

Network-based prediction for sources of transcriptional dysregulation using latent pathway identification analysis

Lisa Pham^a, Lisa Christadore^b, Scott Schaus^b, and Eric D. Kolaczyk^{c,1}

^aProgram in Bioinformatics, and Departments of ^bChemistry and ^cMathematics and Statistics, Boston University, Boston, MA 02215

Edited by Peter J. Bickel, University of California, Berkeley, CA, and approved June 21, 2011 (received for review January 19, 2011)

Understanding the systemic biological pathways and the key cellular mechanisms that dictate disease states, drug response, and altered cellular function poses a significant challenge. Although high-throughput measurement techniques, such as transcriptional profiling, give some insight into the altered state of a cell, they fall far short of providing by themselves a complete picture. Some improvement can be made by using enrichment-based methods to, for example, organize biological data of this sort into collections of dysregulated pathways. However, such methods arguably are still limited to primarily a transcriptional view of the cell. Augmenting these methods still further with networks and additional -omics data has been found to yield pathways that play more fundamental roles. We propose a previously undescribed method for identification of such pathways that takes a more direct approach to the problem than any published to date. Our method, called latent pathway identification analysis (LPIA), looks for statistically significant evidence of dysregulation in a network of pathways constructed in a manner that implicitly links pathways through their common function in the cell. We describe the LPIA methodology and illustrate its effectiveness through analysis of data on (i) metastatic cancer progression, (ii) drug treatment in human lung carcinoma cells, and (iii) diagnosis of type 2 diabetes. With these analyses, we show that LPIA can successfully identify pathways whose perturbations have latent influences on the transcriptionally altered genes.

centrality | microarray | pathway network

Understanding systemic biological pathways and the key cellular mechanisms that dictate disease states, drug response, and altered cellular function is a significant challenge. What is clear is that no single factor determines the response of a cell. A complex picture has emerged to include traditional genetics, epigenetics, signal transduction, and biochemical processes as well as other factors known and unknown. The result is a dynamic system of biological variables that culminate in an altered cellular state. The challenge is deciphering the factors that play key roles in determining the fate of a cell. High-throughput measurement techniques, such as transcriptional profiling, aid the process by providing a snapshot of the gene transcript levels. However, analyzing the sheer amount of information provided becomes a daunting task alone, which is exacerbated by the possible dependency of gene regulation. Organizing the data according to biological collections, such as gene ontologies, facilitates the analysis but fails to provide the necessary systemic understanding. Ultimately, it is necessary to view the cell as a complex collection of biochemical pathways with an inherent interdependency defined by the cell. By combining our knowledge of biochemical function and cellular pathways with global cellular measurements, it may be possible to create an integrated understanding of the biological factors that dictate the cellular state.

Computational methods exist that facilitate our understanding of altered cellular states. A popular method (now with various extensions as well), developed by Subramanian et al. (1), is gene set enrichment analysis (GSEA). GSEA measures the degree of

differential gene expression in a gene set across binary phenotypes. GSEA scores predefined gene sets according to how well the genes within the set will cluster at the top or bottom of a list of genes ranked by differential gene expression scores. There are also approaches that incorporate networks of protein-to-protein interactions (PPIs) into the task of finding transcriptionally altered gene sets (2, 3). This type of approach reports transcriptionally altered regions in the PPI network across binary conditions, effectively constraining the nature of the gene sets reported by information on the manner and extent to which their protein products interact. Such combinations of gene transcription and protein interaction data have been further augmented recently through the inclusion of biological function information (4, 5), bringing to bear additional “context” information. All these methods have been used successfully to find gene sets containing significant amounts of dysregulation, constrained by varying degrees of additional information on the biology of the cell.

Although these methods are capable of organizing the data into prioritized collections of dysregulated gene sets, they were not designed to provide a comprehensive picture of the underlying mechanism(s) by themselves. In fact, the cause of dysregulation is rarely transcriptionally altered but, rather, is typically attributable to the effects of latent pathways positioned to cause or significantly influence the cascade of transcriptionally altered pathways (6, 7). Accordingly, we choose to focus on the identification of these latent pathways. Finding these pathways should provide a greater understanding of the biological mechanisms underlying a given condition or perturbation. For instance, in the case of a disease, this type of discovery may help to uncover the pathogenic mechanisms of the disease; thus, the manipulation of these pathways can be explored for novel treatments.

