Using Protein Interaction Networks to Understand Complex Diseases

Mehmet Koyutürk, *Case Western Reserve University*

Recent developments in biotechnology have enabled interrogation of the cell at various levels, leading to many types of "omic" data that provide valuable information on multiple genetic and environmental factors and their interactions.

One of biology's major challenges in the postgenomic era is systems-level characterization of complex human diseases—that is, diseases that result from the interplay among multiple genetic and environmental factors. With significant advances in high-throughput screening technologies such as next-generation sequencing, it is now possible to develop computational techniques for understanding disease development and progression.

Figure 1 shows the commonly used technologies for interrogating cellular systems and the data they generate. Such technologies generate genome-scale data on biomolecular sequences and structures, the abundance of individual molecules under various conditions, and the physical interactions among two or more molecules.

A common approach to analyzing disease-specific molecular data is to compare the abundance—the gene or protein expression—of each molecule in an affected sample against a control or healthy sample. These analyses reveal individual molecules potentially associated with disease that can serve as candidate biomarkers for use in diagnosis, prognosis, risk assessment, treatment choice, and therapeutic intervention.

However, because the data used in such analyses represents a snapshot of a complex dynamic system, it can be quite noisy, and the signatures of individual markers might be highly variable across different patients.[1] Furthermore, disease mechanisms often involve complex interactions among multiple molecules, and the patient population might not be homogeneous in terms of these mechanisms.[2] Consequently, functionally relevant differences in affected and control samples could present as subtle differences in the properties of individual molecules, which may not be detected by analyzing each molecule in isolation from other molecules.[3] Today, the scientific community seems to be in agreement that systems-level approaches are required to better understand the mechanisms of complex diseases.[4]

Several algorithms aim to overcome the limitations of individual gene based approaches by utilizing established biological pathway information or high-throughput data on protein-protein interactions (PPIs) to identify groups of functionally related genes. This article focuses on identifying *dysregulated subnetworks* through the integration of gene expression and protein-protein interaction (PPI) data.

DIFFERENTIAL GENE EXPRESSION

Many complex diseases are associated with genomic sequence alterations. Therefore, comparative analysis of genomic sequences in healthy and affected populations is quite effective in locating the possible genetic sources of disease phenotype. However, genomic data does not directly capture the regulatory mechanisms that mediate the link between genetics and disease progression. To this end, gene expression, measured in terms of the abundance of mRNA molecules in a tissue sample, proves useful in elucidating the variation in cellular system activity. Indeed, in the past decade, researchers often used genome-wide monitoring of gene expression—the

0018-9162/12/$31.00 © 2012 IEEE Published by the IEEE Computer Society

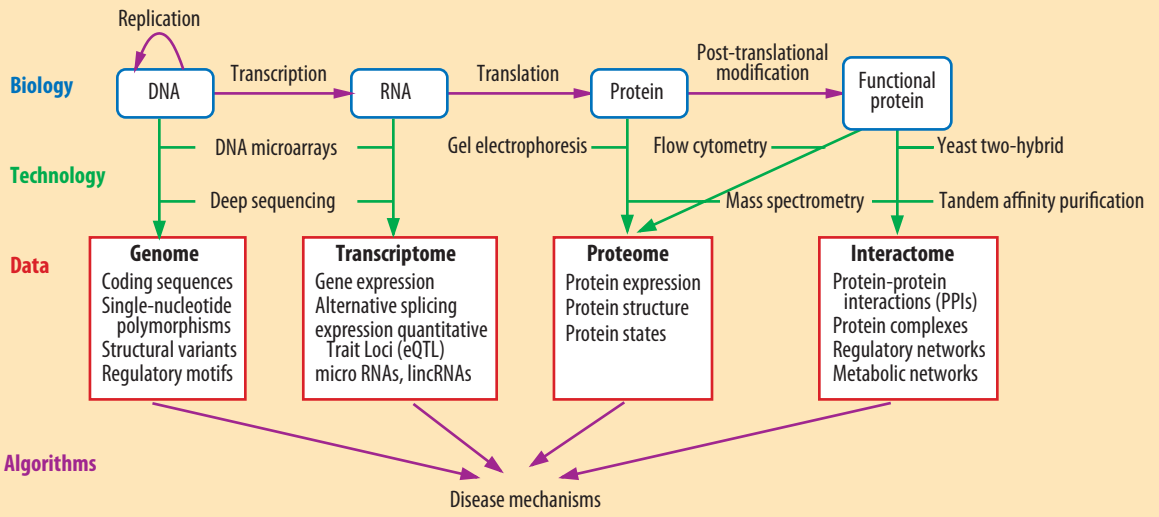

Figure 1. Large-scale interrogation of cellular systems at multiple levels generates various types of "omic" data. For each data type, a corresponding box provides common examples of genome-scale data, but the lists are not exhaustive. Computational algorithms facilitate innovative integration of omic datasets to gain insights into the underlying mechanisms of complex diseases.

transcriptome—enabled via DNA microarray technology, to investigate disease mechanisms.[5] Today, with the advent of next-generation sequencing (NGS) techniques, it is possible to monitor gene expression even more efficiently and reliably.

