

# A community-driven global reconstruction of human metabolism

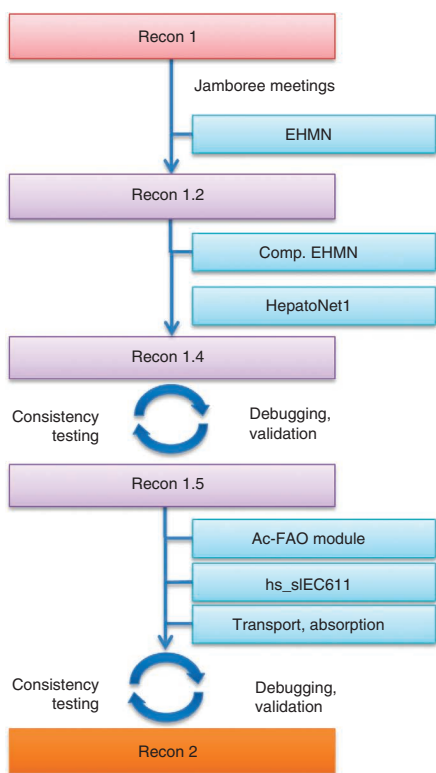
Ines Thiele<sup>1,2,37</sup>, Neil Swainston<sup>3,4,37</sup>, Ronan M T Fleming<sup>1,5</sup>, Andreas Hoppe<sup>6</sup>, Swagatika Sahoo<sup>1</sup>, Maike K Aurich<sup>1</sup>, Hulda Haraldsdottir<sup>1</sup>, Monica L Mo<sup>7</sup>, Ottar Rolfsson<sup>1</sup>, Miranda D Stobbe<sup>8,9</sup>, Stefan G Thorleifsson<sup>1</sup>, Rasmus Agren<sup>10</sup>, Christian Bölling<sup>6</sup>, Sergio Bordel<sup>10</sup>, Arvind K Chavali<sup>11</sup>, Paul Dobson<sup>12</sup>, Warwick B Dunn<sup>3,13</sup>, Lukas Endler<sup>14</sup>, David Hala<sup>15</sup>, Michael Hucka<sup>16</sup>, Duncan Hull<sup>4</sup>, Daniel Jameson<sup>3,4</sup>, Neema Jamshidi<sup>7</sup>, Jon J Jonsson<sup>5</sup>, Nick Juty<sup>17</sup>, Sarah Keating<sup>17</sup>, Intawat Nookaew<sup>10</sup>, Nicolas Le Novère<sup>17,18</sup>, Naglis Malys<sup>3,19,20</sup>, Alexander Mazein<sup>21</sup>, Jason A Papin<sup>11</sup>, Nathan D Price<sup>22</sup>, Evgeni Selkov, Sr<sup>23</sup>, Martin I Sigurdsson<sup>1</sup>, Evangelos Simeonidis<sup>22,24</sup>, Nikolaus Sonnenschein<sup>25</sup>, Kieran Smallbone<sup>3,26</sup>, Anatoly Sorokin<sup>21,27</sup>, Johannes H G M van Beek<sup>28–30</sup>, Dieter Weichart<sup>3,31</sup>, Igor Goryanin<sup>21,32</sup>, Jens Nielsen<sup>10</sup>, Hans V Westerhoff<sup>3,28,33,34</sup>, Douglas B Kell<sup>3,35</sup>, Pedro Mendes<sup>3,4,36</sup> & Bernhard Ø Palsson<sup>1,7</sup>

Multiple models of human metabolism have been reconstructed, but each represents only a subset of our knowledge. Here we describe Recon 2, a community-driven, consensus ‘metabolic reconstruction’, which is the most comprehensive representation of human metabolism that is applicable to computational modeling. Compared with its predecessors, the reconstruction has improved topological and functional features, including ~2× more reactions and ~1.7× more unique metabolites. Using Recon 2 we predicted changes in metabolite biomarkers for 49 inborn errors of metabolism with 77% accuracy when compared to experimental data. Mapping metabolomic data and drug information onto Recon 2 demonstrates its potential for integrating and analyzing diverse data types. Using protein expression data, we automatically generated a compendium of 65 cell type-specific models, providing a basis for manual curation or investigation of cell-specific metabolic properties. Recon 2 will facilitate many future biomedical studies and is freely available at <http://humanmetabolism.org/>.

An understanding of metabolism is fundamental to comprehending the phenotypic behavior of all living organisms, including humans, where metabolism is integral to health and is involved in much of human disease. High quality, genome-scale ‘metabolic reconstructions’ are at the heart of bottom-up systems biology analyses and represent the entire network of metabolic reactions that a given organism is known to exhibit<sup>1</sup>. The metabolic-network reconstruction procedure

is now well-established<sup>2</sup> and has been applied to a growing number of model organisms<sup>3</sup>. Metabolic reconstructions allow for the conversion of biological knowledge into a mathematical format and the subsequent computation of physiological states<sup>1,4,5</sup> to address a variety of scientific and applied questions<sup>3,6</sup>. Reconstructions enable network-wide mechanistic investigations of the genotype-phenotype relationship. A high-quality reconstruction of the metabolic network is thus

<sup>1</sup>Center for Systems Biology, University of Iceland, Reykjavik, Iceland. <sup>2</sup>Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Reykjavik, Iceland. <sup>3</sup>Manchester Centre for Integrative Systems Biology, University of Manchester, Manchester Institute of Biotechnology, Manchester, UK. <sup>4</sup>School of Computer Science, University of Manchester, Manchester, UK. <sup>5</sup>Department of Biochemistry and Molecular Biology, University of Iceland, Reykjavik, Iceland. <sup>6</sup>Computational Systems Biochemistry Group, Charité– Universitätsmedizin Berlin, Berlin, Germany. <sup>7</sup>Department of Bioengineering, University of California, San Diego, La Jolla, California, USA. <sup>8</sup>Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands. <sup>9</sup>Netherlands Bioinformatics Centre, Nijmegen, the Netherlands. <sup>10</sup>Department of Chemical and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden. <sup>11</sup>Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia, USA. <sup>12</sup>Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, UK. <sup>13</sup>Centre for Advanced Discovery and Experimental Therapeutics (CADET), Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, UK. <sup>14</sup>Institute for Theoretical Chemistry, University of Vienna, Vienna, Austria. <sup>15</sup>Department of Biology, University of North Texas, Denton, Texas, USA. <sup>16</sup>Computing and Mathematical Sciences Department, California Institute of Technology, Pasadena, California, USA. <sup>17</sup>European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK. <sup>18</sup>Babraham Institute, Babraham Research Campus, Cambridge, UK. <sup>19</sup>Faculty of Life Sciences, University of Manchester, Manchester, UK. <sup>20</sup>School of Life Sciences, Gibbet Hill Campus, University of Warwick, Coventry, UK. <sup>21</sup>School of Informatics, University of Edinburgh, Edinburgh, UK. <sup>22</sup>Institute for Systems Biology, Seattle, Washington, USA. <sup>23</sup>Genome Designs, Inc., Walnut Creek, California, USA. <sup>24</sup>Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Campus Belval, Esch-sur-Alzette, Luxembourg. <sup>25</sup>School of Engineering and Science, Jacobs University Bremen, Bremen, Germany. <sup>26</sup>School of Mathematics, University of Manchester, Manchester, UK. <sup>27</sup>Institute of Cell Biophysics, Russian Academy of Sciences, Moscow region, Pushchino, Russia. <sup>28</sup>Department of Molecular Cell Physiology, Vrije Universiteit, Amsterdam, the Netherlands. <sup>29</sup>Section Medical Genomics, Department of Clinical Genetics, Vrije Universiteit University Medical Centre, Amsterdam, the Netherlands. <sup>30</sup>Netherlands Consortium for Systems Biology, Amsterdam, The Netherlands. <sup>31</sup>School of Dentistry, The University of Manchester, Manchester, UK. <sup>32</sup>Okinawa Institute Science and Technology, Okinawa, Japan. <sup>33</sup>School of Chemical Engineering and Analytical Science, University of Manchester, Manchester, UK. <sup>34</sup>Swammerdam Institute for Life Sciences, Faculty of Science, University of Amsterdam, Amsterdam, The Netherlands. <sup>35</sup>School of Chemistry, The University of Manchester, Manchester, UK. <sup>36</sup>Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia, USA. <sup>37</sup>These authors contributed equally to this work. Correspondence should be addressed to I.T. ([ines.thiele@gmail.com](mailto:ines.thiele@gmail.com)).



