Computational and Structural Biotechnology Journal 11 (2014) 59–65

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Mini Review

# Methods for integration of transcriptomic data in genome-scale metabolic models

Min Kyung Kim [a], Desmond S. Lun [a,b,*]

[a] Center for Computational and Integrative Biology, Rutgers University, Camden, NJ 08102, USA
[b] Phenomics and Bioinformatics Research Centre and School of Mathematics and Statistics, University of South Australia, Mawson Lakes, SA 5095, Australia

ARTICLE INFO

Available online 3 September 2014

*Keywords:*
Flux balance analysis
Contraint-based model
Omics

ABSTRACT

Several computational methods have been developed that integrate transcriptomic data with genome-scale metabolic reconstructions to infer condition-specific system-wide intracellular metabolic flux distributions. In this mini-review, we describe each of these methods published to date with categorizing them based on four different grouping criteria (requirement for multiple gene expression datasets as input, requirement for a threshold to define a gene's high and low expression, requirement for a priori assumption of an appropriate objective function, and validation of predicted fluxes directly against measured intracellular fluxes). Then, we recommend which group of methods would be more suitable from a practical perspective.



**Contents**



## 1. Introduction

Intracellular metabolic reactions provide a cell with basic biochemical building blocks, energy, and a thermodynamically favorable environment to sustain its life. Because of the large connectivity inherent to metabolic networks via metabolites participating in multiple metabolic reactions, determination of system-level changes in intracellular metabolic fluxes of organisms is important for understanding the fundamental mechanisms of their metabolic responses to environmental or genetic perturbations [1,2].

$^{13}$C metabolic flux analysis ($^{13}$C-MFA) allows intracellular fluxes to be quantified experimentally. In this approach, cells are grown on $^{13}$C-labeled substrates until the cells are at both metabolic steady state (i.e. when concentrations of metabolites remain stable over time) and isotopic steady state (i.e. when the isotope label is distributed throughout the network, and all isotopomer fractions are constant over time). Then the level of $^{13}$C enrichment in metabolites of the cells is measured by mass spectrometry (MS) or nuclear magnetic resonance (NMR). Intracellular flux distribution is reconstituted from the $^{13}$C enrichment patterns [3–8]. System-wide quantification of intracellular metabolic fluxes using $^{13}$C-MFA, however, is challenging not only because of the extensive instrumentation required but also because of the limited number of fluxes and conditions that can be experimentally measured. Typically, $^{13}$C-MFA focuses on central carbon metabolism [7–10].

An alternative method that is widely used for system-level studies of metabolism is a computational modeling approach called flux balance analysis (FBA). FBA predicts metabolic flux distributions at steady state by making use of in silico genome-scale metabolic models [11]. These genome-scale metabolic models are assembled and manually-curated from annotated genome, biochemical, genetic, and cell phenotype data [11–13]. To use FBA, a genome-scale metabolic model is converted into a $m \times n$ stoichiometric matrix, $S$, where the rows in $S$

* Corresponding author at: Center for Computational and Integrative Biology, Rutgers University, Camden, NJ 08102, USA. Tel.: +1 856 225 6094; fax: +1 856 225 6624.
*E-mail addresses:* mk1034@rutgers.edu (M.K. Kim), dslun@rutgers.edu (D.S. Lun).

http://dx.doi.org/10.1016/j.csbj.2014.08.009

correspond to the $m$ metabolites of the metabolic network, and the columns represent the $n$ reactions (Fig. 1a). Each matrix element $s_{ij}$, indicates a stoichiometric coefficient, that is, the number of molecules of the $i$th metabolite participating in the $j$th reaction. $s_{ij} = 0$ means that the $i$th metabolite is not involved, and a positive or a negative $s_{ij}$ indicates that the $i$th metabolite is a product or a reactant of the $j$th reaction, respectively. Under the steady state assumption, the metabolic flux distribution can be represented mathematically by $S \cdot v = 0$, where $v$ is a column vector whose elements are the unknown reaction rates (fluxes) through each of the reactions of $S$ (Fig. 1b). Since genome-scale metabolic models include all possible metabolic reactions implied by the genome annotation regardless of whether the annotated metabolic genes are expressed in a given environment, the resulting system $S \cdot v = 0$, is in general underdetermined [14,15]. Thus, physiologically meaningful flux solutions need to be narrowed down from all the possible flux distributions by imposing additional constraints on the system and by optimizing certain objective functions when performing FBA (Fig. 1c) [16]. The standard FBA involves solving the following linear optimization problem:

$$\max \; f'v \quad \text{subject to} \begin{cases} Sv = 0 \\ lb \le v \le ub \end{cases} \tag{1}$$

where $v$ is a flux vector representing the reaction rates of the $n$ reactions in the network, $f$ is a coefficient vector defining the organism's objective function, $S$ is the stoichiometric matrix, and $lb$ and $ub$ are the minimum and maximum reaction rates through each reaction in $v$.

