Molecular BioSystems

Cite this: DOI: 10.1039/c2mb25075f

www.rsc.org/molecularbiosystems

PAPER

A compendium of inborn errors of metabolism mapped onto the human metabolic network†‡

Swagatika Sahoo,[a] Leifur Franzson,[b] Jon J. Jonsson[bc] and Ines Thiele*[ad]

Received 6th March 2012, Accepted 27th April 2012
DOI: 10.1039/c2mb25075f

Inborn errors of metabolism (IEMs) are hereditary metabolic defects, which are encountered in almost all major metabolic pathways occurring in man. Many IEMs are screened for in neonates through metabolomic analysis of dried blood spot samples. To enable the mapping of these metabolomic data onto the published human metabolic reconstruction, we added missing reactions and pathways involved in acylcarnitine (AC) and fatty acid oxidation (FAO) metabolism. Using literary data, we reconstructed an AC/FAO module consisting of 352 reactions and 139 metabolites. When this module was combined with the human metabolic reconstruction, the synthesis of 39 acylcarnitines and 22 amino acids, which are routinely measured, was captured and 235 distinct IEMs could be mapped. We collected phenotypic and clinical features for each IEM enabling comprehensive classification. We found that carbohydrate, amino acid, and lipid metabolism were most affected by the IEMs, while the brain was the most commonly affected organ. Furthermore, we analyzed the IEMs in the context of metabolic network topology to gain insight into common features between metabolically connected IEMs. While many known examples were identified, we discovered some surprising IEM pairs that shared reactions as well as clinical features but not necessarily causal genes. Moreover, we could also re-confirm that acetyl-CoA acts as a central metabolite. This network based analysis leads to further insight of hot spots in human metabolism with respect to IEMs. The presented comprehensive knowledge base of IEMs will provide a valuable tool in studying metabolic changes involved in inherited metabolic diseases.

Introduction

Inborn errors of metabolism (IEMs) are individually rare but collectively numerous, affecting the metabolism of many human organs. These metabolic defects are congenital and represent single or multiple enzyme deficiencies, which if left untreated can lead to life threatening conditions, with a current incidence rate of 1 : 800 live births.[1] IEMs can be grouped into three diagnostically meaningful groups:[2] (i) disorders giving rise to intoxication, *via* accumulation of intracellular compounds over time, (ii) disorders involving energy metabolism, and (iii) disorders involving metabolism of complex molecules (Fig. S1, ESI‡). Moreover, IEMs can be present at any age – from fetal life to old age – and the symptoms vary between the groups as well as within the groups. Not all IEMs can be easily diagnosed, and treatment does not exist for many of them, although many disorders of the first group can now be treated by changing the patient's diet.[2] In recent years, numerous reviews have been published analyzing IEMs from various aspects, including newborn screening programs and systems biology,[3–5] highlighting their importance and the general interest in understanding the molecular basis of IEMs.

The use of tandem mass spectrometry (MS/MS) has enabled improvements in metabolic screening methods to more reliably detect IEMs among newborns with a low false positive rate.[6] Newborn screening is mandatory in most countries; however, the measured biomarkers differ significantly between them, with 42 (20 primary and 22 secondary disorders) metabolic conditions screened for newborn screening in some states within the USA and countries participating in the Stork program (Iceland being one of the participating country).[7,8] The substantial increase in the number of patients diagnosed with IEMs[4,9] can be majorly attributed to improved and standardized measurements of amino acids and acylcarnitine concentrations.[7,10] For the diagnosis of some IEMs, ratios of biomarker provide more

[a] *Center for Systems Biology, University of Iceland, Iceland. E-mail: ines.thiele@gmail.com*
[b] *Department of Biochemistry & Molecular Biology, Faculty of Medicine, University of Iceland, Iceland*
[c] *Department of Genetics and Molecular Medicine, Landspitali, National University Hospital of Iceland, Iceland*
[d] *Faculty of Industrial Engineering, Mechanical Engineering & Computer Science, University of Iceland, Iceland*
† Published as part of a themed issue dedicated to Emerging Investigators.
‡ Electronic supplementary information (ESI) available. See DOI: 10.1039/c2mb25075f

insight about pathological states than absolute biomarker concentrations (*e.g.*, tyrosine to phenylalanine ratio for the diagnosis of phenylketonuria, OMIM: 261600).

Fatty acid oxidation defects are highly prevalent IEMs with a collective incidence rate of 1 : 9000.[11] In the case of infants, fatty acid oxidation serves as energy source within the first 12 hours of fasting; thus, they are more likely to manifest fatty acid oxidation defects.[12] Acylcarnitines tend to rise in blood and urine on the first day of the birth and then decline gradually.[6] By profiling numerous acylcarnitines, MS/MS technology has expanded the range of detection for fatty acid disorders (*e.g.*, of medium chain acyl-CoA dehydrogenase deficiency) and for organic acid disorders (*e.g.*, propionic academia and glutaric aciduria type-I).

Metabolic network reconstructions are assembled in a bottom-up approach based on genome annotation, biochemical, and physiological data.[13] They summarize current knowledge about the target organism in a structured, stoichiometric manner. In many cases, manual curation including in-depth literature search is performed to ensure high quality and coverage of the reconstruction.[14] Once reconstructed, metabolic networks can be queried for their biochemical information, be employed to map various data types (*e.g.*, transcriptomic data or metabolomic data), or serve as a starting point for computational modelling (see ref. 15 for a recent review).

Results and discussion

The aim of this study was to compile a compendium of IEMs that map onto the available reconstruction of human metabolism[16] to enable the analysis of IEMs in a metabolic modelling context (Fig. 1). Therefore, we first reconstructed the fatty acid oxidation pathways to enable the mapping of routinely measured acylcarnitines, which were missing in the published reconstruction of human metabolism.[16] Subsequently, we mapped IEMs onto the expanded human metabolic network based on known affected genes, and assembled important phenotypic and clinical features of each IEM. Finally, we investigated the IEMs in the context of the expanded human metabolic reconstruction by identifying IEMs, which shared reactions and metabolites, therefore, representing hot spots in human metabolism.

Detailed reconstruction of acylcarnitine metabolism

In this study, we employed the published human metabolic reconstruction, Recon 1,[16] as a starting point for the acylcarnitine and fatty acid oxidation (AC/FAO) reconstruction module. The human metabolic reconstruction accounts for metabolic reactions occurring in any human cell, as defined by the human genome. As such, the reactions present in Recon 1 may not occur in all cells, but a particular cell type may only express a subset of metabolic enzymes captured by Recon 1. For example, the liver completely detoxifies ammonia to urea *via* urea cycle, however, in kidney the urea cycle is only partially active, *i.e.*, kidney accounts for arginine synthesis. The majority of the reactions are mass and charge balanced, and many of them are assigned with structured Boolean relationships between genes, proteins, and reactions. Recon 1 captures all amino acids (except methylhistidine) measured in the newborn screening but it only accounts for the metabolism of six of the 39 acylcarnitines (Table S1, ESI‡). Moreover, Recon 1 accounts for most of the relevant reactions in beta fatty acid oxidation, while alpha and omega fatty acid oxidation reactions are only partially captured. After collecting supporting evidence from more than 150 peer reviewed articles and books, we reconstructed an AC/FAO module that accounts for 352 reactions, 139 metabolites, and 14 genes, with the reactions distributed within the endoplasmic reticulum, peroxisomes, mitochondria, and cytosol (Fig. 2B). The majority of the reactions are involved in fatty acid metabolism (Fig. 2A). Note that this reconstruction module is only functional in conjunction with Recon 1[16] or its successor.[17]

Acylcarnitines in the AC/FAO module. The AC/FAO module accounts for 39 acylcarnitines routinely measured in newborn screening (Table S1, ESI‡), which include dicarboxylic-acylcarnitines, hydroxy-acylcarnitines, and other short chain acylcarnitines (Fig. 2B). For instance, succinylcarnitine, a four carbon fatty acid is a dicarboxylic-acylcarnitine and generated *via* the valine, isoleucine and methionine pathway. Hydroxy-isovalerylcarnitine is a five carbon fatty acid attached to a carnitine moiety, has a hydroxy group at the third carbon atom and is generated *via* the leucine pathway. These two dicarboxylic-acylcarnitines are used as biomarkers for organic acidemias and more recently also for biotin deficiency.[18–20] A total of five reactions were added describing the synthesis of succinylcarnitine and hydroxy-isovalerylcarnitine, and subsequent processing by the carnitine palmitoyltransferase 1 enzyme (CPT-1, E.C. 2.3.1.21), which is encoded by three gene isoforms *CPT1A* (GeneID: 1374), *CPT1B* (GeneID: 1375), and *CPT1C* (GeneID: 126129). Another dicarboxylic-acylcarnitine is glutarylcarnitine, which is an important biomarker in glutaric aciduria,[21,22] generated as an intermediate in lysine and tryptophan metabolism. Likewise, 3-hydroxy-butyrylcarnitine, a hydroxy-acylcarnitine and biomarker for short-chain-acyl-CoA-dehydrogenase deficiency,[23] is generated *via* butanoate metabolism. Hydroxyacylcarnitines generally serve as biomarkers for several mitochondrial fatty acid beta oxidation disorders.[24]

Other short chain acylcarnitines include acetylcarnitine, tiglylcarnitine, and isovalerylcarnitine. Acetylcarnitine is produced from acetyl-CoA, which is an important end product of fatty acid oxidation, ketogenic amino acid metabolism, and a precursor of cholesterol metabolism. An increased level of acetylcarnitine may reflect CPT-2 deficiency.[25] Tiglylcarnitine and isovalerylcarnitine are intermediates of isoleucine and leucine metabolism, respectively, and are formed from their respective precursor fatty acyl-CoAs, tiglyl-CoA and isovaleryl-CoA, by the action of the CPT-1 enzyme. These short chain acylcarnitines are biomarkers for methylacetoacetyl-CoA deficiency, isovaleric acidemia, and short-chain acyl-CoA dehydrogenase deficiency.[26–28] The metabolic reactions of these important short chain acylcarnitines were added to the AC/FAO module.

