Commentary

Short time-series microarray analysis: Methods and challenges

Xuewei Wang[1], Ming Wu[2], Zheng Li[1,4] and Christina Chan*[1,2,3]

Address: [1]Department of Chemical Engineering and Material Science, Michigan State University, East Lansing, MI 48824, USA, [2]Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA, [3]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA and [4]Biomedical Engineering Department, Boston University, Boston, 02215, USA

Email: Xuewei Wang - xwang@egr.msu.edu; Ming Wu - wuming1@msu.edu; Zheng Li - lizheng1@msu.edu; Christina Chan* - krischan@egr.msu.edu

* Corresponding author

Abstract

The detection and analysis of steady-state gene expression has become routine. Time-series microarrays are of growing interest to systems biologists for deciphering the dynamic nature and complex regulation of biosystems. Most temporal microarray data only contain a limited number of time points, giving rise to short-time-series data, which imposes challenges for traditional methods of extracting meaningful information. To obtain useful information from the wealth of short-time series data requires addressing the problems that arise due to limited sampling. Current efforts have shown promise in improving the analysis of short time-series microarray data, although challenges remain. This commentary addresses recent advances in methods for short-time series analysis including simplification-based approaches and the integration of multi-source information. Nevertheless, further studies and development of computational methods are needed to provide practical solutions to fully exploit the potential of this data.

Background

Microarray technology has enabled the interrogation of gene expression data in a global and parallel fashion, and has become the most popular platform in the era of systems biology [1]. A majority of the microarray analysis thus far has focused on elucidating disease mechanisms [2]. More recently, with the rapid growth in research and development of biofuels [3], a new challenge of manipulating plant cell-wall biosynthesis has led to further applications of microarrays [3]. The detection and analysis of steady-state mRNA expression have become routine [4-7], with applications in many areas of biology (i.e., plants, yeast, insects, and mammals). Increasing efforts are focused on deciphering the multidimensional *dynamic* behaviours of complex biological systems, including complex regulation schemes, such as the crosstalk between multiple pathways [3,8,9], and interactions among more than 1000 genes in plant cell wall biogenesis, developmental biology, and human diseases [10-14]. Thus, time-series microarray data, and its analysis, are of growing interest to several research communities [15].

Time-series microarrays capture multiple expression profiles at discrete time points (i.e., minutes, hours, or days) of a continuous cellular process. These data can characterize the complex dynamics and regulation in the form of differential gene-expressions as a function of time. Numerous time-series microarray experiments have been performed to study such biological processes as the biological rhythms or circadian clock of *Arabidopsis*, flowering time, abiotic stress, disease progression, and drug responses [2,16-20]. Many of the methods of analyzing time-series data originated from various disciplines, such as signal processing, dynamic system theory, machine

learning and information theory, and have been applied to detect differentially expressed genes, identify expression patterns, and construct gene networks [15,21-23], nevertheless challenges remain.

A significant challenge in dealing with time-series data comes from the limited sampling or number of time points taken, giving rise to short time-series data. In the growing pool of temporal microarray datasets, a typical time-series record has fewer than ten time-points [24]. The most common type of temporal data available is short time-series data, which arises from the difficulty in obtaining samples for many time points, often times due to the high costs of the arrays or limited biological samples, especially in animal or clinical studies [25,26]. "Short" time-series could signify the time-scale or the number of discrete time-points. Typically, it refers to the latter, which more appropriately should be *sparse* time-series data.

Limited sampling accentuates the difficulties associated with static or standard time-series analyses. First, the problems arising due to high dimensionality accompanied by a small sample size, such as matrix singularity and model over-fitting [27], in analyzing static or long time-series microarray data, become more pronounced in the case of short time-series data. Second, the unavoidable noise has more influence on the analysis of short time-series than on long time-series data, enhancing the difficulty in distinguishing real from random patterns and increasing the potential of misleading analyses [28].

Improving short time-series analysis requires addressing the problems that arise due to limited sampling. Recent efforts by investigators to overcome the difficulties associated with limited sampling include decreasing the complexity of continuous time-series data based on simplification strategies [29,30] or enriching the information content of the data by incorporating multi-source information [31,32], see Figure 1 for a summary of possible options.

