Integrated visualization of a multi-omics study of starvation in mouse intestine

Martijn P. van Iersel[1,2,3], Milka Sokolović[4,5], Kaatje Lenaerts[6], Martina Kutmon[1,2], Freek G. Bouwman[7], Wouter H. Lamers[8], Edwin C.M. Mariman[7] and Chris T. Evelo[1,2,*]

[1]Department of Bioinformatics BiGCaT, NUTRIM School for Nutrition, Toxicology, and Metabolism, Maastricht University, Maastricht, The Netherlands

[2]Netherlands Consortium for Systems Biology (NCSB), The Netherlands

[3]Current address: General Bioinformatics, Reading, UK

[4]Department of Medical Biochemistry, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands

[5]European Food Information Council, Brussels, Belgium

[6]Department of Surgery, NUTRIM School for Nutrition, Toxicology, and Metabolism, Maastricht University, Maastricht, The Netherlands

[7]Department of Human Biology, NUTRIM School for Nutrition, Toxicology, and Metabolism, Maastricht University, Maastricht, The Netherlands

[8]Tytgat Institute for Liver and Intestinal Research (formerly AMC Liver Center), Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

Summary

Our understanding of complex biological processes can be enhanced by combining different kinds of high-throughput experimental data, but the use of incompatible identifiers makes data integration a challenge. We aimed to improve methods for integrating and visualizing different types of omics data. To validate these methods, we applied them to two previous studies on starvation in mice, one using proteomics and the other using transcriptomics technology.

We extended the PathVisio software with new plugins to link proteins, transcripts and pathways. A low overall correlation between proteome and transcriptome data was detected (Spearman rank correlation: 0.21). At the level of individual genes, correlation was highly variable. Many mRNA/protein pairs, such as fructose biphosphate aldolase B and ATP Synthase, show good correlation. For other pairs, such as ferritin and elongation factor 2, an interesting effect is observed, where mRNA and protein levels change in opposite

*To whom correspondence should be addressed. Email: chris.evelo@maastrichtuniversity.nl

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

directions, suggesting they are not primarily regulated at the transcriptional level. We used pathway diagrams to visualize the integrated datasets and found it encouraging that transcriptomics and proteomics data supported each other at the pathway level.
Visualization of the integrated dataset on pathways led to new observations on gene-regulation in the response of the gut to starvation. Our methods are generic and can be applied to any multi-omics study. The PathVisio software can be obtained at http://www.pathvisio.org. Supplemental data are available at http://www.bigcat.unimaas.nl/data/jib-supplemental/, including instructions on reproducing the pathway visualizations of this manuscript.

1 Introduction

The intestine plays an important role in the response of the body to (prolonged) starvation. In two previous publications, the transcriptome [1] and the proteome [2] of the murine intestine were studied after 0, 12, 24 and 72 hours of starvation. In the present study we combine and compare data from the two earlier studies. The goal is twofold. First, to develop bioinformatics tools needed to make an integrated omics approach feasible. Second, to get a more comprehensive view of regulation of pathways involved in the starvation response of the gut.

1.1 "Omics" integration

Microarray technology can be used to measure the expression of thousands of genes at the same time. Methods for processing, analyzing and interpreting microarray data are well established, and off-the-shelf microarrays are available with enough capacity to measure the majority of the genes in a genome.

Not transcripts but proteins are directly involved in biological activities. Microarray studies usually assume implicitly that there is a correlation between transcript and protein abundance, but only moderate levels of correlation have been reported [3, 4, 5, 6, 7]. A lack of correlation could have several causes such as variation in protein turnover rates and post-transcriptional regulation [8]. This lack of correlation between transcript and protein levels is an important argument for measuring protein expression directly. Two-dimensional gel electrophoresis (2DE), still one of the most common techniques for quantitative proteomics, has continued to mature in recent years and has achieved higher standards of data quality, reproducibility and protein identification [9].

Proteomic technologies have, nevertheless, a number of disadvantages. In a standard 2DE proteomics experiment in mammals, typically only ~100 proteins are identified and measured, or roughly 0.5% of the genome, far less than what is typical for microarray studies. Moreover, 2DE studies suffer from problems of bias. Proteins vary more widely than mRNA molecules in physical properties such as hydrophobicity, electric charge and size. The subset of measured proteins is biased towards proteins that are easily separable by 2DE and abundant enough to be identified. Furthermore, the spots that show the clearest response to experimental conditions are picked first. Protein identification is, therefore, a bottleneck in 2DE. New gel-free proteomics techniques measure a wider range of protein chemistries and abundances and suffer less from

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

bias problems [10]. In spite of these issues, 2DE remains commonly used for its maturity and relative simplicity [9].

As both proteomics and transcriptomics methods have their advantages, it could be beneficial to combine them. It appears that only some genes are primarily regulated at the transcriptional level, showing higher transcript/protein correlation than what would be expected from global correlation levels [3], whereas other genes are regulated post-transcriptionally, showing much lower correspondence. Therefore, we believe that integrated analysis of omics datasets must take into account a-priori knowledge about the regulation of genes and proteins, for example in the form of pathway diagrams. Pathway diagrams contain biological context, such as which entities are related, what is their cellular location, which proteins are interaction partners, and what is the nature of those interactions (stimulation or repression). By visualizing the combined dataset on a pathway diagram one can interpret the biological context more easily. As part of this study we aimed to develop easy-to-use software to make this possible.

1.2 Regulation of the intestinal response to fasting

We combined two experimental datasets to validate our data integration methodology. The experimental data relates to the intestinal changes in response to fasting. A better understanding of fasting could lead to better understanding of malnourishment and better treatment of cachexia (wasting syndrome) caused by chronic disease [11].