Addition of carnitine shuttle to mitochondria and ABC transport reactions to the peroxisome. Specific transporters are required to deliver acyl-CoAs to their destined sub-cellular compartments for their oxidation. The three step carnitine shuttle system transports acyl-CoAs into the mitochondrion and this transport is reversible.[29] It involves reactions catalyzed by the CPT-1 enzyme,

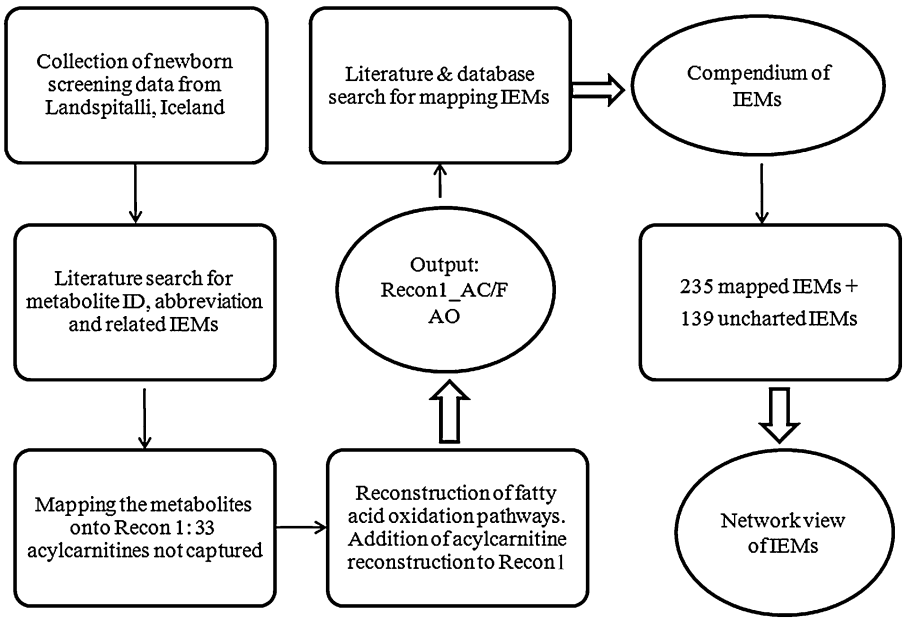


Fig. 1 Overview of the workflow that we used in this study. The metabolites, which are measured in the newborn screening program, were identified along with their chemical names, chemical formula, charge, normal physiological concentration, and related IEMs. These metabolites were then mapped onto Recon1, using name and formula, which led to the mapping of 20 amino acids and six acylcarnitines. For the remaining metabolites (*i.e.*, 33 acylcarnitines), which could not be mapped, we reconstructed their respective metabolic pathways based on the literature, and added the corresponding reactions to the AC/FAO module. Subsequently, we combined the AC/FAO module with Recon 1, resulting in Recon1_AC/FAO. Thereafter, we employed the information retrieved from the literature to map the IEMs onto this extended reconstruction. A total of 235 IEMs could be mapped and subsequently classified along with their different physiological and clinical features. Another set of 139 IEMs were identified, termed as future IEMs, which could not be mapped onto Recon1_AC/FAO due to missing genes in the reconstruction. These 374 IEMs constitute the IEM compendium, which we mapped onto the human metabolic network to obtain a holistic network perspective of the captured IEMs.

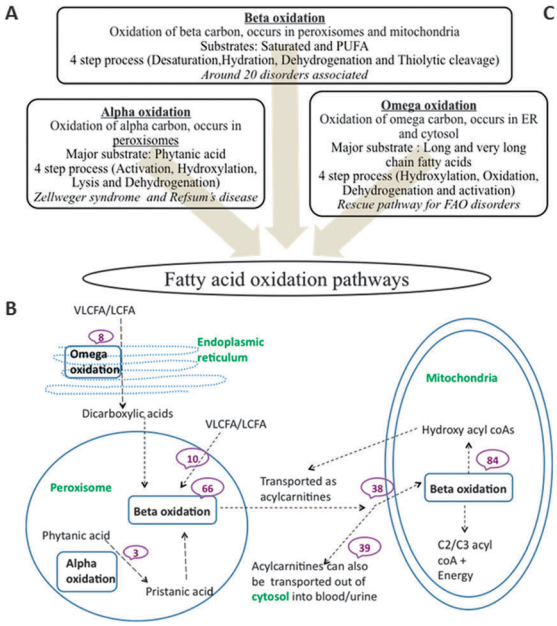
the carnitine/acylcarnitine translocase protein (*SLC25A20*, GeneID: 788) and the carnitine *O*-palmitoyltransferase 2 (*CPT2*, GeneID: 1376, E.C. 2.3.1.21). The transport is affected in some IEMs, such as the carnitine uptake defect, CPT-1 deficiency, and CPT-2 deficiency.[29] The AC/FAO module contains a refined carnitine shuttle system with 38 reactions accounting for the reversible transport of a wide range of metabolites. Peroxisomes contain ATP-dependent transporters for the import of very long chain acyl-CoAs (Fig. 2B).[30] A total of ten ABC transport reactions and their associated genes were identified, including three genes of the ATP-binding cassette superfamily (*ABCD1*, GeneID: 215; *ABCD2*, GeneID: 225; *ABCD3*, GeneID: 5825). The human metabolic reconstruction accounts for the transport of 4,8-dimethylnonanoyl-CoA by the carnitine O-octanoyltransferase (*CROT*, GeneID: 54677, E.C. 2.3.1.137). The same protein also transports butyryl-CoA and hexanoyl-CoA into the peroxisome.[29] We included this extended substrate specificity in the AC/FAO module.

Addition of new fatty acid oxidation genes

To capture all fatty acid oxidation related IEMs, 84 mitochondrial beta-oxidation reactions and 63 metabolites were added to the AC/FAO module. We also added the necessary genes and their respective gene products, including hydroxyacyl-CoA dehydrogenase (*HADH*, GeneID: 3033, E.C. 1.1.1.35) and the acetyl-CoA acyltransferase 2 (*ACAA2*, GeneID: 10449, E.C. 2.3.1.16), which have substrate specificity for short and medium chain acyl-CoAs and catalyze the dehydrogenation and thiolytic cleavage step, respectively.[31] The 2,4-dienoyl-CoA reductase 1 (*DECR1*, GeneID: 1666, E.C. 1.3.1.34) was another auxiliary enzyme of the mitochondrial beta-oxidation of unsaturated acyl-CoAs that was missing in the human metabolic reconstruction. The addition of these three genes enabled the mapping of further five IEMs (Table S2, ESI‡). Moreover, four additional genes were added to the AC/FAO module, whose products are mainly involved in activation and transport of fatty acids.[32,33] These were fatty acid transporter (*SLC27A6*, GeneID: 28965), bubblegum (*ACSBG1*, GeneID: 23205, E.C. 6.2.1.3), bubblegum related protein (*ACSBG2*, GeneID: 81616, E.C. 6.2.1.3), and diazepam binding inhibitor (*DBI*, GeneID: 1622).

Long and very long chain acyl-CoAs are oxidized in peroxisome before entering the mitochondrion.[34,35] Therefore, 66 peroxisomal beta-oxidation reactions and 59 metabolites were included in the AC/FAO module (Fig. 2B) permitting the mapping of X-linked adrenoleukodystrophy (OMIM: 300100), acyl-CoA oxidase-1 (OMIM: 264470), and D-bifunctional-protein deficiencies (OMIM: 261515). The peroxisome contains the complete enzymatic machinery for handling unsaturated fatty acids.[36,37] In the human metabolic reconstruction, important auxiliary proteins, reactions, and their respective genes were missing and were thus included in the AC/FAO module. This addition comprises the functions of the peroxisomal 2,4-dienoyl-CoA reductase 2 (*DECR2*, GeneID: 26063, E.C. 1.3.1.34) and the enoyl-CoA delta isomerase 2 (*PECI*, GeneID: 10455, E.C. 5.3.3.8).

Alpha oxidation of phytanic acid was also reconstructed in detail in the AC/FAO module (Fig. 2B). Two new genes and



C

Components	Recon 1	Recon1_ AC/FAO
Genes	1496	1510
Transcripts	1905	1919
Compartments	8	8
Reactions	3743	4095
Exchange reactions	405	444
Unique metabolites	1509	1648
Citations	1587	+150**

Fig. 2 General properties of the AC/FAO module. (A) Overview of the fatty acid oxidation pathways, along with their definitions, substrate preferences, key steps involved, and associated disorders. (B) Schematic map of the key reactions in the AC/FAO module and their sub-cellular location. The numbers indicate the number of reactions per metabolic pathway. The substrates for the omega oxidation pathway are very long chain fatty acids and long chain fatty acids, while the products are dicarboxylic acids, which further undergo beta oxidation in the peroxisome. Phytanic acid, which is a substrate for the alpha fatty acid oxidation, produces pristanic acid. This product is further beta-oxidized in the proxisomes before finally entering the mitochondrion. The fatty acyl-CoA molecules are transported between the peroxisomes and mitochondria in the form of acylcarnitines, and once inside the mitochondria they may either undergo complete beta oxidation generating acetyl-CoA (in the case of even chain fatty acid) or propionyl-CoA (in the case of odd chain fatty acid) and finally produce energy through the oxidative phosphorylation in the mitochondria. However, if the beta oxidation is incomplete, it produces hydroxy fatty acids, which are then shuttled out of the mitochondrion as acylcarnitines. Substrates for peroxisomal beta oxidation are long chain and vary long chain fatty acids entering the peroxisome *via* an ATP dependent transport system. Short, medium, and long chain fatty acids enter mitochondria as acylcarnitines. (C) Key characteristics of Recon 1 and Recon1_AC/FAO. A reversible version of the Recon 1 reactions, *i.e.*, 64 reactions of the carnitine shuttle system, was also added (not shown here, see Table S9 and supplement text for details). *Taken from ref. 103. **Corresponds to primary literature, peer review articles, and books that were used only for the acylcarnitine reconstruction. Abbreviations used: FA – fatty acid; PUFA – poly unsaturated fatty acid; ER – endoplasmic reticulum; FAO – fatty acid oxidation; VLCFA – very long chain fatty acid; LCFA – long chain fatty acid; MCFA – medium chain fatty acid; SCFA – short chain fatty acid; C2 – acetyl-CoA; C3 – propionyl-CoA.

the reactions catalyzed by the respective gene product were added: fatty acid transporter (*SLC27A2*, GeneID: 11001, E.C. 6.2.1.-) and phytanoyl-CoA 2-hydroxylase 1 (*HACL1*, GeneID: 26061, E.C. 4.1.-.-). Reduced activity of the SLC27A2 transporter protein has been reported to be responsible for the biochemical pathology in X-linked adrenoleukodystrophy in mouse models.[38] Omega oxidation is important for the generation of dicarboxylic acids; therefore, eight reactions were added to AC/FAO (Fig. 2B).

