

Article

Integrative Gene Expression and Metabolic Analysis Tool *IgemRNA*

Kristina Grausa¹, Ivars Mozga¹, Karlis Pleiko^{2,3}  and Agris Pentjuss^{4,*} 

¹ Department of Computer Systems, Latvia University of Life Sciences and Technologies, Liela Street 2, LV-3001 Jelgava, Latvia; kristina.grausa@gmail.com (K.G.); ivars.mozga@gmail.com (I.M.)

² Laboratory of Precision- and Nanomedicine, Institute of Biomedicine and Translational Medicine, University of Tartu, 50411 Tartu, Estonia; karlis.pleiko@lu.lv

³ Faculty of Medicine, University of Latvia, LV-1586 Riga, Latvia

⁴ Institute of Microbiology and Biotechnology, University of Latvia, Jelgavas Street 1, LV-1004 Riga, Latvia

* Correspondence: agris.pentjuss@gmail.com

Abstract: Genome-scale metabolic modeling is widely used to study the impact of metabolism on the phenotype of different organisms. While substrate modeling reflects the potential distribution of carbon and other chemical elements within the model, the additional use of omics data, e.g., transcriptome, has implications when researching the genotype–phenotype responses to environmental changes. Several algorithms for transcriptome analysis using genome-scale metabolic modeling have been proposed. Still, they are restricted to specific objectives and conditions and lack flexibility, have software compatibility issues, and require advanced user skills. We classified previously published algorithms, summarized transcriptome pre-processing, integration, and analysis methods, and implemented them in the newly developed transcriptome analysis tool *IgemRNA*, which (1) has a user-friendly graphical interface, (2) tackles compatibility issues by combining previous data input and pre-processing algorithms in MATLAB, and (3) introduces novel algorithms for the automatic comparison of different transcriptome datasets with or without Cobra Toolbox 3.0 optimization algorithms. We used publicly available transcriptome datasets from *Saccharomyces cerevisiae* BY4741 and H4-S47D strains for validation. We found that *IgemRNA* provides a means for transcriptome and environmental data validation on biochemical network topology since the biomass function varies for different phenotypes. Our tool can detect problematic reaction constraints.

Keywords: genome-scale metabolic modeling; transcriptomics; software engineering; Cobra Toolbox 3.0; MATLAB; flux balance analysis; flux variability analysis; omics data analysis



Citation: Grausa, K.; Mozga, I.; Pleiko, K.; Pentjuss, A. Integrative Gene Expression and Metabolic Analysis Tool *IgemRNA*. *Biomolecules* **2022**, *12*, 586. <https://doi.org/10.3390/biom12040586>

Academic Editors: Hongwu Ma and Igor Groyanin

Received: 17 March 2022

Accepted: 14 April 2022

Published: 16 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Advancements in genomics in the last decade have led to a rapid increase in published genome-scale sequences of different organisms. Currently, there are over 2.2×10^8 available genome sequence data samples (<https://www.ncbi.nlm.nih.gov/genbank/statistics/on> (accessed on 6 May 2021)). The growing genome data availability and technological advancements in sequencing and mass spectrometry have contributed to increasing multi-omics dataset generation. From a purely mathematical and statistical point of view, multi-omics dataset analysis is still very challenging and lacks proper methods [1,2]. Genome-scale metabolic models are comprehensive collections of known biochemical reactions catalyzed by specific associated proteins coded by genes. They have been successfully used to explain genotype–phenotype relationships for prokaryotic, eukaryotic, unicellular, and multi-tissue organisms. Genome-scale metabolic modeling (GSM) has already proven to be successful, leading to several significant breakthroughs in health and systems medicine fields [3–5], biotechnology [6–8], and many other life science fields [9–11]. GSM is a method for analyzing the phenotype responses of an organism by calculating flux distribution and other parameters [12] to see the carbon distribution potential and environmental perturbation impact on

the metabolism [13,14]. Despite the knowledge of genotype in GSM, phenotypic responses are far from being fully understood. With advances in omics technologies, it has become possible to quantitatively monitor gene transcription (high throughput sequencing), protein expression (mass-spectrometry proteomics), and metabolomics data, narrowing this gap between genotype and phenotype [15,16]. In contrast to gene set enrichment analysis, where genes exhibiting similar biological characteristics are sorted and classified in clusters [17], GSM-based integration methods also consider the interconnectivity of a biochemical network, the steady-state assumption, and gene–protein–reaction (GPR) associations. This allows the analysis of transcriptomics datasets to be performed in an interconnected manner on the biochemical network topology [18] using the gene–protein–reaction rule. There are two fundamental approaches to expression data integration in GSM: (1) directly integrating transcriptome data into model flux bounds (DIRECT) or (2) distributing the reactions into different categories based on transcriptome levels (DISTRIBUTE). The first DIRECT approach was developed by Åkesson [19]. They assumed that very low gene expression levels are associated with non-flux reactions. Published later, E-Flux [20] allowed to integrate quantitative transcriptome measurements directly into the model as the maximum possible flux value (flux upper bound). However, the Gene Inactivity Moderated by Metabolism and Expression (GIMME) [21] algorithm was made to distribute reactions into groups of active and inactive by comparing the corresponding enzymatic gene expression levels to a user-specified threshold and minimizing fluxes through inactive reactions.

The tool *IgemRNA* allows to use DIRECT and DISTRIBUTE transcriptomics data integration approaches. *IgemRNA* is an open-access toolbox created for transcriptomics data statistical and biochemical network topology-based analysis. *IgemRNA* was developed in the MATLAB environment to utilize the up-to-date and most commonly distributed GSM tool Cobra Toolbox 3.0 [22] and spreadsheet file (xls, xlsx) capabilities. *IgemRNA* is designed to analyze the quantitative genome-scale transcriptome data measurements like RNA-sequencing [23] and a targeted gene group transcriptome data, for example, gene expression microarray analysis [24]. *IgemRNA* allows to analyze transcriptome data directly or to integrate it into a metabolic model and perform optimization methods, such as FBA [25,26] or FVA [27]. Metabolic models without omics data cannot reveal distinct phenotype properties and can predict only the theoretical carbon and other chemical element distribution within the model. Applying transcriptome data to a metabolic model makes it possible to explain the phenotype properties in different stress and environmental conditions [28]. Although various tools for omics data integration into GSM already exist and have shown to be practical [29], they are not compatible. They use multiple standards and programming environments (Supplementary file S1). These problems have raised the need for a tool that combines most of the previously published basic functionalities, can be used in conjunction with already available GSM tools, and allows to select a variety of data integration, processing, analysis, and storage options in a user-friendly way. *IgemRNA* has been designed to combine several functionalities of previously published tools (Supplementary file S1), e.g., metabolic flux optimizations, transcriptome data initialization, and integration methods. The access is provided via a graphical user interface. Additionally, *IgemRNA* allows the inclusion of growth medium composition data along with the transcriptome data from the same environmental conditions. However, the main novelty is the built-in functions for transcriptome data pre-processing, including gene mapping and thresholding, transcriptome non- and post-GSM-based optimization statistical analysis, and the automatic comparison of data between different phenotypes.

