The Journal of Nutrition
**International Research Conference on Food, Nutrition, and Cancer**

# Metabolomics of a Superorganism[1–3]

Royston Goodacre*

School of Chemistry and Manchester Interdisciplinary Biocentre, University of Manchester, Manchester M1 7ND, United Kingdom

## Abstract

The human can be thought of as a human-microbe hybrid, and the health of this superorganism will be affected by intrinsic properties such as human genetics, diurnal cycles, and age and by extrinsic factors such as lifestyle choices (food and drink, drug intake) and the acquisition of a stable ''healthy'' gut microflora (the so-called microbiome). Alterations in this superorganism will be manifest in the metabolite complement within its serum and urine samples. The unraveling of this metabolic compartmentalization in this complex ecosystem will certainly be a challenge for systems biology and necessary for defining human health at the molecular level. Within the systems biology framework, functional analyses at the level of gene expression (transcriptomics), protein translation (proteomics), and, more recently, the metabolite network (metabolomics) have become increasingly popular. Metabolomics experiments aim to quantify all metabolites in a cellular system (cell or tissue) under defined states and at different time points so that the dynamics of any biotic, abiotic, or genetic perturbation can be accurately assessed. This article provides an overview of metabolomics and discusses how data are generated and analyzed within a systems biology framework. The role of metabolomics in nutrigenomics is also discussed, as are the concepts of the human being a superorganism and the complexities required to be overcome to understand human health and disease. J. Nutr. 137: 259S–266S, 2007.

## Metabolomics and systems biology

When the human genome sequencing projects drew their conclusions (1,2), what was surprising was that just as in the microbiologist's pet organism *Escherichia coli* (3), a vast number of genes had never been seen before in classical (molecular) genetics. For *E. coli* K-12, a jaw dropping 38% of the total 4288 open reading frames (ORFs)[4] had never been observed or studied before; of the remaining 60%, many of these were given gene function by association with DNA sequence matches to other organisms' genes held in the genomic databases and not by direct functional analyses. Despite the continuing deluge of genome sequences (http://www.genomesonline.org/), including those for microorganisms that are yet to be cultured in the laboratory (4), this situation has not improved significantly. For a whole organism, typically 20–40% of the genes can not be ascribed a function by sequence analogy, largely because although the databases are peppered with *A*s, *C*s, *G*s, and *T*s, they do not contain accurate information about what these ORFs actually do.

Clearly something needs to be done to assign biochemical function to these so-called orphan genes and to validate them as molecular targets for therapeutic intervention. In addition, many diseases have no clear diagnoses, and even fewer have any prognostic tests. Therefore, the search for biomarkers from body fluids that can serve as indicators of disease progression or response to therapeutic intervention has also increased.

Perhaps the way we were taught science is to blame for our forgetting the biological phenotype and the now urgent need to bridge the genotype-to-phenotype gap. Our love affair with molecular biology has certainly blinkered us, until recently, into thinking only in terms of its central dogma. In this concept, it is generally considered that there is a linear flow of information within a cell that goes from gene to transcript to protein. Enzymes would then affect metabolic pathways and thereby lead to changes in the phenotype of the organism (**Fig. 1** *A*). However, this traditional thinking is no longer accepted. The cellular processes are in reality intimately networked with many feedback loops and thus should be represented as dynamic protein complexes interacting with neighborhoods of metabolites (5). The construction, visualization, and understanding of these networks (6) certainly present big challenges for systems biology, as does a full understanding of the fluxes through metabolic neighborhoods and their control (7).

Molecular biology has been bogged down by hypothetical-reductionist thinking where small parts of the jigsaw have been

[1] Published in a supplement to *The Journal of Nutrition*. Presented as part of the International Research Conference on Food, Nutrition, and Cancer held in Washington, DC, July 13–14, 2006. This conference was organized by the American Institute for Cancer Research and the World Cancer Research Fund International and sponsored by (in alphabetical order) the California Walnut Commission; Campbell Soup Company; Cranberry Institute; Hormel Institute; IP-6 International, Inc.; Kyushu University, Japan Graduate School of Agriculture; National Fisheries Institute; and United Soybean Board. Guest editors for this symposium were Vay Liang W. Go, Susan Higginbotham, and Ivana Vucenik. *Guest Editor Disclosure:* V.L.W. Go, no relationships to disclose; S. Higginbotham and I. Vucenik are employed by the conference sponsor, the American Institute for Cancer Research.

[2] Author Disclosure: No relationships to disclose.

[3] The author received financial support from the Engineering and Biological Systems committee of the UK BBSRC.

[4] Abbreviations used: HCA, hierarchical cluster analysis; MCA, metabolic control analysis; MVA, multivariate analysis; ORF, open reading frames; PCA, principal components analysis.

* To whom correspondence should be addressed. E-mail: roy.goodacre@manchester.ac.uk.

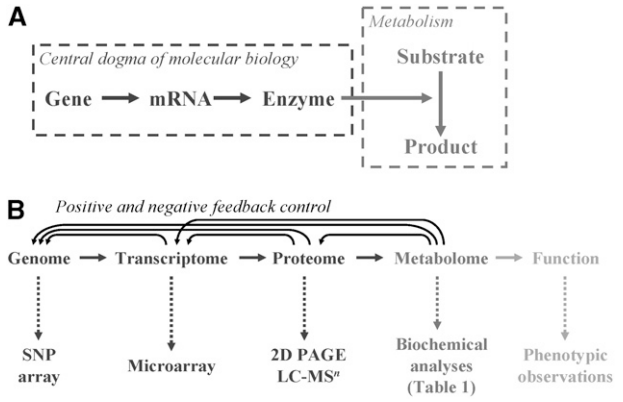


**Figure 1** (*A*) Traditional central dogma of molecular biology where the flow of information goes from gene to transcript to protein; also shown are sites where enzymes act on metabolism. (*B*) General schematic of the 'omic organization where the flow of information is from genes to transcripts to proteins to metabolites to function (or phenotype).

studied in isolation; although this approach has yielded useful information, albeit in isolation, the whole picture is often missed. Even functional analyses (Fig. 1 *B*) have emphasized analyses at the level of gene expression (transcriptomics), protein translation (proteomics, including posttranslational modifications), and more recently the metabolite network (metabolomics). The repopularization of systems biology constitutes a paradigm shift for molecular biology and will initially be dominated by the integrative analyses of these 'omics to generate predictive and hypothesis-generating mathematical models to better understand the cell at the systems level (8).

