The intricate nature of human physiology renders its study a difficult undertaking, and a systems biology approach is necessary to understand the complex interactions involved. Network reconstruction is a key step in systems biology and represents a common denominator because all systems biology research on a target organism relies on such a representation. With the recent development of genome-scale human metabolic networks, metabolic systems analysis is now possible and has initiated a shift towards human systems biology. Here, we review the important aspects of reconstructing a bottom-up human metabolic network, the network’s role in modeling human physiology and the necessity for a community-based consensus reconstruction of human metabolism to be established.

The human genome: a system defined
The human genome sequence is generally touted as a blueprint for the physiological functions of the human body. However, the translation from the annotated sequence to cellular physiology remains elusive. The annotated sequence contains a wealth of information about the gene products (proteins) and the biochemical processes that they mediate. This information essentially provides a ‘parts’ list of biological components that exist in human cells, but it is the interplay among these components that governs physiological behavior.

A system is a collection of individual parts that work synergistically as a single, functional unit. Much of human physiology can be described in a synonymous manner; for example, tissue-level functions are the result of interactions among multiple cell types, and the human body is comprised of interconnected organ and tissue systems that enable various whole-body functions. This overall systems view on biology is the conceptual foundation for the construction and use of networks to understand human physiologic functions.

The complex nature of biological interactions has led to the need for and development of systems biology, an emerging field of research that combines high-throughput experimentation with computational tools for the systematic analysis of biological systems. Two approaches are commonly used in systems biology, top-down and bottom-up; both aim at understanding the interactions arising under various physiological conditions but from different perspectives. The top-down strategy analyzes from a ‘global’ point-of-view by dissecting the overall system into its smaller interacting parts and generally requires statistical measures. Alternatively, the bottom-up approach first specifies in detail the individual components and interactions of a system. The components are then pieced together into a larger system, thus providing a mechanistic basis for studying its underlying biology based on the known, specified parts. The data type largely determines the approach that is most amenable to answering a particular biological question, which is in itself dependent on the level of interaction detail available.

In the post-genomic era, the ‘genome-to-life’ concept has ignited (molecular) systems biology and the implementation of its central paradigm: molecular components networks computational models physiological studies. Because much of the component data are genome-derived, this paradigm is tantamount to developing a mechanistic genotype–phenotype relationship. The prospects of such a development would indeed be transformative in biology and, if successfully realized, would have a broad impact on the life sciences and life-science-based industries. The shift towards human systems biology has been initiated with the emergence of bottom-up, genome-scale human metabolic networks (2–4) (Table 1). This review will outline recent advancements in the reconstruction and analysis of genome-based human metabolic networks and their emerging role in studying human health and disease.
Metabolism: foundational to human health and disease

Metabolism is widely known to play an important part in human physiology. Its function is important for understanding disease states and progression [5,6], aging and nutrition [7–10] and improving the performance of individuals such as athletes, astronauts and soldiers [11–13]. In particular, metabolism has been known to be involved in many major disease states, such as diabetes, obesity and cardiovascular disease [14,15]. Cancers display highly abnormal metabolic phenotypes, and metabolic targets have long been used in cancer chemotherapy [16–18]. More recently, evidence is growing that the effects of metabolism on physiological and pathophysiological brain functions are significant, from schizophrenia to neurodegenerative disorders [19–23]. Successful implementation of molecular systems biology of human metabolism is thus likely to have broad consequences.

The reconstruction of genome-scale metabolic networks in microorganisms is now well developed [24,25] and has been successfully implemented in computing microbial

<table>
<thead>
<tr>
<th>Metabolic network</th>
<th>Description</th>
<th>Publication</th>
<th>Reconstruction approach</th>
</tr>
</thead>
</table>
| HumanCyc        | An automated draft reconstruction | *Genome Biology* (2005) [2] | • Automated mapping from Ensembl [61] and LocusLink (now Entrez Gene [62]) databases  
• PathoLogic algorithm to fill pathway 'holes' with candidate reactions by sequence comparison |
| Recon 1          | A manually curated network reconstruction and functional *in silico* model | *PNAS* (2007) [3] | • Automated mapping from databases including KEGG [63], Ensembl [61], H-Invitational [64] and Entrez Gene [62]  
• Manual curation of mapped data and literature  
• Manual reaction gap assessment  
• 288 Flux Balance Analysis (FBA)-based, functional validation tests |
| Edinburgh Human Metabolic Network | Automated genome annotation mapping and inclusion of biochemical legacy data | *Molecular Systems Biology* (2007) [4] | • Automated mapping and cross validation from databases including KEGG [63], UniProt, HGNC [65] and Ensembl [61]  
• Enzymes and Metabolic Pathways (EMP) database used to obtain and integrate biochemical literature-based information |

