

A Critical Evaluation of Methods for the Reconstruction of Tissue-Specific Models

Sara Correia and Miguel Rocha^(✉)

Centre of Biological Engineering, University of Minho, Braga, Portugal
mrocha@di.uminho.pt

Abstract. Under the framework of constraint based modeling, genome-scale metabolic models (GSMMs) have been used for several tasks, such as metabolic engineering and phenotype prediction. More recently, their application in health related research has spanned drug discovery, biomarker identification and host-pathogen interactions, targeting diseases such as cancer, Alzheimer, obesity or diabetes. In the last years, the development of novel techniques for genome sequencing and other high-throughput methods, together with advances in Bioinformatics, allowed the reconstruction of GSMMs for human cells. Considering the diversity of cell types and tissues present in the human body, it is imperative to develop tissue-specific metabolic models. Methods to automatically generate these models, based on generic human metabolic models and a plethora of omics data, have been proposed. However, their results have not yet been adequately and critically evaluated and compared.

This work presents a survey of the most important tissue or cell type specific metabolic model reconstruction methods, which use literature, transcriptomics, proteomics and metabolomics data, together with a global template model. As a case study, we analyzed the consistency between several omics data sources and reconstructed distinct metabolic models of hepatocytes using different methods and data sources as inputs. The results show that omics data sources have a poor overlapping and, in some cases, are even contradictory. Additionally, the hepatocyte metabolic models generated are in many cases not able to perform metabolic functions known to be present in the liver tissue. We conclude that reliable methods for *a priori* omics data integration are required to support the reconstruction of complex models of human cells.

1 Introduction

Over the last years, genome-scale metabolic models (GSMMs) for several organisms have been developed, mainly for microbes with an interest in Biotechnology [6, 20]. These models have been used to predict cellular metabolism and promote biological discovery [17], under constraint-based approaches such as Flux Balance Analysis (FBA) [18]. FBA finds a flux distribution that maximizes biomass production, considering the knowledge of stoichiometry and reversibility of reactions, and taking some simplifying assumptions, namely assuming quasi steady-state conditions.

Recently, efforts on model reconstruction have also addressed more complex multicellular organisms, including humans [5,9,25]. In biomedical research, they have been used, for instance, to elucidate the role of proliferative adaptation causing the Warburg effect in cancer [23], to predict metabolic markers for inborn errors of metabolism [24] and to identify drug targets for specific diseases [11].

However, the human organism is quite complex, with a large number of cell types/tissues and huge diversity in their metabolic functions. This led to the need of developing tissue/cell type specific metabolic models, which could allow studying in more depth specific cell phenotypes. Towards this end, it was imperative to better characterize specific cell types, gathering relevant data. Indeed, an important set of technological advances in the last decades greatly increased available biological data through high-throughput studies that allow the identification and quantification of cell components (gene expression, proteins and metabolites). These are collectively known as ‘omics’ data and have generated new fields of study, such as transcriptomics, proteomics and metabolomics.

The most widely available omics data are transcriptomics, the quantification of gene expression levels in a cell, using DNA microarrays or sequencing (e.g. RNA-seq). The most significant databases for gene expression data are the Gene Expression Omnibus (GEO) [3] and the ArrayExpress [19]. Other resources use those databases as references to synthesize their information, such as the Gene Expression Barcode [16], which provides absolute measures of expression for the most annotated genes, organized by tissue, cell-types and diseases.

In the cells, mRNA is not always translated into protein, and the amounts of protein depend on gene expression but also on other factors. Thus, knowledge about the amounts of proteins in the cell, provided by proteomics data, is of foremost relevance. These data can confirm the presence of proteins and provide measurements of their quantities for each protein within a cell. For human cells, a database is available with millions of high-resolution images showing the spatial distribution of protein expression profiles in normal tissues, cancer and cell lines - the Human Protein Atlas (HPA) portal [26].

Another source of information is provided by metabolomics data that involve the quantification of the small molecules present in cells, tissues, organs and biological fluids using techniques such as Nuclear Magnetic Resonance spectroscopy or Gas Chromatography combined with Mass Spectrometry [13]. The Human Metabolome Database (HMDB) [28] contains spectroscopic, quantitative, analytic and molecular-scale information about human metabolites, associated enzymes or transporters, their abundance and disease-related properties.

