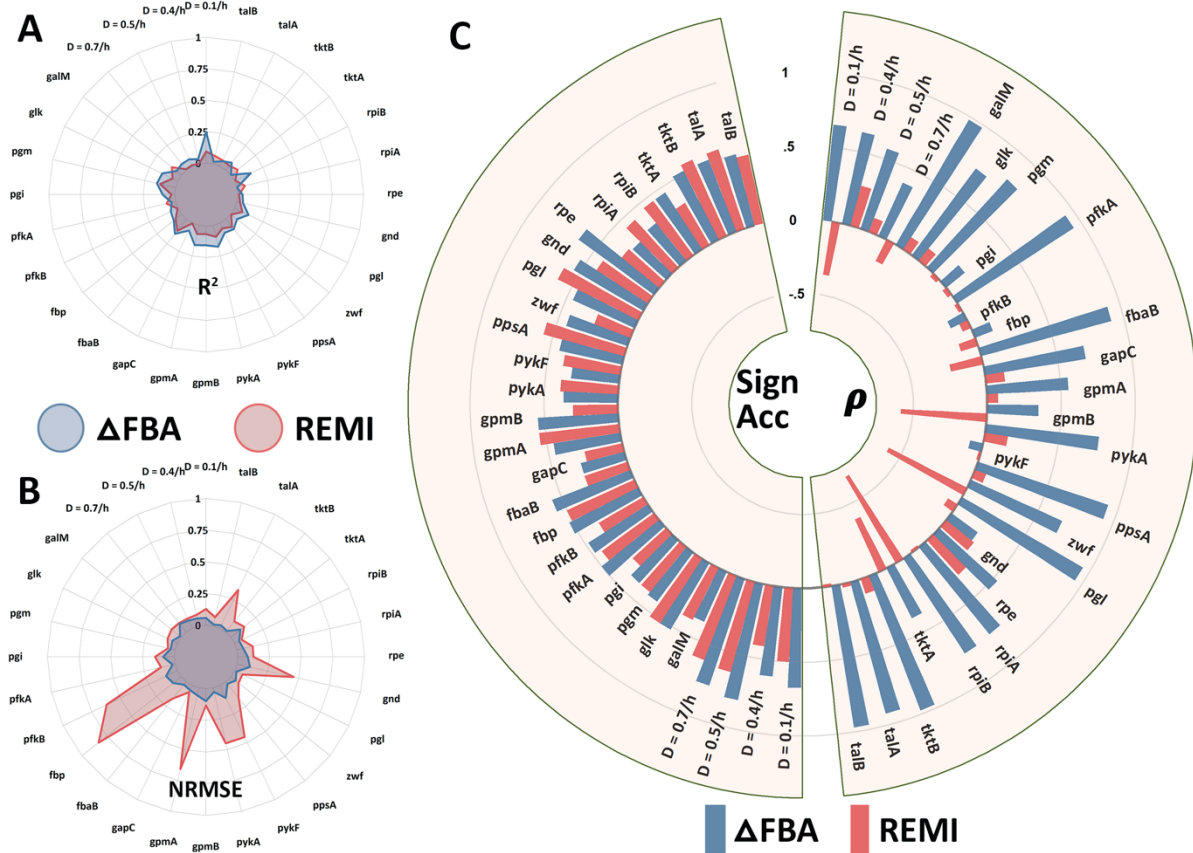


Figure 1: Comparison of the performance of  $\Delta$ FBA and REMI in predicting *E. coli* metabolic response to environmental (dilution rates) and genetic (single gene deletions) perturbations. (A) The coefficient of determination ( $R^2$ ) between the measured flux ratios and the reaction expression ratios shows that differential gene expressions do not directly inform the metabolic flux changes. (B) Normalized Root Mean Square Error (NRMSE) of the predicted flux change – smaller indicates higher accuracy (mean NRMSE:  $\Delta$ FBA = 0.14, REMI = 0.54). (C) Directional (Sign Accuracy) agreement and uncentered Pearson's Correlation Coefficient ( $\rho$ ) between the predicted and measured flux difference for the 46 reactions (mean sign accuracy:  $\Delta$ FBA = 0.49, REMI = 0.43; mean  $\rho$ :  $\Delta$ FBA = 0.61; REMI = -0.06).

