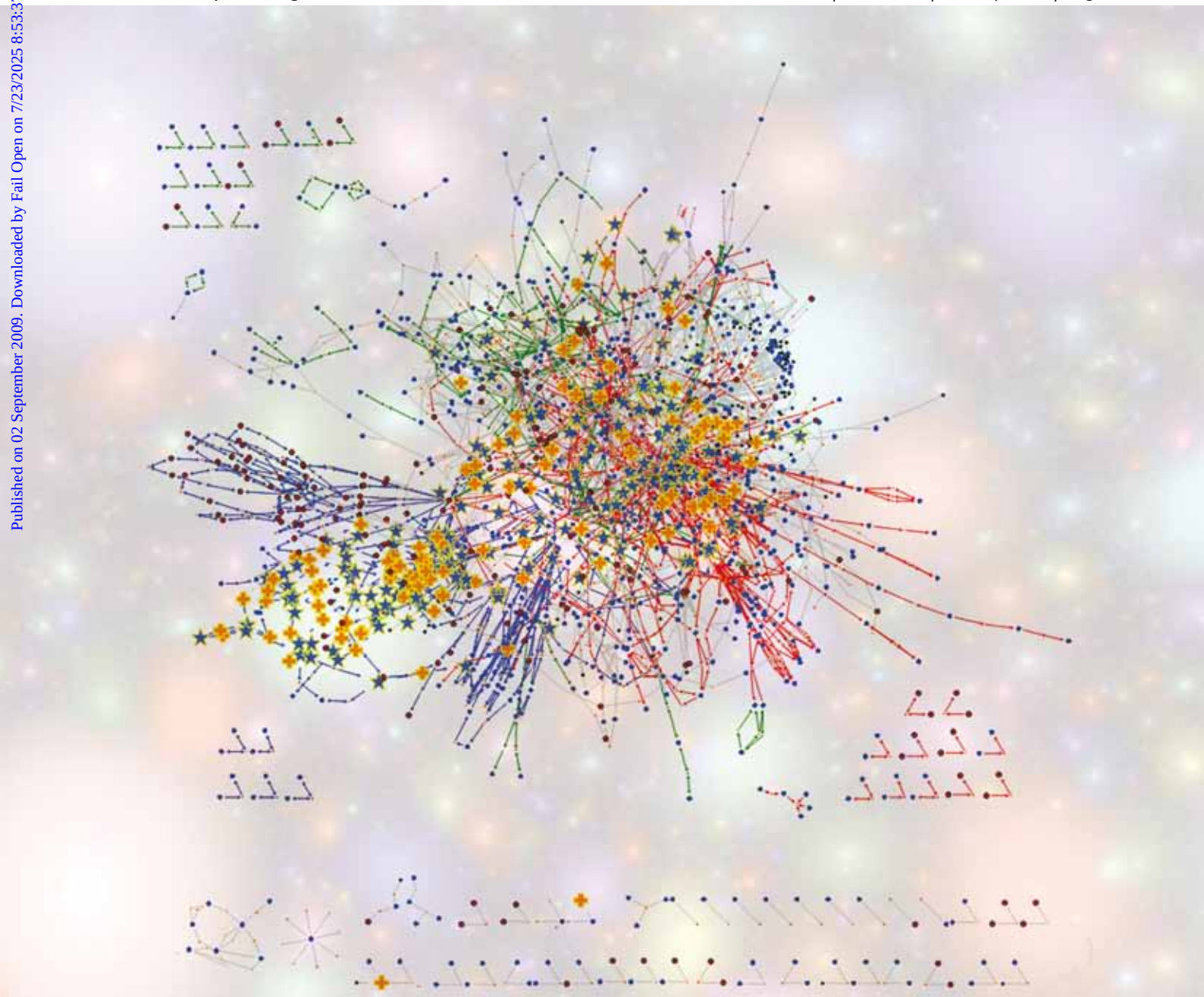
# Molecular BioSystems

www.molecularbiosystems.org

Volume 6 | Number 1 | January 2010 | Pages 1–276





ISSN 1742-206X

RSC Publishing

**PAPER**

Dong-Yup Lee *et al.*

Genome-scale modeling and in silico analysis of mouse cell metabolic network

# Genome-scale modeling and *in silico* analysis of mouse cell metabolic network†

**Suresh Selvarasu,[ab] Iftekhar A. Karimi,[a] Ghi-Hoon Ghim[c] and Dong-Yup Lee*[ab]**

***Received 1st July 2009, Accepted 10th August 2009***
***First published as an Advance Article on the web 2nd September 2009***
**DOI: 10.1039/b912865d**



Genome-scale metabolic modeling has been successfully applied to a multitude of microbial systems, thus improving our understanding of their cellular metabolisms. Nevertheless, only a handful of works have been done for describing mammalian cells, particularly mouse, which is one of the important model organisms, providing various opportunities for both biomedical research and biotechnological applications. Presented herein is a genome-scale mouse metabolic model that was systematically reconstructed by improving and expanding the previous generic model based on integrated biochemical and genomic data of *Mus musculus*. The key features of the updated model include additional information on gene–protein-reaction association, and improved network connectivity through lipid, amino acid, carbohydrate and nucleotide biosynthetic pathways. After examining the model predictability both quantitatively and qualitatively using constraints-based flux analysis, the structural and functional characteristics of the mouse metabolism were investigated by evaluating network statistics/centrality, gene/metabolite essentiality and their correlation. The results revealed that overall mouse metabolic network is topologically dominated by highly connected and bridging metabolites, and functionally by lipid metabolism that most of essential genes and metabolites are from. The current *in silico* mouse model can be exploited for understanding and characterizing the cellular physiology, identifying potential cell engineering targets for the enhanced production of recombinant proteins and developing diseased state models for drug targeting.

## Introduction

Recently, genome-driven *in silico* modeling and analysis have been recognized as the promising approach for characterizing complex cellular systems and for deriving novel strategies in biomedical and biotechnological applications in the context of systems biology and biotechnology.[1–5] Such a genome-scale modeling approach and concomitant analysis are initiated by reconstructing the metabolic network of a given organism, describing its genotype–phenotype relationship on the basis of a primary set of biological information, *i.e.*, genome annotation, biochemical and cell physiological data.[6,7] The reconstructed model is then mathematically expressed by incorporating metabolite balance and reaction reversibility constraints within the model formulation.[8] As such, the phenotypic behavior and metabolic states of the organism can be assessed and predicted by resorting to various optimization techniques such as linear programming,[9] quadratic programming,[10] and mixed integer linear programming.[11]

Among a multitude of model organisms demonstrating the predictive power of the *in silico* models, *Escherichia coli* is one of the well established bacterial systems: significant improvements and updates in the genome-scale metabolic model of *E. coli* have been consistently made for the past two decades.[12] Moreover, this model has been effectively exploited for characterizing the internal metabolism,[4,13–15] examining the cellular behavior under various genetic manipulations and environmental conditions,[16] and enabling hypothesis generation.[17,18] *Saccharomyces cerevisiae* is another well-developed model organism; its genome-scale model can be potentially employed to characterize yeast metabolism, understand complex human diseases and to identify targets for strain improvement.[19–22] Currently, such genome-scale models are being reconstructed and made available for more than 30 microbes as reviewed elsewhere.[23,24] Nevertheless, only a handful of works have been done for reconstructing genome-scale metabolic models of mammalian cells. This is mainly attributable to the complexity involved in describing their metabolisms as well as lack of sufficient information.[25] Sheikh *et al.*[26] first developed a generic model of mouse, based on the annotated genome data of *Mus musculus* and described the industrially important hybridoma cells producing monoclonal antibody and recombinant proteins. Selvarasu *et al.*[27] further improved the mouse model and applied it to characterize the metabolism in mouse hybridoma cells during fed-batch culture. Recently a

[a] *Department of Chemical and Biomolecular Engineering, National University of Singapore, Engineering Drive 4, Singapore 117576. E-mail: cheld@nus.edu.sg; Fax: +65 6779 1936; Tel: +65 6478 8900*
[b] *Bioprocessing Technology Institute, Agency for Science, Technology and Research (A*STAR), 20 Biopolis Way, #06-01, Centros, Singapore 138668*
[c] *Network Research Institute, Cyram Co. Ltd., SNU Research Park Innovation Center, Bongchun 7-dong, Gwanak-gu, Seoul 151-818, South Korea*

† Electronic supplementary information (ESI) available: Details on reactions, metabolites, dead ends and biomass compositions. See DOI: 10.1039/b912865d

genome-scale human model was also reconstructed using genomic and bibiliomic data[28] and combined with the tissue specific gene expression data to elucidate the metabolic states of different human tissues.[29] Wahl *et al.*[30] derived a lumped metabolic model for vaccine-producing mammalian cell line, Madin-Darby Canine Kidney (MDCK) cells from annotated sequences of *Canis familiaris* and explained the characteristic behavior during different growth phases. Hence, these initial and recent developments clearly indicate that there has been an increasing interest and demand towards genome-scale modeling of mammalian cells.