Differential analysis of gene expression facilitates the identification of dysregulated genes in the disease of interest—that is, those genes that exhibit significant difference in the amount of mRNA transcripts present in a range of disease and control samples. To date, systematic analyses of differential gene expression have led to identification of the genetic markers associated with many cancers, as well as genes associated with tumor grade, metastasis, and disease recurrence.[1] Because knowledge of individual dysregulated genes provides limited insights into the dysregulation of cellular processes as a system, researchers increasingly focus on interpreting these findings in the context of cellular systems.[5]

PATHWAYS AND ENRICHMENT ANALYSIS

A popular approach to systems-level interpretation of differential gene expression is to identify biological pathways that are significantly enriched in products of differentially expressed genes. In general, a pathway is defined as a sequence of biochemical reactions or interactions that describe a particular biological process with specified inputs and outputs. Because pathways generally describe relatively well-characterized biological processes, those pathways enriched in products of differentially expressed genes offer immediate clues to biological processes affected in the disease phenotype.

Many statistical and computational methods are available to systematically interpret differential gene expression via pathway-based enrichment analysis. Among these, gene-set enrichment analysis is commonly used.[6] GSEA takes as input a gene expression dataset and a set of genes—for example, genes that code for proteins in a particular pathway—and assesses the overall rank of the genes in the set among all genes in the dataset in terms of their differential expression in the disease of interest. If the genes that code for proteins in a pathway rank significantly higher compared to other genes in the entire genome, then the pathway is considered dysregulated in the disease.

Following GSEA, researchers developed many other methods to improve statistical procedures for pathway-based differential expression analysis, but these methods are conceptually similar. Today, medical scientists often use enrichment analysis as a follow-up procedure for gene expression assays comparing phenotypes. Many commercial software products and databases also provide functionalities for enrichment analysis—for example, MetaCore and Ingenuity Pathway Analysis—but the algorithmic/statistical frameworks that such tools utilize generally are not transparent to the user.

While being quite useful and directly interpretable, pathway-based analysis of differential expression has important limitations. In particular, pathway-based approaches restrict the functional relationships among genes and proteins to established biological knowledge on well-characterized pathways. Therefore, these approaches generally cannot characterize the differential expression of relatively less studied genes, discover novel

functional links among genes, identify disease-specific crosstalk between different pathways, or elucidate potentially combinatorial relationships among multiple genes. Protein-protein interaction (PPI) networks offer an invaluable resource in this regard.

PPI NETWORKS

High-throughput experiments and computational prediction techniques can identify physical and functional interactions among proteins at a large scale.[7] Although high-throughput screening methods are noisy and incomplete, and computational methods have limited prediction accuracy, the data that all these methods generate can be integrated to construct networks of PPIs that serve as maps of the cell's functional organization.[8] This data is organized into public databases, including the Human Protein Reference Database[9] and BioGrid.[10]

Because PPI networks are derived from a diverse range of experimental sources, they contain potential functional links that have not yet been characterized in detail. Furthermore, because they provide a comprehensive map of functional interactions in the cell, PPI networks are useful for analyses that take into account the cell's global organization. For these reasons, network-based analyses of differential expression can prove more promising than pathway-based analyses in discovering novel relationships among genes in the context of disease mechanisms. Note that data on molecular interactions is not limited to PPIs: transcriptional regulatory networks, genetic interactions, and metabolic networks also provide genome-wide information on the functional links among genes and proteins.[11]

DYSREGULATED SUBNETWORKS

Systematic studies of differential gene expression in several complex diseases show that dysregulated genes in similar diseases are likely to interact with each other in PPI networks.[12] Motivated by this observation, researchers have developed many methods to identify dysregulated subnetworks—that is, connected subgraphs of the human PPI network that exhibit collective differential expression with respect to the disease phenotype.

