Inferring Gene Regulatory Networks from Asynchronous Microarray Data

David Oviatt[1*], Mark Clement[1], Quinn Snell[1], Kenneth Sundberg[1], Jared Allen[2], Randall Roper[2]
[1]Department of Computer Science, Brigham Young University, Provo, UT, USA
[2]Department of Biology, Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA
*Corresponding author

***Abstract*—Modern approaches to treating genetic disorders, cancers and even epidemics rely on a detailed understanding of the underlying gene signaling network. Previous work has used time series microarray data to infer gene signaling networks given a large number of accurate time series samples. Microarray data available for many biological experiments is limited to a small number of arrays with little or no time series guarantees. Asynchronous Inference of Regulatory Networks (AIRnet) provides gene signaling network inferrence using more practical assumptions about the microarray data. By learning correlation patterns from all pairs of microarray samples, accurate network reconstructions can be performed with data that is normally available in microarray experiments.**

Keywords: gene network infer microarray

I. Introduction

Sequencing the human genome is one of the great accomplishments in recent history. The knowledge gained through sequencing the human genome is vast and holds great implications for medical practice[2]. No single gene, however, decides how an organism grows and matures. Genes form regulatory networks where many genes interact to produce an observable phenotype[3], [4]. An understanding of gene regulatory networks is the key that will open the door to major breakthroughs in fields as diverse as agriculture[5], [6], [7] and medicine[8], [9], [10], [11], [12].

Gene regulatory networks are complex. Many factors can influence each gene's expression at any moment. One or more proteins produced by other genes within the regulatory network can promote or inhibit the expression of a particular gene. Asynchronous Inference of Regulatory Networks (AIRnet) unravels the complexity of these regulatory networks using unsynchronized microarray data that is generally available to researchers to create a network based on the correlation of gene expression changes between microarray experiements. One of the key functions of AIRnet is to compare two of these networks, and easily highlight the differences.

AIRnet has produced promising results, inferring realistic *in-silico* regulatory networks.

II. Related Works

There are many different strategies that have been formulated to deduce gene regulatory networks from microarray data. In a paper written by Wang *et al.*[13], a strategy is proposed that uses multiple microarray samples from different experiments to find a gene regulatory network. Each of these data sets represents a unique experiment. Each experiment is assumed to represent a unique perturbation to the gene regulatory network. Gene regulatory networks are also assumed to be sparse. Differential equations are used to derive a general solution that is the best representation of the invariant parts of the different microarray data sets. Their results show that they are successful in reconstructing small networks. Our algorithm, unlike Wang *et al.*, does not utilize differential equations to form a model of the regulatory network, but employs a much simpler method that can be extended to whole genome studies.

Another popular strategy developed by Liao *et al.*[14], called network component analysis, makes assumptions about power law relationships between genes and the factors that influence their expression. They explain that microarray data is frequently given as a log ratio, thus being pseudo-linear. Then, based on these premises, a regulatory network can be written as $E = A * P$. Where E is the microarray data, A represents prior information about the network, and P represents samples of regulatory signals. When there is no noise associated with this relationship, there is a unique analytic solution that can be found. In real applications, there is noise, and through the use of simulated and real data they are able to reconstruct gene relationships with acceptable accuracy. Their results depend largely on the amount of noise present. A shortcoming of this approach is that prior information about a network needs to be known and expressed in matrix form. Another problem is that there are very stringent constraints on the characteristics of matrix A. A has to have full column and row rank, and if any connections are removed, A still has to have full column and row rank. These restrictions make this method cumbersome to use and limits the datasets that it can be applied to. AIRnet's algorithm can be applied to any set of microarray

Table I: Discretization of data by k-means clustering for an *in-silico* network consisting of 10 genes [1] - genes are divided by row, samples are divided by column.