Of fully sequenced mammalian organisms at present, mouse cells have been well studied for understanding mammalian genetic functions. They possess high degree of homology with human genome,[31] playing key roles in both biomedical research and therapeutical applications.[32–34] However, currently available genome-scale mouse model[26] is incomplete in describing the cellular metabolism: some of the key metabolic interactions are still missing, which may lead to inaccurate predictions. Thus, in this work, we updated the previous model by revising and improving the network connectivity, thus enhancing its predictive capability. Structural and functional features of the model were also examined using various computational techniques, which allows for better understanding of model characteristics.

## Results and discussion

### Genome-scale reconstruction of mouse metabolic network

The genome-scale metabolic network of mouse was systematically reconstructed based on the previous model and relevant information from various resources (Fig. 1 and Table 1; see Methods for details). Compared to the previous model, 490 reactions are newly added, providing updated information on gene–protein-reaction (GPR) association and detailed description on lipid, amino acids, carbohydrate and nucleotide metabolisms (Fig. 2). The model is comprised of 724 genes, 715 enzymes, 1162 internal metabolites, and 1494 reactions (Table 2); 1246 reactions are biochemical conversions within cytosol (1085) and mitochondria (161), and 248 are exchange reactions describing the metabolite transport between intra- and extra-cellular membrane (171) and cytosol and mitochondria (77). We also compared the current mouse model with other model organisms such as *E. coli* and *S. cerevisiae* (supplementary 1†). In addition to biochemical reactions, we derived one balance equation for expressing the cell biomass from the drain of biosynthetic precursors such as proteins, lipids, carbohydrates, DNA, RNA, and other cellular components at their experimental composition and relevant energy cofactors for their conversion and assembly (supplementary 2†). The full list of reactions, metabolites and their abbreviations are given in supplementary 3.†

During the reconstruction process, manual curation of the resulting network was iteratively performed by checking the consistency, accuracy, and completeness of the model until simulated results were consistent with experimental observation both quantitatively and qualitatively. It allows us to find knowledge gaps for refinining the model. For example, to fill identified metabolic gap in tryptophan metabolism, we included an enzymatic reaction ACMSC (EC:4.1.1.45) which could be obtained through literature mining.[31,35] Similarly we added a transport reaction of spermine which was reported as an important precursor for spermidine synthesis,[36,37] thus resulting in the enhanced network connectivity.

### *In silico* model validation

The predictive capability of the current mouse model was tested using constraints-based flux analysis, based on batch cultural data of mouse hybridoma cells producing anti-F monoclonal antibody, grown in a DMEM media supplemented with proline, asparagine and aspartate.[38] The biomass production was maximized to simulate the cell growth condition, constraining the measured specific consumption/production rates of nutrients/products during the culture (Fig. 3). The resultant growth rate (0.048 $h^{-1}$) was higher than the average specific growth rate (0.0362 $h^{-1}$) in the entire batch culture. We believe that the growth prediction can be improved when relevant measurements for *in silico* simulation are used to reflect more realistic operational condition during exponential growth phase.

For qualitative model prediction, we conducted *in silico* analysis on minimal media requirements for the cell growth and finally identified required medium components. They include essential amino acids, folate and phosphate which are almost consistent with experimentally observed essential components[39] and the nutrition requirements for laboratory animals[40] (see supplementary 4†). However, *in silico* analysis could not identify some minimal medium components such as growth factors, cofactors, and minerals (biotin, thiamine, vitamins, calcium and magnesium ions, *etc.*). Not surprisingly, the predicted growth of the mouse cell was not directly affected only by glucose uptake. Instead, it was determined by the uptake of essential amino acids, thus confirming previous observation that under glucose-deprived or limited conditions, unlike microbial cells mammalian system can survive by utilizing other nutrients like essential amino acids.[41]

Gene essentiality analysis also allowed us to validate and improve the current mouse model in an iterative way. All predicted essential genes using current and previous models were compared with experimentally reported essential genes from KOMP (KnockOut Mouse Project) database. As summarized in Table 3, most *in silico* essential genes are experimentally confirmed while we also found some false positive predictions for genes *dnmT1*, *ada*, *afmId*, *alaS2*, *gyk* and *acsL4*. This implies the presence of knowledge gaps (dead end metabolites) around those genes within the current network, suggesting further investigation for model improvement. For example, the model incorrectly identified the gene *acsL4* as essential. In mouse, *acsL4* codes for acyl-coA synthetase long-chain family member 4 (EC:6.2.1.3), which is involved in fatty acid metabolism. Cho *et al.*[42] also reported that *acsL4* was non-essential in mouse as the deficiency of *acsL4* was compensated by another gene, *acsL3* (acyl-coA synthetase long-chain family member 3), thus leading to a normal phenotype. Such information can be newly included in the

