Review

What mRNA Abundances Can Tell us about Metabolism

Andreas Hoppe

Institute for Biochemistry, Charité University Medicine Berlin, Charitéplatz 1, Berlin 10117, Germany;
E-Mail: hoppe@bioinformatics.org; Tel.: +49-30-450-528176

Abstract: Inferring decreased or increased metabolic functions from transcript profiles is at first sight a bold and speculative attempt because of the functional layers in between: proteins, enzymatic activities, and reaction fluxes. However, the growing interest in this field can easily be explained by two facts: the high quality of genome-scale metabolic network reconstructions and the highly developed technology to obtain genome-covering RNA profiles. Here, an overview of important algorithmic approaches is given by means of criteria by which published procedures can be classified. The frontiers of the methods are sketched and critical voices are being heard. Finally, an outlook for the prospects of the field is given.

Keywords: transcriptomics; metabolic network; metabolism; metabolic function; Virtual-Liver Network

1. Introduction

Genetic regulation is a major control mechanism of the activity of the cell's metabolic functions, especially in the frame of longer times where the metabolic function is described as a specific metabolic input/output behavior of the cell. Its activity is defined as the metabolic flux, *i.e.*, the consumption and production rate of specific metabolites, related to this function. A general description of the flow of information from the genome to metabolism is as follows: the process is initiated by transcription factors, RNA polymerase transcribes genes into RNA, RNA is transported to the ribosome and translated into proteins, and after folding, post-transcriptional modification, and transport to the site of action, proteins act as enzymes and transporters catalyzing biochemical reactions fluxes of the molecules in the cell (quantified by the reaction *flux*, the net number of converted molecules by time by cell volume). The mechanisms of this control have already been recognized as being complex and multi-level—setting

up predictive quantitative models is difficult [1]. Transcription factors are controlled by mechanisms on different layers of the cellular system: their own transcription, translation, and post-translational modification, their localization, their activation by external signaling substances or the concentration of internal metabolites, and their combination with other transcription factors. Some mRNA species [2] and transcription factors [3] are directionally transported by microtubuli in a controlled manner while others rely on diffusion. The efficiency of the translation may dramatically differ between different genes [4]. Finally, the catalytic efficiency of different enzymes varies along six orders of magnitude [5]. The metabolic flux rate is not only determined by the enzymes' concentrations but by a multitude of regulators, some of which change the reaction rate by several orders of magnitude [5]. For many processes the modifying factors have been discovered but on the scale of the whole genome, most of them are unknown.

It might seem presumptuous to propose that transcript data can be used to predict metabolic functions. However, these predictions have received much interest because:

- the layer of RNA transcripts (as opposed to the layer of proteins, the layer of reaction fluxes, and the layer of metabolites) is the only layer where a complete quantitative snapshot of all molecular species is currently feasible. Reaction flux estimations currently cover only a tiny share of all reactions. Metabolite concentrations are currently measured for some 100 s metabolites but specific classes of metabolites such as lipids still present large challenges. Protein amount estimations at the genome-scale are now being done but the effort necessary is huge. The layer of DNA, whose information is a precondition for any transcription-based analysis, is not mentioned as qualitative information.

- transcript arrays are moderately priced in relation to the amount of data gathered,

- the experimental effort for the researcher is moderate due to an highly automatized process,

- the technology provides a low ambiguity and accurate estimates of the RNA amount changes [6]. The high number of probes allows to distinguish between the RNA of separate genes, with only few exceptions. Ambiguity of the peaks is the main problem of the estimation of metabolite concentrations by mass spectrometry. Ambiguity is also the largest challenge in flux estimations based on ^{13}C marked substrates, and

- well-curated genome-scale reconstructions of the metabolic networks are available [7–10].

The measurement of reaction fluxes, metabolite concentrations, enzyme activities, and protein amounts are currently undertaken for a subset of all molecular species. The measurement of protein amounts is just becoming feasible with the advent of techniques such as single-shot ultra HPLC [11]. *If* all enzyme activities and metabolite concentrations *were* available, a much more accurate prediction would be possible, but that is not the case on the large scale. Thus, to judge the results of the reviewed studies squarely it must be stressed that the *expectations must be lowered* accordingly. The systems biologist faces the trade-off between coverage versus accuracy versus the data being *closer* to the enzymatic activity, *i.e.*, that quantitative proteomics data would provide a better indication of enzymatic activity, but the technique does not have the coverage provided by transcriptome data.

Here, studies with the primary focus on metabolism are reviewed. Other major areas of application of transcript data [12] are not covered, such as (i) detection of transcriptional co-regulation leading to (ii) detection of transcription factor binding sites [13] and (iii) transcriptional biomarkers [14].

2. Fundamental Studies

To demonstrate the difficulty of the task bridging several layers of cellular interaction, selected studies of the relation from one layer to the next will be sketched.

2.1. Gene Chip Intensities→mRNA

DNA microarray read-outs depend on the RNA concentrations but also on the varying affinity of the RNA to the probes, which is unknown on the large scale, thus special care is needed when analyzing the data [15]. Nevertheless, it is a very dependable technique [6]. For a comparison of different gene chip techniques, see Baldwin *et al.* [16]. Often, a genome-scale gene chip analysis is coupled with a more accurate qPCR for selected genes as a means of validation [17]. Advanced experimental techniques such as RNA-seq [18] and SAGE [19] allow a more accurate genome-scale quantification of RNA than gene array readouts and will eventually replace them [20], but effort and price currently restrict its widespread use [21].

2.2. mRNA→Protein

In a pioneering study by Gygi *et al.* [22], the correlation between 106 studied mRNA levels to their coded protein levels showed a high value of 0.935. Gygi noted that the number is far lower if the extremely highly abundant proteins are disregarded; then it can be as low as 0.1. In a further note, the relation of protein levels below the detection limit to their respective mRNA levels is obviously unknown, thus, for the numerous proteins that only occur in very small quantities, the relation to RNA levels is unknown. In a subsequent study Griffin *et al.* advocates the combined consideration of both mRNA and protein levels to understand the regulation of central metabolic functions in yeast [23,24]. Tuller *et al.* predicted protein abundances from mRNA expression levels by taking into account additional information on the genes [25]. The results on the test set showed a good correlation of 0.76. Further studies on the relation of RNA levels and protein abundances have been reviewed by Meier *et al.* [26]. In particular the study in human cell lines [1] should be mentioned. In an experimental analysis of Arabidopsis, among 319 protein/transcript pairs, 56% showed concurrence between transcript and protein, and it was suggested that for the others post-transcriptional modification takes place [27].