Metabolomics has a special place within systems biology, functional genomics, and biomarker discovery (9,10). The theory of metabolic control analysis (MCA) (7,11), as well as practical evidence (12), shows that although changes in the levels of individual enzymes may be expected to have little effect on metabolic fluxes, they can and do have significant effects on the concentrations of a variety of individual metabolites. Furthermore, Figure 1 *B* highlights the fact that metabolites are further down the line from gene to function and so more closely reflect the activities of the cell at the phenotypic (functional) level. Indeed, it has been shown (13) that changes in the metabolome are expected to be amplified relative to changes in the transcriptome and the proteome, as changes in the metabolite pool are the downstream result of gene expression. A consideration of the number of transcripts, proteins, and metabolites in humans also suggests that metabolomics is "the way forward." The human genome contains 31,897 genes (http://eugenes.org/). The expression and alternative splicing of the mRNAs indicate that humans may be able to produce $10^6$ different proteins (14)! By contrast, a recent prediction by Palsson suggests that there are only 2645 metabolites (15). Certainly 2645 is a more manageable number to measure accurately. Finally, because many diseases are a result of metabolic disorders, it would be sensible to measure metabolism directly.

## Metabolome analyses: considerations and approaches

As with any emerging science, there are a plethora of approaches being developed and improved on to measure quantitative changes in a cell's metabolism caused by either abiotic, biotic, or genetic perturbation, typically as a function of time. Although this area is constantly evolving, a number of recent reviews have extensively summarized the current technological approaches that are being developed within the metabolomics field (10,16–21), and so only a brief synopsis of the pertinent points is given here.

Metabolites by their nature are generally labile species, chemically very diverse, and often present in a very wide dynamic range. In addition, if one is analyzing a living system, the enzymes are still active, and so metabolite turnover will continue [for example, in yeast this has been estimated to occur in seconds (22)]. Therefore, it is important to quench metabolism or else the system will equilibrate, and the metabolite information will be compromised. Ideally metabolism should be stopped immediately. For unicellular organisms, this can be achieved by spraying the sample into very cold ($<-40°C$) buffered methanol (23). By contrast, for tissues from animals or plants, it is usual to snap freeze these samples in liquid nitrogen ($N_2$), after which mechanical disruption is employed to release metabolites (24).

The above quenching methods are, however, realistic only in a laboratory setting. For the analysis of human biofluids (e.g., blood, urine, tears, breath, and saliva), it may not always be possible to spray these immediately into cold methanol. If a patient is referred to a clinical biochemical laboratory at a local hospital, then immediate cessation of metabolic turnover from a freshly drawn blood sample is possible. By contrast, if the patient has been to see a doctor or nurse in a general practitioner's clinic, then it may be hours before the serum can be separated from the blood sample and frozen. The latter scenario is of course closer to reality for mass screening, and as part of the HUSERMET consortium (http://www.husermet.org/), this stability is being assessed and will form part of a wider program to investigate and then define the human serum metabolome in health and disease.

After metabolism has been quenched, most metabolomics analyses that use chromatography require that the metabolites be extracted. A variety of methods are being used (17,23,25), and the majority are based on acid, alkali, or ethanolic extraction. In addition, Folch-based extractions (26) using methanol:chloroform:$H_2O$ are also employed to separate polar metabolites from lipophilic ones before analysis.

Although there are no universally accepted metabolomics strategies, it is possible to summarize the most popular ones (16,27,28); these are detailed in **Table 1** and include metabolite target analysis, metabolite profiling, metabolomics, metabolite flux analysis, and metabolic fingerprinting (and footprinting). Table 1 includes a brief description of what these 5 strategies aim to achieve and which analytical approaches are commonly used. Clearly, to perform detailed metabolomics analyses, one must be suitably tooled up. Although this is certainly costly in terms of equipment and manpower, the actually running costs per sample are generally lower and of higher throughput than a typical microarray or proteomics experiment. This is especially true for the metabolic fingerprinting methods, which take in the order of a few seconds to minutes to generate information-rich metabolic fingerprints from each sample.

## Informatics: needs and requirements

All of the biochemical analyses that are used for metabolomics generate data torrents; when multiple time-course measurements are made to measure any dynamics, these become data floods, and when one thinks spatially (e.g., location within the liver), then tsunamis of data are easily produced! And it is perhaps no coincidence that one refers to the information overload in terms of natural disasters! As the French philosopher Jules Henri Poincaré (1854–1912) said in *La Science et l' Hypothèse*: "Science is built up with facts, as a house is with