2. Materials and Methods

IgemRNA is developed in the MATLAB programming environment. All functionality parts of the *IgemRNA* toolbox are open-access and freely available under the MIT License. GUI window is created with MATLAB *dialog()* function, and all user interface controls are added to the window using MATLAB *uicontrol()*. *IgemRNA* has dependencies for the MATLAB-based GSM tool Cobra Toolbox 3.0 [22] functionality and spreadsheet

file (xls, xlsx) capabilities to store the results. All *IgemRNA* functionalities are compatible with MATLAB versions not older than 2014 (<https://opencobra.github.io/cobratoolbox/latest/installation.html> (accessed on 6 May 2021)). *IgemRNA* is available on GitHub (<https://github.com/BigDataInSilicoBiologyGroup/IgemRNA> (accessed on 11 April 2022)) and is designed for transcriptome data analysis with or without a metabolic model network topology application.

2.1. *IgemRNA* Architecture Description

Cobra Toolbox 3.0 and spreadsheet data files (transcriptomics, metabolic model, optional medium data) in xls or xlsx format are required to access all the functionality of *IgemRNA* (Figure 1). *IgemRNA* uses different implemented methods and saves results in spreadsheet files, depending on the user-selected transcriptome data processing methods and analysis tasks. *IgemRNA* can also use the built-in methods Flux Balance Analysis (FBA) and Flux Variability Analysis (FVA) in Cobra Toolbox 3.0 [30] to perform more complex analysis on metabolic model network topology.

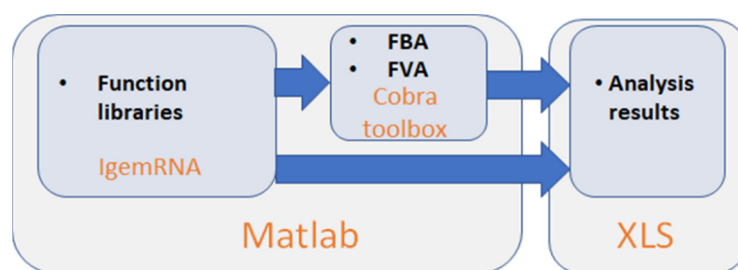


Figure 1. *IgemRNA* toolbox architecture scheme.

2.2. Tools Functionality Description

IgemRNA has seven different functional modules (Figure 2). *IgemRNA* modules include access to the user-friendly interface with data input options, data initialization, and pre-processing steps (gene mapping and thresholding) and executes user-selected non- and post-optimization analysis tasks. Post optimization tasks utilize some Cobra Toolbox 3.0 functions (FBA, FVA). *IgemRNA* can be started by opening the MATLAB software, navigating to the directory containing *IgemRNA.m* script, and running it. If any post-optimization tasks have been selected in the graphical interface, *IgemRNA* will launch Cobra Toolbox 3.0. The most time-consuming step is Cobra Toolbox 3.0 initialization with the update function. Thereby, it is set to run without updates by default.

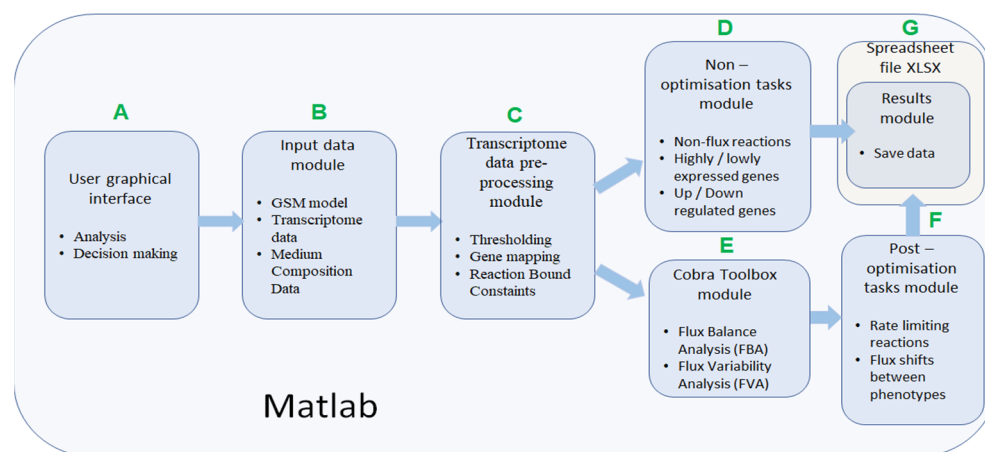


Figure 2. Description of *IgemRNA* function modules. (A) User graphical interface, (B) Input data module, (C) Transcriptome data pre-processing module, (D) Non-optimisation tasks module, (E) Cobra Toolbox module, (F) Post-optimisation tasks module, (G) Spreadsheet file module.

The *IgemRNA* software allows choosing which transcriptome data, medium composition, and metabolic model files to load. Further, desirable GPR mapping and thresholding procedures and optimization and analysis tasks can be selected. When all necessary options have been selected, *IgemRNA* automatically determines which non- or post-optimization functionalities will be executed. An essential part of *IgemRNA* is the graphical user interface (GUI) module (Figure 2A), which collects information about uploaded data files and selections for transcriptome data analysis and optimization tasks and passes these parameters to matching functional modules. The GUI window is opened using the MATLAB *dialog()* function. User interface controls are added with the MATLAB *uicontrol()* function. Depending on the input data files, specific functional modules are called afterward. The optimization and post-optimization tasks modules will not run if only a transcriptome data file is provided since they require a metabolic model (Figure 2).

The input data module accepts transcriptome, metabolic model, and medium composition data from uploaded files (Figure 2B). All input files must meet specified standards and criteria to be recognized by *IgemRNA*. Transcriptome data must be located in a spreadsheet *xlsx* file, where gene names are under the *GeneId* column, and the measured transcriptome values are under the *Data* column. Gene names must be the same as defined in the metabolic model. Allowed metabolic model data file types are *sbml*, *mat*, and *xlsx*, and the structure should be the same as defined in Cobra Toolbox 3.0 [22]. Using BioPax, GPML, and SBML files older than the 3.0 version is possible by converting them to the newest SBML 3.2 version using the online Systems Biology Format Converter (SBFC <https://www.ebi.ac.uk/biomodels/tools/converters/> (accessed on 11 April 2022)) beforehand. Uptake rates for substrates can also be defined by manual input. The medium composition file contains data concerning the medium in which the organism grows, the substrate uptake, and product reaction rates ($\text{mmol}\cdot\text{g}^{-1}\cdot\text{h}^{-1}$), and optionally a specific growth rate (h^{-1}) (Figure 4B).

The transcriptome pre-processing functionality module (Figure 2C) is responsible for preparing transcriptome and metabolic model data for analysis and optimization tasks. Pre-processing of transcriptome data relies mainly on decision parameters: gene mapping and thresholding approach, and different combinations of these decisions influence data processing results and biological data interpretation.