If the complete regulatory structure of an organism were known, it would be possible to produce context-specific constraints by computing which cellular components may be expressed in a given condition. However, the regulatory structure is unknown even for the relatively simple and extensively-studied bacterium, *Escherichia coli*, partly due to the lack of comprehensive transcription unit information and because of the lack of information on the relationship between genotype and phenotype [17].

Recent advances in omics technologies have enabled quantitative monitoring of the abundance of biological molecules at various levels in a high-throughput manner [18]. In the absence of complete information on regulatory rules, omics data can be integrated with genome-scale metabolic models to improve their predictive power [19,20]. For this purpose, transcriptomic data, i.e. genome-wide mRNA expression profiling data, is useful in some points compared to other omics platforms. Fluxomics (i.e. $^{13}$C-MFA) is the most direct measurement of metabolic phenotype, but has the disadvantages in that it is difficult to make measurements and only a limited number of fluxes can be determined as mentioned above. Metabolomics can also be useful, but typically fluxes are more informative than metabolite concentrations themselves, and it is challenging to determine fluxes from metabolite concentrations partly because each metabolite participates in multiple metabolic reactions. Similar to fluxes, specific classes of metabolites such as lipids or labile chemicals easily metabolized are still demanding to measure [21,22]. Unlike the first two omics data that cover a small share of all reactions in a genome-scale model, transcriptomics and proteomics are the platforms where a quantitative snapshot of molecular species at system-level is currently possible [23]. However, proteomics is a relatively immature technology compared to transcriptomics. The accuracy with which protein concentrations can be determined is much lower than that with which mRNA concentrations can be determined. On the other hand, RNA amount changes can be precisely measured in a highly automated process at low cost in comparison with the amount of data gathered [24,25]. By integrating transcriptomics data

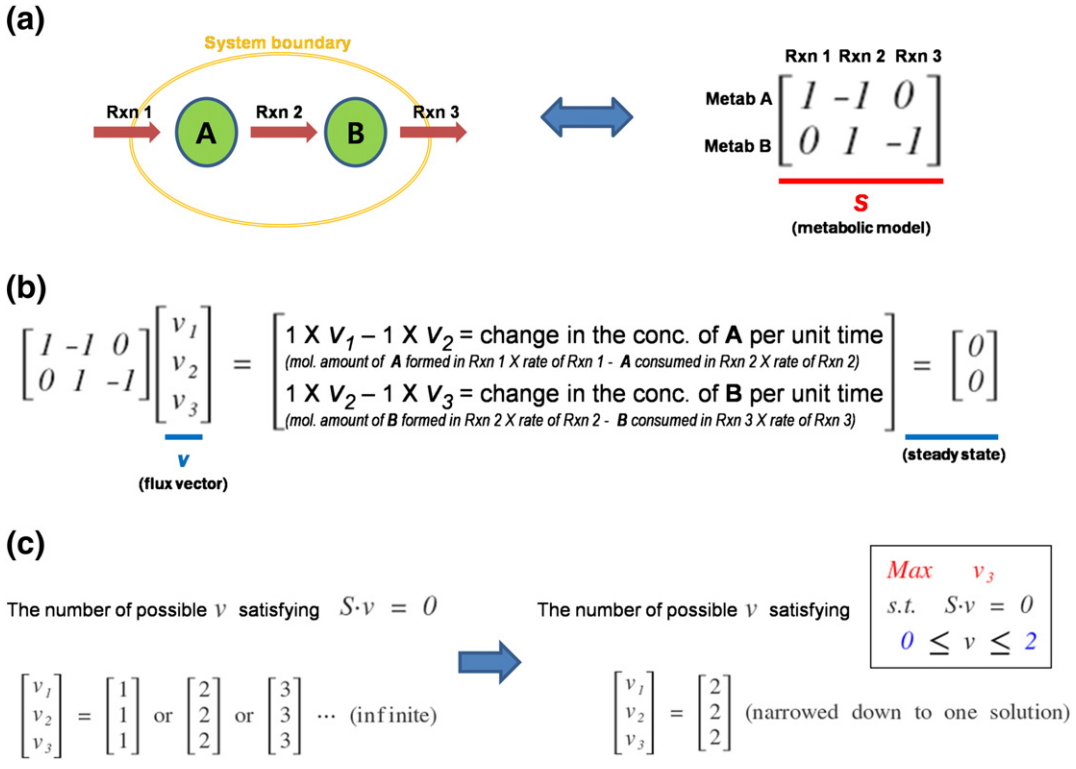


**Fig. 1.** Flux balance analysis (FBA). This figure illustrates how FBA works with an example of the simple network below consisting of two metabolites, A and B, and three metabolic reactions. (a) To use FBA, the network is converted into a stoichiometric matrix, $S$, where the rows in $S$ correspond to the metabolites of the metabolic network, and the columns represent the reactions. Each matrix element $s_{ij}$, indicates a stoichiometric coefficient, that is, the number of molecules of the $i$th metabolite participating in the $j$th reaction. $s_{ij} = 0$ means that the $i$th metabolite is not involved, and a positive or a negative $s_{ij}$ indicates that the $i$th metabolite is a product or a reactant of the $j$th reaction, respectively. (b) Under the steady state assumption, the metabolic flux distribution can be represented mathematically by $S \cdot v = 0$, where $v$ is a column vector whose elements are the unknown reaction rates (fluxes) through each of the reactions of $S$.(c) Since the resulting system, $S \cdot v = 0$, is usually underdetermined, physiologically meaningful flux solutions need to be narrowed down from all the possible flux distributions by imposing additional constraints on the system (e.g. $0 \le v \le 2$ in the figure) and by optimizing certain objective functions (e.g. *Max* $v_3$ in the figure).