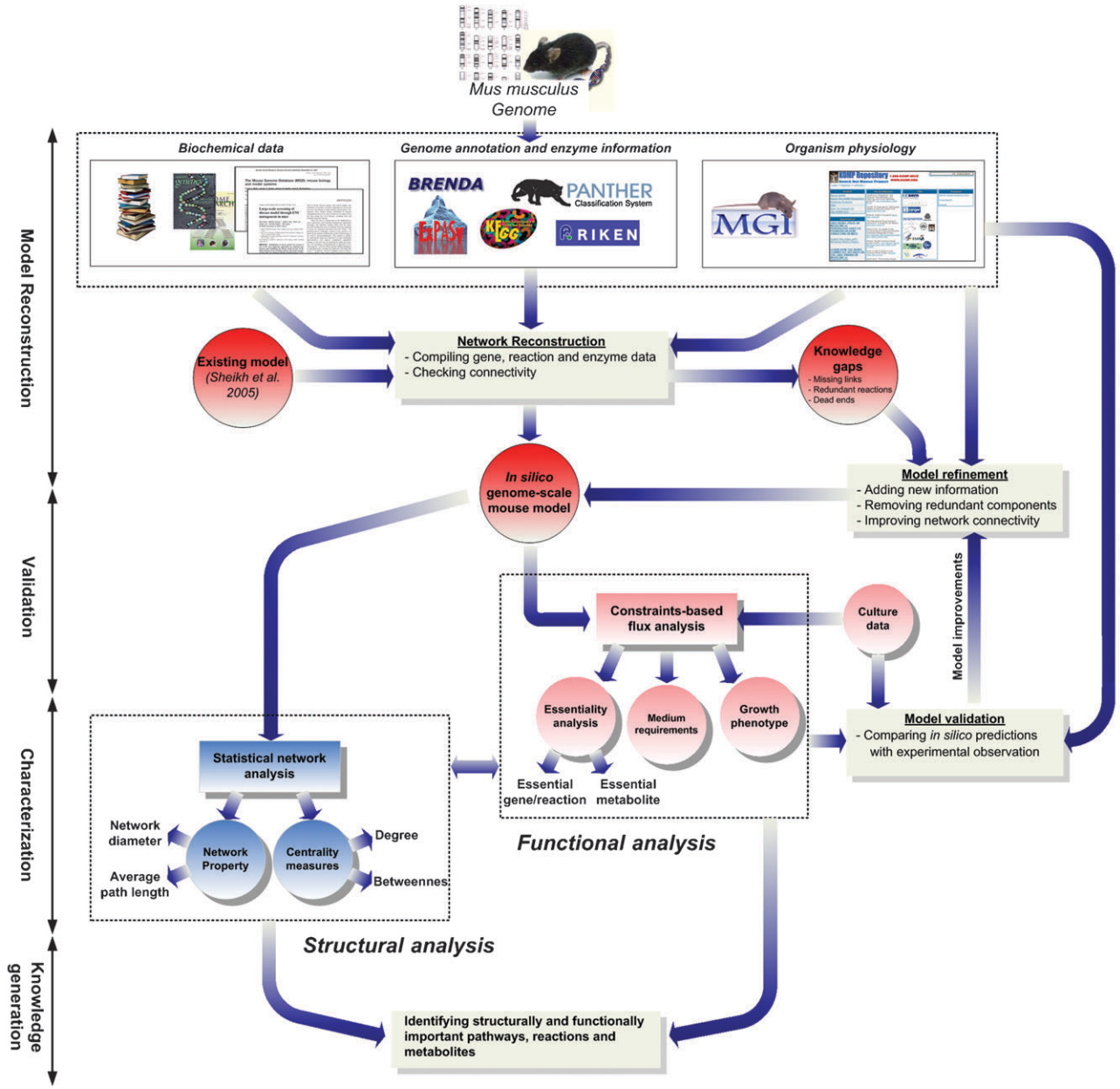


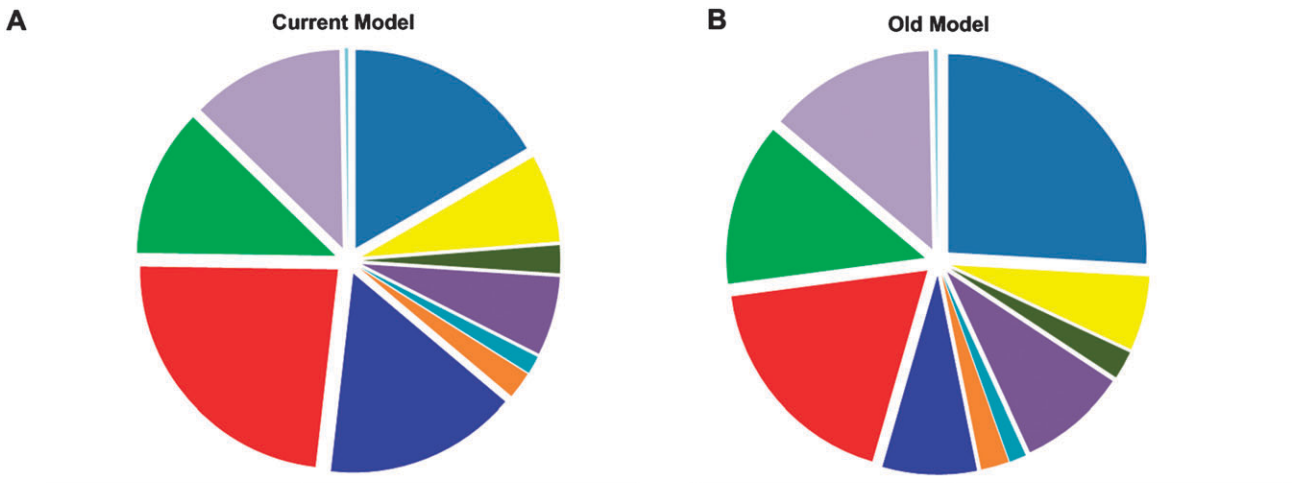

**Fig. 1** Schematic representation of the iterative approach employed in the reconstruction and analysis of genome-scale mouse model. The existing model was used as a template and the network was expanded by compiling the information (genome, biochemical and mouse physiological data). Missing links and redundant reactions were then identified to refine the model with such available resources. The resultant expanded model underwent the validation process using constraints-based flux analysis with cell culture and *in vivo* gene essentiality data for verifying the prediction. The presence of knowledge gaps was explored and again the model could be improved interactively. Subsequently, the model was analyzed both structurally and functionally to characterize mouse metabolism and identify key pathways, reactions and metabolites.

model to improve its predictions. Note that a list of dead end metabolites are given in supplementary 3.†

### Structural and functional characterization of mouse metabolism

The characteristic features of the reconstructed model were explored from its structural and functional points of view (Fig. 1). First, the statistical network analysis identified a large cluster of weakly connected reactions (89% of total reactions) and 119 small clusters with 1 to 17 connecting reactions (Fig. 4). We then calculated the network diameter while the cofactor metabolites (*e.g.*, ATP, $H_2O$, $CO_2$, *etc.*) were excluded to prevent biologically meaningless results of identifying them as major hubs in the network.[43,44] The resulting network diameter for the large cluster was measured to be 40. The average path length (APL) was also calculated as 8.51, revealing that most of the metabolites in the network can be converted between each other by approximately 3∼4 reactions. Similar analysis was conducted for three major sub-networks, which were significantly improved from the previous model (Fig. 2), carbohydrate, amino acids and lipid metabolisms, resulting in different network diameters and APLs as illustrated in Fig. 4.




| Pathways | All reactions | | Gene associations | |
|---|---|---|---|---|
| | No of rxns | % of total | No of rxns | % of total |
| Amino acids metabolism | 350 | 23.43 | 338 | 23.43 |
| Carbohydrates metabolism | 180 | 12.05 | 176 | 12.05 |
| Energy metabolism | 33 | 2.21 | 25 | 2.21 |
| Lipid metabolism | 234 | 15.66 | 203 | 15.66 |
| Membrane transport | 248 | 16.60 | 30 | 16.60 |
| Metabolism of cofactors and vitamins | 107 | 7.16 | 98 | 7.16 |
| Metabolism of complex carbohydrates | 34 | 2.28 | 34 | 2.28 |
| Metabolism of complex lipids | 96 | 6.43 | 93 | 6.43 |
| Metabolism of other amino acids | 22 | 1.47 | 22 | 1.47 |
| Nucleotide metabolism | 186 | 12.45 | 184 | 12.45 |
| Others | 4 | 0.27 | 0 | 0 |
| Total | 1494 | 100 | 1203 | 80.52 |

| Pathways | All reactions | | Gene associations | |
|---|---|---|---|---|
| | No of rxns | % of total | No of rxns | % of total |
| Amino acids metabolism | 185 | 18.43 | 165 | 16.43 |
| Carbohydrates metabolism | 133 | 13.25 | 109 | 10.86 |
| Energy metabolism | 23 | 2.29 | 13 | 1.29 |
| Lipid metabolism | 77 | 7.67 | 72 | 7.17 |
| Membrane transport | 260 | 25.90 | 30 | 2.99 |
| Metabolism of cofactors and vitamins | 61 | 6.08 | 46 | 4.58 |
| Metabolism of complex carbohydrates | 23 | 2.29 | 22 | 2.19 |
| Metabolism of complex lipids | 89 | 8.86 | 79 | 7.87 |
| Metabolism of other amino acids | 14 | 1.39 | 14 | 1.39 |
| Nucleotide metabolism | 136 | 13.55 | 130 | 12.95 |
| Others | 3 | 0.30 | 0 | 0 |
| Total | 1004 | 100 | 680 | 67.72 |

**Fig. 2** Functional classification of metabolic reactions in mouse genome-scale model, (A) current updated model and (B) old model. Number of reactions in each subsystem is shown in the tables. Metabolic subsystems with number of gene and non-gene associated reactions are detailed in the table.