IgemRNA requires a threshold parameter that defines a border between differentially expressed genes. Two types of thresholds exist: local and global. Local thresholds are automatically calculated for each gene individually, given at least two samples of transcriptomics data of the same condition. However, a global threshold is used as a unique parameter that applies to all genes [31]. Different combinations of local and global thresholds lead to different analysis and optimization results. *IgemRNA* allows users to choose between three different thresholding approaches in the data pre-processing step (Table 1) [31]:

- Global T1 (GT1): is designed to analyze the transcriptome datasets using one global threshold. Example case shows that all transcriptome levels above 130 sequencing reads per gene (for a detailed description, see Section 2.3) are considered expressed, and others are considered suppressed. The global T1 threshold approach can be used for one or several phenotype transcriptome datasets.
- Local T1 (LT1): is designed to analyze transcriptome datasets having one global threshold value and a local rule. Local thresholds set a strict border for particular genes based on their varying gene expression levels across multiple samples to determine whether a gene is expressed or suppressed in a specific dataset. Local thresholds are only applied to those genes with expression levels above the global threshold since genes below the global threshold are automatically seen as suppressed. The example case shows that all transcriptome levels defined by the global threshold above 130 sequencing reads per gene are considered as possibly expressed, and below 130 sequencing reads per gene are considered suppressed. Local thresholds for specific

genes determine expression or suppression for genes with expression levels above the global threshold.

- Local T2 (LT2) is designed to use two global threshold values: upper and lower thresholds. Transcriptome levels higher than the upper global threshold are considered expressed genes and are active, and transcriptome levels below the lower global threshold are considered inactive genes. All genes with expression levels between the upper and lower global thresholds are considered possibly active. Local rules for these genes are calculated across multiple gene expression datasets and applied to determine their activity levels. The Local T2 thresholding approach can be used if several transcriptome datasets are available. An example of Local T2 shows that all gene expression levels above the upper global threshold of 130 sequencing reads per gene are considered active. Gene expressions lower than the global threshold of 50 sequencing reads per gene are considered suppressed.

Table 1. Thresholding Options.

| Approach | Threshold Input | Visual Representation | Examples | | | | | | | | | | | | | | |
|-----------|---|-----------------------|---|---------|-----|---------|-----|---------|-----|---------|----|---------|----|---------|---|---------|-----|
| Global T1 | <p>One global threshold:</p> <ol style="list-style-type: none"> exact value input percentile input (25th, 75th, 90th ...) | | <p>Lower threshold: 130</p> <table> <tr><td>YDL227C</td><td>871</td></tr> <tr><td>YDL226C</td><td>126</td></tr> <tr><td>YDL225W</td><td>319</td></tr> <tr><td>YDL224C</td><td>56</td></tr> <tr><td>YDL223C</td><td>13</td></tr> <tr><td>YDL222C</td><td>3</td></tr> <tr><td>YDL221W</td><td>135</td></tr> </table> | YDL227C | 871 | YDL226C | 126 | YDL225W | 319 | YDL224C | 56 | YDL223C | 13 | YDL222C | 3 | YDL221W | 135 |
| YDL227C | 871 | | | | | | | | | | | | | | | | |
| YDL226C | 126 | | | | | | | | | | | | | | | | |
| YDL225W | 319 | | | | | | | | | | | | | | | | |
| YDL224C | 56 | | | | | | | | | | | | | | | | |
| YDL223C | 13 | | | | | | | | | | | | | | | | |
| YDL222C | 3 | | | | | | | | | | | | | | | | |
| YDL221W | 135 | | | | | | | | | | | | | | | | |
| Local T1 | <p>One global threshold:</p> <ol style="list-style-type: none"> exact value input percentile input (25th, 75th, 90th ...) <p>Multiple local thresholds: calculated automatically across multiple transcriptomics data samples of the same conditions</p> | | <p>Lower threshold: 130</p> <p>Local for YDL227C: 600</p> <p>Local for YDL225W: 350</p> <table> <tr><td>YDL227C</td><td>871</td></tr> <tr><td>YDL226C</td><td>126</td></tr> <tr><td>YDL225W</td><td>319</td></tr> <tr><td>YDL223C</td><td>56</td></tr> <tr><td>YDL223C</td><td>13</td></tr> <tr><td>YDL222C</td><td>3</td></tr> <tr><td>YDL221W</td><td>135</td></tr> </table> | YDL227C | 871 | YDL226C | 126 | YDL225W | 319 | YDL223C | 56 | YDL223C | 13 | YDL222C | 3 | YDL221W | 135 |
| YDL227C | 871 | | | | | | | | | | | | | | | | |
| YDL226C | 126 | | | | | | | | | | | | | | | | |
| YDL225W | 319 | | | | | | | | | | | | | | | | |
| YDL223C | 56 | | | | | | | | | | | | | | | | |
| YDL223C | 13 | | | | | | | | | | | | | | | | |
| YDL222C | 3 | | | | | | | | | | | | | | | | |
| YDL221W | 135 | | | | | | | | | | | | | | | | |
| Local T2 | <p>Two global thresholds:</p> <ol style="list-style-type: none"> exact value input percentile input (25th, 75th, 90th ...) <p>Multiple local thresholds: calculated automatically across multiple transcriptomics data samples of the same conditions</p> | | <p>Lower Threshold: 50</p> <p>Upper Threshold: 130</p> <p>Local for YDL224C: 60</p> <p>Local for YDL226C: 70</p> <table> <tr><td>YDL227C</td><td>871</td></tr> <tr><td>YDL226C</td><td>126</td></tr> <tr><td>YDL225W</td><td>319</td></tr> <tr><td>YDL224C</td><td>56</td></tr> <tr><td>YDL223C</td><td>13</td></tr> <tr><td>YDL222C</td><td>3</td></tr> <tr><td>YDL221W</td><td>135</td></tr> </table> | YDL227C | 871 | YDL226C | 126 | YDL225W | 319 | YDL224C | 56 | YDL223C | 13 | YDL222C | 3 | YDL221W | 135 |
| YDL227C | 871 | | | | | | | | | | | | | | | | |
| YDL226C | 126 | | | | | | | | | | | | | | | | |
| YDL225W | 319 | | | | | | | | | | | | | | | | |
| YDL224C | 56 | | | | | | | | | | | | | | | | |
| YDL223C | 13 | | | | | | | | | | | | | | | | |
| YDL222C | 3 | | | | | | | | | | | | | | | | |
| YDL221W | 135 | | | | | | | | | | | | | | | | |

Each threshold approach has unique properties and is included in *IgemRNA*. Global threshold value input can be manual or automatic. In a manual input scenario, the user provides an exact expression value, whereas in an automated global threshold input scenario, threshold values are calculated based on a user-provided percentile. Local thresholds are always calculated automatically across multiple gene expression datasets during analysis. The selection of a thresholding approach is required for both post-optimization and non-optimization task cases.

Other important choice parameters are the Gene mapping approach and Constraining options used only for post-optimization task execution. A metabolic model with GPR

data is required, and gene mapping is performed using the GPR association. The most straightforward case is one gene encodes one protein that catalyzes one reaction (one gene, one protein, one reaction) in a metabolic model. However, the presence of enzyme complexes (multiple genes, one protein), isozymes (multiple proteins, one function), and promiscuous enzymes (one protein, multiple functions) in GSM makes these associations more complex, and they are defined with Boolean AND/OR rules. For transcriptome data implementation in GSM, several in silico gene mapping functions can be performed (Table 2):

Table 2. Gene Mapping Options.

| Requirement | Options |
|-------------------------------|--|
| Reaction constraining options | Only irreversible reactions All reactions Non—essential gene deletion only Meet minimum growth requirements |
| Gene mapping approach | AND/MIN and OR/MAX AND/MIN and OR/SUM AND/geometric mean and OR/MAX AND/geometric mean and OR/SUM |

- *Minimum (MIN)* AND operands in the GPR association are calculated by taking the lowest gene expression value.
- *Geometric mean (GM)* [32] AND operands in the GPR association are calculated as the geometric mean of the gene expression values.
- *Maximum (MAX)* OR operands in the GPR association are calculated by taking the highest gene expression value.
- *Sum (SUM)* OR operands in the GPR association are calculated as the sum of all the gene expression values.