**Figure 1** Overview of the community-driven reconstruction approach to assemble Recon 2. Comp. EHMN, compartmentalized EHMN; and hs\_sIEC611, small intestinal enterocyte reconstruction.

of interest to the community of researchers focused on the systems biology of metabolism of a target organism.

Of the reconstructions of human metabolism that have appeared to date, perhaps the most widely used is Recon 1 (ref. 7), which represents a knowledgebase and has also been converted into many predictive models. These models have been used for various biomedical applications, including the prediction of biomarkers for inborn errors of metabolism (IEMs)<sup>8</sup>, cancer drug targets<sup>9,10</sup> and off-target drug effects<sup>11</sup>. Moreover, they have been used to evaluate missing metabolic functions systematically<sup>12,13</sup> and to model host-microbe interactions<sup>14,15</sup>. These studies demonstrated the potential of metabolic modeling to advance understanding of human metabolism in health and disease.

Various reconstructions of the human metabolic network exist, with only partially overlapping content (ref. 16 and **Supplementary Note 1**). In addition to Recon 1, a global human metabolic reconstruction, EHMN (Edinburgh Human Metabolic Network)<sup>17</sup>, has also been published. Manually curated cell type-specific reconstructions are also available, including the comprehensive reconstruction of human hepatocytes, HepatoNet1 (ref. 18), a small intestinal enterocyte reconstruction<sup>19</sup> and other metabolic models for macrophages<sup>14</sup>, hepatocytes<sup>20</sup> and kidney cells<sup>11</sup>. Moreover, a module for Recon 1

that models acylcarnitine and fatty-acid oxidation (Ac-FAO) has recently been published<sup>8</sup>, which includes the metabolic surroundings of many biomarkers measured in the worldwide newborn screening program<sup>21</sup>. Recon 1 has also been used for the automated generation of cell-specific and tissue-specific models using various 'omics' data sets<sup>22,23</sup>, and to generate metabolic reconstructions semiautomatically for other mammals, such as the mouse<sup>24</sup>.

It is clear that 'competing' (that is, different) reconstructions and reconstruction approaches coexist, but all have the common goal of providing an up-to-date, comprehensive and high-quality reconstruction, either at the global or cell-specific scale. Rather than continuing to duplicate efforts, a substantial fraction of the community has pooled resources to generate a consensus human metabolic reconstruction from many of the sources cited above.

Here we describe Recon 2, a community-driven expansion of the global human metabolic reconstruction, Recon 1. Much of this expansion was performed at reconstruction 'jamboree' meetings<sup>25</sup>, focused events at which domain experts apply their knowledge to refine and consolidate biochemical knowledge from existing reconstructions and published literature. Members of the *Saccharomyces cerevisiae*<sup>26,27</sup> and *Salmonella typhimurium* LT2 (ref. 28) communities have used such a jamboree approach. The jamboree events provided the opportunity to establish common standards (and suitable links to other databases) for the consensus reconstruction and the format of its content, to simplify its reuse and extension, and to increase its transparency and commitment to its development via participation of many stakeholders in the community. Here we also demonstrate the improved predictive capability of Recon 2 over that of its predecessor. We mapped exometabolomic data<sup>29</sup> onto Recon 2 and used proteomics data to generate cell type-specific metabolic models, which we then investigated for their functional properties.

## RESULTS

### Reconstruction approach

To assemble Recon 2, we added metabolic information present in four different resources (EHMN<sup>17</sup>, HepatoNet1 (ref. 18), Ac-FAO module<sup>8</sup>

**Table 1** Comparison of major features of Recon 2 and its predecessor Recon 1

Property	Recon 1	Recon 2
Total number of reactions	3,744	7,440
Total number of metabolites	2,766	5,063
Number of unique metabolites	1,509	2,626
Number of metabolites in extracellular space	404	642
Number of metabolites in cytoplasm	995	1,878
Number of metabolites in mitochondrion	393	754
Number of metabolites in nucleus	95	165
Number of metabolites in endoplasmic reticulum	235	570
Number of metabolites in peroxisome	143	435
Number of metabolites in lysosome	217	302
Number of metabolites in Golgi apparatus	284	317
Number of transcripts	1,905	2,194
Number of unique genes	1,496	1,789
Number of subsystems	90 <sup>a</sup>	99
Number of blocked reactions (% of all reactions)	1,270 (34%)	1,603 (22%)
Number of dead-end metabolites	339 <sup>b</sup>	1,176 <sup>b</sup>
Size of $S^c$ ( $m$ ; $n$ )	2,766; 3,744	5,063; 7,440
Number of linearly independent mass balances	1,070	2,774
Sparsity (% of nonzero entries in $S$ )	0.1385	0.0837
Number of accomplished metabolic tasks	294	354
Mapped IEMs (% of all IEMs)	233 (71%)	248 (76%)
Number of unique genes causing IEMs	255	272
Number of IEMs affecting metabolic tasks (% of effective IEMs)	71 (36%)	98 (44%)

<sup>a</sup>Based on the subsystems categories defined for Recon 2. <sup>b</sup>All but 32 dead-end metabolites Recon 1 have been resolved in Recon 2. Addition of new reactions and pathways (mostly from EHMN) led to the introduction of new dead-end metabolites, which was a starting point for subsequent curation efforts. <sup>c</sup>The stoichiometric matrix ( $S$ ) consists of  $m$  rows, each representing a single metabolite and  $n$  columns, each representing a single reaction. When a metabolite participates in a reaction the corresponding cell in the  $S$  matrix has a negative nonzero entry for substrates and a positive nonzero entry for products.



**Figure 2** Pathway coverage in Recon 1 and Recon 2. Only pathways with added reaction content in Recon 2 are shown. The 29 pathways that have identical reaction content in both reconstructions are not shown.

and the human small intestinal enterocyte reconstruction<sup>19</sup>) to the content of Recon 1 following a step-wise process (Fig. 1). We added more than 370 transport and exchange reactions, based on a review of literature. We applied unambiguous third-party identifiers for cellular compartments, metabolites, enzymes and reactions. We mapped the content from the DrugBank<sup>30</sup> database, which lists experimental and US Food and Drug Administration–approved drugs, to individual enzymes and reactions. Ninety-five percent of metabolic reactions were mass-balanced and charge-balanced<sup>5,31</sup> (Supplementary Table 1), except for those containing metabolites that have either no defined chemical formula or a generic formula. We tested Recon 2 for self-consistency<sup>2</sup>, a process that included gap analysis<sup>2</sup> and leak tests<sup>32</sup>.