(a) *Pre-discretized data for an* in-silico *regulatory network consisting of 10 genes*

	wt	G1(-/-)	G2(-/-)	G3(-/-)	G4(-/-)	G5(-/-)	G6(-/-)	G7(-/-)	G8(-/-)	G9(-/-)	G10(-/-)
G1	0.105	0.034	0.927	0.088	0.015	0.049	0.102	0.105	0.018	0.124	0.171
G2	0.877	0.804	0.000	0.864	0.870	0.981	0.837	0.873	0.797	0.860	0.890
G3	0.054	0.000	0.838	0.000	0.103	0.000	0.069	0.000	0.085	0.026	0.048
G4	0.386	0.310	0.611	0.243	0.083	0.432	0.440	0.394	0.364	0.531	0.358
G5	0.801	0.808	0.748	0.903	0.793	0.000	0.880	0.741	0.686	0.321	0.802
G6	0.118	0.051	0.463	0.006	0.167	0.046	0.058	0.082	0.103	0.149	0.133
G7	0.339	0.359	0.116	0.476	0.423	0.381	0.384	0.012	0.728	0.396	0.779
G8	0.894	0.893	0.965	0.874	0.920	0.805	0.968	0.904	0.002	0.873	0.904
G9	0.870	0.885	0.859	0.787	0.825	0.933	0.862	0.888	0.819	0.016	0.792
G10	0.898	0.908	0.903	0.882	0.798	0.905	0.865	0.828	0.854	0.912	0.065

(b) *Post-discretized data for an* in-silico *regulatory network consisting of 10 genes,* $k = 2$.

	wt	G1(-/-)	G2(-/-)	G3(-/-)	G4(-/-)	G5(-/-)	G6(-/-)	G7(-/-)	G8(-/-)	G9(-/-)	G10(-/-)
G1	0	0	1	0	0	0	0	0	0	0	0
G2	1	1	0	1	1	1	1	1	1	1	1
G3	0	0	1	0	0	0	0	0	0	0	0
G4	1	0	1	0	0	1	1	1	1	1	1
G5	1	1	1	1	1	0	1	1	1	0	1
G6	0	0	1	0	0	0	0	0	0	0	0
G7	0	0	0	0	0	0	0	0	1	0	1
G8	1	1	1	1	1	1	1	1	0	1	1
G9	1	1	1	1	1	1	1	1	1	0	1
G10	1	1	1	1	1	1	1	1	1	1	0

(c) *Post-discretized data for an* in-silico *regulatory network consisting of 10 genes,* $k = 4$.

	wt	G1(-/-)	G2(-/-)	G3(-/-)	G4(-/-)	G5(-/-)	G6(-/-)	G7(-/-)	G8(-/-)	G9(-/-)	G10(-/-)
G1	0	0	1	0	0	0	0	0	0	0	0
G2	1	1	0	1	1	1	1	1	1	1	1
G3	0	0	1	0	0	0	0	0	0	0	0
G4	1	0	1	0	0	1	1	1	1	1	1
G5	1	1	1	1	1	0	1	1	1	0	1
G6	0	0	1	0	0	0	0	0	0	0	0
G7	0	0	0	0	0	0	0	0	1	0	1
G8	1	1	1	1	1	1	1	1	0	1	1
G9	1	1	1	1	1	1	1	1	1	0	1
G10	1	1	1	1	1	1	1	1	1	1	0

data and will extract as much of the signal that is present in the data.

A third method for discovering gene regulatory networks is being attempted by Nathan Barker [15] of the University of Utah. His method first divides the microarray expression data into three categories; high, medium and low. Each gene is assigned one of these values based on its relative expression level when compared to each of the other genes' expression levels. This categorization of genes assigns an approximately equal number of genes to each of the three categories. Barker's algorithm then uses the changes between these categories to build an influence vector that shows the degree of influence each gene has on every other gene. Barker's algorithm assumes it has time series data when comparisons between samples are made. This assumption allows the algorithm to decide which gene is promoting or inhibiting another, rather than just find that there is a promoting or inhibiting relationship between two genes.

The main problem with Barker's algorithm is in the use of time series data. The simple fact that several samples are taken sequentially does not guarantee that they are sequential as far as the biological model is concerned. The time taken for genes interact with one another is too small for us to accurately measure, and certainly too small to isolate the system at each time step, which would be necessary to fulfill Barker's time series assumption. While Barker's ideas may be theoretically sound, in practice, the actual acquisition of time series microarray data that is accurate with respect to the biological model is impractical. Since AIRnet does not expect time series data, it can be applied to more practical microarray datasets.