In this paper, we introduce a computational method for identifying pathways as putative sources of transcriptional dysregulation, called latent pathway identification analysis (LPIA). Because individual pathways are part of a larger biological network of interactions, we use a network-based approach to find these aberrant pathways. In constructing our network, we use three distinct but complementary sources of biological data: (i) biological pathways [taken from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (8, 9)], (ii) biological functions [as compiled in the Gene Ontology (GO) database (10)], and (iii) gene transcriptional response (in the form of mRNA microarray expression profiles) for a pair of binary con-

Author contributions: S.S. and E.D.K. designed research; L.P. and L.C. performed research; L.P. and L.C. analyzed data; and L.P., S.S., and E.D.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE5434).

¹To whom correspondence should be addressed. E-mail: kolaczyk@bu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1100891108/-DCSupplemental.

ditions (e.g., case/control, normal/diseased). The network we construct is a network of pathways, in which links and their weights reflect both the extent to which incident pathways function in ways that produce similar biological outcomes and the differential transcriptional activity of genes coding for proteins common to those pathways. Pathways are then ranked using a measure of their centrality in this network.

Effectively, a pathway is identified as important in our methodology if the evidence in its corresponding differential transcription measurements, as interpreted in the context of its role in the cell biology, suggests that its proteins are strongly associated with the biological outcome of interest (e.g., disease vs. normal). In other words, our LPIA method is designed to identify pathways latent to observed dysregulatory transcription. We demonstrate that LPIA successfully does so in the context of (i) metastatic cancer progression, (ii) drug treatment in human lung carcinoma cells, and (iii) type 2 diabetes.

Results

We analyzed each of the three datasets using LPIA. To provide a point of comparison, we also analyzed each using GSEA (1), a standard method for finding dysregulated gene sets using transcriptional data. We comment on this comparison explicitly only in the first analysis below, because the relative performance of the two methods was qualitatively similar in the other two analyses. The full lists of ranked pathways reported by both methods may be found in [Datasets S1, S2, and S3](#) for the metastatic, geldanamycin, and diabetes analyses, respectively. A description of the implementation details for each method is included in [SI Text S2](#).

Prostate Cancer Metastasis. Our first example was chosen to illustrate the effectiveness of LPIA in identifying pathway dysregulation associated with disease, in part, through its comparison with GSEA. The factors that influence cancer progression can be characterized by cellular signal initiation and transduction, angiogenesis, proliferation, and morphological changes that promote cell adhesion [a review of these processes in metastatic cancer progression is included in the study by Bacac and Stamenkovic (11)]. The process is multifaceted, which is the result of a series of cellular processes operating in concert. We sought to identify the collection of processes differentiating metastatic cancer in the progression of prostate carcinomas.

At a false discovery rate (FDR) control level of 0.20, LPIA and GSEA identified one and four pathways, respectively, as statistically significant. Under stringent prioritization of significance, LPIA uniquely identified the wingless/integration-1 (Wnt) pathway, a causative signaling pathway in the progression of prostate cancers

(reviewed in 12), whereas GSEA identified cell cycle, focal adhesion, regulation of actin cytoskeleton, and T-cell receptor signaling pathways that, although important in the biology of prostate cancer, are more peripheral to the central signal that initiates the process of metastasis.

More generally, consider Table 1, which shows pathways implicated in metastatic cancer progression that were identified among the top 10 pathways by either LPIA or GSEA. All 10 pathways reported by LPIA are implicated in metastatic prostate cancer, whereas 7 of the top 10 pathways reported by GSEA have been shown to be of comparable significance. LPIA and GSEA shared an overlap of 5 pathways. Of the cancer-related pathways, GSEA uniquely identified JAK-STAT signaling and regulation of actin cytoskeleton, whereas LPIA uniquely identified MAPK, adherens junction, ErbB signaling, tight junction, and TGF- β signaling. We discuss below those pathways identified by only LPIA and their implication in metastatic cancer.