### Benchmarking Recon 2 against Recon 1

Recon 2 accounts for 1,789 enzyme-encoding genes, 7,440 reactions and 2,626 unique metabolites distributed over eight cellular compartments, which is a large increase in comprehensiveness relative to Recon 1 (Table 1). Such an increase in scope does not necessarily constitute an

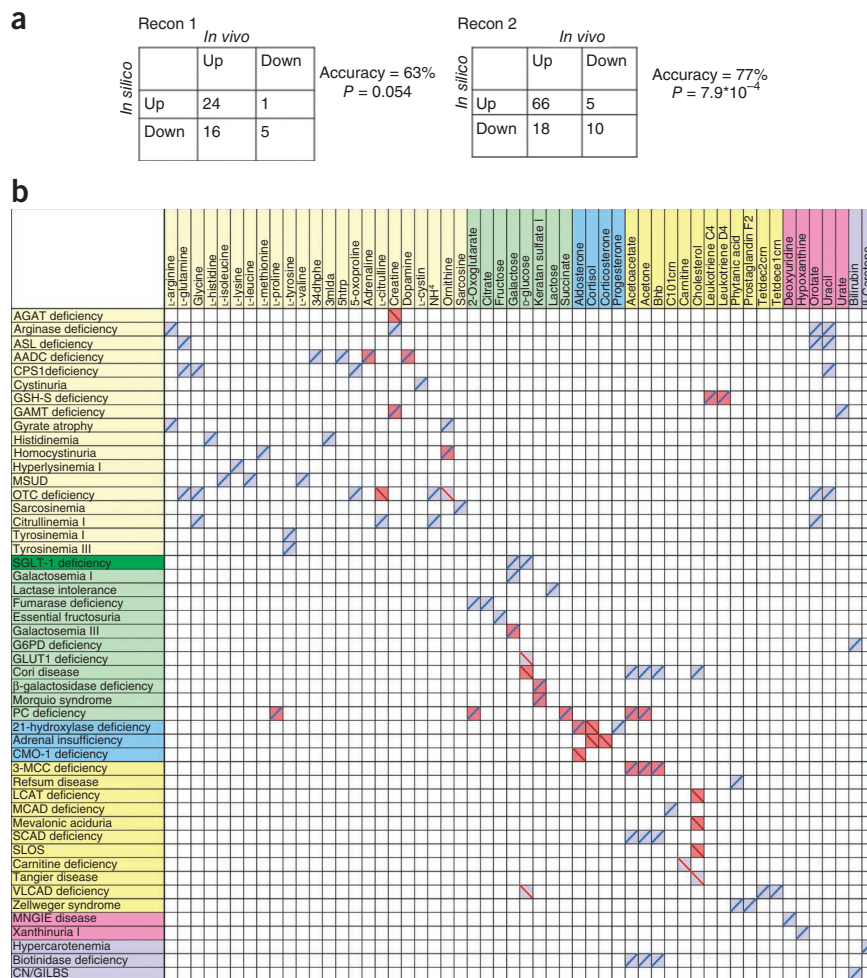
improvement in utility over the previous version: expanding the reconstruction to resolve existing gaps and dead-end metabolites may introduce additional gaps and dead ends elsewhere. To demonstrate an improvement of the network, both coverage and functional improvements must be considered.

To quantify the overall improvements achieved through the community-driven expansion and refinement in the global human metabolic reconstruction, we compared the information coverage, topological and functional properties of Recon 2 with those of Recon 1 (Table 1). The reaction content was almost doubled, much of which belonged to one of the nine new pathways (Fig. 2). Moreover, 62% (61/99) of the existing pathways have been expanded in Recon 2, and reaction coverage in 29 pathways, accounting for 16.5% (1,231/7,440) of the reactions, remained unchanged. A total of 307 dead-end metabolites (metabolites that are either only produced or only consumed in the reconstruction) from Recon 1 were resolved in Recon 2, whereas 32 remained as participants in only one reaction. As a result of the expansion, 1,144 new dead-end metabolites, mostly from EHMN, were introduced. These will need to be connected to the rest of the network in subsequent efforts. Blocked reactions cannot carry a nonzero flux in any steady-state condition because they contain one or more dead-end metabolites or are in a linear pathway with such reactions. The expanded coverage of metabolic information resolved 827 blocked reactions present in Recon 1, and 443 blocked reactions remained (Table 1). The number of remaining and new blocked reactions and dead-end metabolites highlights that this current update is not intended to be the final compendium of human metabolism, but it is a major advance over Recon 1 and represents our current, continually evolving knowledge.

A metabolic task is defined as a nonzero flux through a reaction or through a pathway leading to the production of a metabolite B from a metabolite A. Examples of such tasks include the synthesis of all known precursors to produce a cell (biomass reaction; Supplementary Note 2) and the generation of energy via oxidative phosphorylation or fermentation (Supplementary Table 2). A total of 354 metabolic tasks were defined. Although a particular cell type is not capable of fulfilling all these metabolic tasks, Recon 2 should be able to fulfill these tasks because it is a global metabolic reconstruction. Recon 2 carried a nonzero flux for all tasks, compared with Recon 1, which achieved this functionality for only 83% of the tasks (Table 1).

To benchmark the models derived from both reconstructions against an independent data set, we used a manually assembled compendium of IEMs<sup>8</sup> as a gold standard. This compendium accounts for 330 IEMs, such as phenylketonuria and orotic aciduria, along with their known metabolite biomarkers. As Recon 2 captured more metabolic genes, more IEMs could be mapped (Table 1).

**Figure 3** Predicted biomarkers for IEMs. (a) Comparison of the prediction accuracy of Recon 1 and Recon 2 against the gold standard<sup>8</sup>. (b) Correct and incorrect predictions. IEMs and biomarkers are sorted by subsystem. Bright yellow, amino-acid metabolism; green, central metabolism; blue, hormones; yellow, lipid metabolism; pink, nucleotide metabolism; lilac, vitamin and cofactor metabolism. Blue and red shading corresponds to predicted increase and decrease in biomarker, respectively. Blue and red lines represent reported increase and decrease of the biomarker in plasma, respectively. 34dhphe, 3,4-dihydroxy-L-phenylalanine; 3mlda, 3-methylimidazoleacetic acid; 5htrp, 5-hydroxy-L-tryptophan; bhh, (R)-3-hydroxybutanoate; tetdec2crn, tetradecadienoyl carnitine; tetdece1crn, tetradecenoyl carnitine. (See **Supplementary Table 5** for complete names of the IEMs.)



