Integrative Biology

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/ibiology

FRONTIER

Reconstruction of genome-scale human metabolic models using omics data

Jae Yong Ryu,‡[a] Hyun Uk Kim‡[a,b,d] and Sang Yup Lee*[a,b,c,d]

[a] Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Center for Systems and Synthetic Biotechnology, Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea.

[b] BioInformatics Research Center, KAIST, Daejeon 305-701, Republic of Korea.

[c] BioProcess Engineering Research Center, KAIST, Daejeon 305-701, Republic of Korea.

[d] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm, Denmark.

*e-mail: leesy@kaist.ac.kr

‡ These authors contributed equally to this work.

Abstract

The impact of genome-scale human metabolic models on human systems biology and medical sciences is becoming greater, thanks to increasing volumes of model building platforms and publicly available omics data. The genome-scale human metabolic models started with Recon 1 in 2007, and have since been used to describe metabolic phenotypes of healthy and diseased human tissues and cells, and to predict therapeutic targets. Here we review recent trends in genome-scale human metabolic modeling, including various generic and tissue/cell type-specific human metabolic models developed to date, and the methods, databases and platforms used to construct them. For generic human metabolic models, we pay attention to Recon 2 and HMR 2.0 with emphasis on data sources used to construct them. Draft and high-quality tissue/cell type-specific human metabolic models have also been generated using these generic human metabolic models. Integration of tissue/cell type-specific omics data with the generic human metabolic models is the key step, and we discuss omics data and their integration methods to achieve this task. Initial version of the tissue/cell type-specific human metabolic models can further be computationally refined through gap filling, reaction directionality assignment and subcellular localization of metabolic reactions. We review relevant tools for this model refinement procedure as well. Finally, we suggest the direction of further studies on reconstructing an improved human metabolic model.

Introduction

Genome-scale models (GSMs) of metabolism continue to be an important tool in systems biology by providing snapshots of the global metabolism under given genetic and environmental conditions. A fundamental principle of GSM reconstruction lies in mass balance of metabolites, the assumption of pseudo-steady state and the use of a stoichiometric matrix to run simulations using numerical optimization[1]. This modeling approach is called constraints-based modeling or more formerly constraint-based reconstruction and analysis (COBRA) because genetic, environmental and physicochemical variables (*e.g.*, gene inactivation, specific nutrient availability and reaction directionality, respectively) are incorporated into the GSMs in the form of *constraints* which then are numerically considered during optimization procedure for the prediction of intracellular flux values. Due to its ease of implementation and relatively high predictive power, especially for microorganisms, constraints-based metabolic modeling and simulation have contributed to a diverse array of applications in the fields of systems biology and metabolic engineering, for example prediction of gene manipulation targets in metabolic engineering[2], and prediction of drug targets in microbial pathogens[3] and abnormal human cells (*e.g.*, hepatocytes from patients with non-alcoholic fatty liver disease)[4].

Advances in metabolic modeling and the increasing availability of high-quality omics information have enabled construction of models not only for prokaryotes, but also for higher organisms, including eukaryotes. The GSM of the most widely employed bacterium *Escherichia coli* was reported in 2000 for the first time[5], while the first GSM of eukaryotic metabolism came from *Saccharomyces cerevisiae* in 2003[6]. In addition to many more GSMs for a number of organisms developed since then, the first GSM of human metabolism (Recon 1) was released in 2007. This human model was created by thorough manual curation of

biochemical data generated over more than half a century[7]. Recon 1 subsequently ignited further studies of constraints-based metabolic modeling specific to human, in particular development of omics data integration methods to build tissue/cell type-specific metabolic models, as well as model refinement methods to improve the quality of the model. It has become apparent that novel methods of constraints-based modeling and simulation are increasingly being developed for human systems to address the biochemical and genetic complexity of human metabolism because conventional constraints-based modeling and simulation methods primarily developed for microorganisms cannot be directly applied to the human models[8]. Among many relevant challenges, omics data integration with metabolic models is one of more important challenges in human metabolic modeling in order to generate tissue/cell type-specific metabolic models. Subcellular localization of metabolic reactions is also critical step in this field, and gene-protein-reaction (GPR) information needs to be more carefully refined with consideration of alternative splicing of each genes. Such challenges need to be explicitly discussed.

Here we review recent trends in human metabolic modeling with emphasis on tissue/cell-specific human metabolic models developed so far, and the methods used for their construction and refinement; high-throughput tools employed to build functional human metabolic models through the use of omics data and the methods to computationally fine-tune the initial version of the model are reviewed. Also, we discuss current challenges to further improving the human metabolic modeling approaches using omics data. More refined human metabolic models will certainly contribute to better understanding and treatment of various metabolic disorders (*e.g.*, diabetes) and cancers that are highly associated with abnormal expression patterns of metabolic genes.

Current status of genome-scale human metabolic models

A unique aspect of human metabolic models in comparison with those of microorganisms is the requirement of *generic* metabolic models, which intend to cover all the known metabolic reactions according to the entire human genome. An obvious reason for this is that cells of different tissues from one person possess the same genomic information, but show differentiated, tissue-specific gene expression patterns. Such feature suggests the importance of developing a high quality generic metabolic model, as this generic version serves as a template model to derive tissue/cell-type-specific metabolic models based on the gene expression and other omics data. Context-specific models refer to such metabolic models derived from the original generic model, or models developed and simulated under specified genetic and/or environmental conditions. Obviously, the context-specific metabolic models will show poor performance if the generic models used are of low quality. Consequently, the early versions of genome-scale human metabolic models, Recon 1[7] and Edinburgh Human Metabolic Network (EHMN)[9], were constructed through thoughtfully planned manual curation in order to validate each metabolic reaction to the greatest extent.