Resources for omics data, together with generic human metabolic models, have been used to generate context-specific models. This has been achieved through the development of methods, such as the Model Building Algorithm (MBA) [12], the Metabolic Context-specificity Assessed by Deterministic Reaction Evaluation (mCADRE)[27] and the Task-driven Integrative Network Inference for Tissues (tINIT)[2].

The reconstructed models have allowed, for instance, to find metabolic targets to inhibit the proliferation of cancer cells [29], to study the interaction

between distinct brain cells [14], and to find potential therapeutic targets for the treatment of non-alcoholic steatohepatitis [15].

However, the aforementioned methods have not yet been critically and systematically evaluated on standardized case studies. Indeed, each of the methods is proposed and validated with distinct cases and taking distinct omics data sources as inputs. Thus, the impact of using different omics datasets on the final results of those algorithms is a question that remains to be answered. Here, we present a critical evaluation of the most important methods for the reconstruction of tissue-specific metabolic models published until now.

We have developed a framework where we implemented different methods for the reconstruction of tissue-specific metabolic models. In this scenario, the algorithms use sets of metabolites and/or maps of scores for each reaction as input. So, in our framework the algorithms are independent from the omics data source, and the separation of these two layers allows to use different data sources in each algorithm for the generation of tissue-specific metabolic models. As a case study, to compare the three different approaches implemented, metabolic models were reconstructed for hepatocytes, using the same set of data sources as inputs for each algorithm. Moreover, distinct combinations of data sources are evaluated to check their influence on the final results.

2 Materials and Methods

2.1 Human Metabolic Models

Modeling metabolic systems requires the analysis and prediction of metabolic flux distributions under diverse physiological and genetic conditions. The human organism is one of the most complex organisms to build a metabolic model since the number of genes, types of cells and their diversity are huge. In the last years, a few human metabolic models were proposed [5,9,15,25]. In this work, we will use the Recon 2 human metabolic model that accounts for 7440 reactions, 5063 metabolites and 1789 enzyme-encoding genes. This model is a community-driven expansion of the previous human reconstruction, Recon 1 [5], with additional information from different resources: EHMN [9], Hepatonet1[8], Ac-FAO module[21] and the small intestinal enterocyte reconstruction [22].

2.2 Algorithms for Tissue-Specific Metabolic Models Reconstruction

Although there are several applications of the human GSMs, the specificity of cell types requires the reconstruction of tissue-specific metabolic models. Some approaches have been proposed based on existing generic human models. Here, we present three of the most well-known approaches for this task that will be used in the remaining of this work.

MBA. The Model-Building Algorithm (MBA) [12] reconstructs a tissue-specific metabolic model from a generic model and two sets of reactions, denoted as core reactions (C_H) and reactions with a moderate probability to be carried out in the specific tissue (C_M). These sets were previously built according to evidence levels based on omics data, literature and experimental knowledge. In general, the C_H set includes human-curated tissue-specific reactions and the C_M set includes reactions certified by omics data. The algorithm iteratively removes one reaction from the generic model, in a random order, and validates if the model remains consistent. The process ends when the removal of all reactions, except the ones in C_H , is tried. As a result, this algorithm reconstructs a model containing all the C_H reactions, as many as possible C_M reactions, and a minimal set of other reactions that are required for obtaining overall model consistency (for each reaction there is a flux distribution in which it is active).

Since reactions are scanned in a random order, the authors recommend to run the algorithm a large number of times to generate intermediate models. After this step, a score per each reaction is calculated, according to the number of times it appears in these models. The final model is built starting from C_H and iteratively adding reactions ordered by their scores, until a final consistent model is achieved.