Cobra Toolbox 3.0 has implemented several previously published transcriptome analysis methods [22]. *IgemRNA* uses 4 methods to implement transcriptomics data on reactions in the metabolic model:

- *Only irreversible reactions function.* Enzymatic reactions have three different directions in metabolic models: irreversible, reversible, and backward irreversible. This approach constrains only irreversible and backward irreversible reactions in the respective direction.
- *All reactions* function constrains all reactions: irreversible and backward irreversible reactions in an oriented direction, but reversible reactions are constrained in both directions.
- *Growth not affecting gene deletion only* option allows for the deletion of only those genes with expression values below the given threshold and which deletion does not affect growth. Cobra Toolbox 3.0 *singleGeneDeletion* analysis with the FBA method is performed before executing gene deletion for those genes. Only if the returned output *grRatio* by *singleGeneDeletion* function is equal to 1 (meaning that the wild type growth is equal to the deletion strain growth) does the gene get deleted.
- *Meet minimum growth requirements* option allows constraining only those reactions where the gene mapping end value (which is set as a reaction bound) is not below the minimum growth requirements for that reaction. Minimum growth requirements are obtained by creating another context-specific model where only the gene deletion and medium exchange reaction constraining is applied to calculate the Cobra Toolbox 3.0 FBA (*optimizeCbModel*) minimization of growth.

Non-optimization tasks module is the most straightforward transcriptome analysis module, which does not require additional Cobra Toolbox 3.0 functionality or a metabolic model. This module includes three predefined different transcriptome data analysis methods:

- *Filter high- and low-expression genes:* this method uses chosen threshold data and sorts genes into high-expression and low-expression datasets.

- *Filter low-expression genes*: this method uses chosen threshold parameters, filters genes with expression levels below the supplied thresholds, and returns them as non-expressed datasets.
- *Filter up-/down-regulated genes between phenotypes*: This method uses chosen threshold data and filters up- and down-regulated genes from two or more transcriptome datasets. The gene names must match in all datasets. The resulting data are passed to the Spreadsheet module (Figure 2G).
- *Cobra Toolbox module* is called before post-optimization tasks to calculate FBA and FVA results using Cobra Toolbox 3.0 functions (Figure 2E). This module requires a metabolic model.

Post-optimization task module is the *IgemRNA* advanced transcriptome analysis module that uses metabolic models and analyses them with Cobra Toolbox 3.0. The module applies all transcriptome data pre-processing functions (Figure 2C). Each function then has several options, resulting in a different analytical approach. The novelty of *IgemRNA* compared to other tools (Supplementary file S1) is the post-optimization module that has several functions for analyzing context-specific models:

- *Filter non-flux reactions*: this functionality filters out enzymatic reactions that do not carry a flux because the coded gene transcription levels are below the chosen threshold value in the pre-processing module (Figure 2C).
- *Filter rate-limiting reactions*: This functionality finds maximum reaction rates equal to the calculated GPR value based on gene expression data. The function uses the FVA optimization method to calculate the minimal and maximal rate value for each reaction and then filters reactions with upper bounds of the same value as the FVA maximal results.
- *Flux shifts between phenotypes*: this function compares minimal and maximal fluxes (calculated by FVA) between different phenotypes or datasets, calculating ratios between them.

All functionalities generate results, which are passed to the *Spreadsheet module*.

Results Module saves all non- and post-optimization analysis results in spreadsheet files (xls or xlsx).

IgemRNA is available on GitHub (<https://github.com/BigDataInSilicoBiologyGroup/IgemRNA>) (accessed on 11 April 2022).

2.3. RNA Sequencing Data Analysis

A publicly available and previously published [33] gene expression dataset from *Saccharomyces cerevisiae* BY4741 strain and mutant strain H4-S47D were used for analysis. Reads were aligned using STAR aligner [34] and assigned to genomic features using *featureCounts* [35]. Statistical comparisons between samples were made using *edgeR* [36]. The workflow is publicly available on the Galaxy platform [37,38]: <https://usegalaxy.eu/u/karlisleiko/w/rna-seq-kp-fromgeosingle-read> (accessed on 11 April 2022).

3. Results

3.1. The Comparison of Available Transcriptome Data Integration Tools

Before designing and implementing the *IgemRNA* software, we reviewed existing transcriptome analysis methodologies. We discovered many previously published methods and sorted them according to a variety of parameters. To classify transcriptome analysis methods, we chose to categorize them by the following attributes: method name (1), does the method return a context-specific model (2), thresholding method (3), gene mapping approach (4), requirements to run (5), if it is actively maintained (6), 3rd party software availability (7), and accessibility to build-in statistical analysis methods (8).

These methods differ in terms of their functionality and data pre-processing approaches. The method proposed by Åkesson [19] assumes that very low gene expression levels are associated with no flux reactions. The GIMME (Gene Inactivity Moderated by Metabolism and Expression) [21] method compares two transcriptome datasets, determines

active and inactive genes, and minimizes low-expression reactions, thus keeping objective function above a set value. The iMAT (Integrative Metabolic Analysis Tool) [39] allows the integration of transcriptomic and proteomic data into metabolic models. This method groups reactions into high-, moderate- and low-expression and maximizes the high- while minimizes the low-expression reactions. MADE (Metabolic Adjustment by Differential Expression) [40] uses two or more datasets of gene expression data across multiple conditions, creates a sequence of binary expression states to find statistically significant changes in gene expression measurements, and determines high-/low-expression reactions. The TIGER (Toolbox for Integrating Genome-scale Metabolism, Expression, and Regulation) [41] software platform facilitates the conversion of GPR associations into a mixed-integer linear program (MILP), which is used to constrain a metabolic model.

The resulting model combines GPR associations with a transcriptional regulatory network and can be further specified using gene expression data. E-Flux [20] constraints upper bounds for reactions classified as low-expression based on a given threshold and expression data. PROM (probabilistic regulation of metabolism) [42] is used for integrating genome-scale transcriptional regulatory networks into metabolic networks. The algorithm calculates the probability that a gene is active with respect to its transcription factor as specified by expression data. It then constrains the maximum reaction flux by a factor of this probability. The application of transcriptional regulatory networks and in metabolic modeling has also been outlined by TRFBA (transcriptional regulated flux balance analysis) [43]. This tool uses gene expression data from various perturbations as continuous variables and constrains reaction upper bounds to link transcriptional regulatory and metabolic networks. The INIT (Integrative Network Inference for Tissues) algorithm [44] maximizes reactions based on a qualitative confidence score and minimizes reactions associated with low expression. In addition, this method allows a small net accumulation rate for internal metabolites to prevent the removal of necessary reactions. Lee-12 [45] integrates absolute gene expression data directly into the objective function of a constraint-based model instead of constraining the fluxes. The biological objective function is replaced by a function that minimizes the deviation between gene expression levels and the fluxes. mCADRE (Context-Specificity Assessed by Deterministic Reaction Evaluation) [46] instead introduces a core set of reactions that must be present and active based on gene expression data and non-core reactions determined based on gene expression and connectivity evidence. This method also runs a test to ensure the basic functionality of the generated models and was successfully applied to reconstruct 126 human tissue genome-scale draft models. Fang-12 [47] is used to predict flux distribution for a perturbed state based on the differences in gene expression levels relative to a reference condition with precalculated flux distribution. This method also allows minor variations in biomass composition for the perturbed state.