Table 1 shows a comparison of pathway-based and protein-protein interaction network-based approaches. The key difference between network-based algorithms and pathway enrichment analysis is that the functional relationships among proteins are not restricted to well-characterized pathways in network-based analysis. Rather, any group of proteins functionally linked through interactions in the global PPI network is considered as a potential group of proteins with collective dysregulation. Therefore, in addition to dysregulated pathways, these algorithms also have the potential to discover dysregulated parts of pathways and dysregulated functional links across multiple pathways.[13]

Table 1. Comparison of pathways-based and protein-protein interaction network-based approaches.

Characteristic	Pathways	PPI network
Source	Literature driven	Experimentally identified
Scale	Process specific	Genome wide
Reliability	Well characterized	Noisy, incomplete
Procedure	Statistical tests	Combinatorial algorithms
Outcome	Dysregulated pathways	Dysregulated parts of pathways, disease-associated crosstalk among different pathways, combinatorial logic of disease mechanisms

However, since this formulation requires searching the PPI network's entire subnetwork space, it leads to intractable computational problems and requires development of sophisticated computational algorithms for efficient discovery of dysregulated subnetworks.

Algorithms for identifying dysregulated subnetworks generally integrate two different sources of molecular data: a *transriptomic* dataset that includes measurement of gene expression from disease and control samples, and an *interactomic* dataset that includes interactions among proteins coded by genes in the gene expression dataset. Here, we use the following notation to describe these algorithms:

- V, the set of genes for which gene expression data is available;
- e_i, the m-dimensional expression profile of $g_i \in V$, such that $e_i(j)$ denotes the expression of g_i in sample j for $1 \leq j \leq m$ and m denotes the number of samples in the dataset;
- C, an m-dimensional phenotype vector, such that $C(j) = 1$ indicates that sample j is a phenotype sample (for example, taken from tumor tissue), while $C(j) = 0$ indicates that sample j is a control sample (for example, taken from a normal tissue); and
- $G = (V, E)$, the undirected graph that describes the PPI network, where V represents the genes that code for the proteins in the network and E represents the set of interactions among these proteins.

Because we construct PPI networks by integrating data from multiple sources, the edges of G are generally weighted, representing each interaction's reliability. However, most of the algorithms discussed here do not directly use these weights; rather, they apply a threshold on the reliability scores and represent the network as an unweighted graph. Recent work demonstrates that incorporating reliability scores can improve algorithm performance.[14]

Additive coordinate dysregulation

Early algorithmic approaches identify dysregulated or "active" subnetworks by searching for connected subgraphs of the PPI network with a high aggregate significance in the differential gene expression.[15] While these approaches can extract functional links among genes that exhibit individual differential expression, they are limited in capturing coordination in the dysregulation of multiple genes, since they assess differential expression individually for each gene. Recognizing this limitation, Han-Yu Chuang and colleagues[16] proposed an information-theoretic formulation of subnetwork dsyregulation that takes into account the sample-specific variation in the expression of genes in a subnetwork. For this purpose, they introduced the notion of subnetwork activity, defined as the aggregate expression of gene products in the subnetwork in each sample—that is, the activity of subnetwork $S \subseteq V$ is defined as

$$e(S) = \sum g_i \in S^{\frac{e_i}{\sqrt{|S|}}}. \quad (1)$$

Subsequently, they quantify the dysregulation of subnetwork S as

$$\Delta_{additive}(S) = I(\overline{e}(S); C) = H(C) - H(C|\overline{e}(S)). \quad (2)$$

Here, $I(\overline{e}(S); C)$ denotes the mutual information between the phenotype vector C and the activity of subnetwork S, that is, it is the reduction in phenotype uncertainty upon observing the aggregate expression of the genes coding for the proteins in S. Phenotype uncertainty is quantified by entropy $H(C)$, and the uncertainty after observation of subnetwork activity is quantified by conditional entropy $H(C\overline{e}(S))$. Here, $\Delta_{additive}$ is referred to as additive coordinate dysregulation because it is based on additive assessment of the coordination among multiple genes.