At the moment, the most comprehensive generic human metabolic models are Recon 2[10] and HMR 2.0[4]. These two generic models were created by integrating previously developed models and incorporating additional and updated biochemical content obtained from databases and literature. Recon 2 was developed by merging the metabolic contents of EHMN, HepatoNet1[11], an acyl carnitine and fatty-acid oxidation module[12], and a human small intestinal enterocyte model[13] with those of Recon 1. In a similar manner, HMR 2.0 was developed by merging metabolic information from several comprehensive human metabolic models such as Recon 1, EHMN, HepatoNet1, iHuman1512[14] and iAdipocytes1809[15], as well as from the four major metabolic databases: KEGG[16], HumanCyc[17], LIPID MAPS

Lipidomics Gateway[18] and REACTOME[19]. In particular, the HMR 2.0 intended to strengthen its description of lipid metabolism. The objective behind the development of these two generic human metabolic models is to complete the models as comprehensively as possible by covering virtually all the experimentally and theoretically known metabolic reactions. As of January 2015, Recon 2 and HMR 2.0 cover 1789 genes (7440 reaction and 5063 metabolites) and 3765 genes (8181 reactions and 5546 metabolites), respectively. The Recon and HMR series are regularly updated, and their COBRA-compliant SBML and .mat files are provided through their repositories, Recon X (http://humanmetabolism.org/) and Human Metabolic Atlas (http://humanmetabolism.org/), respectively. In addition, the Human Metabolic Atlas provides various context-specific human models derived from the HMR 2.0.

Integration of tissue/cell type-specific omics data with the generic human metabolic model has become a key practice to generate high-quality, context-specific metabolic models[20] (Fig. 1). Such metabolic models generated from Recon 1 include: hepatic metabolic models by Jerby et al.[21] and HepatoNet1[11]; a multi-cellular brain metabolic model consisting of three neuron cells and astrocyte[22]; a kidney metabolic model[23]; the alveolar macrophage model iAB-AMØ-1410[24]; a multi-tissue metabolic model consisting of hepatocyte, myocyte and adipocyte[25]; a heart specific model by Zhao and Huang[26]; a cardiomyocyte model CardioNet[27]; and a Human Embryonic Kidney (HEK) cell culture model[28]. Those derived from HMR include iAdipocytes1850[15,29] and iHepatocytes2322[4] (Fig. 2).

Application of tissue/cell type-specific omics data to the generic human metabolic models has also led to the construction of large sets of context-specific metabolic models, largely draft versions, for various human tissues and cells[10,14,30], cancer metabolic models[14,20,31,32] and personalized metabolic models of hepatocellular carcinoma patients[33] (Fig. 2). Such context-specific metabolic models have addressed some of medically important

diseases in modern society. Cancer metabolism has in particular been an ideal application target of human metabolic models because of the rapid biomass generation, mutations in metabolic genes and the Warburg effect (aerobic glycolysis) in cancers which can be effectively simulated using constraints-based flux analysis. A representative application was the generation of six personalized genome-scale hepatocellular carcinoma metabolic models[33] using the HMR 2.0 and personal proteome data. This investigation led to the identification of antimetabolites that effectively inhibit cancer cells by binding to enzymes consuming structurally similar native substrates. More recently, oncometabolites, excessively accumulated natural metabolites or abnormal metabolites typically absent in healthy cells, which cause the onset of cancers, have been predicted using nine tissue-specific cancer and normal metabolic models using Recon 2 as a template[34]. Other metabolic diseases such as obesity and non-alcoholic fatty liver disease have also been rigorously studied using adipocyte and hepatocyte metabolic models, iAdipocytes1809 and iHepatocytes2322, respectively[4,15]. In both studies, metabolic profiles emerged from diseased and normal condition-specific omics data were compared, thereby identifying potential biomarkers and therapeutic targets. It is obvious that more of metabolic diseases will be subjected to these kinds of systematic analyses using human GSMs.

Increasing availability of high-quality omics data has greatly facilitated studies generating both high-quality and draft context-specific models, especially since 2010. Such draft context-specific metabolic models transform into high-quality models upon manual curation of the relevant literature and/or based on the simulation results obtained with the predefined tissue/cell-type-specific metabolic tasks. Metabolic tasks are defined to be (in)activation status of specific metabolic reactions that are experimentally verified under a particular condition (*e.g.*, conversion of ammonia into urea, glutamine and alanine in liver),

and the reconstructed metabolic models are supposed to correctly predict the defined reaction activity status. An important pattern is that a greater number of metabolic tasks have been considered to be correctly predicted in upgraded versions of Recon series and hepatocyte metabolic models, suggesting the improved model quality in latter versions (Fig. 3). The results reported so far suggest that the selection of omics data and the methods for their integration with the human generic metabolic models are key determinants of the quality of context-specific metabolic models.

Use of state-of-the-art high-throughput techniques and data for human metabolic modeling

High-throughput methods to build functional context-specific human metabolic models

Several context-specific modeling and simulation methods have been developed and are important assets to human metabolic modeling and simulation (Fig. 1). At the moment, the COBRA Toolbox[35], the RAVEN Toolbox[36] and the COBRApy[37] package serve as representative platforms for building, manipulating, and/or simulating metabolic models. They provide built-in functions to generate context-specific models from the generic metabolic models through previously developed algorithms, including GIMME[38] and iMAT[39] in the COBRA Toolbox, and INIT[14] and tINIT[33] in the RAVEN Toolbox. These algorithms allow building context-specific models aiming at maximizing the degree of consistency between the activity status of reactions and their respective gene/protein expression levels in the examined omics data.

These algorithms are different in their characteristics as follows. The GIMME algorithm directly edits the model by removing and re-inserting reactions based on their gene/protein expression profiles and predefined target metabolic functionalities, whereas

iMAT uses mixed-integer linear programming (MILP) to constrain reaction fluxes to maximize their consistency with the expression profiles without model editing. Recently, metabolic phenotype analysis (MPA)[32] was developed to build a breast cancer metabolic model which categorizes reactions into high, low and moderate activities (as with iMAT), and additionally employs a scoring scheme to obtain flux profiles maximally consistent with examined omics data. Similar to the iMAT method, INIT uses MILP and experimental omics data, but takes human proteome data as the primary input and additionally considers tissue-specific gene expression and metabolome data[14]. The tINIT algorithm is an extended version of INIT by enabling the target context-specific models to accomplish tissue/cell type-specific metabolic functions[33].