Detailed reconstruction of fatty acid oxidation reactions. The long chain fatty acids were only partially captured in the human metabolic reconstruction. Gene-protein-reaction (GPR) associations for three acyl-CoA dehydrogenases: C-2 to C-3 short chain (*ACADS*, GeneID: 35, E.C. 1.3.99.2), C-4 to C-12 straight chain (*ACADM*, GeneID: 34, E.C. 1.3.99.3), and very long chain (*ACADVL*, GeneID: 37, E.C. 1.3.99.-), were incorporated according to their substrate specificities. We also added the reactions catalyzed by the trifunctional hydroxyacyl-CoA dehydrogenase (*HADHA*, GeneID: 3030 and *HADHB*, GeneID: 3032; E.C. 4.2.1.17/1.1.1.211/ 2.3.1.16),[31] which catalyses the steps after desaturation of long acyl-CoAs (>14 carbon units). The mitochondrial short chain enoyl-CoA hydratase 1 (*ECHS1*, GeneID: 1892, E.C. 4.2.1.17) catalyzes the hydration step for medium and short chain fatty acyl-CoAs, but in Recon 1, the gene product catalyzes only reactions involved in tryptophan and beta-alanine metabolism. Hence, the hydration steps are needed to be added to account for the relevant fatty acid oxidation reactions. We also expanded the substrate specificity of the enoyl-CoA delta isomerase 1 (*DCI*, GeneID: 1632, E.C. 5.3.3.8), which has only one reaction associated in Recon 1. These comprehensive additions to the fatty acid oxidation pathway enabled the mapping of numerous IEMs, including peroxisomal acyl-CoA oxidase deficiency/adrenoleukodystrophy pseudoneonatal (OMIM: 264470), Zellweger syndrome, neonatal adrenoleukodystrophy and infantile Refsum disease, D-bifunctional protein deficiency (OMIM: 261515), and pseudo-Zellweger syndrome.

All enzymes involved in omega oxidation were present in Recon 1, but catalyzed only reactions involving either xenobiotic or eicosanoid metabolism, while reactions of the omega oxidation of fatty acids were missing. The AC/FAO module accounts for all known reactions of the omega oxidation

pathway[39–43] along with the relevant GPR association, linking the corresponding genes to these reactions. The participating genes include the cytochrome P450 family genes (*CYP4F2*, GeneID: 8529, E.C. 1.14.13.30; *CYP4F3*, GeneID: 4051, E.C. 1.14.13.30), the aldehyde dehydrogenase 3 family (*ALDH3A2*, GeneID: 224, E.C. 1.2.1.3), the alcohol dehydrogenase (*ADH5*, GeneID: 128, E.C. 1.1.1.1), and the acyl-CoA synthetase (*ACSL5*, GeneID: 51703, E.C. 6.2.1.3). A detailed description of all fatty acid oxidation pathways is provided in ESI.‡

Mapping of inborn errors of metabolism onto the human metabolic reconstruction

A major aim of this study was to create a comprehensive compendium of IEMs that maps onto the human metabolic reconstruction. We therefore joined our AC/FAO module with the published human metabolic reconstruction (Recon 1)[16] and called this extended network Recon1_AC/FAO. To map defective genes (*i.e.*, gene mutations that cause an IEM) reported in the literature and databases for the various IEMs to reactions, we employed the GPR associations defined for the 1510 metabolic genes accounted for in Recon1_AC/FAO. After extensive literature search, we identified 235 IEMs, which had all causal genes captured in Recon1_AC/FAO (Table S2, ESI‡). In 139 cases, not all genes known to cause an IEM were present in the reconstruction and we collected them in a separate list (Table S3, ESI‡), which represents a starting point for future extension and gap filling of functionalities included in the human metabolic reconstruction.[44,45] The following information was retrieved for each IEM from the literature and databases: affected organs, mode of inheritance, biomarkers, biomarker concentrations, and phenotype. While we illustrate some interesting cases below, a complete list of mapped IEMs and all relevant information can be found in Table S2, ESI.‡

Pathway classification. The 235 mapped IEMs could be grouped into 16 different central metabolic pathways (Fig. 3A). The highest number of IEMs was found to be in carbohydrate metabolism (65), followed by amino acid metabolism (54), and lipid metabolism (51). Carbohydrate metabolism includes a wide range of metabolic pathways, more so, since we counted also glycoprotein and glycolipid metabolism to this subsystem. Moreover, this class of inherited disorders is well studied.[46] The 235 IEMs were caused by defects in 250 unique metabolic genes, which encode enzymes catalyzing 1067 reactions in total (this includes 750 unique and 317 shared metabolic reactions). A similar distribution pattern was observed for the affected genes, when categorizing them according to their biochemical pathway. The higher number of reactions relative to the number of genes was caused by the broad substrate specificity of many enzymes, which was particularly true for enzymes involved in the fatty acid oxidation metabolism. This analysis highlights that a systems biology modelling approach may facilitate significantly the assessment of an IEM's contribution to its global metabolic phenotype.

Mode of inheritance and phenotype analysis. IEMs follow a particular mode of inheritance.[47] We grouped the IEMs based on five different modes of inheritance: (i) autosomal recessive, (ii) autosomal dominant, (iii) autosomal recessive or autosomal dominant (as seen in hypophosphatasia, OMIM: 241500, OMIM: 241510, and OMIM: 146300), (iv) X-linked or autosomal dominant (as seen in familial gynecomastia, OMIM: 139300 and OMIM: 107910), and (v) X-linked pattern (X-linked recessive and X-linked dominant IEMs treated as one mode, *i.e.*, X-linked pattern) (Table S2 and S4, ESI‡). Inheritance patterns were identified for 199 (61%) of the IEMs, as the majority of the remaining IEMs have not been studied sufficiently or have been only identified in very few individuals so that the pattern of inheritance could not be established. This latter group includes acetyl-CoA carboxylase deficiency (OMIM: 200350), AICA-ribosiduria (OMIM: 608688), and hepatic lipase deficiency (OMIM: 151670). The majority of the IEMs with the reported inheritance pattern were autosomal recessive (163 IEMs, 82%).

IEMs can also be classified according to the number of phenotypic patterns observed. Different phenotypes arise due to modifier genes, allelic variations, complex genetic and environmental interactions, or environmental factors. Therefore, specific effects may be seen at different age (*i.e.*, appearance of specific morphological features), or may lead to difference in the severity of the disorder.[48,49] We were able to collect such information for 90 IEMs (38%) (Table S2 and S5, ESI‡). The majority of these IEMs present two phenotypic forms. For example, a classic form and a mild form (*e.g.*, in medium chain acyl-CoA dehydrogenase deficiency, OMIM: 201450), an acute and a chronic form (*e.g.*, in isovaleric academia, OMIM: 243500), an early and a delayed onset (*e.g.*, in carbamoyl phosphate synthetase I deficiency, OMIM: 237300), a severe and a moderate form (*e.g.*, in glucose-6-phosphate dehydrogenase deficiency, OMIM: 305900), a type 1 and a type 2 form (*e.g.*, in 3-methylcrotonyl-CoA carboxylase deficiency, OMIM: 210200 and OMIM: 210210). Other IEMs have a more complex phenotypic pattern (Table S2, ESI‡).

Organ classification. Clinical manifestations observed in affected patients with particular IEMs were also noted. A total of 28 organ systems were affected in 217 distinct IEMs. Most of the remaining IEMs, such as essential pentosuria (OMIM: 260800), erythrocyte AMP deaminase deficiency (OMIM: 612874), and acatalasemia (OMIM: 115500), present no apparent clinical dysfunction. As one may expect, the organ systems were not equally affected by IEMs (Fig. 3B). A criterion for calling an organ 'affected' was either (i) when a clinical sign was reported by a clinical practitioner (*e.g.*, *tongue* in mucopolysaccharidosis VI, OMIM: 253200), (ii) when a symptom was described by the patient, (iii) an identified pathological report, (iv) an established biochemical defect (*e.g.*, *kidneys* in renal glucosuria, OMIM: 233100, due to inability of the renal tubules to reabsorb glucose), or (v) a combination of these criteria. For example, cherry red spot signs on the eye have been reported as abnormal feature in Sandhoff disease (OMIM: 268800) and Tay–Sach's disease (OMIM: 272800); thus, the *eye* was noted as being affected. For some IEMs, organs can be affected at a later stage, as it is the case for the CPT-2 deficiency (lethal neonatal form, OMIM: 608836), where postmortem examination showed diffuse lipid accumulation in the liver, heart, kidneys, adrenal cortex,