with genome-scale metabolic models, we can potentially determine metabolic fluxes through a relatively simple and low-cost omics technology. If other omics technology especially proteomics technology becomes as mature (e.g. wide coverage at lower cost with less effort) as that of transcriptomics, most of the methods introduced in this paper could be applied to other omics data, too.

Not only do genome-scale models benefit from transcriptomic data in creating condition- and tissue-specific models, but transcriptomic data itself can also benefit by being integrated onto the models. Although a large amount of transcriptomic data is continuously being generated, gaining meaningful insight into the functioning of cellular processes from mRNA levels is challenging because of the functional layers in between the two, such as translation, post-translational modifications, mRNA/protein degradation, and enzyme activity regulation by effectors (inhibitors or activators) [14,23,26]. Genome-scale metabolic models are well-suited to inferring metabolic phenotype from genotype using transcriptomic data, since the models are comprehensive repositories of biochemical data for organisms that enable the description of gene–protein–reaction relationships [13,19]. Whereas correlations between mRNA and fluxes have been often found to be poor, approaches taking into account for the large connectivity of metabolites inherent to metabolic networks have been successful in linking gene expression level to metabolites [2,27–29]. This implies that the consideration of the metabolic network is essential to draw a predictive relation from transcript abundances to fluxes [23].

For these reasons, there have been previous studies to integrate transcriptomic data with genome-scale metabolic models, and some of these methods have been covered in recent reviews [12,14,18,30–33]. However, most of these reviews broadly introduce methods inferring metabolic fluxes from various kinds of omics data and are not focused specifically on transcriptomic data. In addition, some of the reviews do not include the most recent methods since transcriptomic data-driven metabolic modeling methods are being developed at a fast pace [33]. In this mini-review, we focus on introducing methods for integrating transcriptomic data in genome-scale metabolic models, and we give a brief description of each one published to date. We exclude methods that require multi-omics datasets as input for an analysis even if they use transcriptomic data, because multi-omics studies are not common [34–37]. We categorize all methods that are covered in this paper based on four different grouping criteria, and we evaluate which group of methods is more suitable from a practical perspective. Lastly, we discuss several limitations of existing methods that new methods need to overcome.

## 2. Grouping criterion 1: requirement for multiple gene expression datasets as input

As the first criterion, methods for estimating metabolic flux from transcriptomic data can be grouped by how many gene expression datasets are required as input. There are two representative methods that need multiple transcriptomic datasets measured under two or more conditions for an analysis.

First, Probabilistic Regulation Of Metabolism (PROM) published in 2010 is a method that integrates regulatory and metabolic networks [38]. It calculates the probability of a metabolic target gene being expressed relative to the activity of its regulating transcription factor from a large dataset of gene expression data, and the flux maxima of the metabolic reaction associated with the metabolic target gene is constrained by a factor of this probability (Fig. 2a). It has several advantages such as its ability to account for the presence of noise in the data, and to differentiate between a strong transcriptional regulator and a weak one. However, this method requires a large number of experimental datasets to calculate the probability of regulatory interactions between transcription factors and their target genes. It also requires a priori knowledge on transcription factor–target gene pairs. In the original paper, around 1300 microarrays and 2000 transcription factor–target interactions were used for *E. coli* and *Mycobacterium tuberculosis*.

Second, Metabolic Adjustment by Differential Expression (MADE) published in 2011, was developed to overcome the issue of selecting a subjective user-supplied threshold in defining a gene's high and low expression states [39]. MADE creates a sequence of binary expression states using several datasets for differential gene expression so as to find the model that most closely reproduces the observed expression changes (Fig. 2b). The principle of this method is that if the activity of a gene drastically changes from one condition to the other, the flux through the reaction controlled by that gene will change accordingly [40]. Using this method, the authors examined the metabolic effects of the transition from glucose- to glycerol-based growth in *Saccharomyces cerevisiae* over the course of time. They showed that the binary expression state changes calculated by MADE matched 98.7% of the feasible observed gene expression transitions (83.5% of all expression transitions). They also showed that, accompanied by these expression state changes, the flux variability of the model was increased after the shift to glycerol.

The other methods described below use a single gene expression dataset for each experimental condition. One of the possible concerns of using a single transcriptomic dataset may be the lack of proportionality between transcript and flux levels. Accounting for relative gene expression changes from multiple datasets as an indicator of the flux reconfiguration might seem to provide a more meaningful description. However, a recent research paper shows that the methods that use relative expression levels does not necessarily give more accurate flux predictions [33]. Although both methods have advantages, the requirement for multiple sets of input data such as transcription regulatory information or different gene expression datasets to perform the analysis is more onerous from a practical point of view.