There also exist context-specific modeling methods that are not built-in functions of the abovementioned platforms; nevertheless, many of them are compatible with these platforms. For instance, GIM^{3}E run in COBRApy considers metabolomics data to make sure that detected metabolites in the metabolomics data are used in the simulated metabolic flux distributions. Subsequently, transcriptomics data is used to further adjust flux values by imposing penalties on reactions whose fluxes are not consistent with gene expression profiles[40]. However, this method is yet to be demonstrated with human models. Another set of algorithms including MBA[21], mCADRE[30], and FASTCORE[41] are also available to generate context-specific models. They all have in common that they first define a *core* set of reactions that have strong support for being active in a specific tissue or cell, and therefore include them in the target context-specific models created. The major difference among them lies in their subsequent strategies to generate the final context-specific metabolic models. Finally, a recently reported PRIME algorithm generates context-specific models by identifying genes whose expression profiles are significantly associated with the specific cell's phenotype (i.e.,

a linear relationship) and modifying the upper bounds of the respective reaction fluxes[42].

Among these several tools to generate context-specific metabolic models, it is difficult to assess which one is the best due to their different logics, as discussed by Vlassis et al[41]. Rather, it is important that the same method is consistently used throughout the process of constructing target context-specific models in order to perform comparative studies (*e.g.*, normal *versus* cancerous cells). More recently developed algorithms including mCADRE and FASTCORE are designed to overcome limitations observed in the methods developed in the past, such as long computing time to generate context-specific models. Finally, although manual curation is an important procedure to improve the model quality, the use of appropriate context-specific modeling methods with omics data can still achieve high-quality models even without manual curation. For example, application of MBA to Recon 1 resulted in a liver metabolic model with reasonably good quality as quantitatively validated[21], while use of the iMAT with Recon 2 produced 65 draft cell type-specific metabolic models[10] (Fig. 2). More detailed discussions on the omics data integration methods are available elsewhere[43-47].

Further computational refinement of context-specific human metabolic models

Initial versions of functional context-specific models can be refined through computational tools (Fig. 4). First, improvement can be made through gap filling procedure, which is necessary due to our incomplete knowledge on a target organism's metabolism. In fact, gap filling procedures have been a longstanding topic in metabolic modeling to improve model fidelity[48]. Both the COBRA and RAVEN Toolboxes provide built-in functions to fill in biochemical gaps of metabolic pathways. GapFind and GapFill are built-in functions of the COBRA Toolbox, which fulfill this gap-filling procedure. GapFind identifies every gap by

searching for blocked (or dead-end) metabolites, and GapFill minimizes the number of gaps with minimal modifications through changing of reaction reversibilities and/or addition of gap-filling reactions[49]. The RAVEN Toolbox provides similar built-in functions that check metabolites that can be consumed and/or produced, identify internal loops in model, and remove the biochemical gaps by adding a minimal set of reactions. A recently reported fastGapFill, which runs on the COBRA Toolbox, is the first algorithm designed for compartmentalized and large-scale metabolic models such as Recon 2[50]. Application of fastGapFill to the Recon 2 model resulted in successful identification of 400 gap-filling reactions.

Another important element of the GSM reconstruction is the assignment of reaction directionality that highly affects the prediction of intracellular flux distributions. A relevant software package is the von Bertalanffy series (current version 2.0) that runs under the COBRA Toolbox. This package quantitatively assigns reaction directionality in multi-compartment metabolic models by considering experimental and estimated thermodynamic data, metabolite structures, temperature, pH, electrical potential, ionic strength and metabolomic data[51]. It was applied to determine reaction directionality of the Recon 1 model, including intercompartmental transport reactions[52]. The recently reported von Bertalanffy 2.0 adopts a new framework called component contribution method that appears to better estimate the standard Gibbs energy of metabolic reactions compared to previous methods[53]. Another complementary approach to assign reaction directionality is the analysis of metabolite patterns in the metabolic network[54], but its predictive power in comparison with the component contribution method is yet to be studied.

Because of the presence of intracellular organelles in human cells, it is important to accurately assign metabolic reactions to their respective organelles and to provide

intercompartmental transporter reactions to link reactions occurring in different organelles in the human metabolic model. The process of assigning metabolic reactions to the correct organelles, also known as subcellular localization, can be predicted from gene sequences. There exist several organelle-specific prediction tools, including WoLF PSORT[55] and CELLO[56] in RAVEN Toolbox. Moreover, BRENDA[57] provides literature-based subcellular localization information. Furthermore, the recently released COMPARTMENTS[58] database also provides integrated localization information obtained from literature, as well as high-throughput microscopy-based and theoretically predicted data (*e.g.*, search of signal sequences and text mining of relevant literatures). The Human Protein Atlas (HPA; http://www.proteinatlas.org/)[59,60] is another important database that provides experimental subcellular localization data obtained by immunostaining experiments. However, it should be noted that both intercompartmental and membrane transporter reactions that link reactions across different organelles and between the extracellular space and the intracellular cytosol, respectively, cannot be precisely predicted from gene sequences alone. Thus, manual curation is critical during the process of adding transporter reactions. In case of the Recon 2 model, 70 transporter reactions were newly added and GPR associations of 24 transporter reactions were updated through manual curation of literature[61].

Finally, in addition to the aforementioned three stand-alone platforms, BioMet Toolbox (http://www.biomet-toolbox.org/)[62] and MetaNetX (http://metanetx.org/)[63] are web-based platforms that can also be used to improve the quality of context-specific metabolic models in the web environment. Among several useful functions for metabolic models and omics data analysis, BioMet Toolbox has functions to check the elemental balance within a reaction and to identify dead-end reactions and metabolites in GSMs. Likewise, MetaNetX provides functions to identify dead-ends, fill in biochemical gaps of metabolic pathways and predict

the direction of each reaction in metabolic models. Both web-based platforms are intuitive and can be particularly useful for biologists who are not familiar with the command-line interface.