**TABLE 1** Metabolomics strategies and common analytical platforms

| | |
|---|---|
| *Metabolite target analysis* | |
| An approach that is restricted to metabolites of, for example, a particular enzyme system that would be directly affected by abiotic, biotic, or genetic perturbation. | |
| Metabolite extraction approach is selective to those metabolites affected by the specific enzymes under study. | |
| HPLC | High-performance liquid chromatography |
| GC-MS | Gas chromatography-mass spectrometry (76) |
| LC-MS | Liquid chromatrography-MS |
| *Metabolite profiling* | |
| Focuses on a specific group of metabolites (e.g., lipids), within clinical and pharmaceutical analysis. Within the pharmaceutical sector this is often called *metabolic profiling,* which is used to trace the fate of a drug or metabolite. | |
| Metabolite extraction approach is selective to particular class of metabolites. | |
| HPLC-MS | HPLC coupled to electrospray ionization MS (77,78) |
| HILIC | Hydrophobic interaction liquid chromatography (79) |
| CE-MS | Capillary electrophoresis-MS (80,81) |
| LC-NMR | LC coupled to nuclear magnetic resonance (82) |
| LC-EC | Liquid chromatography using an electrochemical array (83) |
| *Metabolomics* | |
| Comprehensive analysis of the entire metabolome (all measurable metabolites) under a given set of conditions; this is often confused with *metabonomics,* which seeks to measure the fingerprint of biochemical perturbations caused by disease, drugs, and toxins. | |
| Many metabolite extraction approaches are used to be as comprehensive as possible. | |
| In addition to those used for metabolite profiling: | |
| GC(xGC)-MS | 2-dimensional GC coupled to MS (84) |
| LC$^n$-MS | 2D or parallel LC coupled to MS |
| UPLC | Ultraperformance LC-MS (78) |
| FT-ICR-MS | Fourier transform ion cyclotron resonance MS (85) |
| *Metabolite flux analysis* | |
| Commonly referred to as mass isotopomer analysis and by others as fluxomics. | |
| Labeled ($^{13}$C or $^{15}$N) metabolites are fed to tissue cultures (mammalian, plant, yeast, or bacterial), and the destination of these metabolites is assessed. Temporal studies may help reveal novel metabolic pathways and networks. | |
| NMR | For highly abundant metabolites isotope patterns can be investigated by NMR spectroscopy (86). |
| MS | Isotope distributions are investigated by chromatography linked to MS (87,88). This MS-based approach has often been referred to as SIDMAP [stable isotope-based dynamic metabolic profiling (89)] |
| *Metabolic fingerprinting/footprinting* | |
| Classification of samples based on provenance of either their biological relevance or origin using a fingerprinting technology that is rapid but does not necessarily give specific metabolite information. This approach is generally aimed at measuring the cell or tissue samples directly, where information on the intracellular metabolome is generated. | |
| Typically samples are analyzed directly or with very little extraction and without lengthy chromatography. | |
| The footprinting (90) techniques are variants on this approach in which extracellular metabolites are measured. In this case GC-MS or LC-MS may be used to generate comprehensive metabolic footprint profiles. | |
| NMR | Nuclear magnetic resonance spectroscopy (12,91) which can incorporate magic angle spinning (MAS) |
| DIMS | Direct infusion electrospray ionization-MS (92,93) |
| LDI-MS | Laser desorption ionization-MS (94) |
| MSLDI-MS | Matrix suppressed LDI-MS (95) |
| FT-IR | Fourier transform infrared spectroscopy (96,97) |
| Raman | Inelastic light scattering spectroscopy (97) |

stones. But a collection of facts is no more a science than a heap of stones is a house." Therefore, in addition to strategies for data storage, curation, and retrieval of metabolomics data, robust statistical analyses are needed to turn these data into information and hence understand changes in cellular metabolism.

Each metabolomics experiment contains 2 general data sets. The first are the metabolite data themselves and the second a summary of the experimental setup. The latter metadata (data about the data) are just as important as the metabolite data and need to be captured and stored so that database mining is easily achieved (e.g., one might want to retrieve all the GC-MS serum profiles from diabetic females who are nonsmokers, nonpregnant, aged 21–30, with a BMI of 20–25). With reference to data storage, curation, and retrieval, the reader is referred to **Table 2**, which summarizes the main sorts of database (29) and the current international efforts in progress to address the structure and contents of these metabolite databases.

After the data have been collected, the multivariate analysis (MVA) is tailored to the question that is being asked (*vide infra*). Metabolomics data are multivariate in that they consist of the results of observations of many different metabolites (variables; so-called *x*-data) for a number of individuals (objects). In MVA, each variable may be regarded as constituting a different dimension, such that if there are *n* variables (metabolites), each object may be said to reside at a unique position in an abstract entity referred to as *n*-dimensional hyperspace (30). This hyperspace is necessarily difficult to visualize, and the underlying theme of MVA is thus simplification or dimensionality reduction. This dimensionality reduction occurs in 1 of 2 ways, using either unsupervised or supervised learning algorithms.

**TABLE 2** Metabolomics database types and structures

| | |
|---|---|
| *Database types* | |
| Laboratory specific databases | Store raw (primary) data and metadata. |
| | Specific and narrow in topic. |
| | Contain much detailed information. |
| | Able to export data in standard formats to allow for interoperability with other databases. |
| Species-based databases | Store relatively simple metabolite profiles of collections of all experiments published for 1 species. |
| | Provide a data source for other types of experiments (phenotypic ones, sequencing, transcriptomics, and proteomics). |
| | Act as primary point of entry for species-related information. |
| Generic metabolite profiles | Do not store raw data but reference(s) to it. |
| | Because these are generic they must allow comparisons between different databases and between different metabolomics platforms. |
| | These are complex databases in that they contain all published metabolic profiles from many species in many different physiological states. |
| | Likely to be few of these, preferably all mirroring the same data. |
| Known metabolites for each biological species | Species specific. |
| | List all metabolites observed in that organism, that encompass different physiological states. |
| | Equivalent to gene databases in that they contain lists of all potential metabolites that could be seen in an organism. |
| | E.g., Human metabolome project: http://www.metabolomics.ca/ |
| All known metabolites | Not species specific. |
| | Contain a list of all metabolites ever likely to see, with a means for their identification (98). |
| | Data deposition is likely to be at the single organism level and organized taxonomically. |
| Reference biochemical databases | Represent established biochemical facts (i.e., reference information from the literature). |
| | Many exist already (KEGG is the most popular): |
| | KEGG: Kyoto encyclopedia of genes and genomes: http://www.genome.jp/kegg/ |
| | ExPASy: Biochemical pathways: http://www.expasy.ch/cgi-bin/search-biochem-index |
| | PUMA2: Evolutionary analysis of metabolism: http://compbio.mcs.anl.gov/puma2/ |
| | EcoCyc: Encyclopedia of *E. coli* K12 genes and metabolism: http://ecocyc.org/ |
| | BRENDA: The comprehensive enzyme information system: http://www.brenda.uni-koeln.de/ |
| | Reactome: Curated knowledgebase of biological pathways: http://www.reactome.org/ |
| *Metabolome databases and international standards* | |
| Armet | Architecture for metabolomics: http://www.armet.org/ |
| | Data schema for metabolomics including the basis for storage and transmission of data via UML (99). |
| SMRS | Standard metabolic reporting structure: http://www.smrsgroup.org/ |
| | Discussion of the details of what experimental metadata need to be captured (100). |
| MeMo | Metabolomic modeling: http://dbkgroup.org/memo/ |
| | Hybrid SQL/XML approach to metabolomic data management for functional genomics (101) |
| MSI | Metabolomics standards initiative: http://msi-workgroups.sourceforge.net/ |
| | International workgroups aimed at defining the minimal information required throughout a metabolomics study (102). |