**Table 1** Online resources for reconstructing genome-scale mouse metabolic network

| Online resources | URL |
|---|---|
| Mouse genome informatics | http://www.informatics.jax.org |
| NCBI mouse genome resource | http://www.ncbi.nih.gov/genome/guide/mouse |
| PANTHER | http://www.pantherdb.org |
| KEGG database | http://www.genome.jp/kegg/pathway.html |
| ExPaSy | http://au.expasy.org |
| Brenda | http://www.brenda.uni-koeln.de |
| RIKEN | http://fantom2.gsc.riken.go.jp/metabolome/ |
| KOMP (Knock Out Mouse Project) database | http://www.komp.org |

We also explored the network topology by calculating degree and betweenness centrality of metabolites, thus identifying highly connected and critical (bridge-acting) components within the network. The detailed results are given in supplementary 3.† We further investigated the topological properties of the network by comparing the essential metabolites with their centrality scores (Fig. 5). The essential metabolites for the cell growth were obtained using flux sum approach as described in Methods. It was observed that the average centrality scores of essential metabolites (degree: 6.37 and betweenness centrality: 0.00198) were much higher than the non-essential ones (degree 2.55 and betweenness centrality: 0.00039). Interestingly, L-glutamate (GLU) was identified as one of highly linked as well as key bridging essential metabolites with the highest degree and betweenness centrality scores (Fig. 5A). It was reported that this metabolite plays versatile roles in trans-amination, deamination, and neurotransmission in the cellular metabolism.[45,46] Similarly, acetyl-coA (ACCOA) participates in many reactions, connecting glycolysis, lipid metabolism, and TCA cycle as visualized in Fig. 6. Topologically, the excess carbohydrates and fatty acid may lead to the production of ACCOA toward fatty acid synthesis and citrate acid cycle for respiration, respectively, to maintain energy requirements within the cell. Thus, it is suggested to consider those bridging and highly connected metabolites as cell engineering targets for enhancing the cell viability.

Unexpectedly, metabolite centrality was not clearly correlated with metabolite essentiality (Fig. 5). For example, our metabolite essentiality analysis identified AKG as non-essential inspite of its high degree and betweenness centrality values. In order to explore the phenotypic effect of AKG deletion in mouse metabolism, we conducted *in silico* analysis by removing the AKG metabolite from the network and compared both resulting flux distributions and metabolite flux sum with the reference

**Table 2** Characteristics of the mouse genome-scale metabolic network and its comparison with the previous generic model

| | Current model | Previous model[26] |
|---|---|---|
| Genome size (base pairs, bp) | $2.5 \times 10^9$ | |
| Genes | 724 | 473 |
| Enzymes | 715 | — |
| Metabolites | 1285 | 915 |
| Cytosol | 930 | 619 |
| Mitochondrial | 232 | 173 |
| Extracellular | 123 | 123 |
| Reactions | 1494 | 1004 |
| Cytosol | 1085 | 618 |
| Mitochondrial | 161 | 122 |
| Transport | 248 | 260 |
| Membrane[a] | 171 | 185 |
| Mitochondrial[b] | 77 | 75 |
| Reactions with ORF assignment | 1203 | 680 |
| Inferred reactions | 291 | 324 |
| Knowledge gaps[c] | 367 | 335 |

[a] Transport reactions between extracellular and cytosol. [b] Transport between cytosol and mitochondria. [c] Knowledge gap refers to the dead-end metabolites in the network.

that was determined under the normal growth condition (Fig. 7A and B). Surprisingly, the effect of AKG deletion on the cell growth was negligible: it was observed that required biomass precursors were synthesized *via* other alternate pathways such as pentose phosphate pathway and nucleotide metabolism, thus leading to unchanged phenotypic state. Similarly, PYR, one of the highly connected metabolites in the central metabolic pathway was also rendered non-essential even though it has high degree and betweenness centrality scores. It should be noted that in mouse metabolic network, PYR is available in both cytosol and mitochondria compartments. Thus, no phenotypic change was observed upon its removal in the cytosol due to the compensational role by its counterpart from the other compartment.

We identified a set of essential genes for the cell growth in a defined medium. Initially, single-gene reaction association was assumed to perform gene deletion analysis under rich medium (RM) as well as minimal medium (MM) conditions. Of 109 essential reactions under RM condition, 93 were gene-associated, 6 non-gene-associated, and 10 for the transport of amino acids. Interestingly, the highest percentage (59%) of essential reactions is from lipid metabolism (fatty acid biosynthesis and fatty acid metabolism), indicating that it may be one of the most vulnerable sub-systems to environmental disturbances. The additional 6 reactions under MM condition are from amino acids (5) and carbohydrate (1) metabolism (details in supplementary 3†). When GPR associations were considered, only 72 essential genes were identified as there were many isozymes and multifunctional proteins in the current genome-scale model. For example, fatty acids synthase (*fasN*), one of the multifunctional proteins, alone catalyzed 37 reactions in lipid metabolism, while other genes or proteins are associated with at least two or more reactions in the metabolic network.

The presence of low percentage (<10%) of essential reactions implies that mouse metabolism is highly flexible and robust upon internal changes to attain the same phenotype through alternate pathways. For instance, a necessary precursor for nucleotides synthesis, 5-phospho-α-D-ribose-1-diphosphate (PRPP), can be produced by two reactions PBEF1 (EC:2.4.2.12) and RPPK (EC:2.7.6.1). When one of the reactions was deleted *in silico*, the other produced PRPP, thus rendering two reactions/genes non-essential and making the network flexible. In the view of exploring such combinatorial genes/reactions, we conducted double-knockout analysis. From more than $9.5 \times 10^6$ pairs of 1385 non-essential reactions, we could identify only 139 lethal pairs involving 114 unique reactions. Most essential pairs belong to two categories: (i) two reactions producing the same metabolite (*e.g.* both reactions, PBEF1 and RPPK, producing same metabolite PRPP), and (ii) subsequent two reactions producing and consuming same metabolite (*e.g.* reaction PC1 produces metabolite PE, which is then consumed by PSS1). Similar analysis has been successfully applied and the functional features have been elucidated for *H. pyroli*, as such demonstrating the cellular robustness and suggesting multiple deletion analysis for identifying drug targets.[47]

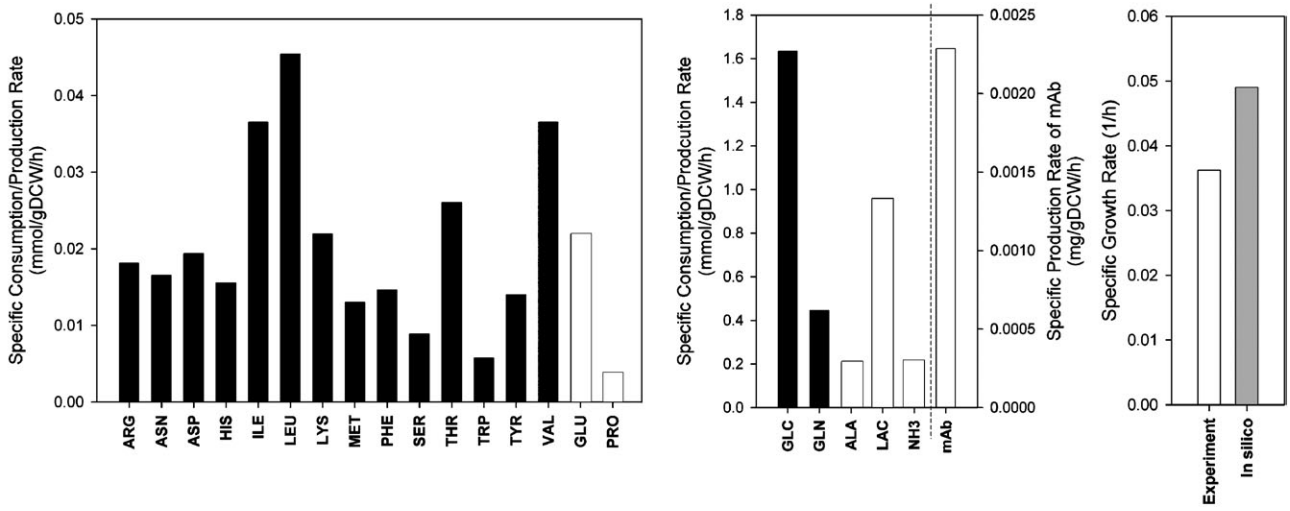



**Fig. 3** Comparison of *in silico* growth rate with experimentally observed growth rate during batch culture. Specific growth rate is in $hr^{-1}$; mAb production rate in mg $gDCW^{-1}$ $h^{-1}$. The bars with black and white colours represent specific consumption and production rates, respectively.