**Table 1. List of available genome-scale human metabolic networks.**

**Figure 1.** Incorporation of genomic and biochemical knowledge derived from the genome annotation and experimental literature into a BiGG-structured knowledge base network. High-throughput annotation data provides information on gene products, transcript variants and their associated functions, as well as localization (i.e. cellular compartment and tissue). Literature documents specific biochemical details from experiments on the gene product functions, such as reaction mechanism and substrate specificity.
metabolic phenotypes for a variety of biological applications [26,27]. Although the biochemical makeup of enzymes and metabolites varies, the general framework of metabolism is consistent across all organisms, and metabolic studies have been enabled for a wide range of species using similar methodologies. Thus, a logical next step is to extend the successful development and analysis of microbial metabolic networks to an analogous effort for human metabolism. Reconstructing metabolism provides a basis for building large-scale, mechanistically accurate networks for human physiology because the biochemical transformations of metabolism are well studied and documented at both pathway and mechanistic levels. Metabolic systems might hence become the first process in human cells and tissues where the application of molecular systems biology will bear fruit.

Collecting metabolic knowledge: a reaction network point-of-view

The annotated human genome, along with literature (or ‘bibliome’) data, defines the known biological components present in the human body. Because this information exists in many different domains, there is a need to compile the data in a structured format that catalogs genes, their associated protein products and related biological functions. A component-by-component, or ‘bottom-up’, approach to network reconstruction results in a biochemically, genetically and genomically (BiGG) structured format that serves as a knowledge base of genome-derived biological components and a framework for computational modeling. This part of the reconstruction process has been described extensively for microbial metabolic networks [24,27] and can thus be readily applied towards human metabolism.

Figure 1 illustrates the general information that is derived and considered from annotation databases and the bibliome as it is incorporated into a BiGG reconstruction. The information is structured such that the whole genome is dissected into its gene parts (genomically), genes are related to its encoded proteins (genetically) and protein enzymes are linked to their catalytic reaction functions and the metabolite species in which they interconvert (biochemically). Such information is structured to describe the connections among genes, proteins and their respective metabolic functions. Complex relationships, such as reactions catalyzed by more than one enzyme (i.e. isozymes) and multifunctional proteins, can be textually and graphically described as Boolean logic relationships as an additional layer of reaction information [2,3]. This multi-level structure distinguishes a knowledge base network from a standard database by providing an integrative view of disparate data types and placing them in a relevant biological context.

Automation and manual curation approaches can be used to reconstruct a network. Automated methods have been used to enable quicker mining and cross-comparison of data from different resources and automatically assign potential functions to annotated genes. However, this essentially generates a rough draft of the knowledge content that requires additional refinement to enhance its content quality. Further investigation and consideration of experimental data based on literature reading provides evidence for a biochemical reaction’s addition to the network and has been emphasized as an integral part of the reconstruction process [3,4]. Although manual verification of literature evidence is a timely and laborious process, it is a crucial quality control procedure that a reconstruction process must undergo to ensure content quality. Such procedures include ensuring charge- and mass-balanced reactions, determining the localization of reaction activity, identifying substrate and cofactor specificity and, more specific to human metabolic reconstructions, incorporating alternatively spliced variants. These data workflows have previously been defined and have been successfully implemented in microbial and human metabolic networks [24,25,28].

Although the genome annotation can be used to derive a majority of human metabolic reactions, there are generally missing reactions where the genome has not been fully elucidated, and additional evaluation is needed to fill in these gaps. Gap-filling can be done algorithmically; for example, one approach proposes candidate reactions that might fill gaps in a human metabolic pathway by projecting known reactions in pathways from other organisms [2]. Alternatively, reactions can be inferred directly from network topology [29]. A more arduous approach involves manual assessment of published literature and adding those reactions that were not automatically identified from the annotated genome. Literature mining and review not only increases the confidence of adding a reaction but also places the physiological functions that each reaction fulfills into context (e.g. at the metabolic pathway level).