Types of database are adapted from Mendes (29).

Unsupervised algorithms are used when one wants to answer a question such as "Are the metabolite profiles from the same patient taken throughout the day different?" In this case one wants an algorithm that will cluster the metabolite data into groups (31). For MVA this optimization procedure computes dimensionality reduction (30,32), where a large body of metabolite data (*x*-data) are summarized by means of a few parameters with minimal loss of information. The most used approaches are principal components (PCA) and hierarchical cluster analyses (HCA). After clustering (or what is also called explanatory analysis), the ordination plots or dendrograms then have to be interpreted. Alternative approaches include nested algorithms (e.g., PCA followed by HCA), soft independent modeling of class analogy (SIMCA) and k-nearest neighbors (kNN).

By contrast, for supervised learning algorithms, one generally wants to place a new metabolite pattern into a class that one has already encountered. For example "Does this serum sample come from someone who has prostate cancer or not?" For these cases there must be some gold standard data of metabolic profiles taken from patients with prostate cancer and healthy matched controls. With supervised learning, then, one knows the desired responses (*y*-data, or traits or classes; in this case prostate cancer or not) associated with each of the metabolite data inputs (*x*-data). The goal of supervised analyses is to find a mathematical transformation (model) that will correctly associate all or some of the inputs with the target traits (16,33,34). This is usually achieved by minimizing the error between the known target and the model's response (output), with artificial neural networks (ANNs), discriminant algorithms (DAs), and partial least squares (PLS) currently being the most popular (33,35,36).

The example trait to be predicted above is categorical (e.g., disease vs. healthy), but one may of course want a quantitative prediction to be made. For example, the Gleason grade for prostate cancer is important, as this would dictate therapeutic action. ANNs and PLS are ideally suited for quantification (33,37) and so will also play a role for predictions of the severity of disease.

Finally with respect to supervised learning, there also exist special types of algorithms that affect explanatory analyses; that is to say, the mathematical transformation from input to output data is transparent. These inductive reasoning methods include decision tree approaches in which predictive segregation (branching) of the data produces decision boundaries allowing

the discovery of which metabolites are important. CART (classification and regression trees) (38), FuRES (fuzzy rule-building expert system) (39), and C4.5/C5 (40) are the most popular decision tree algorithms. In addition, there are a range of evolutionary computation (EC) algorithms (41) that also affect inductive reasoning, and include genetic algorithms (GAs) (42,43), genetic programming (GP) (44–47), and genomic computing (GC) (48). These are based on the concepts of Darwinian selection and are programmed to evolve the desired mapping between input and output variables; again these elucidate which metabolites are important.

All of the methods used must be validated properly, and this is an important step within good modeling practice (GMP) (49). Brown et al. (50) have recently described a metabolomics data analysis pipeline. In this process, one goes from the design of good experiments through instrumental optimization, data storage, and manipulation, the chemometric data-processing methods in common use, and the necessary means of validation and cross-validation for giving conclusions that are credible (50). This pipeline is likely to help in the validation of the model in terms of its biological relevance (especially when inductive algorithms are used), which can be tested by a complementary approach using transcriptomics and proteomics.

## Superorganisms and their biological complexity

Thanks to the genome sequence projects, the number of genes in the human has been calculated to be a modest 31,897 (http://eugenes.org/), a relatively small number when one considers that yeast has 7,547, thale cress (*Arabidopsis thaliana*) contains 29,388 ORFs, and a measly little worm (*Caenorhabditis elegans*) has 23,399. Indeed, *Mus musculus* (the mouse) has ~6,000 genes more than humans, and one may wonder if we really are that complex after all!

Gill et al. have used small subunit ribosomal DNA sequences to estimate that the human intestinal microflora are composed of between $10^{13}$ and $10^{14}$ microorganisms (10 times the number of our own cells, comprising >1000 bacterial species). The metagenome of this so-called microbiome has at least 100 times as many genes as our own genome (51). The microbiome can be thought of as an additional organ, which is estimated to weigh 1 kg in an adult human and is mutalistic to humans (and the commensal microflora that inhabit the host) (52). In mice and rats, it has been shown that *Lactobacillus* species have antidiabetic effects on noninsulin-dependent diabetes mellitus (53,54). Ordovas and Mooser have suggested that the microbiome may play an important role in maintaining human health (55), and it is likely therefore that the intestinal microflora (as well as those bacteria found on the skin, in airways, and in the urogenital tract) play a vital role in our well-being.

This interaction of the microbiome with humans suggests that the human be considered as a superorganism (56), where we are in fact a human-microbe hybrid (57). Eckburg et al. have shown that the diversity in the microbiome is huge and that there is significant intersubject variability (58). The womb is sterile, and so babies are born with gnotobiotic gastrointestinal tracts, and the new soon-to-be superorganism needs to acquire its commensal microflora (59). These bacteria are mainly maternally acquired, and this process is largely achieved in the first year of life. We may be born 100% human but will die 90% bacterial—a truly complex organism!

