

Visualization of complementary systems biology data with parallel heatmaps

R. M. Podowski
B. Miller
W. W. Wasserman

The interpretation of large-scale biological data can be aided by the use of appropriate visualization tools. Heatmaps—pattern-revealing aggregate views of data—have emerged as a preferred technique for the display of genomics data, since they provide an extra dimension of information in a two-dimensional display. However, an increasing focus on the integration of data from multiple sources has created a need for the display of additional dimensions. To improve the identification of relationships between co-expressed genes identified in microarray experiments, a parallel dataset heatmap viewer has been developed for four-dimensional data display. The flexible data entry structure of the parallel heatmap viewer facilitates the display of both continuous and discrete data. Specific examples are presented for the analysis of diverse functional genomics yeast data related to gene regulation, expression, and annotation. The parallel heatmap viewer enables knowledgeable life science researchers to observe patterns and properties within high-throughput genomics data in order to rapidly identify biologically logical relationships.

Background

Systems biology research generates large, complex datasets. While clustering algorithms can identify subsets of genes that behave similarly, the interpretation of inter-gene relationships can be difficult. In only a small subset of cases can a biological theme be accurately ascribed to a statistical grouping of genes. Interpretation is complicated by the fact that popular clustering algorithms are guaranteed to produce clusters, even if no underlying biological process or statistical motivation exists. Assessment and interpretation of clusters can be simplified when data from multiple sources is correlated. Examples of bioinformatics tools that facilitate such integrative approaches include GoMiner [1], EASE [2], FunSpec [3], and methods based on comparative analysis across species (co-expression networks and interologs) [4–6].

Visualization tools for assessment of correlations offer an alternative approach based on accessing the cumulative knowledge of human specialists—knowledge

that can be difficult to replicate computationally [7]. The scientific community has recognized the benefits of visualization in data exploration and interpretation. Numerous publications, workshops, and conferences addressing these issues can be found through the IEEE Computer Society [8] and ACM [9] Web sites. Improved coordination across multiple views has received recent emphasis [10, 11]. There exist tools designed to explore a single source of information, as exemplified by Cytoscape [12] and Osprey [13] for navigation of molecular interaction data, Pathway Voyager [14] for KEGG [15] metabolic pathway visualization, and the Reactome knowledge base of biochemical pathways and biological processes browser [16].

As the availability of complementary high-throughput data is growing, we sought means to visually discover new relationships within large and complex data. Heatmaps are well established in genomics, provide a means to rapidly identify relationships across large datasets, and conveniently display continuous data through

©Copyright 2006 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

0018-8646/06/\$5.00 © 2006 IBM

color intensity [17–20]. Most gene expression analysis packages—from academic tools such as Cluster and TreeView [17], Hierarchical Clustering Explorer (HCE) [21, 22], GDS Browser [23], and Prism [24] to commercial products such as the LION ArraySCOUT**, Agilent GeneSpring**, and Spotfire DecisionSite** for functional genomics—provide heatmap visualization tools that are linked to clustering algorithms [25]. Layering of complementary data into the visual display, however, has been limited. The above-mentioned tools provide visualization and annotation enhancements to support cluster analysis, including dendrograms, scatterplots, line graphs, detailed row and column descriptions, and links to external annotations. Similarities between the Gene Ontology project [26] annotations assigned to individual genes [21, 27–29] can provide a useful, albeit limited, hint at inter-gene relationships. While multidimensional visualization tools for database exploration have long been established [30, 31], we are not aware of established bioinformatics tools that facilitate the visualization of functional relationships from unrelated sources on a global scale.

Our single most important objective is to enable biologists to compare multiple gene-centric data sources to discover significant functional relationships between genes and characteristics. Examples of such gene-centric data classes include gene expression profiles, binding sites for transcription factors (TFs), Gene Ontology (GO) terms, disease and pathway annotations, literature-based associations, and subcellular localization. A thorough search and examination of existing tools failed to identify a ready solution. Possibilities of enhancements to the existing GeneSpring and Spotfire tools were judged to be neither practical nor affordable. In order to endow researchers with the capacity to seek correlations between such disparate classes of data, we developed the parallel heatmap (PHM) viewer.