Next generation omics data for high-quality metabolic models

Along with omics data integration methods, the omics data itself is becoming more important to improve the quality of context-specific human metabolic models. With remarkable advances in high-throughput techniques, several public databases provide high-quality omics data across the genome, transcriptome, proteome and metabolome. In the field of cancer genomics that has been one of the major topics of human systems biology studies, The Cancer Genome Atlas (TCGA; http://cancergenome.nih.gov/) and Catalogue Of Somatic Mutations In Cancer (COSMIC; http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/)[64] provide useful cancer genomic data that can be used to predict the effects of cancer-associated genetic mutations on human metabolism using metabolic models[34]. TCGA database contains data on 34 types of tumors from more than 11,000 patients covering genome sequences, expression profiles (*e.g.*, RNA-Seq) and somatic mutations. In addition to the cancer genomics, RNA-Seq profiles of 16 normal human tissues (Gene Expression Omnibus accession code: GSE30611) have been released from the Illumina Human Body Map 2.0 project.

As for the human proteome, the HPA[59,60] and ProteomicsDB (https://www.proteomicsdb.org) are useful databases; proteome data from the HPA served as primary input data for the tINIT algorithm to create personalized human metabolic models[33]. The HPA database contains protein expression data for normal and tumor tissues/cells in humans, obtained from a variety of proteomics techniques such as immunohistochemistry, immunofluorescence, western blot, protein arrays (*i.e.*, antigen microarrays), RNA-Seq and

manual curation of literature. Meanwhile, ProteomicsDB provides proteomic data obtained using mass-spectrometry that shows a high coverage of human proteins and their quantitative expression levels. At present, ProteomicsDB covers 93% and 24% of all known human proteins and splice isoforms, respectively, obtained from various tissues, cell lines, and body fluids.

The availability of such a large volume of omics data and its effective use in human metabolic modeling requires several considerations. First, GPR association of the generic human metabolic model needs to be regularly updated according to the up-to-date contents of major metabolic databases and literature. This is the common objective of the Recon X and Human Metabolic Atlas. Second, heterogeneous data from several different databases, such as KEGG, MetaCyc, HumanCyc, REACTOME, and SMPDB[65], need to be effectively incorporated into the generic human metabolic model by resolving any inconsistencies existing among them. To accomplish this, the use of consistent identifiers for metabolites and reactions can be useful. Recently, the database MNXref was released through MetaNetX, providing reconciled information for metabolites, biochemical reactions and compartments adopted from several different databases using a chemoinformatics approach[66]. A recent example of using MNXref is the draft metabolic models generated from the Path2Models project having their model contents described with MNXref identifiers[67]. Third, human metabolic models need to have more detailed genetic information (*i.e.*, GPR associations) up to a splice isoform level (Fig. 4). Current human metabolic models have gene-level information, but human genes are additionally spliced into alternative isoforms that have important implications in pathologies. For instance, compared to healthy cells, cancerous cells have different expression levels of each splice isoform for particular genes, and effects of the splice isoform-level changes cannot be predicted using metabolic models having only

gene-level information. Because current omics techniques such as RNA-Seq provide splice isoform-level expression profiles, human metabolic models with refined GPR association accordingly will make predictions of human metabolism more precise. In this step, the APPRIS database, which provides annotation of human splice isoforms, can be useful to define the splice isoform-level GPR association of the human metabolic models[68]. The annotations in APPRIS currently cover splice isoforms generated from 85% of human protein-encoding genes. Fourth, next generation omics data available at diverse databases, including those mentioned above, should be more actively used to generate and simulate context-specific metabolic models. A large fraction of the RNA-Seq data publicly available still remains unused in human metabolic modeling despite many databases providing many sets of such data in various forms. Examples of data useful for building context-specific metabolic models include quantitative expression levels of individual splice isoforms for each gene from the TCGA, and probability of changes in expression levels of splice isoforms between tumor and non-tumor cells from the TCGA SpliceSeq database. In particular, personalized context-specific metabolic modeling will benefit from patient-specific omics data available at the TCGA, potentially providing patient-specific therapeutic targets. Finally, although genomic and transcriptome data for specific cell and tissue types are increasingly available, omics data for blood vessels, preferably large-scale metabolome data, are relatively very rare. This will certainly bottleneck the advance of whole body metabolism modeling because individual cell and tissue models can be connected via metabolites flowing through the blood vessel[22,25].

Prediction of human metabolic fluxes

Upon generation of context-specific metabolic models, they can be subjected to constraints-

based flux analysis to predict intracellular flux distributions. In contrast to microorganisms, however, human metabolism lacks clear objectives that can be considered for the constraints-based simulation, especially for healthy cells and tissues, because their growths can be hardly defined and various objective functions have not been systematically examined in human system[69]. For this reason, human metabolic models are often simulated to examine whether they achieve the predefined metabolic tasks (Fig. 3). It should be noted that metabolic tasks cannot be considered as a true representation of genome-scale flux distributions because they mostly focus on single reactions rather than a set of many reactions (*e.g.*, biomass generation)[10]. In the past, minimization of internal fluxes was used to predict genome-scale flux distributions while reflecting the near-zero growth rate of healthy human cells[11]. This method is based on an assumption that cells are evolved to achieve specific metabolic tasks with efficient use of energy[70]. Cancer metabolism having a rapid growth rate is an exceptional case and biomass maximization is considered to be a safe assumption[71]. Finally, some of the aforementioned context-specific modeling algorithms can also be used to predict intracellular fluxes using tissue/cell type-specific omics data in addition to generating context-specific models, including iMAT, GIMME, GIM^{3}E, INIT and tINIT[44]. However, these methods have not been extensively used to conduct pathway-level analysis using genome-scale flux distributions of healthy and diseased cells/tissues, and await more rigorous experimental validations.

Conclusions

Generation of context-specific metabolic models using generic human metabolic models has now become common practice for prediction of metabolic phenotypes of specific tissues and cells under varied health statuses, and for identification of therapeutic targets. With

continually updated generic human metabolic models, increasing volumes of next generation omics data and improvements in the metabolic modeling platforms, the quality and number of resulting context-specific metabolic models will also increase accordingly. We therefore reviewed various human metabolic models developed so far, as well as recent tools available for each process of context-specific human metabolic modeling. Furthermore, we emphasize the use of newly available omics data by incorporating them into the context-specific modeling process, for example splice isoform-level gene expression profiles form RNA-Seq. Through this approach, high-quality context-specific metabolic models will be more efficiently generated, which will be useful for studying human metabolism in genome-scale.