To identify network markers of breast cancer metastasis, Chuang and colleagues searched for subnetworks of the human PPI network that maximize $\Delta_{additive}(S)$ using a greedy algorithm. Next, they used the activity of significantly dysregulated subnetworks as features for classification. As compared to single gene markers, these subnetwork markers exhibited improved accuracy in predicting breast cancer metastasis.[16]

Coordinate dysregulation and set cover

Although the concept of coordinate dysregulation is promising for capturing the interplay among multiple interacting proteins, the problem of identifying subnetworks with significant coordinate dysregulation is intractable. Furthermore, bottom-up heuristics that grow subnetworks to greedily maximize the objective function may lack global awareness. Motivated by these considerations, recent work formulates the problem of identifying coordinately dysregulated subnetworks as a variation of the well-known set-cover problem.[17,18] Although set-cover is also an NP-hard problem, this formulation provides interpretable insights into the coordination among multiple proteins, enabling development of more effective heuristic algorithms.[18]

To see the relationship between coordinate dysregulation and set-cover, consider gene expression data from paired samples. A gene g_i is said to cover sample j positively/negatively if it is up-regulated/down-regulated in the phenotype sample with respect to control—for example, $\hat{e}_i(j) = H$ and $\hat{e}_i(j') = L$, where $\hat{e}_i(j) \in \{L, H\}$ represents the binarized expression of gene g_i in sample j and j' denotes the control sample paired with phenotype sample j. Subsequently, the set of samples covered positively/negatively by g_i is called the positive/negative cover set of g_i and respectively denoted P_i/N_i.

Based on this formulation, as Equation 2 shows, Δadditive($\{g_i\}$), the dysregulation of a single gene, is a monotonically increasing function of $||P_i| - |N_i||$. Motivated by this insight, the NETCOVER algorithm searches for subnetworks comprising genes that together cover all samples consistently, either positively or negatively.[17] Here, for a given subnetwork $S \subseteq V$, the positive and negative cover sets of S are

$$P(S) = \bigcup_{g_i \in S} P_i \text{ and } N(S) = \bigcup_{g_i \in S} P_i. \quad (3)$$

Figure 2 illustrates this concept. In the context of human colorectal cancer, NETCOVER identifies subnetworks that provide better classification accuracy than subnetworks identified by a greedy algorithm that aims to explicitly maximize additive coordinate dysregulation.[17]

Synergistic dysregulation

Dimitris Anastassiou[19] further delineated the concept of coordinate dysregulation to assess the synergistic dysregulation among two or more genes. For a pair of genes g_i and g_j, the synergistic dysregulation of g_i and g_j is defined as

$$\Delta_{synergistic}(\{g_i, g_j\}) = I(\{\hat{e}_i, \hat{e}_i\}; C) - (I(\hat{e}_i; C) + I(\hat{e}_j; C)). \quad (4)$$

Observe that synergistic dysregulation is different from additive coordinate dysregulation in two aspects. First, the dysregulation of the subnetwork comprising g_i and g_j is quantified in terms of the mutual information between phenotype and the binary expression state of the genes, which is a two-dimensional binary vector (as opposed to the average expression of the genes). Second, the dysregulation of each individual gene is subtracted from the overall dysregulation of the subnetwork to capture the pair's ability to distinguish phenotype and control beyond what each individual gene can distinguish.

In this respect, if two genes exhibit correlated differential expression in phenotype samples, their synergistic dysregulation will be negative. On the other hand, if the two genes are complementary with each other in distinguishing phenotype and control, their synergistic dysregulation will be positive. Therefore, the concept of synergy provides a useful measure for quantifying the complementarity and redundancy of two genes in distinguishing phenotype and control.

We can extend the concept of synergistic dysregulation to a subnetwork comprising multiple genes by defining the expression state of a set S of k genes as $F_S = \{\hat{e}_1, \hat{e}_2 \ldots, \hat{e}_k\} \in \{L, H\}^k$, that is, a random variable represents the combination of binary expression states of the genes in S. However, computation of the synergy of k genes requires exponential time in k because computing the synergy requires considering all subsets of S. Consequently, synergy computation for a given subnetwork of arbitrary size becomes an intractable problem, not to mention the problem of identifying subnetworks with high synergy. For this reason, in the context of prostate cancer, John Watkinson and colleagues[20] limited their attention to pairs of genes and constructed a synergy network by representing identified synergistic relationships as interactions between pairs of genes.

Combinatorial coordinate dysregulation

To overcome the computational difficulties in computing synergy but still capture the combinatorial relationship in the dysregulation of multiple interacting genes, Salim Chowdhury and colleagues[21] defined combinatorial coordinate dysregulation of a subnetwork S as follows:

$$\Delta_{combinational}(S) = I(F_S, C) = H(C) - H\left(C \middle| \hat{e}_1, \hat{e}_2 \ldots, \hat{e}_k\right) \quad (5)$$

The difference between synergistic and combinatorial dysregulation is that combinatorial dysregulation does not account for the dysregulation of parts of the subnetwork. However, in contrast to additive coordinate dysregulation, which requires all genes in the subnetwork to be dysregulated in the same direction, combinatorial coordinate dysregulation can discover combinatorial patterns of dysregulation among the genes in a subnetwork—for example, it can discover that patterns such as g_1 and g_3 are down-regulated in tumor samples, while g_2 is up-regulated. Such patterns can shed light on the combinatorial logic of disease mechanisms.

