Eurographics/ IEEE-VGTC Symposium on Visualization 2008
A. Vilanova, A. Telea, G. Scheuermann, and T. Möller
(Guest Editors)

Volume 27 (2008), Number 3

Visualizing Genome Expression and Regulatory Network Dynamics in Genomic and Metabolic Context

M. A. Westenberg[1], S. A. F. T. van Hijum[2,3], O. P. Kuipers[2], J. B. T. M. Roerdink[1]

[1]Institute for Mathematics and Computing Science,
University of Groningen, P.O. Box 407, 9700 AK Groningen, The Netherlands
[2]Department of Genetics, Groningen Biomolecular Sciences and Biotechnology Institute,
University of Groningen, P.O. Box 14, 7950 AA Haren, The Netherlands
[3]Interfacultary Centre for Functional Genomics, Ernst-Moritz-Arndt-Universität, Greifswald 17489, Germany

Abstract

DNA microarrays are used to measure the expression levels of thousands of genes simultaneously. In a time series experiment, the gene expressions are measured as a function of time. We present an application for integrated visualization of genome expression and network dynamics in both regulatory networks and metabolic pathways. Integration of these two levels of cellular processes is necessary, since it provides the link between the measurements at the transcriptional level (gene expression levels approximated from microarray data) and the phenotype (the observable characteristics of an organism) at the functional and behavioral level. The integration requires visualization approaches besides traditional clustering and statistical analysis methods. Our application can (i) visualize the data from time series experiments in the context of a regulatory network and KEGG metabolic pathways; (ii) identify and visualize active regulatory subnetworks from the gene expression data; (iii) perform a statistical test to identify and subsequently visualize pathways that are affected by differentially expressed genes. We present a case study, which demonstrates that our approach and application both facilitates and speeds up data analysis tremendously in comparison to a more traditional approach that involves many manual, laborious, and error-prone steps.

Categories and Subject Descriptors (according to ACM CCS): I.3.8 [Computer Graphics]: Applications, J.3 [Life and Medical Sciences]: Biology and genetics

1. Introduction

Integration of biological interaction processes, which not only take place at genomic, proteomic, and metabolomic levels, but also between these levels, is important in systems biology. Software frameworks are established that visualize such interaction networks, and which offer interactive exploration to a researcher [BBO*06].

In this paper, we consider visualization of gene regulatory networks and pathways. Gene regulatory networks can be represented by graphs, in which nodes represent genes, and edges represent interactions between a gene product (a regulator protein) and its target genes. The nodes may have several attributes, such as position on the chromosome, and gene expression attributes for multiple time points together with p-values indicating statistical significance.

Biologists also study how genes, gene products, enzyme reactions, and compounds form together a pathway, which is made up of chemical reactions that catalyze transformation of compounds. Compounds are central building blocks in cellular components, energy storage, and are often used in intra- and extra-cellular signalling. A well-established annotation system of pathways is provided by the Kyoto Encyclopedia of Genes and Genomes (KEGG) [OGK*00]. It covers a wide range of organisms, and its manually constructed pathway drawings are similar to textbook pictures.

Despite the existence of many tools for visualizing the various types of biological networks, a recent review paper

© 2008 The Author(s)
Journal compilation © 2008 The Eurographics Association and Blackwell Publishing Ltd.
Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

by Suderman & Hallett [SH07] shows that there is a need for tools that incorporate dynamic information in the networks, and for tools that support both data analysis and visualization, so that a researcher does not have to switch between various applications that perform these tasks separately. Visualization of gene expression data from time series has been limited to displaying either the whole time series in a node or coloring a node based on a single time point. Since no information is mapped to the edges, it is very difficult to understand which substructures of the network are active at any point in time. The complexity of the network connection pattern also makes it hard to determine whether two or more active genes are actually connected.

Here, we present an application for integrated visualization of genome expression and network dynamics in both regulatory networks and metabolic pathways. Integration of these two levels of cellular processes is necessary, since it provides the link between the measurements at the transcriptional level (gene expression levels approximated from microarray data) and the phenotype (the observable characteristics of an organism) at the functional and behavioral level. The novel aspects are:

- glyph-based visualization of gene expression data from time series experiments and associated statistical data
- visualization of the time series at multiple levels of abstraction, specifically in the contexts of regulatory networks and metabolic pathways
- visualization of regulatory network dynamics based on the time series
- identification of pathways affected by active networks

The organization of this paper is as follows. In Section 2, we review related work in regulatory network and pathway visualization. We then describe our visualization techniques and application in Section 3. Section 4 contains some results from a case study performed by an expert in the bioinformatics field, and, finally, we draw conclusions in Section 5.

2. Related work

There exist many tools in bioinformatics for visualization of biological networks, see Saraiya et al. [SND05] and Suderman & Hallett [SH07] for recent overviews. This section reviews the most important of these applications.