**Table 3** Comparison of mouse *in silico* essential genes with the previous model and *in vivo* reported genes

| Genes | Enzymes | EC No | Comparison with experiment[a] | | Reference[b] |
|---|---|---|---|---|---|
| | | | Current model | Previous model | |
| *acaCb* | Acetyl CoA carboxylase | 6.4.1.2 | + | + | Harada *et al.* 2007 |
| *arg1* | Arginase 1, Liver | 3.5.3.1 | + | – | Iyer *et al.* 2002 |
| *cds2* | Phosphatidate cytidylyltransferase 1 | 2.7.7.41 | + | + | Lexicon Genetics, 2005 |
| *fasN* | Fatty acid synthase | 2.3.1.85/2.3.1.38/2.3.1.39 2.3.1.41/ 1.1.1.100/4.2.1.61/1.3.1.10/3.1.2.14 | + | – | Chirala *et al.* 2003; Chakravarthy *et al.* 2005 |
| *gys1* | Glycogen synthase 1, muscle | 2.4.1.11 | + | – | Pederson *et al.* 2004 |
| *hk2* | Atp: D-Fructose 6-phosphotransferase | 2.7.1.1 | + | – | Heikkinen *et al.* 1999 |
| *hmgCr* | 3-Hydroxy-3-methylglutaryl-coenzyme A reductase | 1.1.1.34 | + | + | Ohashi *et al.* 2003 |
| *mthFr* | 5,10-Methylenetetrahydrofolate reductase | 1.5.1.20 | + | – | Chen *et al.* 2001 |
| *odc1* | Ornithine decarboxylase, structural | 4.1.1.17 | + | – | Pendeville *et al.* 2001 |
| *pcyT1a* | Phosphate cytidylyltransferase 1, choline, alpha isoform | 2.7.7.15 | + | + | Wang *et al.* 2005 |
| *pemT* | Phosphatidylethanolamine *N*-methyltransferase | 2.1.1.17 | + | – | Walkey *et al.* 1997 |
| *pgm2* | Phosphoglucomutase | 5.4.2.2 | + | + | Greig *et al.* 2007 |
| *pisD* | Phsophatidyl-L-serine carboxy-lyase | 4.1.1.65 | + | + | Steenbergen *et al.* 2005 |
| *srm /sSms* | Spermidine synthase | 2.5.1.16/2.5.1.22 | + | – | Strom *et al.* 1997 |
| *acsL4* | Acyl-coA synthetase long chain family member 4 | 6.2.1.3 | – | NA | Cho *et al.* 2001 |
| *ada* | Adenosine deaminase | 3.5.4.4 | – | + | Wakamiya *et al.* 1995 |
| *afmId* | Kynurenine formamidase | 3.5.1.9 | – | – | Dobrovolsky *et al.* 2005 |
| *alas2* | Glycine *C*-succinyl-transferase-(decarboxylating) | 2.3.1.27 | – | NA | Nakajima *et al.* 1999 |
| *dnmT1* | *S*-Adenosyl-L-methionine DNA (CYTS-5-)-methyltransferase | 2.1.1.37 | – | NA | Li *et al.* 1993 |
| *gyk* | Glycerol kinase | 2.7.1.30 | – | – | Huq *et al.* 1997 |

[a] '+'and '–' stand for 'consistent' and 'not consistent' with experiments, respectively. [b] Details of references are available in supplementary 3.† NA: Not available.

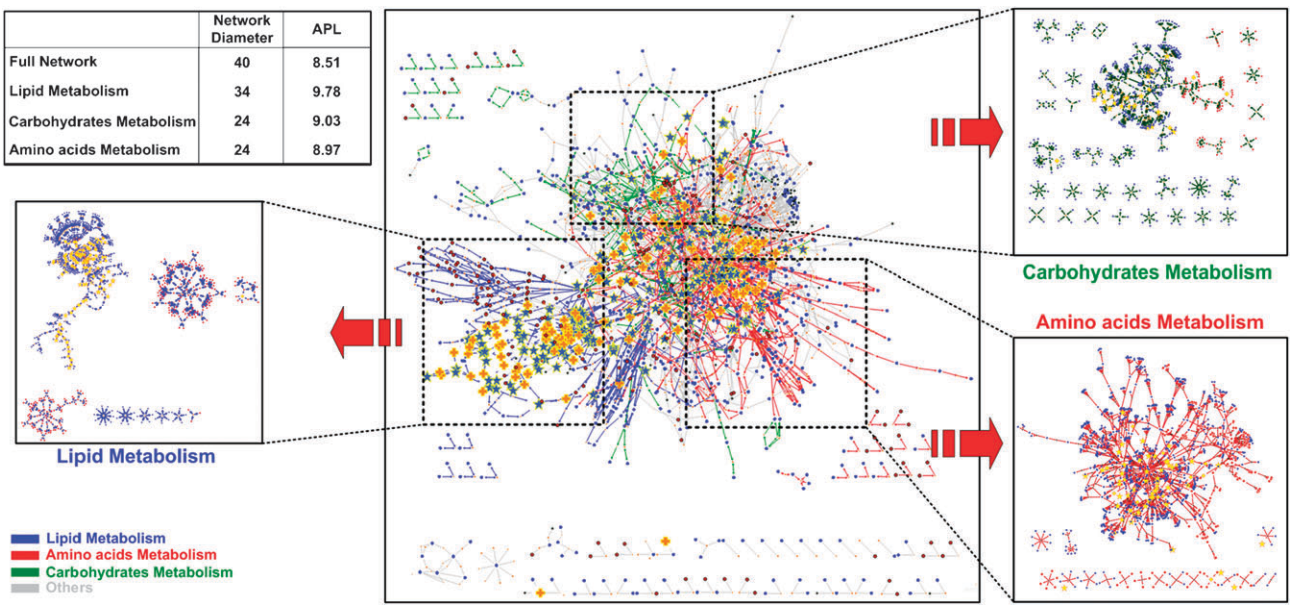



**Fig. 4** The connectivity of metabolites in different reactions in the metabolic network. The reactions involved in significantly improved metabolic subsystems such as carbohydrates, lipids and amino acids metabolisms are indicated by their edge colours: green, blue and red, respectively. Metabolites colours: blue-cytosol, red-mitochondria, green-extracellular and yellow-cofactors. Metabolites and reactions from amino acids, lipids and carbohydrates metabolism were extracted to draw individual edge generated graphs. Essential reactions and metabolites in the subnetworks are highlighted using cross and star-shaped nodes. Network diameter and average path lengths (APL) for the main network and the three sub-networks are also shown.