The latent pathway analysis of the dataset (Table 1) uniquely identified five key pathways in the metastatic progression of prostate cancer beginning with cellular signaling processes responsible for cell proliferation. Dysregulation of ErbB signaling corresponds with androgen-independent cell proliferation in prostate cancers (reviewed in 13). Specific to prostate cancer, early stage cancer cells depend on androgen for growth and survival but advanced prostate cancer cells become androgen-independent (14, reviewed in 15). Moreover, three of the four families of ErbB receptors have been shown to enhance proliferation of prostate cancer cells via distinct mechanisms of action (reviewed in 13). Coincident with androgen-independent progression of prostate cancers is the activation of the MAPK pathway and higher levels of the necessary kinases (16). LPIA uniquely identified TGF- β signaling. In metastatic prostate cancer, TGF- β also serves as a tumor promoter (reviewed in 17). Tu et al. (18) demonstrated that dysregulation of TGF- β signaling in prostate cancer plays a causal role in promoting tumor metastasis. Catenin, mediated by Wnt signaling, serves two cellular roles: the first in conjunction with T-cell factors (TCF) in the transcription control of proteins key to cell proliferation and the second as a part of a complex with E-cadherin, an extracellular binding matrix that promotes intercellular adhesion (19, 20). Adherens junction is the process by which E-cadherin mediates cell adhesion in tissue required for morphological changes leading to changes in the phenotype (11, 19, 20). Lastly, tight junction mediates cell motility, and research [these mechanisms are reviewed in the study by Martin and Jiang (21)] shows a significant role for tight junctions in maintaining cell-to-cell integrity, such that perturbations can lead to invasion and metastasis of cancer cells.

Table 1. Pathways fundamental to metastatic prostate cancer that were identified in the top 10 ranking pathways by either LPIA or GSEA

Pathway	LPIA	GSEA	Mechanisms involved in metastatic prostate cancer
ErbB signaling	✓		Increased ErbB expression correlates with androgen-independent cellular proliferation (13–15)
TGF- β signaling	✓		TGF- β is a tumor promoter in metastatic prostate cancers (17)
Adherens junction	✓		Cell dissociation via loss of adherens junction; E-cadherin/ β -catenin complex and cell-to-cell adhesion (11, 19, 20)
Tight junction	✓		Process that mediates cell motility in metastatic cancers (21)
MAPK signaling	✓		Mediates signaling pathways involved with cell growth and proliferation in metastatic cancers (11, 16)
Wnt signaling	✓	✓	Cell-to-cell signaling pathway that mediates β -catenin-induced cell proliferation (11, 19, 20)
Focal adhesion	✓	✓	Positive regulator of prostate cancer cell adhesion, migration, and invasion (22)
Cell adhesion molecules	✓	✓	Extracellular proteins that promote migration and mediate cell-to-cell adhesion (23)
p53 signaling	✓	✓	Dysregulated p53 correlates with prostate cancer pathogenesis (24)
Cell cycle	✓	✓	Genes involved in progression of the cell through the cell cycle
Regulation of actin cytoskeleton		✓	Key processes that promote the migratory phenotype of metastatic cancer cells (11)
Jak/Stat signaling		✓	Growth factor-mediated cellular proliferation pathway (25)

Furthermore, changes in the tight junctions may lead to uncontrolled cell proliferation and detachment/invasion of cancer cells (21). Uniquely identified by LPIA, this collection of dysregulated pathways promotes cell differentiation in metastatic cancer.

Heat Shock Protein 90 Inhibition via Geldanamycin Treatments. Our second example was chosen to illustrate the effectiveness of LPIA in identifying pathway dysregulation resulting from drug treatment. We selected an antiproliferative compound that alters cellular signaling by disrupting a signal transduction pathway: geldanamycin, an ansamycin natural product that inhibits the biochemical activity of heat shock protein 90 (Hsp90) (26, 27). Hsp90 is responsible for facilitating normal protein folding; intracellular disposition; and proteolytic turnover of regulators of cell growth, differentiation, and survival (reviewed in 28). Of particular interest is the interaction of Hsp90 with growth factor receptor proteins and the MAP kinases. We compared geldanamycin-exposed human lung carcinoma cells with untreated cells (*Methods*), using a 50% inhibitory concentration (IC₅₀) for a duration of 24 h and 48 h as well as a 20% inhibitory concentration (IC₂₀) at 48 h. The rationale for selecting the times for cell harvesting and drug concentrations was twofold. First, the cellular response in a collection of cells that are asynchronous with respect to cell cycle could potentially produce heterogeneity in the transcriptional response. Given the typical cell cycle time of 20 h for A549 cells, we chose times for cell collection that were on that time scale. Second, there were two responses that could occur as the result of an antiproliferative agent, the primary response to inhibition of Hsp90 and a secondary response as the result of the apoptotic transcriptional program. Using a concentration lower than the IC₅₀ could potentially minimize the response determined by the induction of apoptosis.