Based on the rate of weight loss, nitrogen excretion, concentration of plasma metabolites and resting metabolic rate, the body passes through three successive adaptive phases during fasting, often defined as postabsorptive, coping and preterminal [12]. In mammals, these phases have been associated with the primary fuel that is putatively available to the tissues [13, 14, 15, 16]. Based on whole-body energy expenditure, the “sugars-fats-proteins” succession of energy substrates during fasting was proposed [12, 13], and this model was then extrapolated to all organs separately. However, transcriptomic studies in rodents that have prospected the adaptive response to fasting of different organs [1, 17, 18, 19, 20, 21, 22] reveal a different scenario. These studies have shown that already in the postabsorptive phase organs liberally increase protein and lipid catabolism, fuelling hepatic and renal gluconeogenesis and ketogenesis. Only during prolonged fasting ($\geq$ 24h in mice) is protein catabolism minimized.

It is clear that the various organs (liver, intestine, adipose tissue, muscle, kidney, brain) play different roles in this response, but the whole picture is not yet completely elucidated. The small intestine contributes to the body’s adaptation to fasting by a biphasic response of carbohydrate metabolism, which peaks during short and again after prolonged fasting [1]. Early changes are associated with glutamine conservation, inhibition of pyruvate oxidation, stimulation of glutamate catabolism via aspartate and phosphoenolpyruvate to lactate, and enhanced fatty-acid oxidation and ketone-body synthesis. Changes upon continued fasting implied the production of glucose rather than lactate from carbohydrate backbones and downregulation of fatty-acid oxidation. In addition, cell turnover is progressively downregulated by inhibiting cell cycling and apoptosis to reduce the high energy costs of constant enterocyte turnover [1].

The question that this study tries to answer is whether protein expression data reinforce the pic-

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

ture that arises from transcriptomics pathway analysis, and to what extent post-transcriptional regulation plays a role in these pathways.

2 Methods

The experimental procedure for treatment of animals, microarray and 2DE was described before [1, 2], but is briefly summarized here.

Animals. Male FVB mice from Charles River (Maastricht, The Netherlands) were housed at 20-22°C, 50-60% humidity on a 12 hours light/dark cycle. They ingested food and water ad libitum until the age of 6 weeks. Groups of 6 mice were fasted for 0, 12, 24, and 72 hours, after which the animals were killed by cervical dislocation. The small intestine was removed and separated from adjacent tissue, and both protein and RNA were isolated. The same mice were used for both microarray and proteomics experiments.

Microarrays. Samples of the intestine were applied to 60-mer Mouse Development 22k Oligo Microarray G4120A (Agilent). Three arrays per experimental condition were used. Per microarray, 20 μg mRNA, pooled from 2 intestines, was labeled with Cy3. RNA Pooled from 6 fed animals was labeled with Cy5 and used as a common reference across all arrays. After hybridization the arrays were scanned with Agilent's dual-laser microarray slide scanner and processed with Agilent's Feature Extraction software 6.1.1. Quantile normalization was applied to background-subtracted intensities.

2DE. The procedure for generating 2D protein gel images was as described before [23]. Protein samples were isolated from equal quantities of proximal and distal parts of the intestine, pooled per mouse. 1 gel was made for each mouse. 100 μg of total protein was separated by isoelectric focusing using IPG strips, and then placed onto 12.5% SDS-polyacrylamide gels for protein separation in the second dimension. Gels stained with SYPRO Ruby Protein stain were scanned with the Molecular Imager FX (Bio-Rad Laboratories). Analysis of differentially expressed proteins was performed using PDQuest 7.3 (Bio-Rad Laboratories). A number of spots were selected for identification, with a preference for spots with a significant intensity difference. Selected spots were excised and subjected to tryptic in-gel digestion and MALDI-TOF MS (Waters, Manchester, UK), generating peptide mass fingerprints which were subsequently identified using the MASCOT search engine against the SwissProt database.

Microarray annotation. The microarray type used was the 60-mer Mouse Development 22k Oligo Microarray G4120A (Agilent). The microarray contains 20,280 probes of 60 nucleotides each. The annotation file provided by Agilent (Version of Dec.16 2009) associates only 9,616 probes to Ensembl gene identifiers. The probes were designed based on a 5-year old genome build, which could have diverged in the intervening years. On the assumption that 60-mer probes do not require a complete sequence match to hybridize to transcripts, we investigated if a BLAST search with less stringent settings would increase probe annotation coverage. We selected the best BLAST hits against a more recent Ensembl (release 57) Mouse cDNA database, with a minimum e value of 1.0e-6. The BLAST resulted in annotation for 10,696 probes, an increase of 11%. We opted to employ the BLAST results instead of Agilent annotations for all further analysis. Identifiers in both data sets were mapped to pathways using the BridgeDb frame-

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

work [24]. Because none of the standard identifier mapping resources included in BridgeDb contained mappings for the Agilent array, we prepared a custom mapping table that allowed proper interpretation of the BLAST results by BridgeDb.

Analysis and Pathway visualization. Correlation and expression plots were created with R/BioConductor. Pathway visualization was performed using the pathway editing and visualization tool PathVisio which was first published in 2008 [25]. We developed two new plugins for PathVisio: the Gex plugin, which manages data import and visualization, and the BridgeDb-Config plugin, which enables configuration of identifier mapping resources.

The Gex plugin handles the import of expression data in PathVisio and the mapping of the provided identifiers in the dataset to the identifiers used in the pathways. This plugin is now a core module and does not need to be installed separately. To enable more advanced options of identifier mapping, e.g. the usage of different identifier mapping tables together, the BridgeDb-Config plugin was developed. The HTML export plugin allows the export of a single pathway diagram as an HTML page as well as the export of a complete pathway statistics result including the colored pathway diagrams. Since PathVisio 3, the plugin manager provides a fast and simple way to install plugins from a central plugin repository. The BridgeDbConfig plugin and the HTML export plugin are available in the central plugin repository since March 2013.

The mouse pathway set, obtained from WikiPathways [26] March 2010, was used for visualization. This pathway set covers 3,975 unique genes, or 14% of the whole mouse genome (using all genes and pseudo genes of Ensembl release 57 as the reference). We counted the overlap between the list of measured genes and proteins and the list of genes occurring in at least one pathway. 66% of measured proteins occur in at least one pathway (51 out of the 77 unique measured proteins). 24% of measured genes occur in at least one pathway (2,083 out of 8,648 unique measured genes).