It was stated earlier that, from human genome sequence analyses, there are predicted to be 2645 metabolites in the human metabolic network (15). This number will inevitably need revising when one considers the likelihood that one will detect prokaryotic-derived metabolites in humans. Indeed, Nicholson et al. have detected microbial metabolites from the intestinal microflora in human serum and urine using nuclear magnetic resonance (NMR) spectroscopy (60). There are likely to be more examples forthcoming from these and other researchers, and this added layer of complexity needs careful consideration. **Figure 2** shows a metabolic network from a superorganism, where secreted metabolites from humans may be metabolically changed and reabsorbed across the intestinal cell wall. Thus, not only will environmental effects (*vide infra*) have impact on the hundreds of functionally specialized cell types found in humans (61), but understanding both the interactions among many different organisms in a single complex system and their perturbation to environmental changes is very important, especially when this is highly person specific (58).

## Nutrigenomics and metabolomics

One environmental factor that is very important is what happens to the health of an individual when he or she eats. The link between nutrition and health is obvious, and the role of bioactive food components in the protection against disease is also well documented (e.g., 62–64).

The area of nutrigenomics has recently emerged where the aim is to generate a picture of how gene expression changes when a human is exposed to various nutrients (65–68). This approach may allow the discovery of bioprotective foods. Metabolomics is expected to play a pivotal role here (http://www.nugo.org/metabolomics) because nutrition by definition is aimed at maintaining cellular and organism homeostasis. In addition, it is known that some metabolic diseases can be prevented by nutrition (69), and food and drink contain many metabolites; thus, measuring metabolites directly would be sensible. Within this context nutrigenomics can be expanded and should include defining tissue, cellular, or biofluid-specific nutritional metabolomes (70–73).

The homeostasis of human metabolism is affected by a combination of intrinsic and extrinsic factors (74). The intrinsic

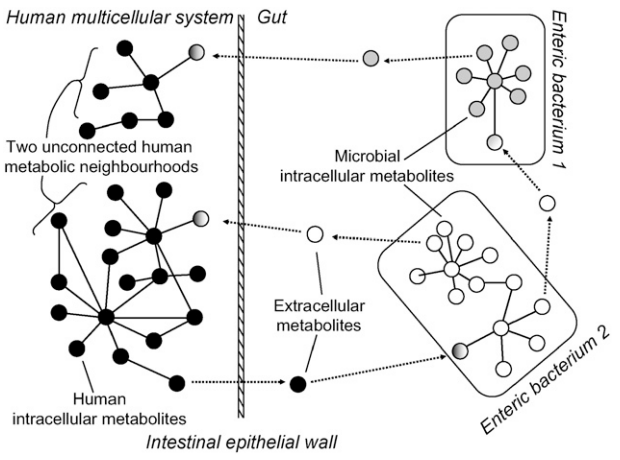


**Figure 2** A complex metabolic network from a superorganism showing metabolites derived from the enzymatic action of proteins encoded by genes in the human genome (*black circles*). One of these metabolites has been secreted into the gut, where it has been used as a substrate by a microorganism resident in the gut (*enteric bacterium 2*). This bacterium has metabolically transformed this metabolite (*white circles*) using its own microbially derived enzymes. Two of these products are secreted; 1 crosses the intestinal barrier and is used by the human, while the other is absorbed by a second enteric microbe (whose metabolites are represented by *gray circles*), leading to so called cross-feeding. Note in the schematic shown that areas of metabolism in humans that are not connected could become linked by microbial transformation.

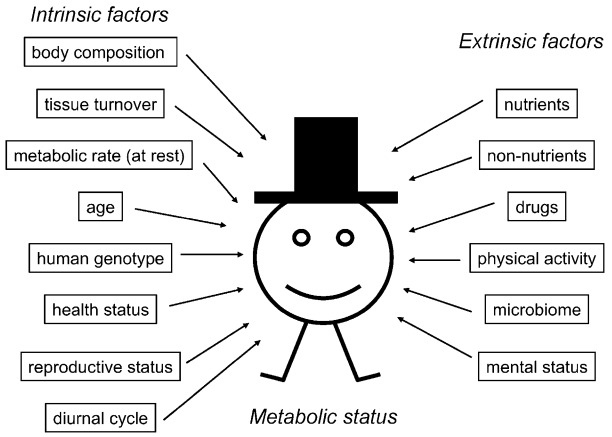


**Figure 3** Intrinsic and extrinsic factors that affect the metabolic status of the human. The status of these can be measured using metabolomics.

factors include body composition, tissue turnover, resting metabolic rate, age, human genotype, health status, reproductive status, and diurnal cycle. The extrinsic factors include diet (in terms of nutrients and nonnutrients), drugs (prescribed and lifestyle/leisure activities), physical activity, microbiome, and stress (**Fig. 3**). Metabolomics would aid in the interpretation of disease processes because a baseline healthy metabolome under different nutritional conditions could be defined leading to personalized metabolic assessment (75). Controlled perturbation from this baseline would lead to an understanding of how to keep human metabolism in homeostasis by tailoring nutritional intake (69).

### Concluding remarks

Metabolomics is the functional analysis method aimed at acquiring robust and reproducible quantitative information on intracellular and extracellular metabolites. It is gaining increasing interest across a wide variety of disciplines including functional genomics, integrative and systems biology, nutrigenomics, pharmacogenomics, and biomarker discovery for disease prognoses, diagnoses, and therapy monitoring.

Man can be thought of as a superorganism, and the health of this human-microbe hybrid will be affected by intrinsic properties such as human genetics, diurnal cycles, and age and by extrinsic factors such as lifestyle choices (food, drink, and drug intake) and the acquisition of a stable "healthy" gut microbiome. Alterations in this superorganism will be manifested in the metabolite complement as revealed by its serum and urine samples. The unraveling of this metabolic compartmentalization in this complex ecosystem will certainly be a challenge for systems biology approaches and necessary for defining human health at the molecular level. Understanding how to keep human metabolism in healthy homeostasis will lead to lifestyle choices regarding nutritional intake (you are what you eat!) and will allow the tailoring of immediate and future personalized healthcare.