## 3. Grouping criterion 2: requirement for a threshold to define a gene's high/low expression

As the second criterion, methods can be grouped by whether they use a user-supplied threshold. Some methods require discretization (e.g. −1, 0, 1), binarization (e.g. 1, 0), or classification (e.g. below/above threshold) of gene expression measurement data according to user-defined arbitrary thresholds to distinguish active and inactive states of the corresponding reactions. In addition to PROM, which is mentioned in the previous section, the following three methods also require thresholds.

An approach suggested by Åkesson et al. in 2004 is one of the earliest methods to integrate genome-wide expression data into genome-scale metabolic models [41]. In this method, the fluxes of reactions whose corresponding genes are not expressed are constrained as zero (Fig. 2c). A probe set for a gene is considered absent if it is undetected in all three replicates from independent cultures of the same condition. Using this principle, they combined microarray measurements of gene expression from chemostat and batch cultivations of *S. cerevisiae* with a genome-scale model for yeast, *i*FF708 [42]. The computed metabolic flux distributions were compared to experimental values from $^{13}C$-labeling experiments. The integration of expression data resulted in improved predictions of metabolic behavior in batch culture. Due to the Boolean nature of this method, failure in correctly detecting presence of lowly expressed genes may give rise to erroneous predictions.

Gene Inactivity Moderated by Metabolism and Expression (GIMME) introduced in 2008, creates a context-specific metabolic model that predicts the subset of reactions a cell is likely to use under particular conditions using gene expression data [43]. This method consists of a two-step procedure (Fig. 2d). First, the method finds a flux distribution that optimizes a given biological objective such as growth and/or ATP production using FBA. Then, the method minimizes the utilization of 'inactive' reactions whose corresponding mRNA transcript levels are below a given