Some applications are generic network visualization tools, and do not distinguish between, e.g., gene transcription networks and metabolic pathways. Examples are Osprey [BST03], which was one of the first biological interaction network visualization tools, and Cytoscape [SMO*03], which is a popular data analysis and visualization tool. Both support superimposing gene expression data on the nodes in the network by color-coding, however, this is limited to a single time point from a time series experiment. VisANT [HNY*07] is a tool for biological network analysis and visualization. It focuses strongly on the analysis of networks of various types, such as protein-protein interaction networks, gene transcription networks, metabolic pathways, and interconnections between these. Network analysis is performed by calculating topological statistics and features, or querying a server-side database for functional gene annotations. In its latest version, VisANT supports visualization of gene expression data from time series experiments by either color-coding nodes or embedding an expression profile plot inside the node.

Other tools are more specific to pathway analysis and visualization. BiologicalNetworks [BSRG06] has a strong focus on data integration and analysis. It supports visualization of time series data in matrix form and by plots, but only allows the user to overlay one time point at a time on a pathway. VANTED [JKS06] visualizes experimental data inside pathway nodes by bar or line charts. It also provides several data analysis techniques such as clustering by gene expression profile and extraction of correlation networks. Some other tools provide time series expression data overlays on pathway nodes [MSH*05, WES07, WGdS00].

None of the above tools map statistical properties of the expression data to a meaningful visual attribute. Statistical properties are very important in microarray data analysis, because they provide a measure of confidence for replicated measurements. Some tools allow varying attributes of a node other than the fill color. However, browsing through time points is then cumbersome, since two distinct attributes of a node need to be changed (usually manually). It is also hard to incorporate statistical information into an expression plot embedded in a node. The first tool to combine expression value and associated statistical data into a single glyph is GENeVis [WvHL*07]. All time points are mapped to rectangular glyphs, which are embedded in the nodes.

3. Visualization Design

This section first describes design choices for visualizing gene expression time series, regulatory networks, and pathways. This is followed by the method to identify active subnetworks, and the method to find pathways affected by network activity. Finally, we describe the application that integrates these components.

3.1. Genome Representation

The genes of a given genome are represented in a table in which each row corresponds to a single gene. The table columns contain for each gene the locus tag, common gene name, start and end position on the genome, UniProt ID [Uni07], and a data field to store additional information, such as the function of a gene. The locus tag is a unique identifier for a gene within one organism. The UniProt ID is used to link a gene to the KEGG pathway database. The remaining columns serve as annotation information. We use

© 2008 The Author(s)
Journal compilation © 2008 The Eurographics Association and Blackwell Publishing Ltd.

a rectangular box annotated with the gene name to represent a gene as a node in the network visualizations.

3.2. Time Series Visualization

Gene expression is measured by DNA microarrays and gives a measure of the amount of messenger RNA (mRNA) produced after transcription of each gene. Gene expression values can either be absolute levels of expression or ratios of expression levels between a test and a reference condition. Ratios are usually log-transformed to obtain comparable scales for ratios above and below 1. To each expression value, a statistical value is associated, which expresses the reliability of the measurement. Commonly, the coefficient of variation is used in the case of expression levels and a p-value is used in the case of expression ratios.

We follow the approach as used in GENeVis [WvHL*07], and draw each time point as a colored rectangular glyph. The expression value is mapped to a color, and the reliability value mapped to the glyph's height: the more reliable, the higher the rectangle. High expression values that are also reliable are most important for the interpretation. Therefore, reliability is mapped to the stronger perceptual cue of size to give reliable data more emphasis than unreliable data. The expression glyphs for all time points are drawn inside the gene boxes, which facilitates the visual identification of outliers and node comparison [SLN05]. To support analysis of the network at a single time point, the gene expression value can also be used as a fill color for the gene box.

Expression levels are mapped to colors that range from white to black via yellow and red. Expression ratios are mapped to colors that range from green to red via black. The use of these colormaps is standard practice in the bioinformatics field. The expression values are not mapped linearly to colors, instead, we divide the expression value range in a number of quantiles. Each quantile is assigned a color from the color map, which provides the user with some insight in the statistical distribution of the data.

3.3. Regulatory Network Visualization

A regulatory network is a network of proteins that control the expression of their respective target genes. These interactions form the edges in the graphical representation. A regulator either inhibits or activates its target, which is represented by decorating the target end of an edge by a bar or an arrow head, respectively. These symbols have become fairly standard in regulatory network visualization. The edges are also colored, green for activation and red for inhibition, to make identification of the interaction type of an edge that extends over a large distance easier. Since biological data are often incomplete, an unknown interaction type is implemented: the edge is not decorated and has a grey color.