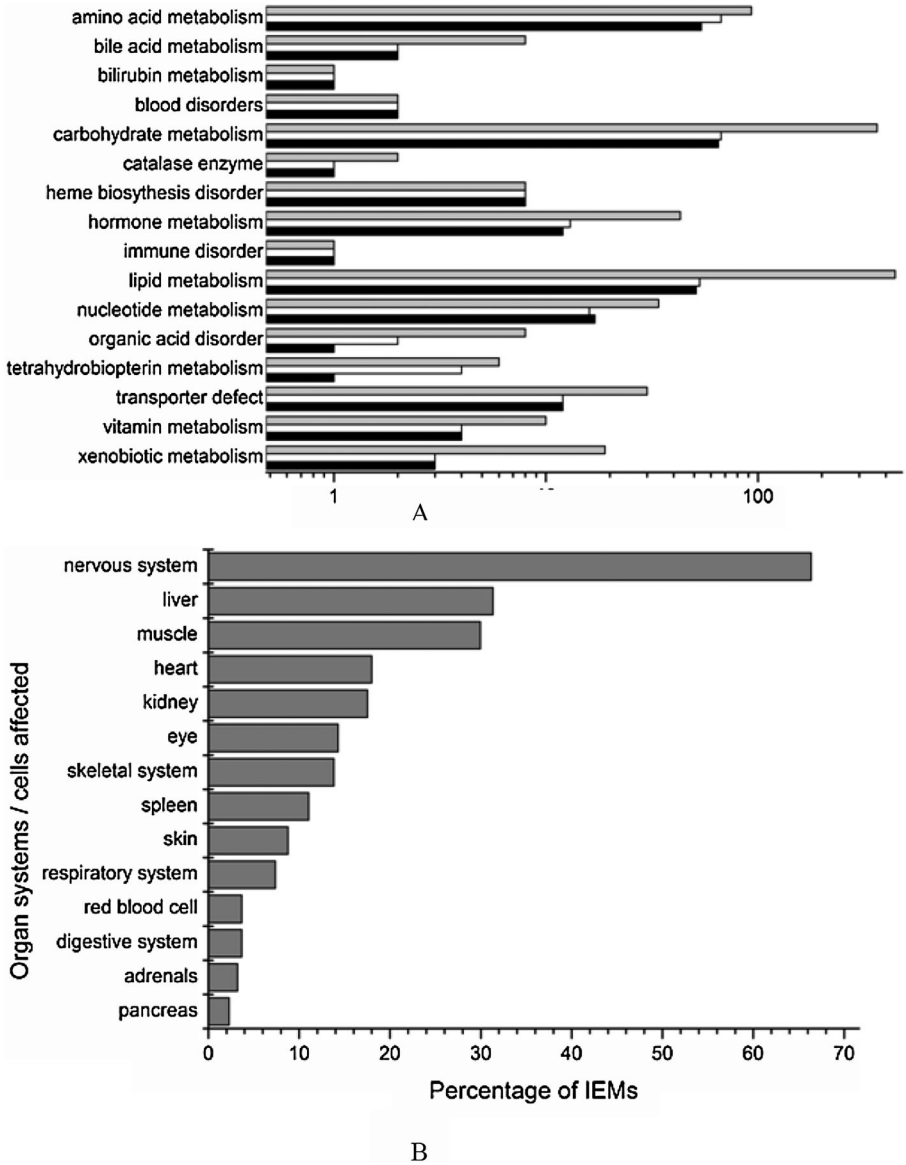


Fig. 3 Distribution of IEMs based on various properties. (A) The 235 IEMs mapped onto Recon1_AC/FAO were categorized into 16 different metabolic categories. The comparison between the number of IEMs (black), number of genes (white), and number of reactions (grey) per IEM category is plotted on a logarithmic scale. A total of 250 unique genes encode for gene products catalyzing 750 unique metabolic reactions, which are affected in at least one IEM. (B) Percentage of IEMs that affect different organ systems. For simplicity we only show highly affected organs.

skeletal muscle, and lungs. In these cases, the affected organs were noted based on these data. Certain organs, such as the *liver*, *pancreas*, *bile canaliculi*, and *gall bladder* were deliberately treated separately from the digestive system since these organs are specifically affected in certain disorders without involvement of the digestive system as a whole. For example, glycogen storage disease type 6/Hers disease affects only the liver,[50] while the HMG-CoA lyase deficiency (OMIM: 246450) affects only liver and pancreas and the familial hypercholanemia (OMIM: 607748) affects the liver and bile canaliculi. The category *digestive system* mainly represents the small and large intestine in our classification scheme. The *nervous system* includes the central nervous system and the peripheral nervous system. The nervous system was also deemed to be affected in IEMs with hypotonia as a clinical sign (*e.g.*, in 3-methylcrotonyl-CoA carboxylase deficiency, propionic academia, and aromatic L-amino acid decarboxylase deficiency[51]) since hypotonia, or reduced muscle tone, is related to motor nerve dysfunction, which is controlled by the brain. Using this classification scheme, we found the nervous system to be the most commonly affected organ by IEMs. This result can be explained in part on the basis that in almost all aminoacidopathies (second most prevalent category of IEMs, Fig. 3A), blood concentration of one or more amino acids is increased, which then competitively inhibits the transport of essential amino acids across the blood–brain-barrier.[52] Amino acids and their important derivative metabolites also serve as neurotransmitters, which are serine (produces the choline part of acetylcholine), tyrosine (produces L-3,4-dihydroxyphenylalanine, dopamine, and catecholamines), tryptophan (produces serotonin), glycine, glutamate (produce gama-aminobutyric acid), and histidine (produces histamine). In many aminoacidurias, there may be

observed deficits of neurotransmitters, particularly of *N*-methyl-D-aspartate, and of their receptors.[52] Furthermore, low levels of *N*-acetylaspartate and high levels of lactate in the brain, besides high levels of branched chain amino acids in both the brain and blood, are observed during metabolic de-compensation, and indicate mitochondrial dysfunction leading to brain energy failure.[53,54] These examples illustrate how a defective amino acid metabolism can dramatically affect the nervous system. The category '*muscles*' includes skeletal and smooth muscles but not cardiac muscles, since we considered the heart as a separate organ category. Carbohydrates and lipids provide energy to muscles while amino acids help in forming muscle proteins, which explains why these metabolic pathway defects affect muscle function. The *liver* is the central metabolic unit of the human body accounting for majority of the biochemical transformations. A disturbance in any metabolic pathway will eventually affect the liver's function. Energy supply of the *heart* is mostly provided by fatty acids; however, glucose, lactate, ketone bodies, and amino acids are also used in different proportions.[55]

Biomarker classification. Enzyme assays, which test for enzyme activity, and identification of specific mutations, are frequently used as diagnostic tools for IEMs.[2] The diagnosis of IEMs experienced a tremendous rise with the advent of high-throughput detection techniques, such as mass spectrometry.[56] Metabolic biomarkers arise due to the presence of an IEM and are extensively used for diagnosis. For example, a disease either causes accumulation of a biomarker (*e.g.*, elevated triglycerides in Tangier disease, OMIM: 205400) or the decrease/absence of a specific metabolite (*e.g.*, carnitine in the carnitine uptake defect, OMIM: 212140). In fact, 17 different metabolites are used as biomarkers for the diagnosis of 130 of the 235 IEMs (Fig. S2, ESI‡). Metabolic biomarkers are, of course, acylcarnitines and amino acids, which also include amino acid derivatives, such as 5-oxoproline, 3-methoxytyrosine, and 5-hydroxytryptophan, but also include the other classes of compounds (see Table S2 (ESI‡) for a complete list). For example, *carbohydrates* account for a variety of compounds, most notably dolichol, oligosaccharides, galactitol, galactose, glucose, xylulose, arabitol, ribitol, glycogen, and glycan molecules. *Cholesterol* includes free and esterified cholesterol and its steroid derivatives (lipoproteins like HDL-cholesterol). *Fatty acids* include both straight and branched chain fatty acids as well as their derivatives, such as 2-ethyl-3-keto-hexanoic acid and 3-hydroxyisovaleric acid. The *lipids* class includes, for example, triglycerides, prostagladins (leukotrienes), and glycerol. The *organic acids* class represents a wide range of compounds, such as glycolic acid, glyoxylic acid, glyceric acid, and oxalic acid (see Table S2 (ESI‡) for the detailed classification scheme).

To assess the use of different metabolic biomarkers and enzyme assays employed for diagnosis of IEMs, we compared the reported biomarkers across the mapped IEMs. We found that enzyme assays are used for the diagnosis of 34% of the mapped IEMs (Fig. S2, ESI‡). Moreover, we found that the measurement of amino acid and acylcarnitine concentrations are primary laboratory tests, followed by tests for carbohydrates, organic acids, and fatty acids. In fact, acylcarnitine profiling is one of the emerging trends[57] due to the numerous roles of carnitine in the body.[58] Abnormal concentrations of acylcarnitines may indicate aminoacidopathies, fatty acid oxidation defects, and organic acid disorders. However, the absence of acylcarnitine abnormalities does not rule out disorders and needs to be confirmed by enzyme assays.[57] Moreover, all diagnostic criteria need confirmation for the presence of known mutations in the corresponding gene(s).[57,59,60] The DNA analysis may serve as sole diagnostic test for some IEMs, including infantile neuroaxonal dystrophy (OMIM: 256600) and leber congenital amaurosis type 3 (OMIM: 612712).

Types of therapies. Six different types of therapies are used to treat 158 of the 235 mapped IEMs (67%) (Table S6, ESI‡). The different therapeutic measures include dietary interventions, medications, organ transplantation, enzyme replacement therapy, and gene therapy. Enzyme replacement therapy is used to treat, for example, adenosine deaminase deficiency (OMIM: 102700), Fabry disease (OMIM: 301500), Pompe disease (OMIM: 232300), and mucopolysaccharidosis type VI/Maroteaux–Lamy syndrome (OMIM: 253200). Organ transplantation usually involves either bone marrow transplant or a specific organ, and is generally employed in cases of mucopolysaccharidosis type I (OMIM: 607014) and methylmalonic academia.[61] So far, enzyme replacement therapy has been promising, *e.g.*, for lysosomal storage disorders, while organ transplantation has only shown limited success due to graft rejection.[2] Additionally, gene therapy depends on the possible extent and duration of the gene expression without leading to toxicity.[2] On the other hand, dietary measures and specific medications have been preferred methods for the treatment of many IEMs. Therefore, we considered diet and medications as separate categories for IEMs, which are treated solely by these measures (see Table S2 (ESI‡) for a complete list).