Mechanistically, the relation between RNA and protein concentrations can be seen as the interplay of three aspects: (i) the life span of RNA and (ii) proteins as well as (iii) the translation efficiency at the ribosome [4]. In a groundbreaking work, Schwanhäusser studied the life cycles of RNA and protein translation in mammalian fibroblasts and found "that the cellular abundance of proteins is predominantly controlled at the level of translation" [28]. The rates of mRNA synthesis and decay in yeast in response to stress have been measured [29] The life span of proteins *in vivo* has been assessed on a large scale in yeast [30] and for selected glycolytic enzymes in mammalian cells [31].

2.3. Enzyme Concentration → Enzyme Activity

The enzyme activity (the maximal catalytic rate $v_{\max}$ for a given cell volume) depends on the enzyme concentration. Mostly, the relation is approximately linear in a predefined environment—the ratio is called turnover number. The turnover numbers of enzymes (together with other kinetic parameters) have been estimated for many enzymes, comprehensively reviewed and made available in public databases [5,32]. With respect to the set of all enzymes, this information is far from complete. Turnover numbers have been measured for different conditions (pH, temperature, and the concentrations of activators and inhibitors) and the resulting values vary considerably for one enzyme. Some enzymes are nine orders of magnitude more efficient than others (minimal *vs.* maximal turnover numbers in [5]). Considering this data, the variability of this step in the chain from RNA to metabolic flux is greater than of any of the other steps.

2.4. Enzyme Activity → Metabolic Flux

The prediction of metabolic fluxes from enzyme activity information (and concentration of reactants, products, and other metabolic species) has been extensively studied in the field of kinetic modeling and its results are available in public databases [33,34]. A main challenge in the understanding is the interplay of metabolite concentration, enzyme levels, and reaction fluxes in a highly connected network. The network effect, defined as the difference of the simultaneous flow of chemical reactions compared with the isolated flow of reactions, modifies the activity–flux relation. It is studied in metabolic control analysis [35–37]. In extreme cases, it can lead to paradoxical situations where an increased enzyme amount leads to a lower flux in the same metabolic reaction.

2.5. Crossing Several Layers

Hancock *et al.* analyzed the relation of RNA abundance to metabolite concentrations in combination with the topological structure of the network. Based on clustering of correlated genes, their approach allows the identification of hub reactions depending on a specific change of condition, which subsequently leads to a minimal set of commonly controlled metabolites. Their results support the hypothesis that the gene expression response (on different forms of stress on *E. coli* in this case) targets a small number of metabolites which consequently entails a large-scale change in the metabolism [38].

Kharchenko *et al.* found that the highest co-expression of metabolic genes is arranged in simple motifs in the metabolic network, in other words, "regulation of metabolic genes is local" [39]. Cakir *et al.* studied the transcriptional adaption of yeast on growth media. They calculate optimal transcript ratios on the basis of elementary flux modes [40] and the comparison to real transcript ratios showed a high agreement [41]. This result, in comparison with other studies showing less agreement, leads to the conclusion that the adaptation on cellular substrates is a distinguished case.

Hajduch *et al.* compared the proteome of different oilseed to reveal differences in the intermediary metabolism, and their analysis showed a diverging use of malate as a precursor for lipids [42]. Saito *et al.* reviewed studies using transcript and metabolite co-occurrence for various applications in

plant biology [43]. Ishihama *et al.* performed a large-scale proteomic screening of *E. coli* and found that, among the enzymes, only proteins involved in energy metabolism are highly abundant [44].

Of particular interest are studies which measured RNA, protein, fluxes, and metabolite concentrations in parallel in the same experiment [45,46]. The common finding in these studies is that there is not a high overall correlation between the abundance of RNA and the coded protein, between the enzyme and the catalyzed flux, and between the metabolite concentrations and the level of enzymes that catalyze them. However, looking at the regulation of selected metabolic paths and functions, in almost all cases the pattern of abundance changes of RNA and protein is in accordance with the observed changes in reaction fluxes and metabolite concentrations. To sum it up, although there is little direct predictivity of RNA to the reaction fluxes, the transcriptional regulation of the metabolic function can still be observed in the RNA abundance data.

2.6. mRNA → Fluxes

As a summary of an early attempt to relate transcript values to metabolic fluxes, ter Kuile expressed "strong doubts on whether transcriptome and proteome analysis suffices to assess biological function" [47]. The conclusion has been drawn by the authors of subsequent approaches that transcript profiles must be used in conjunction with other information to yield meaningful results.

Moxley *et al.* [48] correlated the fluxes (estimated by tracer experiments) to the respective RNA levels and found a mere correlation of 0.07, which could be increased to 0.8 by the use of a network-based model from which a parameter called "metabolite interaction density" is calculated. This density is used as a modifier for the flux prediction from RNA levels. The conclusion of this study is that the consideration of the metabolic network is *essential* to draw a predictive relation from transcript abundances to fluxes.

Yang *et al.* studied gene expression in Synechocystis in combination with ^{13}C isotope-based flux measurements and emphasizes the importance of integrating transcript and flux data for the understanding of regulatory mechanisms [49].

Daran-Lapujade *et al.* studied the role of "hierarchical" flux regulation (by changed enzyme activity, e.g., transcriptional regulation) versus metabolic regulation (change of flux due to changed metabolite concentrations) for glycolytic enzymes in yeast [50]. Factor analysis showed that transcriptional regulation was only responsible for 20%–50% of the observed flux changes. A similar analysis [51] led to the assignment of different *roles* to the regulated enzymes in glycolysis in yeast: regulation of some is predominately hierarchical, for others it is metabolically. For some, the regulation is cooperative between both, and for others it is antagonistic. In an earlier study they compared other central metabolism pathways and found strong qualitative correspondence between transcript and flux changes for the maltose metabolism, partial correspondence for triose-phosphate cycle and pentose-phosphate pathway, and little correspondence for glycolysis [52]. Their results put the prediction methods reviewed in the next chapter into perspective. However, glycolysis is a quite special pathway due to the large enzyme concentrations. Its fast response (for instance, to the sudden loss of membrane potential due to a rupture) is absolutely necessary as ATP depletion leads to rapid cell death. The transcriptional regulation is too slow for this life-saving response. Furthermore, the rapid growth of yeast on a glucose-rich media is an

extreme condition rarely found *in vivo*, thus, it is likely that the structure of the metabolic system is not optimized to this situation. So their findings regarding glycolysis do not seem to be sufficient to discard the idea of observing metabolic changes from transcript data for the entirety of the metabolism.

2.7. Regulation of Metabolic Genes

Which metabolic genes are regulated at all? Wessely *et al.* analyzed transcript profiles of *E. coli* and found that pathways (*i.e.*, the set of biochemical reactions necessary to perform a specific metabolic conversion) associated with high protein cost are "controlled by fine-tuned transcriptional programs" and those with low protein cost are only regulated in key reactions [53].