We first approached the analysis of the resulting data as an exercise in drug target deconvolution, irrespective of any previous knowledge or a priori identification of the biological target of geldanamycin. In so doing, the drug treatment needed to be evaluated in light of the following: cell culture heterogeneity and the time necessary to progress through the cell cycle. A549 cells were simply treated with geldanamycin. An asynchronous collection of cells then continued to progress through the cell cycle, halting at the G2 stage, as a result of geldanamycin treatment. Characteristically at 24 h, the highest KEGG pathway ranked by LPIA was cell cycle; by this time, point cells begin to halt at the G2 phase (29, 30). The IC₂₀ analysis at 48 h appears not to yield a significantly high enough alteration at the transcriptional level in the most significant Hsp90-mediated pathways to identify those most affected by LPIA. In contrast, in the analysis at 48 h of the IC₅₀ treatment, ErbB was ranked at number 1. ErbB is a key pathway targeted by geldanamycin in lung cancer caused by inhibition of Hsp90 and resulting in the initial stages of apoptosis (31). We thus show that LPIA is able to identify the cause of the altered cellular state, namely, changes in the competency of the ErbB signaling pathway as the result of Hsp90 inhibition, cellular death.

Taken together, these results illustrate how LPIA is capable of divining key pathways of dysregulation as the result of drug perturbation. However, because the actual target of geldanamycin, Hsp90, does not appear in a KEGG pathway of cellular functions or processes, we next performed an analysis of Hsp90 client proteins within the KEGG perturbed by treatment with geldanamycin to assess the global cellular effects.

As the target of geldanamycin, Hsp90 and its client proteins have been well documented, easily facilitating analysis of results and consequences of treatment (reviewed in 28). Here, we will use the extent to which pathways are enriched with proteins directly interacting with Hsp90 as a measure of their importance [i.e., being indicative of pathways that are substantially perturbed on a proteomic level (via chemical perturbations) rather than

a transcriptional level alone]. Interacting proteins and protein complexes of Hsp90AA1 and Hsp90AB1 that have been experimentally validated either in vivo or in vitro were compiled from the Human Protein Reference Database (release 9) (32). We will assess the performance of LPIA, with respect to determining the effect of geldanamycin on chaperone function at a global cellular level, by evaluating how successful it is at identifying pathways enriched with Hsp90 interactors, from both aggregate and rank-ordered perspectives.

For our aggregate assessment, we first collapsed the top m pathways called by LPIA into a single gene set, for $m = 5, 10, 15,$ and 20 pathways. Each gene set was then compared with the full set of Hsp90 interactors, computing an enrichment P value according to a hypergeometrical distribution. The results show that in all three experiments, the collapsed gene set reported by LPIA is highly enriched with Hsp90 interactors (P values $<10^{-4}$ for all cases but IC₅₀ at 24 h with $m = 5$, which was 0.19). Thus, as a whole, the collection of pathways identified by LPIA is strongly linked to the set of Hsp90 interactors. We then looked at Hsp90 enrichment with respect to individual pathway ranking. We computed a P value for each KEGG pathway, summarizing its enrichment with Hsp90 interactors, and ranked pathways according to their P values. This ranked list, acting as a gold standard, was then compared, in turn, with each of the ranked lists produced by LPIA for all three treatments. Comparisons between the top m pathways of the gold standard and those of the method being assessed were made using both Euclidean distance of rank vectors and number of true-positive findings as a function of m . The results are summarized in Fig. 1. Our analysis illustrates that as we increase the time to 48 h from 24 h for both concentrations, LPIA shows a considerable improvement, particularly at the IC₅₀ concentration. Therefore, in the latter treatments, LPIA provides a better indicator of key pathways of dysregulation as the result of chemical perturbation.