Other methods of microarray analysis are based on statistical significance tests. Gene Set Enrichment Analysis, proposed by Subramanian *et al.*[16], and Significance Analysis of Microarray Gene Sets, proposed by Dinu *et al.*[17], are two methods that apply statistical tests to previously

Table II: Activation State Change Examples - genes are divided by row, samples are divided by column

(a) *equal, non-zero activation state changes - v_{xy} incremented*

	wt	G1(-/-)	G2(-/-)	G3(-/-)	G4(-/-)	G5(-/-)	G6(-/-)	G7(-/-)	G8(-/-)	G9(-/-)	G10(-/-)
G1	0	0	1	0	0	0	0	0	0	0	0
G2	1	1	0	1	1	1	1	1	1	1	1
G3	0	0	1	0	0	0	0	0	0	0	0
G4	1	0	1	0	0	1	1	1	1	1	1
G5	1	1	1	1	1	0	1	1	1	0	1
G6	0	0	1	0	0	0	0	0	0	0	0
G7	0	0	0	0	0	0	0	0	1	0	1
G8	1	1	1	1	1	1	1	1	0	1	1
G9	1	1	1	1	1	1	1	1	1	0	1
G10	1	1	1	1	1	1	1	1	1	1	0

(b) *equal magnitude, opposing-signed activation state changes - v_{xy} decremented*

	wt	G1(-/-)	G2(-/-)	G3(-/-)	G4(-/-)	G5(-/-)	G6(-/-)	G7(-/-)	G8(-/-)	G9(-/-)	G10(-/-)
G1	0	0	1	0	0	0	0	0	0	0	0
G2	1	1	0	1	1	1	1	1	1	1	1
G3	0	0	1	0	0	0	0	0	0	0	0
G4	1	0	1	0	0	1	1	1	1	1	1
G5	1	1	1	1	1	0	1	1	1	0	1
G6	0	0	1	0	0	0	0	0	0	0	0
G7	0	0	0	0	0	0	0	0	1	0	1
G8	1	1	1	1	1	1	1	1	0	1	1
G9	1	1	1	1	1	1	1	1	1	0	1
G10	1	1	1	1	1	1	1	1	1	1	0

(c) *activation state changes equal to zero - q_{xy} incremented*

	wt	G1(-/-)	G2(-/-)	G3(-/-)	G4(-/-)	G5(-/-)	G6(-/-)	G7(-/-)	G8(-/-)	G9(-/-)	G10(-/-)
G1	0	0	1	0	0	0	0	0	0	0	0
G2	1	1	0	1	1	1	1	1	1	1	1
G3	0	0	1	0	0	0	0	0	0	0	0
G4	1	0	1	0	0	1	1	1	1	1	1
G5	1	1	1	1	1	0	1	1	1	0	1
G6	0	0	1	0	0	0	0	0	0	0	0
G7	0	0	0	0	0	0	0	0	1	0	1
G8	1	1	1	1	1	1	1	1	0	1	1
G9	1	1	1	1	1	1	1	1	1	0	1
G10	1	1	1	1	1	1	1	1	1	1	0

selected groups of genes, such as genes in the same signaling pathway. These methods attempt to identify whether the gene sets are associated with a particular phenotype. Similarly, AIRnet can highlight a subset of genes within the entire network, either treating the subset as the entire network, or showing only genes, which are highly correlated with one or more genes in the specified subset. AIRnet will also compare networks to identify phenotype-specific gene correlations.

III. Methods

AIRnet infers a gene regulatory network, given reasonable assumptions about microarray data. Non-time series data from different samples are used to find correlated patterns in gene expression. The correlation of gene expression changes is used to create influence vectors, which highlight the genes with the highest probability of being correlated. Ultimately, AIRnet is designed to compare two networks of different genotypes, in order to draw out differences that will assist researchers with the development of treatments.

The first step in analyzing microarray data is discretizing the data (*table Ia*). While discretizing, the data for a single gene across all the samples is considered a single dataset. Each dataset is clustered using k-means clustering. After all the data sets have been discretized, the data for each gene in every sample is classified as a number between 0 and $k-1$ (*table Ib*). The new value represents the relative level of activation for a particular gene, as compared between samples.