Similarly, Δ FBA (deltaFBA) [48] also estimates differences between two conditions only by calculating the flux differences between the conditions, and it does not require the specification of an objective. RELATCH (relative change) [49] uses gene expression and fluxomic data from a reference state to estimate metabolic changes in a perturbed state for which there is no expression data available. Flux distribution for the perturbed state is calculated by minimizing the adjustment to the reference state. The TEAM (Temporal Expression-based Analysis of Metabolism) [50] method estimates time-course flux profiles using temporal gene expression patterns by combining dFBA (Dynamic Flux Balance Analysis) and the GIMME algorithm. It calculates the flux distribution at each time point and uses flux sum minimization to find the optimal solution. The GX-FBA (Gene-expression FBA) [51] method that integrates gene expression data into flux balance analysis uses the deviation in gene expression levels between a reference state and a perturbed state to define flux constraints for the perturbed state.

Akkeson, GIMME, E-Flux, and Δ FBA use only one global threshold for gene activity determination. PROM employs a predefined 33rd percentile threshold from the average value. INIT has an optional minimum flux threshold and positive/negative weights for each reaction. TEAM calculates the threshold from the M3D microarray dataset database.

iMAT sets two different global thresholds as lower and upper bounds. Meanwhile, MADE and TRFBA do not have thresholding.

All transcriptome analysis and optimization methods require a metabolic model and one or more transcriptome datasets. However, some algorithms require more complex experimental data or software to run. For iMAT and GIMME, a specified objective function is needed, while MADE needs access to a mixed-integer linear program solver (like GUROBI) and more than one transcriptome dataset. Multiple gene expression datasets states are also necessary to perform PROM, whereas TRFBA and Δ FBA E-Flux requires a function to convert gene expression levels into fluxes. PROM, TIGER, and TRFBA are developed to apply not only transcriptome environmental and genetic perturbation datasets, but also transcriptional regulatory networks. RELATCH method requires transcriptome and fluxomics datasets from the same conditions. The more advanced method TEAM uses initial composition data and temporal transcriptome and biomass composition data.

Along with different data integration approaches, an attribute not less important is the software availability and regular maintenance. For example, Lee-12 [45] has claimed its availability in Cobra Toolbox, but it is not found in the latest Cobra Toolbox 3.0 version. TIGER serves as a software platform for three previously published methods: GIMME, iMAT, and MADE. However, unfortunately, the platform is not maintained anymore. To our best knowledge, GIMME, iMAT, INIT, TEAM (only for microarray experiments), mCADRE, TRFBA, and Δ FBA are maintained regularly.

Most of the published methods are designed for particular modeling cases. They are not flexible for other scenarios since they require complex experimental data (fluxomics, multiple transcriptome datasets, or temporal data) or the integration of transcriptional regulatory networks. Later published transcriptome analysis approaches are considered ineligible for unified manuscript classification criteria.

IgemRNA is a novel tool that has combined many previously listed pre-processing and different transcriptome analysis methods. It allows the analysis of transcriptomics data together with or separately from metabolic models to find high- and low- expression genes or reactions and compare them with other phenotype datasets. *IgemRNA* has implemented more advanced pre-processing functions by combining several previously listed methods. It includes three thresholding, four gene mapping, and four reaction constraining options.

IgemRNA is compatible with MATLAB-based software and optionally uses Cobra Toolbox 3.0 functionality. Compared to previously mentioned methods, *IgemRNA* facilitates multiple thresholding and gene mapping approaches and several constraining options for transcriptome data integration into metabolic models. The additional feature is the possibility for the end-user to select different types of reactions (based on their flux and constraints) to extract from the result context-specific models. More advanced analysis methods in *IgemRNA* also include data comparison between different phenotypes. On top of that, all previously mentioned options can easily be selected via the graphical user interface.

3.2. *IgemRNA* Demonstration

The main novelty within this paper is the *IgemRNA* tool itself, combining multiple critical decisions and options for transcriptomics data analysis and integration in metabolic models. The graphical interface of *IgemRNA* makes data entry and pre-processing (thresholding, gene mapping) easier for the end-user (Figure 3). Another valuable feature of *IgemRNA* is the possibility for the user to choose which data analysis tasks to perform, including both non-optimization and post-optimization tasks.

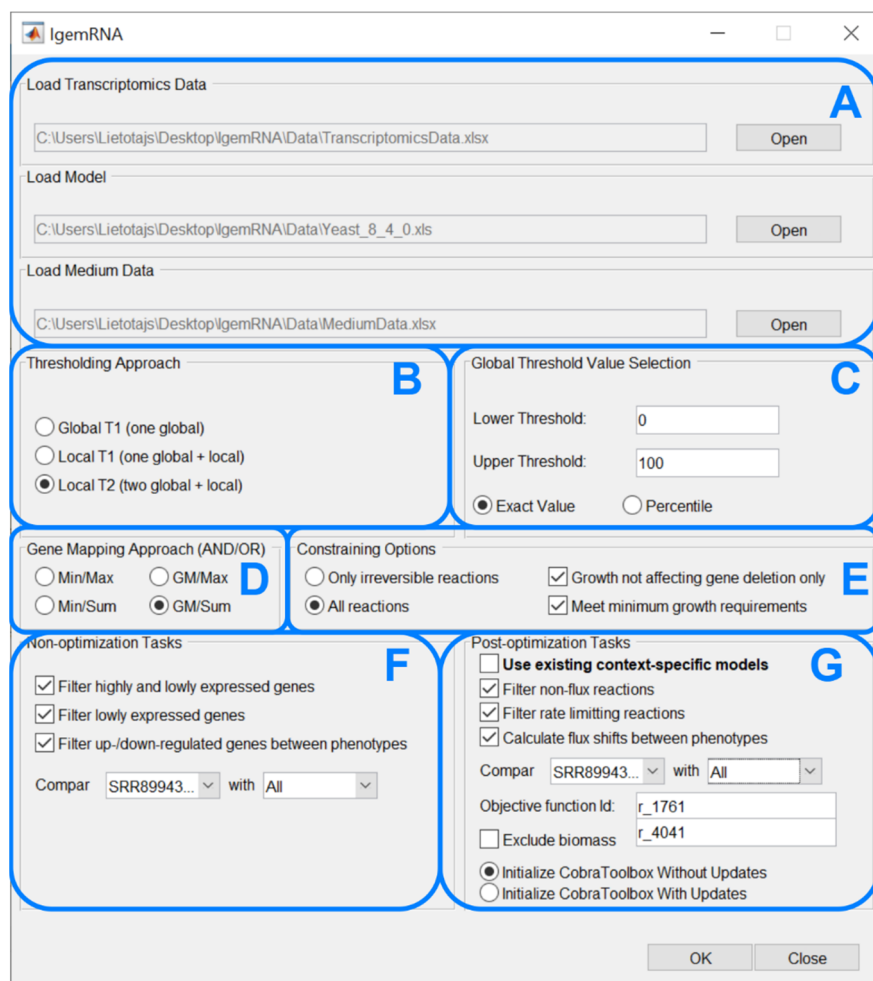


Figure 3. IgemRNA’s graphical interface. (A) Input data window, (B) Thresholding window, (C) Global thresholding value window, (D) Gene mapping window, (E) Reactions and transcriptomics data mapping window, (F) Non—optimization task window, (G) Post—optimization window.