Simplification strategies
Simplification strategies reduce time-series data from continuous to discrete representations prior to analysis. These strategies usually transform the raw temporal profiles into a set of symbols [29,30,33] or nominal values [31,34] that are used to categorize qualitatively the gene expression data into different states or trends, that is, in terms of phases (early or late), magnitudes (high or low), or directions (up- or down-regulation). Based on this concept, a "quantization" method introduced by Di Camillo et al[35], whereby the expression of a gene at a particular time-point is quantized (discretized) into three patterns of "states", representing under-expressed, not differentially expressed or over-expressed with respect to a base-

line pre-defined by a hypothetical distribution. After such discretization, the Dynamic Bayesian Network algorithm performed better in terms of precision and recall in reconstructing the regulatory network from synthetic expression data generated from differential equations based on a series of defined rules of regulation. Similarly, Kim [33] developed a difference-based clustering algorithm (DIB-C) in which the profile of short time-series data was discretized to symbolic patterns, but according to the differences between adjacent time-points. These patterns or "trend" simplified the profile of a gene from numerical values to direction of change, that is, "*I* (Increase), *D* (Decrease) or *N* (No change)", and rate of change, that is, "*V* (conVex), *A* (concAve) or *N* (No change)". Inevitably information is lost through this simplification. Even so, such conceptual discretization helped achieve more interpretable and biologically meaningful clusters [33].

Simplifications methods have a side benefit in reducing the noise in the original data to some degree when decreasing the dimension of the time-series data, thus making the subsequent analysis more robust to noise. This was demonstrated by Sacchi et al. [30] with their adaptation of the Temporal Abstractions (TA)-clustering method from the field of artificial intelligence to gene expression analysis. Here, the temporal expression profiles were described in terms of trends of "Increasing", "Decreasing", or "Steady". A reduced rate of misclassification in computational experiments was observed for simulated data using TA-clustering with pre-defined patterns and noise than with the clustering approach without such simplification strategies, particularly when the noise level was high [30].