3 Results

3.1 Supplemental data

The section below refers to four supplemental data files, all of which can be found at the following location: http://www.bigcat.unimaas.nl/data/jib-supplemental/. These supplemental files contain the necessary data to reproduce the analysis presented here, as well as a tutorial on how to recreate the pathway visualizations of this manuscript.

3.2 Identifier mapping

A prerequisite for integration of multiple omics studies is the ability to map identifiers from various sources [27]. Each data point in the microarray dataset was identified with an Agilent probe identifier (such as A_65_P03556), and each protein was associated with a Uniprot identifier (such as P09528). One of the difficulties in identifier mapping is that there is a one-to-many

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

relation between genes and measurements. Because microarray designs often include some redundancy, there could be more than one probe identifier per gene. Similarly, one protein might give rise to multiple spots on the 2D gel, depending on the protein modification state [9], so there are frequently multiple spot numbers per gene.

The BridgeDb framework [24] was used to map probe and protein identifiers to gene identifiers and integrate the two datasets. This framework was used both in the PathVisio pathway visualization software [25] and in the correlation analysis in R.

Three sources of information were used for identifier mapping. By using BridgeDb, the three resources were unified into a single mapping resource. First, the results of BLAST of microarray sequences for mapping Agilent probes to genes in the transcriptomics dataset; second, a regular BridgeDerby database for mapping proteins to genes; and third, a manual override table to fix three errors in the protein dataset.

Three proteins were annotated with erroneous protein identifiers by the identification software (in one case a GenBank accession number, in another an identifier for the human homologue, and in the third a deprecated Uniprot identifier). Rather than fixing the original data, BridgeDb allowed us to create a separate "manual override" mapping table and combine it with the rest of the mapping resources. The advantage of doing so, rather than simply fixing up the original dataset, was that the modifications remained separated and could be re-examined and adjusted later.

3.3 Correlation

The proteomics dataset contained 130 identified spots. Since more than one spot may arise from the same protein in different modification states, those 130 spots corresponded to only 77 unique protein identifiers. Moreover, due to limitations of the microarray used, for some of them no mRNA data was available, resulting in only 59 having both the mRNA and protein abundance determined. For each of the 59 pairs, the base-2 logarithm (log2) of the ratio relative to 0 hours was calculated for 12, 24 and 72 hours of fasting, and correlations between the two types of data were calculated. Taken together, there was very little overall correlation, with a Spearman rank correlation coefficient of 0.21 (Figure 1). The positive value of the coefficient nevertheless points to a slight positive overall correlation between changes in mRNA and protein levels. A list of all log2 ratios can be found in *Supplemental file 1* (see section 3.1).

The picture was very different for specific genes. There were highly correlating, differentially expressed transcript-protein pairs, such as *Aldob* and *Atp5h* (Figure 1) and *Vim* (not shown). Others, on the other hand, showed a negative correlation, with the transcript changing in the opposite direction of the protein. In most of these cases, the transcript was up- and the corresponding protein down-regulated. Examples are *Eef2* (elongation factor 2, Figure 1), *Aldh1b1*, *Arhgdia*, *Hnrnpa2b1* and *Uqcrc1*. For *Eef2*, a similar divergence of mRNA and protein levels upon fasting was reported earlier in liver and muscle [28].

Some of the proteins identified in multiple spots showed variable expression of different isoforms, indicating that post-translational modifications could play a role in regulation of their activity. The proteins that were potentially post-translationally modified included metabolic