Although computation of combinatorial coordinate dysregulation is straightforward, the problem of identifying subnetworks with high combinatorial coordinate dysregulation is intractable. Motivated by this consideration, Chowdhury and colleagues[21] decomposed the combinatorial coordinate dysregulation of a subnetwork into components associated with individual subnetwork states. Namely, they defined

$$J(f_S; C) = p(f_S) \sum_{c \in \{0,1\}} p\left(c \middle| f_S\right) \log\left(p\left(c \middle| f_S\right) / p(c)\right), \quad (6)$$

where

$$I(F_S; C) = \sum_{f_S \in \{H,L\}^k} J(f_S; C). \quad (7)$$

Here, $f_S \in \{H, L\}^k$ denotes an observation of the random variable F_S, that is, f_S is a specific combination of the expression states of the genes in S, c denotes an observation of the random variable C (a specific phenotype) and $p(x)$ denotes $P(X = x)$, the probability that random variable X is equal to x (similarly, $p(x|y)$ denotes $P(X = x|Y = y)$).

In biological terms, we can consider $J(f_S; C)$ to be a measure of the information provided by subnetwork state f_S on phenotype C. Based on this definition, Chowdhury and colleagues derived a useful bound on $J(f_S; C)$ that can be computed using statistics of smaller subnetworks. Using this bound, they developed CRANE, an efficient branch-and-bound algorithm, to identify subnetworks and associated state functions. Subsequently, they trained neural networks to use these subnetworks for predicting colon cancer metastasis on independent datasets.

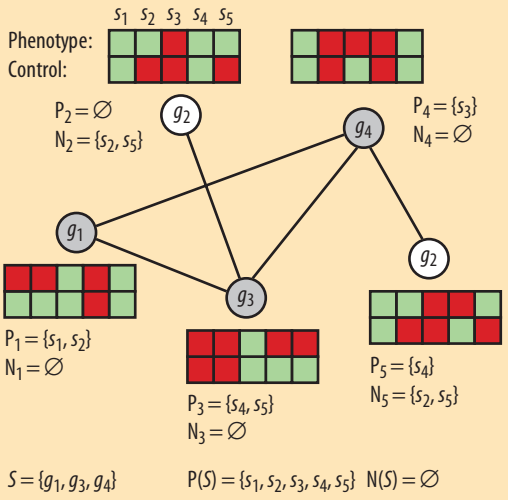

Figure 2. Set-cover based formulation of coordinate dysregulation in complex phenotypes. Circles represent proteins, and connecting lines represent interactions between proteins. The matrix near each protein shows its mRNA-level expression in phenotype (upper row) and control (lower row) samples. Dark red indicates high expression, and light green indicates low expression. In this example, $P_1 = \{s_1, s_2\}$, that is, g_1 covers samples s_1 and s_2 positively because it is up-regulated in the phenotype samples compared to the control samples. The subnetwork comprising the gray proteins covers all phenotype samples positively.