After running the main *IgemRNA* script *IgemRNA.m*, a dialog box is opened where it is necessary to select the transcriptomics data file (xls or xlsx format) in the file upload section (Figure 3A). The transcriptomics data should be organized into two columns *GeneId* and *Data* (Figure 4B), where *GeneId* corresponds to the genes in the metabolic model if one is supplied. The transcriptomics data file can consist of multiple RNA-seq samples of the same conditions as well as different phenotypes. Therefore, the phenotype and sample name must appear as the sheet name since this name will be used in the result files and for selecting phenotypes for comparison (Figure 3F,G).

| | A | B | C | | A | B | C | D | E | F | G | H | I |
|----|------------|------------|------------|--|---|-----------|------|---|---|---|---|---|---|
| 1 | ReactionId | LowerBound | UpperBound | | 1 | GeneId | Data | | | | | | |
| 2 | r_1714 | -15 | -0,01 | | 2 | YDL248W | 1 | | | | | | |
| 3 | r_1761 | 8 | 12 | | 3 | YDL247W-A | 0 | | | | | | |
| 4 | | | | | 4 | YDL247W | 0 | | | | | | |
| 5 | | | | | 5 | YDL246C | 5 | | | | | | |
| 6 | | | | | 6 | YDL245C | 10 | | | | | | |
| 7 | | | | | 7 | YDL244W | 6 | | | | | | |
| 8 | | | | | 8 | YDL243C | 25 | | | | | | |
| 9 | | | | | 9 | YDL242W | 64 | | | | | | |
| 10 | | | | | | | | | | | | | |

Figure 4. Input data file structure; (A) Medium data file structure; (B) Transcriptomics data file structure.

Having selected a transcriptomics data file, the non-optimization tasks section (Figure 3F) becomes visible. Having chosen a GSM of Cobra Toolbox 3.0 available formats, the post-optimization tasks section (Figure 3G) becomes visible. Optionally, a medium data file (xls or xlsx format), where all the needed reactions and their upper and lower bounds are listed (Figure 4A), can be supplied in the file upload section (Figure 3A).

The following four dialog sections (Figure 3B–E) contain options for data pre-processing. *The thresholding approach* (Figure 3B) and *global threshold value selection* (Figure 3C) options determine methods for transcriptomics data pre-processing to split genes into high- and low-expression groups. It is possible to choose one of three thresholding approaches, namely Global T1 (GT1), Local T1 (LT1), or Local T2 (LT2) (see 2.2 threshold parameter). Depending on the selected approach, the user is then asked to enter one or two global threshold values, which can be supplied as a percentile or an exact value. *Gene mapping approach* (Figure 3D) and *constraining options* (Figure 3E) determine methods for transcriptomics data integration in a metabolic model. The gene mapping approach specifies what operations to use for mapping gene expression data to gene-protein-reaction (GPR) associations (see 2.2 Gene mapping), whereas constraining options allow to constrain only irreversible or all reactions. Moreover, it is possible to constrain reactions (see 2.2 Constraining options) by deleting only non-essential genes and the satisfy biomass objective function minimal resources value.

Non-optimization tasks (Figure 3F) allows to filter high- and/or low-expressed genes, export results in spreadsheet and compare gene expression data between phenotypes.

Post-optimization tasks combine transcriptomics data and a metabolic model to extract data on reaction levels. It is possible to filter non-flux and rate-limiting reactions and calculate reaction flux shifts between different phenotypes. An objective function is required to perform optimizations on the metabolic model. Since post-optimization tasks use some of the Cobra Toolbox 3.0 functions, the user is also asked to choose how to start Cobra Toolbox: with or without updates.

To validate *IgemRNA*, we used the *Saccharomyces cerevisiae* genome-scale metabolic model [52] version 8.4.0 [53] and transcriptome datasets [33]. We performed high-throughput genetic screenings that provide a novel global map of the histone residues required for transcriptional reprogramming in response to heat and osmotic stress in steady-state growth conditions. All validation details are found in Supplementary file S2.

To perform test cases provided in this user manual, simply run scripts via MATLAB environment. For example, *TestCase_determineGeneActivity.m* script will run a non-optimization task filter for high- and low-expression genes in Supplementary file S2.

4. Discussion and Conclusions

Having summarized several transcriptome data processing tools, we found that many are case-specific and allow users to select only a few possible data pre-processing and analysis functions. Recently proposed tools are not widely used due to the complexity of required input data, such as tools that facilitate time-series or genome-scale metabolomics data that are not widely accessible or use their own unique FBA modified functions or different 3rd party software.

We sorted out methods and algorithms providing flexibility and the possibility of application in a wide range of experimental scenarios and implemented them in the newly developed transcriptome analysis tool, *IgemRNA*. The main contributions of the gene expression and metabolic analysis tool are a user-friendly graphical interface, the tackling of compatibility issues by combining several data pre-processing and integration methods in MATLAB environment, allowing the end-user to select different types of reactions (based on their flux or constraints) to extract and filter from the result models, and the novel algorithms for the automatic comparison of transcriptome data from different phenotypes with or without Cobra Toolbox 3.0 optimization algorithms.

We found that *IgemRNA* provides a means for transcriptome and environment data validation on biochemical network topology since the biomass function varies for different

phenotypes. The software can detect problematic reaction constraints to loosen their bounds and achieve a steady state. *IgemRNA*, in contrast to gene set enrichment analysis, additionally validates transcriptomics measurement quality, where minimal metabolic network connectivity and flux requirements must be fulfilled. Otherwise, transcriptome data quality is questionable.

In general, the *IgemRNA* tool combines unique data entry, initialization, optimization, and analysis algorithms that significantly facilitate manual work for the end user with the integration of transcriptomics and environmental data into metabolic models, and thus presents great potential to increase novel scientific discoveries in research concerning different organisms.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biom12040586/s1>, Supplementary file S1: Classifying previously published transcriptome data integration in GSM methods. Supplementary file S2: User manual documentation for the *IgemRNA*. All *IgemRNA* test data and scripts are found at <https://github.com/BigDataInSilicoBiologyGroup/IgemRNA> (accessed on 11 April 2022).

Author Contributions: Conceptualization, A.P.; methodology, A.P. and K.G.; RNA sequencing methodology, K.P.; software, K.G.; validation, K.G. and A.P.; investigation, A.P. and K.G.; data curation, K.G. and I.M.; writing—original draft preparation, A.P.; writing—review and editing, A.P., K.G., K.P. and I.M.; supervision, A.P. All authors have read and agreed to the published version of the manuscript.