INIT/ tINIT. The Integrative Network Inference for Tissues (INIT) [1] uses the Human Protein Atlas (HPA) as its main source of evidence. Expression data can be used when proteomic evidence is missing. It also allows the integration of metabolomics data by imposing a positive net production of metabolites for which there is experimental support, for instance in HMDB. The algorithm is formulated using mixed integer-linear programming (MILP), so that the final model contains reactions with high scores from HPA data. This algorithm does not impose strict steady-state conditions for all internal metabolites, allowing a small net accumulation rate. A couple of years later, a new version of this algorithm was proposed, the Task-driven Integrative Network Inference for Tissues (tINIT) [2], which reconstructs tissue-specific metabolic models based on protein evidence from HPA and a set of metabolic tasks that the final context-specific model must perform. These tasks are used to test the production or uptake of external metabolites, but also the activation of pathways that occur in a specific tissue. Another improvement from the previous version is the addition of constraints to guarantee that irreversible reactions operate in one direction only.

mCADRE. The Metabolic Context specificity Assessed by Deterministic Reaction Evaluation (mCADRE) [27] method is able to infer a tissue-specific network based on gene expression data, network topology and reaction confidence levels. Based on the expression score, the reactions of the global model, used as template, are ranked and separated in two sets - core and non-core. All reactions with expression-based scores higher than a threshold value are included in the core set, while the remaining reactions make the non-core set. In this method, the expression score does not represent the level of expression, but

rather the frequency of expressed states over several transcript profiles. So, it is necessary to previously binarize the expression data. Thus, it is possible to use data retrieved from the Gene Expression Barcode project that already contains binary information on which genes are present or not in a specific tissue/cell type. Reactions from the non-core set are ranked according to the expression scores, connectivity-based scores and confidence level-based scores. Then, sequentially, each reaction is removed and the consistency of the model is tested. The elimination only occurs if the reaction does not prevent the production of a key-metabolite and the core consistency is preserved. Comparing with the MBA algorithm, mCADRE presents two improvements: it allows the definition of key metabolites, i.e. metabolites that have evidence to be produced in the context-specific model reconstruction, and relaxes the condition of including all core reactions in the final model.

Table 1 shows the mathematical formulation and pseudocode for all algorithms described above.

Table 1. Formulation and description of algorithms of MBA, tINIT and mCADRE. In the table R_G represents the list of reaction from the global model, R_C the set of core reaction on mCADRE algorithm, C_H and C_M the core and moderate probability sets used in MBA algorithm, r a reaction and the $for(i)$ and the $rev(i)$ represent the i -th reaction direction (forward and reverse).

MBA	tINIT	mCADRE
$generateModel(R_G, C_H, C_M)$	$min \sum_{i \in R} w_i * y_i$	$generateModel(R_G, threshold)$
$R_P \leftarrow R_G$	s.t.	$R_P \leftarrow R_G$
$R_S \leftarrow R_P \setminus (C_H \cup C_M)$	$Sv = b$	$R_C \leftarrow score(R_P) > threshold$
$P \leftarrow randomPermutation(R_S)$	$ v_i \leq v_{max}$	$coreActiveG \leftarrow flux(r) = 0, r \in R_C$
$for(r \in P)$	$0 < v_i + (v_{max} * y_i) \leq v_{max}$	$R_{NC} \leftarrow R_P \setminus R_C$
$inactiveR \leftarrow CheckModel(R_P, r)$	$b_j \geq \delta \quad j \in Metabolomics$	$for(r \in order(R_{NC}))$
$e_H \leftarrow inactiveR \cap C_H$	$b_j = 0 \quad j \notin Metabolomics$	$inactiveR \leftarrow CheckModel(R_P, r)$
$e_M \leftarrow inactiveR \cap C_M$	$y_{for(i)} + y_{rev(i)} \leq 1$	$s1 = inactiveR \cap R_C $
$e_X \leftarrow inactiveR \setminus (C_H \cup C_M)$	$y_{for(i)} \geq \delta, i \in RequiredReac$	$s2 = inactiveR \cap R_{NC} $
$if(e_H == 0 \text{ AND } e_M < \delta * e_X)$	$y_i \in \{0, 1\}$	$if(r \notin withExpressionValues \text{ AND } s1 \setminus s2 <= RACIO \text{ AND } checkModelFunction(R_P \setminus inactiveR))$
$R_P \leftarrow R_P \setminus (e_M \cup e_X)$	$w_i, score \text{ for } i \in R$	$R_P \leftarrow R_P \setminus inactiveR$
$endif$		$elseif(s1 == 0 \text{ AND } checkModelFunction(R_P \setminus inactiveR))$
$endifor$		$R_P \leftarrow R_P \setminus inactiveR$
$returnR_P$		$endif$
$endfunction$		$returnR_P$
		$endfunction$

2.3 Omics Data

Proteomics data used in this work were retrieved from the Human Protein Atlas (HPA) [26], which contains the profiles of human proteins in all major human healthy and cancer cells. We collected information for the liver tissue (hepatocytes) from HPA version 12 and Ensembl [7] version 73.37. After a conversion from Ensembl gene identifiers to gene symbols, duplicated genes with different evidence levels were removed (Table S1 from supplementary data)¹.