A key challenge with simplification strategies is how to pre-define these *a priori* representative temporal trends or patterns of gene expression in the discretization step. Defining these patterns have largely depended on the expertise of the researchers, for example, Gerber et al defined six temporal expressions trends in terms of phase (early, middle and late) and direction (increase and decrease) [31], similarly, Wu et al. proposed 27 possible temporal patterns to group gene expression data for CD8 T cell differentiation [34]. However, this may introduce bias in the patterns that are pre-defined and, in turn, the analysis and results obtained. Data-driven approaches could extract potentially novel gene expression patterns in an objective and reasonably unbiased fashion [36]. Thus, developing methods to automatically define temporal trends could alleviate this limitation or bias. Ernst et al. proposed a procedure to generate potential trends which describe the directions and magnitudes of the expression changes with respect to time [24,28]. Attempts at automatic abstraction of temporal features have met with some success in providing easily interpretable clusters,
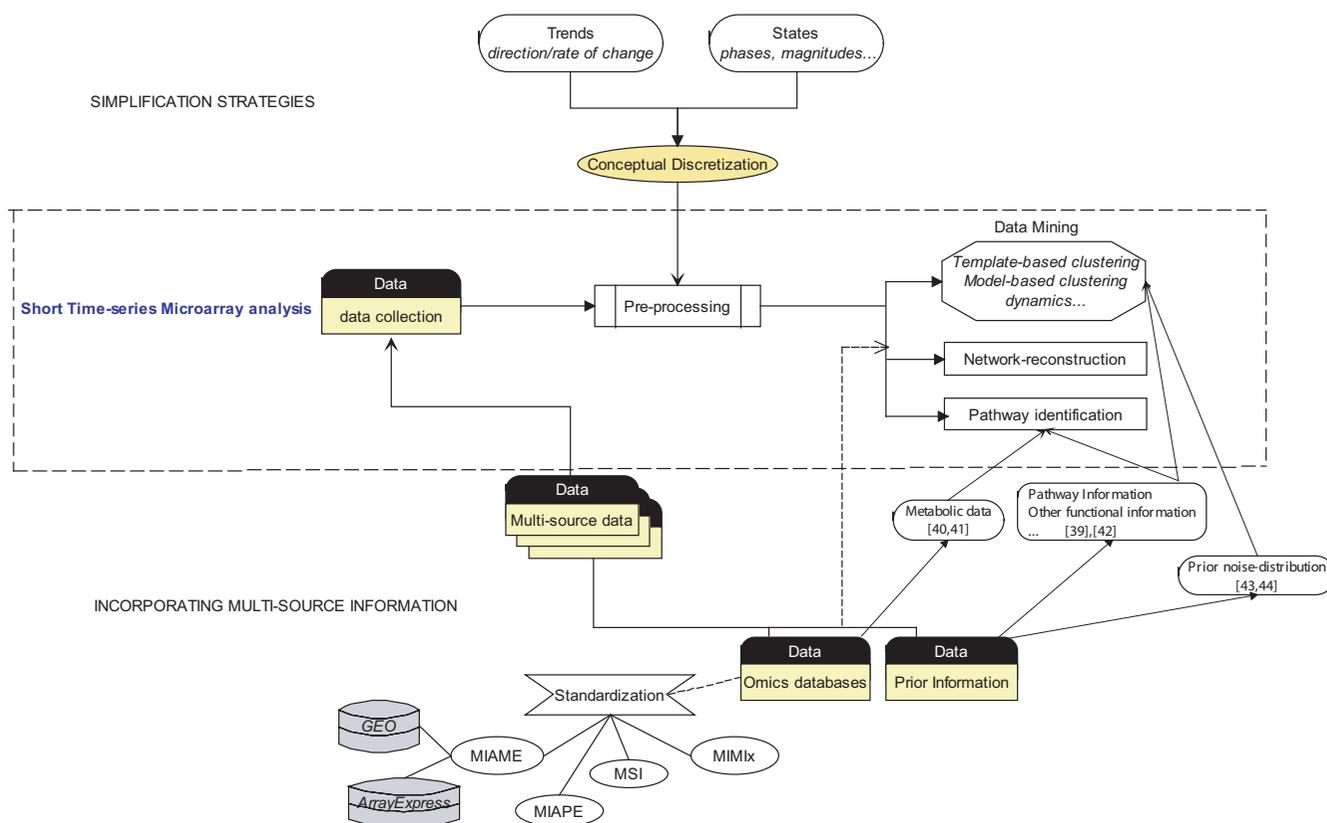
Figure 1
The general process of *time-series expression analysis* starts with *data collection* from microarray experiments.
The data then undergoes *pre-processing* procedures, such as normalization and quality evaluation. Next *data mining* techniques are used to *discover patterns or characteristics*, *identify related pathways* or *reconstruct systems network* for biological processes from short-time series data. To address the limited sampling in short-time series data, two strategies are introduced in the general process of microarray analysis. *Simplification strategies* reduce the data to discrete representations based on *trends* or *states* with respect to time to achieve more interpretable and biologically meaningful clusters. Such *conceptual discretization* is part of the pre-processing step, prior to data mining. *Incorporating multi-source information* takes a different strategy. In this strategy *multi-source data*, including various *omics databases* and *prior biological information*, are collected and integrated to obtain a comprehensive dataset and enhance the information content. To minimize the heterogeneity of *omics data* from different experiments, *standardization* can and have been imposed on *omics databases*. Current standards for high-through-put database include *MIAME, MIAPE, MSI, MIMIx*. MIAME has been implemented with *GEO* and *ArrayExpress* microarray databases. The integration of various *omics databases* or *prior biological information* can enhance the effectiveness and efficiency of mining and interpretation of short-time series data to achieve biological discoveries. For example, multi-source *prior biological information*, i.e., *prior noise-distribution* has been proposed to enhance the performance of the data mining and network inference [43,44]. In addition, *pathway and functional knowledge* and *metabolic data* from different databases have also enhanced the clustering results and pathway identification [39-42]. These studies are discussed and referenced in the text.

examples include the temporal abstraction-based method that defines trends (i.e., *Increasing, Decreasing and Steady*) over subintervals [30], and the difference-based method that uses the first and second order differences in expression values to detect the direction and rate of change of the temporal expression [33]. Although simplification strategies make the raw expression profiles coarse-grained, which could somewhat ameliorate the noise in the data, inevitably the simplification leads to loss of information, which may exacerbate the situation of limited sampling.