In Recon 2, almost all of the mapped IEMs affected the reaction activity, as no complementary isoenzymes are known for the absent enzymes (consistent with their occurrence as IEMs). We compared the predictive potential of Recon 2 and Recon 1 for associated biomarkers for the mapped IEMs (Fig. 3), in a process analogous to gene-deletion studies in microbial modeling. Recon 2 predicted 54 reported biomarkers for 49 different IEMs, with an accuracy of 77%. The coverage of predicted biomarkers and the accuracy was much lower for Recon 1, with 31 reported biomarkers for 29 IEMs and an accuracy of 63% (Fig. 3). This comparison demonstrates that the increased scope of Recon 2 led to a higher coverage of IEM-related biomarkers mapped and to an increase in predictive power.

Based on the accurate predictive capability of Recon 2 for biomarkers and for the metabolic tasks, the benchmarking demonstrated an increase in both scope and predictive accuracy of Recon 2 relative to its predecessor.

### Recon 2 captures the majority of known exometabolites

Recon 2 accounts for 642 extracellular metabolites, which should be found in cell culture medium and in biofluids such as plasma and urine. When comparing these extracellular metabolites with a reported subset of a cancer exometabolome<sup>33</sup> consisting of 140 metabolites, the majority of the metabolites were indeed present in the extracellular compartment (Fig. 4a and Supplementary Table 3). Using a flux variability analysis<sup>34</sup>, we predicted *in silico* the uptake and release profile of Recon 2. Recon 2 had high sensitivity in predicting correctly the uptake and release for 91 metabolites (sensitivity values of 0.92 and 0.94, respectively; Supplementary Table 3). However, the capability of Recon 2 to predict true negatives (that is, metabolites that cannot be taken up or released) was very low. For instance, 14% (13/91) of the metabolites could not be released by any of the tested cancer cell lines, whereas all but one metabolite was released *in silico*. As Recon 2 represents the combined metabolic capability of all cells in the human body, not just of neoplastic cells, the low sensitivity is expected. The mismatches could be used to guide the assembly of cancer-specific metabolic models or to refine an existing cancer metabolic model<sup>9</sup>.

It is noteworthy that Recon 2 contains many more extracellular metabolites (551), which were not seen in the experimental exometabolome, presumably in part because of the targeted liquid chromatography–tandem mass spectrometry method used<sup>33</sup>, as more than 1,000 metabolite peaks can easily be observed in human serum<sup>35</sup>.

We also compared the Recon 2 exometabolome and the metabolites reported in the Human Metabolome Database (HMDB)<sup>36</sup> as being detectable in biofluids (Fig. 4b). Biofluid information could be found for about half of the metabolites in Recon 2 identified in the HMDB. About 44% of these metabolites were also present in the extracellular compartment, indicating that there are still some transport and metabolic routes missing in Recon 2.

In summary, Recon 2 captured many of the reported exometabolites and biofluid metabolites, illustrating its comprehensiveness and the value of annotating with multiple metabolite identifiers, permitting the integration of data from many sources.

### Generation of draft, cell type-specific models

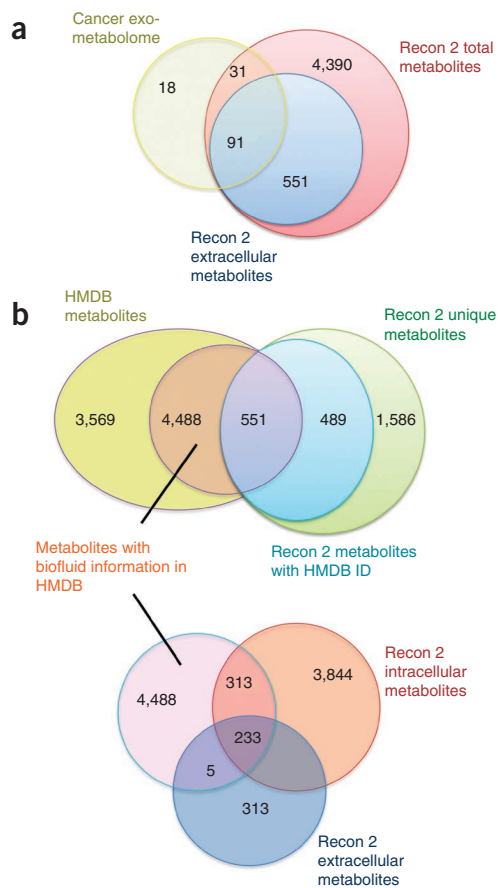
A metabolic reconstruction is unique to a genome, and thus to an organism, but condition-specific constraints can be applied to create a condition-specific model from the reconstruction. Thus, one reconstruction can give rise to many models. We mapped expression data from the Human Protein Atlas<sup>37</sup> for 65 cell types, capturing information for 25% (451/1,789) of the unique gene products in Recon 2 (Supplementary Fig. 1). These data were used together with a published algorithm<sup>5,38</sup> to generate 65 draft cell type-specific metabolic



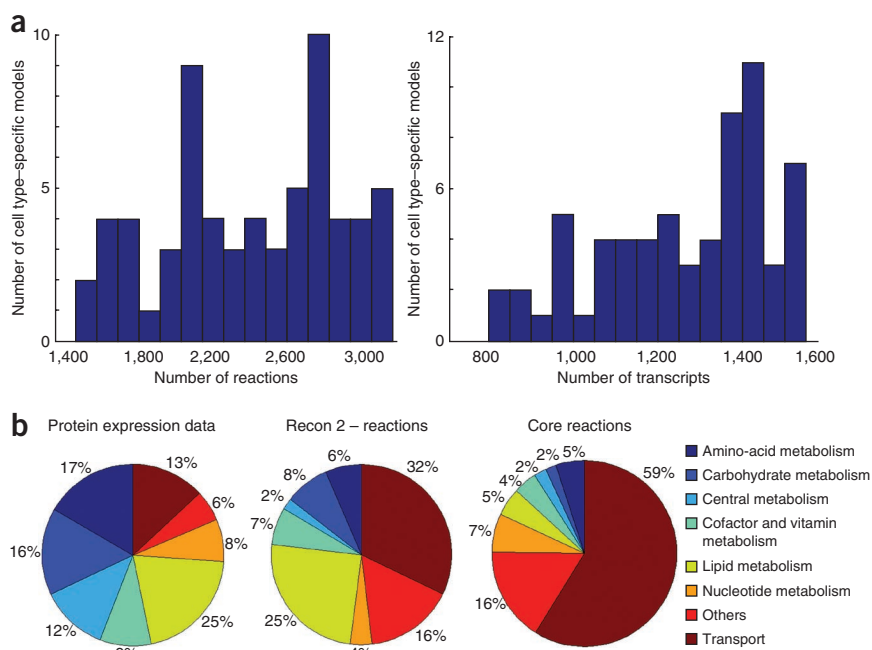
**Figure 4** Comparison of metabolomic data with the extracellular metabolites present in Recon 2 metabolites. (a) Comparison with the cancer exometabolome<sup>33</sup> (see **Supplementary Table 3** for a breakdown of the 18 cancer exometabolites that are absent from Recon 2, the 31 for which no exchange reaction is present, and the 91 for which uptake or release profile comparisons could be performed). (b) Comparison with the HMDB<sup>36</sup> (see **Supplementary Fig. 6**).

models consisting of  $2,426 \pm 467$  reactions ( $\pm$  s.d.) and  $1,262 \pm 204$  transcripts (**Fig. 5**). Of the 593 core reactions present in all cell-type models, more than half of them were transport reactions (**Supplementary Table 4**). Furthermore, 33% (2,463/7,440) of the Recon 2 reactions appeared in none of the cell-type models, with almost 40% (968/2,463) of those belonging to the subsystem ‘lipids’ (**Supplementary Fig. 2**). Based on the cell type-specific models, 26% (457/1,789) of the genes in Recon 2 represent the core genes (that is, they were found in all cell type-specific models), and 58% of the genes were part of two or more models. Only 32 genes were specific to a particular cell type-specific model. The remaining 14% (245/1,789) of the genes were not captured by any cell type-specific model. The glandular-cell models were highly correlated based on presence and absence of genes and reactions, whereas epithelial-cell models were less correlated (**Supplementary Figs. 3,4**). Overall, we observed higher correlation on the subsystems level with the ‘transport’ subsystem being the most correlated one, highlighting the importance of nutrient uptake and secretion for cellular function (**Supplementary Figs. 3,5**).