Funding: Supported by European Regional Development Fund Postdoctoral research aid 1.1.1.2/VIAA/2/18/278 and FACCE SURPLUS and supported by No. 23-11.17e/20/173 “Sustainable up-cycling of agricultural residues: modular cascading waste conversion system” (UpWaste).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: *IgemRNA* MATLAB scripts and test cases for transcriptome data integration in genome-scale metabolic models are available at <https://github.com/BigDataInSilicoBiologyGroup/IgemRNA> (accessed on 11 April 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. López de Maturana, E.; Alonso, L.; Alarcón, P.; Martín-Antoniano, I.A.; Pineda, S.; Piorno, L.; Calle, M.L.; Malats, N. Challenges in the Integration of Omics and Non-Omics Data. *Genes* **2019**, *10*, 238. [\[CrossRef\]](#)
2. Krassowski, M.; Das, V.; Sahu, S.K.; Misra, B.B. State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing. *Front. Genet.* **2020**, *11*, 610798. [\[CrossRef\]](#)
3. Mardinoglu, A.; Boren, J.; Smith, U.; Uhlen, M.; Nielsen, J. Systems Biology in Hepatology: Approaches and Applications. *Nat. Rev. Gastroenterol. Hepatol.* **2018**, *15*, 365–377. [\[CrossRef\]](#)
4. Benfeitas, R.; Uhlen, M.; Nielsen, J.; Mardinoglu, A. New Challenges to Study Heterogeneity in Cancer Redox Metabolism. *Front. Cell Dev. Biol.* **2017**, *5*, 65. [\[CrossRef\]](#)
5. Mardinoglu, A.; Nielsen, J. New Paradigms for Metabolic Modeling of Human Cells. *Curr. Opin. Biotechnol.* **2015**, *34*, 91–97. [\[CrossRef\]](#)
6. Choi, S.; Song, C.W.; Shin, J.H.; Lee, S.Y. Biorefineries for the Production of Top Building Block Chemicals and Their Derivatives. *Metab. Eng.* **2015**, *28*, 223–239. [\[CrossRef\]](#)
7. Kalnietieks, U.; Pentjuss, A.; Rutkis, R.; Stalidzans, E.; Fell, D.A. Modeling of *Zymomonas Mobilis* Central Metabolism for Novel Metabolic Engineering Strategies. *Front. Microbiol.* **2014**, *5*, 42. [\[CrossRef\]](#)
8. Kalnietieks, U.; Balodite, E.; Ströhler, S.; Strazdina, I.; Rex, J.; Pentjuss, A.; Fuchino, K.; Bruheim, P.; Rutkis, R.; Pappas, K.M.; et al. Improvement of Acetaldehyde Production in *Zymomonas Mobilis* by Engineering of Its Aerobic Metabolism. *Front. Microbiol.* **2019**, *10*, 2533. [\[CrossRef\]](#)
9. McNally, C.P.; Borenstein, E. Metabolic Model-Based Analysis of the Emergence of Bacterial Cross-Feeding via Extensive Gene Loss. *BMC Syst. Biol.* **2018**, *12*, 69. [\[CrossRef\]](#)
10. Borer, B.; Ataman, M.; Hatzimanikatis, V.; Or, D. Modeling Metabolic Networks of Individual Bacterial Agents in Heterogeneous and Dynamic Soil Habitats (IndiMeSH). *PLoS Comput. Biol.* **2019**, *15*, e1007127. [\[CrossRef\]](#)

11. Marinos, G.; Kaleta, C.; Waschina, S. Defining the Nutritional Input for Genome-Scale Metabolic Models: A Roadmap. *PLoS ONE* **2020**, *15*, e0236890. [[CrossRef](#)] [[PubMed](#)]
12. Stalidzans, E.; Seiman, A.; Peebo, K.; Komasilovs, V.; Pentjuss, A. Model-Based Metabolism Design: Constraints for Kinetic and Stoichiometric Models. *Biochem. Soc. Trans.* **2018**, *46*, 261–267. [[CrossRef](#)] [[PubMed](#)]
13. Pentjuss, A.; Odzina, I.; Kostromins, A.; Fell, D.A.; Stalidzans, E.; Kalnenieks, U. Biotechnological Potential of Respiring *Zymomonas Mobilis*: A Stoichiometric Analysis of Its Central Metabolism. *J. Biotechnol.* **2013**, *165*, 1–10. [[CrossRef](#)] [[PubMed](#)]
14. Pentjuss, A.; Stalidzans, E.; Liepins, J.; Kokina, A.; Martynova, J.; Zikmanis, P.; Mozga, I.; Scherbaka, R.; Hartman, H.; Poolman, M.G.; et al. Model-Based Biotechnological Potential Analysis of *Kluyveromyces marxianus* Central Metabolism. *J. Ind. Microbiol. Biotechnol.* **2017**, *44*, 1177–1190. [[CrossRef](#)]
15. Ramata-Stunda, A.; Valkovska, V.; Boroduškis, M.; Livkiša, D.; Kaktiņa, E.; Silamikele, B.; Boroduške, A.; Pentjušs, A.; Rostoks, N. Development of Metabolic Engineering Approaches to Regulate the Content of Total Phenolics, Antiradical Activity and Organic Acids in Callus Cultures of the Highbush Blueberry (*Vaccinium corymbosum* L.). *Agron. Res.* **2020**, *18*, 1860–1872. [[CrossRef](#)]
16. Subramanian, I.; Verma, S.; Kumar, S.; Jere, A.; Anamika, K. Multi-Omics Data Integration, Interpretation, and Its Application. *Bioinform. Biol. Insights* **2020**, *14*, 117793221989905. [[CrossRef](#)]
17. 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.; et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [[CrossRef](#)]
18. Zhang, C.; Lee, S.; Bidkhor, G.; Benfeitas, R.; Lovric, A.; Chen, S.; Uhlen, M.; Nielsen, J.; Mardinoglu, A. RMetD2: A Tool for Integration of Relative Transcriptomics Data into Genome-Scale Metabolic Models. *bioRxiv* **2019**, 663096. [[CrossRef](#)]
19. Åkesson, M.; Förster, J.; Nielsen, J. Integration of Gene Expression Data into Genome-Scale Metabolic Models. *Metab. Eng.* **2004**, *6*, 285–293. [[CrossRef](#)]
20. Colijn, C.; Brandes, A.; Zucker, J.; Lun, D.S.; Weiner, B.; Farhat, M.R.; Cheng, T.-Y.; Moody, D.B.; Murray, M.; Galagan, J.E. Interpreting Expression Data with Metabolic Flux Models: Predicting Mycobacterium Tuberculosis Mycolic Acid Production. *PLoS Comput. Biol.* **2009**, *5*, e1000489. [[CrossRef](#)]
21. Becker, S.A.; Palsson, B.O. Context-Specific Metabolic Networks Are Consistent with Experiments. *PLoS Comput. Biol.* **2008**, *4*, e1000082. [[CrossRef](#)] [[PubMed](#)]
22. 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. Creation and Analysis of Biochemical Constraint-Based Models Using the COBRA Toolbox v.3.0. *Nat. Protoc.* **2019**, *14*, 639–702. [[CrossRef](#)] [[PubMed](#)]
23. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A Revolutionary Tool for Transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63. [[CrossRef](#)] [[PubMed](#)]
24. Tarca, A.L.; Romero, R.; Draghici, S. Analysis of Microarray Experiments of Gene Expression Profiling. *Am. J. Obstet. Gynecol.* **2006**, *195*, 373–388. [[CrossRef](#)]
25. Orth, J.D.; Thiele, I.; Palsson, B.Ø. What Is Flux Balance Analysis? *Nat. Biotechnol.* **2010**, *28*, 245–248. [[CrossRef](#)]
26. Fell, D.A.; Small, J.R. Fat Synthesis in Adipose Tissue. An Examination of Stoichiometric Constraints. *Biochem. J.* **1986**, *238*, 781–786. [[CrossRef](#)]
27. Gudmundsson, S.; Thiele, I. Computationally Efficient Flux Variability Analysis. *BMC Bioinform.* **2010**, *11*, 489. [[CrossRef](#)]
28. Nogales, J.; Agudo, L. A Practical Protocol for Integration of Transcriptomics Data into Genome-Scale Metabolic Reconstructions. In *Hydrocarbon and Lipid Microbiology Protocols*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 135–152. [[CrossRef](#)]
29. Blazier, A.S.; Papin, J.A. Integration of Expression Data in Genome-Scale Metabolic Network Reconstructions. *Front. Physiol.* **2012**, *3*, 299. [[CrossRef](#)]
30. Lewis, N.E.; Nagarajan, H.; Palsson, B.O. Constraining the Metabolic Genotype–Phenotype Relationship Using a Phylogeny of in Silico Methods. *Nat. Rev. Microbiol.* **2012**, *10*, 291–305. [[CrossRef](#)]
31. Richelle, A.; Joshi, C.; Lewis, N.E. Assessing Key Decisions for Transcriptomic Data Integration in Biochemical Networks. *PLoS Comput. Biol.* **2019**, *15*, e1007185. [[CrossRef](#)]
32. Salvy, P.; Hatzimanikatis, V. The ETFL Formulation Allows Multi-Omics Integration in Thermodynamics-Compliant Metabolism and Expression Models. *Nat. Commun.* **2020**, *11*, 30. [[CrossRef](#)] [[PubMed](#)]
33. Viéitez, C.; Martínez-Cebrián, G.; Solé, C.; Böttcher, R.; Potel, C.M.; Savitski, M.M.; Onnebo, S.; Fabregat, M.; Shilatifard, A.; Posas, F.; et al. A Genetic Analysis Reveals Novel Histone Residues Required for Transcriptional Reprogramming upon Stress. *Nucleic Acids Res.* **2020**, *48*, 3455–3475. [[CrossRef](#)] [[PubMed](#)]
34. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29*, 15–21. [[CrossRef](#)]
35. Liao, Y.; Smyth, G.K.; Shi, W. FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics* **2014**, *30*, 923–930. [[CrossRef](#)]
36. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* **2010**, *26*, 139–140. [[CrossRef](#)] [[PubMed](#)]
37. Afgan, E.; Baker, D.; van den Beek, M.; Blankenberg, D.; Bouvier, D.; Čech, M.; Chilton, J.; Clements, D.; Coraor, N.; Eberhard, C.; et al. The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2016 Update. *Nucleic Acids Res.* **2016**, *44*, W3–W10. [[CrossRef](#)] [[PubMed](#)]