Acknowledgements

We would like to thank Martin Gustavsson for his critical comments on the manuscript. This work was supported by the Bio-Synergy Research Project (2012M3A9C4048759) of the Ministry of Science, ICT and Future Planning through the National Research Foundation. This work was also supported by the Novo Nordisk Foundation.

Competing financial interests

The authors declare no competing financial interests.

References

1 Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245-248 (2010).

2 Park, J. M. *et al.* Flux variability scanning based on enforced objective flux for identifying gene amplification targets. *BMC Syst. Biol.* **6**, 106 (2012).

3 Kim, H. U. *et al.* Integrative genome-scale metabolic analysis of *Vibrio vulnificus* for drug targeting and discovery. *Mol. Syst. Biol.* **7**, 460 (2011).

4 Mardinoglu, A. *et al.* Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat. Commun.* **5**, 3083 (2014).
5 Edwards, J. S. & Palsson, B. O. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA* **97**, 5528-5533 (2000).
6 Forster, J., Famili, I., Fu, P., Palsson, B. O. & Nielsen, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**, 244-253 (2003).
7 Duarte, N. C. *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA* **104**, 1777-1782 (2007).
8 Mardinoglu, A., Gatto, F. & Nielsen, J. Genome-scale modeling of human metabolism - a systems biology approach. *Biotechnol. J.* **8**, 985-996 (2013).
9 Hao, T., Ma, H. W., Zhao, X. M. & Goryanin, I. Compartmentalization of the Edinburgh Human Metabolic Network. *BMC Bioinformatics* **11**, 393 (2010).
10 Thiele, I. *et al.* A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* **31**, 419-425 (2013).
11 Gille, C. *et al.* HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol. Syst. Biol.* **6**, 411 (2010).
12 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**, 2545-2558 (2012).
13 Sahoo, S. & Thiele, I. Predicting the impact of diet and enzymopathies on human small intestinal epithelial cells. *Hum. Mol. Genet.* **22**, 2705-2722 (2013).
14 Agren, R. *et al.* Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput. Biol.* **8** (2012).
15 Mardinoglu, A. *et al.* Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Mol. Syst. Biol.* **9**, 649 (2013).
16 Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109-114 (2012).
17 Romero, P. *et al.* Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.* **6**, R2 (2005).
18 Fahy, E. *et al.* Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.* **50** S9-14 (2009).
19 Croft, D. Building models using Reactome pathways as templates. *Methods Mol. Biol.* **1021**, 273-283 (2013).
20 Jerby, L. & Ruppin, E. Predicting drug targets and biomarkers of cancer via genome-scale metabolic modeling. *Clin. Cancer Res.* **18**, 5572-5584 (2012).
21 Jerby, L., Shlomi, T. & Ruppin, E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol. Syst. Biol.* **6**, 401 (2010).
22 Lewis, N. E. *et al.* Large-scale *in silico* modeling of metabolic interactions between cell types in the human brain. *Nat. Biotechnol.* **28**, 1279-1285 (2010).
23 Chang, R. L., Xie, L., Xie, L., Bourne, P. E. & Palsson, B. O. Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comput. Biol.* **6**, e1000938 (2010).
24 Bordbar, A., Lewis, N. E., Schellenberger, J., Palsson, B. O. & Jamshidi, N. Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic

reconstructions. *Mol. Syst. Biol.* **6**, 422 (2010).
25 Bordbar, A. *et al.* A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. *BMC Syst. Biol.* **5**, 180 (2011).
26 Zhao, Y. & Huang, J. Reconstruction and analysis of human heart-specific metabolic network based on transcriptome and proteome data. *Biochem. Biophys. Res. Commun.* **415**, 450-454 (2011).
27 Karlstadt, A. *et al.* CardioNet: a human metabolic network suited for the study of cardiomyocyte metabolism. *BMC Syst. Biol.* **6**, 114 (2012).
28 Quek, L. E. *et al.* Reducing Recon 2 for steady-state flux analysis of HEK cell culture. *J. Biotechnol.* **184**, 172-178 (2014).
29 Mardinoglu, A. *et al.* Defining the human adipose tissue proteome to reveal metabolic alterations in obesity. *J. Proteome. Res.* **13**, 5106-5119 (2014).
30 Wang, Y. L., Eddy, J. A. & Price, N. D. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst. Biol.* **6** (2012).
31 Frezza, C. *et al.* Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature* **477**, 225-228 (2011).
32 Jerby, L. *et al.* Metabolic associations of reduced proliferation and oxidative stress in advanced breast cancer. *Cancer Res.* **72**, 5712-5720 (2012).
33 Agren, R. *et al.* Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol. Syst. Biol.* **10** (2014).
34 Nam, H. *et al.* A systems approach to predict oncometabolites via context-specific genome-scale metabolic networks. *PLoS Comput. Biol.* **10**, e1003837 (2014).
35 Schellenberger, J. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* **6**, 1290-1307 (2011).
36 Agren, R. *et al.* The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Comput. Biol.* **9**, e1002980 (2013).
37 Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **7**, 74 (2013).
38 Becker, S. A. & Palsson, B. O. Context-specific metabolic networks are consistent with experiments. *PLoS Comput. Biol.* **4**, e1000082 (2008).
39 Shlomi, T., Cabili, M. N., Herrgard, M. J., Palsson, B. O. & Ruppin, E. Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.* **26**, 1003-1010 (2008).
40 Schmidt, B. J. *et al.* GIM^{3}E: condition-specific models of cellular metabolism developed from metabolomics and expression data. *Bioinformatics* **29**, 2900-2908 (2013).
41 Vlassis, N., Pacheco, M. P. & Sauter, T. Fast reconstruction of compact context-specific metabolic network models. *PLoS Comput. Biol.* **10**, e1003424 (2014).
42 Yizhak, K. *et al.* Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer. *eLife* **3** (2014).
43 Machado, D. & Herrgard, M. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput. Biol.* **10**, e1003580 (2014).
44 Robaina Estevez, S. & Nikoloski, Z. Generalized framework for context-specific metabolic model extraction methods. *Front. Plant Sci.* **5**, 491 (2014).
45 Kim, M. K. & Lun, D. S. Methods for integration of transcriptomic data in genome-scale metabolic models. *Comput. Struct. Biotechnol. J.* **11**, 59-65 (2014).