Implementation

To support the efficient manipulation of gene expression data, the PHM viewer is embedded within the TG Services GenePilot** microarray data analysis package.¹ In the current version of GenePilot, the PHM viewer is accessed via the hierarchical clustering (HC) results screen. A freeware version, GenePilot–Lite Edition, retains the HC algorithm and the full complement of visualization features, allowing all researchers to conduct rapid comparisons of parallel datasets.

Figure 1 shows the parallel heatmap viewer interface. The left pane contains the complete hierarchically clustered gene expression data and dendrogram for

the elutriation-synchronized yeast cell-cycle time-series dataset [32], with corresponding predictions of TF binding sites generated with the MSCAN software [33] using a collection of yeast-binding profiles [34]. A selected cluster (red dendrogram) is magnified in the right pane. The detailed view contains clear column and row headings and color-coded cell-cycle stages for the expression data. The cell-cycle color tag legend in the top right corner identifies individual M, G1, S, and G2 cell-cycle stages. The top heatmap row in the expanded view represents the average column values for all genes in the selected cluster. Finally, the twelve most commonly shared GO annotations for the selected group of genes are shown between the TF heatmap and the row headings. This view is dynamic and will change for each subselection of genes in the expanded view cluster.

The cluster visualization tools can support display of complete genome-scale data and selected subsets of genes. In the latter case, researchers can magnify a specific cluster for a detailed view with annotations and row and column names. Gene-centric annotation information can be accessed automatically from a number of sources, including Stanford SOURCE [35]. Columns of a selected cluster can be reordered by user-predefined categories (e.g., converting between treatment types and cell types) or clustered on the basis of Pearson Correlation Coefficients or Euclidean distance. Data normalization and color palettes can be dynamically applied to enhance interpretation.

In the PHM viewer, a complementary dataset can be displayed. Any selected cluster can then be examined in detail from two perspectives, and the columns of each set individually reordered. The PHM viewer manages the correct alignment and order of rows (genes) for both panes and between both datasets. To facilitate this, matching identifiers must be available between the annotation fields for each set (e.g., official gene names or accession numbers from a common database). A complete one-to-one match is not required. In the case of multiple rows with identical names in the parallel dataset, only the first instance is displayed. Once relationships are observed, the user has the option to output bitmap image files for archival and communication purposes.

Results and discussion

Datasets

Diverse classes of high-throughput genomics data are available for the single-celled yeast *Saccharomyces cerevisiae*. Results from several of the yeast genomics studies have been prepared for direct import into GenePilot and are available at <http://www.cisreg.ca/raf/PHM/>. These include gene expression datasets [32], Gene Ontology annotations generated from the *Saccharomyces*

¹A complete description of GenePilot features and download instructions may be found on the official GenePilot website (<http://www.genepilot.com/>). A detailed guide to the use of the PHM viewer is available on the project website (<http://www.cisreg.ca/raf/PHM/>).