The layout of the network is computed by a force-directed algorithm. This produces satisfactory layouts for this type of network, since the grouping of nodes corresponds quite well with the biological concept of a regulon (a collection of genes under regulation by the same protein). The layout algorithm can be run interactively, and a user can drag nodes to other positions until satisfied with the layout. The application allows the user to save a network as a GraphML file, so that the same layout can be used in subsequent runs.

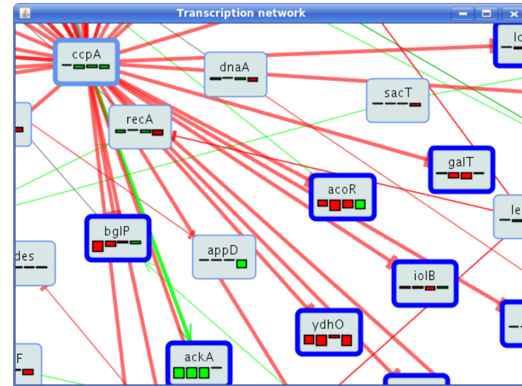

Figure 1: *The network window shows the transcription network and gene expression time series. The data shown correspond to a part of the gene regulatory network of the bacterium B. subtilis, and an expression ratio time series of four time points. Neighbor highlighting (of the gene ccpA) assists the user in understanding network structure. The interaction type is mapped to a color: red for inhibition, green for activation, and grey for other cases.*

The regulatory network is displayed in the application's network window, see Fig. 1. It has several interactive features, besides zoom and pan. Pointing the mouse to a gene box highlights a node and its direct neighbors, which makes it easier to see the network structure. The image shows highlighting for the gene *ccpA* (the biological convention is that gene names are written in italics with the first letter in lower case). The network window supports gene selection by (shift-)clicking the mouse on nodes, by a rubberband method, and by regulon (node and its neighbors). Nonselected nodes are made semi-transparent so that they cause less distraction. Hovering over a gene box also pops up a tooltip that contains information from the genome table.

3.4. KEGG Pathway Visualization

KEGG provides a web service interface to the database, which ensures that a user has always access to the latest updates. However, remote database access is very slow, and it can take up to several seconds to complete a single query, such as retrieving all pathways that involve a certain gene. Since it is not acceptable to wait a long time for results in an

© 2008 The Author(s)
Journal compilation © 2008 The Eurographics Association and Blackwell Publishing Ltd.

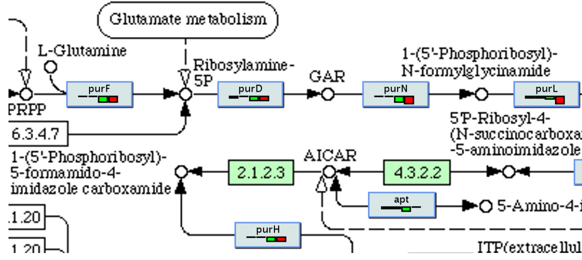

Figure 2: *Visualization of an expression ratio time series of four time points in the context of a part of the purine metabolism pathway in B. subtilis.*

interactive application, we construct a local cache of a part of the KEGG database for faster access.

The cache is constructed once for a given organism. It stores mappings from genes to pathways and conversely, pathway images in GIF format, and pathway descriptions. The description contains a KEGG identifier, a textual description, and the list of genes in that pathway and their coordinates in the corresponding pathway drawing. The pathway data are obtained by parsing the HTML formatted data returned by the database.

A pathway is visualized in the KEGG Viewer by drawing the GIF image with the gene boxes drawn on top of this image. An example is shown in Fig. 2, which contains a part of the purine metabolism pathway in the bacterium *B. subtilis* overlaid with a four point expression ratio time series.

3.5. Active Network Visualization

We make use of a trace-back algorithm [LBY*04] to identify subnetworks that are active at a given time point. Denote by $L_{g,t}$ and $R_{g,t}$ the expression level and expression ratio (log-transformed) of the gene g at time point t, respectively. The expression level $L_{g,t}$ is *low* if $L_{g,t} < T_m$, *medium* if $T_m \leq L_{g,t} < T_h$, and *high* if $L_{g,t} \geq T_h$, where T_m and T_h are thresholds so that $0 \leq T_m < T_h$. A gene is differentially expressed if $|R_{g,t}| \geq T_r$, where $T_r > 0$ is a threshold. All thresholds can be adjusted, since they may be data dependent. The detection algorithm proceeds as follows:

1. Determine active regulators (nodes with at least one outgoing edge) in the network. A regulator gene g is active at time point t if one of the following conditions holds:

$$L_{g,t} < T_m \text{ and } R_{g,t} \geq T_r \quad (1)$$

$$T_m \leq L_{g,t} < T_h \text{ and } R_{g,t} \geq 0 \quad (2)$$

$$L_{g,t} \geq T_h \quad (3)$$

2. Determine active nonregulator genes (nodes with only incoming edges or with degree zero). A nonregulator gene g is active at time point t if

$$|R_{g,t}| \geq T_r \quad (4)$$

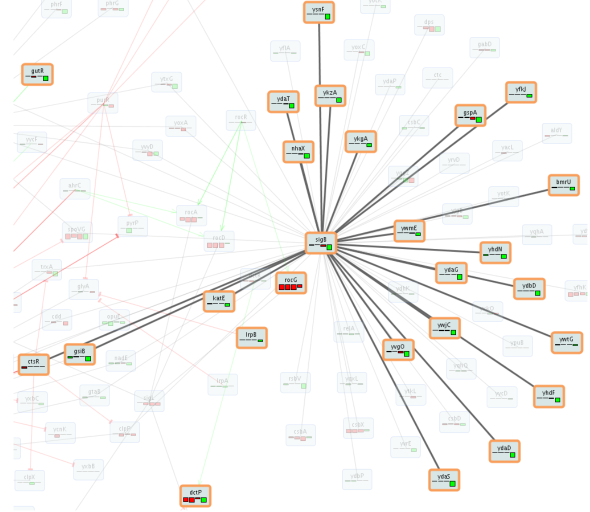

Figure 3: *Part of the active network in B. subtilis at the last time point of a four-point time series. Active nodes and edges are drawn opaque and highlighted.*

3. Determine active links. A link between a source gene and target gene is active if both endpoints are active.

After the active subnetwork has been highlighted by an increased border width, we render all inactive genes and edges semi-transparent. We draw the active network in the context of the complete network to support the user in maintaining a mental map of the entire network. It would be possible to relayout the subnetwork, but this would make it more difficult to follow network activity over time. Leaving gene and edge locations static lowers the cognitive load on the user.

Figure 3 shows a part of the active network in *B. subtilis* at the last time point of a four-point time series. The expression levels range from 0 to 0.004, and the level and ratio thresholds were set to $T_m = 0.0005$, $T_h = 0.002$, and $T_r = 1.5$.

3.6. Pathway Selection

For a given active subnetwork, it is interesting to identify the pathways that contain a large number of active genes. Let N^a and N^i be the total number of active and inactive genes, respectively. For a given pathway P, we count N^a_P and N^i_P, the numbers of active and inactive genes in the pathway, respectively. From these numbers, we compute the probability p that an enrichment of active genes in a pathway can be attributed to chance by Fisher's exact test. Pathways for which $p \leq T_p$, where T_p is a threshold, are considered significant.

The significant pathways are sorted increasingly by p-value, and are shown in a drop-down box of the KEGG Viewer. The user can select a pathway from the list, and the viewer highlights the genes that have been marked active.

© 2008 The Author(s)
Journal compilation © 2008 The Eurographics Association and Blackwell Publishing Ltd.

Other genes are rendered without highlighting. This information can be used to infer new interactions between genes: when a gene from the latter group shows a connection to another active gene in the pathway, it is possible that there exists a regulatory interaction as well. It would be necessary to manually verify this potential link either experimentally or by a literature survey.

3.7. Application Design and Workflow

Data analysis starts by loading a genome file, which contains a plain-text table with tab-separated columns according to the description in subsection 3.1. After loading the data, the application lets the user select the organism from a list presented in a dialog box. The next step is to load an interaction network from a file. The network is either provided by a file in GraphML format, or as a plain-text file containing a table with three columns: source gene locus tag, target gene locus tag, and interaction type. The layout algorithm will be run if necessary. The rendered regulatory network is shown in a separate window. The final step is to load one or more gene expression files. The expression file format is also a plain-text table containing $2N+1$ columns: the first column contains the locus tag, the next N columns contain expression ratios or levels, and the final N columns contain coefficient of variation values or p-values. The application determines whether it loaded expression level or ratio data, performs quantization, and constructs a color map.

Figure 4 shows the main application window after loading the *B. subtilis* genome data, its transcription network, and two four-point time series. One time series contains gene expression ratios, and the other one contains gene expression levels. The top part of the window displays a tab containing basic network information, such as the name of the organism and the total number of nodes and edges. The lower left corner shows the overview display of the entire network; a rectangle indicates the current view in the network window (see Fig. 1). The rectangle in the overview display can be dragged and zoomed, which updates the view in the linked network window. The panel in the lower right corner contains the color legend and some controls for gene expression visual mapping. The tabs allow switching between expression ratio and expression level rendering. The time point selection menu can be used to select a single time point, which then determines the gene box fill color. The sliders below control the data value and reliability ranges, and a standard fill color (not in the colormap) is applied when gene expressions values are beyond the limits set by the sliders.