¹ All supplementary files are provided in <http://darwin.di.uminho.pt/epia2015>

Transcriptomics data were collected from Gene Expression Barcode (GEB) [16] (HG133plus2 (Human) cells v3). The conversion to gene expression levels was done considering the average level of probes for each gene. The mapping between probes and gene symbols was performed using the library “hgu133plus2.db” [4] from Bioconductor. The gene expression is classified as *High*, *Moderate* and *Low* if the gene expression evidence on that tissue is greater than 0.9, between 0.5 and 0.9, and between 0.1 and 0.5, respectively. The genes with expression evidence below 0.1 were considered not expressed in hepatocytes.

The reaction scores were obtained through the Gene-Protein-Rules present in the Recon2 model, based on the scores associated with each gene in the data. The reaction scores were calculated by taking the maximum (minimum) value of expression scores for genes connected by an “OR” (“AND”). If one of the gene scores is unknown, the other gene score is assumed in the conversion rule.

3 Results

To compare the metabolic models generated by the different algorithms and the effects of distinct omics data sources, we chose the reconstruction of hepatocytes metabolic models as our case study. Hepatocytes are the principal site of the metabolic conversions underlying the diverse physiological functions of the liver [10]. The hepatocytes metabolic models were generated using Recon2 as a template model and the GEB, HPA and the sets C_H and C_M from [12] as input data, for the three methods described in the previous section.

In the experiments, we seek to answer two main questions: Are omics data consistent across different data sources? What is the overlap of the resulting metabolic models obtained using different methods and different data sources? In 2010, a manually curated genome-scale metabolic network of human hepatocytes was presented, the HepatoNet1 [8], used as a reference in the validation process.

3.1 Omics Data Consistency

The HPA has evidence information related with 16324 genes in hepatocytes. The reliability of the data is also scored as “supportive” or “uncertain”, depending on similarity in immunostaining patterns and consistency with protein/gene characterization data. On the other hand, the GEB transcriptome has information for 20149 genes, of which 5772 have evidence of being expressed in hepatocytes.

Together, these two data sources have information for 21921 genes, but only 14552 are present in both (Figure 1A). Moreover, the number of genes with evidence of being expressed in the tissue in both sources is only of 3549, around 24% of all shared genes (Figure 1B). These numbers decrease significantly if using only HPA information marked as “supportive”. In this scenario, only 3868 genes are present also in GEB and only 1294 of them have expression evidence.

Next, evidence levels frequencies (*High*, *Moderate*, *Low*) were calculated across the GEB and HPA, as shown in Figure 2. Only a small number of genes have similar evidence levels in both data sources. Furthermore, a significant

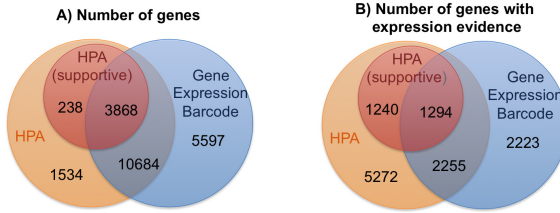


Fig. 1. A) Number of genes present in Gene Expression Barcode and Human Protein Atlas. In HPA, the number of genes with reliability “supportive” and “uncertain” are shown. B) Number of genes with evidence level “Low”, “Moderate” or “High” in HPA and gene expression evidence higher than 0 in Gene Expression Barcode.