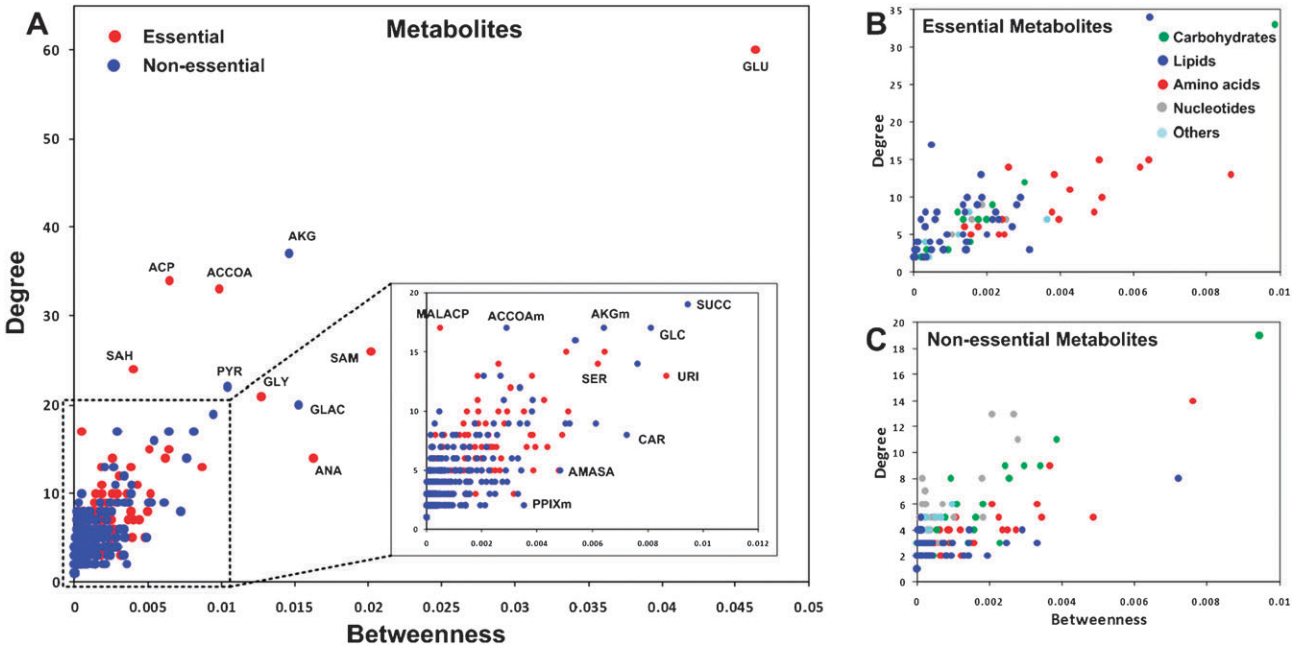



**Fig. 5** Correlation between metabolite degree and betweenness centrality for (A) all metabolites, (B) essential metabolites and (C) non-essential metabolites. The metabolite can be identified as essential when its removal leads to no growth. Highly-connected, bridging metabolites are highlighted in (A). ACP: acyl carrier protein, ACCOA: acetyl-coA, ACCOAm: acetyl-coA mitochondiral, AKG: α-ketoglutarate, AKGm: α-ketoglutarate mitochondrial, AMASA: L-2-aminoadipate-6-semialdehyde, ANA: *N*-acetylneuraminate, CAR: carnitine, GLAC: D-galactose, GLC: D-glucose, GLU: L-glutamate, GLY: glycine, MALACP: malonyl-[acyl-carrier-protein], PPIXm: Protoporphyrin mitochnondrial, PYR: pyruvate, SAH: S-adenosyl-L-homocysteine, SAM: S-adenosyl-L-methionine, SUCC: succinate, SER: L-serine and URI: uridine.

### Important role of lipid pathway in mouse metabolism

The structural and functional analysis of the genome-scale mouse model revealed that lipid metabolic pathways play a key role in supporting the cell growth. It was evident from the results of essentiality analysis: more than 50% of the essential reactions/genes were from the lipid metabolism (Fig. 8A); among 177 essential metabolites out of 1162 total, nearly 50% (86 out of 177) were directly linked to lipid metabolism (Fig. 8B). In addition, centrality measurements from statistical analysis identified some of key metabolites in lipid metabolism that structurally form a major hub in metabolic network (Fig. 5B). Thus, such highly connected lipid metabolites are functionally important as well. Although their functional role in cellular activities should be examined and verified experimentally, the emergence of 'lipidomics' allows us to gain new insight into the lipid metabolism, providing large sets of data on lipid metabolites and their functional interactions with other proteins, lipids and metabolites within the cell.[48] Hence, this new information can be potentially integrated with the metabolic network to develop a prototype model for describing the regulation of lipid metabolites under different growth conditions. Such observations can also be informative in understanding lipid metabolism related human diseases such as obesity, hypertension, diabetes, *etc.*, and for examining mutants or disease causing agents.[49]

### Links to glycosylation for model improvement

Glycosylation is the enzymatic process in mammalian systems to form complex glycans and their eventual attachment to cell proteins and lipids, thereby activating their interactions with the cell environment through metabolism, regulation and signaling.[50,51] It involves transport of dietary sugar molecules such as glucose, mannose, fucose and galactose into the cell, and their conversion along the metabolic pathways by utilizing more than 2–3% of genes and many high-energy metabolic intermediates. Such monosaccharides are generally synthesized within cytosol and assembled inside endoplasmic reticulum (ER) and golgi compartments. The complexity pertaining to the glycosylation mainly comes from the structural diversity of generated complex glycans and the involvement of more than 250 isozymic and multifunctional enzymes.[52] So far, this posed a major obstacle in investigating the glycosylation process and is evident from the limited number of works available on the modeling aspect of glycosylation.[53,54] The current genome-scale mouse model offers a good platform for integrating the intricate glycosylation process and understanding its link to the cell metabolism since it describes the synthesis of mono-saccharide and nucleotide sugars well. However, the formation of complex glycan structures in ER and golgi compartments has been overlooked in the model due to lack of relevant data and information. Moreover, it is highly necessary to determine the amount of sugars and the associated glycans required for glycoproteins and glycolipids. In this regard, the emerging field of 'glycomics' can provide more information on sugar molecules and their concentration inside the cell.[55] We can determine amount of required monosaccharides for the assembly of glycans and subsequently examine the activity of the relevant enzymes and, more importantly, their spatial localization during the carbohydrate synthesis. Therefore,

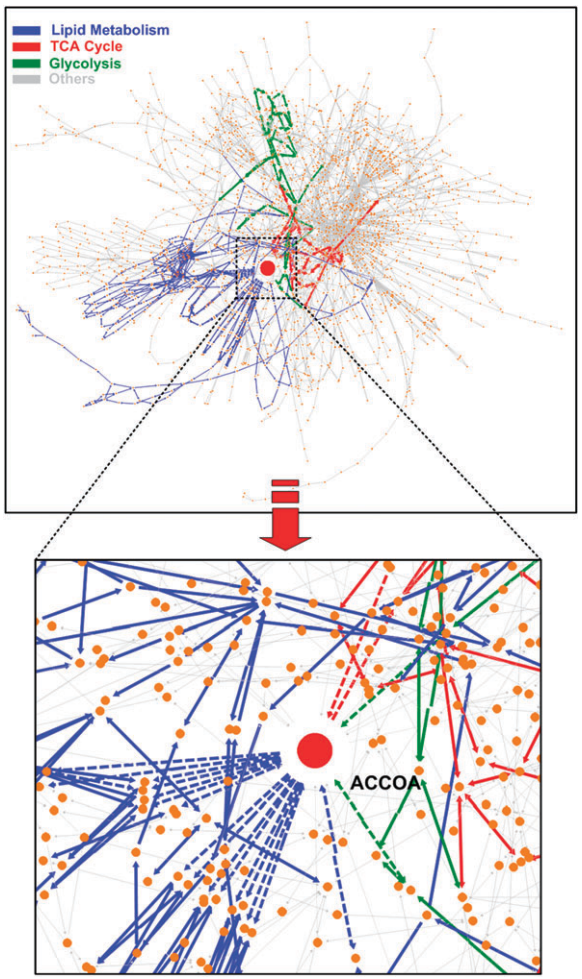



**Fig. 6** Visualization of the ACCOA interaction across lipid metabolism, TCA cycle and glycolysis. The enlarged section shows the high connectivity and bridging characteristics of ACCOA. Blue edges: lipid metabolic reactions, green: TCA cycle and red: glycolysis. ACCOA: acetyl-coA.

such data can be combined with the carbohydrate, protein and lipid compositions in the cell biomass equation, allowing us to better understand the glycosylation process along with additional sugars availability constraint in the model.

## Concluding remarks

The main objectives of the current work are to reconstruct a genome-scale mouse metabolic network by updating a previous generic model and structurally and functionally characterize the mouse metabolism. The key features of the updated model include improved network connectivity through the addition of reactions in lipid, amino acids, carbohydrates and nucleotides metabolisms and new information on GPR association. The predictability of the reconstructed model was validated both quantitatively and qualitatively using batch culture data of mouse hybridoma cells and *in vivo* gene essentiality data. The structural analysis of the network identified highly connected and bridging metabolites, revealing the high correlation between metabolite degree and betweenness centrality. Subsequent functional analysis enabled us to characterize mouse metabolism and identify essential metabolites, enzymes and reactions from the lipid metabolism, which indicated its important role in the cellular growth. Indeed, there were some disagreements between model predications and experimental observation, thus suggesting potential areas for model improvements and expansion in future. We believe that continuing efforts of genome-scale mouse modeling and analysis would help us to understand the mechanism underlying the cellular behavior of mouse and extract valuable information in conjunction with new omics data such as glycomics and lipidomics within the context of mammalian systems biology and biotechnology.

## Methods

### Metabolic network reconstruction

Fig. 1 depicts a schematic overview of the genome-scale reconstruction process and subsequent analyses for characterizing the mouse metabolism. The previous generic model of mouse[26] was considered as a starting point for our enhancement effort. Initially the repeated or redundant reactions in the model were identified and removed. Then, various simulations of the model were performed to verify its ability to produce each cellular component defining the biomass from different carbon sources. It allows us to find missing links or gaps in the network and subsequently fill them by adding relevant enzymatic and transport reactions obtained from several online resources (KEGG, RIKEN, MGI, BRENDA, and ExPaSy) (Table 1) and relevant literature to *M. musculus*. Additionally, information on new open reading frames (ORFs) and GPR association were also included, thus significantly expanding the scope of the model.