The top part of the main window also contains a search tab. It contains a panel with a text field to enter a gene name. If the network contains the gene, the network view pans to center that gene in the window. The last tab contains a panel to set the parameters for active network identification.

Finally, there are three windows that provide linked views

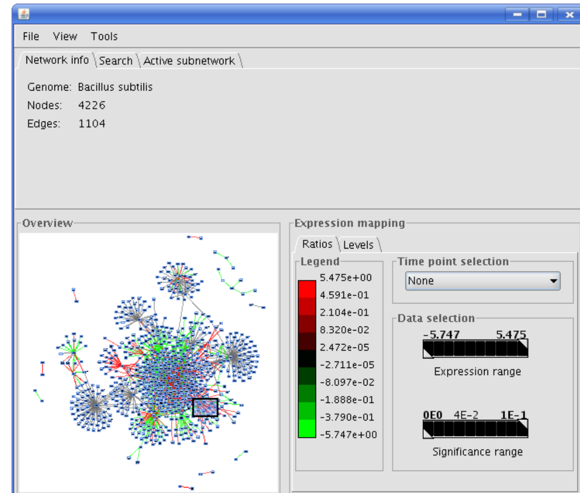

Figure 4: *Main application window after loading B. subtilis genome, its transcription network, a two time series containing gene expression ratios and gene expression levels, respectively. See text for more details of the user interface.*

to the current set of selected genes. This selection is either made in the network window by selecting genes manually or by applying the active subnetwork detection algorithm. The table window provides the raw gene expression data in tabular format, and the plot window displays these data as a line chart. Optionally, gene expression profiles for which each time point has a reliability outside the limits set by the significance slider can be omitted from both the plot and table. Manipulating the slider causes immediate updates in the table window and plot window. The final dependent display is the KEGG Viewer: the set of currently selected genes is checked for overlap with pathways as described in subsection 3.6. All linked view displays are updated automatically when the set of selected genes changes.

4. Case Study

To demonstrate our visualization approach, we make use of a time series dataset from the bacterium *Escherichia coli* [CSC02]. The dataset contains gene expression data from 17 time points, during which *E. coli* is grown on a mixture of glucose and lactose. The bacterium grows preferentially on glucose until that energy source is depleted, resulting in growth arrest while the cells adjust to growth on lactose. This shift, called the diauxic shift, takes places at about time point 6. At time point 14, the stationary phase is entered in which the organism stops growing due to the lack of nutrients. During this phase, many processes are shut down by the bacterial cell in order to save energy. Here, we will focus on the diauxic shift and the changes occurring in stationary phase.

© 2008 The Author(s)
Journal compilation © 2008 The Eurographics Association and Blackwell Publishing Ltd.