And how are the genes (resp. transcription factors) controlled? In the transcription factor network of *E. coli*, a hierarchy of general and specific transcription factors has been found, and each metabolic function is controlled by a distinct combination of them. Enzymes catalyzing sequential reactions are co-regulated by the same transcription factors, while the regulation at junctions in the metabolic network is more complex [54]. An interesting fact has been found by Notebaart *et al.* which provides an argument to analyze a metabolic network with respect to metabolic functions and not the graph structure alone: "The co-regulation of metabolic genes is better explained by flux coupling than by network distance" in *E. coli* [55].

2.8. Genetic Interactions

In the studies reviewed so far, the focus was the correlation between an individual RNA and the protein, flux, or concentrations. There is also another form of interactions called epistasis, which has also been modeled in the context of metabolic networks. An epistatic interaction occurs if the phenotypic impact of the knockout of one gene depends on the knockout of another gene [56]. Such an interaction might be caused by redundant reaction paths in the metabolic network in which case it can be predicted by network-based approaches [57–60]. One common finding is that most epistatic interactions are restricted to certain conditions [57,59]. Potentially, the verified set of epistatic interactions can be used for the more accurate interpretation of transcript profiles. Szappanos *et al.* studied genetic interactions for the metabolic genes using the flux-balance framework for yeast [61]. They found many "instances of genetic interactions ... not explained by the structure of the metabolic network", indicating that this is one more complicating factor that has to be taken into account for the mechanistic description of the transcriptional regulation of the metabolism.

3. Systematic Comparison of Methods

To systematically assess the multitude of studies relating RNA profiles to the metabolism, criteria will be given to distinguish how the profiles are used.

3.1. Absolute/Relative/Coexpression

Expression profiles can either be used in several ways. (i) Expression profiles can be directly used to assess a single state, which is called *absolute*, e.g., to decide whether a gene is active [62]. (ii) Differential expression profiles can be used to differentiate between states (changed conditions, time series)—normally logarithmic expression values are subtracted, which is called *relative*, e.g., to quantify changed metabolic activities [63]. (iii) A third alternative is to analyze the correlation of expression changes for each pair of genes, called *co-expression*, e.g., to assert which metabolic paths are controlled concertedly [64].

Absolute expression profiles are widely used to predict the active regions in metabolic networks [62,65–69]. Absolute expression profiles are also used for network reconstruction [7,8,70]: if a particular gene is expressed in at least one of a large number of expression profiles in a particular cell type, then the reaction catalyzed by or the transport process facilitated by its gene product can be considered as a part of the network [71].

Relative expression profiles are often simply analyzed by counting the number of up- or down-regulated genes using a threshold on the ratio (e.g., more than 2-fold change) or the significance level (e.g., using t-test) with respect to classifications such as gene ontology [72] or KEGG maps [73]. However, a quantitative prediction of the change of the metabolic mode of operation has also been demonstrated [63,74]. To cope with the non-linear relationship of transcript change and enzyme activity change, a ranking approach called Differential Rank Conservation (DIRAC) has been successfully applied [75].

Expression correlations are used to determine which genes are commonly regulated, for instance to predict transcription factors. Metabolic pathways with a high correlation of genes coding the necessary enzymes can be considered as a functional mode of operation in a particular cell type [64,76,77]. Ihmels *et al.* analyzed the co-expression of genes coding enzymes and found higher correlations along linear reaction paths between branch points and a hierarchical modularity of the regulation [78]. Loraine demonstrates the use of the gene clustering tool CressExpress for metabolic genes [79].

3.2. Thresholds

The distinction between active and inactive genes is crucial for all methods using *absolute* expression profiles.

Hebenstreit *et al.* gave clear evidence that in reality there is a clear distinction between genes which are expressed and those which are not expressed (in the sense that the gene product is present in sufficient abundance to take effect in the cell) [80]. The observable concentrations of RNA species is distributed in a bimodal distribution reflecting a normal distribution for both expressed and non-expressed genes. To decide whether a gene is considered active or not, a *threshold* is the method of choice. As there is an overlap of RNA abundance levels of *inactive* and *active* genes [80], methods applying the threshold must be robust enough to cope with a certain fraction of wrongly assigned activities. This robustness will also allow the use of transcript data which are not accurately representing RNA counts. Based on a comprehensive analysis of a gene chip in conjunction with proteomics data, an optimal threshold can

be calculated. However, mostly such experiments are considered too elaborate and the threshold is set heuristically. Instead, the approach is validated by the overall predictivity.

The negative effect of uncertainty of the optimal threshold value is reduced by its "soft" application. For instance, in the GIMME algorithm [62] the threshold is applied in such a way that an expression below the threshold entails a gradual (linear) penalty for an activity of the assigned reaction. Thus, a reaction assigned to a gene expressed at a lower level than the threshold can still be considered active but the total amount of these errors is minimized. In the iMAT approach [67,68,81] the threshold application is softened by the introduction of two threshold values. The upper threshold separates the genes highly likely to be active while the lower threshold separates the genes highly likely to be inactive, leaving a range of expression values without a clear attribution. As there is still no guarantee to avoid incorrect gene assignment, an optimization is used where the *clearly active* genes receive a bonus, the *clearly inactive* genes a penalty. More sophisticated is the MADE approach [74] that avoids the arbitrariness of the heuristic threshold setting. For each gene, a single but flexible threshold is calculated from a set of expression profiles by identifying the largest gap of values.

In other approaches, the setting of a threshold is completely circumvented and the expression values are used in a continuous way [69,82,83].

3.3. Representation of the Metabolic System

The way the metabolic system is represented is another important aspect of the methods. Mostly, the system is represented by the metabolic network which consists of the metabolites and the biochemical reactions which convert the metabolites in fixed quantities, the stoichiometric factors—thus, it is called stoichiometric model. Often a stoichiometric model is used to compute flux distributions in the flux-balance framework [62,65,69,74]. A different approach is to use metabolic paths (small linear chains of reactions) which do not necessarily form a complete network [64,84]. An alternative way to represent the metabolic system is to compute the set of elementary flux modes first [40] and perform the analysis using these flux distributions [41]. Also the decomposition of the total flux as convex sum minimal flux modes [85] parameterized by gene expression has been proposed [63]. The flux balance framework is equivalent to a bipartite graph (Petri net [86]) but also simplified graphs have been used such as the adjacency graph [87].

In a simplification of the stoichiometric model, the stoichiometric factors are ignored [88]. Hancock *et al*. use such a graph representation where the nodes are the metabolites, and for every biochemical reaction an edge is drawn from each substrate to each product [38].