In particular, some important patterns may be lost when the raw expression profiles are oversimplified, for example, simplifications that consider only monotonously expressing genes [31] may not capture some of the complex temporal patterns, such as oscillatory gene expression profiles [37].

Incorporating multi-source information
Incorporating multi-source information, including prior knowledge (i.e., pathway information) [38,39], multi-

scale or different levels of information [40-42], or additional time-series datasets from other sources [31,32], is another approach to address the limited sampling and to improve the computational analysis and interpretation of short time-series microarray data.

Different types of prior knowledge have been used to improve the computational analysis of short time-series data. They include applying a prior noise distribution to the expression data [43]. For example, by incorporating a prior noise-distribution to improve the parameter estimation in the commonly used CAGED model (Cluster Analysis of Gene Expression Dynamic), Wang et al. achieved more functional and meaningful clusters, as validated by Gene Ontology [43]. This approach was advanced further by Wang et al. [44] to a stochastic dynamic model where the gene expression profile is modelled with the addition of noisy "measurements". The authors try to explicitly separate the real pattern of expression from the Gaussian noise imposed onto the expression data. Based on such a model, they applied Expectation Maximization (EM) algorithm to estimate both the parameters for the noise model and the actual values of the expression levels, and efficiently reconstructed the gene regulatory network. Thus defining a prior noise-distribution in analyzing time series microarrays is both biologically relevant and computationally efficacious especially when the time series is too short to satisfy the requirements of traditional multivariate methods for parameter estimation [44].

In addition, pre-defined gene sets involving specific pathways or functional categories have focused on pattern changes of sets of genes rather than individual genes and helped to enhance our understanding of cellular processes [38,39]. Similarly, incorporating multi-level biological information, such as metabolic data or prior knowledge about the genes and pathways, has improved interpretation of the data. For example, metabolic data [40,41] and pathway information [40,42] have been integrated with short time-series gene expression data to identify liver toxicity pathways in HepG2 cells. Likewise, protein-DNA interaction data and promoter motif information have been integrated with short time series data to reconstruct the dynamic gene regulatory network of *Saccharomyces cerevisiae* response to stress [45], and to identify targets of known transcription factors in cold acclimation of *Arabidopsis thaliana* [46], respectively. Furthermore, metabolic profiles have been integrated with short time-series gene expression data to characterize the dynamics of metabolic changes during oxidative stress [47], the effect of elevated CO_2 on the physiology of *A. thaliana* [48], and to reconstruct the temporal sequence of events during bud development [49]. Similarly, integrating multiple time-series datasets has become increasingly popular with the growing pool of publicly available datasets [50]. Combining

multiple time-series datasets has been shown to improve the confidence of the gene regulatory relationships that are inferred [51], as well as identify regulatory relationships [32] and functional gene clusters [31] under different treatment conditions.

A key challenge with integrating different datasets is the heterogeneity of the data, that is, each set may have a unique set of sampling rates, time-scales, cell types, and sample populations, as well as varying measurement noise levels, etc. The heterogeneity across the datasets increases the difficulty in extracting meaningful results. To maximize the usefulness and minimize the heterogeneity of the publicly available data, stricter standardization methods should be defined and imposed on procedures such as data collection and pre-processing. Indeed, standards such as MIAME (Minimum information about a microarray experiment), MIAPE (Minimum information about a preoteomics experiment), MSI (Metabolomics standards initiative), MIMIx (Minimum information required for reporting a molecular interaction experiment) have been proposed and implemented for presenting and exchanging gene expression [52], proteomics [53], metabolomics [54] and interaction data [55], respectively. Thus far, standardizing gene expression data is the most mature and hence, most successful compared to the standardization of the other data types. Therefore, integrating gene expression data from various sources is now readily achievable with public databases, such as GEO [56] and ArrayExpress [57], where the quality of the data is controlled with the MIAME score.

Conclusion

In summary, analysis of short time-series microarrays is still at an early stage. Most studies using short time-series data have applied methods that had been developed for static or long time-series microarray data, and which tend to perform poorly with limited temporal sampling. Current efforts, including simplification approaches and the integration of multi-source information, have shed promising light on improving the analysis of short time-series microarray data.