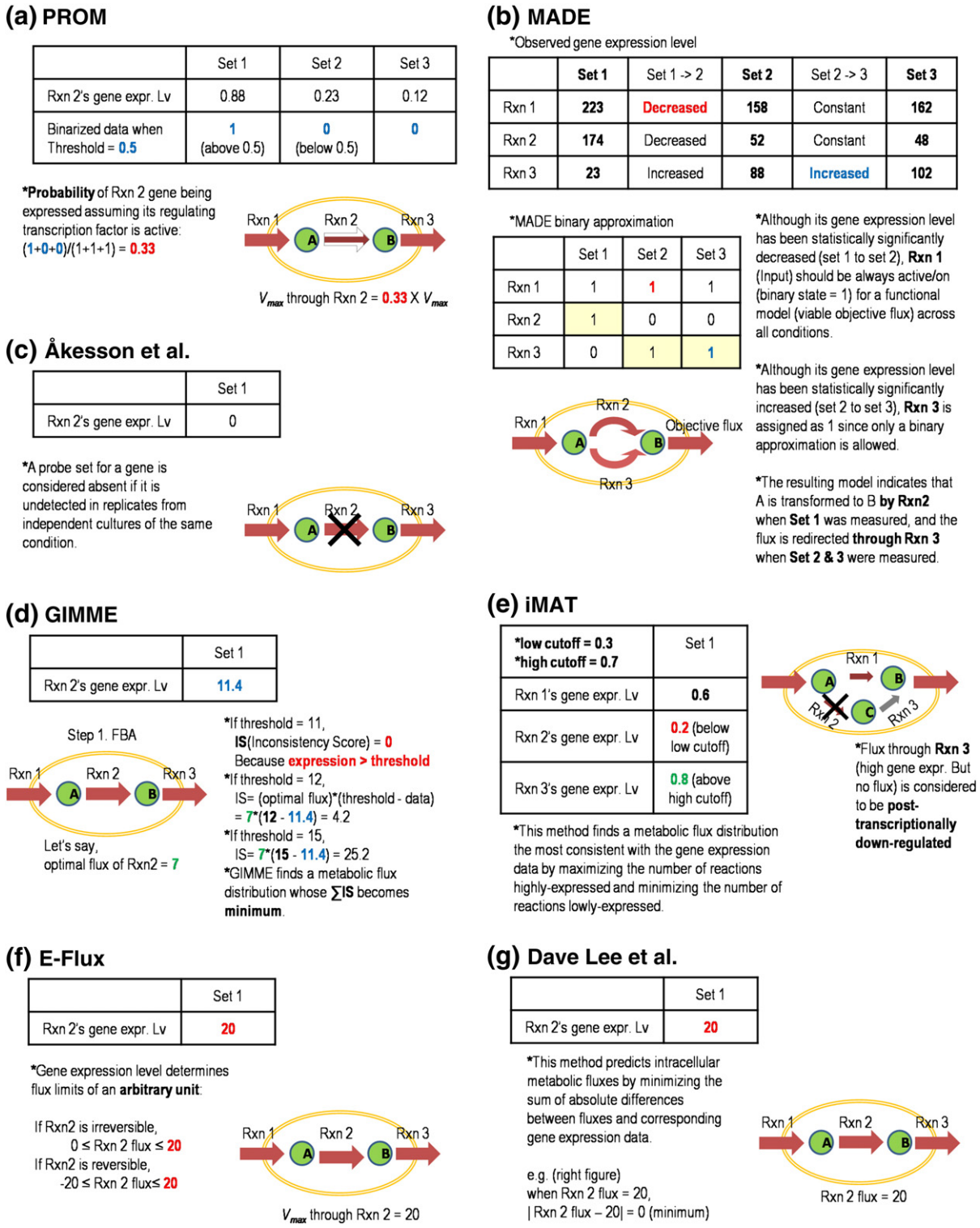


**Fig. 2.** Representative methods currently available for integration of transcriptomic data in genome-scale metabolic models. (a)–(g) show how each method integrates gene expression data onto the models. (a) PROM binarizes the gene expression data according to a user-supplied threshold. Then, it calculates the probability of a metabolic target gene being expressed relative to the activity of its regulating transcription factor from a large dataset of gene expression data. The flux maxima of the metabolic reaction associated with the metabolic target gene is constrained by a factor of this probability. (b) MADE creates a sequence of binary expression states using several datasets for differential gene expression so as to find the model that most closely reproduces the observed expression changes. (c) Åkesson's method is one of the earliest methods to integrate genome-wide expression data into genome-scale metabolic models. In this method, the fluxes of reactions whose corresponding genes are not expressed are constrained as zero. (d) GIMME consists of a two-step procedure. First, the method finds a flux distribution that optimizes a given biological objective such as growth and/or ATP production using FBA. Then, the method minimizes the utilization of 'inactive' reactions whose corresponding mRNA transcript levels are below a given threshold. (e) iMAT discretized gene expression data into tri-valued expression states, representing either low, moderate or high expression in the condition studied according to a user-specified threshold. Then, the method finds an optimal metabolic flux distribution that is the most consistent with the discrete gene expression data by maximizing the number of flux-carrying reactions associated with highly expressed enzymes and minimizing the number of flux-carrying reactions that correspond to lowly-expressed enzymes. (f) E-Flux maps continuous gene expression levels into flux bound constraints according to gene–protein–reaction (GPR) associations. It uses transcriptomic data to set upper and lower bounds on metabolic fluxes so that reactions associated with more highly expressed genes will be allowed to have higher absolute flux values. (g) Dave Lee's method uses transcriptomic data in the objective function. This method predicts intracellular metabolic fluxes by minimizing the deviation between the flux distribution and the transcriptomic data. The deviation was calculated by the sum of absolute differences between fluxes and corresponding gene expression data.

threshold. By avoiding the use of below-threshold reactions that are inconsistent with the flux distribution of the first step, the method was used to find context-specific metabolic flux distributions that best fit physiological data in *E.coli* and human skeletal muscle cells.

The integrative Metabolic Analysis Tool (iMAT) implements a method proposed by Shlomi et al. in 2008, which was developed for tissue-specific modeling of metabolism in mammalian cells [44,45]. In this method, gene expression data is discretized into tri-valued expression states, representing either low, moderate or high expression in the condition studied according to a user-specified threshold (Fig. 2e). Then, iMAT finds an optimal metabolic flux distribution that is the most consistent with the discrete gene expression data by maximizing the number of flux-carrying reactions associated with highly expressed enzymes and minimizing the number of flux-carrying reactions that correspond to lowly-expressed enzymes. This method does not require information on biomass composition or metabolite exchange. By integrating transcriptomic data with a global human metabolic model using this method, they predicted tissue-specific metabolic activity in ten different tissues. A method called EXAMO (EXploration of Alternative Metabolic Optima) is an extended version of iMAT that builds a context-specific model [46].

Tailored gene expression using user-defined thresholds may avoid data normalization issues [33]. However, using arbitrary thresholds may lead to subjective results that lose the fine-grained information for individual genes. This is because the specific threshold above which the level of gene expression indicates physiological activeness of corresponding reactions may vary across genes, conditions, or organisms. The following two methods incorporate continuous gene expression values without using thresholds.

E-Flux (as a combination of flux and expression) published in 2009 is a method that maps continuous gene expression levels into flux bound constraints according to gene–protein–reaction (GPR) associations [47,48]. It uses transcriptomic data to set upper and lower bounds on metabolic fluxes so that reactions associated with more highly expressed genes will be allowed to have higher absolute flux values (Fig. 2f). The rationale behind E-flux is that, given a limited translational efficiency and a limited accumulation of enzyme over the time, the level of mRNA can be used as an approximate upper bound on the maximum amount of metabolic enzymes, and hence as a bound on reaction rates. Using this method, the authors correctly predicted decreased mycolic acid synthesis by seven of the eight known fatty acid inhibitors in *M. tuberculosis*. In a follow-up study [48], they identified preferred carbon sources of *E. coli* that are not influenced by expression derived constraints.