After discretizing the data, AIRnet performs a pairwise comparison of the change in activation state between samples for all genes.

Table III: Reporting AIRnet results for 3 network sizes from the DREAM3 competition. The empty network row shows values for graphs with 0 edges and provide a baseline to interpret the scores for other networks. The empty network scores were generated by submitting an empty file with no predictions to the DREAM score evaluator. The other rows correspond to the type of data used to infer the networks. In the score column, larger values are better. In the other two columns, smaller values are better. Scores reported are produced using an 80% threshold parameter for AIRnet.

(a) *Reporting values for 10-gene networks.*

	score	AUPR p-value average	AUROC p-value average
empty network	1.1816e+00	8.5675e-03	5.0578e-01
trajectories	1.6298e+00	2.1759e-03	2.5279e-01
heterozygous	2.2401e+00	3.6441e-04	9.0845e-02
null-mutant	2.8198e+00	4.1550e-05	5.5198e-02

(b) *Reporting values for 50-gene networks.*

	score	AUPR p-value average	AUROC p-value average
empty network	2.4438e+00	2.6065e-05	4.9687e-01
trajectories	2.6700e+00	6.1865e-06	7.3901e-01
heterozygous	2.6207e+00	7.7215e-06	7.4297e-01
null-mutant	1.4152e+01	5.2984e-26	9.3634e-04

(c) *Reporting values for 100-gene networks.*

	score	AUPR p-value average	AUROC p-value average
empty network	5.2312e+00	6.8572e-11	5.0297e-01
trajectories	3.8264e+00	4.7395e-08	4.6923e-01
heterozygous	5.2881e+00	6.3523e-11	4.1762e-01
null-mutant	3.7911e+01	1.0263e-71	1.4694e-05

$$v_{xy} = \frac{q_{xy} + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f(x_i, x_j, y_i, y_j)}{n * (n-1)/2} \quad (1)$$

$$f(x_i, x_j, y_i, y_j) = \begin{cases} 1 & \text{if } (x_i > x_j) \wedge (y_i > y_j) \text{ or} \\ & \quad (x_i < x_j) \wedge (y_i < y_j); \\ 0 & \text{if } (x_i = x_j) \vee (y_i = y_j); \\ -1 & \text{if } (x_i > x_j) \wedge (y_i < y_j) \text{ or} \\ & \quad (x_i < x_j) \wedge (y_i > y_j). \end{cases}$$

Comparing genes x and y, AIRnet calculates an influence vector, v_{xy}, representing how correlated x and y appear to be (*equations 1 and 2*). Negative values of v_{xy} indicate a inhibiting relation between x and y, while a positive value of v_{xy} indicates a promoting relation.

$$q_{xy} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} g(x_i, x_j, y_i, y_j) \quad (2)$$

$$g(x_i, x_j, y_i, y_j) = \begin{cases} 1 & \text{if } x_i = x_j, y_i = y_j, \\ & \quad \text{and } x_i - y_i < k/2; \\ -1 & \text{if } x_i = x_j, y_i = y_j, \\ & \quad \text{and } x_i - y_i >= k/2; \\ 0 & \text{otherwise.} \end{cases}$$

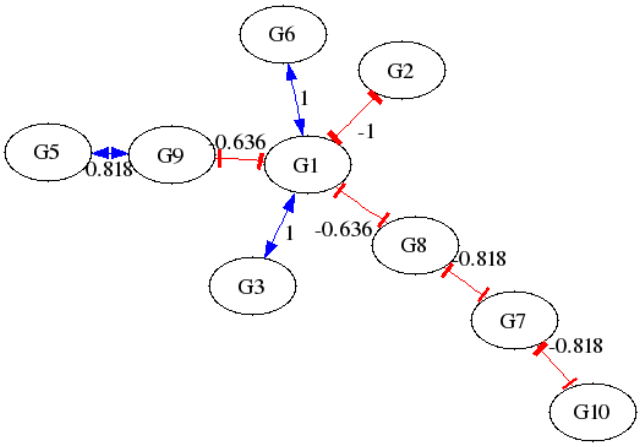
Following the calculation of v_{xy} for all values of x and y, AIRnet reconstructs the regulatory network by including edges that have the highest correlation.