Future studies could combine both of these strategies to simultaneously decrease the complexity of continuous time-series representations, yet minimize the information loss with the simplification-based approaches by increasing the information content of the data. Gene-module-level analysis could be a potential solution, in which the concept of modularity not only plays a central role in incorporating multi-source biological information, but also reflect a simplification strategy focusing on groups of genes rather than individual ones. Gene-module-level analysis could efficiently combine both strategies.

A recent study by Hirose et al [58] used a statistical inference method to reconstruct a module-level gene network based on time-series data, rather than networks of individual genes. They concentrated on groups of genes and the correlations between them, thus the transcription modules extracted could be building blocks of the regulatory networks. Such module-based network construction overcomes, in part, the problem of limited sampling. The modules in the study are calculated by a vector regressive approach based on the state space model, which essentially simplifies the data by including only the significant temporal relationships between the modules. Unfortunately, their modules are defined based on statistical criteria and thus are limited in their biological significance. The integration of multi-source biological information to identify modules from short-time series microarray data should enhance understanding and interpretation of biological systems and disease processes.

Thus far, the predominant focus has still been on lower levels of analyses, such as detecting differently expressed genes or clustering genes with similar temporal profiles, whereas few higher levels of analysis, i.e. network construction, have been reported. With the rapid growth in availability of short time-series data, more theoretical and technical studies are urgently needed to provide practical solutions to exploit fully the potential of this wealth of data.