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

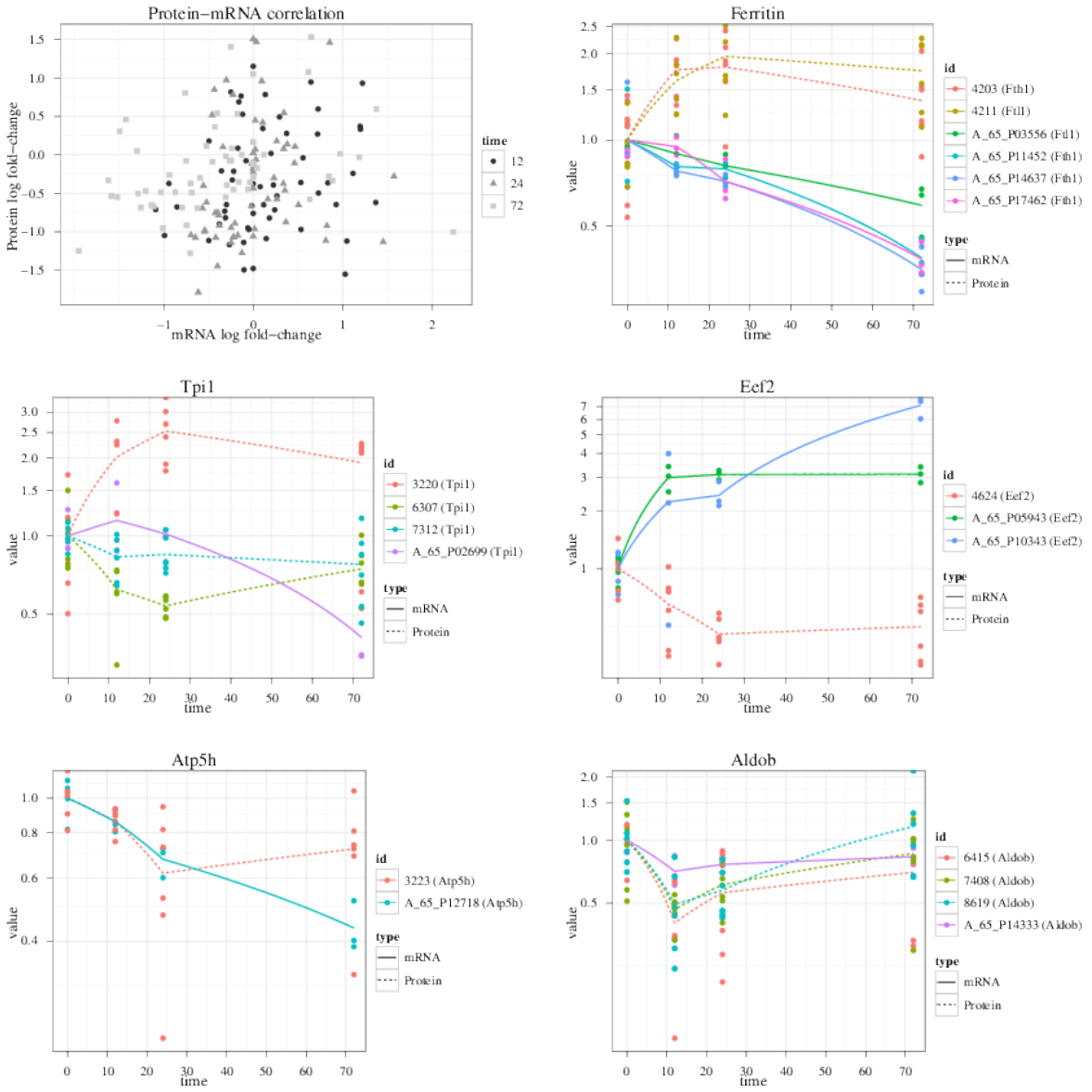

Figure 1: Combined mRNA and protein expression plots. In the top-left plot, the fold-change of protein and mRNA expression are plotted against each other. Fold-changes are calculated for each time point against t=0. In cases where multiple protein spots correspond to the same gene, the average was used. The overall correlation plot shows that there was very little agreement between protein and gene expression. The five other plots show individual genes in detail. Each dot represents a measurement of a spot or probe. Lines connect the average of each probe or spot per time point. Solid lines represent transcripts, dashed lines represent proteins. The average intensities at 0 hours have been normalized to 1; all other values are relative to this average. Top-right: Ferritin heavy (*Fth1*) and light chain (*Ftl1*), showing opposite trends for transcript and protein expression levels. Center-left: Triose phosphate isomerase (*Tpi1*), which is down-regulated at 72 hours with the exception of one protein spot. Center-right: Elongation Factor 2 (*Eef2*), an example of a gene that shows opposite effects of fasting on transcript and protein expression. Bottom-left: ATP synthase D chain (*Atp5h*) and bottom-right: fructose biphosphate aldolase B (*Adlob*), are examples of two genes that show good correlation between protein and transcript levels.

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

enzymes (TPI1 and ATP5H), proteins related to protein folding (CALR, HSPA8 and HSPA5) and cytoskeletal proteins (KRT19, ACTB, ACTG2 and VIL1).

3.4 Pathway visualization

To visualize protein and gene expression data side-by-side in pathway context, we used the PathVisio program [25]. High-throughput datasets in tab-delimited text format are imported using the expression data import wizard plugin. Data should be normalized and preferably log-transformed before import. *Supplemental file 2* contains the data from this study in a format suitable for PathVisio. During data import, the user can select a column that contains gene or protein identifiers. PathVisio allows direct import of mixed identifiers without the need to pre-process the data. After the import step, the user can configure the color representation of the data. PathVisio provides rule-based and color gradient-based visualization options. In the rule-based visualization the user defines a Boolean expression to specify a color for elements that evaluate as true. The gradient-based visualization maps numerical values to a color gradient. Colors are displayed in the gene boxes. Multiple conditions can be displayed side-by-side, and asterisks can be added to the diagram to indicate significantly changed measurements. *Supplemental file 3* is a tutorial, containing step-by-step instructions to reproduce these visualizations.

A feature of particular interest for omics integration is the fact that if multiple data points map to the same box, it can be divided horizontally for each corresponding row in the dataset. Thus, the box is used as a small heat map representation of a subset of the data, where each column represents a condition, and each row represents a probe. The problem of a variable number of data points per box occurs when visualizing data from microarrays that have more than one probe per gene, but it is especially important for omics integration, which often deals with two datasets of very unequal size (in this case 130 proteins versus 10,696 transcripts), which automatically means that some boxes have more data than others. Although this subdivision means that the boxes can get cramped, we find this approach superior to summarization, which inevitably means discarding data. In the example of TPI1 (Figure 2), the two rows are clearly differentially expressed, but this insight would be lost if the average value of these measurements was used. If needed, the boxes can be manually enlarged. Data can be visualized together for direct comparison, but it is also possible to create separate visualizations and toggle between them using a drop-down list. Visualizing different time points separately can prevent confusion and decrease the chance of an erroneous comparison of a gene at one time point with another gene at a different time point. Separating the time points into different visualizations also helps to combat the information overload in a single box.

The resulting images of pathways with overlayed data can be exported as a set of HTML image maps, which can be viewed in any browser without the need to install PathVisio. The HTML image maps that resulted from the present study can be found in *Supplemental file 4* (see section 3.1).