The automatically generated hepatocyte model captured 61% (1,109/1,823) of the reactions present in a published hepatocyte model<sup>20</sup>, whereas 64% (1,932/3,041) of this model’s reactions were not in the published model. When comparing our draft hepatocyte model with the published hepatocyte model, HepatoNet1 (ref. 18), 1,098 reactions were present in both, whereas 64% (1,943/3,041) were unique to our model and 57% (1,441/2,539) of reactions were unique to HepatoNet1. Such discrepancies between our draft model and the published cell-type models may serve as an indication of areas of attention for subsequent manual curation and assessment.



We investigated the metabolic tasks that each cell type-specific model can perform. On average, the cell type-specific models had nonzero flux for  $174 \pm 32$  of 354 metabolic tasks. Thirty-one metabolic tasks could be carried out by all such models, many of which tested different aspects of amino-acid metabolism, and 44 metabolic tasks had a zero flux for all models. Of all of these models, only the hepatocyte model could generate urea, via the urea cycle, which is consistent with current knowledge. Small intestinal enterocytes are also capable of urea synthesis but are currently not captured by the Human Protein Atlas<sup>37</sup>. Notably, 25% (16/65) of the models had a nonzero flux through the biomass reaction, which means that these models contain all necessary reactions either to take up or synthesize all the precursors defined in the biomass reaction.



**Figure 5** Summary properties of the 65 draft cell type-specific models. (a) Distribution of reaction (left) and genes (right) across the models. (b) The protein expression data captured proteins from all subsystems almost evenly. In contrast, most of reactions in the core reaction set, consisting of 593 reactions, belonged to the transport subsystem. The second largest subsystem in Recon 2, lipid metabolism, represented only a small fraction of the core reaction set.



## Recon 2 includes mappings of drug actions to enzymes

Recon 2 includes 2,657 metabolic enzymes and 1,052 enzymatic complexes, many of which are known drug targets. We queried DrugBank<sup>30</sup>, a comprehensive resource that includes drug-to-enzyme mappings for >6,000 small-molecule and peptide or protein drugs, to allow drugs (and their actions) to be mapped to the enzymes of Recon 2. We found that 1,290 drugs were mapped to 308 enzyme and enzymatic complexes. This equated to 3,168 drug-enzyme or drug-complex interactions, of which 841 were specified as inhibitory. These mappings are included in both the global reconstruction and the cell type-specific models, providing a starting point for the simulation of drug actions with both constraint-based and kinetic modeling<sup>1,39,40</sup>.

## DISCUSSION

Our results illustrate that (i) Recon 2 is a comprehensive metabolic resource and serves as an effective predictive model; (ii) mapping of exometabolomic data onto Recon 2 provides a starting point for further iterative expansion and refinement; (iii) protein expression data and Recon 2 can be used to generate draft cell type-specific metabolic models; and (iv) comparative analysis of cell type-specific models provides insight into alternate metabolic strategies.

The metabolic reconstruction process is inherently iterative, as increasing biochemical and genomic knowledge is generated about the target organism over time. This calls for periodic updates and expansion in the coverage and content of a reconstruction<sup>2</sup>; thus we adopted the 'Recon 1' and 'Recon 2' naming convention in analogy with the 'build X' convention for the assembly of the human genome sequence and similar conventions used in the naming of iteratively released metabolic reconstructions in yeast<sup>27</sup>. The published resources for human metabolism differ in syntax and content; for instance, a comparison of five different resources revealed that only a small number of overlapping reactions were present in all resources (ref. 16 and **Supplementary Note 1**). These discrepancies make it difficult to compare and combine metabolic reconstructions. We overcame this issue by manually curating overlapping entries. The presented consensus reconstruction of human metabolism is fully semantically annotated<sup>41</sup> with references to persistent and publicly available chemical and gene databases, unambiguously identifying its components and increasing its applicability for third-party users (including automated processing by software). Moreover, the work expanded beyond combining existing resources, by adding transport and absorption reactions known to occur in epithelial cells of the gastrointestinal tract and the renal tubules. Drug information was mapped from DrugBank<sup>30</sup>, providing a comprehensive starting point to investigate off-target effects of drugs<sup>11</sup> or to obtain information on known drugs for drug target predictions studies<sup>9</sup>.

We improved the predictive potential of Recon 2 through the addition of new content. The number of IEMs that Recon 2 captured and the accurate prediction of known biomarkers (**Fig. 3**) demonstrate the substantial functional advancement achieved through the expansion and refinement. We illustrated the potential of the constraint-based modeling approach by generating a global model, not tailored to a particular application, which can nonetheless predict accurately many distinct biomarkers for many IEMs. In addition, Recon 2 also predicted some previously undescribed biomarkers, which have not yet been measured<sup>8</sup>. With the increasing sophistication of targeted metabolomic approaches<sup>42</sup>, biomarker predictions may help to guide such analyses. Additionally, large cohort studies have started to connect genotype with metabolotype<sup>43</sup>. As Recon 2 captures most of the measured metabolites, it provides a resource for investigating the connection between genotype and metabolotype and ultimately phenotypes based on these data.

We mapped two large-scale exometabolome resources to Recon 2, showing that Recon 2 captures most of the exometabolites and biofluid metabolites reported therein. Although this illustrated the comprehensiveness of the reconstruction, it also highlighted that the information content in the reconstruction is still not complete. Many algorithms exist that may assist in proposing missing metabolic and transport reactions in Recon 2 and thus aid subsequent manual curation (**Supplementary Note 3**)<sup>12,13,44</sup>. It is also clear that reported biofluid metabolites<sup>36</sup> might be of dietary and/or microbial origin, as the mammalian gut microbiota has been found to affect blood composition<sup>45</sup> and metabolism<sup>46</sup> substantially. To determine the origin of the biofluid metabolites, a comprehensive computational model accounting for microbial metabolism, human metabolism and dietary composition is needed.

Recon 1 has been used together with omics data to generate comprehensive sets of draft cell type-specific metabolic reconstructions<sup>23,38</sup>. We mapped data from the Human Protein Atlas<sup>37</sup> onto Recon 2 to generate 65 cell type-specific metabolic models automatically. Although protein expression data were available for only one-quarter of the Recon 2 gene products, to our surprise the draft cell type-specific models contained many reactions and genes (**Fig. 5**). The size of many published cell type-specific models, which have been at least in part generated automatically, is comparable<sup>23,37</sup>. In a comparison of our automatically generated hepatocyte model with two published models<sup>18,20</sup>, we found reasonable overlap, even though these models were assembled using different methods and information. Our automatically generated hepatocyte model contained alternate reactions, which reflect the proteomic input data and the algorithm used. For instance, the protein expression data were limited to those that had a high confidence level, but by including lower-confidence expression data, combined with a weighting scheme, one may obtain alternate draft models. Moreover, the algorithmic approach currently does not consider transcriptional regulation, thermodynamic constraints and the synthesis cost for each enzyme and its building blocks. Ultimately, manual inspection and curation will be necessary to obtain more versatile, comprehensive, predictive cell type-specific models.