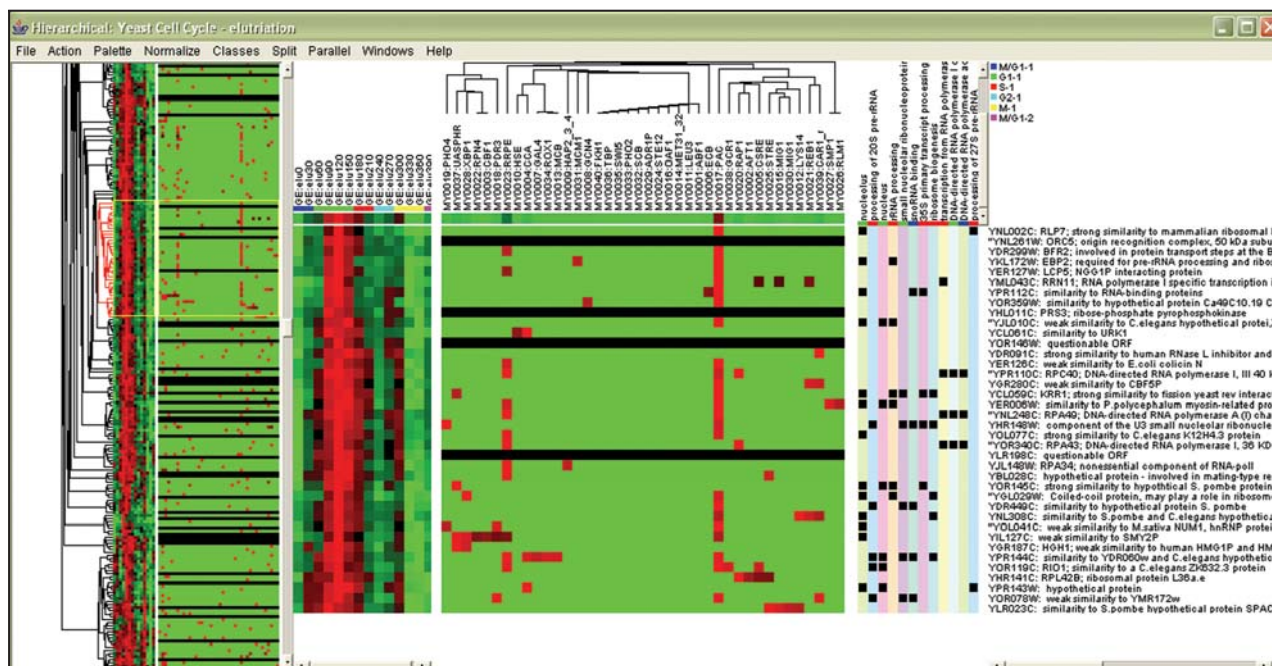


Figure 1

Parallel heat map viewer interface.

Genome Database (SGD**) Gene Ontology Term Finder [36], transcription factor binding site predictions produced by the MSCAN software [33] using a collection of yeast-binding profiles [34], and chromatin immunoprecipitation (ChIP) results generated with microarrays (“ChIP on Chip”) [37]. In all datasets, the rows represent yeast genes, uniquely referenced by the systematic name obtained from the SGD [38]. GenePilot permits the inclusion of multiple column classification and supervision vectors, which makes it possible to tag each column with multiple labels. This enables the user to visualize the information in multiple ways (for instance, grouping columns by array identification numbers). In the sample gene expression data, two vectors are included representing the stages of cell-cycle progression. In addition to the expression data, a file is provided containing the GO terms associated with each yeast gene—a file that may be used with any yeast dataset.

Correlations between in vivo and in silico binding sites

The identification of regulatory elements in the promoters of co-expressed genes can be a challenge. Computational prediction of transcription factor binding sites produces numerous false predictions. Alternatively, high-throughput ChIP methods produce complicated results that may reflect the inclusion of false positives or

instances in which TFs are present at a specific location because of protein–protein interactions rather than direct binding to DNA. Because the two classes of data appear independent and complementary, we sought to determine whether there were relationships that could be visually distinguished in genome-scale data. A database of binding profiles for yeast transcription factors [34] was screened against all yeast gene promoters using the MSCAN algorithm [33] in order to produce a combined probability score for all potential binding sites for each TF in each gene.

Figure 2 shows the correlations of transcription factor binding-site predictions generated by the MSCAN software [33] using a collection of five distinct yeast-binding profiles (*Gal4*, *Gcn4*, *Leu3*, *Reb1*, and *Ste12*) [34], and microarray-assessed chromatin immunoprecipitation results (“ChIP on Chip”) [37]. The figure is a compilation of PHM views of different gene groups with identical column orders, saved as bitmap images. TFs from the ChIP dataset are color-coded and grouped according to the primary functions of the target genes via a column classification vector. The dark gray rows under MSCAN predictions for TF *Gcn4* indicate that no TF binding was predicted for prediction thresholds used in the analysis. Arrows indicating the location of the matching individual TF in both datasets were added to ease interpretation.