### Statistical network analysis

The visualization and statistical analysis of reconstructed genome-scale mouse network were all performed using the network analysis software, BioNetMiner (http://bio.netminer.com). A large-size mouse network can be efficiently visualized by BioNetMiner embedding graph layout algorithms, Force-Directed Kamada-Kawai[56] and GEM:[57] Kamada-Kawai generates a balanced metabolic network graph with uniform distribution and symmetry[55] while GEM draws an aesthetically pleasing layout with the lower number of edge crossings.[58] It also allows us to measure a variety of network properties such as node-degree, shortest paths and clustering coefficients. In addition, the network topology can be statistically analyzed by identifying highly-connected and bridging metabolites using degree and betweenness centrality, respectively (details are given in supplementary 5†).

### Constraints-based flux analysis

Once reconstructed genome-scale metabolic network is stoichiometrically balanced, the predictive capabilities of the model can be examined in both quantitative and qualitative manners by resorting to constraints-based flux analysis.[59] Initially, under stationary assumption during cell growing phase, cell biomass production can be considered as plausible cellular

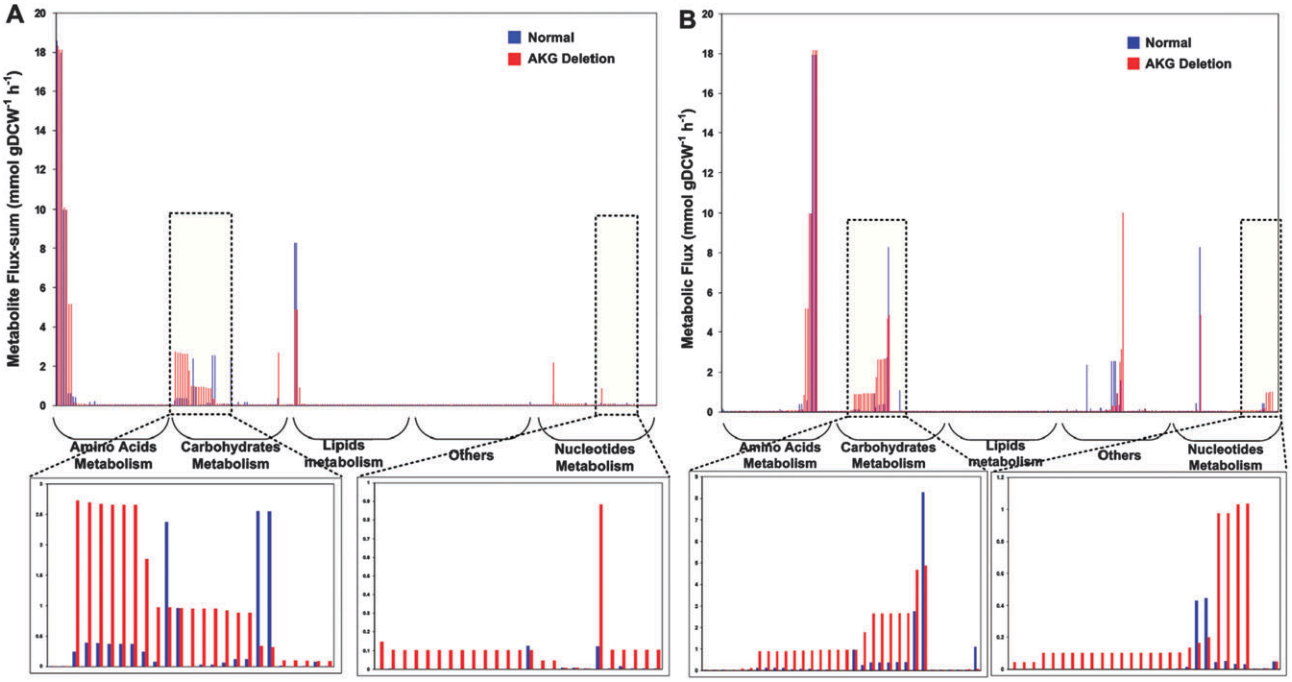



**Fig. 7** Comparison of (A) metabolite flux-sum and (B) metabolic flux distribution during cell growth under normal and AKG deletion conditions. Metabolites flux sum and flux distributions in carbohydrates and nucleotides metabolisms are shown in the enlarged sections. Blue and red colour bars represent normal and AKG deletion conditions, respectively. AKG: α-ketoglutarate.

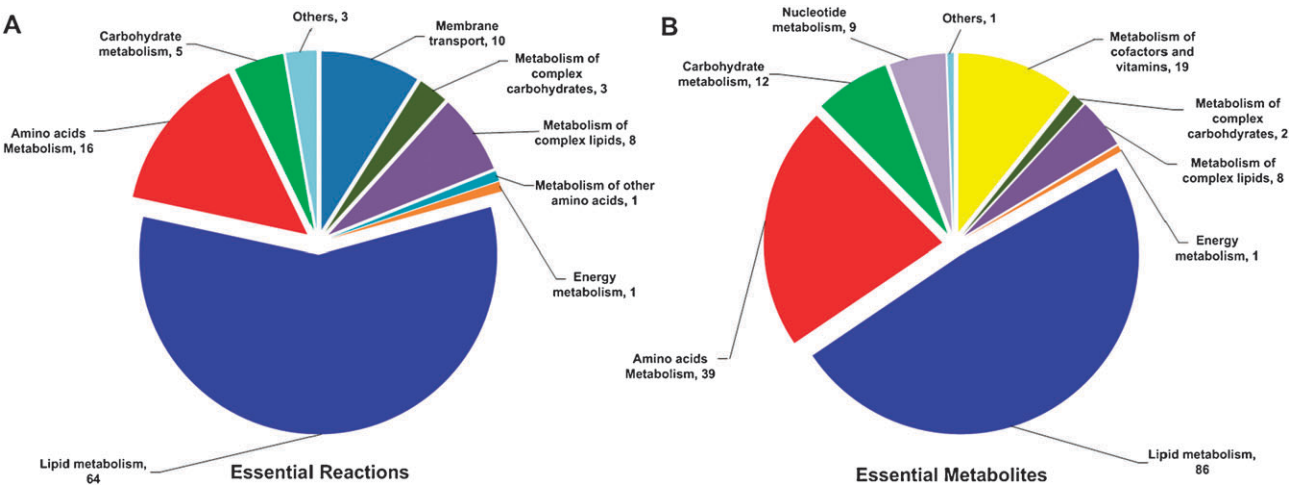



**Fig. 8** Classification of essential (A) reactions and (B) metabolites according to different metabolic subsystems in the mouse metabolism.

objective to be maximized for quantifying the cellular growth phenotype. The resulting growth rate was then compared with experimentally observed specific growth rate. Subsequently, the model can be qualitatively assessed by simulating minimal media requirements and gene deletion analysis. The minimal nutrient components can be determined by minimizing the summation of all consumed substrates from the medium; under the determined minimal medium condition, the cell growth was maximized, constraining each reaction flux to be zero. The reaction and corresponding gene were deemed essential when their removal resulted in zero growth. Similarly, essential metabolites can be identified by forcing the flux sum across each metabolite as zero under cell growth condition as described in Kim *et al.*[60] Finally, the functional organization of the mouse metabolism can be investigated on the basis of gene/metabolite essentiality and its correlation with structural characteristics of the network. All these linear optimization problems were solved by MetaFluxNet[61] and GAMS/CPLEX 10.0.[62]

## Acknowledgements

The work was supported by the Academic Research Fund (R-279-000-258-112) from the National University of Singapore and the Biomedical Research Council of A*STAR (Agency for Science, Technology and Research), Singapore.