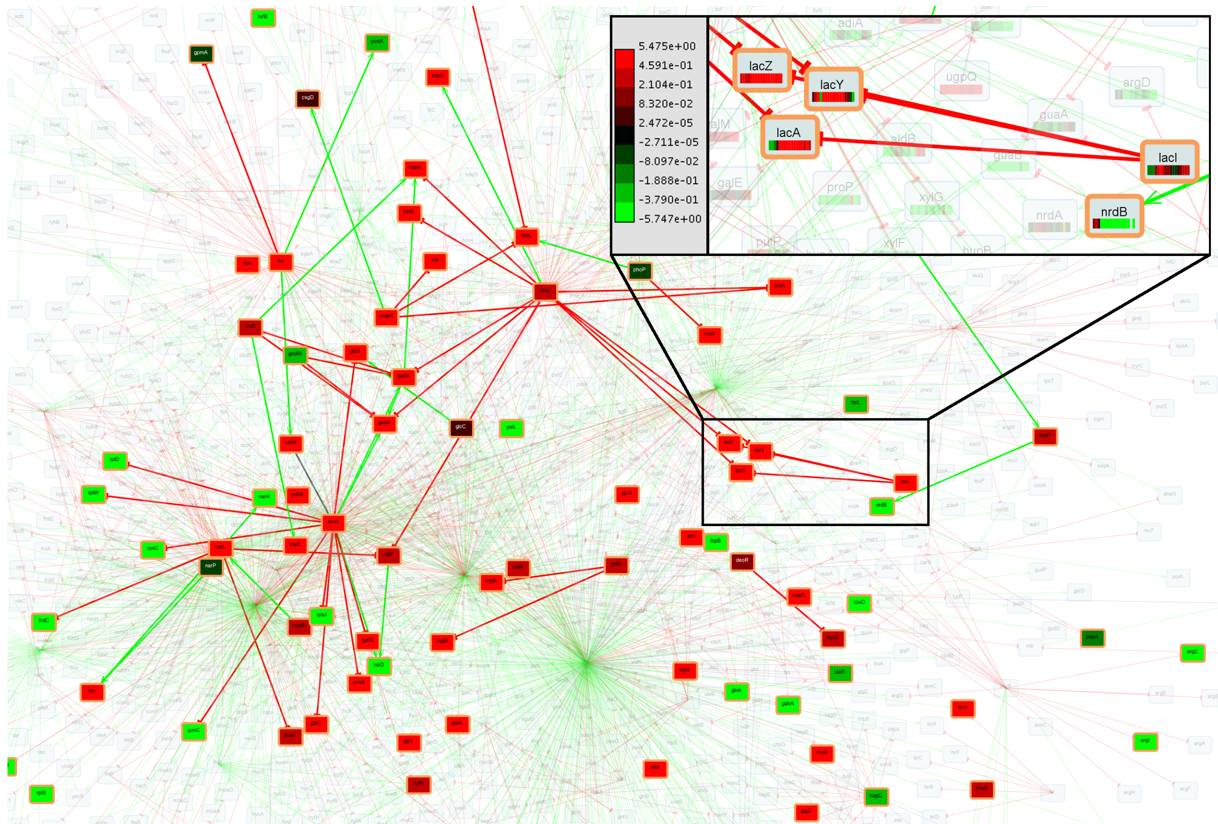

Figure 5: *Active E. coli subnetwork at time point 6; the expression ratios determine the gene box fill colors. The LacI regulon is shown enlarged in the top right corner. There, a standard fill color is used, and the whole time series is shown with the expression glyphs colored by ratio. The genes lacI, lacA, lacY, and lacZ are all upregulated, as is indicated by red-colored expression glyphs, which reveals that the organism adjusts to growing on lactose.*

The traditional approach to analyze such a data set involves many laborious steps, most of which need to be performed manually. Analysis consists of browsing lists of differentially expressed genes, identifying interactions with other genes by database mining, generating lists of affected pathways and accessing these through web pages, clustering by various algorithms resulting in yet more lists of genes, making expression profile plots of several subsets of genes, and so on. In general, subnetworks of affected genes are only determined by manual mining of the various data tables. During the analysis, many different tools are used, and most steps are not automated.

The expression levels for this data set range from 0 to 14.97. We set the medium and high expression level thresholds to $T_m = 5.0$ and $T_h = 10.0$, respectively. The ratio threshold was set to $T_r = 1.0$, and we consider pathways significant at the 10% level, i.e., we set $T_p = 0.1$.

Using our application, a domain expert can quickly identify network activity and corresponding affected pathways from time point to time point. It takes the user only a couple of minutes to reach time point 6, at which the ribosome pathway is disturbed ($p = 3.0 \cdot 10^{-13}$), see Fig. 6. Figure 5 shows the active network at that time point. A ribosome is a protein that is involved in the translation of mRNA into proteins. The image shows that the pathway is almost completely shut down at the diauxic shift phase: the expression glyphs inside the gene boxes change to a green color around time point 6, indicating strong downregulation. This happens because the glucose energy source is depleted (cell growth stops), and the cell needs to adjust to growing on lactose. At the same time, we can see changes in the LacI regulon, which is shown in more detail in the inset in Fig. 5. The genes *lacI*, *lacA*, *lacY*, and *lacZ* are all upregulated, indicated by red-colored expression glyphs. Finding this effect would have been very difficult and time-consuming with traditional analysis methods (see above). It would also be very easy to miss, since it would involve testing explicitly whether this regulon shows activity differences in this time span. In our visualization, the small subnetwork formed by these genes is easily identified.

Near the end of the time series, the bacterium enters the

© 2008 The Author(s)
Journal compilation © 2008 The Eurographics Association and Blackwell Publishing Ltd.