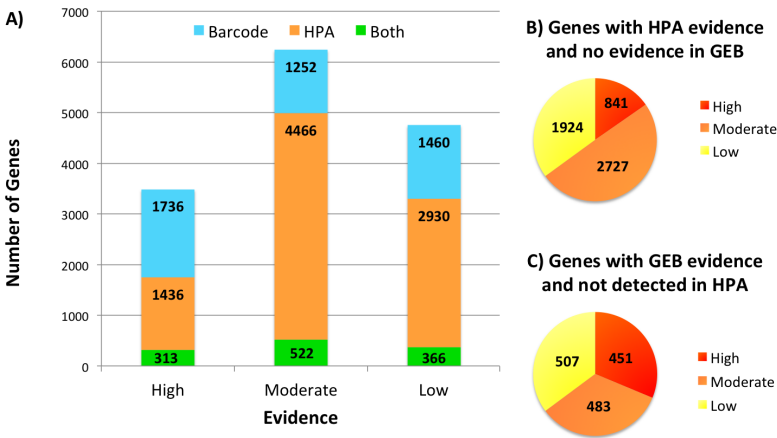


Fig. 2. A) Distribution of genes from Gene Expression Barcode project and Human Protein Atlas across the evidence levels - “High”, “Moderate” and “Low”. The ranges $[0.9, 1]$, $[0.5, 0.9[$ and $[0.1, 0.5[$ were used to classify the data into “Low”, “Moderate” and “High” levels. B) Genes with no evidence to be present in hepatocytes from GEB, but with evidence in the HPA. C) Genes with no evidence to be present in hepatocytes from HPA, but with evidence in GEB.

number of genes have contradictory levels of evidence - genes with expression evidence in one data source and not expressed in the other. If we focus only in the genes present in the model Recon2 with information in GEB and HPA (supportive), there are 15% of genes with “High” or “Moderate” evidence in one of the sources and not expressed in the other. This number increases to 22% if we also consider “Low” evidence level (Supplementary Figure S1).

The methods to reconstruct tissue-specific metabolic models use reaction scores calculated based on omics data to determine their inclusion in the final models. So, we analyzed the impact of these omics discrepancies in the values of reaction scores and compared those with the manually curated set C_H from Jerby et al. [12]. In Figure 3A, the poor overlap of the reaction scores calculated based

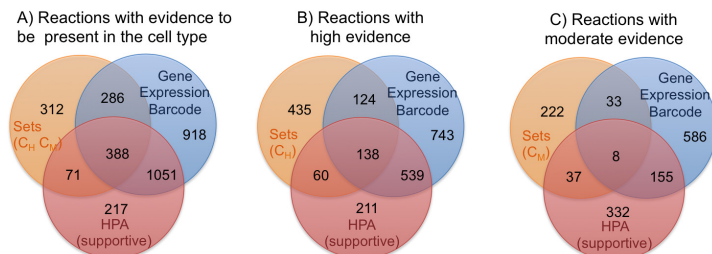


Fig. 3. A) Reactions with evidence that support their inclusion in the hepatocytes metabolic model. B) Number of reactions that have a high level of evidence of expression for each data source. C) Number of reactions that have a moderate evidence of expression for each data source

on different sources can be observed. Considering all data sources, 3243 reactions show some evidence that support their inclusion in the hepatocytes metabolic model, but only 388 are supported by all sources. The numbers are further dramatically reduced if we consider only moderate or high levels of evidence (Figure 3 B-C).

3.2 Metabolic Models

We applied each of the three algorithms to each omics data source, resulting in nine metabolic models for hepatocytes. In the application of mCADRE, we consider the list of key metabolites as published in the original article and a threshold of 0.5 to calculate the core set. A set of core metabolic tasks, that should occur in all cell types, was retrieved from [2] and used in the tINIT algorithm. The final MBA models were constructed based on 50 intermediate metabolic models. According to [12], a larger number would be desirable, but the time needed to generate each model prevented larger numbers of replicates. The detailed list of reactions that compose each metabolic model are available in supplementary material.

In Figure 4 A-C, we observe the consistency of the intermediate models generated by MBA, as well as the number of occurrences of reactions present in the final model. Moreover, Figure 4D shows the relations between the nine metabolic models generated through hierarchical clustering. The models obtained using the C_H and C_M sets as input data group together. Regarding the remaining, the mCADRE and MBA resulting models group according to their data (HPA and GEB), while the models created by tINIT cluster together. Overall, the data used as input seems to be the most relevant factor in the final result.