Type 2 Diabetes. The third dataset we chose to evaluate was an analysis of the transcriptional differences of skeletal muscle tissue among patients with type 2 diabetes, impaired glucose tolerance, and normal glucose tolerance, as reported by Gallagher et al. (33). We performed the analysis on the two populations possessing the strongest binary comparison: gene transcription of skeletal muscle cells of people diagnosed with type 2 diabetes (45 subjects) compared with that of skeletal muscle cells exhibiting normal glucose tolerance (47 subjects). Under this comparison, LPIA uniquely identified the oxidative phosphorylation pathway after multiple test correction with an FDR level of 0.20. Dysregulation of genes in the mitochondrial oxidative phosphorylation pathway is characteristic of both human diabetic skeletal muscle and liver samples of patients with type 2 diabetes (34, 35). Additionally, muscle biopsies in patients with type 2 diabetes showed decreased activity of mitochondrial oxidative enzymes (36), and analysis of healthy patients revealed that increased levels of intramyocellular lipid content, an indicator of insulin resistance, were caused by inherited mitochondrial oxidative phosphorylation defects (37). Takamura et al. (38) further explored the correlation between obesity and diabetes, showing that oxidative gene expression significantly correlated with insulin resistance and reactive oxygen species generation in liver specimens. Oxidative phosphorylation was not identified at the transcriptional level in the same dataset by Gallagher et al. (33). These results demonstrate that LPIA may be implemented as a diagnostic tool capable of identifying disease characteristics, such as insulin resistance caused by dysfunctional oxidative phosphorylation in patients with type 2 diabetes.

Discussion

The analyses performed in these three biological contexts highlight the ability of LPIA to provide effective biological insight into

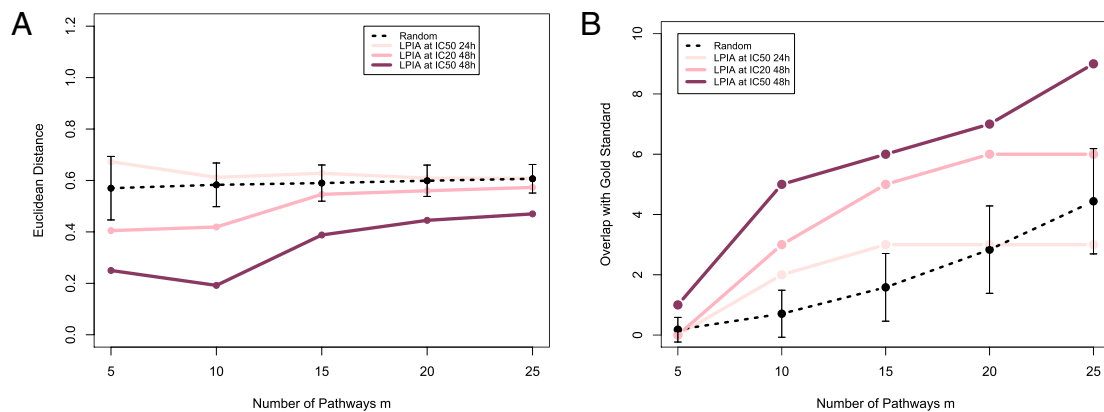


Fig. 1. (A) Euclidean distances of the top m pathways assessed by LPIA from the gold standard list. (B) Number of true-positive findings as a function of the top m pathways in the gold standard list. The dotted lines in A and B indicate the mean expected by random chance (via simulating random ranked lists) with 1-SD error bars.

alterations of cellular function. The first example demonstrates how use of LPIA resolves the multifactorial process of cancer metastasis in the identification of cellular signaling pathways key to initiation, cell growth, and propagation. The KEGG pathways identified are instrumental as an aggregate leading to the progression of the tumor. The second example illustrates how the analysis may be used for drug target deconvolution strategies. Although not providing a direct protein target, the cellular processes governing the antiproliferative effects were identified. Finally, the type 2 diabetes analysis demonstrates how LPIA may be used as a diagnostic tool in a clinical context, identifying a clearly dysregulated biological pathway not readily resolved by using transcriptional measurements alone. Although each biological scenario uses transcriptional alterations as an analytical technique, each one is unique in the type of biological questions needing to be addressed.