3.4. Type of Inference

A flux distribution or an active subnetwork can be computed by penalizing fluxes belonging to *inactive* genes and/or bonusing nonzero fluxes belonging to *active* genes (in other words, the binary compliance to the expression profile) [62,67,81,89]. This approach has also been used as a secondary criterion in constrained flux-balance optimization [68]. Expression data has been used to define upper bounds on fluxes in a flux-balance computation [69]. Another possibility is to use the expression values to define target values for the fluxes and minimizing the quadratic deviation [82]. A similar method, based on

error minimization and developed for protein levels [90], can in principle also be applied to expression data. A multi-layer probabilistic framework, called PROM, mainly integrates a metabolic network with a transcriptional regulatory network but is also capable of using transcript data [83]. Its basic idea is to assign a probability value to gene states. Expression profiles have been used to *rank* reaction paths [64] or similarly "metabolic modules" [91]. Based on the textbook pathway definition (e.g., implemented in KEGG [73]), expression values have been used to score pathways in a framework called differential rank conservation [75,92,93]. The clustering of sets of genes, very common in the elucidation of transcription factors, has also been applied in conjunction with metabolic functions [87]. Also, graph theoretical inference has been used [94]. The topology of the metabolic network is the starting point to find the so-called regulatory signatures, patterns of gene changes indicating a diseased state (type 2 diabetes mellitus in this case) [95].

Gene set analysis [96] can be applied to metabolic pathways [97] as a distinct approach to use transcript correlation. A common technique to evaluate transcript profiles is to count up-/down-regulated genes (with a significance threshold); this can also be applied to KEGG pathway maps [73] or GO terms [72] to estimate the emphasis on certain functional characterizations [98].

3.5. Biological Focus of Studies

Methods applying transcript data to the metabolism can have many different aims. As a distinguishing characteristic, some studies are directed to lay theoretical foundations, while others are directly targeted to answer specific biological question.

For the first category, the reconstruction of a metabolic network for a specific cell with the help of transcript data can be mentioned [71]. Once a universal metabolic network is reconstructed (such as the universal human cell [7]), the subnetwork of reactions in specific cell type can be obtained with the same approach [8,67,81,89]. Similarly, transcript data is also used to estimate the set of active reactions in a particular state [68,69,82,99]. From the set of active reactions in a particular state, the essential information can be extracted in a further processing step, such as the so-called flux phenotypes [48] or, similarly, the metabolic state [90]. The detection of novel metabolic pathways [100,101] is an application in the area of fundamental biochemistry.

There are a number of studies that try to understand the regulation patterns by analyzing the co-expression of metabolic genes in a large number of transcript profiles [38,39,54,55,78]. These regulation patterns can lead to the prediction of transcription factors of one or several genes. Reed and Palsson analyzed the connection between correlated genes and coupled reactions [102].

Some applications of transcript data are directly related to clinical questions such as the prediction of biomarkers [17,94], the prediction of drug targets [67,68,103], identification of reporter metabolites in type 2 diabetes [95], the study of the effects of a drug such as baicalein [104], and identifying diet effects [105]. The search for target metabolites of regulation (*i.e.*, concerted regulation of genes to change the concentration of a certain metabolite) was the focus of another study [38]. If the underlying hypothesis of this study was true also for organisms other than *E. coli*, then this method would open a path to identify biomarkers in biotechnology and medicine.

Often, studies have an explicit biotechnological focus, for instance plant strain optimization [106], bacterial production rate optimization [107,108], optimization for algal growth [109,110], and understanding seed filling of oilseeds [42].

4. Available Software

The threshold-based activity prediction GIMME [62] (and the closely related iMAT [81], see Section 3.2 for the difference) is widely used, as it requires only minimal preconditions: a functional stoichiometric model and a few transcript profiles suffice. Without any further requirement it can be applied to predict the exometabolic fluxes. As these fluxes are often known, they can be used to increase the reliability of the model, in a multi-step algorithm to ensure the concordance to the input/output fluxes [62,81] or directly in the flux-balance optimization [68].

These expression-based prediction methods have been implemented in the universal flux computation frameworks COBRA [111] and FASIMU [112]. For the iMAT method [81] a standalone implementation is available [113]. The software for quantitative application of transcript data for flux prediction by Lee *et al.* is also freely available [82]. The TIGER [114] toolbox can be recommended if transcriptional regulation should also be taken into account. If a large number of transcript profiles are available and transcriptional networks should also be modeled, the freely available probabilistic framework PROM [83] can be recommended. The threshold value can be adjusted if quite a number of transcript profiles are available. It can be calculated individually with an optimization using MADE, which is also freely available [74].

To analyze correlations of the expression of different genes from transcript profiles with respect to metabolic paths, the PathRanker method [64] offers a freely available implementation. It does not require a functional stoichiometric model but needs large profile sets to work reliably.

5. Conclusions and Outlook

Inferring metabolic activity changes from transcript profiles is justified in two ways: mechanistically and by the assumption of evolutionary optimality. The former is based on the fact that RNA is translated into proteins then working as enzymes or transporters, thus modifying the metabolic flux related to the function. The latter is based on the argument: if the cell undertakes the effort to increase the mRNA production rate, it does so only with a purpose (related to the philosophical concept of *final cause*). The most likely purpose is to enhance a function for which the coded protein is required.

As the direct correlation of transcript profiles to metabolic reaction fluxes is not high, there is a wide range of applied methods having different strengths and weaknesses. The critical question is whether a particular method is suited to a particular application.

There is a clearly recognizable trend that the applied methods are increasingly enriched with available knowledge as the only way to increase the predictivity.

For the outlook of the field, it is foreseeable that large-scale metabolomics, proteomics, fluxomics, and enzyme characterization will become more manageable and affordable and the need to cover the wide distance from transcript to metabolism will vanish. The methods can then be improved with mechanistic descriptions of the underlying processes as soon as they are discovered. The methods crossing several

layers will have to include more components as it will be possible to parametrize them using experimental data. Genome-scale quantitative proteomics is on the brink of being widely available and feasible [11]. Quantitative metabolomics has reached the level of feasibility for hundreds of species. The developers of the reviewed methods and the users of their results will adopt this data when the coverage, cost, or accuracy makes it viable to do so. The application of mRNA data is, at the current time, just the most applicable means.

Acknowledgments

References

1. Gry, M.; Rimini, R.; Strömberg, S.; Asplund, A.; Pontén, F.; Uhlén, M.; Nilsson, P. Correlations between RNA and protein expression profiles in 23 human cell lines. *BMC Genomics* **2009**, *10*, doi:10.1186/1471-2164-10-365.