A more detailed comparison between the models reconstructed using the same algorithm or the same data source is available in Figure 5, A and B respectively. Considering the models generated by the same algorithm, it is observed that mCADRE has a smaller overlap (only 812 reactions) compared to the other methods. This could be explained by the possibility of removing core reactions during the mCADRE reconstruction process. Note that both reactions with

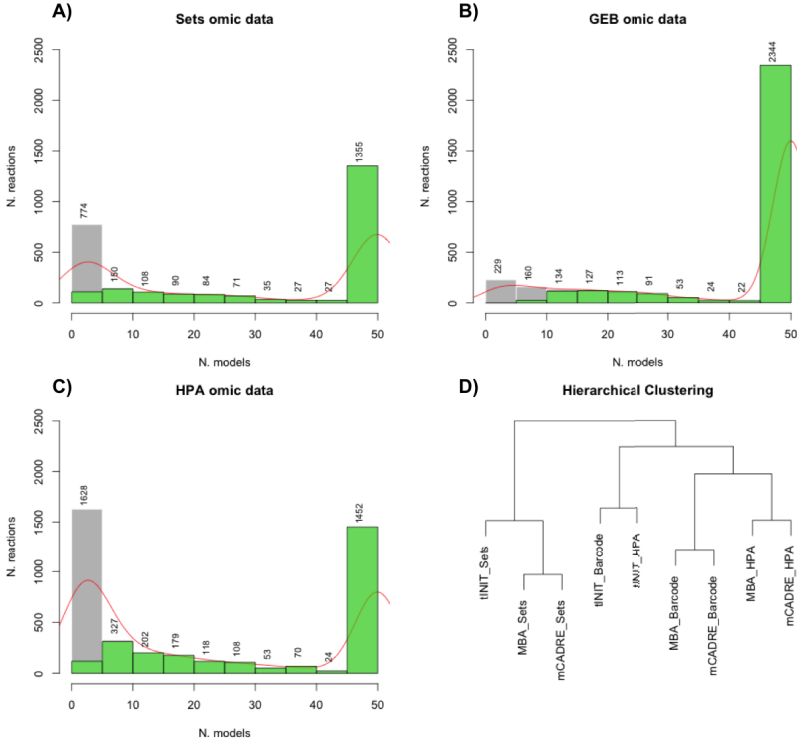


Fig. 4. A-C) Distribution of reactions across the 50 models for each data type. Grey bars show an histogram with the number of reactions present in a certain number of models. Green bars show the reactions that are present in the final model. D) Results from hierarchical clustering of the resulting nine models.

“High” and “Moderate” evidence levels, and from C_H and C_M sets, are all considered as belonging to the core. Furthermore, the mean of reactions that belong to all models of the same algorithm is around 45%. When the comparison is made by grouping models with the same input data, the variance between models is lower than grouping by algorithm. Here, the mean of reactions common to all models with the same data source is around 67% (Supplementary Table S3). Again, the variability of the final results seems to be dominated by the data source factor.

The quality of the metabolic models was further validated using the metabolic functions that are known to occur in hepatocytes [8]. The generic Recon2 human metabolic model, used as template in the reconstruction process, is able to satisfy 337 of the 408 metabolic functions available. Metabolic functions related with disease or involving metabolites not present in Recon2 were removed from the original list.