## Literature Cited


1. The –International –Human Genome-Mapping –Consortium. A physical map of the human genome. Nature. 2001;409:934–41.
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, et al. The sequence of the human genome. Science. 2001;291:1304–51.
3. Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, et al. The complete genome sequence of *Escherichia coli* K-12. Science. 1997;277:1453–74.
4. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, et al. Environmental genome shotgun sequencing of the Sargasso Sea. Science. 2004;304:66–74.
5. Cornish-Bowden A, Cárdenas ML. Complex networks of interactions connect genes to phenotypes. Trends Biochem Sci. 2001;26:463–5.
6. Barabasi A-L, Oltvai ZN. Network biology: understanding the cell's functional organisation. Nat Rev Genet. 2004;5:101–13.
7. Fell DA. Understanding the control of metabolism. London: Portland Press; 1996.
8. Westerhoff HV, Palsson BO. The evolution of molecular biology into systems biology. Nat Biotechnol. 2004;22:1249–52.
9. Kell DB. Metabolomics and systems biology: making sense of the soup. Curr Opin Microbiol. 2004;7:296–307.
10. van der Greef J. The role of analytical sciences in medical systems biology. Curr Opin Chem Biol. 2004;8:559–65.
11. Kell DB, Westerhoff HV. Metabolic control theory: its role in microbiology and biotechnology. FEMS Microbiol Revs. 1986;39:305–20.
12. Raamsdonk LM, Teusink B, Broadhurst D, Zhang NS, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, et al. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat Biotechnol. 2001;19:45–50.
13. Urbanczyk-Wochniak E, Luedemann A, Kopka J, Selbig J, Roessner-Tunali U, Willmitzer L, Fernie AR. Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. EMBO Rep. 2003;4:989–93.
14. Oh JE, Krapfenbauer K, Fountoulakis M, Frischer T, Lubec G. Evidence for the existence of hypothetical proteins in human bronchial epithelial, fibroblast, amnion, lymphocyte, mesothelial and kidney cell lines. Amino Acids. 2004;26:9–18.
15. Palsson BO. The Human Metabolic Network. Consortium for Post Genomic Science, Manchester, UK; 2006.
16. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. Metabolomics by numbers—acquiring and understanding global metabolite data. Trends Biotechnol. 2004;22:245–52.
17. Dunn WB, Bailey NJC, Johnson HE. Measuring the metabolome: current analytical technologies. Analyst. 2005;130:606–25.
18. Dunn WB, Ellis DI. Metabolomics: Current analytical platforms and methodologies. Trends Anal Chem. 2005;24:285–94.
19. Weckwerth W, Morgenthal K. Metabolomics:from pattern recognition to biological interpretation. Drug Discov Today. 2005;10:1551–8.
20. Hall RD. Plant metabolomics: from holistic hope, to hype, to hot topic. New Phytol. 2006;169:453–68.
21. Hollywood K, Brison DR, Goodacre R. Metabolomics: current technologies and future trends. Proteomics. 2006;6:4716–23.
22. de la Fuente A, Snoep JL, Westerhoff HV, Mendes P. Metabolic control in integrated biochemical system. Eur J Biochem. 2002;269:4399–408.
23. Tweeddale H, Notley-McRobb L, Ferenci T. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("Metabolome") analysis. J Bacteriol. 1998;180:5109–16.
24. Viant MR, Bundy JG, Pincetich CA, de Ropp JS, Tjeerdema RS. NMR-derived developmental metabolic trajectories: an approach for visualizing the toxic actions of trichloroethylene during embryogenesis. Metabolomics. 2005;1:149–58.
25. Villas-Boas SG, Højer-Pedersen J, Akesson M, Smedsgaard J, Nielsen J. Global metabolite analysis of yeast: evaluation of sample preparation methods. Yeast. 2005;22:1155–69.
26. Folch J, Lees M, Sloane Stanley GH. A simple method for the isolation and purification of total lipids from animal tissues. J Biol Chem. 1957;226:497–509.
27. Fiehn O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. Comp Funct Genomics. 2001;2:155–68.
28. Harrigan GG, Goodacre R, editors. Metabolic profiling: its role in biomarker discovery and gene function analysis. Boston: Kluwer Academic Publishers; 2003.
29. Mendes P. Emerging bioinformatics for the metabolome. Brief Bioinform. 2002;3:134–45.
30. Chatfield C, Collins AJ. Introduction to multivariate analysis. London: Chapman & Hall; 1980.