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

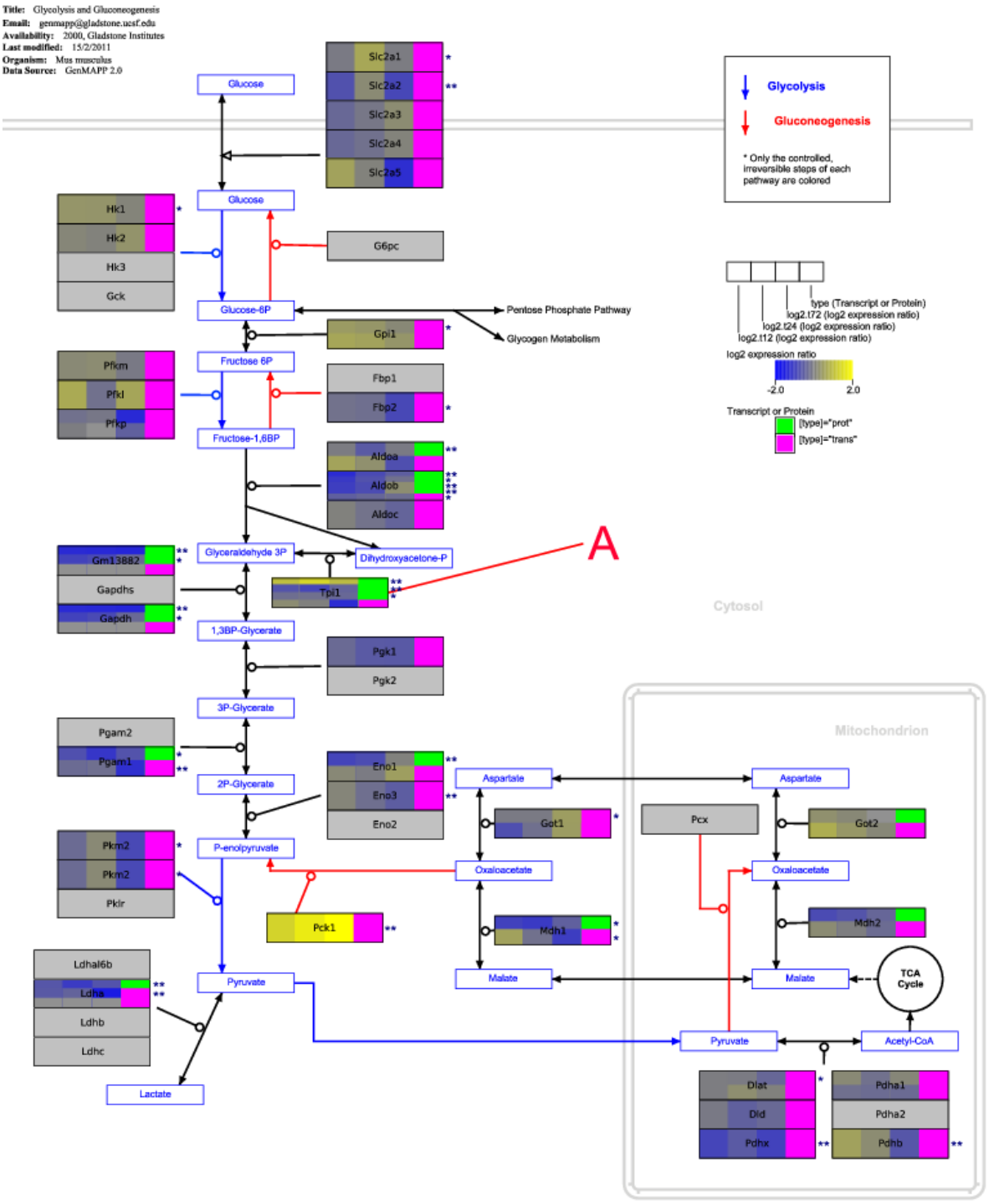

Figure 2: Glycolysis and Gluconeogenesis. Visualization of expression data on the glycolysis / gluconeogenesis pathway. Each colored box represents a gene product. Blue indicates decreased expression levels, yellow increased. Each box is a heat map with rows representing probes or spots, and columns representing time points. The rightmost column is a flag that indicates if the given row is a protein spot (green) or microarray probe (pink). The asterisks denote significance. Multiple probes and/or spots can be shown in a box. *Tpi1* is marked with the letter A. More pathway diagrams can be found in *supplemental file 2*.

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

3.5 Integration at the gene / protein level

The combination of proteomics and transcriptomics data provides us more information than just one of the two datasets on its own. Global correlation is not high, but on a gene-by-gene level we see a very varied picture. Gene/protein pairs that do not show high correlation present leads for investigation into post-transcriptional regulation.

Although proteomics data have fewer data points than transcriptomics data, there are a few instances where important transcripts were not measured, due to absence of a corresponding probe on the array. In those cases, protein measurements can fill important gaps in pathways. For example, no mRNA expression levels were measured for *Otc2* and *Arg2* due to absence of probes for these genes on the microarray, but their protein concentrations were measured and show a clear down-regulation, in particular in the early (12 hours) response. This is consistent with suppression of citrulline synthesis without an additional increase in arginine catabolism and with glutamine conservation via suppression of *Oat*, *Pycs* and *Gls*, and the upregulation of *Gln* (see Figure 3).

Similarly, the down-regulation of ACAA2 (only measured as protein) is completely consistent with the reduction of fatty acid biosynthesis, also supported by the down-regulation of genes such as *Hadh1* (*Supplemental file 4*). Additionally, down-regulation of the GAPDH protein, indicating an overall decrease in glycolytic and gluconeogenic activity, is entirely consistent with lower expression of other genes in the same pathway, such as *Pgam1* and *Eno3* (Figure 2).

Analysis of the transcriptome has revealed strong effects on cell cycle and apoptosis, by down-regulation of cyclins and upregulation of their inhibitors, cyclin-dependent kinases ([1]; *Supplemental file 4*). Cell turnover in intestine is thought to be a major source of energy expenditure, and its (down)regulation in fasting is in good agreement with a need for energy preservation. Unfortunately, no cyclins or other proteins related to cell cycle regulation and apoptosis could be identified, most likely because their level of expression was too low for detection by the 2DE technique. The comparison of the two datasets underlines the problem of bias in proteomics techniques. Proteomics analysis alone would have missed a major regulatory effect of starvation in the gut.

The importance of glucose metabolizing pathway in fasting (Figure 2) is stressed by sheer number of genes differentially expressed. Out of 11 gene/protein couples found in the pathway, only in 4 both transcript and protein were significantly differentially expressed (*Pgam1*, *Ldha*, *Aldob* and *Mdh1*). In all 4, interestingly, the direction of the change was the same. The direction was also the same in case of *Got2* and *Mdh2*, in which changes occurred in both data types, but have not reached significance. For two glyceraldehyde dehydrogenases only changes in protein expression were detectable, while for *Aldoa* and *Eno1* the direction of the change differed between gene and protein. *Tpi1* had different direction of change between different protein isoforms (discussed in more detail below). Data integration and visualisation convey therefore a clear message that in fasting some of the proteins in this pathway must be regulated by other means than just transcription rate.