Functional network view of the IEM compendium

The combined Recon1_AC/FAO metabolic network accounted for 3032 metabolites, 4095 reactions, and 1919 transcripts, corresponding to 1510 unique genes, with metabolites and reactions distributed over eight compartments (Fig. 2C). We investigated the 235 mapped IEMs from a functional network topology perspective. First, we determined the number of reactions associated with each IEM (Fig. 4A). Interestingly, six IEMs were associated with more than 50 reactions each in Recon1_AC/FAO (Table S2, ESI‡), including CPT-1 deficiency and Zellweger syndrome (both having 71 reactions associated with their respective defective genes). These IEMs with high reaction association represent hubs in the metabolic network, which are expected to have a great impact on the network functionality when perturbed. In fact, many of these IEMs are known to have a severe phenotype (Table S2, ESI‡). While 97 (41%) of the IEMs had only one reaction associated, we were interested in whether there were any IEMs, which had multiple associated reactions, that shared reactions and exhibited similar or overlapping phenotype. We identified 53 IEM pairs, consisting of 52 unique IEMs, which share at least one metabolic reaction in Recon1_AC/FAO. The reaction overlap, and potential similarities of disease phenotypes,

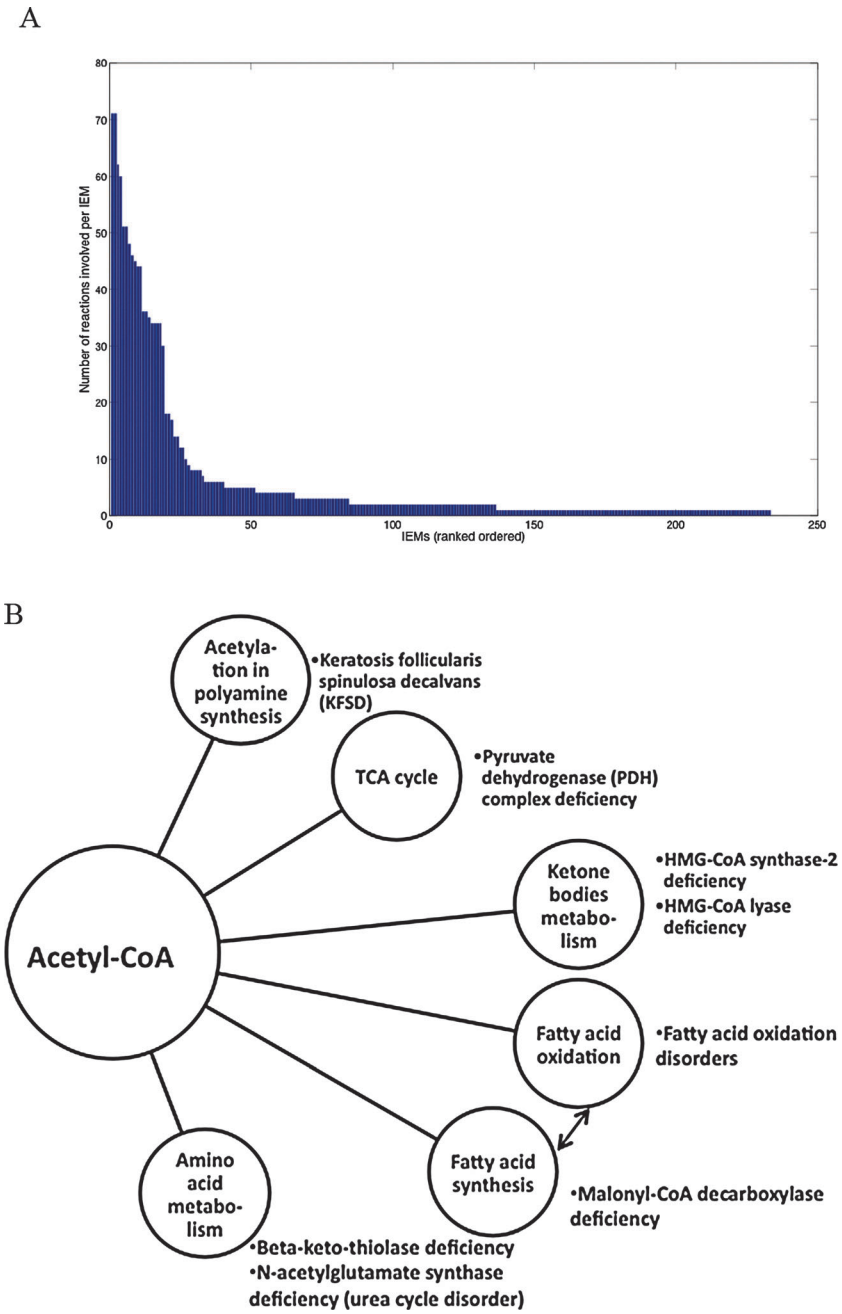


Fig. 4 Network properties of the IEM compendium. (A) The graphic shows the number of reactions affected per IEM (rank ordered). We used the gene-protein-reaction associations present in Recon1_AC/FAO to determine the number of reactions catalyzed by enzymes in Recon1_AC/FAO that is known to be deficient in IEMs (see also Table S2, ESI‡). (B) Many IEMs are connected in the metabolic reconstruction through shared metabolites. We use the highly connected, important central metabolite acetyl-CoA to illustrate the metabolic connectivity of the IEMs (shown adjacent to the respective metabolisms) within the human metabolic network. Malonyl-CoA, which is formed from acetyl-CoA *via* the action of acetyl-CoA carboxylase (E.C. 6.4.1.2), participates in fatty acid synthesis, and a potent inhibitor of CPT-I enzyme (E.C. 2.3.1.21). Malonyl-CoA levels are regulated by malonyl-CoA decarboxylase (E.C. 4.1.1.9). All together, these enzymes maintain fine-tuning between fatty acid breakdown and synthesis.

may also hamper the diagnosis of these IEMs. At the same time, this network driven view on IEMs permits us to correlate diseases that may not be associated with each other otherwise. We were particularly interested in those IEMs that had all reactions in common since these IEMs and knowledge about their phenotypes and biomarkers could be used in subsequent studies to assess the model's predictive capabilities and may also underline the importance of tissue- and cell-type specific reconstructions for metabolic modelling.

Furthermore, we identified seven reaction pairs in Recon1_AC/FAO that shared all reactions (Table 2). Interestingly, these IEM pairs also shared clinical features. For example, the Tay–Sachs disease (OMIM: 272800) and the Sandhoff disease (OMIM: 268800) are caused by a defect in the

beta-hexosaminidase (E.C. 3.2.1.52).[62] The beta-hexosaminidase catalyzes the glycosyl bond hydrolysis, thus removing amino sugars from the non-reducing ends of oligosaccharides.[63] Three forms of the enzyme have been described (A, B, and S) with distinct structures and subunit compositions.[64] The gene *HEXA* (Gene ID: 3073) encodes the alpha subunit, whereas the gene *HEXB* (Gene ID: 3074) encodes the beta subunit of the enzyme. Though different mutations in these two genes lead to Tay–Sachs disease and Sandhoff disease, these diseases are clinically indistinguishable (OMIM: 268800).[65] Nonetheless, a mouse model study has reported differences in phenotypes.[66] Another interesting case was the IEM pair of sialidosis (OMIM: 256550) and Morquio syndrome (OMIM: 253000), which are both defects in carbohydrate metabolism and exhibit similar neurologic abnormalities[67] (Table 2).

The network view of IEM compendium does not only allow us to determine reactions that are shared between IEMs but also assists in the systematic identification of IEMs, which are connected through common metabolites. This analysis was triggered by the observation that many IEMs were adjacent on the metabolic map (Fig. 5). Metabolic connectivities between IEMs may also highlight co-morbidities.[68] When excluding highly connected metabolites, such as adenosine triphosphate, water, and protons, we found that acetyl-CoA in particular acted as a common link between the metabolic pathways, which is a well-known fact in biochemistry. Acetyl-CoA is (i) used to acetylate various metabolites, including polyamines, (ii) generated as an end-product of leucine and isoleucine catabolism, (iii) a precursor for cholesterol and fatty acid synthesis, and (iv) required to form *N*-acetyl-glutamate, which is an allosteric activator of carbamoyl-phosphate-synthase-I (E.C. 6.3.4.16) catalyzing the first reaction of the urea cycle.[69] Under normal physiological conditions, acetyl-CoA enters the TCA cycle after being produced by beta fatty acid oxidation. However, under prolonged fasting conditions, acetyl-CoA enters ketone body synthesis, and they appear in urine.[69,70] Our network-based analysis further revealed that acetyl-CoA serves a hub linking different important IEMs (Fig. 4B), whereas the metabolic fate of acetyl-CoA varies depending on pathways and pathophysiological conditions.

Simulation of phenylketonuria with the metabolic model of Recon1_AC/FAO

Phenylketonuria (PKU, OMIM: 261600) is caused by mutations in *PAH* (GeneID: 5053) that may lead to complete or partial loss of the phenylalanine hydroxylase activity (E.C. 1.14.16.1). The enzyme catalyzes hydroxylation of phenylalanine to tyrosine (Table 1). We chose PKU to demonstrate how a model derived from Recon1_AC/FAO can be used to simulate enzymopathies and to predict potential changes in biofluid concentrations. Therefore, we deleted the model reactions associated with the *PAH* gene, as defined through the GPRs, to simulate the consequences of complete loss of enzyme activity. The model was allowed the uptake and secretion of all metabolites with defined exchange reactions. We then performed flux balance analysis,[71] where the exchange reaction for L-tyrosine or L-phenylalanine was used as an objective function. As expected, L-tyrosine secretion rate was zero in the PKU model, while L-tyrosine secretion was possible in the model corresponding to the healthy state. This result is consistent as PKU patients usually have low plasma tyrosine levels (OMIM: 261600), which is most often seen due to low enzyme activity. In contrast, no change in L-phenylalanine uptake or secretion was calculated in the PKU model compared to the healthy model. Thus, the model could not accurately represent the accumulation of phenylalanine (in form of phenylalanine secretion), which has been observed in plasma, urine, and cerebrospinal fluid

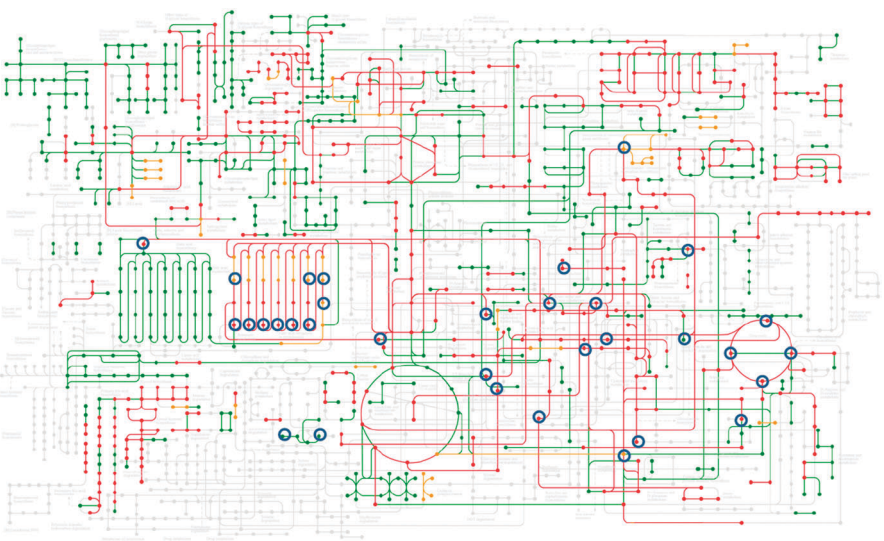


Fig. 5 Visualization of 375 IEMs in our IEM compendium on the human metabolic map. This metabolic map, obtained from Kegg,[104] gives an overview of current knowledge about human metabolism based on the human genome and known biochemistry. We map the genes from Recon1_AC/FAO (green), the genes underlying the 235 mapped IEMs (red), and the genes corresponding to the 139 uncharted IEMs (orange). Note that 245 Recon1_AC/FAO genes and 55 uncharted IEM genes could not be mapped. The blue circles represent the different amino acids and acylcarnitines, which are routinely measured in the newborn screening program (see Table S1 (ESI‡) for details). This map represents only a subset of the metabolic pathways captured in Recon1_AC/FAO; we could only map 14 acyl-CoAs and 17 amino acids.