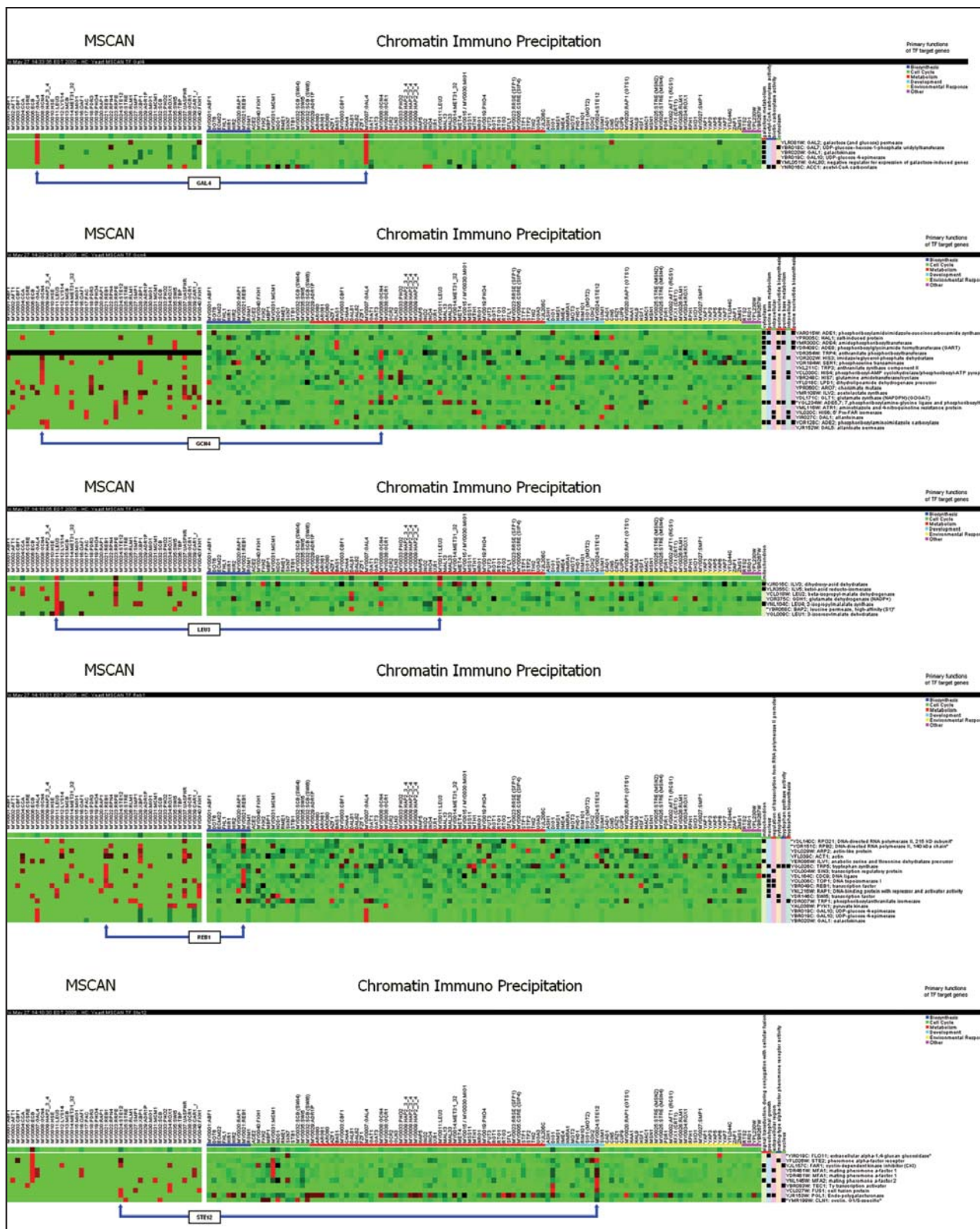


Figure 2

Correlations between *in vivo* and *in silico* binding sites.