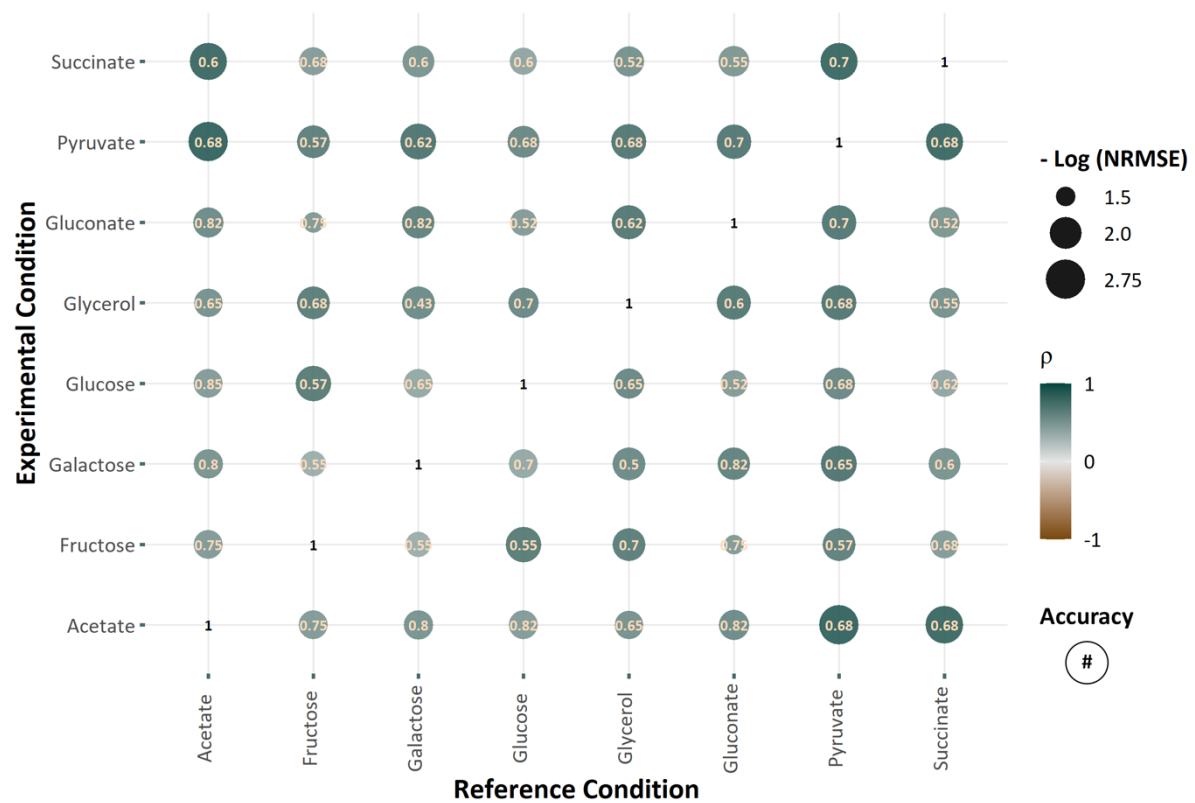


Figure 2: Prediction of metabolic flux changes in *E. coli* caused by changes in the carbon source using  $\Delta$ FBA. The horizontal axis reports the reference carbon source (control) and the vertical axis shows the altered (perturbed) carbon source. Uncentered Pearson's Correlation Coefficient ( $\rho$ ) is shown by the color of the markers. NRMSE is represented by the size of the markers – the larger the markers, the smaller is the NRMSE. Finally, the directional (sign) accuracy of the flux perturbation predictions is shown by the numbers inside the markers.