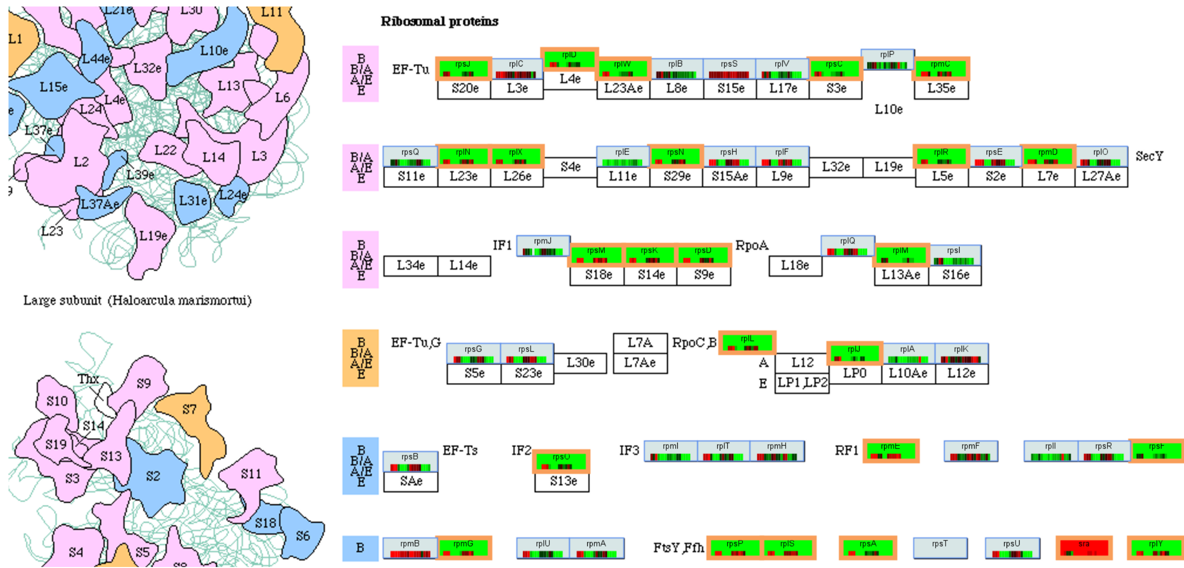

Figure 6: *Many genes active at time point 6 are part of the ribosome pathway. The expression ratio at that time point was used to determine the fill color of a gene box. A green color indicates strong downregulation.*

stationary phase. At time point 16, a much larger part of the network is active, and more pathways are affected (not all shown for space considerations). Since there are no more nutrients, the cell invokes a general stress response, and shuts down many cellular processes to conserve energy. The most affected pathways are ribosome ($p = 9.2 \cdot 10^{-16}$), two-component system ($p = 1.9 \cdot 10^{-7}$), bacterial chemotaxis ($p = 5.5 \cdot 10^{-4}$), and flagellar assembly ($p = 0.02$). The flagellar assembly drives the motility organs of the cell, the flagella, which the bacterium uses to move around. Chemotaxis is the process by which the bacterium moves towards the highest concentration of nutrients. This pathway is shut down at time point 16, see Fig. 7.

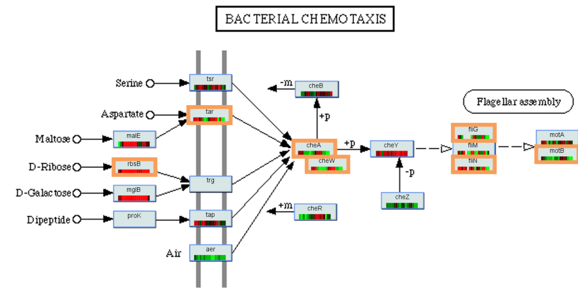

Figure 7: *There is strong downregulation of the genes in involved in the chemotaxis pathway at time point 16.*

5. Conclusions

We have presented an application for visualization of gene expression data from time series experiments in both a gene regulatory network and metabolic pathway context. This integration provides the link between the measurements at the transcriptional level and the phenotype. In addition, we have described a method to identify and visualize active parts of the regulatory network, and an analysis method to identify affected pathways. This approach was used to perform a brief analysis of a time series experiment involving the *E. coli* bacterium. With our application, a biologist can explore the time series by stepping through the time points, and inspecting the subnetworks that are active and the pathways that are affected. This approach both facilitates and speeds up data analysis tremendously in comparison to a traditional approach that involves many manual, laborious, and error-prone steps. The application also facilitates a user to identify regulons within the active network, and to switch from a network view to a list of pathways and a pathway view.

Part of future work includes a more extensive usability study of our approach. We would also like to address dynamic visualization of pathways. An advantage of the static KEGG pathway drawings is that their presentation has been fine-tuned by a curator. However, in many cases a user may wish to have more freedom in drawing a pathway, combining a number of pathway drawings into one picture, and to visualize the actual network of metabolic compounds and associated chemical reactions. We plan to extend our approach by the visualization methods proposed by Klukas & Schreiber [KS07], because they allow combining several pathway diagrams into one visualization. In this way, a user

© 2008 The Author(s)
Journal compilation © 2008 The Eurographics Association and Blackwell Publishing Ltd.

can study how multiple pathways are affected in an experimental condition and also how these pathways are related.

It is also interesting to relate pathway structure to transcription network structure, i.e., to see where the genes involved in a pathway are located in the transcription network and vice versa. To tackle this, we consider the VisLink [CC07] approach.

We have only considered analysis and visualization of one organism at a time. However, when confronted with an uncharacterized gene and uncharacterized function, a biologist would like to explore a network of known interactions in a different species for conserved features to infer gene function and interactions. We plan to address this in future work.