For several TFs, there was a clear relationship between the computational predictions and genome-scale ChIP data. For instance, the binding data for both *Gal4* and *Leu3* shows strong agreement between computational predictions and ChIP results. In the cases of *Gcn4* and *Reb1*, a subset of the genes shows good agreement between the two data classes; however, a significant portion is not supported. Visualization does not facilitate uniformly convincing observations, as demonstrated for the set of genes known to be regulated by *Ste12* and where both MSCAN and ChIP data implicate other TFs.

Correlation between TF binding sites and gene expression patterns

Numerous algorithms have emerged to identify over-represented sequence motifs in the promoters of co-expressed genes. While many methods identify the sequence patterns *de novo*, the set of TFs for which binding properties have been defined continues to grow. Therefore, new methods are emerging to identify an overabundance of predicted sites for characterized TFs (e.g., Toucan [39] and oPOSSUM [40]). Visual inspection appears to be a powerful means of assessing these relationships while the statistical methods mature.

Figure 3 shows the correlation between gene expression and MSCAN predictions for a gene set derived from a co-expressed set of yeast genes related to cell-cycle gene expression with an unknown mediating TF, as defined by Getz et al. [41]. Gene expression data was derived from the elutriation-synchronized, yeast cell-cycle gene expression data collection [32] and shows an induction peak at 120 minutes during the G1 stage, followed by a secondary peak at 270 minutes during the G2 stage. It was previously indicated that there is a strong correlation between an over-represented pattern in the regulatory regions and the binding of the TF PAC [34]. Using the PHM viewer, we clustered genes on the basis of the cell-cycle pattern of expression and displayed the computational prediction of binding sites. This view provides strong support for the link of PAC to G1 stage-specific expression in the cell cycle; however, it is clear that not all genes in the expression cluster are linked with PAC binding sites. From the visualization, it appears that additional transcription factors may play a role. Specifically, regulation by RRPE is predicted for a majority of genes showing elevated expression at 120 minutes (G1 stage) and 270 minutes (G2 stage). This visual correlation is not apparent for the genes that are not up-regulated in G2.

Summary

The traditional use of heatmaps for visual verification of gene expression profile relationships and dependencies is just one approach to deriving knowledge from genomic

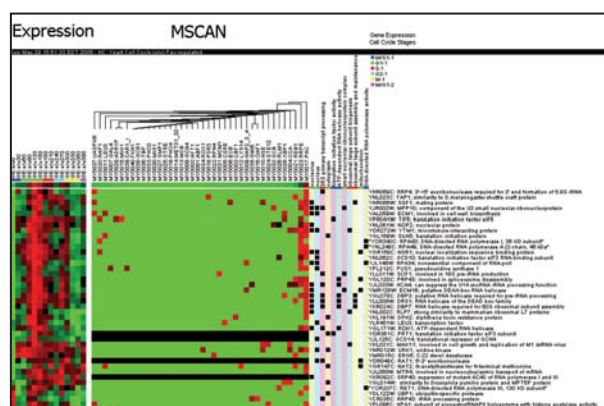


Figure 3

Correlation between gene expression and MSCAN predictions.

data. With the PHM viewer, we have demonstrated that diverse data sources, when examined together, can be used to increase understanding. The examples presented here sampled information from continuous sources, including gene expression time series and experimental and computational protein–DNA binding predictions.

Discrete data sources can be used with equal ease. A discrete example dataset is provided on the PHM project website [42] in the form of GO annotations [36]. Other informative discrete information can be used—for example, disease associations derived from literature, tissue specificity, and metabolic pathway associations. Our experiences with both wet-lab and computational genomics data led us to develop this simple yet powerful tool to readily cross-examine complementary information. It can help build confidence, form a consensus, and provide a basis to question assumptions resting upon highly specialized approaches or sources.