2. Zarnack, K.; Feldbrügge, M. mRNA trafficking in fungi. *Mol. Genet. Genomics* **2007**, *278*, 347–359.

3. Muresan, V.; Muresan, Z. Unconventional functions of microtubule motors. *Arch. Biochem. Biophys.* **2012**, *520*, 17–29.

4. Gingold, H.; Pilpel, Y. Determinants of translation efficiency and accuracy. *Mol. Syst. Biol.* **2011**, *7*, doi:10.1038/msb.2011.14.

5. Schomburg, I.; Chang, A.; Hofmann, O.; Ebeling, C.; Ehrentreich, F.; Schomburg, D. BRENDA: A resource for enzyme data and metabolic information. *Trends Biochem. Sci.* **2002**, *27*, 54–56.

6. Couzin, J. Genomics: Microarray data reproduced, but some concerns remain. *Science* **2006**, *313*, doi:10.1126/science.313.5793.1559a.

7. Duarte, N.C.; Becker, S.A.; Jamshidi, N.; Thiele, I.; Mo, M.L.; Vo, T.D.; Srivas, R.; Palsson, B.Ø. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1777–1782.

8. Gille, C.; Bölling, C.; Hoppe, A.; Bulik, S.; Hoffmann, S.; Hübner, K.; Karlstädt, A.; Ganeshan, R.; König, M.; Rother, K.; Weidlich, M.; Behre, J.; Holzhütter, H.G. HepatoNet1: A comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol. Syst. Biol.* **2010**, *6*, doi:10.1038/msb.2010.62.

9. 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.* **2012**, *6*, doi:10.1186/1752-0509-6-55.

10. Orth, J.D.; Conrad, T.M.; Na, J.; Lerman, J.A.; Nam, H.; Feist, A.M.; Palsson, B.Ø. A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. *Mol. Syst. Biol.* **2011**, *7*, doi:10.1038/msb.2011.65.

11. Nagaraj, N.; Kulak, N.A.; Cox, J.; Neuhauser, N.; Mayr, K.; Hoerning, O.; Vorm, O.; Mann, M. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by

single-shot ultra HPLC runs on a bench top Orbitrap. *Mol. Cell Proteomics* **2012**, *11*, doi:10.1074/mcp.M111.013722.

12. Hackl, H.; Cabo, F.S.; Sturn, A.; Wolkenhauer, O.; Trajanoski, Z. Analysis of DNA microarray data. *Curr. Top. Med. Chem.* **2004**, *4*, 1357–1370.

13. Ladunga, I. An overview of the computational analyses and discovery of transcription factor binding sites. *Methods Mol. Biol.* **2010**, *674*, 1–22.

14. Boone, D.N.; Lee, A.V. Targeting the insulin-like growth factor receptor: Developing biomarkers from gene expression profiling. *Crit. Rev. Oncog.* **2012**, *17*, 161–173.

15. Churchill, G.A. Fundamentals of experimental design for cDNA microarrays. *Nat. Genet.* **2002**, *32 Suppl*, 490–495.

16. Baldwin, D.; Crane, V.; Rice, D. A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. *Curr. Opin. Plant Biol.* **1999**, *2*, 96–103.

17. Chang, R.L.; Xie, L.; Xie, L.; Bourne, P.E.; Palsson, B.Ø. Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comput. Biol.* **2010**, *6*, e1000938, doi:10.1371/journal.pcbi.1000938.

18. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63.

19. Matsumura, H.; Reich, S.; Ito, A.; Saitoh, H.; Kamoun, S.; Winter, P.; Kahl, G.; Reuter, M.; Kruger, D.H.; Terauchi, R. Gene expression analysis of plant host-pathogen interactions by SuperSAGE. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15718–15723.

20. Shendure, J. The beginning of the end for microarrays? *Nat. Methods* **2008**, *5*, 585–587.

21. Kawasaki, E.S. The end of the microarray Tower of Babel: Will universal standards lead the way? *J. Biomol. Tech.* **2006**, *17*, 200–206.

22. Gygi, S.P.; Rochon, Y.; Franza, B.R.; Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.* **1999**, *19*, 1720–1730.

23. Griffin, T.J.; Gygi, S.P.; Ideker, T.; Rist, B.; Eng, J.; Hood, L.; Aebersold, R. Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. *Mol. Cell Proteomics* **2002**, *1*, 323–333.

24. Greenbaum, D.; Colangelo, C.; Williams, K.; Gerstein, M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* **2003**, *4*, doi:10.1186/gb-2003-4-9-117.

25. Tuller, T.; Kupiec, M.; Ruppin, E. Determinants of protein abundance and translation efficiency in S. cerevisiae. *PLoS Comput. Biol.* **2007**, *3*, e248, doi:10.1371/journal.pcbi.0030248.

26. Maier, T.; Güell, M.; Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **2009**, *583*, 3966–3973.

27. Hajduch, M.; Hearne, L.B.; Miernyk, J.A.; Casteel, J.E.; Joshi, T.; Agrawal, G.K.; Song, Z.; Zhou, M.; Xu, D.; Thelen, J.J. Systems analysis of seed filling in Arabidopsis: Using general linear modeling to assess concordance of transcript and protein expression. *Plant Physiol.* **2010**, *152*, 2078–2087.

28. Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global quantification of mammalian gene expression control. *Nature* **2011**, *473*, 337–342.

29. Miller, C.; Schwalb, B.; Maier, K.; Schulz, D.; Dümcke, S.; Zacher, B.; Mayer, A.; Sydow, J.; Marcinowski, L.; Dölken, L.; Martin, D.E.; Tresch, A.; Cramer, P. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol. Syst. Biol.* **2011**, *7*, doi:10.1038/msb.2010.112.

30. Belle, A.; Tanay, A.; Bitincka, L.; Shamir, R.; O'Shea, E.K. Quantification of protein half-lives in the budding yeast proteome. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13004–13009.

31. Zimran, A.; Forman, L.; Suzuki, T.; Dale, G.L.; Beutler, E. *In vivo* aging of red cell enzymes: Study of biotinylated red blood cells in rabbits. *Am. J. Hematol.* **1990**, *33*, 249–254.

32. Wittig, U.; Kania, R.; Golebiewski, M.; Rey, M.; Shi, L.; Jong, L.; Algaa, E.; Weidemann, A.; Sauer-Danzwith, H.; Mir, S.; Krebs, O.; Bittkowski, M.; Wetsch, E.; Rojas, I.; Müller, W. SABIO-RK–database for biochemical reaction kinetics. *Nucleic Acids Res.* **2012**, *40*, D790–D796.