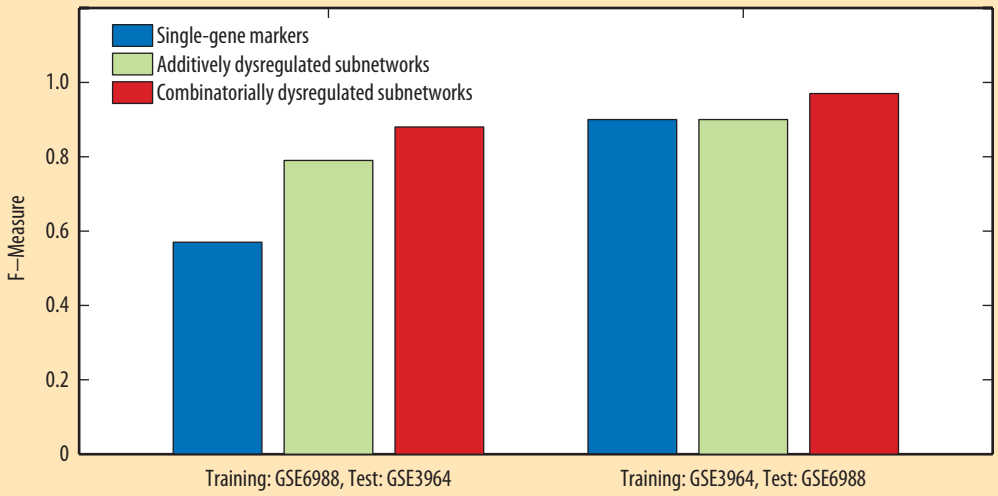

Figure 3. Performance of dysregulated subnetworks in predicting metastasis of colon cancer when used as features for classification. Two different datasets (GSE3964, GSE6988) from the Gene Expression Omnibus are used for cross-classification. Top-scoring single-gene markers, additively dysregulated subnetworks, and combinatorially dysregulated subnetworks are identified on the training dataset; neural networks are trained on the training dataset using a variable number of features (ranging from 1 to 20); and the resulting classifiers are tested on the test dataset. This figure shows the best F-measure achieved by each set of features, with F-measure defined as the harmonic mean of precision and recall.

Figure 3 shows the results of these cross-classification experiments. As the figure shows, combinatorially dysregulated subnetworks outperform additively dysregulated subnetworks and single gene markers in predicting metastasis.

Figure 4 shows a sample subnetwork that illustrates how the dysregulated subnetworks identified by computational algorithms can generate novel hypotheses for experimental biology. This subnetwork contains membrane-bound proteins TNFSF11, MMP1, BCAN, MMP2, TBSH1, and SPP1. For this subnetwork, the state function LLLLLH (in respective order) indicates metastatic phenotype with J-value 0.33. The combinatorial coordinate dysregulation of this subnetwork is 0.72, while its additive coordinate dysregulation is 0.37; this subnetwork would likely escape detection by the additive algorithm.

As seen, SPP1, TBSH1, and MMP2 interact with integrin alpha-chain V (ITGAV). Integrins are heterodimeric integral membrane proteins comprising an alpha chain and a beta chain. They are known to play a major role in mediating cell adhesion and cell motility, processes known to be involved in metastasis of colorectal cancer.[22] However, alpha-v/beta-5 integrin does not exhibit significant differential expression at the mRNA-level, suggesting that the subnetwork identified by CRANE may be a signature of its post-translational dysregulation in metastatic cells.

This subnetwork therefore suggests several follow-up experiments, including investigation of the post-translational dysregulation of integrin heterodimers in metastatic samples and pharmacological inhibition or siRNA interference of the integrin dimers to evaluate the role of these proteins in maintaining the metastatic phenotype.

Discriminative subnetworks

The methods discussed so far use information theoretic measures to formulate subnetwork dysregulation. Recent methods take an alternate approach and formulate the ability of a subnetwork to discriminate phenoype and control within the framework of a specific classification technique.

In particular, Phuong Dao and colleagues[23] developed an algorithm to identify optimally discriminative subnetworks in the context of response to chemotherapy in breast cancer patients. Their formulation treats each sample as a point in the $|V|$-dimensional space represented by all genes. Based on this representation, they defined a discriminative subnetwork as a set of at most k interacting gene products such that the total distance within each phenotype class is minimized and the total distance between different phenotype classes is maximized in the subspace characterized by these genes. Subsequently, they used a color-coding-based randomized algorithm to identify optimally discriminative subnetworks. Their results showed that this algorithm is more effective than greedy algorithms in discovering subnetworks that can accurately predict response to chemotherapy and are reproducible across multiple datasets.

Other researchers subsequently proposed developing network-guided forests by extending random forests to construct decision trees with nodes corresponding to interacting proteins in a subnetwork.[24]

As the various studies discussed here demonstrate, network algorithms are very promising for uncovering the combinatorial relationships among multiple genes in the development and progression of complex diseases. These algorithms offer an excellent platform for life and medical scientists to use in extracting information from high-throughput data, which they can then use to filter out candidates for further, detailed experimental studies.