An approach suggested by Lee et al. uses transcriptomic data in the objective function [49]. This method predicts intracellular metabolic fluxes by minimizing the deviation between the flux distribution and the transcriptomic data (Fig. 2g). The deviation was calculated by the sum of absolute differences between fluxes and corresponding gene expression data. The assumption behind this method is that enzymatic transcript concentrations and metabolic fluxes can be related to each other, albeit in a complex manner, since the existence of a transcript is necessary but not sufficient for the presence or activity of its corresponding enzyme [50]. They compared this method against FBA, GIMME, and iMAT, showing a better accuracy in predicting experimentally measured exometabolic flux for *S. cerevisiae* cultures under two growth conditions. FALCON (Flux Assignment with Least absolute deviation Convex Objectives and Normalization) is a recently published, related method with improvements in time efficiency [51].

## 4. Grouping criterion 3: requirement for a priori assumption of an appropriate objective function

The third feature that can distinguish the methods is whether a method requires the a priori assumption of an appropriate biological objective function.

Except for the method of Lee et al. and iMAT, the other methods described here need a priori knowledge of an appropriate objective function of the system such as biomass production rate. The biomass flux (i.e. the growth rate) is the most widely used objective function for FBA optimization problems since it is commonly assumed that, under given resources, efficient growth of a certain microorganism compared to its competitors is beneficial for its survival from an evolutionary perspective [52,53]. Indeed, the assumption of biomass flux maximization in FBA has successfully predicted metabolic behavior of various organisms in a number of studies [54,55]. Nevertheless, biomass flux may be unsuitable as an objective function for some organisms such as microorganisms with variable biomass composition, pathogens in dormancy or in latent phase, or cells of a multi-cellular organism [56]. Thus, in practical applications, we sometimes need methods like the method of Lee et al. and iMAT whose objective functions can be universally applied to a variety of organisms in cases where knowledge of the biological objective function is uncertain.

## 5. Grouping criterion 4: validation of predicted fluxes directly against measured intracellular fluxes

The last distinction among the methods is the utilization of measured intracellular fluxes for the purpose of validation. Basically, the output of the methods described here is predicted intracellular metabolic flux distribution. With the exception of the method of Åkesson et al., none of these methods, however, have tested their predictive accuracy against experimentally measured intracellular fluxes. Lee et al. did attempt to validate their predictions for the intracellular fluxes indirectly using exometabolomic data by measuring changes in the concentration of extracellular metabolites. Nevertheless, considering that detailed information on the underlying mechanisms of metabolic responses is not accessible from extracellular physiological data, it would be preferable to validate predictive accuracy using measured intracellular fluxes [57].

Table 1 summarizes the features of the presented methods with regard to the four grouping criteria described so far.

## 6. Summary and outlook

Given its many advantages, the integration of transcriptomic data in a genome-scale model is a promising method for predicting system-level intracellular metabolic fluxes. From a practical perspective, we suggest that an ideal method satisfies all of the following criteria: a method that needs a single gene expression dataset as input; that utilizes continuous gene expression values without using arbitrary thresholds; that can be used even when an appropriate objective function is unknown; and whose predictive accuracy is validated against measured intracellular fluxes data.

Yet none of the surveyed methods satisfies all of the practical conditions. Lee's method seems to be the most practical method among them in that it achieves three of the four criteria for a practically ideal method. An important limitation of the currently available methods including Lee's method is that, except for the Åkesson's method, their predictive accuracy has not been validated directly against experimentally measured intracellular fluxes. Considering that the major purpose of developing these methods is to accurately predict context-specific intracellular metabolic flux distribution, it would be better if existing or new methods prove how accurately they predict intracellular metabolic distribution by comparing their results with in vivo intracellular flux data.

Importantly, the most practical method does not guarantee the best or the most accurate method. The choice of the most appropriate method would depend on various factors such as biological systems of interest, primary objective of study, and the availability of experimental data. For instance, if we study fast-growing microorganisms such as *E. coli* and *S. cerevisiae* of which the assumption of biomass flux maximization

**Table 1**
Summary of the features of previous methods according to four grouping criteria described in this paper. Desirable features from a practical perspective are shaded in green.

| Method | Requirements for multiple transcriptomic datasets as input | Requirement for a threshold to define a gene's high/low expression state | Requirement for a priori assumption of an appropriate objective function | Validation of predicted fluxes directly against measured intracellular fluxes |
|---|---|---|---|---|
| E-Flux | No | No | Yes | No |
| Lee et al. | No | No | No | No |
| Åkesson et al. | No | Yes | Yes | Yes (4 fluxes were used for validation) |
| GIMME | No | Yes | Yes | No |
| iMAT | No | Yes | No | No |
| PROM | Yes | Yes | Yes | No |
| MADE | Yes | No | Yes | No |