33. Li, C.; Donizelli, M.; Rodriguez, N.; Dharuri, H.; Endler, L.; Chelliah, V.; Li, L.; He, E.; Henry, A.; Stefan, M.I.; *et al.* BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst. Biol.* **2010**, *4*, doi:10.1186/1752-0509-4-92.

34. Olivier, B.G.; Snoep, J.L. Web-based kinetic modelling using JWS Online. *Bioinformatics* **2004**, *20*, 2143–2144.

35. Reder, C. Metabolic control theory: A structural approach. *J. Theor. Biol.* **1988**, *135*, 175–201.

36. Hofmeyr, J.H.; Kacser, H.; van der Merwe, K.J. Metabolic control analysis of moiety-conserved cycles. *Eur. J. Biochem.* **1986**, *155*, 631–641.

37. Small, J.R.; Fell, D.A. The matrix method of metabolic control analysis: Its validity for complex pathway structures. *J. Theor. Biol.* **1989**, *136*, 181–197.

38. Hancock, T.; Wicker, N.; Takigawa, I.; Mamitsuka, H. Identifying neighborhoods of coordinated gene expression and metabolite profiles. *PLoS One* **2012**, *7*, e31345, doi:10.1371/journal.pone.0031345.

39. Kharchenko, P.; Church, G.M.; Vitkup, D. Expression dynamics of a cellular metabolic network. *Mol. Syst. Biol.* **2005**, *1*, doi:10.1038/msb4100023.

40. Schuster, S.; Hilgetag, C. On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* **1994**, *2*, 165–182.

41. Cakir, T.; Kirdar, B.; Ulgen, K.O. Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks. *Biotechnol. Bioeng.* **2004**, *86*, 251–260.

42. Hajduch, M.; Matusova, R.; Houston, N.L.; Thelen, J.J. Comparative proteomics of seed maturation in oilseeds reveals differences in intermediary metabolism. *Proteomics* **2011**, *11*, 1619–1629.

43. Saito, K.; Matsuda, F. Metabolomics for functional genomics, systems biology, and biotechnology. *Annu. Rev. Plant Biol.* **2010**, *61*, 463–489.

44. Ishihama, Y.; Schmidt, T.; Rappsilber, J.; Mann, M.; Hartl, F.U.; Kerner, M.J.; Frishman, D. Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics* **2008**, *9*, doi:10.1186/1471-2164-9-102.

45. Ishii, N.; Nakahigashi, K.; Baba, T.; Robert, M.; Soga, T.; Kanai, A.; Hirasawa, T.; Naba, M.; Hirai, K.; Hoque, A.; *et al.* Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* **2007**, *316*, 593–597.

46. Wisselink, H.W.; Cipollina, C.; Oud, B.; Crimi, B.; Heijnen, J.J.; Pronk, J.T.; van Maris, A.J.A. Metabolome, transcriptome and metabolic flux analysis of arabinose fermentation by engineered Saccharomyces cerevisiae. *Metab. Eng.* **2010**, *12*, 537–551.

47. Ter Kuile, B.H.; Westerhoff, H.V. Transcriptome meets metabolome: Hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett.* **2001**, *500*, 169–171.

48. Moxley, J.F.; Jewett, M.C.; Antoniewicz, M.R.; Villas-Boas, S.G.; Alper, H.; Wheeler, R.T.; Tong, L.; Hinnebusch, A.G.; Ideker, T.; Nielsen, J.; Stephanopoulos, G. Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 6477–6482.

49. Yang, C.; Hua, Q.; Shimizu, K. Integration of the information from gene expression and metabolic fluxes for the analysis of the regulatory mechanisms in Synechocystis. *Appl. Microbiol. Biotechnol.* **2002**, *58*, 813–822.

50. Daran-Lapujade, P.; Rossell, S.; van Gulik, W.M.; Luttik, M.A.H.; de Groot, M.J.L.; Slijper, M.; Heck, A.J.R.; Daran, J.M.; de Winde, J.H.; Westerhoff, H.V.; *et al.* The fluxes through glycolytic enzymes in Saccharomyces cerevisiae are predominantly regulated at posttranscriptional levels. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 15753–15758.

51. Rossell, S.; van der Weijden, C.C.; Lindenbergh, A.; van Tuijl, A.; Francke, C.; Bakker, B.M.; Westerhoff, H.V. Unraveling the complexity of flux regulation: A new method demonstrated for nutrient starvation in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 2166–2171.

52. Daran-Lapujade, P.; Jansen, M.L.A.; Daran, J.M.; van Gulik, W.; de Winde, J.H.; Pronk, J.T. Role of transcriptional regulation in controlling fluxes in central carbon metabolism of Saccharomyces cerevisiae. A chemostat culture study. *J. Biol. Chem.* **2004**, *279*, 9125–9138.

53. Wessely, F.; Bartl, M.; Guthke, R.; Li, P.; Schuster, S.; Kaleta, C. Optimal regulatory strategies for metabolic pathways in Escherichia coli depending on protein costs. *Mol. Syst. Biol.* **2011**, *7*, doi:10.1038/msb.2011.46.

54. Seshasayee, A.S.N.; Fraser, G.M.; Babu, M.M.; Luscombe, N.M. Principles of transcriptional regulation and evolution of the metabolic system in *E. coli*. *Genome Res.* **2009**, *19*, 79–91.

55. Notebaart, R.A.; Teusink, B.; Siezen, R.J.; Papp, B. Co-regulation of metabolic genes is better explained by flux coupling than by network distance. *PLoS Comput. Biol.* **2008**, *4*, e26, doi:10.1371/journal.pcbi.0040026.

56. De Visser, J.A.G.M.; Cooper, T.F.; Elena, S.F. The causes of epistasis. *Proc. Biol. Sci.* **2011**, *278*, 3617–3624.

57. Snitkin, E.S.; Segrè, D. Epistatic interaction maps relative to multiple metabolic phenotypes. *PLoS Genet.* **2011**, *7*, e1001294, doi:10.1371/journal.pgen.1001294.

58. He, X.; Qian, W.; Wang, Z.; Li, Y.; Zhang, J. Prevalent positive epistasis in Escherichia coli and Saccharomyces cerevisiae metabolic networks. *Nat. Genet.* **2010**, *42*, 272–276.

59. Harrison, R.; Papp, B.; Pál, C.; Oliver, S.G.; Delneri, D. Plasticity of genetic interactions in metabolic networks of yeast. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2307–2312.

60. Segrè, D. The regulatory software of cellular metabolism. *Trends Biotechnol.* **2004**, *22*, 261–265.

61. Szappanos, B.; Kovács, K.; Szamecz, B.; Honti, F.; Costanzo, M.; Baryshnikova, A.; Gelius-Dietrich, G.; Lercher, M.J.; Jelasity, M.; Myers, C.L.; *et al.* An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat. Genet.* **2011**, *43*, 656–662.