A difference in any of 500 known posttranslational modifications [29] (e.g. phosphorylation states) can lead to different spots in a 2D gel. Such a change could then mean either that the

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

experimental condition has led to a change in total quantity of the corresponding protein, or to a change in the functional state of the protein (or both), but due to the incompleteness of proteomics technology, these two possibilities can hardly be distinguished. Unfortunately, a decrease at the spot level can therefore not be straightforwardly interpreted as a decrease in functional activity. A clear example of differentiation at the spot level is TPI1, an important enzyme of the gluconeogenesis pathway, which is increased in spot 3,220 but decreased in spots 6,307 and 7,312 throughout the fasting period (Figure 1). From the estimated mass as well as the mass spectrum it appears that the protein in spot 3,220 lacks an N-terminal fragment. Alternative splicing or proteolysis could play a role in the regulation of this protein. Assuming that the partial TPI1 protein has a reduced activity, this finding is consistent with a reduced activity in the glycolytic pathway. However, to determine the exact nature of the observed change in TPI1 and its consequence for enzyme activity, follow-up experiments are necessary.

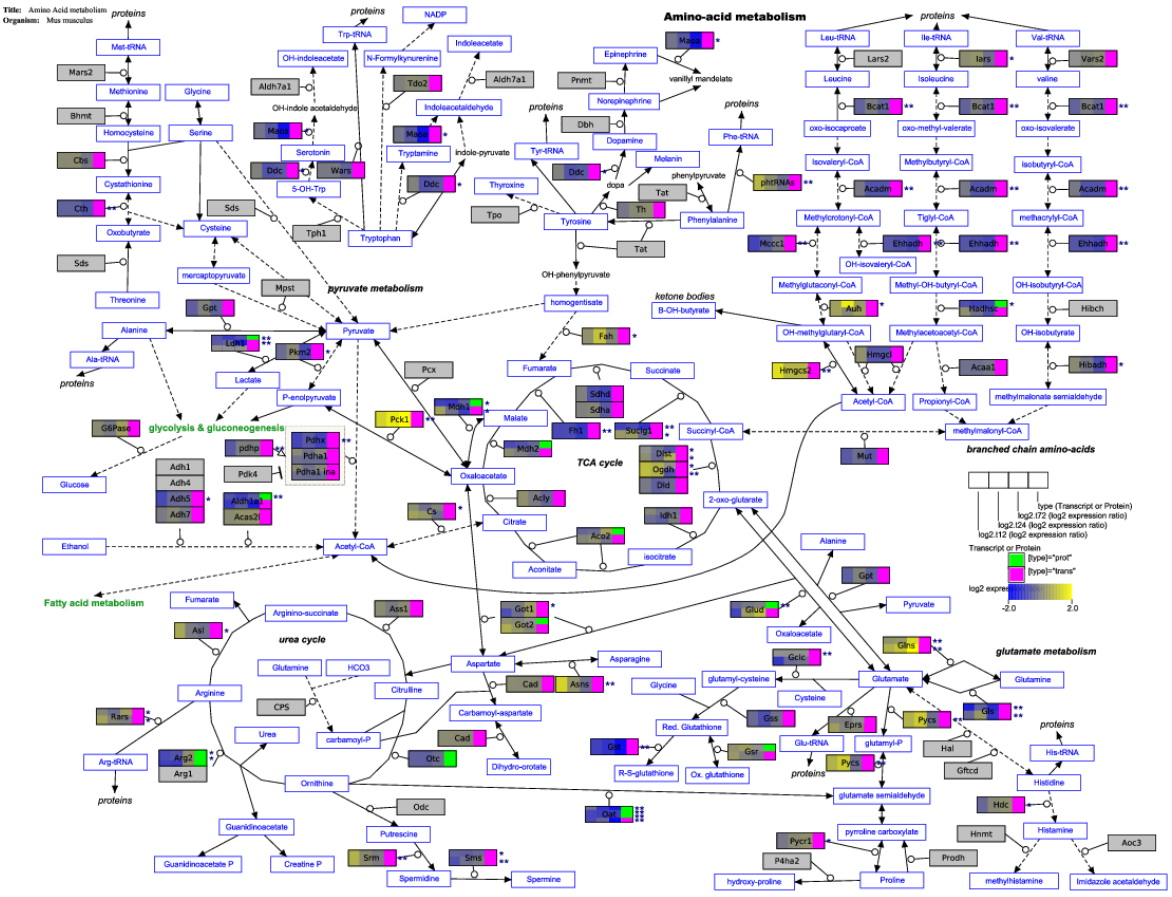

Figure 3: Amino acid metabolism. Visualization of expression data on the amino acid metabolism pathway. Coloring is identical to Figure 2. The genes *Otc2*, *Arg2* and *Oat*, mentioned in the main text, can be found in the bottom-left quadrant of the image. *Gls*, *Glns* and *Pycs* can be found in the bottom-right quadrant.

Another protein affected in an interesting way was ferritin, a protein necessary for the storage of iron in tissues. Ferritin was down-regulated at the gene level, but up-regulated at the protein level throughout starvation (Figure 1), and this was the case for both its heavy and light chain. Though dietary iron is unavailable in fasting, the prevention of iron release from the existing stores would limit its availability to the invading microorganisms, especially in the small intes-

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

tine. The need for such a response is accentuated by the impressive downregulation of immune response seen in the intestine in our recent study [22]. The increased protein abundance in spite of lower transcript levels is harder to explain. It has been shown that mRNA turnover, regulated by the iron responsive element binding protein (*Ireb2*, not measured in this study), has a strong effect on ferritin [30], which could explain the discrepancy between protein and mRNA abundance. Another possible explanation is that the increased values are caused by a change occurring in erythrocytes, which can never be fully excluded from the protein sample. Other possibilities, such as the presence of other ferritin spots in the gel that have not yet been identified, cannot be ruled out.