in FBA has successfully predicted metabolic behavior, using the methods such as E-Flux and Åkesson's method that need a priori knowledge of an appropriate objective function of the system would not be a problem. However, in order to study a broad range of systems including microorganisms with variable biomass composition, pathogens in dormancy or in latent phase, or cells of a multi-cellular organism, the methods such as Lee's method and iMAT whose objective functions can be universally applied to a variety of conditions are more desirable for such practical applications. In addition, if we focus on examining clear changes in metabolic behavior of a system, and want to avoid data normalization issues, using the methods that require binarized gene expression data would be appropriate. However, if we need to see more finely grained information, and if it is hard to define the specific threshold above which the level of gene expression indicates physiological activeness of corresponding reactions, using the methods which incorporate continuous gene expression values would be useful. Lastly, although PROM is sorted as an impractical method in Table 1 mainly due to its requirements for a large number of experimental datasets with regulatory information, PROM identified knock-out phenotypes for *E. coli* and *M. tuberculosis* with accuracies as high as 95% [38]. Still, as a recent research paper shows that the methods that use multiple gene expression datasets does not necessarily give more accurate flux predictions [33], the requirement for a large amount of input data to perform the analysis, which might make the job more onerous, could be considered as another limitation of some of the existing methods from a practical point of view.

In this paper, we introduced four different grouping criteria which enable to categorize methods for integration of transcriptomic data in genome-scale metabolic models. Based on these criteria, we suggested features of a practically ideal method. Then, we discussed about which group of the existing methods is more suitable to use for different cases from a practical perspective. Considering that none of the surveyed methods satisfies all of the practical conditions, efforts to develop a new method that overcomes the limitations of the existing methods should be continued.

## Acknowledgments

This work was supported in part by the Samsung Advanced Institute of Technology (SAIT) through the Samsung Global Research Outreach (GRO) program.

## References

[1] Stephanopoulos G. Metabolic fluxes and metabolic engineering. Metab Eng 1999;1:1–11.
[2] Zelezniak A, Sheridan S, Patil KR. Contribution of network connectivity in determining the relationship between gene expression and metabolite concentration changes. PLoS Comput Biol 2014;10:e1003572.
[3] Wiechert W. $^{13}$C metabolic flux analysis. Metab Eng 2001;3:195–206.
[4] Zamboni N, Fendt S-M, Rühl M, Sauer U. $^{(13)}$C-based metabolic flux analysis. Nat Protoc 2009;4:878–92.
[5] Beurton-Aimar M, Beauvoit B, Monier A, Vallée F, Dieuaide-Noubhani M, Colombié S. Comparison between elementary flux modes analysis and $^{13}$C-metabolic fluxes measured in bacterial and plant cells. BMC Syst Biol 2011;5:95.
[6] Nielsen J. It is all about metabolic fluxes. J Bacteriol 2003;185:7031–5.
[7] Krömer J, Quek L-E, Nielsen L. $^{13}$C-fluxomics: a tool for measuring metabolic phenotypes. Aust Biochem 2009;40:17–20.
[8] Sauer U. Metabolic networks in motion: $^{13}$C-based flux analysis. Mol Syst Biol 2006;2:62.
[9] Celton M, Sanchez I, Goelzer A, Fromion V, Camarasa C, Dequin S. A comparative transcriptomic, fluxomic and metabolomic analysis of the response of Saccharomyces cerevisiae to increases in NADPH oxidation. BMC Genomics 2012:317.
[10] Winter G, Krömer JO. Fluxomics — connecting 'omics analysis and phenotypes. Environ Microbiol 2013;15:1901–16.
[11] Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol 2010;28:245–8.
[12] Blazier AS, Papin JA. Integration of expression data in genome-scale metabolic network reconstructions. Front Physiol 2012;3(299).
[13] Chavali AK, D'Auria KM, Hewlett EL, Pearson RD, Papin JA. A metabolic network approach for the identification and prioritization of antimicrobial drug targets. Trends Microbiol 2012:113–23.
[14] Hyduke DR, Lewis NE, Palsson BØ. Analysis of omics data with genome-scale models of metabolism. Mol Biosyst 2013;9:167–74.
[15] Reed JL, Famili I, Thiele I, Palsson BO. Towards multidimensional genome annotation. Nat Rev Genet 2006;7:130–41.
[16] Price ND, Reed JL, Palsson BØ. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol 2004;2:886–97.
[17] Cho B-K, Zengler K, Qiu Y, Park YS, Knight EM, Barrett CL, et al. The transcription unit architecture of the *Escherichia coli* genome. Nat Biotechnol 2009;27:1043–9.
[18] Zhang W, Li F, Nie L. Integrating multiple "omics" analysis for microbial biology: application and methodologies. Microbiology 2010;156:287–301.
[19] Palsson B. In silico biology through "omics". Nat Biotechnol 2002;20:649–50.
[20] Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. Nat Rev Microbiol 2012;10:291–305.
[21] German JB, Gillies LA, Smilowitz JT, Zivkovic AM, Watkins SM. Lipidomics and lipid profiling in metabolomics. Curr Opin Lipidol 2007;18:66–71.
[22] Mayr M. Metabolomics: ready for the prime time? Circ Cardiovasc Genet 2008;1:58–65.
[23] Hoppe A. What mRNA abundances can tell us about metabolism. Metabolites 2012:614–31.
[24] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10:57–63.
[25] Malone JH, Oliver B. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol 2011;9:34.
[26] Palsson B, Zengler K. The challenges of integrating multi-omic data sets. Nat Chem Biol 2010;6:787–9.
[27] Patil KR, Nielsen J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proc Natl Acad Sci U S A 2005;102:2685–9.
[28] Bradley PH, Brauer MJ, Rabinowitz JD, Troyanskaya OG. Coordinated concentration changes of transcripts and metabolites in *Saccharomyces cerevisiae*. PLoS Comput Biol 2009;5.
[29] Moxley JF, Jewett MC, Antoniewicz MR, Villas-Boas SG, Alper Alper H, Wheeler RT, et al. Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. Proc Natl Acad Sci U S A 2009;106:6477–82.
[30] Reed JL. Shrinking the metabolic solution space using experimental datasets. PLoS Comput Biol 2012:e1002662.
[31] Joyce AR, Palsson BØ. The model organism as a system: integrating 'omics' data sets. Nat Rev Mol Cell Biol 2006;7:198–210.
[32] Saha R, Chowdhury A, Maranas CD. Recent advances in the reconstruction of metabolic models and integration of omics data. Curr Opin Biotechnol 2014;29C:39–45.