Such frameworks are useful in identifying drug targets, as well as in developing detailed dynamic models for disease mechanisms. Furthermore, the combinatorial patterns discovered by these algorithms serve as markers for classification that repeatedly demonstrate improved success compared with single gene markers in predicting disease progression[16,21] and response to therapy.[23] Using these algorithms to generate concise and effective sets of biomarkers will be particularly useful in personalized medicine. Although the discussion here is limited to integration of transcriptomic and interactomic data, algorithms that also utilize genomic and proteomic data could lead to the generation of more detailed disease models.[25-28] C

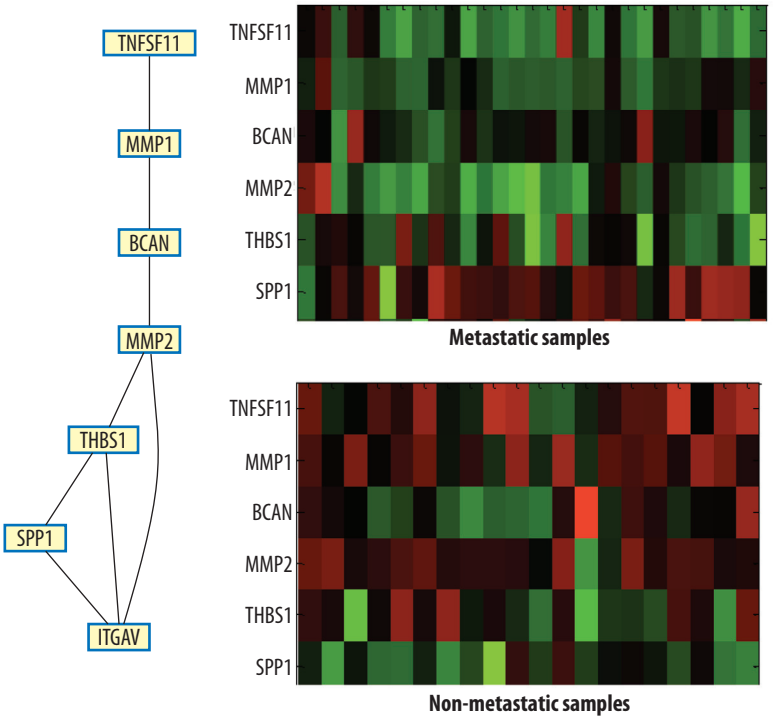

Figure 4. Algorithms for dysregulated subnetwork discovery drive experimental biology by identifying the patterns that lead to novel hypotheses. (a) A combinatorially dysregulated subnetwork identified by CRANE in the context of colorectal cancer metastasis and the interactions between this subnetwork and integrin alpha-chain V (ITGAV). (b) Heatmap of gene expression for these proteins in metastatic and nonmetastatic colon tumor samples; tone of green and red, respectively, show down- and up-regulation.

Acknowledgments

I thank Salim A. Chowdhury (CMU), Sinan Erten (CWRU), Vishal Patel (CWRU), Rod K. Nibbe (CWRU), Jill Barnholtz-Sloan (CWRU), and Mark R. Chance (CWRU) for their contribution to the projects described here and for many useful discussions. This work was supported in part by NSF CAREER Award CCF-0953195.

REFERENCES

1. W.G. Hatfield, S.P. Hung, and P. Baldi, "Differential Analysis of DNA Microarray Gene Expression Data," *Molecular Microbiology*, vol. 47, no. 4, 2003, pp. 871-877.
2. D.F. Ransohoff, "Rules of Evidence for Cancer Molecular-Marker Discovery and Validation," *Nature Reviews Cancer*, vol. 4, no. 4, 2004, pp. 309-314.
3. E E. Schadt, "Molecular Networks as Sensors and Drivers of Common Human Diseases," *Nature*, vol. 461(7261), 2009, pp. 218-223.
4. E.E. Eichler et al., "Missing Heritability and Strategies for Finding the Underlying Causes of Complex Disease," *Nature Reviews Genetics*, vol. 11, no. 6, 2010, pp. 446-450.
5. D.K. Slonim, "From Patterns to Pathways: Gene Expression Data Analysis Comes of Age," *Nature Genetics*, vol. 32 (suppl), 2002, pp. 502-508.
6. A. Subramanian et al., "Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles," *Proc. Nat'l Academy Sciences*, vol. 102, no. 43, 2005, pp. 15545-15550.
7. M. Koyutürk, "Algorithmic and Analytical Methods in Network Biology," *WIREs Systems Biology Medicine*, vol. 2, no. 3, 2010, pp. 277-292.
8. I. Lee et al., "A Probabilistic Functional Network of Yeast Genes," *Science*, vol. 306(5701), 2004, pp. 1555-1558.
9. T.S. Keshava et al, "Human Protein Reference Database–2009 Update," *Nucleic Acids Research*, vol. 37, D767-D772, 2009.
10. C. Stark et al., "The BioGRID Interaction Database: 2011 Update," *Nucleic Acids Research*, vol. 39 (Database issue), D698-D704, 2011.
11. A. Zelezniak et al., "Metabolic Network Topology Reveals Transcriptional Regulatory Signatures of Type 2 Diabetes," *PLoS Computational Biology*, vol. 6, no. 4, e1000729, 2010.
12. D.R. Rhodes and A.M. Chinnaiyan, "Integrative Analysis of the Cancer Transcriptome," *Nature Genetics*, vol. 37, S31-S37, 2005.
13. G. Gibson, "Decanalization and the Origin of Complex Disease," *Nature Reviews Genetics*, vol. 10, no. 2, 2009, pp. 134-140.
14. P. Dao et al., "Inferring Cancer Subnetwork Markers Using Density-Constrained Biclustering," *Bioinformatics*, vol. 26,