6. Acknowledgement

This work was supported by the BioRange programme of the Netherlands Bioinformatics Centre (NBIC), which is supported by a BSIK grant through the Netherlands Genomics Initiative (NGI).

References

[BBO*06] BOSMAN D. W. J., BLOM E.-J., OGAO P. J., KUIPERS O. P., ROERDINK J. B. T. M.: MOVE: A multi-level ontology-based visualization and exploration framework for genomic networks. *In Silico Biology 7* (2006), 0004.

[BSRG06] BAITALUK M., SEDOVA M., RAY A., GUPTA A.: BiologicalNetworks: Visualization and analysis tool for systems biology. *Nucleic Acids Res. 34* (2006), W466–W471.

[BST03] BREITKREUTZ B.-J., STARK C., TYERS M.: Osprey: A network visualization system. *Genome Biology 4*, 3 (2003), R22.

[CC07] COLLINS C., CARPENDALE S.: VisLink: Revealing relationships amongst visualizations. *IEEE Trans. Visualization and Computer Graphics 13*, 6 (2007), 1192–1199.

[CSC02] CHANG D. E., SMALLEY D. J., CONWAY T.: Gene expression profiling of Escherichia coli growth transitions: an expanded stringent response model. *Molecular Microbiology 45*, 2 (2002), 289–306.

[HNY*07] HU Z., NG D. M., YAMADA T., CHEN C., KAWASHIMA S., MELLOR J., LINGHU B., KANEHISA M., STUART J. M., DELISI C.: VisANT 3.0: new modules for pathway visualization, editing, prediction and construction. *Nucleic Acids Res. 35* (2007), W625–W632.

[JKS06] JUNKER B. H., KLUKAS C., SCHREIBER F.: VANTED: A system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics 7* (2006), 109.

[KS07] KLUKAS C., SCHREIBER F.: Dynamic exploration and editing of KEGG pathway diagrams. *Bioinformatics 23*, 3 (2007), 344–350.

[LBY*04] LUSCOMBE N. M., BABU M. M., YU H., SNYDER M., TELCHMANN S. A., GERSTEIN M.: Genomic analysis of regulatory network dynamics reveals large topological changes. *Letters to Nature 431*, 7006 (2004), 308–312.

[MSH*05] MLECNIK B., SCHEIDELER M., HACKL H., HARTLER J., SANCHEZ-CABO F., TRAJANOSKI Z.: PathwayExplorer: Web service for visualizing high-throughput expression data on biological pathways. *Nucleic Acids Res. 33* (2005), W633–W637.

[OGK*00] OGATA H., GOTO S., K K. S., FUJIBUCHI W., BONO H., KANEHISA M.: KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res. 27*, 1 (2000), 29–34.

[SH07] SUDERMAN M., HALLETT M.: Tools for visually exploring biological networks. *Bioinformatics 23*, 20 (2007), 2651–2659.

[SLN05] SARAIYA P., LEE P., NORTH C.: Visualization of graphs with associated timeseries data. In *Proc. IEEE Symp. Information Visualization (InfoVis'05)* (2005), pp. 225–232.

[SMO*03] SHANNON P., MARKIEL A., OZIER O., BALIGA N. S., WANG J. T., RAMAGE D., AMIN N., SCHWIKOWSKI B., IDEKER T.: Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res. 13*, 11 (2003), 2498–2504.

[SND05] SARAIYA P., NORTH C., DUCA K.: Visualizing biological pathways: Requirements analysis, systems evaluation and research agenda. *Information Visualization 4*, 3 (2005), 191–205.

[Uni07] UNIPROT CONSORTIUM: The universal protein resource (UniProt). *Nucleic Acids Res. 35* (2007), D193–D197.

[WES07] WENIGER M., ENGELMANN J. C., SCHULTZ J.: Genome expression pathway analysis tool – analysis and visualization of microarray gene expression data under genomic, proteomic and metabolic context. *BMC Bioinformatics 8* (2007), 179.

[WGdS00] WOLF D., GRAY C. P., DE SAIZIEU A.: Visualizing gene expression in its metabolic context. *Briefings in Bioinformatics 1*, 3 (2000), 297–304.

[WvHL*07] WESTENBERG M. A., VAN HIJUM S. A. F. T., LULKO A. T., KUIPERS O. P., ROERDINK J. B. T. M.: Interactive visualization of gene regulatory networks with associated gene expression time series data. In *Visualization in Medicine Life Sciences*, Linsen L., Hagen H., Hamann B., (Eds.), Visualization and Mathematics. Springer Verlag, Berlin, Germany, 2007, pp. 293–312.

© 2008 The Author(s)
Journal compilation © 2008 The Eurographics Association and Blackwell Publishing Ltd.