4 Discussion

A number of existing applications can perform visualization of high-throughput datasets, such as KEGG Atlas [31], ProMeTra [32], Vanted [33] and Reactome SkyPainter [34], and reviewed in [35, 36]. Some of those have demonstrated the capability to perform pathway visualization with multiple omics datasets together in one pathway, such as ProMeTra, which is focused on combining metabolomics and transcriptomics data. The automated mapping of mixed identifiers is a distinguishing feature of PathVisio that makes it particularly suited for integration and simultaneous visualization of datasets from different sources. A table comparing the identifier mapping and data visualization functionality of the different tools can be found in the *Supplemental file 5*.

Our examples showed that proteomics data can reinforce the conclusions deduced from transcriptomics data, and simultaneously indicate areas where post-transcriptional regulation plays a role. The 2DE technique by itself does not provide enough data for an overview on systems biology level. In this study, if only proteomics data had been available, important pathways such as apoptosis and cell cycle regulation would have been missed entirely. Nevertheless, protein expression data do provide interesting insights into regulation on a gene-by-gene basis. The proteins that do not correlate well are of particular interest, at the very least for generating hypotheses for follow-up experiments. With the maturation of proteomics (and metabolomics) technology, the issue of number of measurements will lose the impact and significance, only increasing the importance of being able to visualize the different datasets simultaneously.

The interpretation is not straightforward, and the correlation of gene and protein expression levels (or lack thereof) must be interpreted on a case-by-case basis. Pathway visualization can serve as a useful aid, given the role of pathways as a knowledge base of biological information. Flexible identifier mapping is essential for data integration. The relation between genes, microarray probes and protein spots is not straightforward, which makes software support for automated identifier mapping essential.

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Acknowledgements

This work was supported by the Dutch Ministry of Economic Affairs through the Innovation Oriented Research Program on Genomics: IOP Genomics IGE01016, and by the transnational University of Limburg (tUL). This work was carried out within the research program of the Netherlands Consortium for Systems Biology (NCSB), which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research.

List of abbreviations

2DE	2D Gel Electrophoresis
Acaa2	acetyl-Coenzyme A aminotransferase 2
Actb	actin, beta
Actg2	actin, gamma 2
Aldh1b1	aldehyde dehydrogenase 1 family, member B1
Aldob	fructose biphosphate aldolase B
Arg2	arginase-2
Arhgdia	Rho GDP dissociation inhibitor alpha
Atp5h	ATP synthase D chain
BLAST	Basic Local Alignment Search Tool
Calr	Calreticulin
Eef2	elongation factor 2
Eno3	enolase 3
Fth1	ferritin heavy chain 1
Ftl1	ferritin light chain 1
Gapdh	glyceraldehyde-3-phosphate dehydrogenase
Glns	glutamine synthetase
Gls	glutaminase
Hadh1	hydroxyacyl-Coenzyme A dehydrogenase
Hnrnpa2b1	heterogeneous nuclear ribonucleoprotein A2/B1
Hspa5	heat shock protein 5
Hspa8	heat shock protein 8
Ireb2	iron responsive element binding protein 2
Krt19	keratin 19
Oat	ornithine aminotransferase
Otc2	ornithine carbamoyltransferase
Pgam1	phosphoglycerate mutase 1
Pycs	pyrroline 5 carboxylate synthetase
Tpi1	triose phosphate isomerase 1
Uqcrc1	ubiquinol-cytochrome c reductase core protein 1
Vil1	villin 1
Vim	vimentin

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

References

[1] M. Sokolović, D. Wehkamp, A. Sokolović et al. Fasting induces a biphasic adaptive metabolic response in murine small intestine. *BMC Genomics*, 8(1):361, 2007.

[2] K. Lenaerts, M. Sokolovic, F. G. Bouwman, W. H. Lamers, E. C. Mariman and J. Renes. Starvation induces phase-specific changes in the proteome of mouse small intestine. *Journal of Proteome Research*, 5(9):2113–2122, 2006.

[3] L. Nie, G. Wu, D. E. Culley, J. C. Scholten and W. Zhang. Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications. *Critical Reviews in Biotechnology*, 27(2):63–75, 2007.

[4] L. Anderson and J. Seilhamer. A comparison of selected mrna and protein abundances in human liver. *Electrophoresis*, 18(3-4):533–537, 1997.

[5] S. P. Gygi, Y. Rochon, B. R. Franza and R. Aebersold. Correlation between protein and mrna abundance in yeast. *Molecular and Cellular Biology*, 19(3):1720–1730, 1999.

[6] T. Ideker, V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold and L. Hood. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292(5518):929–934, 2001.

[7] M. P. Washburn, A. Koller, G. Oshiro, R. R. Ulaszek, D. Plouffe, C. Deciu, E. Winzeler and J. R. Yates. Protein pathway and complex clustering of correlated mrna and protein expression analyses in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences U.S.A.*, 100(6):3107–3112, 2003.

[8] B. Pradet-Balade, F. Boulmé, H. Beug, E. W. Müllner and J. A. Garcia-Sanz. Translation control: bridging the gap between genomics and proteomics? *Trends in Biochemical Sciences*, 26(4):225–229, 2001.

[9] F. Chevalier. Review highlights on the capacities of Gel-based proteomics. *Proteome Science*, 8(1):23, 2010.

[10] M. R. Roe and T. J. Griffin. Gel-free mass spectrometry-based high throughput proteomics: Tools for studying biological response of proteins and proteomes. *Proteomics*, 6(17):4678–4687, 2006.

[11] J. M. Argilés. Cancer-associated malnutrition. *European Journal of Oncology Nursing*, 9(Supplement 2):S39–S50, 2005.

[12] Y. Le Maho, H. Vu Van Kha, H. Koubi, G. Dewasmes, J. Girard, P. Ferre and M. Cagnard. Body composition, energy expenditure, and plasma metabolites in long-term fasting geese. *American Journal of Physiology-Endocrinology And Metabolism*, 241(5):E342–E354, 1981.

[13] M. Caloin. Modeling of lipid and protein depletion during total starvation. *American Journal of Physiology-Endocrinology and Metabolism*, 287(4):E790–E798, 2004.