To assess the functionality of the automatically generated cell type-specific models, we tested 354 metabolic tasks. One-quarter of the models could produce all defined biomass precursors, thus enabling them to simulate cell growth. A review of literature revealed that some of these cell types are indeed known to divide *in vivo* upon injury or induction by specific growth factors. The protein expression data are generated in many cases on tissue biopsies. The growth capabilities could not have been determined from the omics data alone, and this highlights the importance of computational modeling to analyze experimental data. Moreover, the metabolic activity profile, combined with the set of active exchange reactions, could be compared with cell type-specific literature to evaluate the correctness of the metabolic content of each model and to refine the models by enabling or disabling additional tasks, exchange reactions and metabolic content.

We anticipate that, as a result of the improvements over its predecessor, Recon 2 will be widely used and will enable the exploration of new frontiers in research in human metabolism and its role in health and disease. The global model is available via a database at <http://humanmetabolism.org/>, in SBML format at <http://identifiers.org/biomodels.db/MODEL1109130000> and as **Supplementary Data 1**.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

Note: Supplementary information is available in the online version of the paper.

#### ACKNOWLEDGMENTS

I.T. was supported, in part, by a Marie Curie International Reintegration Grant (249261) within the 7th European Community Framework Program. I.T., O.R., S.G.T. and M.A. were supported by the European Research Council grant proposal number 232816. S.S. and H.H. were supported by a Rannis research grant (100406022). Authors from Manchester thank the Biotechnology and Biological Sciences Research Council (BBSRC), and Engineering and Physical Sciences Research Council for their funding of the Manchester Centre for Integrative Systems Biology (grant BB/C008219/1), H.W. for Bioprocessing Research Industry Club grants, and P.M. and N. Swainston for support from the European Union FP7 project UNICELLSYS (grant agreement 201142). The Knut and Alice Wallenberg Foundation supported R.A., S.B. and I.N.; D.J. thanks the BBSRC for funding of Systems Approaches to Biological Research grants BB/F005938 and BB/F00561X. A.H. and C.B. were supported by the German Federal Ministry for Education and Research within the Virtual Liver Network (grant numbers 0315756 and 0315741). A.K.C. and J.A.P. acknowledge funding from the US National Institutes of Health (grant GM088244), National Science Foundation (grant 0643548) and Cystic Fibrosis Research Foundation (grant 1060). N.D.P. was supported by a National Cancer Institute to Independence Award in Cancer Research. I.G. thanks the Science and Technology Facilities Council for Scottish Bioinformatics Research Network funding. M.H. and P.M. thank the US National Institute of General Medical Sciences for support under grants R01GM070923 and R01GM080219. M.D.S. thanks the BioRange programme (project SP1.2.4) of The Netherlands Bioinformatics Centre for support under a Besluit Subsidies Investeren Kennisinfrastructuur grant through The Netherlands Genomics Initiative.

#### AUTHOR CONTRIBUTIONS

I.T. and N. Swainston led the project and developed the reconstruction. I.T., N. Swainston, B.Ø.P., P.M. and D.K. wrote the manuscript. B.Ø.P., D.B.K. and P.M. conceived the project. R.M.T.F. and A.H. performed validation of the reconstruction. S.S. and M.L.M. performed substantial manual curation of the reconstruction. M.K.A., H.H., O.R., M.D.S. and S.G.T. contributed to the analysis of the reconstruction and its models. I.T., N. Swainston, R.M.T.F., A.H., S.S., M.K.A., H.H., M.L.M., O.R., M.D.S., S.G.T., R.A., C.B., S.B., A.K.C., P.D., W.B.D., L.E., D. Hala, M.H., D. Hull, D.J., N. Jamshidi, J.J., N. Juty, S.K., I.N., N.L.N., N.M., A.M., J.A.P., N.D.P., E. Selkov, M.I., E. Simeonidis, N. Sonnenschein, K.S., A.S., J.H.G.M.v.B., D.W., I.G., J.N., H.V.W., D.B.K., P.M. and B.Ø.P. attended one or more jamboree meetings and provided manual curation of the reconstruction.