62. Becker, S.A.; Palsson, B.Ø. Context-specific metabolic networks are consistent with experiments. *PLoS Comput. Biol.* **2008**, *4*, e1000082, doi:10.1371/journal.pcbi.1000082.

63. Hoffmann, S.; Holzhütter, H.G. Uncovering metabolic objectives pursued by changes of enzyme levels. *Ann. N.Y. Acad. Sci.* **2009**, *1158*, 57–70.

64. Hancock, T.; Takigawa, I.; Mamitsuka, H. Mining metabolic pathways through gene expression. *Bioinformatics* **2010**, *26*, 2128–2135.

65. Shlomi, T.; Cabili, M.N.; Ruppin, E. Predicting metabolic biomarkers of human inborn errors of metabolism. *Mol. Syst. Biol.* **2009**, *5*, doi:10.1038/msb.2009.22.

66. Shlomi, T. Metabolic network-based interpretation of gene expression data elucidates human cellular metabolism. *Biotechnol. Genet. Eng. Rev.* **2010**, *26*, 281–296.

67. Folger, O.; Jerby, L.; Frezza, C.; Gottlieb, E.; Ruppin, E.; Shlomi, T. Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* **2011**, *7*, doi:10.1038/msb.2011.35.

68. Huthmacher, C.; Hoppe, A.; Bulik, S.; Holzhütter, H.G. Antimalarial drug targets in Plasmodium falciparum predicted by stage-specific metabolic network analysis. *BMC Syst. Biol.* **2010**, *4*, doi:10.1186/1752-0509-4-120.

69. Colijn, C.; Brandes, A.; Zucker, J.; Lun, D.S.; Weiner, B.; Farhat, M.R.; Cheng, T.Y.; Moody, D.B.; Murray, M.; Galagan, J.E. Interpreting expression data with metabolic flux models: Predicting Mycobacterium tuberculosis mycolic acid production. *PLoS Comput. Biol.* **2009**, *5*, e1000489, doi:10.1371/journal.pcbi.1000489.

70. Manichaikul, A.; Ghamsari, L.; Hom, E.F.Y.; Lin, C.; Murray, R.R.; Chang, R.L.; Balaji, S.; Hao, T.; Shen, Y.; Chavali, A.K.; *et al.* Metabolic network analysis integrated with transcript verification for sequenced genomes. *Nat. Methods* **2009**, *6*, 589–592.

71. Thiele, I.; Palsson, B.Ø. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc* **2010**, *5*, 93–121.

72. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; *et al.* Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29.

73. Kanehisa, M. The KEGG database. *Novartis Found Symp.* **2002**, *247*, 91–101; Discussion 101–103, 119–128, 244–252.

74. Jensen, P.A.; Papin, J.A. Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics* **2011**, *27*, 541–547.

75. Eddy, J.A.; Geman, D.; Price, N.D. Relative Expression Analysis for Identifying Perturbed Pathways. In *Proceedings of the Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, Minneapolis, MN, USA, 3–6 September 2009; pp. 5456–5459.

76. Mamitsuka, H.; Okuno, Y.; Yamaguchi, A. Mining biologically active patterns in metabolic pathways using microarray expression profiles. *SIGKDD Explor.* **2003**, *5*, 113–121.

77. Edelman, L.B.; Toia, G.; Geman, D.; Zhang, W.; Price, N.D. Two-transcript gene expression classifiers in the diagnosis and prognosis of human diseases. *BMC Genomics* **2009**, *10*, doi:10.1186/1471-2164-10-583. .

78. Ihmels, J.; Levy, R.; Barkai, N. Principles of transcriptional control in the metabolic network of Saccharomyces cerevisiae. *Nat. Biotechnol.* **2004**, *22*, 86–92.

79. Loraine, A. Co-expression analysis of metabolic pathways in plants. *Methods Mol. Biol.* **2009**, *553*, 247–264.

80. Hebenstreit, D.; Fang, M.; Gu, M.; Charoensawan, V.; van Oudenaarden, A.; Teichmann, S.A. RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Mol. Syst. Biol.* **2011**, *7*, doi:10.1038/msb.2011.28.

81. Shlomi, T.; Cabili, M.N.; Herrgård, M.J.; Palsson, B.Ø.; Ruppin, E. Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.* **2008**, *26*, 1003–1010.

82. Lee, D.; Smallbone, K.; Dunn, W.B.; Murabito, E.; Winder, C.L.; Kell, D.B.; Mendes, P.; Swainston, N. Improving metabolic flux predictions using absolute gene expression data. *BMC Syst. Biol.* **2012**, *6*, doi:10.1186/1752-0509-6-73.

83. Chandrasekaran, S.; Price, N.D. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 17845–17850.

84. Hancock, T.; Mamitsuka, H. Active pathway identification and classification with probabilistic ensembles. *Genome Inf.* **2010**, *22*, 30–40.

85. Hoffmann, S.; Hoppe, A.; Holzhütter, H.G. Composition of metabolic flux distributions by functionally interpretable minimal flux modes (MinModes). *Genome Inf.* **2006**, *17*, 195–207.

86. Voss, K.; Heiner, M.; Koch, I. Petri nets for steady state analysis of metabolic systems. *Stud. Health Technol. Inf.* **2011**, *162*, 56–76.

87. König, R.; Schramm, G.; Oswald, M.; Seitz, H.; Sager, S.; Zapatka, M.; Reinelt, G.; Eils, R. Discovering functional gene expression patterns in the metabolic network of Escherichia coli with wavelets transforms. *BMC Bioinforma.* **2006**, *7*, doi:10.1186/1471-2105-7-119.

88. Arita, M. From metabolic reactions to networks and pathways. *Methods Mol. Biol.* **2012**, *804*, 93–106.

89. Jerby, L.; Shlomi, T.; Ruppin, E. Computational reconstruction of tissue-specific metabolic models: Application to human liver metabolism. *Mol. Syst. Biol.* **2010**, *6*, doi:10.1038/msb.2010.56.

90. Yizhak, K.; Benyamini, T.; Liebermeister, W.; Ruppin, E.; Shlomi, T. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics* **2010**, *26*, i255–i260.

91. Deo, R.C.; Hunter, L.; Lewis, G.D.; Pare, G.; Vasan, R.S.; Chasman, D.; Wang, T.J.; Gerszten, R.E.; Roth, F.P. Interpreting metabolomic profiles using unbiased pathway models. *PLoS Comput. Biol.* **2010**, *6*, e1000692, doi:10.1371/journal.pcbi.1000692.