AIRnet produces a graph, G representation of the regulatory network, where each node, x, represents a single gene, and each edge, $(\{x, y\}, v)$ represents an interaction between x and y. The sign of v shows the interaction between x and y as either promoting or inhibiting, while $|v|$ shows the probability of x interacting with y. To form the graph, AIRnet adds the edges $(\{x, y\}, w_{xy})$, where $w_{xy} = 1 - |v_{xy}|$, for all values of x and y. To prune edges out of the graph, Kruskal's Algorithm is used to find the minimum cost spanning tree of the graph G, with the addition of stopping the production of the minimum cost spanning tree when the value of $|w_{xy}|$ for the next edge to be added falls below a user-defined threshold. The w_{xy} values are then exchanged with their corresponding v_{xy} values. This effectively deals with the mutual information problem[18].

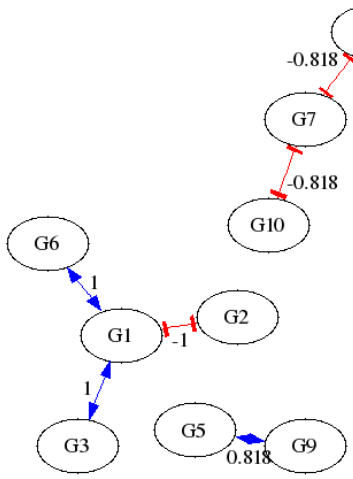
IV. Results

Tests are run on data from *in-silico* regulatory networks, originally created for the DREAM3 competition[1]. Three types of data are used to test AIRnet's accuracy for each of the *in-silico* regulatory networks. The data types are labeled as heterozygous knock-down, null-mutant, and trajectory. The heterozygous knock-down and null-mutant data sets each contain data for the steady states of the wild-type as well as knock-down or knock-out data for each gene.

(a) Network inferred using 40% threshold



(b) Network inferred using 80% threshold



(c) True *in-silico* network

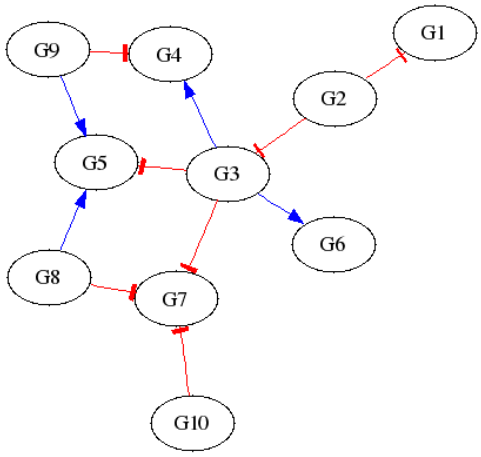


Figure 1: Example *in-silico* regulatory network, as inferred by AIRnet.

Trajectory data sets are comprised of time series data, with 21 time points, for each network recovering from external perturbations. Each network is subjected to, and has data for, a number of perturbations equal to 46% of the number of genes within the network.

Figures 0a and 0b depict regulatory networks inferred by AIRnet using data produced by one of the networks generated for DREAM3. The *in-silico* network is shown in figure 0c. Figures 0a and 0b show that higher threshold values produce more selective networks by excluding connections for which the correlation between the two genes is not great enough. Visual verification, however, is not always the best method for measuring the accuracy of an inferred regulatory network, especially if the network is large.

Scoring metrics from the DREAM3 competition are used to verify the statistical significance of AIRnet's reconstructed regulatory networks. The DREAM3 metrics calculate the AUROC and AUPR values and compare the resulting values with the AUROC and AUPR of 100,000 randomly generated networks to compute the probability of randomly creating a network with equal or greater AUROC and AUPR values, producing a p-value for both the AUROC and AUPR. The AUROC p-values are combined by averaging the scores for same-sized networks. The same is done for the AUPR p-values. The averaged AUROC and AUPR p-values are subsequently combined as a log-transformed average, $-log_{10}(AUROC_p * AUPR_p)/2$. Each log-transformed average provides a single value, which summarizes AIRnet's accuracy for five individual, same-sized networks.