The challenge in using transcriptional changes as an analytical tool of cellular function is to discern comprehensive systemic alterations of biological states. One would like to see the signal initiation underlying the resultant transcriptional alterations. To date, enrichment methods and biological gene set categorization have proven useful and more efficient at identifying transcriptionally dysregulated gene collections than simple clustering methods. However, the underlying biology and the effectors of the altered disease state proved to be illusive without the guidance of cellular systems and pathways to provide the necessary insight into linked cellular processes and understanding of the centrality of signal initiation. By incorporating these concepts, our latent pathway analysis is capable of augmenting existing technologies for analysis by resolving the aberrant pathways that give rise to the aberrant phenotype, recognizing that the disease state is more comprehensive than simply a transcriptionally altered pathway (39). In so doing, the utility of our method to characterize disease progression, to facilitate clinical diagnosis, and to provide systemic evaluation of drug-induced cellular alterations as an enabling technology in biomedical research is manifold.

From a mathematical perspective, the nature of the problem we address in this paper is not unlike that of “deconvolution” in image processing, a similarity that has been noted by others in this area [e.g., “drug target deconvolution” (40)]. In the image processing version of this problem, an image, say f , is of interest but one has available only blurred and noisy measurements, say $y = Kf + e$. Although denoising y can be relatively straightforward, it only leaves one with an estimate of the blurred image, Kf . To recover f itself, the effect of the blurring operator, K , must be inverted. However, even in the ideal case where K is known, this inversion can be ill-posed and the recovery of f can be se-

verely degraded by the corresponding inflation of the noise, e . When K is unknown, the degradation can be arbitrarily worse.

In the context of this paper, the transcriptional measurements are analogous to y , the underlying biological state, f , and the cascade of biological processes that produce the transcriptional responses as a result of an altered biological state, K . Although it is now standard to report transcriptionally dysregulated gene sets as part of the analysis of transcriptional measurements (and these may, in fact, show high levels of statistical significance), this is arguably analogous only to denoising y through the mechanism of averaging over gene sets. In contrast, LPIA is more analogous to inverting the effect of K (i.e., the effect of the underlying biological processes, although at the expected cost of an inflation in noise (i.e., note the relative magnitude of P values for LPIA and GSEA in [Datasets S1, S2, and S3](#)).

Further progress on this problem may potentially be had from a model-based statistical analog of LPIA, based on formal principles of statistical modeling and inference. For such, latent factor models are a natural tool. Such models have been used extensively in the bioinformatics literature for the analysis of microarray data. However, to date, they have been implemented almost uniformly in the spirit of so-called “exploratory factor analysis,” which lacks the incorporation of known biological structure to the extent that we have used it in LPIA. More consistent with LPIA would be something in spirit of so-called “confirmatory factor analysis,” incorporating, for example, our pathway/function bipartite network into a model for the covariance of latent factors. We are currently exploring this avenue of research. A preliminary version of these models is described in [SI Text, S3](#), where we use them as the basis for generating data in a simulation study aimed at establishing some initial *in silico* notions of the sensitivity of LPIA to the strength of the sources of dysregulation.

Methods

LPIA. The algorithm at the heart of our proposed method of LPIA is depicted in Fig. 2 and consists of four steps. In steps 1–3, a network graph representation of pathways is constructed that reflects the inherent interdependency among pathways as defined by the cell, through their participation in common biological functions but augmented with measurements of transcriptional dysregulation specific to a pair of binary conditions of interest (e.g., disease/normal). We then calculate, in step 4, how central each pathway is in this network. Finally, the significance of these centrality scores is assessed using a bootstrap-based randomization method, running the above algorithm repeatedly for various bootstrap resamplings of the data (not shown in Fig. 1). Those pathways with significant centralities are reported as potentially important in their latent influence on the cascade of transcriptionally altered pathways.