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Palsson, B. Systems biology: properties of reconstructed networks. (Cambridge University Press, 2006).
- Thiele, I. & Palsson, B.O. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **5**, 93–121 (2010).
- Oberhardt, M.A., Palsson, B.O. & Papin, J.A. Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* **3**, 20 (2009).
- Orth, J.D., Thiele, I. & Palsson, B.O. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
- Schellenberger, J. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* **6**, 1290–1307 (2011).
- Bordbar, A. & Palsson, B.O. Using the reconstructed genome-scale human metabolic network to study physiology and pathology. *J. Intern. Med.* **271**, 131–141 (2012).
- Duarte, N.C. *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA* **104**, 1777–1782 (2007).
- Sahoo, S., Franzson, L., Jonsson, J.J. & Thiele, I. A compendium of inborn errors of metabolism mapped onto the human metabolic network. *Mol. Biosyst.* **8**, 2545–2558 (2012).
- Folger, O. *et al.* Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* **7**, 501 (2011).
- Frezza, C. *et al.* Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature* **477**, 225–228 (2011).
- Chang, R.L., Xie, L., Bourne, P.E. & Palsson, B.O. Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comput. Biol.* **6**, e1000938 (2010).
- Rolfsson, O., Palsson, B.O. & Thiele, I. The human metabolic reconstruction Recon 1 directs hypotheses of novel human metabolic functions. *BMC Syst. Biol.* **5**, 155 (2011).
- Rolfsson, O., Paglia, G., Magnusdottir, M., Palsson, B.O. & Thiele, I. Inferring the metabolism of human orphan metabolites from their metabolic network context affirms human gluconokinase activity. *Biochem. J.* **449**, 427–435 (2013).
- Bordbar, A., Lewis, N.E., Schellenberger, J., Palsson, B.O. & Jamshidi, N. Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions. *Mol. Syst. Biol.* **6**, 422 (2010).
- Heinken, A., Sahoo, S., Fleming, R.M. & Thiele, I. Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. *Gut Microbes* **4**, 28–40 (2013).
- Stobbe, M.D., Houten, S.M., Jansen, G.A., van Kampen, A.H. & Moerland, P.D. Critical assessment of human metabolic pathway databases: a stepping stone for future integration. *BMC Syst. Biol.* **5**, 165 (2011).
- Hao, T., Ma, H.W., Zhao, X.M. & Goryanin, I. Compartmentalization of the Edinburgh Human Metabolic Network. *BMC Bioinformatics* **11**, 393 (2010).
- Gille, C. *et al.* HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol. Syst. Biol.* **6**, 411 (2010).
- Sahoo, S. & Thiele, I. Predicting the impact of diet and enzymopathies on human small intestinal epithelial cells. *Human Mol. Genet.* (in the press).
- Jerby, L., Shlomi, T. & Ruppin, E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol. Syst. Biol.* **6**, 401 (2010).
- McHugh, D.M. *et al.* Clinical validation of cutoff target ranges in newborn screening of metabolic disorders by tandem mass spectrometry: a worldwide collaborative project. *Genet. Med.* **13**, 230–254 (2011).
- Blazier, A.S. & Papin, J.A. Integration of expression data in genome-scale metabolic network reconstructions. *Front Physiol.* **3**, 299 (2012).
- Agren, R. *et al.* Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput. Biol.* **8**, e1002518 (2012).
- Sigurdsson, M.I., Jamshidi, N., Steingrimsson, E., Thiele, I. & Palsson, B.O. A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. *BMC Syst. Biol.* **4**, 140 (2010).
- Thiele, I. & Palsson, B.O. Reconstruction annotation jamborees: a community approach to systems biology. *Mol. Syst. Biol.* **6**, 361 (2010).
- Herrgard, M.J. *et al.* A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat. Biotechnol.* **26**, 1155–1160 (2008).
- Heavner, B.D., Smallbone, K., Barker, B., Mendes, P. & Walker, L.P. Yeast 5—an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network. *BMC Syst. Biol.* **6**, 55 (2012).
- Thiele, I. *et al.* A community effort towards a knowledge-base and mathematical model of the human pathogen *Salmonella Typhimurium* LT2. *BMC Syst. Biol.* **5**, 8 (2011).
- Kell, D.B. *et al.* Metabolic footprinting and systems biology: the medium is the message. *Nat. Rev. Microbiol.* **3**, 557–565 (2005).
- Wishart, D.S. *et al.* DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **36**, D901–D906 (2008).
- Swainston, N., Smallbone, K., Mendes, P., Kell, D. & Paton, N. The SuBliMinaL Toolbox: automating steps in the reconstruction of metabolic networks. *J. Integr. Bioinform.* **8**, 186 (2011).
- Hoppe, A., Hoffmann, S., Gerasch, A., Gille, C. & Holzhtuter, H.G. FASIMU: flexible software for flux-balance computation series in large metabolic networks. *BMC Bioinformatics* **12**, 28 (2011).
- Jain, M. *et al.* Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. *Science* **336**, 1040–1044 (2012).
- Gudmundsson, S. & Thiele, I. Computationally efficient flux variability analysis. *BMC Bioinformatics* **11**, 489 (2010).
- Zelena, E. *et al.* Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Anal. Chem.* **81**, 1357–1364 (2009).
- Wishart, D.S. *et al.* HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **37**, D603–D610 (2009).
- Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010).
- Shlomi, T., Cabili, M.N., Herrgard, M.J., Palsson, B.O. & Ruppin, E. Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.* **26**, 1003–1010 (2008).
- Smallbone, K., Simeonidis, E., Broomhead, D.S. & Kell, D.B. Something from nothing: bridging the gap between constraint-based and kinetic modelling. *FEBS J.* **274**, 5576–5585 (2007).
- Smallbone, K., Simeonidis, E., Swainston, N. & Mendes, P. Towards a genome-scale kinetic model of cellular metabolism. *BMC Syst. Biol.* **4**, 6 (2010).
- Le Novère, N. *et al.* Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.* **23**, 1509–1515 (2005).
- Paglia, G. *et al.* Monitoring metabolites consumption and secretion in cultured cells using ultra-performance liquid chromatography quadrupole-time of flight mass spectrometry (UPLC-Q-ToF-MS). *Anal. Bioanal. Chem.* **402**, 1183–1198 (2012).
- Suhre, K. *et al.* A genome-wide association study of metabolic traits in human urine. *Nat. Genet.* **43**, 565–569 (2011).
- Reed, J.L. *et al.* Systems approach to refining genome annotation. *Proc. Natl. Acad. Sci. USA* **103**, 17480–17484 (2006).
- Wikoff, W.R. *et al.* Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl. Acad. Sci. USA* **106**, 3698–3703 (2009).
- Claus, S.P. *et al.* Systemic multicompartmental effects of the gut microbiome on mouse metabolic phenotypes. *Mol. Syst. Biol.* **4**, 219 (2008).

## ONLINE METHODS

**Reconstruction approach.** The reconstruction of the expanded and refined global human metabolic network, Recon 2, was performed in multiple stages. Intermediate versions of Recon 2 are referred to as 'Recon 1.x'.

**Jamboree work.** The jamboree meetings were used to discuss strategies combining the content of the two reconstructions and the required quality control of the finished consensus. The strategy was to start with the compartmentalized Recon 1 reconstruction and incrementally add reactions from the EHMN reconstruction.

Initially, automated approaches<sup>31</sup> were applied to Recon 1 (ref. 7) and the EHMN<sup>17</sup> to solve the problem of inconsistent naming of the components (compartments, metabolites, genes and reactions). The remaining components that could not be automatically matched to existing database entries were manually annotated during the jamboree meetings. Cellular compartments were annotated with Gene Ontology (GO) terms. Metabolites were annotated with terms from the resources Chemical Entities of Biological Interest (ChEBI), Kyoto Encyclopedia of Genes and Genomes (KEGG) Compound, PubChem Compound and HMDB<sup>36</sup>, and also IUPAC International Chemical Identifier (InChI) terms where possible. Where metabolites were not present in any existing data resources, many were submitted as new entries to ChEBI. Enzymes were annotated with Enzyme Classification (EC) terms, US National Center for Biotechnology Information (NCBI) Gene identifiers and UniProt terms.

The jamboree meetings also focused on manual curation of the reconstruction content. Reactions were curator-validated and annotated with PubMed literature references, standardized GO evidence codes (**Supplementary Table 6**) and a confidence scoring system ranging from 0 (no evidence) to 4 (biochemical evidence)<sup>2</sup>. Metabolites and enzymes were assigned to appropriate cellular compartments. Metabolic reactions were checked to ensure correct stoichiometry, irreversibility, correct assignment of gene association and enzyme rules, and mass and charge balancing, and appropriate transport reactions were added.

**Post-jamboree work.** For reactions that occurred in both Recon 1 and EHMN, the co-occurrence was considered to be evidence for their inclusion in Recon 1.x. The jamboree teams had manually evaluated reactions from Recon 1, and reactions unique to the compartmentalized EHMN were added to Recon 1.x only if they were mass- and charge-balanced. The charge state of metabolites was calculated at an assumed pH of 7.2. Reactions were mass- and charge-balanced where possible, as defined in Recon 1. Where additional metabolites and reactions were added from sources such as EHMN, the SuBliMinaL Toolbox<sup>31</sup> and COBRA Toolbox<sup>5</sup> were used to apply mass and charge balancing. In cases where reactions contained 'generic' compounds (containing -R and -X groups), which were used to represent a set of specific compounds to decomplexify the reconstruction, mass and charge balancing was not possible. An investigation into thermodynamic consistency of Recon 1 reactions<sup>47</sup> proposed changes to a small subset of reactions, which were also incorporated. We also updated the gene list based on changes introduced from build 35 to build 37, which includes the removal or replacement of obsolete gene entries.