38. McCarthy, D.J.; Chen, Y.; Smyth, G.K. Differential Expression Analysis of Multifactor RNA-Seq Experiments with Respect to Biological Variation. *Nucleic Acids Res.* **2012**, *40*, 4288–4297. [[CrossRef](#)]
39. Jerby, L.; Shlomi, T.; Ruppin, E. Computational Reconstruction of Tissue-specific Metabolic Models: Application to Human Liver Metabolism. *Mol. Syst. Biol.* **2010**, *6*, 401. [[CrossRef](#)]
40. Jensen, P.A.; Papin, J.A. Functional Integration of a Metabolic Network Model and Expression Data without Arbitrary Thresholding. *Bioinformatics* **2011**, *27*, 541–547. [[CrossRef](#)]
41. Jensen, P.A.; Lutz, K.A.; Papin, J.A. TIGER: Toolbox for Integrating Genome-Scale Metabolic Models, Expression Data, and Transcriptional Regulatory Networks. *BMC Syst. Biol.* **2011**, *5*, 147. [[CrossRef](#)]
42. 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. USA* **2010**, *107*, 17845–17850. [[CrossRef](#)] [[PubMed](#)]
43. Motamedian, E.; Mohammadi, M.; Shojaosadati, S.A.; Heydari, M. TRFBA: An Algorithm to Integrate Genome-Scale Metabolic and Transcriptional Regulatory Networks with Incorporation of Expression Data. *Bioinformatics* **2017**, *29*, 1399–1406. [[CrossRef](#)] [[PubMed](#)]
44. Agren, R.; Bordel, S.; Mardinoglu, A.; Pornputtapong, N.; Nookaew, I.; Nielsen, J. Reconstruction of Genome-Scale Active Metabolic Networks for 69 Human Cell Types and 16 Cancer Types Using INIT. *PLoS Comput. Biol.* **2012**, *8*, e1002518. [[CrossRef](#)] [[PubMed](#)]
45. Lee, D.; Smallbone, K.; Dunn, W.B.; Murabito, E.; Winder, C.L.; Kell, D.B.; Mendes, P.; Swainston, N. Improving Metabolic Flux Predictions Using Absolute Gene Expression Data. *BMC Syst. Biol.* **2012**, *6*, 73. [[CrossRef](#)]
46. Wang, Y.; Eddy, J.A.; Price, N.D. Reconstruction of Genome-Scale Metabolic Models for 126 Human Tissues Using MCADRE. *BMC Syst. Biol.* **2012**, *6*, 153. [[CrossRef](#)]
47. Fang, X.; Wallqvist, A.; Reifman, J. Modeling Phenotypic Metabolic Adaptations of *Mycobacterium Tuberculosis* H37Rv under Hypoxia. *PLoS Comput. Biol.* **2012**, *8*, e1002688. [[CrossRef](#)]
48. Ravi, S.; Gunawan, R. Δ FBA—Predicting Metabolic Flux Alterations Using Genome-Scale Metabolic Models and Differential Transcriptomic Data. *PLoS Comput. Biol.* **2021**, *17*, e1009589. [[CrossRef](#)]
49. Kim, J.; Reed, J.L. RELATCH: Relative Optimality in Metabolic Networks Explains Robust Metabolic and Regulatory Responses to Perturbations. *Genome Biol.* **2012**, *13*, R78. [[CrossRef](#)]
50. Collins, S.B.; Reznik, E.; Segrè, D. Temporal Expression-Based Analysis of Metabolism. *PLoS Comput. Biol.* **2012**, *8*, e1002781. [[CrossRef](#)]
51. Navid, A.; Almaas, E. Genome-Level Transcription Data of *Yersinia Pestis* Analyzed with a New Metabolic Constraint-Based Approach. *BMC Syst. Biol.* **2012**, *6*, 150. [[CrossRef](#)]
52. Lu, H.; Li, F.; Sánchez, B.J.; Zhu, Z.; Li, G.; Domenzain, I.; Marcišauskas, S.; Anton, P.M.; Lappa, D.; Lieven, C.; et al. A Consensus *S. cerevisiae* Metabolic Model Yeast8 and Its Ecosystem for Comprehensively Probing Cellular Metabolism. *Nat. Commun.* **2019**, *10*, 3586. [[CrossRef](#)] [[PubMed](#)]
53. Sánchez, B.J.; Quan, C.W.; Mendoza, S.; Li, F.; Lu, H.; Beber, M.; Christian, L.; Kerkhoven, E.J. SysBioChalmers/Yeast-GEM: Yeast 8.4.0 (Version v8.4.0). 2020. Available online: <https://zenodo.org/record/3894510#.Y1d8N0hByUk> (accessed on 11 April 2022). [[CrossRef](#)]