2010, pp. i625-i631.
15. T. Ideker et al., "Discovering Regulatory and Signalling Circuits in Molecular Interaction Networks," *Bioinformatics*, vol. 18 (suppl), 2001, pp. 233-240.
16. H.-Y. Chuang et al., "Network-Based Classification of Breast Cancer Metastasis," *Molecular Systems Biology*, vol. 3, 2007; www.ncbi.nlm.nih.gov/pmc/articles/PMC2063581.
17. S.A. Chowdhury and M. Koyutürk, "Identification of Coordinately Dysregulated Subnetworks in Complex Phenotypes, *Proc. Pacific Symp. Biocomputing*, 2010; http://psb.stanford.edu/psb-online/proceedings/psb10/chowdhury.pdf.
18. I. Ulitsky, R.M. Karp, and R. Shamir, "Detecting Disease-Specific Dysregulated Pathways via Analysis of Clinical Expression Profiles," *Proc. Ann. Int'l Conf. Research Computational Molecular Biology* (RECOMB 08), LNCS 4955, Springer, 2008, pp. 347-359.
19. D. Anastassiou, "Computational Analysis of the Synergy among Multiple Interacting Genes," *Molecular Systems Biology*, vol. 3, article no. 83, 2007; www.nature.com/msb/journal/v3/n1/full/msb4100124.html.
20. J. Watkinson et al., "Identification of Gene Interactions Associated with Disease from Gene Expression Data Using Synergy Networks," *BMC Systems Biology*, 2008; www.biomedcentral.com/1752-0509/2/10.
21. S.A. Chowdhury et al., "Subnetwork State Functions Define Dysregulated Subnetworks in Cancer," *J. Computational Biology*, vol. 18, no. 3, 2011, pp. 263-281.
22. N. Reinmuth et al., "Alphabeta3 Integrin Antagonist S247 Decreases Colon Cancer Metastasis and Angiogenesis and Improves Survival in Mice," *Cancer Research*, vol. 63, no. 9, 2003, pp.2079-2087.
23. P. Dao et al., "Optimally Discriminative Subnetwork Markers Predict Response to Chemotherapy," Bioinformatics, vol. 27, no. 13, 2011, pp. i205-i213.
24. J. Dutkowski and T. Ideker, "Protein Networks as Logic Functions in Development and Cancer," *PLoS Computational Biology*, vol. 7, no. 9, 2011, e1002180, 09.
25. Y.-A. Kim, S. Wuchty, and T. M. Przytycka, "Identifying Causal Genes and Dysregulated Pathways in Complex Diseases," *PLoS Computational Biology*, vol. 7, no. 3, 2011, e1001095.
26. R.K. Nibbe, M. Koyuturk, and M.R. Chance, "An Integrative -omics Approach to Identify Functional Subnetworks in Human Colorectal Cancer," *PLoS Computational Biology*, vol. 6, no. 1, 2010, e1000639.
27. E.E. Schadt, S.H. Friend, and D.A. Shaywitz, "A Network View of Disease and Compound Screening," *Nature Reviews Drug Discovery*, vol. 8, no. 4, 2009, pp. 286-295.
28. O. Vanunu et al., "Associating Genes and Protein Complexes with Disease via Network Propagation," *PLoS Computational Biology*, vol. 6, no. 1, 2010, e1000641.

Mehmet Koyutürk *is the T. & D. Schroeder Assistant Professor of Computer Science and Engineering at Case Western Reserve University. His research interests include applied algorithms, network biology, and computational genomics. Koyutürk received a PhD in computer science from Purdue University. He is a member of the IEEE Computer Society, ACM, and the International Society for Computational Biology. Contact him at mxk331@case.edu.*