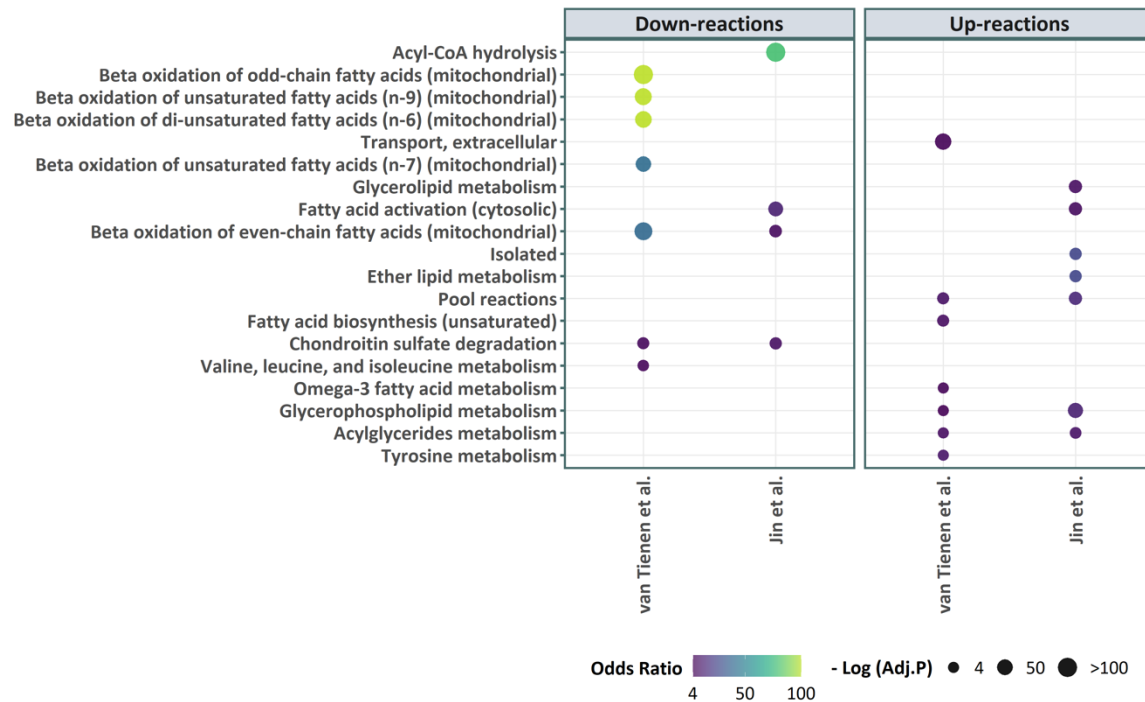


Figure 3: Enriched metabolic subsystems (FDR<0.05) among the in T2D patients based on flux changes predicted using  $\Delta$ FBA. The flux changes were computed based on the transcriptome datasets from two T2D studies: van Tienen *et al.*<sup>41</sup> (GSE19420) and Jin *et al.*<sup>42</sup> (GSE25462). The statistical significance of the over-representation is shown by the size of the markers – larger markers have smaller adjusted  $p$ -values – while the odds ratio is shown by the color of the markers.

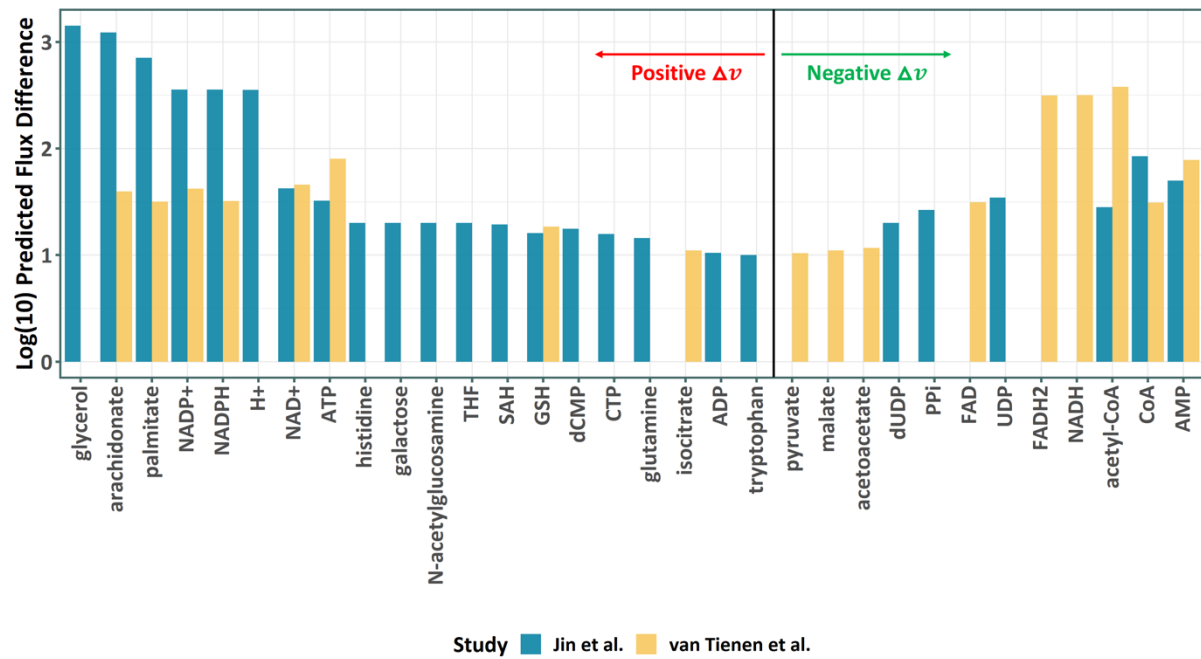


Figure 4: Alterations in metabolite flux throughput in T2D patients as predicted by  $\Delta$ FBA.