Because the graphs AIRnet produces are signed and undirected, the standards, against which AIRnet is being measured, were modified to be undirected as well.

The score, along with the AURR and AUROC p-values, are displayed in table III. The first row in tables IIIa - IIIc, the empty network, report the values obtained from reporting a network with zero edges, or a network which assumes genes do not interact in any way with each other. The empty network is not produced by AIRnet, but is included as a baseline for comparing AIRnet's accuracy using the supplied data types. Each other row corresponds to a type of data used to infer the networks, as specified by the first column.

As seen in table III, the null-mutant data produces significantly better results than either of the other two data types. The networks AIRnet infers using null-mutant data appear to be only marginally better when inferring small networks, however, as the network size grows, the null-mutant produced networks' accuracy grows at a much faster rate than the accuracy for networks produced by either the heterozygous data, or the trajectory (time series) data.

It is interesting to note, using the trajectories data to infer networks gave the lowest scores of all the data types, in one case, scoring even lower than the empty network even though the data was produced by a simulator with accurate time series outputs (table IIIc).

Comparing the values in table III with the 300 submissions to the DREAM3 competition ranks the AIRnet in the top 5 performers. This comparison ignores directionality, which would probably lower the AIRnet ranking. The results are promising and are obtained using microarray assumptions that can be met by most biological experiments.

Experiments were also performed using microarrays from mice with Down Syndrome phenotypes. An average of 60% of the connections predicted by AIRnet were validated using the KEGG[19] database. The KEGG database is biased towards more studied regulatory pathways, so actual accuracy of AIRnet is probably higher than this number.

V. Conclusion

AIRnet uses influence vector to infer regulatory networks from microarray data with practical assumptions. The microarray data does not have to have time-series characteristics and no constraints are placed on the structure of the matrices. Networks inferred by AIRnet are comparable in accuracy to the best algorithms participating in the DREAM3 competition even though many of these algorithms were more restrictive on the kind of data they could use. Edges predicted by AIRnet compare favorably with experimentally validated regulatory networks found in KEGG. AIRnet can perform predictions on microarrays with 20,000 genes in less than 24 hours, making it practical for most analysis needs.

AIRnet can provide the understanding of gene regulatory networks necessary for the impending major breakthroughs in agriculture and medicine.

References

[1] D. Marbach, T. Schaffter, C. Mattiussi, and D. Floreano, "Generating Realistic in silico Gene Networks for Performance Assessment of Reverse Engineering Methods," *Journal of Computational Biology*, 2009.

[2] F. S. Collins and V. A. McKusick, "Implications of the human genome project for medical science," *JAMA*, vol. 285, no. 5, pp. 540–544, 2001. [Online]. Available: http://jama.ama-assn.org/cgi/content/abstract/285/5/540

[3] E. H. Davidson, J. P. Rast, P. Oliveri, A. Ransick, C. Calestani, C.-H. Yuh, T. Minokawa, G. Amore, V. Hinman, C. Arenas-Mena, O. Otim, C. T. Brown, C. B. Livi, P. Y. Lee, R. Revilla, A. G. Rust, Z. j. Pan, M. J. Schilstra, P. J. C. Clarke, M. I. Arnone, L. Rowen, R. A. Cameron, D. R. McClay, L. Hood, and H. Bolouri, "A genomic regulatory network for development," *Science*, vol. 295, no. 5560, pp. 1669–1678, 2002. [Online]. Available: http://www.sciencemag.org/cgi/content/abstract/295/5560/1669

[4] M. Levine and E. H. Davidson, "From the cover: Gene regulatory networks special feature: Gene regulatory networks for development," *Proceedings of the National Academy of Sciences*, vol. 102, no. 14, pp. 4936–4942, 2005. [Online]. Available: http://www.pnas.org/cgi/content/abstract/102/14/4936

[5] F. Carrari, E. Urbanczyk-Wochniak, L. Willmitzer, and A. R. Fernie, "Engineering central metabolism in crop species:: learning the system," *Metabolic Engineering*, vol. 5, no. 3, pp. 191–200, 7 2003.