46 Saha, R., Chowdhury, A. & Maranas, C. D. Recent advances in the reconstruction of metabolic models and integration of omics data. *Curr. Opin. Biotechnol.* **29**, 39-45 (2014).

47 Hyduke, D. R., Lewis, N. E. & Palsson, B. O. Analysis of omics data with genome-scale models of metabolism. *Mol. Biosyst.* **9**, 167-174 (2013).

48 Orth, J. D. & Palsson, B. O. Systematizing the generation of missing metabolic knowledge. *Biotechnol. Bioeng.* **107**, 403-412 (2010).

49 Satish Kumar, V., Dasika, M. S. & Maranas, C. D. Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics* **8**, 212 (2007).

50 Thiele, I., Vlassis, N. & Fleming, R. M. fastGapFill: efficient gap filling in metabolic networks. *Bioinformatics* **30**, 2529-2531 (2014).

51 Fleming, R. M. & Thiele, I. von Bertalanffy 1.0: a COBRA toolbox extension to thermodynamically constrain metabolic models. *Bioinformatics* **27**, 142-143 (2011).

52 Haraldsdottir, H. S., Thiele, I. & Fleming, R. M. Quantitative assignment of reaction directionality in a multicompartmental human metabolic reconstruction. *Biophys. J.* **102**, 1703-1711 (2012).

53 Noor, E., Haraldsdottir, H. S., Milo, R. & Fleming, R. M. Consistent estimation of Gibbs energy using component contributions. *PLoS Comput. Biol.* **9**, e1003098 (2013).

54 Ganter, M., Kaltenbach, H. M. & Stelling, J. Predicting network functions with nested patterns. *Nat. Commun.* **5**, 3006 (2014).

55 Horton, P. *et al.* WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* **35**, W585-587 (2007).

56 Yu, C. S., Chen, Y. C., Lu, C. H. & Hwang, J. K. Prediction of protein subcellular localization. *Proteins* **64**, 643-651 (2006).

57 Scheer, M. *et al.* BRENDA, the enzyme information system in 2011. *Nucleic Acids Res.* **39**, D670-676 (2011).

58 Binder, J. X. *et al.* COMPARTMENTS: unification and visualization of protein subcellular localization evidence. *Database (Oxford)* **2014**, bau012 (2014).

59 Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248-1250 (2010).

60 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

61 Sahoo, S., Aurich, M. K., Jonsson, J. J. & Thiele, I. Membrane transporters in a human genome-scale metabolic knowledgebase and their implications for disease. *Front. Physiol.* **5**, 91 (2014).

62 Garcia-Albornoz, M. *et al.* BioMet Toolbox 2.0: genome-wide analysis of metabolism and omics data. *Nucleic Acids Res.* **42**, W175-181 (2014).

63 Ganter, M., Bernard, T., Moretti, S., Stelling, J. & Pagni, M. MetaNetX.org: a website and repository for accessing, analysing and manipulating metabolic networks. *Bioinformatics* **29**, 815-816 (2013).

64 Forbes, S. A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* (2014).

65 Jewison, T. *et al.* SMPDB 2.0: big improvements to the Small Molecule Pathway Database. *Nucleic Acids Res.* **42**, D478-484 (2014).

66 Bernard, T. *et al.* Reconciliation of metabolites and biochemical reactions for metabolic networks. *Brief. Bioinform.* **15**, 123-135 (2014).

67 Buchel, F. *et al.* Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst. Biol.* **7**, 116 (2013).

68 Rodriguez, J. M. *et al.* APPRIS: annotation of principal and alternative splice isoforms. *Nucleic Acids Res.* **41**, D110-117 (2013).
69 Schuetz, R., Kuepfer, L. & Sauer, U. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Mol. Syst. Biol.* **3**, 119 (2007).
70 Holzhutter, H. G. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *European journal of biochemistry / FEBS* **271**, 2905-2922 (2004).
71 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 Comput. Biol.* **7**, e1002018 (2011).
72 Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207-210 (2002).

Figures

Fig. 1 Schematic overview of the procedure for building functional, context-specific human metabolic models. See Table 1 for the details.

Fig. 2 Current status of reported generic and context-specific human metabolic models. Model statistics linked to human organs correspond to high-quality context-specific human metabolic models. For these models, left, middle and right numbers indicate the number of genes, reactions and metabolites, respectively. Numbers with an asterisk were obtained from direct processing of the respective SBML files because the model statistics were not obvious in the original literature.

Fig. 3 Model statistics of Recon series and hepatocyte metabolic models. A greater number of metabolic tasks are considered in the upgraded versions, indicating the improved model quality.

Fig. 4 Various components of human cells that heavily influence the quality of context-specific human metabolic models. Membrane and intercompartmental transporter reactions are representative components that largely depend on manual curation rather than computational analysis. Other cellular components including metabolic reactions, reaction directionality, subcellular localization and incorporation of splice isoform-level information for the GPR associations can be computationally refined.

Tables

Table 1 Databases and tools useful for building and refining genome-scale human metabolic models.

Category	Name	Description			Refs / URL
Transcriptome	The Cancer Genome Atlas (TCGA)	Repository for comprehensive cancer genomics			http://cancergenome.nih.gov/
	GEO database	Repository for high-throughput gene expression data			72
	Illumina Human Body Map 2.0	A project that provided RNA-Seq data for 16 human tissue types			
Proteome	Human Protein Atlas (HPA)	Repository for human protein expression data from immunostaining experiments			59,60
	ProteomicsDB	Repository for proteomic data obtained using mass-spectrometry			https://www.proteomicsdb.org
Omics integration methods	**Algorithms**	**Optimization problem**	**Platform**	**Objective function**	
	iMAT	MILP	Matlab (COBRA)	Not required	39
	GIMME	LP	Matlab (COBRA)	Required	38
	GIM^{3}E	MILP	Python (COBRApy extension)	Required	40
	INIT / tINIT	MILP	Matlab (RAVEN)	Not required	14, 33
	PRIME	LP	Matlab	Required	42
	MBA	MILP	-	Not required	21
	mCADRE	MILP	Matlab	Not required	30
	FASTCORE	LP	Matlab	Not required	41
Platforms	COBRA Toolbox	Run in the MATLAB environment			35
	COBRApy	Run in the Python environment			37
	RAVEN Toolbox	Run in the MATLAB environment			36