The parallel heatmap viewer enables knowledgeable life science researchers to observe patterns and properties within high-throughput genomics data in order to rapidly identify biologically logical relationships. The viewer is written in Java** and is platform-independent. It is free for academic use, and use of the Lite Edition is unrestricted. For non-academic use and for GenePilot program availability, see [43]. Supplemental figures to this paper can be found at [42].

Acknowledgments

We are grateful to William S. Hayes (AstraZeneca) for identifying the need for a simple, parallel heatmap visualization tool, and also thank Gavin Fischer (OmniViz), Brian Middleton (AstraZeneca), and Gemma Satherwaite (AstraZeneca) for their input. Wyeth W. Wasserman acknowledges financial support from the

Canadian Institutes for Health Research and the Michael Smith Foundation for Health Research.

**Trademark, service mark, or registered trademark of LION bioscience AG, Agilent Technologies, Inc., Spotfire, Inc., TG Services, Inc., The Board of Trustees, Leland Stanford Junior University, and Sun Microsystems, Inc. in the United States, other countries, or both.

References

1. B. R. Zeeberg, W. Feng, G. Wang, M. D. Wang, A. T. Fojo, M. Sunshine, S. Narasimhan, D. W. Kane, W. C. Reinhold, S. Lababidi, K. J. Bussey, J. Riss, J. C. Barrett, and J. N. Weinstein, "GoMiner: A Resource for Biological Interpretation of Genomic and Proteomic Data," *Genome Biol.* **4**, No. 4 (2003); see <http://genomebiology.com/2003/4/4/R28>.
2. D. A. Hosack, G. Dennis, Jr., B. T. Sherman, H. C. Lane, and R. A. Lempicki, "Identifying Biological Themes Within Lists of Genes with EASE," *Genome Biol.* **4**, No. 10 (2003); see <http://genomebiology.com/2003/4/10/R70>.
3. M. D. Robinson, J. Grigull, N. Mohammad, and T. R. Hughes, "FunSpec: A Web-Based Cluster Interpreter for Yeast," *BMC Bioinform.* **3**, No. 1, 35 (2002).
4. L. R. Matthews, P. Vaglio, J. Reboul, H. Ge, B. P. Davis, J. Garrels, S. Vincent, and M. Vidal, "Identification of Potential Interaction Networks Using Sequence-Based Searches for Conserved Protein-Protein Interactions or 'Interologs,'" *Genome Res.* **11**, No. 12, 2120–2126 (2001).
5. J. M. Stuart, E. Segal, D. Koller, and S. K. Kim, "A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules," *Science* **302**, No. 5643, 249–255 (2003).
6. V. van Noort, B. Snel, and M. A. Huynen, "Predicting Gene Function by Conserved Co-Expression," *Trends Genet.* **19**, No. 5, 238–242 (2003).
7. R. D. Meyer and D. Cook, "Visualization of Data," *Current Opin. Biotechnol.* **11**, No. 1, 89–96 (2000).
8. IEEE Computer Society; see <http://www.computer.org/portal/site/ieeeecs/index.jsp>.
9. Association for Computing Machinery; see <http://www.acm.org>.
10. N. Boukhelifa and P. J. Rodgers, "A Model and Software System for Coordinated and Multiple Views in Exploratory Visualization," *Info. Visualiz.* **2**, No. 4, 258–269 (2003).
11. *Proceedings of the 2nd International Conference on Coordinated and Multiple Views in Exploratory Visualization*, London, England (2004); see <http://csdl.computer.org/comp/proceedings/cmv/2004/2179/00/2179toc.htm>.
12. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks," *Genome Res.* **13**, No. 11, 2498–2504 (2003).
13. B.-J. Breitkreutz, C. Stark, and M. Tyers, "Osprey: A Network Visualization System," *Genome Biol.* **4**, No. 3, R22 (2003).
14. E. Altermann and T. R. Klaenhammer, "PathwayVoyager: Pathway Mapping Using the Kyoto Encyclopedia of Genes and Genomes (KEGG) Database," *BMC Genom.* **6**, No. 1, 60 (2005).
15. KEGG: Kyoto Encyclopedia of Genes and Genomes; see <http://www.genome.jp/kegg/>.