31. Duda RO, Hart PE, Stork DE. Pattern classification, 2nd ed. London: John Wiley; 2001.
32. Krzanowski WJ. Principles of multivariate analysis: a user's perspective. Oxford: Oxford Univeristy Press; 1988.
33. Martens H, Næs T. Multivariate calibration. Chichester: John Wiley; 1989.
34. Beavis RC, Colby SM, Goodacre R, Harrington PB, Reilly JP, Sokolow S, Wilkerson CW. Artificial intelligence and expert systems in mass spectrometry. In: Meyers RA, editor. Encyclopedia of analytical chemistry. 2000. Chichester: John Wiley & Sons. p. 11558–97.
35. Manly BFJ. Multivariate statistical methods: a primer. London: Chapman & Hall; 1994.
36. Bishop CM. Neural networks for pattern recognition. Oxford: Clarendon Press; 1995.
37. Goodacre R, Neal MJ, Kell DB. Rapid and quantitative analysis of the pyrolysis mass spectra of complex binary and tertiary mixtures using multivariate calibration and artificial neural networks. Anal Chem. 1994;66:1070–85.
38. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Pacific Grove, CA: Wadsworth; 1984.
39. Harrington PB. Fuzzy rule-building expert systems: minimal neural networks. J Chemom. 1991;5:467–86.
40. Quinlan JR. C4.5: programs for machine learning. San Mateo, CA: Morgan Kaufmann; 1993.
41. Bäck T, Fogel DB, Michalewicz Z. Handbook of evolutionary computation. Oxford: IOPPublishing/Oxford University Press; 1997.
42. Holland JH. Adaption in natural and artifcial systems, 2 ed. Cambridge, MA: MIT Press; 1992.
43. Reeves CR. Genetic algorithms—principles and perspectives: a guide to GA theory. Dordrecht: Kluwer Academic Publishers; 2002.
44. Koza JR. Genetic programming: on the programming of computers by means of natural selection. Cambridge, MA: MIT Press; 1992.
45. Koza JR. Genetic programming II: automatic discovery of reusable programs. Cambridge, MA: MIT Press; 1994.
46. Koza JR, Bennett FH, Keane MA, Andre D. Genetic programming III: Darwinian invention and problem solving. San Francisco: Morgan Kaufmannn; 1999.
47. Koza JR, Keane MA, Streeter MJ, Mydlowec W, Yu J, Lanza G. Genetic programming: routine human-competitive machine intelligence. New York: Kluwer; 2003.
48. Kell DB, Darby RM, Draper J. Genomic computing. Explanatory analysis of plant expression profiling data using machine learning. Plant Physiol. 2001;126:943–51.
49. Kell DB, Sonnleitner B. GMP—Good modelling practice: an essential component of good manufacturing practice. Trends Biotechnol. 1995;13:481–92.
50. Brown M, Dunn WB, Ellis DI, Goodacre R, Handl J, Knowles JD, O'Hagan S, Spasic I, Kell DB. A metabolome pipeline: from concept to data to knowledge. Metabolomics. 2005;1:39–51.
51. Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE. Metagenomic analysis of the human distal gut microbiome. Science. 2006;312:1355–9.
52. Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. Science. 2005;307:1915–20.
53. Matsuzaki T, Yamazaki R, Hashimoto S, Yokokura T. Antidiabetic effects of an oral administration of *Lactobacillus casei* in a non-insulin-dependent diabetes mellitus (NIDDM) model using KK-A(y) mice. Endocr J. 1997;44:357–65.
54. Yadav H, Jain S, Sinha PR. Effect of dahi containing *Lactococcus lactis* on the progression of diabetes induced by a high-fructose diet in rats. Biosci Biotechnol Biochem. 2006;70:1255–8.
55. Ordovas JM, Mooser V. Metagenomics: the role of the microbiome in cardiovascular diseases. Curr Opin Lipidol. 2006;17:157–61.
56. Wilson DS, Sober E. Reviving the Superorganism. J Theor Biol. 1989;136:337–56.
57. Sekirov I, Finlay BB. Human and microbe: united we stand. Nat Med. 2006;12:736–7.
58. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. Diversity of the human intestinal microbial flora. Science. 2005;308:1635–8.
59. Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell. 2006;124:837–48.
60. Nicholson JK, Holmes E, Lindon JC, Wilson ID. The challenges of modeling mammalian biocomplexity. Nat Biotechnol. 2004;22:1268–74.
61. Nicholson JK, Wilson ID. Understanding 'global' systems biology: Metabonomics and the continuum of metabolism. Nat Rev Drug Discov. 2003;2:668–76.
62. Afman L, Muller M. Nutrigenomics: From molecular nutrition to prevention of disease. J Am Diet Assoc. 2006;106:569–76.
63. Davis CD, Milner J. Frontiers in nutrigenomics, proteomics, metabolomics and cancer prevention. Mutat Res. 2004;551:51–64.
64. Ferguson LR. Nutrigenomics—Integrating genomic approaches into nutrition research. Mol Diagn Ther. 2006;10:101–8.
65. Mariman ECM. Nutrigenomics and nutrigenetics: the 'omics' revolution in nutritional science. Biotechnol Appl Biochem. 2006;44:119–28.
66. Mutch DM, Wahli W, Williamson G. Nutrigenomics and nutrigenetics: the emerging faces of nutrition. FASEB J. 2005;19:1602–16.
67. Ordovas JM, Mooser V. Nutrigenomics and nutrigenetics. Curr Opin Lipidol. 2004;15:101–8.
68. Roche HM. Nutrigenomics—new approaches for human nutrition research. J Sci Food Agric. 2006;86:1156–63.
69. Trujillo E, Davis C, Milner J. Nutrigenomics, proteomics, metabolomics, and the practice of dietetics. J Am Diet Assoc. 2006;106:403–13.
70. van Ommen B, Stierum R. Nutrigenomics: exploiting systems biology in nutrition and health arena. Curr Opin Biotechnol. 2002;13:517–21.
71. van Ommen B. Nutrigenomics: Exploiting systems biology in the nutrition and health arenas. Nutrition. 2004;20:4–8.
72. Kaput J, Rodriguez RL. Nutritional genomics: the next frontier in the postgenomic era. Physiol Genomics. 2004;16:166–77.
73. Zeisel SH, Freake HC, Bauman DE, Bier DM, Burrin DG, German JB, Klein S, Marquis GS, Milner JA, et al. The nutritional phenotype in the age of metabolomics. J Nutr. 2005;135:1613–6.
74. Gibney MJ, Walsh M, Brennan L, Roche HM, German B, van Ommen B. Metabolomics in human nutrition: opportunities and challenges. Am J Clin Nutr. 2005;82:497–503.
75. German JB, Hammock BD, Watkins SM. Metabolomics: Building on a century of biochemistry to guide human health. Metabolomics. 2005;1:3–9.
76. Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey RN, Willmitzer L. Metabolite profiling for plant functional genomics. Nat Biotechnol. 2000;18:1157–61.
77. Buchholz A, Takors R, Wandrey C. Quantification of intracellular metabolites in *Escherichia coli* K12 using liquid chromatographic-electrospray ionization tandem mass spectrometric techniques. Anal Biochem. 2001;295:129–37.
78. Wilson ID, Plumb R, Granger J, Major H, Williams R, Lenz EM. HPLC-MS-based methods for the study of metabonomics. J. Chromatogr B. 2005;817:67–76.
79. Tolstikov VV, Lommen A, Nakanishi K, Tanaka N, Fiehn O. Monolithic silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics. Anal Chem. 2003;75:6737–40.
80. Soga T, Ueno Y, Naraoka H, Ohashi Y, Tomita M, Nishioka T. Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. Anal Chem. 2002;74:2233–9.
81. Soga T, Ohashi Y, Ueno Y, Naraoka H, Tomita M, Nishioka T. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. J Proteome Res. 2003;2:488–94.
82. Wolfender JL, Ndjoko K, Hostettmann K. Liquid chromatography with ultraviolet absorbance-mass spectrometric detection and with nuclear magnetic resonance spectroscopy: a powerful combination for the on-line structural investigation of plant metabolites. J Chromatogr A. 2003;1000:437–55.
83. Gamache PH. Metabolomic applications of electrochemistry/ mass spectrometry. J Am Soc Mass Spectrom. 2004;15:1717–26.
84. Welthagen W, Shellie RA, Spranger JM, Ristow M, Zimmermann R, Fiehn O. Comprehensive two dimensional gas chromatography—time of flight mass spectrometry (GCxGC-TOF) for high resolution