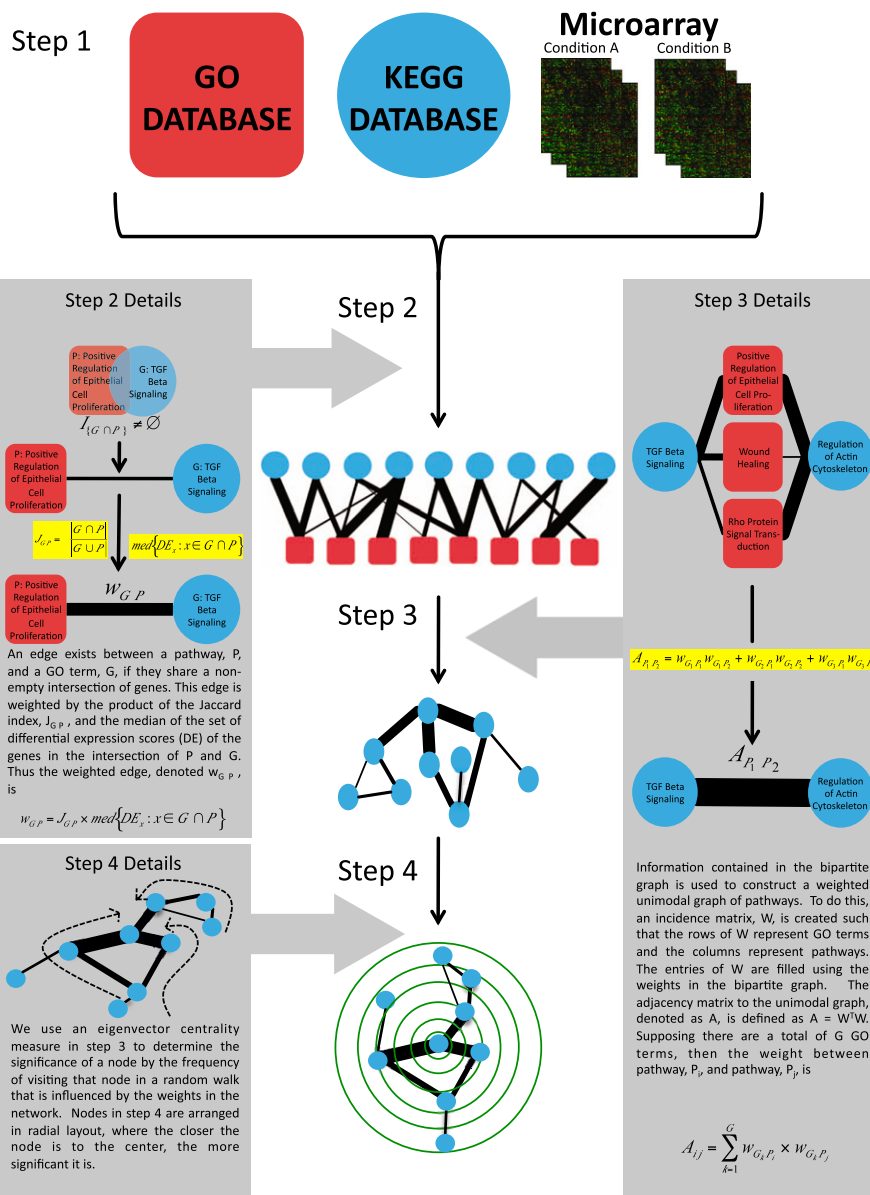


Fig. 2. Schematic illustration of the proposed LPIA method.

More specifically, in step 1, three sources of biological data are assembled: KEGG pathways, GO biological processes, and microarray data from two conditions. In step 2, as an intermediate step to constructing our network of pathways, we first construct a bipartite network, where the two sets of nodes represent (i) KEGG pathways and (ii) GO functions, respectively. An edge exists between a pathway, P, and a GO term, G, in this bipartite network if the intersection of P and G (as gene sets) is nonempty. The edge is weighted by the product of two terms: (i) the relative overlap of G and P (as measured by their Jaccard similarity) and (ii) the differential activity of genes common to both G and P [summarized as the median differential expression (DE) over these genes]. Our network of pathways then results from the projection of this two-mode (i.e., bipartite) network, in step 3, onto the corresponding one-mode network formed by the KEGG pathway node set alone. As a result, two pathways are linked in this network if and only if they share at least one biological process. The corresponding edge weights will be larger for pairs of pathways P_i and P_j that relate more to GO term G_i , whose contributions to both pathways are large. The centrality we calculate in step 4 is an eigenvector centrality, essentially summarizing for each pathway P, the frequency with which it would be visited by a random walk on the network, with movement between neighboring nodes being determined by the relative magnitude of the corresponding edge weights.

A more detailed description of the algorithm may be found in *SI Text, S1*. Implementation details relevant to the analyses reported below in this paper may be found in *SI Text, S2*. A software implementation of the algorithm is available at <http://math.bu.edu/LPIA/>.