Acknowledgements

References

1. Panda S, Sato TK, Hampton GM, Hogenesch JB: **An array of insights: application of DNA chip technology in the study of cell biology.** *Trends in cell biology* 2003, **13(3)**:151-156.
2. Cobb JP, Mindrinos MN, Miller-Graziano C, Calvano SE, Baker HV, Xiao W, Laudanski K, Brownstein BH, Elson CM, Hayden DL, Herndon DN, Lowry SF, Maier RV, Schoenfeld DA, Moldawer LL, Davis RW, Tompkins RG, Baker HV, Bankey P, Billiar T, Brownstein BH, Calvano SE, Camp D, Chaudry I, Cobb JP, Davis RW, Elson CM, Freeman B, Gamelli R, Gibran N, Harbrecht B, Hayden DL, Heagy W, Heimbach D, Herndon DN, Horton J, Hunt J, Laudanski K, Lederer J, Lowry SF, Maier RV, Mannick J, McKinley B, Miller-Graziano C, Mindrinos MN, Minei J, Moldawer LL, Moore E, Moore F, Munford R, Nathens A, O'Keefe G, Purdue G, Rahme L, Remick D, Sailors M, Schoenfeld DA, Shapiro M, Silver G, Smith R, Stephanopoulos G, Stormo G, Tompkins RG, Toner M, Warren S, West M, Wolfe S, Xiao W, Young V: **Application of genome-wide expression analysis to human health and disease.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102(13)**:4801-4806.
3. US Department of Energy, Office of Sciences: **Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda.** 2006.
4. Salunkhe P, Topfer T, Buer J, Tummler B: **Genome-wide transcriptional profiling of the steady-state response of *Pseudomonas aeruginosa* to hydrogen peroxide.** *Journal of bacteriology* 2005, **187(8)**:2565-2572.
5. Rosso D, Ivanov AG, Fu A, Geisler-Lee J, Hendrickson L, Geisler M, Stewart G, Krol M, Hurry V, Rodermel SR, Maxwell DP, Huner NP: **IMMUTANS does not act as a stress-induced safety valve in the protection of the photosynthetic apparatus of Arabidopsis during steady-state photosynthesis.** *Plant physiology* 2006, **142(2)**:574-585.
6. Rawool SB, Venkatesh KV: **Steady state approach to model gene regulatory networks--simulation of microarray experiments.** *Bio Systems* 2007, **90(3)**:636-655.
7. Kocabas AM, Crosby J, Ross PJ, Otu HH, Beyhan Z, Can H, Tam WL, Rosa GJ, Halgren RG, Lim B, Fernandez E, Cibelli JB: **The transcriptome of human oocytes.** *Proc Natl Acad Sci U S A* 2006, **103(38)**:14027-14032.
8. Laule O, Fürholz A, Chang HS, Zhu T, Wang X, Heifetz PB, Gruissem W, Lange M: **Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*.** *Proc Natl Acad Sci U S A* 2003, **100(11)**:6866-6871.
9. Setlur SR, Royce TE, Sboner A, Mosquera JM, Demichelis F, Hofer MD, Mertz KD, Gerstein M, Rubin MA: **Integrative Microarray analysis of pathways dysregulated in metastatic prostate cancer.** *Cancer Res* 2007, **67(21)**:10296-10303.
10. Yong WD, Link B, O'Malley R, Tewari J, Hunter CT, Lu CA, Li XM, Bleecker AB, Koch KE, McCann MC, McCarty DR, Patterson SE, Reiter WD, Staiger C, Thomas SR, Vermerris W, Carpita NC: **Genomics of plant cell wall biogenesis.** *Planta* 2005, **221(6)**:747-751.
11. Carpita N, Tierney M, Campbell M: **Molecular biology of the plant cell wall: searching for the genes that define structure, architecture and dynamics.** *Plant Mol Biol* 2001, **47(1-2)**:1-5.
12. Dozmorov MG, Kyker KD, Saban R, Shankar N, Baghdayan AS, Centola MB, Hurst RE: **Systems biology approach for mapping the response of human urothelial cells to infection by *Enterococcus faecalis*.** *BMC bioinformatics* 2007, **8 Suppl 7**:S2.
13. Hooper SD, Boue S, Krause R, Jensen LJ, Mason CE, Ghanim M, White KP, Furlong EE, Bork P: **Identification of tightly regulated groups of genes during *Drosophila melanogaster* embryogenesis.** *Mol Syst Biol* 2007, **3**:72.
14. Baugh LR, Hill AA, Slonim DK, Brown EL, Hunter CP: **Composition and dynamics of the *Caenorhabditis elegans* early embryonic transcriptome.** *Development (Cambridge, England)* 2003, **130(5)**:889-900.
15. Androulakis IP, Yang E, Almon RR: **Analysis of time-series gene expression data: Methods, challenges, and opportunities.** *Annual Review of Biomedical Engineering* 2007, **9**:205-228.
16. Hsu KL, Pilobello KT, Mahal LK: **Analyzing the dynamic bacterial glycome with a lectin microarray approach.** *Nature chemical biology* 2006, **2(3)**:153-157.
17. McAdams HH, Shapiro L: **A bacterial cell-cycle regulatory network operating in time and space.** *Science* 2003, **301(5641)**:1874-1877.
18. Lan H, Carson R, Provart NJ, Bonner AJ: **Combining classifiers to predict gene function in *Arabidopsis thaliana* using large-scale gene expression measurements.** *BMC bioinformatics* 2007, **8**:358.
19. Welch SM, Roe JL, Dong ZS: **A genetic neural network model of flowering time control in *Arabidopsis thaliana*.** *Agron J* 2003, **95(1)**:71-81.
20. Locke JC, Millar AJ, Turner MS: **Modelling genetic networks with noisy and varied experimental data: the circadian clock in *Arabidopsis thaliana*.** *Journal of theoretical biology* 2005, **234(3)**:383-393.
21. Bar-Joseph Z: **Analyzing time series gene expression data.** *Bioinformatics (Oxford, England)* 2004, **20(16)**:2493-2503.
22. Opgen-Rhein R, Strimmer K: **Learning causal networks from systems biology time course data: an effective model selection procedure for the vector autoregressive process.** *BMC bioinformatics* 2007, **8 Suppl 2**:S3.
23. Opgen-Rhein R, Strimmer K: **From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data.** *Bmc Syst Biol* 2007, **1**:37.
24. Ernst J, Bar-Joseph Z: **STEM: a tool for the analysis of short time series gene expression data.** *BMC bioinformatics* 2006, **7**:191.
25. Ding M, Cui SY, Li CJ, Jothy S, Haase V, Steer BM, Marsden PA, Pippin J, Shankland S, Rastaldi MP, Cohen CD, Kretzler M, Quaggin SE: **Loss of the tumor suppressor Vhlh leads to upregulation of Cxcr4**

and rapidly progressive glomerulonephritis in mice. *Nat Med* 2006, **12(9)**:1081-1087.