16. Reactome, Cold Spring Harbor Laboratory, European Bioinformatics Institute, and GO Consortium; see <http://www.reactome.org>.
17. M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, "Cluster Analysis and Display of Genome-Wide Expression Patterns," *Proc. Natl. Acad. Sci. USA* **95**, No. 25, 14863–14868 (1998).
18. A. P. Gasch, M. Huang, S. Metzner, D. Botstein, S. J. Elledge, and P. O. Brown, "Genomic Expression Responses to DNA-Damaging Agents and the Regulatory Role of the Yeast ATR Homolog Mec1p," *Molec. Biol. Cell* **12**, No. 10, 2987–3003 (2001).
19. J. N. Weinstein, T. G. Myers, P. M. O'Connor, S. H. Friend, A. J. Fornace, Jr., K. W. Kohn, T. Fojo, S. E. Bates, L. V. Rubinstein, N. L. Anderson, J. K. Buolamwini, W. W. van Osdol, A. P. Monks, D. A. Scudiero, E. A. Sausville, D. W. Zaharevitz, B. Bunow, V. N. Viswanadhan, G. S. Johnson, R. E. Wittes, and K. D. Paull, "An Information-Intensive Approach to the Molecular Pharmacology of Cancer," *Science* **275**, No. 5298, 343–349 (1997).
20. X. Wen, S. Fuhrman, G. S. Michaels, D. B. Carr, S. Smith, J. L. Baker, and R. Somogyi, "Large-Scale Temporal Gene Expression Mapping of Central Nervous System Development," *Proc. Natl. Acad. Sci. USA* **95**, No. 1, 334–339 (1998).
21. Hierarchical Clustering Explorer, Human-Computer Interaction Laboratory, University of Maryland; see <http://www.cs.umd.edu/hcil/hce/index.html>.
22. J. Seo and B. Shneiderman, "Interactively Exploring Hierarchical Clustering Results," *IEEE Computer* **35**, No. 7, 80–86 (2002).
23. R. Edgar, M. Domrachev, and A. E. Lash, "Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository," *Nucl. Acids Res.* **30**, No. 1, 207–210 (2002).
24. W. Wu and W. S. Noble, "Genomic Data Visualization on the Web," *Bioinformatics* **20**, No. 11, 1804–1805 (2004).
25. I. M. Dresen, J. Hüsing, E. Kruse, T. Boes, and K.-H. Jöckel, "Software Packages for Quantitative Microarray-Based Gene Expression Analysis," *Current Pharmaceut. Biotechnol.* **4**, 417–437 (2003).
26. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene Ontology: Tool for the Unification of Biology," *Nature Genet.* **25**, 25–29 (2000).
27. S. G. Lee, W. S. Lee, and Y. S. Kim, "GOODIES: GO Based Data Mining Tool for Characteristic Attribute Interpretation on a Group of Biological Entities," *Genome Informat.* **14**, 675–676 (2003).
28. K. Nishimura, K. Abe, S. Ishikawa, S. Tsutsumi, K. Hirota, H. Aburatani, and M. Hirose, "A PCA Based Method of Gene Expression Visual Analysis," *Genome Informat.* **14**, 346–347 (2003).
29. E. Segal, B. Taskar, A. Gasch, N. Friedman, and D. Koller, "Rich Probabilistic Models for Gene Expression," *Bioinformatics* **17**, S 243–S 252 (2001).
30. G. S. Davidson, B. Hendrickson, D. K. Johnson, C. E. Meyers, and B. N. Wylie, "Knowledge Mining With VxInsight: Discovery Through Interaction," *J. Intell. Info. Syst.* **11**, No. 3, 259–285 (1998).
31. D. Keim and H.-P. Kriegel, "VisDB: Database Exploration Using Multidimensional Visualization," *IEEE Computer Graph. & Appl.* **14**, No. 5, 40–49 (1994).
32. P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher, "Comprehensive Identification of Cell Cycle-Regulated Genes of the Yeast *Saccharomyces Cerevisiae* by Microarray Hybridization," *Molec. Biol. Cell* **9**, No. 12, 3273–3297 (1998).
33. Ö. Johansson, W. Alkema, W. W. Wasserman, and J. Lagergren, "Identification of Functional Clusters of Transcription Factor Binding Motifs in Genome Sequences: The MSCAN Algorithm," *Bioinformatics* **19**, Supplement 1, I169–I176 (2003).
34. A. Sandelin, A. Höglund, B. Lenhard, and W. W. Wasserman, "Integrated Analysis of Yeast Regulatory Sequences for Biologically Linked Clusters of Genes," *Funct. & Integrat. Genom.* **3**, No. 3, 125–134 (2003).