	Tools	Logics	Platforms	
Model refinement methods and tools	GapFind	Finds all gaps in a metabolic model by searching for blocked metabolites	COBRA Toolbox	49
	GapFill	Minimizes the number of gaps with minimal modifications	COBRA Toolbox	49
	fastGapFill	Fills the gaps for compartmentalized and large-scale metabolic models	COBRA Toolbox extension	50
	von Bertalanffy	Assigns reaction directionality in multi-compartment metabolic models by considering experimental and estimated thermodynamic data, metabolite structures, temperature, pH, electrical potential, ionic strength and metabolomic data	COBRA Toolbox extension	52, 53
	WoLF PSORT and CELLO	Predicts protein localization sites based on amino acid sequence	RAVEN Toolbox	55, 56
	BRENDA	Provides literature-based subcellular localization information		57
	COMPARTMENTS	Subcellular localization database with different levels including prediction, literature, and experimental data		58
	Human Protein Atlas (HPA)	Provides spatial expression patterns on the subcellular level as well as protein expressions		59,60

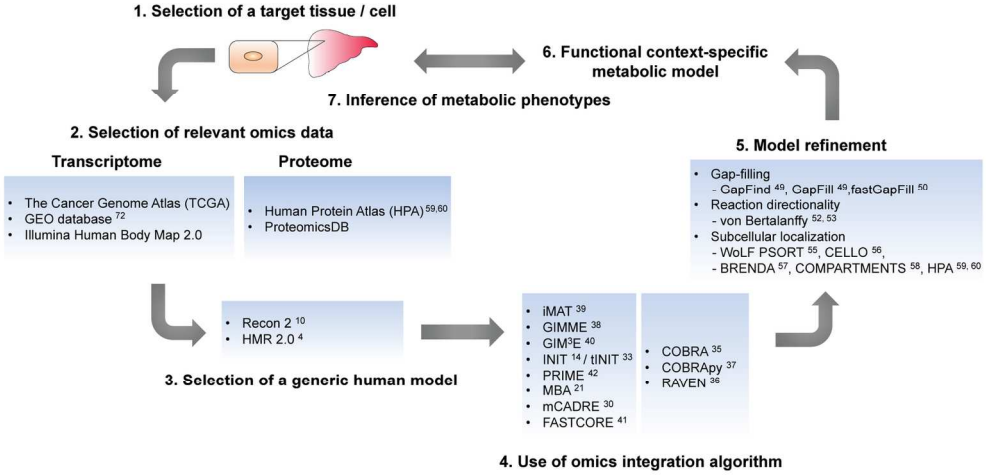


Fig. 1 Schematic overview of the procedure for building functional, context-specific human metabolic models. See Table 1 for the details.
399x191mm (150 x 150 DPI)

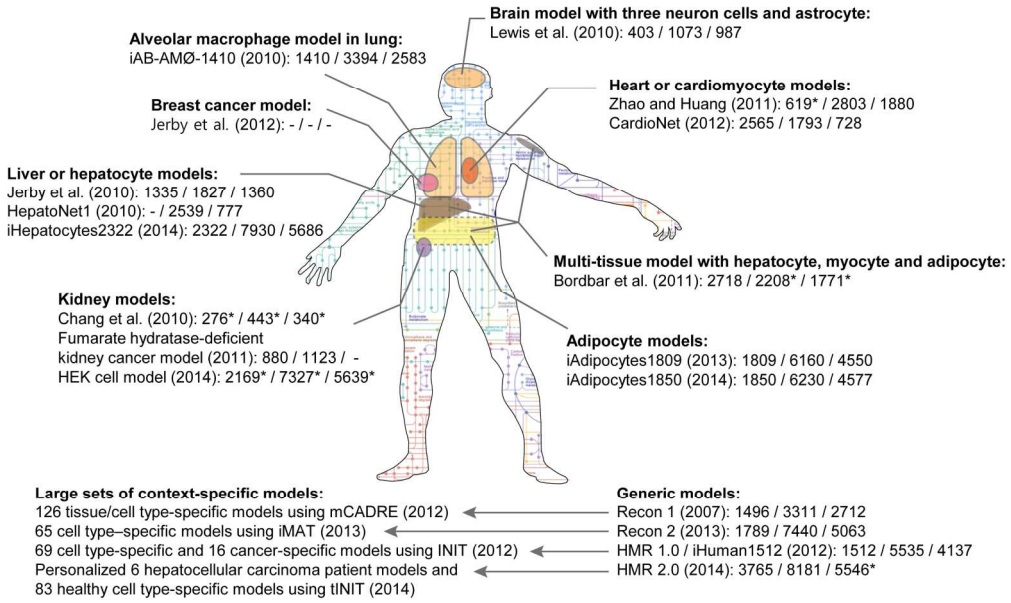


Fig. 2 Current status of reported generic and context-specific human metabolic models. Model statistics linked to human organs correspond to high-quality context-specific human metabolic models. For these models, left, middle and right numbers indicate the number of genes, reactions and metabolites, respectively. Numbers with an asterisk were obtained from direct processing of the respective SBML files because the model statistics were not obvious in the original literature.
399x245mm (150 x 150 DPI)