26. Karpuj MV, Becher MW, Springer JE, Chabas D, Youssef S, Pedotti R, Mitchell D, Steinman L: **Prolonged survival and decreased abnormal movements in transgenic model of Huntington disease, with administration of the transglutaminase inhibitor cystamine.** *Nat Med* 2002, **8(2)**:143-149.

27. Braga-Neto U: **Fads and fallacies in the name of small-sample microarray classification.** *Ieee Signal Proc Mag* 2007, **24(1)**:91-99.

28. Ernst J, Nau GJ, Bar-Joseph Z: **Clustering short time series gene expression data.** *Bioinformatics (Oxford, England)* 2005, **21**:I159-II68.

29. Yang E, Maguire T, Yarmush ML, Berthiaume F, Androulakis IP: **Bioinformatics analysis of the early inflammatory response in a rat thermal injury model.** *BMC bioinformatics* 2007, **8**:10.

30. Sacchi L, Bellazzi R, Larizza C, Magni P, Curk T, Petrovic U, Zupan B: **TA-clustering: Cluster analysis of gene expression profiles through Temporal Abstractions.** *Int J Med Inform* 2005, **74(7-8)**:505-517.

31. Gerber GK, Dowell RD, Jaakkola TS, Gifford DK: **Automated discovery of functional generality of human gene expression programs.** *PLoS Comput Biol* 2007, **3(8)**:e148.

32. Redestig H, Weicht D, Selbig J, Hannah MA: **Transcription factor target prediction using multiple short expression time series from *Arabidopsis thaliana*.** *BMC bioinformatics* 2007, **8(1)**:454.

33. Kim J, Kim JH: **Difference-based clustering of short time-course microarray data with replicates.** *BMC bioinformatics* 2007, **8**:253.

34. Wu H, Yuan M, Kaech S, Halloran M: **A Statistical Analysis of Memory CD8 T Cell Differentiation: An Application of a Hierarchical State Space Model to a Short Time Course Microarray Experiment.** *Annals of Applied Statistics* 2007, **1(2)**:442-458.

35. Di Camillo B, Sanchez-Cabo F, Toffolo G, Nair SK, Trajanoski Z, Cobelli C: **A quantization method based on threshold optimization for microarray short time series.** *Bmc Bioinformatics* 2005, **6**:.

36. Breitling R: **Biological microarray interpretation: the rules of engagement.** *Biochimica et biophysica acta* 2006, **1759(7)**:319-327.

37. Dequeant ML, Glynn E, Gaudenz K, Wahl M, Chen J, Mushegian A, Pourquie O: **A complex oscillating network of signaling genes underlies the mouse segmentation clock.** *Science* 2006, **314(5805)**:1595-1598.

38. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102(43)**:15545-15550.

39. Segal E, Friedman N, Koller D, Regev A: **A module map showing conditional activity of expression modules in cancer.** *Nat Genet* 2004, **36(10)**:1090-1098.

40. Li Z, Srivastava S, Yang X, Mittal S, Norton P, Resau J, Haab B, Chan C: **A hierarchical approach employing metabolic and gene expression profiles to identify the pathways that confer cytotoxicity in HepG2 cells.** *Bmc Syst Biol* 2007, **1**:21.

41. Srivastava S, Li Z, Yang X, Yedwabnick M, Shaw S, Chan C: **Identification of genes that regulate multiple cellular processes/responses in the context of lipotoxicity to hepatoma cells.** *Bmc Genomics* 2007, **8**:364.

42. Li Z, Srivastava S, Findlan R, Chan C: **Using Dynamic Gene Module Map Analysis To Identify Targets That Modulate Free Fatty Acid Induced Cytotoxicity.** *Biotechnology Progress* 2008, **24(1)**:29-37.

43. Wang L, Ramoni M, Sebastiani P: **Clustering short gene expression profiles.** *Lect Notes Comput Sc* 2006, **3909**:60-68.

44. Wang Z, Yang F, Ho DW, Swift S, Tucker A, Liu X: **Stochastic dynamic modeling of short gene expression time-series data.** *IEEE transactions on nanobioscience* 2008, **7(1)**:44-55.