35. SOURCE, Genetics Department, Stanford University; see <http://source.stanford.edu/cgi-bin/source/sourceSearch>.
36. GO Term Finder; see <http://db.yeastgenome.org/cgi-bin/GO/goTermFinder>.
37. T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J.-B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young, "Transcriptional Regulatory Networks in *Saccharomyces Cerevisiae*," *Science* **298**, No. 5594, 799–804 (2002).
38. *Saccharomyces* Genome Database; see <http://www.yeastgenome.org/>.
39. S. Aerts, P. Van Loo, Y. Moreau, and B. DeMoor, "A Genetic Algorithm for the Detection of New Cis-Regulatory Modules in Sets of Coregulated Genes," *Bioinformatics* **20**, No. 12, 1974–1976 (2004).
40. oPOSSUM; see <http://cisreg.ca/cgi-bin/oPOSSUM/opossum>.
41. G. Getz, E. Levine, E. Domany, and M. Q. Zhang, "Super-Paramagnetic Clustering of Yeast Gene Expression Profiles," *Physica A* **279**, 457–464 (2000).
42. Parallel HeatMap Viewer Project; see <http://www.cisreg.ca/raf/PHM/>.
43. GenePilot, TG Services, Inc.; see <http://genepilot.com>.

Received October 6, 2005; accepted for publication January 26, 2006; Internet publication September 15, 2006

Raf M. Podowski *SeeqPod, Inc., 6425 Christie Avenue, Emeryville, California 94608 (raf@seeqpod.com)*. Mr. Podowski is currently completing a Ph.D. degree in bioinformatics at the Karolinska Institute while working as a Chief Scientific Officer at SeeqPod, Inc., an exploratory search technology startup, of which he is also a co-founder. He holds a licentiate of philosophy in molecular biology from Uppsala University, Sweden, and an M.S. degree in engineering physics from Rensselaer Polytechnic Institute. Mr. Podowski has worked as a senior product manager for life sciences at Oracle, as a text-mining consultant at AstraZeneca, and as a senior bioinformatics scientist at X-Mine, Inc., where he developed a number of life science applications in text mining, drug discovery informatics, and genomic data analysis.

Brett Miller *GenePilot, Inc., P.O. Box 21225, El Sobrante, California 94820 (brett@genepilot.com)*. Mr. Miller is a technology consultant whose background includes producing award-winning educational software products as Project Manager and Senior Programmer for Sierra On-Line and Platform Engine Architect for Broderbund Software, Inc. He worked in DNA microarray analysis software development as Director of Engineering for X-Mine, Inc. and in media center control system software for Cisco Systems.

Wyeth W. Wasserman *Center for Molecular Medicine and Therapeutics, Department of Medical Genetics, University of British Columbia, 950 West 28th Avenue, Vancouver, British Columbia, Canada V5Z 4H4 (wyeth@cmmt.ubc.ca)*. Dr. Wasserman is an Associate Professor who leads a research group in computational biology. His research addresses the analysis of regulatory sequences in the human genome through phylogenetic footprinting and analysis of clusters of transcription-factor binding sites. After earning college degrees in both mathematics and cell biology, he completed his doctorate in oncology studying gene regulation in cancer prevention. Dr. Wasserman led a research group at the Karolinska Institute and has worked as a computational biologist in the pharmaceutical industry.