Table 1 Properties of mapped IEMs explained using classical phenylketonuria (PKU, OMIM: 261600) as an example

Criteria	Definition	Characterization of PKU
Pathway classification	IEMs were classified according to their metabolic pathways	Amino acid metabolism
Affected genes	Mutations in the gene lead to (i) non-functional protein, (ii) mutated protein, or (iii) no protein synthesis. All these instances give rise to an IEM	Phenylalanine hydroxylase (*PAH*, GeneID: 5053, E.C. 1.14.16.1)
Affected reactions	Genes encode enzymes that catalyze reactions. IEMs are seen due to enzyme deficiencies, transporter, or co-factor defects; thereby, reactions are either blocked or follow abnormal pathways	O_2 + phenylalanine + tetrahydrobiopterin > tyrosine + dihydrobiopterin + H_2O (PHETHPTOX2 in Recon 1). Accumulated phenylalanine produces phenylacetate and phenyllactate as abnormal metabolites
Mode of inheritance	IEMs are inherited in a dominant or recessive pattern. Majority of the IEMs are autosomal recessive	Phenylketonuria follows an autosomal recessive pattern of inheritance
Phenotypes	A single metabolic defect may present various phenotypic patterns, which may range from two to seven different forms (see text and ESI)	Three major types: phenylketonuria (PKU), non-PKU hyperphenylalaninemia and variant PKU
Affected organs	The severity of the IEMs is assessed by their ability to affect organ systems. While some of the IEMs may be un-noticed, many of them have profound effects on different organ systems	Brain
Biomarkers	Biomarkers enable easy and early diagnosis of IEMs	In plasma: increased phenylalanine and phenylalanine/tyrosine ratio. In urine: increased excretion of 2-hydroxyphenylacetic acid, phenyl lactic acid, and phenyl pyruvic acid
Therapies available	IEMs, when diagnosed at a young age, are often treatable. Various therapeutic measures are available, including dietary interventions, medications, gene therapy, enzyme replacement, and organ transplantation	Phenylalanine restricted diet. In fact, most disorders of amino acid metabolism are controlled by dietary interventions

of PKU patients (OMIM: 261600). The accumulation is due to decreased utilization of phenylalanine by the phenylalanine hydroxylase, while there is a continuous supply of dietary phenylalanine. As we do not explicitly model the dietary intake and human cells cannot synthesize phenylalanine, we did not calculate increase in net phenylalanine secretion.

Conclusions

In this study, we generated an extended metabolic reconstruction, Recon1_AC/FAO, which accounts for all metabolic and transport reactions necessary to enable the mapping of biomarkers measured in newborn screening programs.[7,8] Furthermore, we compiled a comprehensive compendium of 235 IEMs that have all causal genes captured in Recon1_AC/FAO, containing also important phenotypic, genetic, and clinical features for each IEM.

A challenge that we encountered during the reconstruction process was the correct identification of some acylcarnitines, as MS/MS measurements do not always provide sufficient chemical details about the unsaturated fatty acids (*i.e.*, positional isomers). To enable the inclusion of more metabolites and corresponding reactions in the reconstruction, more precise measurements are needed.

The IEMs of the compendium were categorized based on different criteria to assess to which extent they affect whole body and the global metabolism. For instance, the organ classification will be of additional value when tissue- and cell type specific reconstructions and models are employed rather than a generic human metabolic network. A current challenge in computational modelling of metabolism is how one can generate in a (semi-) automated manner tissue- and cell-type specific reconstructions based on 'omics' data (*e.g.*, transcriptomics).[72] The organ classification of IEMs could be used to further refine tissue-specific metabolic reconstruction. Moreover, as the simulation of PKU highlighted, the use of a tissue-specific model rather than the generic model derived from Recon1_AC/FAO and more defined boundary constraints will further increase the predictive potential of the effects of IEMs on the overall metabolism and metabolite biomarkers.

In addition to the mapped 235 IEMs, we identified 139 'uncharted' IEMs (Table S3, ESI‡), which could not be mapped onto Recon1_AC/FAO due to missing genes. These IEMs were classified under metabolic and non-metabolic IEMs (*e.g.*, causing a defect in regulatory or signalling pathways), depending on the chief mechanism involved. The uncharted metabolic IEMs will be a good starting point for further extension and refinement of the human metabolic reconstruction. For instance, the mapping of inherited disorders of lipid metabolism will require further extension of the current human metabolic reconstruction as only 51 IEMs of the lipid metabolism could be mapped. Numerous new disorders have been identified[12] and many of them are fatal, therefore, this class of IEMs will be an important future extension of the human metabolic reconstruction. In contrast, the inclusion of the non-metabolic IEM will require the expansion of the metabolic reconstruction for further cellular processes, such as signalling,[73–75] macromolecular synthesis,[76–78] and transcriptional regulation.[76,79,80]

The presented extension to the current human metabolic reconstruction is not only important and relevant to map and analyze newborn screening data but also to map large-scale metabolomic data sets of plasma and urine from defined patient groups. For instance, serum metabolite concentration data for more than 1000 participants have been recently published, along with their genetic variants, from the KORA population.[81] Thanks to our extension, a majority of these metabolite measurements could be mapped and analyzed within the metabolic network context.

Published on 02 May 2012 on http://pubs.rsc.org | doi:10.1039/C2MB25075F

Table 2 Metabolically linked IEMs with known overlapping phenotypes. The IEM pairs are grouped together

IEM pairs	Genes involved	Pathway involved	Enzyme (E.C. number)	No. of shared reactions	Remarks
Tay–Sachs disease	*HEXA* (GeneID: 3073)	Carbohydrate metabolism	Beta-*N*-acetylhexosaminidase (E.C. 3.2.1.52)	51	*HEXA* gene encodes the alpha subunit, whereas *HEXB* encodes the beta subunit of the enzyme. Even though different mutations in these two genes lead to Tay–Sachs disease and Sandhoff disease, these diseases are clinically indistinguishable[65]
Sandhoff disease	*HEXB* (GeneID: 3074)	Carbohydrate metabolism	Beta-*N*-acetylhexosaminidase (E.C. 3.2.1.52)	51	
Sialidosis	*NEU1* (GeneID: 4758)	Carbohydrate metabolism	Sialidase (3.2.1.18)	34	Common clinical features in both, *i.e.*, coarse features and neurologic abnormalities. Additionally, the structural changes in Morquio syndrome may lead to nervous symptoms, which are also seen in sialidosis[67]
Mucopolysaccharidosis type IV type A/Morquio syndrome	*GALNS* (GeneID: 2588)	Carbohydrate metabolism	Chondroitinase (E.C. 3.1.6.4)	34	
NICCD (Neonatal intrahepatic cholestasis caused by citrin deficiency)	*SLC25A13* (GeneID: 10165)	Transport	Calcium-binding mitochondrial carrier protein Aralar2 deficiency	3	Same gene is affected in NICCD and citrullinemia type 2 Citrullinemia type 2 may develop after several years of being asymptomatic in NICCD patients[91,92]
Type II citrullinemia	*SLC25A13* (GeneID: 10165)	Amino acid metabolism	Calcium-binding mitochondrial carrier protein Aralar2 deficiency	3	
Erythrocyte AMP deaminase deficiency	*AMPD3* (GeneID: 272)	Nucleotide metabolism	AMP deaminase (3.5.4.6)	1	The enzyme is encoded by different isoforms, which depends on tissue specific expression. Low plasma uric acid level but no apparent clinical signs are observed in cases of Erythrocyte AMP deaminase deficiency.[93] The defect in the muscle form is one of the most common types of muscle enzyme defects. It exhibits myalgia[94] and is linked to obesity and diabetes[95]
Myoadenylate deaminase deficiency	*AMPD1* (GeneID: 270)	Nucleotide metabolism	AMP deaminase (3.5.4.6)	1	
Liver glycogen synthase deficiency	*GYS2* (GeneID: 2998)	Carbohydrate metabolism	Glycogen (starch) synthase (E.C. 2.4.1.11)	2	Glycogen synthase has two isozymes (liver and muscle isoform). Liver glycogen synthase deficiency is comparatively rare and associated with hypoglycemia, mental retardation and seizures.[96–98] The muscle form is associated with cardiomyopathy and exercise intolerance[99]
Muscle glycogen synthase deficiency	*GYS1* (GeneID: 2997)	Carbohydrate metabolism	Glycogen (starch) synthase (E.C. 2.4.1.11)	2	
Distal renal tubular acidosis type 1	*SLC4A1* (GeneID: 6521)	Transport	Band 3 anion transport protein	1	Band 3 protein is a multi-spanning membrane protein in the red blood cell.[100] Genetic mutations in the SLC4A1 gene result in two different types of defect, *i.e.* hereditary spherocytosis and distal renal tubular acidosis type 1. These diseases may co-occur in an affected individual.[101] Specific reasons for this co-occurrence is still not known but very frequently reported[101,102]
Hereditary spherocytosis-4	*SLC4A1* (GeneID: 6521)	Transport	Band 3 anion transport protein	1	

Table 2 (*continued*)

IEM pairs	Genes involved	Pathway involved	Enzyme (E.C. number)	No. of shared reactions	Remarks
Renal tubular acidosis, distal, autosomal recessive	*ATP6V0A4* (GeneID: 50617)	Transport	V-type proton ATPase 116 kDa subunit a isoform 4	1	The affected genes encode for different subunits of the V-ATPase pump. This pump mediates intracellular acidification and consists of a V0 (A, B, C, D, E, F, G and H subunits) and a V1 domain (a, c, c′, c″, and d subunits). The ATP6V0A4 gene encodes the a-subunit of the V1 domain, whereas the ATP6V1B1 gene encodes B-subunit of the V0 domain (Gene ID 50617 and 525). Similar clinical features are observed in renal tubular acidosis (distal, autosomal recessive) and RTA
Renal tubular acidosis (RTA)	*ATP6V1B1* (GeneID: 525)	Transport	V-type proton ATPase subunit B, kidney isoform	1	

Metabolism has been found to be abnormal in many diseases. In this study, we identified potential IEM hubs in the metabolic network, which are connected by inter-linking metabolites. The use of comprehensive mathematical metabolic models for the analysis of IEMs is promising, especially, for those IEMs, which have currently no diagnosis. Moreover, computational modelling may be particularly valuable for those IEMs that are very rare and have currently no promising diagnostic and therapeutic regime.