45. Ernst J, Vainas O, Harbison CT, Simon I, Bar-Joseph Z: **Reconstructing dynamic regulatory maps.** *Mol Syst Biol* 2007, **3**:74.

46. Chawade A, Brautigam M, Lindlof A, Olsson O, Olsson B: **Putative cold acclimation pathways in *Arabidopsis thaliana* identified by a combined analysis of mRNA co-expression patterns, promoter motifs and transcription factors.** *Bmc Genomics* 2007, **8**:304.

47. Baxter CJ, Redestig H, Schauer N, Repsilber D, Patil KR, Nielsen J, Selbig J, Liu J, Fernie AR, Sweetlove LJ: **The metabolic response of heterotrophic Arabidopsis cells to oxidative stress.** *Plant physiology* 2007, **143(1)**:312-325.

48. H. Kanani, B. Dutta, J. Quackenbush, Klapa MI: **Time-Series Integrated Metabolomic and Transcriptional Profiling Analyses .** In *Concepts in Plant Metabolomics* Edited by: Basil J. Nikolau, Wurtele ES. Springer Netherlands; 2007:93-110.

49. Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm J, Bhalerao RP, Boerjan W, Rohde A: **A molecular timetable for apical bud formation and dormancy induction in poplar.** *The Plant cell* 2007, **19(8)**:2370-2390.

50. Ng A, Bursteinas B, Gao QO, Mollison E, Zvelebil M: **Resources for integrative systems biology: from data through databases to networks and dynamic system models.** *Brief Bioinform* 2006, **7(4)**:318-330.

51. Shi Y, Mitchell T, Bar-Joseph Z: **Inferring pairwise regulatory relationships from multiple time series datasets.** *Bioinformatics (Oxford, England)* 2007, **23(6)**:755-763.

52. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FCP, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M: **Minimum information about a microarray experiment (MIAME) - toward standards for microarray data.** *Nat Genet* 2001, **29(4)**:365-371.

53. Taylor CF, Paton NW, Lilley KS, Binz PA, Julian RK, Jones AR, Zhu WM, Apweiler R, Aebersold R, Deutsch EW, Dunn MJ, Heck AJR, Leitner A, Macht M, Mann M, Martens L, Neubert TA, Patterson SD, Ping PP, Seymour SL, Souda P, Tsugita A, Vandekerckhove J, Vondriska TM, Whitelegge JP, Wilkins MR, Xenarios I, Yates JR, Hermjakob H: **The minimum information about a proteomics experiment (MIAPE).** *Nat Biotechnol* 2007, **25(8)**:887-893.

54. Fiehn O, Robertson D, Griffin J, van der Werf M, Nikolau B, Morrison N, Sumner LW, Goodacre R, Hardy NW, Taylor C, Fostel J, Kristal B, Kaddurah-Daouk R, Mendes P, van Ommen B, Lindon JC, Sansone SA: **The metabolomics standards initiative (MSI).** *Metabolomics* 2007, **3(3)**:175-178.

55. Orchard S, Salwinski L, Kerrien S, Montecchi-Palazzi L, Oesterheld M, Stumpflen V, Ceol A, Chatr-Aryamontri A, Armstrong J, Woollard P, Salama JJ, Moore S, Wojcik J, Bader GD, Vidal M, Cusick ME, Gerstein M, Gavin AC, Superti-Furga G, Greenblatt J, Bader J, Uetz P, Tyers M, Legrain P, Fields S, Mulder N, Gilson M, Niepmann M, Burgoon L, De Las Rivas J, Prieto C, Perreau VM, Hogue C, Mewes HW, Apweiler R, Xenarios I, Eisenberg D, Cesareni G, Hermjakob H: **The minimum information required for reporting a molecular interaction experiment (MIMIx).** *Nat Biotechnol* 2007, **25(8)**:894-898.

56. **Gene Expression Omnibus** [http://www.ncbi.nlm.nih.gov/geo/]

57. **ArrayExpress** [http://www.ebi.ac.uk/microarray-as/ae/]

58. Hirose O, Yoshida R, Imoto S, Yamaguchi R, Higuchi T, Charnock-Jones DS, Print C, Miyano S: **Statistical inference of transcriptional module-based gene networks from time course gene expression profiles by using state space models.** *Bioinformatics* 2008, **24(7)**:932-942.