Further consideration can be made to ensure reaction gaps are filled in the context of metabolic functions. Cellular biomass growth or energy demands are metabolic functions that are primarily assessed for networks of single-celled, microbial organisms. However, current genome-scale human networks represent metabolic reactions that exist in any human cell and are therefore a ‘global’ depiction of all human metabolic functions. Thus, in reconstructing a network for ‘global’ human metabolism, a variety of basic physiological functions that can exist in any human cell must be considered. In one study describing a global reconstruction of human metabolism, functional validation testing was implemented during the reconstruction procedure and ensured that the 288 basic metabolic processes described were computationally functional [3]. Tested metabolic processes represent a defined list of known physiological functions and include ketogenesis, ATP production and biosynthesis of non-essential amino acids from their respective precursors. Algorithms have now been developed that can be useful during reconstruction to computationally identify gaps and determine candidate reactions that might fulfill a particular metabolic function [30,31].

Once formed, the BiGG knowledge base becomes the basis for a mathematical representation of the network that is reconstructed from it. A network comprises nodes and links, where the nodes are biological components and the links are chemical transactions among components. Thus, a metabolic network is composed of metabolites and biochemical reactions that catalyze transformations between them as the respective nodes and links. Its math-
Mathematical description culminates in a matrix format of stoichiometric coefficients where the rows are the metabolite components and the columns the reaction links, effectively representing a two-dimensional annotation of the genome [32]. This representation thus provides the mathematical context to quantitatively study human metabolism as a whole as well as by compartment (e.g. cell- or tissue-specific).

**Community approach to reconstructing the human metabolic network**

Reconstruction efforts by multiple research groups have resulted in human metabolic network representations that vary in content owing to differences in reconstruction approaches and literature interpretation [2–4]. Thus, the development of a human network reconstruction whose content is agreed upon necessitates a collective community effort to formalize such a network. The notion of a two-dimensional annotation jamboree for a consensus network reconstruction was first articulated at a meeting in 2006 (http://issy25.vtt.fi/) and successfully implemented in 2007 for *Saccharomyces cerevisiae* metabolism. With the ‘jamboree’ approach, a focused meeting of experts in *S. cerevisiae* biology and modeling approaches culminated in the establishment of a consensus metabolic network by early 2008 and its subsequent publication in the fall of 2008 [33].

In the case of *S. cerevisiae*, the consensus network reconstruction approach began with the comparison of two existing network reconstructions to identify existing components that were agreed upon by both networks. This was followed by detailed curation of non-agreeing components and details, such as reaction mechanism and substrate specificity. The consensus meeting also resolved issues dealing with standardizing nomenclature conventions to facilitate easier high-throughput data integration for future network analysis. For instance, species identifiers for metabolites and enzymes were standardized such that metabolomic and proteomic data could be directly mapped onto the network for a user’s analysis. Having undergone a two year process from start to finish, the jamboree approach for yeast metabolism was deemed a success because it resulted from cooperative efforts and input from a wide scientific community. A similar effort for human metabolism is likely to be more extensive, but the resulting yeast consensus network was an encouraging indication of a comparable outcome for its human counterpart.

**Figure 2.** The four general steps towards studying a physiological system *in silico*. Components from various biological sources are first compiled and incorporated into a BiGG-structured format. The assembled information becomes the basis for a network reconstruction specific to the system of interest. The network can be converted into a computational model by imposing mathematical parameters relevant to specific biological conditions (e.g. genetic, environmental). Each step in the process formulates a more detailed *in silico* representation of the studied system, thus increasing the level of physiological focus and refining the genotype-phenotype relationship.
Advancing systems-level analysis of human metabolism

Reconstruction of genome-scale human metabolic networks has initiated development towards studying human physiology *in silico* at a systems level. Four crucial steps to this process have been described [1], with the biological representation becoming more focused and detailed at each level and thus culminating in a systems framework for analyzing and modeling human metabolic phenotypes (Figure 2).

Two general approaches used to study metabolic network systems are (i) topology-based analysis and (ii) *in silico* modeling. Metabolic networks are data structures describing known biochemical interactions among metabolites and, as such, can be used to define structural, or topological, features of metabolism. A key motivation for understanding network topology is that metabolic functions, and corresponding malfunctions, are thought to be determined in part by how their interactions are structured. This topology–function relationship has already been extensively studied in microbial networks [34–39], as well as in smaller-scale human metabolic and signaling networks [40–43], and a similar approach can be readily applied towards understanding human physiological and disease functions at a larger scale.