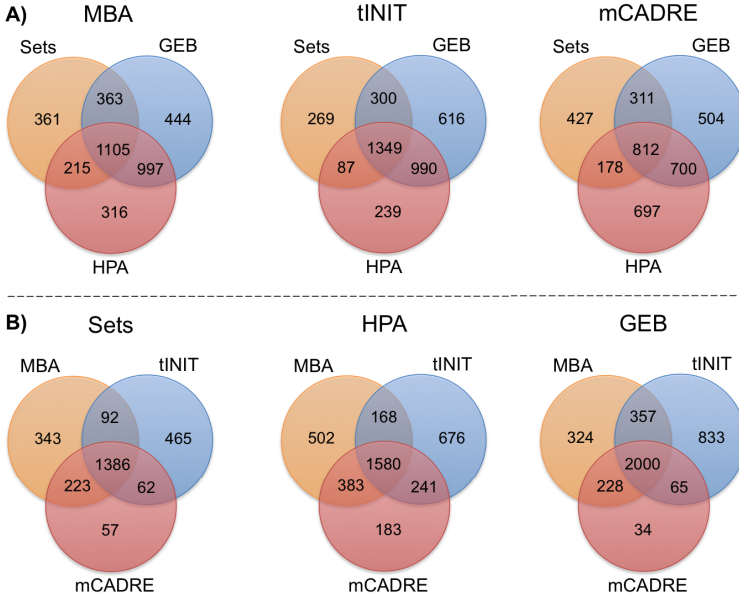


Fig. 5. Metabolic models reaction intersection considering: (A) the same algorithm; (B) the same omics data source.

Table 2. Number of reactions and the percentage of liver metabolic functions that each metabolic model performs when compared with the template model - Recon 2.

Method	Sets		HPA		GEB	
	N. Reac.	Tasks	N. Reac	Tasks	N. Reac.	Tasks
MBA	2044	18%	2633	24%	2909	6%
mCADRE	1728	2%	2387	3%	2327	4%
tINIT	2005	4%	2665	5%	3255	6%

The results of this functional validation, showing also the number of reactions in each metabolic model, are given on Table 2. These show that the number of satisfied metabolic tasks is very low compared with the manual curated metabolic model HepatoNet1. The metabolic model which performs the higher number of metabolic tasks was obtained using the MBA algorithm with the HPA evidence. Nevertheless, the success percentage is less than 25% when comparing with the performance by the template metabolic model - Recon2.

4 Conclusions

In this work, we present a survey of the most important methods for the reconstruction of tissue-specific metabolic models. Each method was proposed to use

different data sources as input. Here, we analyze the consistency of information across important omics data sources used in this context and verify the impact of such differences in the final metabolic models generated by the methods.

The results show that metabolic models obtained depend more on the data sources used as inputs, than on the algorithm used for the reconstruction. To validate the accuracy of the obtained metabolic models, a set of metabolic functions that should be performed in hepatocytes was tested for each metabolic model. We found that the number of satisfied liver metabolic functions was surprisingly low. This shows that methods for the reconstruction of tissue-specific metabolic models based on a single omics data source are not enough to generate high quality metabolic models. Methods to combine several omics data sources to rank the reactions for the reconstruction process could be a solution to improve the results of these methods. Indeed, this study emphasizes the need for the development of reliable methods for omics data integration, which seem to be required to support the reconstruction of complex models of human cells, but also reinforce the need to be able to incorporate known phenotypical data available from literature or human experts.

Acknowledgments. S.C. thanks the FCT for the Ph.D. Grant SFRH/BD/80925/2011. The authors thank the FCT Strategic Project of UID/BIO/04469/2013 unit, the project RECI/BBB-EBI/0179/2012 (FCOMP-01-0124-FEDER-027462) and the project “BioInd - Biotechnology and Bioengineering for improved Industrial and Agro-Food processes”, REF. NORTE-07-0124-FEDER-000028 Co-funded by the Programa Operacional Regional do Norte (ON.2 - O Novo Norte), QREN, FEDER.

References

1. 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 Computational Biology* **8**(5), e1002518 (2012)
2. Agren, R., Mardinoglu, A., Asplund, A., Kampf, C., Uhlen, M., Nielsen, J.: Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Molecular Systems Biology* **10**, 721 (2014)
3. Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., et al.: NCBI GEO: archive for functional genomics data sets - 10 years on. *Nucleic Acids Research* **39**(suppl 1), D1005–D1010 (2011)
4. Carlson, M.: hgu133plus2.db: Affymetrix Human Genome U133 Plus 2.0 Array annotation data (chip hgu133plus2) (2014). r package version 3.0.0
5. Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srivas, R., Palsson, B.O.: Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences of the United States of America* **104**(6), 1777–1782 (2007)
6. Duarte, N.C., Herrgård, M.J., Palsson, B.O.: Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research* **14**(7), 1298–1309 (2004)