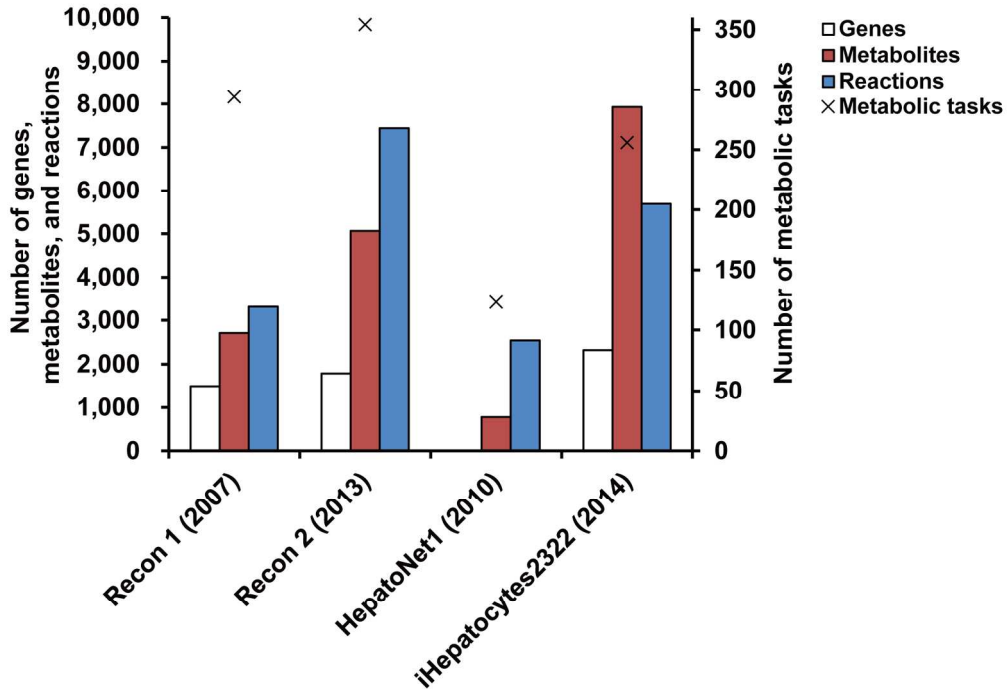


Fig. 3 Model statistics of Recon series and hepatocyte metabolic models. A greater number of metabolic tasks are considered in the upgraded versions, indicating the improved model quality.
335x239mm (150 x 150 DPI)

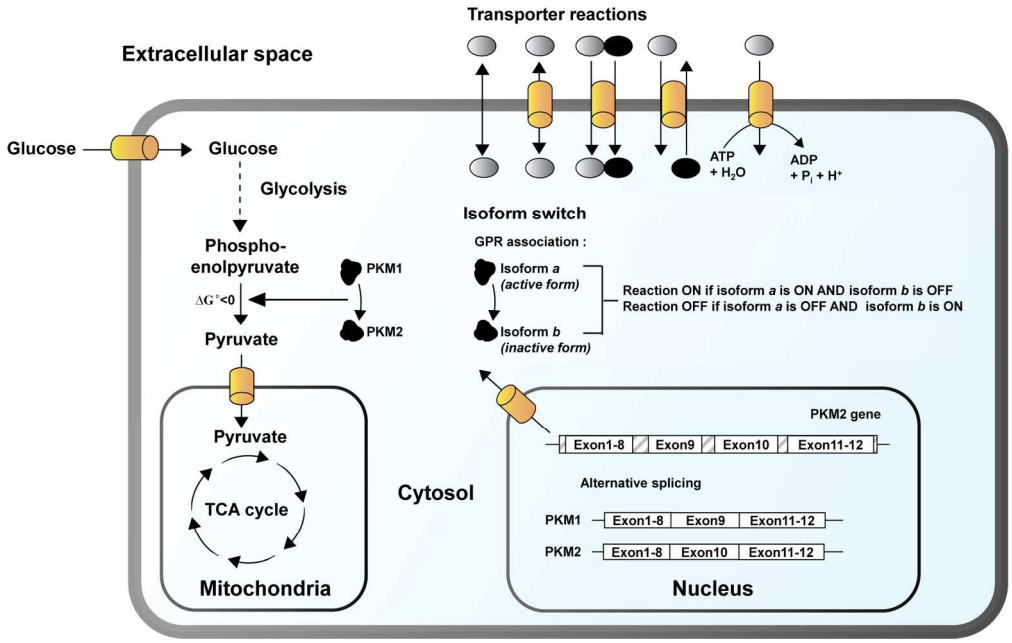
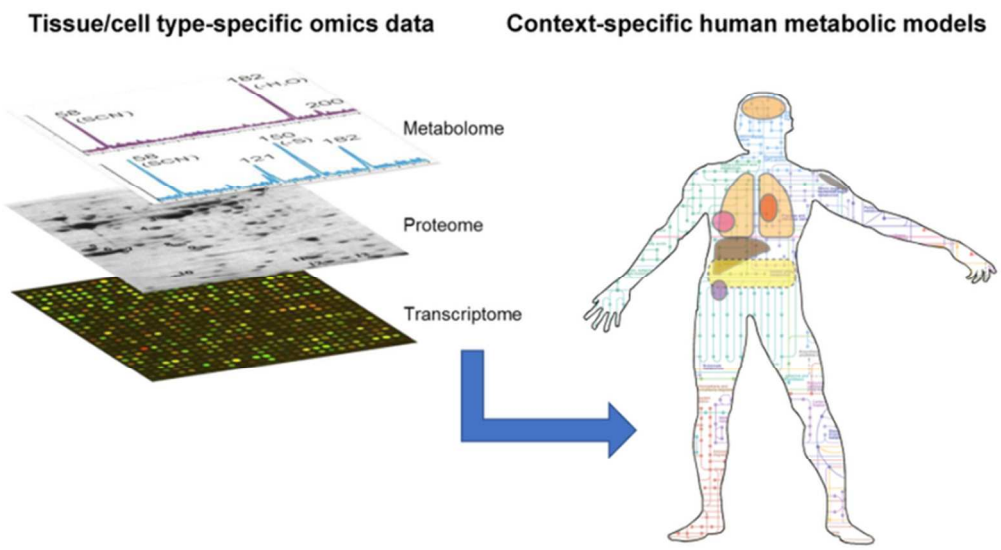


Fig. 4 Various components of human cells that heavily influence the quality of context-specific human metabolic models. Membrane and intercompartmental transporter reactions are representative components that largely depend on manual curation rather than computational analysis. Other cellular components including metabolic reactions, reaction directionality, subcellular localization and incorporation of splice isoform-level information for the GPR associations can be computationally refined.
165x104mm (300 x 300 DPI)



Genome-scale metabolic models of human cells and tissues can be reconstructed using omics data for systematic and personalized medicine
160x87mm (150 x 150 DPI)