## References


1 E. Davidov, J. Holland, E. Marple and S. Naylor, *Drug Discovery Today*, 2003, **8**, 175–183.
2 D. B. Kell, *Drug Discovery Today*, 2006, **11**, 1085–1092.
3 M. L. Yaspo, *Trends Mol. Med.*, 2001, **7**, 494–501.
4 S. Y. Lee, D.-Y. Lee and T. Y. Kim, *Trends Biotechnol.*, 2005, **23**, 349–358.
5 J. D. Trawick and C. H. Schilling, *Biochem. Pharmacol.*, 2006, **71**, 1026–1035.
6 J. L. Reed, I. Famili, I. Thiele and B. O. Palsson, *Nat. Rev. Genet.*, 2006, **7**, 130–141.
7 M. Covert, C. Schilling, I. Famili, J. Edwards, I. Goryanin, E. Selkov and B. Palsson, *Trends Biochem. Sci.*, 2001, **26**, 179–186.
8 N. Price, J. Reed and B. Palsson, *Nat. Rev. Microbiol.*, 2004, **2**, 886–897.
9 J. Edwards, M. Covert and B. Palsson, *Environ. Microbiol.*, 2002, **4**, 133–140.
10 D. Segre, D. Vitkup and G. M. Church, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 15112–15117.
11 A. Burgard, P. Pharkya and C. Maranas, *Biotechnol. Bioeng.*, 2003, **84**, 647–657.
12 A. M. Feist and B. Ø. Palsson, *Nat. Biotechnol.*, 2008, **26**, 659–667.
13 J. Varner and D. Ramkrishna, *Curr. Opin. Biotechnol.*, 1999, **10**, 146–150.
14 A. Varma and B. Palsson, *Appl. Environ. Microbiol.*, 1994, **60**, 3724–3731.
15 S. Selvarasu, D. S.-W. Ow, S. Y. Lee, M. M. Lee, S. K.-W. Oh, I. A. Karimi and D.-Y. Lee, *Biotechnol. Bioeng.*, 2009, **102**, 923–934.
16 S. Kauffman, C. Peterson, B. Samuelsson and C. Troein, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 14796–14799.
17 M. Covert, E. Knight, J. Reed, M. Herrgard and B. Palsson, *Nature*, 2004, **429**, 92–96.
18 D. S. W. Ow, D.-Y. Lee, M. G. S. Yap and S. K. W. Oh, *Biotechnol. Prog.*, 2009, **25**, 61–67.
19 D. Petranovic and J. Nielsen, *Trends Biotechnol.*, 2008, **26**, 584–590.
20 J. Nielsen and M. C. Jewett, *FEMS Yeast Res.*, 2008, **8**, 122–131.
21 A. Graf, M. Dragosits, B. Gasser and D. Mattanovich, *FEMS Yeast Res.*, 2009, **9**, 335–348.
22 N. Duarte, M. Herrgard and B. Palsson, *Genome Res.*, 2004, **14**, 1298–1309.
23 K. R. Patil, M. Akesson and J. Nielsen, *Curr. Opin. Biotechnol.*, 2004, **15**, 64–69.
24 M. Durot, P. Y. Bourguignon and V. Schachter, *FEMS Microbiol. Rev.*, 2009, **33**, 164–190.
25 P. M. O'Callaghan and D. C. James, *Briefings Funct. Genomics Proteomics*, 2008, **7**, 95–110.
26 K. Sheikh, J. Forster and L. K. Nielsen, *Biotechnol. Prog.*, 2005, **21**, 112–121.
27 S. Selvarasu, V. V.T. Wong, I. A. Karimi and D.-Y. Lee, *Biotechnol. Bioeng.*, 2009, **102**, 1494–1594.
28 N. C. Duarte, S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas and B. O. Palsson, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 1777–1782.
29 T. Shlomi, M. N. Cabili, M. J. Herrgard, B. Ø. Palsson and E. Ruppin, *Nat. Biotechnol.*, 2008, **26**, 1003–1010.
30 A. Wahl, Y. Sidorenko, M. Dauner, Y. Genzel and U. Reichl, *Biotechnol. Bioeng.*, 2008, **101**, 135–152.
31 H. Bono, I. Nikaido, T. Kasukawa, Y. Hayashizaki and Y. Okazaki, *Genome Res.*, 2003, **13**, 1345–1349.
32 J. M. Reichert and V. E. Valge-Archer, *Nat. Rev. Drug Discovery*, 2007, **6**, 349–356.
33 J. Rossant and S. Scherer, *GenomeBiology*, 2003, **4**, 109.
34 M. E. Reff, *Curr. Opin. Biotechnol.*, 1993, **4**, 573–576.
35 T. Fukuwatari, S. Ohsaki, S. Fukuoka, R. Sasaki and K. Shibata, *Toxicol. Sci.*, 2004, **81**, 302–308.
36 C. A. Mackintosh and A. E. Pegg, *Biochem. J.*, 2000, **351**, 439–447.
37 C. Porter and R. Bergeron, *Science*, 1983, **219**, 1083–1085.
38 D. De Alwis, R. Dutton, J. Scharer and M. Moo-Young, *Bioprocess Biosyst. Eng.*, 2007, **30**, 107–113.
39 K. Yamamoto and A. Niwa, *Amino Acids*, 1993, **5**, 1–16.
40 National Research Council, *Nutrient requirements of laboratory animals*, National Academy Press, Washington, DC, 1995.
41 H. R. Zielke, C. L. Zielke and P. T. Ozand, *Fed. Proc.*, 1984, **43**, 121–125.
42 Y.-Y. Cho, M.-J. Kang, H. Sone, T. Suzuki, M. Abe, M. Igarashi, T. Tokunaga, S. Ogawa, Y. A. Takei, T. Miyazawa, H. Sasano, T. Fujino and T. T. Yamamoto, *Biochem. Biophys. Res. Commun.*, 2001, **284**, 993–997.
43 H. Ma and A.-P. Zeng, *Bioinformatics*, 2003, **19**, 270–277.
44 D. Zhu and Z. Qin, *BMC Bioinformatics*, 2005, **6**, 8.
45 N. Westergaard, J. Drejer, A. Schousboe and U. Sonnewald, *Glia*, 1996, **17**, 160–168.
46 Z. Kovacevic and J. D. McGivan, *Physiol. Rev.*, 1983, **63**, 547–605.
47 C. Schilling, M. Covert, I. Famili, G. Church, J. Edwards and B. Palsson, *J. Bacteriol.*, 2002, **184**, 4582–4593.
48 M. R. Wenk, *Nat. Rev. Drug Discovery*, 2005, **4**, 594–610.
49 R. Korke, M. d. L. Gatti, A. L. Y. Lau, J. W. E. Lim, T. K. Seow, M. C. M. Chung and W.-S. Hu, *J. Biotechnol.*, 2004, **107**, 1–17.
50 B. Fleischer, *J. Histochem. Cytochem.*, 1983, **31**, 1033–1040.
51 A. Yan and W. J. Lennarz, *J. Biol. Chem.*, 2005, **280**, 3121–3124.
52 A. Dove, *Nat. Biotechnol.*, 2001, **19**, 913–917.
53 G. Liu, D. D. Marathe, K. L. Matta and S. Neelamegham, *Bioinformatics*, 2008, **24**, 2740–2747.
54 F. J. Krambeck and M. J. Betenbaugh, *Biotechnol. Bioeng.*, 2005, **92**, 711–728.
55 M. P. Murrell, K. J. Yarema and A. Levchenko, *ChemBioChem*, 2005, **5**, 1334–1347.
56 T. Kamada and S. Kawai, *Inf. Process. Lett.*, 1989, **31**, 7–15.
57 A. Frick, A. Ludwig and H. Mehldau, in *Proceedings of the DIMACS International Workshop on Graph Drawing*, Springer-Verlag, London, UK, 1994, pp. 388–403.
58 K. Sugiyama, *World Scientific*, 2002, **11**, 87–91.
59 J. S. Edwards and B. O. Palsson, *Biotechnol. Bioeng.*, 1998, **58**, 162–169.
60 P.-J. Kim, D.-Y. Lee, T. Y. Kim, K. H. Lee, H. Jeong, S. Y. Lee and S. Park, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 13638–13642.
61 D.-Y. Lee, H. Yun, S. Park and S. Y. Lee, *Bioinformatics*, 2003, **19**, 2144–2146.
62 A. Brooke, D. Kendrick, A. Meeraus and R. Raman, *GAMS Development Corporation*, 1998.