7. Flicek, P., Amode, M.R., Barrell, D., et al.: Ensembl 2014. *Nucleic Acids Research* **42**(D1), D749–D755 (2014)
8. Gille, C., Bölling, C., Hoppe, A., et al.: HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Molecular Systems Biology* **6**(411), 411 (2010)
9. Hao, T., Ma, H.W., Zhao, X.M., Goryanin, I.: Compartmentalization of the Edinburgh Human Metabolic Network. *BMC Bioinformatics* **11**, 393 (2010)
10. Ishibashi, H., Nakamura, M., Komori, A., Migita, K., Shimoda, S.: Liver architecture, cell function, and disease. *Seminars in Immunopathology* **31**(3) (2009)
11. Jerby, L., Ruppin, E.: Predicting Drug Targets and Biomarkers of Cancer via Genome-Scale Metabolic Modeling. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* **18**(20), 5572–5584 (2012)
12. Jerby, L., Shlomi, T., Ruppin, E.: Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Molecular Systems Biology* **6**(401), 401 (2010)
13. Kaddurah-Daouk, R., Kristal, B., Weinshilboum, R.: Metabolomics: a global biochemical approach to drug response and disease. *Annu. Rev. Pharmacol. Toxicol.* **48**, 653–683 (2008)
14. Lewis, N.E., Schramm, G., Bordbar, A., Schellenberger, J., Andersen, M.P., Cheng, J.K., Patel, N., Yee, A., Lewis, R.A., Eils, R., König, R., Palsson, B.O.: Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nature Biotechnology* **28**(12), 1279–1285 (2010)
15. Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Uhlen, M., Nielsen, J.: Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nature Communications* **5**, Jan 2014
16. McCall, M.N., Jaffee, H.A., Zelisko, S.J., Sinha, N., et al.: The Gene Expression Barcode 3.0: improved data processing and mining tools. *Nucleic Acids Research* **42**(D1), D938–D943 (2014)
17. Oberhardt, M.A., Palsson, B.O., Papin, J.A.: Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology* **5**(320), 320 (2009)
18. Orth, J.D., Thiele, I., Palsson, B.O.: What is flux balance analysis? *Nature Biotechnology* **28**(3), 245–248 (2010)
19. Parkinson, H., Sarkans, U., Shojatalab, M., Abeygunawardena, N., et al.: ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Research* **33**(Database issue), Jan 2005
20. Reed, J.L., Vo, T.D., Schilling, C.H., Palsson, B.O.: An expanded genome-scale model of *Escherichia coli* K-12 (i JR904 GSM / GPR) **4**(9), 1–12 (2003)
21. Sahoo, S., Franzson, L., Jonsson, J.J., Thiele, I.: A compendium of inborn errors of metabolism mapped onto the human metabolic network. *Mol. BioSyst.* **8**(10), 2545–2558 (2012)
22. Sahoo, S., Thiele, I.: Predicting the impact of diet and enzymopathies on human small intestinal epithelial cells. *Human Molecular Genetics* **22**(13), 2705–2722 (2013)
23. Shlomi, T., Benyamini, T., Gottlieb, E., Sharan, R., Ruppin, E.: Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the Warburg effect. *PLoS Computational Biology* **7**(3), e1002018 (2011)
24. Shlomi, T., Cabili, M.N., Ruppin, E.: Predicting metabolic biomarkers of human inborn errors of metabolism. *Molecular Systems Biology* **5**(263), 263 (2009)
25. Thiele, I., Swainston, N., Fleming, R.M.T., et al.: A community-driven global reconstruction of human metabolism. *Nature Biotechnology* **31**(5), May 2013

26. Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., et al.: Towards a knowledge-based Human Protein Atlas. *Nat Biotech* **28**(12), 1248–1250 (2010)
27. Wang, Y., Eddy, J.A., Price, N.D.: Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Systems Biology* **6**(1), 153 (2012)
28. Wishart, D.S., Knox, C., Guo, A.C., Eisner, R., et al.: HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Research* **37**(suppl 1), Jan 2009
29. Yizhak, K., Le Dévédec, S.E., Rogkoti, V.M.M., et al.: A computational study of the Warburg effect identifies metabolic targets inhibiting cancer migration. *Molecular Systems Biology* **10**(8) (2014)