[33] Machado D, Herrgård M. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. PLoS Comput Biol 2014;10:e1003580.
[34] Schmidt BJ, Ebrahim A, Metz TO, Adkins JN, Palsson B, Hyduke DR. GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. Bioinformatics 2013;29:2900–8.
[35] Kim HU, Kim WJ, Lee SY. Flux-coupled genes and their use in metabolic flux analysis. Biotechnol J 2013;8:1035–42.
[36] Kim J, Reed JL. RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. Genome Biol 2012:R78.
[37] Collins SB, Reznik E, Segrè D. Temporal expression-based analysis of metabolism. PLoS Comput Biol 2012;8.
[38] Chandrasekaran S, Price ND. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 2010;107:17845–50.
[39] Jensen PA, Papin JA. Functional integration of a metabolic network model and expression data without arbitrary thresholding. Bioinformatics 2011;27:541–7.
[40] Van Berlo RJP, De Ridder D, Daran JM, Daran-Lapujade PAS, Teusink B, Reinders MJT. Predicting metabolic fluxes using gene expression differences as constraints. IEEE/ACM Trans Comput Biol Bioinform 2011;8:206–16.
[41] Åkesson M, Förster J, Nielsen J. Integration of gene expression data into genome-scale metabolic models. Metab Eng 2004;6:285–93.
[42] Förster J, Famili I, Fu P, Palsson BØ, Nielsen J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res 2003;13:244–53.
[43] Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. PLoS Comput Biol 2008;4:e1000082.
[44] Zur H, Ruppin E, Shlomi T. iMAT: an integrative metabolic analysis tool. Bioinformatics 2010;26:3140–2.
[45] Shlomi T, Cabili MN, Herrgård MJ, Palsson BØ, Ruppin E. Network-based prediction of human tissue-specific metabolism. Nat Biotechnol 2008;26:1003–10.
[46] Rossell S, Huynen MA, Notebaart RA. Inferring metabolic states in uncharacterized environments using gene-expression measurements. PLoS Comput Biol 2013;9.
[47] Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, et al. Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. PLoS Comput Biol 2009;5:e1000489.
[48] Brandes A, Lun DS, Ip K, Zucker J, Colijn C, Weiner B, et al. Inferring carbon sources from gene expression profiles using metabolic flux models. PLoS One 2012:e36947.
[49] Lee D, Smallbone K, Dunn WB, Murabito E, Winder CL, Kell DB, et al. Improving metabolic flux predictions using absolute gene expression data. BMC Syst Biol 2012;6:73.
[50] Reder C. Metabolic control theory: a structural approach. J Theor Biol 1988;135:175–201.
[51] Barker B, Sadagopan N, Wang Y, Smallbone K, Myers CR, Xi H, et al. A robust and efficient method for estimating enzyme complex abundance and metabolic flux from expression data; 2014 27.
[52] Pramanik J, Keasling JD. Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. Biotechnol Bioeng 1997;56:398–421.
[53] Westerhoff HV, Hellingwerf KJ, Van Dam K. Thermodynamic efficiency of microbial growth is low but optimal for maximal growth rate. Proc Natl Acad Sci U S A 1983;80:305–9.
[54] Feist AM, Palsson BO. The biomass objective function. Curr Opin Microbiol 2010;13:344–9.
[55] Raman K, Chandra N. Flux balance analysis of biological systems: applications and challenges. Brief Bioinform 2009;10:435–49.
[56] Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008;5:621–8.
[57] Cocaign M, Monnet C, Lindley N. Batch kinetics of *Corynebacterium glutamicum* during growth on various carbon substrates: use of substrate mixtures to localise metabolic bottlenecks. Appl Microbiol Biotechnol 1993;40:526–30.