Subsequently, the Recon 1.x content was extended to account for the content in the published HepatoNet1 (ref. 18) reconstruction. The same procedure for converting HepatoNet1 to the syntax of Recon 1.x was used as described above. As HepatoNet1 is a pruned and cell-specific stoichiometric model, some conventions were different from that in Recon 1.x. Most importantly, the implementation of lipids as some specific pools with fixed fatty-acid distributions was incompatible with the flexible and less specified R system in Recon 1.x, and therefore the lipid reactions were excluded from the merging process. HepatoNet1 accounts for multiple extracellular compartments. For simplicity, all of the HepatoNet1 extracellular metabolites and reactions were assumed to be present in a single common extracellular compartment, following the convention of Recon 1.

Content from both the recently published acylcarnitine-fatty acid oxidation module<sup>8</sup> and the metabolic reconstruction for human small intestinal enterocytes<sup>19</sup> were also added to Recon 1.x. The latter accounts for two extracellular compartments (luminal and blood side). Again, a single extracellular compartment was assumed.

As a final step of expanding the content of the global human metabolic reconstruction, thorough literature research was performed to identify missing transport and absorption reactions in Recon 1.x. The reconstruction of this module was performed as described previously<sup>2</sup>.

The reconstruction assembly and conversion to a mathematical model was done using rBioNet<sup>48</sup>. The uniqueness of reactions in Recon 1.x was determined, and multiple rounds of self-consistency testing and functional assessment of the model performed using established procedures<sup>2</sup>, which included gap analysis<sup>2</sup>, leak tests<sup>32</sup> and a functional analysis.

The resulting reconstruction was finalized and named Recon 2.

**Leak analysis.** A model of Recon 2 was used for the leak analysis, in which all unbalanced reactions were considered inactive by constraining their lower and upper bound flux bounds to zero. The test of mass leaks was performed with a simulation procedure defined in the tutorial of FASIMU<sup>32</sup>, by looking for steady-state flux distributions that either (i) consume no substrates and generate an output or (ii) consume substrates but do not generate products. The flux distributions representing a leak were analyzed manually to identify causative unbalanced reactions. After balancing or deleting the corresponding reaction(s), the test was repeated until no additional leaks were observed. In all simulations with Recon 1 and Recon 2, all metabolites with defined exchange reactions could be taken up and secreted. Reaction constraints applied to the leak-free Recon 2 model that was used for all computations in this study are listed in **Supplementary Table 7**.

**Topological analysis.** Performing flux variability analysis<sup>34</sup>, while permitting constrained uptake and secretion of all metabolites with defined exchange reactions has been defined, allowed the identification of blocked reactions in Recon 1 and Recon 2, that is, reactions that could not carry any nonzero flux under this simulation condition. Dead-end metabolites were identified as described previously<sup>2</sup>. Compound participation was calculated as described by previously<sup>1</sup> (**Supplementary Note 4**).

**Functional characterization of Recon 1 and Recon 2.** The metabolic capacity of the network was demonstrated by testing nonzero flux values for 354 metabolic tasks, which were based on Recon 1 and ref. 19. For each of the simulations a steady-state flux distribution was calculated. Each metabolic task was optimized individually by choosing the corresponding reaction in either Recon 1 or Recon 2, if present, as objective function and maximized the flux through the reaction (see **Supplementary Table 2**).

**Mapping the compendium of inborn errors of metabolism and predicting biomarkers.** The compendium of IEMs<sup>8</sup> contains information about causative genes and known biomarkers for 330 distinct IEMs. Using gene-reaction associations, the IEMs were mapped onto Recon 1 and Recon 2. A gene-associated reaction was affected if no isoenzyme existed that was not known to cause an IEM. To predict biomarkers for each IEM affecting one or more reaction, the method reported in ref. 49 was followed. Changes in biomarkers resulting from an IEM were compared with reported biomarkers for each IEM. A Fisher's exact test was applied to compute the hypergeometric *P* value. This analysis was performed for Recon 1 and Recon 2. Mapped IEMs are listed in **Supplementary Table 5**.

**Cancer exometabolome mapping.** Information related to a cancer exometabolome<sup>33</sup>, which reported consumed and secreted compounds in the culture medium, was manually mapped to the metabolites of Recon 2 (**Supplementary Table 3**). Flux variability analysis was performed on the exchange reactions, and the uptake-secretion capability was compared with reported consumption and secretion capability of the cancer cells.

**Biofluid metabolome mapping.** Information about the biofluid location of metabolites was obtained from the MNDB (version August 2012)<sup>36</sup>. Mapping was performed through HMDB identifiers, which were recorded during the reconstruction process.

**Protein expression data mapping.** Protein expression data were downloaded from the Human Protein Atlas<sup>37</sup> in May 2012 and Ensembl identifiers were mapped onto



Recon 2 gene products. Gene products with moderate/medium and strong/high levels of expression were assumed to be present, and all others were assumed to be absent. Gene products without data in a cell type were assumed to be absent.

**Generation of a draft cell type-specific model compendium.** For each cell type, the presence and absence information from the protein expression data was used as input for the MinMax algorithm<sup>38</sup>, which was implemented in the COBRA Toolbox<sup>5</sup>. The biomass reaction and a reaction representing ATP hydrolysis were subsequently added to each model. The IEM compendium was mapped onto each cell type-specific draft model as described above. For each model, its capability to perform the defined metabolic tasks was tested.

**SBML models.** Recon 2 is available in the Systems Biology Markup Language format (SBML)<sup>50</sup>, which is compliant with the Minimal Information Required In the Annotation of Models (MIRIAM) standard<sup>41</sup> at <http://humanmetabolism.org/>, Biomodels (<http://identifiers.org/biomodels.db/MODEL1109130000>) and as **Supplementary Data 1** (which also contains all derived cell type-specific versions). The programming library libAnnotationSBML<sup>51</sup> was used to apply unified cross-references in the form of MIRIAM identifiers to most components in the models. Systems Biology Ontology (SBO) terms were applied to specify metabolites, polypeptides and protein complexes, and to make the distinction between biochemical and transport reactions.

**DrugBank data mapping.** DrugBank<sup>30</sup> was queried via its XML file download, and enzymes were mapped to their equivalent in Recon 2 by consideration of both UniProt identifiers and specified intracellular compartmentalization. Drugs (and their actions) were mapped to enzymes and included in the distributed SBML<sup>50</sup> versions of the Recon 2 models as annotations on enzyme species.

**Software.** All computations were carried out in the Matlab programming environment (MathWorks, Inc.) using the COBRA Toolbox<sup>5</sup> and Tomlab cplex as the linear programming solver (TomOpt, Inc.).

47. Haraldsdottir, H.S., Thiele, I. & Fleming, R.M. Quantitative assignment of reaction directionality in a multicompartmental human metabolic reconstruction. *Biophys. J.* **102**, 1703–1711 (2012).
48. Thorleifsson, S.G. & Thiele, I. rBioNet: A COBRA toolbox extension for reconstructing high-quality biochemical networks. *Bioinformatics* **27**, 2009–2010 (2011).
49. Shlomi, T., Cabili, M.N. & Ruppin, E. Predicting metabolic biomarkers of human inborn errors of metabolism. *Mol. Syst. Biol.* **5**, 263 (2009).
50. Hucka, M. *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531 (2003).
51. Swainston, N. & Mendes, P. libAnnotationSBML: a library for exploiting SBML annotations. *Bioinformatics* **25**, 2292–2293 (2009).