[6] C.-j. Li, R. W. Li, Y.-h. Wang, and T. H. Elsasser, "Pathway analysis identifies perturbation of genetic networks induced by butyrate in a bovine kidney epithelial cell line," *Functional and Integrative Genomics*, vol. 7, no. 3, pp. 644–652, 2006.

[7] A. Stephansky and T. Leustek, "Histidine biosynthesis in plants," *Amino Acids*, vol. 30, no. 2, pp. 127–142, 2006.

[8] M. O. Hoque, M. S. Kim, K. L. Ostrow, J. Liu, G. B. A. Wisman, H. L. Park, M. L. Poeta, C. Jeronimo, R. Henrique, A. Lendvai, E. Schuuring, S. Begum, E. Rosenbaum, M. Ongenaert, K. Yamashita, J. Califano, W. Westra, A. G. van der Zee, W. V. Criekinge, and D. Sidransky, "Genome-wide promoter analysis uncovers portions of the cancer methylome," *Cancer Res*, vol. 68, no. 8, pp. 2661–2670, 2008. [Online]. Available: http://cancerres.aacrjournals.org/cgi/content/abstract/68/8/2661

[9] T. Ideker and R. Sharan, "Protein networks in disease," *Genome Res.*, vol. 18, no. 4, pp. 644–652, 2008. [Online]. Available: http://www.genome.org/cgi/content/abstract/18/4/644

[10] K. Rubenstein, *Systems Biology: A Disruptive Technology*. CHI Insight Pharma Reports, May 2008.

[11] M. van Greevenbroek, J. Zhang, C. J. Kallen, P. Schiffers, E. Feskens, and T. de Bruin, "Effects of interacting networks of cardiovascular risk genes on the risk of type 2 diabetes mellitus (the codam study)," *BMC Medical Genetics*, vol. 9, no. 1, p. 36, 2008. [Online]. Available: http://www.biomedcentral.com/1471-2350/9/36

[12] J. Zhu, B. Zhang, and E. E. Schadt, *A Systems Biology Approach to Drug Discovery*, ser. Advances in Genetics,. Academic Press, 2008, vol. Volume 60.

[13] Y. Wang, T. Joshi, X.-S. Zhang, D. Xu, and L. Chen, "Inferring gene regulatory networks from multiple microarray datasets," *Bioinformatics*, vol. 22, no. 19, pp. 2413–2420, 2006. [Online]. Available: http://bioinformatics.oxfordjournals.org/cgi/content/abstract/22/19/2413

[14] J. C. Liao, R. Boscolo, Y.-L. Yang, L. M. Tran, C. Sabatti, and V. P. Roychowdhury, "Network component analysis: Reconstruction of regulatory signals in biological systems," *Proceedings of the National Academy of Sciences*, vol. 100, no. 26, pp. 15 522–15 527, 2003.

[15] N. A. Barker, "Learning genetic regulatory network connectivity from time series data," Ph.D. dissertation, University of Utah, December 2007.

[16] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov, "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15 545–15 550, 2005. [Online]. Available: http://www.pnas.org/content/102/43/15545.abstract

[17] I. Dinu, J. Potter, T. Mueller, Q. Liu, A. Adewale, G. Jhangri, G. Einecke, K. Famulski, P. Halloran, and Y. Yasui, "Improving gene set analysis of microarray data by sam-gs," *BMC Bioinformatics*, vol. 8, no. 1, p. 242, 2007. [Online]. Available: http://www.biomedcentral.com/1471-2105/8/242

[18] P. Meyer, F. Lafitte, and G. Bontempi, "minet: A r/bioconductor package for inferring large transcriptional networks using mutual information," *BMC Bioinformatics*, vol. 9, no. 1, 2008. [Online]. Available: http://dx.doi.org/10.1186/1471-2105-9-461

[19] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, and Y. Yamanishi, "Kegg for linking genomes to life and the environment," *Nucleic Acids Res.*, vol. 36, pp. D480–D484, 2008.