metabolomics: Biomarker discovery on spleen tissue extracts of obese NZO compared to lean C57BL/6 mice. Metabolomics. 2005;1:65–73.
85. Aharoni A, Ric de Vos CH, Verhoeven HA, Maliepaard CA, Kruppa G, Bino R, Goodenowe DB. Nontargeted metabolome analysis by use of Fourier transform ion cyclotron mass spectrometry. OMICS. 2002;6:217–34.
86. Satake M, Dmochowska B, Nishikawa Y, Madaj J, Xue J, Guo Z, Reddy DV, Rinaldi PL, Monnier VM. Vitamin C metabolomic mapping in the lens with 6-deoxy-6-fluoro-ascorbic acid and high-resolution $^{19}$F-NMR spectroscopy. Invest Ophthalmol Vis Sci. 2003; 44:2047–58.
87. Bederman IR, Reszko AE, Kasumov T, David F, Wasserman DH, Kelleher JK, Brunengraber H. Zonation of labeling of lipogenic acetyl-CoA across the liver: implications for studies of lipogenesis by mass isotopomer analysis. J Biol Chem. 2004;279:43207–16.
88. Villas-Bôas SG, Mas S, Åkesson M, Smedsgaard J, Nielsen J. Mass spectrometry in metabolome analysis. Mass Spectrom Rev. 2005;24:613–46.
89. Boros LG. Metabolic targeted therapy of cancer: current tracer technologies and future drug design strategies in the old metabolic network. Metabolomics. 2005;1:11–5.
90. Allen J, Davey HM, Broadhurst D, Heald JK, Rowland JJ, Oliver SG, Kell DB. High-throughput classification of yeast mutants for functional genomics via metabolic footprinting. Nat Biotechnol. 2003; 21:692–6.
91. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999;29:1181–9.
92. Vaidyanathan S, Kell DB, Goodacre R. Flow-injection electrospray ionization mass spectrometry of crude cell extracts for high-throughput bacterial identification. J Am Soc Mass Spectrom. 2002;13:118–28.
93. Overy SA, Walker HJ, Malone S, Howard TP, Baxter CL, Sweetlove LJ, Hill SA, Quick WP. Application of metabolite profiling to the identification of traits in a population of tomato introgression lines. J Exp Bot. 2005;56:287–96.
94. Vaidyanathan S, Jones D, Broadhurst DI, Ellis J, Jenkins T, Dunn WB, Hayes A, Burton N, Oliver SG, et al. A laser desorption ionisation mass spectrometry approach for high throughput metabolomics. Metabolomics. 2005;1:243–50.
95. Vaidyanathan S, Gaskell SJ, Goodacre R. Matrix suppressed laser desorption ionization mass spectrometry and its suitability for metabolome analyses. Rapid Commun Mass Spectrom. 2006; 20:1192–8.
96. Harrigan GG, LaPlante RH, Cosma GN, Cockerell G, Goodacre R, Maddox JF, Ludyenck JP, Ganey PE, Roth RA. Application of high-throughput Fourier-transform infrared spectroscopy in toxicology studies: contribution to a study on the development of an animal model for idiosyncratic toxicity. Toxicol Lett. 2004;146:197–205.
97. Ellis DI, Goodacre R. Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. Analyst. 2006;131:875–85.
98. Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, et al. Potential of metabolomics as a functional genomics tool. Trends Plant Sci. 2004;9:418–25.
99. Jenkins H, Hardy N, Beckmann M, Draper J, Smith AR, Taylor J, Fiehn O, Goodacre R, Bino R, et al. A proposed framework for the description of plant metabolomics experiments and their results. Nat Biotechnol. 2004;22:1601–6.
100. Lindon JC, Nicholson JK, Holmes E, Keun HC, Craig A, Pearce JT, Bruce SJ, Hardy N, Sansone SA, et al. Summary recommendations for standardization and reporting of metabolic analyses. Nat Biotechnol. 2005;23:833–8.
101. Spasic I, Dunn W, Velarde G, Tseng A, Jenkins H, Hardy NW, Oliver SG, Kell DB. MeMo: a hybrid SQL/XML approach to metabolomic data management for functional genomics. BMC Bioinformatics. 2006; 7:281.
102. Castle AL, Fiehn O, Kaddurah-Daouk R., Lindon JC. Metabolomics Standards Workshop and the development of international standards for reporting metabolomics experimental results. Briefings in Bioinformatics. 2006;7:159–65.