The imposition of modeling parameters adds another dimension to analyzing genome-scale networks. By formulating mathematical constraints corresponding to different physiological environments, a network can be adapted into a computational, or *in silico*, model of metabolism for a specified condition. A network itself can be used to look at connectivity properties, and its converted model form can be used to study its functional metabolic states. Various mathematical procedures have already been developed that explore systems properties (in particular the steady-state properties) of genome-scale models [26,44]. Once in its *in silico* form, the genome-scale model can be used to perform several biological studies, as demonstrated by the historical use of the *Escherichia coli* metabolic reconstructions [27,45]. Progress has already been made in adapting computational approaches for the study of smaller human metabolic networks; such approaches include the calculation of candidate metabolic states under diabetic, ischemic and dietary conditions in the cardiomyocyte mitochondria [46] and the simulation of the effects of plasma environment on hepatocyte metabolism [47,48].

With the availability of genome-scale human metabolic networks, the scope of studies and the potential spectrum of relevant biomedical applications will undoubtedly expand. Some broader applications where current human metabolic networks can make an immediate impact include: (i) disease phenotype characterization; (ii) functional genomics; and (iii) high-throughput data analysis. Recent examples of such applications are discussed.

*Pathophysiology: disease phenotype characterization*

Several networks have now been constructed to reflect current genomic and biochemical knowledge and will be useful for systems-level interpretation of disease phenotypes. For example, graph-based network analysis can be used to modularize and define functional relationships that are consistent with known disease-associated reactions [4]. In another study, flux coupling network analysis identified correlated sets, or co-sets, of reaction flux rates that were biologically relevant to connected disease phenotypes and potential drug targets [3]. For instance, a reaction co-set involving glutathione metabolism contained multiple enzyme reactions that were causally related to hemolytic anemia. Reactions correlated in another co-set containing 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, a common target for the anti-lipidemic drug class of statins, can be considered as potential alternative drug targets for hyperlipidemia [3].

The gene–reaction relationships defined in metabolic networks can enable a straightforward implementation and analysis of disease-associated genetic and epigenetic perturbations. More recently, a metabolic disease network was constructed to describe gene–disease phenotype relationships using the Online Mendelian Inheritance in Man (OMIM) database (http://www.ncbi.nlm.nih.gov/omim/) and a human metabolic network [49]. A large-scale comparison with patient records indeed showed that adjacent reaction connectivity is correlated with coexpression of disease phenotypes and is thus a possible predictor of disease comorbidity. Detailed analysis of disease associations provided additional insight into underlying reaction mechanisms contributing to a shared pathophysiology. An interesting example discussed the high association between diabetes and hemolytic anemia, which, upon further analysis, was explained by a NADPH deficiency via glucose-6-phosphate dehydrogenase leading to glutathione deficiency, resulting in hemolytic anemia. Such studies indicate that human-metabolic-network-based studies can expand into mechanistic analysis of disease relevant to diagnostic and therapeutic applications.

*Annotating gene functions*

Network gaps and non-gene-associated reactions are inherently present in human metabolic reconstructions and represent areas where knowledge data are not available. These ‘knowledge gaps’ are attractive in terms of functional genomics because they describe potential metabolic functions and gene annotations that have yet to be discovered. Computational tools that predict candidate reactions based on fulfilling defined metabolic pathways or functions have now been developed [30,31,50]. These can be combined with networks to facilitate a systems approach of annotating missing metabolic functions or genes. For instance, this was successfully applied on a genome-scale *E. coli* network and led to the novel functional assignment and experimental verification of eight genes associated with metabolic and transporter functions [30]. Such a systematic approach can be useful in facilitating a focused, integrated computational and experimental process to further drive ongoing human genome annotation efforts.

*Interpreting high-throughput data*

The vast amount of information generated by high-throughput technologies requires computational tools to facilitate its biological interpretation. Network reconstructions are organized such that the incorporation and
analysis of different data types can be done in the context of a genomically and biochemically structured framework (i.e. ‘context for content’). Because human metabolic networks represent ‘global’ human metabolism, the networks can be further refined by datasets for specified conditions. The gene–protein–reaction structured format allows the integration of gene and protein expression data to determine metabolic reaction activity present for a given condition [28]. A method exploiting this format was recently used to predict tissue-specific metabolism based on gene expression and proteomic information mined from various public databases [51]. By integrating tissue-specific high-throughput data with a global metabolic network, specific metabolic activities for ten different tissues were accurately predicted and determined a significant role for post-transcriptional regulation in tissue-specific metabolic phenotypes.

In addition to gene and protein expression data, metabolomic information can also be directly incorporated because it is explicitly represented in metabolic networks. High-throughput metabolomics has become increasingly available and provides quantitative information on various metabolic states. For example, the Human Metabolome Database (HMDB) contains detailed physiological and disease concentration metabolite levels from several biofluids and tissues [52]. Because metabolic networks describe the mechanistic relationships among metabolites, they can be used to systematically identify underlying pathways associated with the measured metabolites. Such applications are an alternative to the standard, top-down statistical approaches used to analyze high-throughput ‘omics’ data and offer exciting prospects of interpreting high-throughput data in the context of a mechanistically detailed modeling framework.