92. Eddy, J.A.; Hood, L.; Price, N.D.; Geman, D. Identifying tightly regulated and variably expressed networks by Differential Rank Conservation (DIRAC). *PLoS Comput. Biol.* **2010**, *6*, e1000792, doi:10.1371/journal.pcbi.1000792.

93. Eddy, J.A.; Sung, J.; Geman, D.; Price, N.D. Relative expression analysis for molecular cancer diagnosis and prognosis. *Technol. Cancer Res. Treat.* **2010**, *9*, 149–159.

94. Song, W.M.; Matteo, T.D.; Aste, T. Hierarchical information clustering by means of topologically embedded graphs. *PLoS One* **2012**, *7*, e31929, doi:10.1371/journal.pone.0031929.

95. Zelezniak, A.; Pers, T.H.; Soares, S.; Patti, M.E.; Patil, K.R. Metabolic network topology reveals transcriptional regulatory signatures of type 2 diabetes. *PLoS Comput. Biol.* **2010**, *6*, e1000729, doi:10.1371/journal.pcbi.1000729.

96. Song, S.; Black, M.A. Microarray-based gene set analysis: A comparison of current methods. *BMC Bioinforma.* **2008**, *9*, doi:10.1186/1471-2105-9-502.

97. Werner, T. Bioinformatics applications for pathway analysis of microarray data. *Curr. Opin. Biotechnol.* **2008**, *19*, 50–54.

98. Bono, H.; Okazaki, Y. The study of metabolic pathways in tumors based on the transcriptome. *Semin. Cancer Biol.* **2005**, *15*, 290–299.

99. Roberts, S.B.; Robichaux, J.L.; Chavali, A.K.; Manque, P.A.; Lee, V.; Lara, A.M.; Papin, J.A.; Buck, G.A. Proteomic and network analysis characterize stage-specific metabolism in Trypanosoma cruzi. *BMC Syst. Biol.* **2009**, *3*, doi:10.1186/1752-0509-3-52.

100. Craig, T.A.; Zhang, Y.; McNulty, M.S.; Middha, S.; Ketha, H.; Singh, R.J.; Magis, A.T.; Funk, C.; Price, N.D.; Ekker, S.C.; Kumar, R. Research Resource: Whole Transcriptome RNA Sequencing Detects Multiple 1α,25-Dihydroxyvitamin D3-Sensitive Metabolic Pathways in Developing Zebrafish. *Mol. Endocrinol.* **2012**, *26*, 1630–1642.

101. Kaleta, C.; de Figueiredo, L.F.; Werner, S.; Guthke, R.; Ristow, M.; Schuster, S. In silico evidence for gluconeogenesis from fatty acids in humans. *PLoS Comput. Biol.* **2011**, *7*, e1002116, doi:10.1371/journal.pcbi.1002116.

102. Reed, J.L.; Palsson, B.Ø. Genome-scale in silico models of *E. coli* have multiple equivalent phenotypic states: Assessment of correlated reaction subsets that comprise network states. *Genome Res.* **2004**, *14*, 1797–1805.

103. Oberhardt, M.A.; Goldberg, J.B.; Hogardt, M.; Papin, J.A. Metabolic network analysis of Pseudomonas aeruginosa during chronic cystic fibrosis lung infection. *J. Bacteriol.* **2010**, *192*, 5534–5548.

104. Qin, S.; Chen, J.; Tanigawa, S.; Hou, D.X. Gene expression profiling and pathway network analysis of hepatic metabolic enzymes targeted by baicalein. *J. Ethnopharmacol.* **2012**, *140*, 131–140.

105. Nookaew, I.; Gabrielsson, B.G.; Holmäng, A.; Sandberg, A.S.; Nielsen, J. Identifying molecular effects of diet through systems biology: Influence of herring diet on sterol metabolism and protein turnover in mice. *PLoS One* **2010**, *5*, e12361, doi:10.1371/journal.pone.0012361.

106. Diretto, G.; Al-Babili, S.; Tavazza, R.; Scossa, F.; Papacchioli, V.; Migliore, M.; Beyer, P.; Giuliano, G. Transcriptional-metabolic networks in beta-carotene-enriched potato tubers: The long and winding road to the Golden phenotype. *Plant Physiol.* **2010**, *154*, 899–912.

107. Ghosh, A.; Zhao, H.; Price, N.D.   Genome-scale consequences of cofactor balancing in engineered pentose utilization pathways in Saccharomyces cerevisiae. *PLoS One* **2011**, *6*, e27316, doi:10.1371/journal.pone.0027316.

108. Milne, C.B.; Eddy, J.A.; Raju, R.; Ardekani, S.; Kim, P.J.; Senger, R.S.; Jin, Y.S.; Blaschek, H.P.; Price, N.D. Metabolic network reconstruction and genome-scale model of butanol-producing strain Clostridium beijerinckii NCIMB 8052. *BMC Syst. Biol.* **2011**, *5*, doi:10.1186/1752-0509-5-130.

109. Chang, R.L.; Ghamsari, L.; Manichaikul, A.; Hom, E.F.Y.; Balaji, S.; Fu, W.; Shen, Y.; Hao, T.; Palsson, B.Ø.; Salehi-Ashtiani, K.; *et al*.   Metabolic network reconstruction of Chlamydomonas offers insight into light-driven algal metabolism. *Mol. Syst. Biol.* **2011**, *7*, doi:10.1038/msb.2011.52.

110. Sorokina, O.; Corellou, F.; Dauvillée, D.; Sorokin, A.; Goryanin, I.; Ball, S.; Bouget, F.Y.; Millar, A.J.   Microarray data can predict diurnal changes of starch content in the picoalga Ostreococcus. *BMC Syst. Biol.* **2011**, *5*, doi:10.1186/1752-0509-5-36.

111. Becker, S.A.; Feist, A.M.; Mo, M.L.; Hannum, G.; Palsson, B.Ø.; Herrgård, M.J. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox. *Nat. Protoc.* **2007**, *2*, 727–738.

112. Hoppe, A.; Hoffmann, S.; Gerasch, A.; Gille, C.; Holzhütter, H.G. FASIMU: Flexible software for flux-balance computation series in large metabolic networks. *BMC Bioinforma.* **2011**, *12*, doi:10.1186/1471-2105-12-28.

113. Zur, H.; Ruppin, E.; Shlomi, T. iMAT: An integrative metabolic analysis tool. *Bioinformatics* **2010**, *26*, 3140–3142.

114. Jensen, P.A.; Lutz, K.A.; Papin, J.A. TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. *BMC Syst. Biol.* **2011**, *5*, doi:10.1186/1752-0509-5-147.