Materials and methods

Routinely measured biomarkers in newborn screening data

We used information obtained from the Landspítali (Icelandic National University Hospital) for routinely measured biomarkers, which include 18 standard and four non-standard amino acids (*i.e.*, argininosuccinate, methylhistidine, ornithine and citrulline), 36 acylcarnitines, and 23 ratios (of amino acids and acylcarnitines) (see Table S1 (ESI‡) for a complete list). Furthermore, we considered three additional acylcarnitines (C14:1-OH, C16:2-OH, C14:2-OH) to be mapped onto the human reconstruction. Note that measured methylhistidine concentrations cannot be mapped onto the human reconstruction. We retrieved the molecular formulae and chemical structure for each metabolite from the literature and/or databases.

Human metabolic reconstruction

The published human metabolic reconstruction, Recon 1,[16] was obtained from the BiGG database[82] in SBML format[83] and loaded into our reconstruction tool, rBioNet,[84] using the COBRA toolbox.[85] Recon 1 consists of 3743 reactions, 2766 metabolites, and 1905 transcripts (corresponding to 1495 unique genes). Genes are connected to their respective reactions based on Boolean logic by defining gene–protein-reaction (GPR) associations for each network reaction. AND signifies that two gene products are required to carry out one reaction (enzyme complex), while OR stands for either gene product is sufficient for the reaction (isozyme).

AC/FAO reconstruction module

The reconstruction was done using an established protocol[14] and the rBioNet reconstruction extension[84] for the COBRA toolbox.[85] Besides information retrieved from more than 150 primary publications and books, we used numerous databases (Table S7, ESI‡). Reaction directionality information was obtained from the literature and/or from thermodynamic considerations.[86–88] We employed the COBRA toolbox to combine the AC/FAO module with Recon 1, and named the resulting expanded reconstruction Recon1_AC/FAO. The content of the AC/FAO module can be found in Table S9, ESI.‡

IEM collection and analysis

Candidate IEMs were mapped onto Recon1_AC/FAO using GPR associations in Recon1_AC/FAO and reported affected gene(s) of an IEM. Phenotypic and clinical information for each IEM was retrieved from primary literature, books, and databases (Table S8, ESI‡). A complete list of mapped and

future IEMs along with their characteristics can be found in Tables S2 and S3 (ESI‡), respectively.

Network based analysis of IEMs

Using the GPR associations defined in Recon1_AC/FAO, we mapped all 235 IEMs onto the network. The reconstruction was converted into a mathematical model using the COBRA toolbox. The resulting stoichiometric matrix, S, has m rows, one for each metabolite, and n columns, one for each reaction. If a metabolite i participates in a reaction j, then the cell $S(i,j)$ has a non-zero entry, where the number corresponds to the stoichiometric coefficient of the corresponding reaction j. By definition, substrates have a negative coefficient, while products have a positive coefficient. Similarly, the model of Recon1_AC/FAO contains a matrix describing the links between genes and reactions, deemed G. Using G, we mapped the IEMs from genes to reactions, resulting in a new matrix M, in which each row corresponds to an IEM and each column corresponds to a reaction. Using a binary version of M (M_{bin}), we multiplied M_{bin} with its transpose. Each diagonal element of this adjacency matrix of M_{bin} corresponds to the number of reactions that are connected with each IEM. The off-diagonal elements corresponding to the number of reactions shared between two IEMs, which allowed us to identify reaction pairs. Similarly, by multiplying the M_{bin} with the transpose of the binary version of S, S_{bin}, we could determine metabolites shared between two IEMs. For the visualization of all IEMs, we employed the pathway viewer from the KEGG database.[89]

Simulation of PKU

Recon1_AC/FAO was converted into a computational model using the COBRA toolbox.[85] The reaction list of the reconstruction was therefore converted into a so-called stoichiometric matrix, S, consisting of m rows (one row for each metabolite) and n columns (one for each reaction in the network). If a metabolite i participates in a reaction j, then the entry $S_{i,j}$ contains the corresponding non-zero stoichiometric coefficient, which is a negative number for substrates and a positive number for reaction products. Constraint-based modelling assumes that the modelled system is in a quasi-steady state, *i.e.*, the change of concentration of a metabolite i is zero over time t: $\mathrm{d}i/\mathrm{d}t = S \cdot v \equiv 0$ where v is the flux vector containing a flux value v_i for each reaction i. In flux balance analysis, we generally optimize for an objective reaction, thus identifying one possible flux vector with optimal value in the objective function that is consistent with the applied constraints.[71,90] Constraints can be applied to any reaction i in the model, such that $v_{i,\min} \leq v_i \leq v_{i,\max}$. To simulate PKU, we set the constraints of the reaction PHETHPTOX2, catalyzed by *PAH*, to $v_{\mathrm{PHETHPTOX2,min}} = v_{\mathrm{PHETHPTOX2,max}} = 0$ mmol g_{DW}^{-1} h^{-1}, while the constraints through this reaction in the healthy model were unchanged. Additionally, for the healthy and PKU models, we set the lower bounds ($v_{i,\min}$) of the exchange reactions for water, carbon dioxide, proton, bicarbonate, ammonium, inorganic phosphate and sulphate to -100 mmol g_{DW}^{-1} h^{-1}, and the lower bound for oxygen uptake was set to be -40 mmol g_{DW}^{-1} h^{-1}. The lower bounds of the remaining exchange reactions were set to be -1 mmol g_{DW}^{-1} h^{-1}. The upper bounds of all exchange reactions were unconstrained, *i.e.*, $v_{i,\max} = 1000$ mmol g_{DW}^{-1} h^{-1}. In both models, we selected the exchange reaction for L-tyrosine and L-phenylalanine separately as objective reactions for maximisation and minimisation. The calculations were carried out using Matlab (Mathworks, Inc.) as programming environment and TomLab (TomOpt, Inc.) as linear programming solver.

Acknowledgements

This work was supported by the Icelandic Research Fund (No. 100406022) and by an ERC grant agreement (No. 232816 – SYSTEM Us). IT was supported, in part, by a Marie Curie International Reintegration Grant (No. 249261) within the 7th European Community Framework Program. The authors are thankful for valuable discussions with Ms H. Haraldsdottir, and Dr R. M. T. Fleming and Dr O. Rolfsson.

Notes and references

1 T. Pampols, *Adv. Exp. Med. Biol.*, 2010, **686**, 397–431.
2 J. Fernandes, *Inborn metabolic diseases: diagnosis and treatment*, Springer, Heidelberg, 4th, rev. edn, 2006.
3 M. Mamas, W. Dunn and L. Neyses, *et al.*, *Arch. Toxicol.*, 2011, **85**, 5–17.
4 J. V. Leonard and C. Dezateux, *Paediatr. Child Health*, 2011, **21**, 56–60.
5 M. R. B. Edward, *Mol. Genet. Metab.*, 2010, **100**, 1–5.
6 D. H. Chace and T. A. Kalas, *Clin. Biochem.*, 2005, **38**, 296–309.
7 D. M. McHugh, C. A. Cameron and J. E. Abdenur, *et al.*, *Genet. Med.*, 2011, **13**, 230–254.
8 R. R. Howell, *Genet. Med.*, 2011, **13**, 205.
9 H. L. Levy, *Genet. Med.*, 2010, **12**, S213–S214, 210.1097/GIM.1090b1013e3181fe1095d1077.
10 K. R. Chalcraft and P. Britz-McKibbin, *Anal. Chem.*, 2009, **81**, 307–314.
11 M. Lindner, G. F. Hoffmann and D. Matern, *J. Inherited Metab. Dis.*, 2010, **33**, 521–526.
12 R. S. Rector and J. A. Ibdah, *Semin. Fetal Neonat. Med.*, 2010, **15**, 122–128.
13 B. O. Palsson, *Systems Biology – Properties of Reconstructed Networks*, Cambridge University Press, 2006.
14 I. Thiele and B. O. Palsson, *Nat. Protoc.*, 2010, **5**, 93–121.
15 M. A. Oberhardt, B. O. Palsson and J. A. Papin, *Mol. Syst. Biol.*, 2009, **5**, 320.
16 N. C. Duarte, S. A. Becker and N. Jamshidi, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 1777–1782.
17 I. Thiele and *et al.*, submitted; I. Thiele, unpublished data.
18 V. Valayannopoulos, C. Haudry and V. Serre, *et al.*, *Mitochondrion*, 2010, **10**, 335–341.
19 T. D. Horvath, S. L. Stratton and A. Bogusiewicz, *et al.*, *Anal. Chem.*, 2010, **82**, 4140–4144.
20 F. T. Eminoglu, A. A. Ozcelik and I. Okur, *et al.*, *J. Child Neurol.*, 2009, **24**, 478–481.
21 M. Popek, M. Walter and M. Fernando, *et al.*, *Clin. Chim. Acta*, 2010, **411**, 2087–2091.
22 S. W. Sauer, S. Opp and G. F. Hoffmann, *et al.*, *Brain*, 2011, **134**, 157–170.
23 M. D. Nenad Blau and K. Michael Gibson, *Laboratory guide to the methods in biochemical genetics*, Springer, 2008.
24 P. M. Jones, R. Quinn and P. V. Fennessey, *et al.*, *Clin. Chem.*, 2000, **46**, 149–155.
25 T. Hori, T. Fukao and H. Kobayashi, *et al.*, *Tohoku J. Exp. Med.*, 2010, **221**, 191–195.
26 M. Fontaine, G. Briand and N. Ser, *et al.*, *Clin. Chim. Acta*, 1996, **255**, 67–83.
27 C. Castelnovi, K. Moseley and S. Yano, *Clin. Chim. Acta*, 2010, **411**, 2101–2103.