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

[14] Y. Cherel, D. Attaix, D. Rosolowska-Huszcz, R. Belkhou, J.-P. Robin, M. Arnal and Y. Le Maho. Whole-body and tissue protein synthesis during brief and prolonged fasting in the rat. *Clinical Science*, 81(5):611–619, 1991.

[15] Y. Cherel and Y. Le Maho. Refeeding after the late increase in nitrogen excretion during prolonged fasting in the rat. *Physiology & Behavior*, 50(2):345–349, 1991.

[16] C. Habold, C. Chevalier, S. Dunel-Erb, C. Foltzer-Jourdainne, Y. Le Maho and J.-H. Lignot. Effects of fasting and refeeding on jejunal morphology and cellular activity in rats in relation to depletion of body stores. *Scandinavian Journal of Gastroenterology*, 39(6):531–539, 2004.

[17] M. Bauer, A. C. Hamm, M. Bonaus, A. Jacob, J. Jaekel, H. Schorle, M. J. Pankratz and J. D. Katzenberger. Starvation response in mouse liver shows strong correlation with life-span-prolonging processes. *Physiological Genomics*, 17(2):230–244, 2004.

[18] R. T. Jagoe, S. H. Lecker, M. Gomes and A. L. Goldberg. Patterns of gene expression in atrophying skeletal muscles: response to food deprivation. *The FASEB Journal*, 16(13):1697–1712, 2002.

[19] S. H. Lecker, R. T. Jagoe, A. Gilbert, M. Gomes, V. Baracos, J. Bailey, S. R. Price, W. E. Mitch and A. L. Goldberg. Multiple types of skeletal muscle atrophy involve a common program of changes in gene expression. *The FASEB Journal*, 18(1):39–51, 2004.

[20] M. Sokolović, A. Sokolović, D. Wehkamp, E. V. L. van Themaat, D. R. de Waart, L. A. Gilhuijs-Pederson, Y. Nikolsky, A. H. C. van Kampen, T. B. M. Hakvoort and W. H. Lamers. The transcriptomic signature of fasting murine liver. *BMC Genomics*, 9(1):528, 2008.

[21] X. Q. Xiao, K. L. Grove, B. E. Grayson and M. S. Smith. Inhibition of uncoupling protein expression during lactation: role of leptin. *Endocrinology*, 145(2):830–838, 2004.

[22] T. B. M. Hakvoort, P. D. Moerland, R. Frijters et al. Interorgan Coordination of the Murine Adaptive Response to Fasting. *Journal of Biological Chemistry*, 286(18):16332–16343, 2011.

[23] K. Lenaerts, E. Mariman, F. Bouwman and J. Renes. Glutamine regulates the expression of proteins with a potential health-promoting effect in human intestinal Caco-2 cells. *Proteomics*, 6(8):2454–2464, 2006.

[24] M. P. van Iersel, A. R. Pico, T. Kelder, J. Gao, I. Ho, K. Hanspers, B. R. Conklin and C. T. Evelo. The BridgeDb framework: standardized access to gene, protein and metabolite identifier mapping services. *BMC Bioinformatics*, 11(1):5, 2010.

[25] M. P. van Iersel, T. Kelder, A. R. Pico, K. Hanspers, S. Coort, B. R. Conklin and C. Evelo. Presenting and exploring biological pathways with PathVisio. *BMC Bioinformatics*, 9(1):399, 2008.

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

[26] T. Kelder, M. P. van Iersel, K. Hanspers, M. Kutmon, B. R. Conklin, C. T. Evelo and A. R. Pico. WikiPathways: building research communities on biological pathways. *Nucleic Acids Research*, 40(D1):D1301–D1307, 2012.

[27] K. M. Waters, J. G. Pounds and B. D. Thrall. Data merging for integrated microarray and proteomic analysis. *Briefings in Functional Genomics & Proteomics*, 5(4):261–272, 2006.

[28] F. Yoshizawa, Y. Miura, K. Tsurumaru, Y. Kimata, K. Yagasaki and R. Funabiki. Elongation factor 2 in the liver and skeletal muscle of mice is decreased by starvation. *Bioscience, Biotechnology, and Biochemistry*, 64(11):2482–2485, 2000.

[29] R. G. Krishna and F. Wold. Post-translational modifications of proteins. In K. Imahori and F. Sakiyama (editors), *Methods in Protein Sequence Analysis*, pages 167–172. Springer, 1993.

[30] K. J. Hintze and E. C. Theil. Cellular regulation and molecular interactions of the ferritins. *Cellular and Molecular Life Sciences*, 63(5):591–600, 2006.

[31] S. Okuda, T. Yamada, M. Hamajima, M. Itoh, T. Katayama, P. Bork, S. Goto and M. Kanehisa. KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Research*, 36(suppl 2):W423–W426, 2008.

[32] H. Neuweger, M. Persicke, S. P. Albaum, T. Bekel, M. Dondrup, A. T. Hüser, J. Winnebald, J. Schneider, J. Kalinowski and A. Goesmann. Visualizing post genomics datasets on customized pathway maps by ProMeTra–aeration-dependent gene expression and metabolism of Corynebacterium glutamicum as an example. *BMC Systems Biology*, 3(1):82, 2009.

[33] C. Klukas and F. Schreiber. Integration of -omics data and networks for biomedical research with VANTED. *Journal of Integrative Bioinformatics*, 7(2):112, 2010.

[34] D. Croft, A. F. Mundo, R. Haw et al. The Reactome pathway knowledgebase. *Nucleic Acids Research*, 42(Database issue):D472–477, 2014.

[35] N. Gehlenborg, S. I. O'Donoghue, N. S. Baliga et al. Visualization of omics data for systems biology. *Nature Methods*, 7:S56–S68, 2010.

[36] A. R. Joyce and B. Ø. Palsson. The model organism as a system: integrating 'omics' data sets. *Nature Reviews Molecular Cell Biology*, 7(3):198–210, 2006.

Copyright 2014 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).