**Towards comprehensive human modeling**

Human metabolic physiology arises collectively from different levels of biological organization, as illustrated in Figure 3, and thus requires a systems perspective to understand it as a whole. Whole-body metabolic functions are mediated by the interactions and exchange among various compartments (i.e. among different cell or tissue types); thus, a multi-compartment human model can provide a better depiction of physiologically relevant metabolic states. Efforts have been initiated towards developing context-specific metabolic networks from high-throughput data [51,53], providing the basis for constructing specific, segmented networks from a global metabolic network. Accurate representations of individual cellular and tissue compartments will be essential to such higher-level modeling and will require rigorous assessment of their unique metabolic functions and demands. Specified metabolic objectives, such as those described in [3], can be used to query and assess the potential functional capabilities of tissue-specific networks.

Once individual networks have been formulated, they can be integrated to form multi-network models. Recently, a codependent, two-organism microbial system was constructed to model flux interactions between separate metabolic networks and was shown to accurately predict known physiological features [54]. Indeed, similar approaches can also be applied in principle when modeling metabolite exchange and interaction among multiple tissue metabolic networks. With the methods for constructing and integrating context-specific metabolic networks in different cell- or tissue-types at hand, a multi-compartmentalized model of human metabolism could soon be realized.

Genome-scale metabolic networks are a starting point for modeling more complex biological processes in humans. An intuitive next step is to extend the scope of human molecular systems modeling from metabolism to other biological systems as more details on their mechanisms and interactions are elucidated. The application of similar methodologies used in reconstructing metabolism has been prototyped at a smaller scale and applied towards human signal transduction pathways [55,56], as well as at the...
genome-scale for E. coli transcriptional and translational machinery (I. Thiele et al., unpublished observations). The merging of these systems would indeed represent a more accurate and comprehensive description of the inherently dependent nature of cellular processes [57]. Additionally, integration of transcriptional regulation with metabolism has already been achieved for microbial networks in the form of Boolean logic networks [58–60]. Although human transcriptional regulation is markedly more complex than regulation in microorganisms, the general framework has been formulated and can potentially be implemented at a genome-scale as detailed data on regulatory component interactions become available.

Concluding remarks
We have reached a mature stage in the development of metabolic systems biology in microbes, and the extension of this approach to human metabolism has now been initiated. A key component is the bottom-up reconstruction of genome-scale metabolic networks based on our current knowledge, a process that requires detailed curation and incorporation of genomic and biochemical information into a mathematically structured network. Recent network applications have shown that it is possible to analyze the genotype–metabolic phenotype relationship in a systematic and quantitative fashion from both mechanistic and systems standpoints. However, what we can learn from genome-scale human networks is still limited by the content and scope of information that is incorporated into the reconstructed network. Therefore, future advances must be made towards the integration and modeling of multiple tissues and other biological systems to improve the physiological relevance and predictive potential of reconstructed networks. This distribution of information and domain expertise is widely dispersed and necessitates the implementation of a reconstruction jamboree with wide scientific community participation. The first manually curated human metabolic reconstruction – published in 2007 – is beginning to be widely used, so it is clear that such an undertaking is necessary to advance our knowledge of human metabolism, as well as our capability to model it.

Acknowledgements
We thank Neema Jamshidi and Ines Thiele for their valuable comments and suggestions during preparation of the manuscript.

References
2 Romero, P. et al. (2005) Computational prediction of human metabolic pathways from the complete human genome. Genome Biol. 6, R2
5 Matsuzawa-Nagata, N. et al. (2008) Increased oxidative stress precedes the onset of high-fat diet-induced insulin resistance and obesity. Metabolism 57, 1071–1077
14 Fuevray, D. and Darmeliah, A. (2008) Diabetes-related metabolic perturbations in cardiac myocyte. Diabetes Metab. 34 (Suppl. 1), S3–S9
22 Huang, C. et al. (2007) Changes in network activity with the progression of Parkinson’s disease. Brain 130, 1834–1846
31 Satish Kumar, V. et al. (2007) Optimization based automated curation of metabolic reconstructions. BMC Bioinformatics 8, 212
Review

Trends in Biotechnology  Vol.27  No.1

42 Famili, I. and Palsson, B.O. (2003) The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. Biophys. J. 85, 16–26