28 S. Forni, X. Fu and S. E. Palmer, *et al.*, *Mol. Genet. Metab.*, 2010, **101**, 25–32.
29 R. R. Ramsay, R. D. Gandour and F. R. van der Leij, *Biochim. Biophys. Acta*, 2001, **1546**, 21–43.
30 R. J. Wanders, W. F. Visser and C. W. van Roermund, *et al.*, *Pflugers Arch.*, 2007, **453**, 719–734.
31 R. J. Wanders, J. P. Ruiter and L. IJLst, *et al.*, *J. Inherited Metab. Dis.*, 2010, **33**, 479–494.
32 P. A. Watkins, D. Maiguel and Z. Jia, *et al.*, *J. Lipid Res.*, 2007, **48**, 2736–2750.
33 C. Vock, K. Biedasek and I. Boomgaarden, *et al.*, *Cell. Physiol. Biochem.*, 2010, **25**, 675–686.
34 R. J. Wanders, E. G. van Grunsven and G. A. Jansen, *Biochem. Soc. Trans.*, 2000, **28**, 141–149.
35 R. J. Wanders, S. Ferdinandusse and P. Brites, *et al.*, *Biochim. Biophys. Acta*, 2010, **1801**, 272–280.
36 B. V. Geisbrecht, D. Zhang and H. Schulz, *et al.*, *J. Biol. Chem.*, 1999, **274**, 21797–21803.
37 J. M. Street, H. Singh and A. Poulos, *Biochem. J.*, 1990, **269**, 671–677.
38 A. K. Heinzer, S. Kemp and J. F. Lu, *et al.*, *J. Biol. Chem.*, 2002, **277**, 28765–28773.
39 J. E. Pettersen, *Biochim. Biophys. Acta*, 1973, **306**, 1–14.
40 R. J. Wanders and J. M. Tager, *Mol. Aspects Med.*, 1998, **19**, 69–154.
41 S. Ferdinandusse, S. Denis and C. W. Van Roermund, *et al.*, *J. Lipid Res.*, 2004, **45**, 1104–1111.
42 R. J. Sanders, R. Ofman and F. Valianpour, *et al.*, *J. Lipid Res.*, 2005, **46**, 1001–1008.
43 M. Fer, L. Corcos and Y. Dreano, *et al.*, *J. Lipid Res.*, 2008, **49**, 2379–2389.
44 O. Rolfsson, B. O. Palsson and I. Thiele, *BMC Syst. Biol.*, 2011, **5**, 155.
45 I. Thiele and B. O. Palsson, *Mol. Syst. Biol.*, 2010, **6**, 361.
46 H. H. Freeze and V. Sharma, *Semin. Cell Dev. Biol.*, 2010, **21**, 655–662.
47 M. T. Scheuner, P. W. Yoon and M. J. Khoury, *Am. J. Med. Genet., Part C*, 2004, **125**, 50–65.
48 J. H. Nadeau, *Nat. Rev.*, 2001, **2**, 165–174.
49 D. J. Weatherall, *Nat. Rev.*, 2001, **2**, 245–255.
50 A. I. Dagli and D. A. Weinstein, in *GeneReviews [Internet]*, ed. R. A. Pagon, T. D. Bird, C. R. Dolan and K. Stephens, University of Washington, Seattle, Seattle (WA), 2010/03/20 edn, 1993.
51 M. R. Seashore, in *GeneReviews [Internet]*, ed. R. A. Pagon, T. D. Bird, C. R. Dolan and K. Stephens, University of Washington, Seattle, Seattle (WA), 2010/03/20 edn, 1993.
52 G. J. Siegel, W. R. Albers and S. T. Brady, et al., *Basic neurochemistry: molecular, cellular and medical aspects*, Elsevier Academic Press, London, UK, 7th edn, 2006.
53 S. R. Felber, W. Sperl and A. Chemelli, *et al.*, *Ann. Neurol.*, 1993, **33**, 396–401.
54 W. Jan, R. A. Zimmerman and Z. J. Wang, *et al.*, *Neuroradiology*, 2003, **45**, 393–399.
55 R. M. Beadle and M. Frenneaux, *Heart*, 2010, **96**, 824–830.
56 M. Piraud, S. Ruet and S. Boyer, *et al.*, *Methods Mol. Biol.*, 2011, **708**, 25–53.
57 S. Santra and C. Hendriksz, *Arch. Dis. Child Educ. Pract. Ed.*, 2010, **95**, 151–156.
58 L. L. Jones, D. A. McDonald and P. R. Borum, *Prog. Lipid Res.*, 2010, **49**, 61–75.
59 B. Wilcken and V. Wiley, *Pathology*, 2008, **40**, 104–115.
60 A. Boneh, B. S. Andresen and N. Gregersen, *et al.*, *Mol. Genet. Metab.*, 2006, **88**, 166–170.
61 J. V. Leonard, *J. Inherited Metab. Dis.*, 1995, **18**, 430–434.
62 R. L. Proia, A. d'Azzo and E. F. Neufeld, *J. Biol. Chem.*, 1984, **259**, 3350–3354.
63 B. L. Mark, D. J. Mahuran and M. M. Cherney, *et al.*, *J. Mol. Biol.*, 2003, **327**, 1093–1109.
64 F. Norflus, S. Yamanaka and R. L. Proia, *DNA Cell Biol.*, 1996, **15**, 89–97.
65 K. J. Branda, J. Tomczak and M. R. Natowicz, *Genet. Test*, 2004, **8**, 174–180.
66 K. Sango, S. Yamanaka and A. Hoffmann, *et al.*, *Nat. Genet.*, 1995, **11**, 170–176.
67 K. L. Becker, *Principles and practice of endocrinology and metabolism*, Lippincott Williams and Wilkins, Third edn, 2001.
68 D. S. Lee, J. Park and K. A. Kay, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 9880–9885.
69 R. K. Murray, D. A. Bender and K. M. Botham, et al., *Harper's illustrated biochemistry*, Mc Graw Hill, 28th edn, 2009.
70 M. Murray and M. L. Greenberg, *Mol. Microbiol.*, 2000, **36**, 651–661.
71 J. D. Orth, I. Thiele and B. O. Palsson, *Nat. Biotechnol.*, 2010, **28**, 245–248.
72 T. Shlomi, *Biotechnol. Genet. Eng. Rev.*, 2010, **26**, 281–296.
73 F. Li, I. Thiele and N. Jamshidi, *et al.*, *PLoS Comput. Biol.*, 2009, **5**, e1000292.
74 J. A. Papin and B. O. Palsson, *Biophys. J.*, 2004, **87**, 37–46.
75 M. S. Dasika, A. Burgard and C. D. Maranas, *Biophys. J.*, 2006, **91**, 382–398.
76 A. M. Feist, M. J. Herrgard and I. Thiele, *et al.*, *Nat. Rev.*, 2009, **7**, 129–143.
77 I. Thiele, R. M. Fleming and A. Bordbar, *et al.*, *Biophys. J.*, 2010, **98**, 2072–2081.
78 I. Thiele, N. Jamshidi and R. M. Fleming, *et al.*, *PLoS Comput. Biol.*, 2009, **5**, e1000312.
79 E. P. Gianchandani, A. R. Joyce and B. O. Palsson, *et al.*, *PLoS Comput. Biol.*, 2009, **5**, e1000403.
80 E. P. Gianchandani, J. A. Papin and N. D. Price, *et al.*, *PLoS Comput. Biol.*, 2006, **2**, e101.
81 T. Illig, C. Gieger and G. Zhai, *et al.*, *Nat. Genet.*, 2010, **42**, 137–141.
82 J. Schellenberger, J. O. Park and T. M. Conrad, *et al.*, *BMC Bioinformatics*, 2010, **11**, 213.
83 M. Hucka, A. Finney and H. M. Sauro, *et al.*, *Bioinformatics (Oxford, England)*, 2003, **19**, 524–531.
84 S. G. Thorleifsson and I. Thiele, *Bioinformatics (Oxford, England)*, 2011, **27**, 2009–2010.
85 J. Schellenberger, R. Que and R. M. Fleming, *et al.*, *Nat. Protoc.*, 2011, **6**, 1290–1307.
86 R. M. Fleming and I. Thiele, *Bioinformatics (Oxford, England)*, 2011, **27**, 142–143.
87 R. M. Fleming, I. Thiele and H. P. Nasheuer, *Biophys. Chem.*, 2009, **145**, 47–56.
88 H. S. Haraldsdóttir, I. Thiele and R. M. T. Fleming, *Biophys. J.*, 2012, **102**, 1703–1711.
89 S. Okuda, T. Yamada and M. Hamajima, *et al.*, *Nucleic Acids Res.*, 2008, **36**, W423–W426.
90 N. D. Price, J. L. Reed and B. O. Palsson, *Nat. Rev.*, 2004, **2**, 886–897.
91 K. Kobayashi and T. Saheki, *Gene Reviews*, 1993–2011, University of Washington, Seattle, Seattle, WA, 2010/03/20 edn, 1993.
92 T. Saheki, K. Kobayashi and M. Iijima, *et al.*, *Mol. Genet. Metab.*, 2004, **81**, 20–26.
93 D. M. Gross, in *Orphanet encyclopedia*, ed. P. G. V. D. Berghe, 2001.
94 M. Gross, *J. Inherited Metab. Dis.*, 1997, **20**, 186–192.
95 K. Safranow, J. Suchy and K. Jakubowska, *et al.*, *J. Appl. Genet.*, 2011, **52**, 67–76.
96 W. A. Walker, *Pediatric gastrointestinal disease: pathophysiology, diagnosis, management*, 2004.
97 R. Gitzelmann, M. A. Spycher and G. Feil, *et al.*, *Eur. J. Pediatr.*, 1996, **155**, 561–567.
98 D. A. Weinstein, C. E. Correia and A. C. Saunders, *et al.*, *Mol. Genet. Metab.*, 2006, **87**, 284–288.
99 G. Kollberg, M. Tulinius and T. Gilljam, *et al.*, *N. Engl. J. Med.*, 2007, **357**, 1507–1514.
100 E. van den Akker, T. J. Satchwell and R. C. Williamson, *et al.*, *Blood Cells, Mol., Dis.*, 2010, **45**, 1–8.
101 O. Wrong, L. J. Bruce and R. J. Unwin, *et al.*, *Kidney Int.*, 2002, **62**, 10–19.
102 M. L. C. Ribeiro, N. Alloisio and H. Almeida, *et al.*, *Blood*, 2000, **96**, 1602–1604.
103 N. C. Duarte, S. A. Becker and N. Jamshidi, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 1777–1782.
104 M. Kanehisa, M. Araki and S. Goto, *et al.*, *Nucleic Acids Res.*, 2008, **36**, D480–D484.