Datasets for Analysis. To illustrate the effectiveness of LPIA, we used three datasets: (i) comparing localized prostate cancer with prostate cancer metastasis, (ii) comparing geldanamycin-treated lung cancer carcinoma cells with untreated lung cancer carcinoma cells, and (iii) comparing normal glucose tolerance with impaired glucose tolerance in type 2 diabetes. The first dataset was a prostate cancer analysis published by Varambally et al. (41) (accession no. GSE3325). The data analyzed consisted of four replicates of metastatic cancer and five replicates of clinically localized cancer, all of which used Affymetrix U133 Plus version 2.0 microarrays. The second dataset was obtained in our laboratory. We compared lung cancer carcinoma cells treated with an Hsp90 inhibitor, geldanamycin, at various time points and concentrations with mock control groups [GEO database (accession no. GSE26525)]. We describe the methods of this experiment in *SI Text, S2*. The third dataset, obtained from Gallagher et al. (33) [GEO database (accession no. GSE18732)], investigates the transcriptional differences of skeletal muscle tissue among patients with type 2 diabetes, patients who exhibit impaired glucose toler-

ance, and a patient population with normal glucose tolerance. We focused our attention on the binary comparison with the strongest disparity, comparing gene transcription of skeletal muscle of people diagnosed with type 2 diabetes (45 subjects) with that of people with normal glucose tolerance (47 subjects).

1. Subramanian A, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102: 15545–15550.
2. Ideker T, Ozier O, Schwikowski B, Siegel AF (2002) Discovering regulatory and signaling circuits in molecular interaction networks. *Bioinformatics* 18(Suppl 1):S233–S240.
3. Liu M, et al. (2007) Network-based analysis of affected biological processes in type 2 diabetes models. *PLoS Genet* 3:e96.
4. Huttenhower C, et al. (2009) Exploring the human genome with functional maps. *Genome Res* 19:1093–1106.
5. Nariai N, Kolaczyk ED, Kasif S (2007) Probabilistic protein function prediction from heterogeneous genome-wide data. *PLoS ONE* 2:e337.
6. Lucas J, et al. (2006) Sparse statistical modelling in gene expression genomics. *Bayesian Inference for Gene Expression and Proteomics*, eds Do KA, Muller P, Vannucci M (Cambridge Univ Press, Cambridge, UK), pp 155–176.
7. Vogelstein B, Kinzler KW (2004) Cancer genes and the pathways they control. *Nat Med* 10:789–799.
8. Kanehisa M, et al. (2006) From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res* 34(Database issue):D354–D357.
9. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30.
10. Ashburner M (2000) Gene Ontology: Tool for the unification of biology. The Gene Ontology Consortium *Nat Genet* 25:25–29.
11. Bacac M, Stamenkovic I (2008) Metastatic cancer cell. *Annu Rev Pathol* 3:221–247.
12. Yardy GW, Brewster SF (2005) Wnt signalling and prostate cancer. *Prostate Cancer Prostatic Dis* 8:119–126.
13. El Sheikh SS, Domin J, Abel P, Stamp G, Lalani N (2003) Androgen-independent prostate cancer: Potential role of androgen and ErbB receptor signal transduction crosstalk. *Neoplasia* 5:99–109.
14. Shah RB, Ghosh D, Elder JT (2006) Epidermal growth factor receptor (ErbB1) expression in prostate cancer progression: Correlation with androgen independence. *Prostate* 66:1437–1444.
15. Feldman BJ, Feldman D (2001) The development of androgen-independent prostate cancer. *Nat Rev Cancer* 1:34–45.
16. Gioeli D, Mandell JW, Petroni GR, Frierson HF, Jr., Weber MJ (1999) Activation of mitogen-activated protein kinase associated with prostate cancer progression. *Cancer Res* 59:279–284.
17. Bierie B, Moses HL (2006) Tumour microenvironment: TGFbeta: The molecular Jekyll and Hyde of cancer. *Nat Rev Cancer* 6:506–520.
18. Tu WH, et al. (2003) The loss of TGF-beta signaling promotes prostate cancer metastasis. *Neoplasia* 5:267–277.
19. Saha B, et al. (2008) Overexpression of E-cadherin and beta-catenin proteins in metastatic prostate cancer cells in bone. *Prostate* 68:78–84.
20. Jaggi M, et al. (2005) Aberrant expression of E-cadherin and beta-catenin in human prostate cancer. *Urol Oncol* 23:402–406.
21. Martin TA, Jiang WG (2009) Loss of tight junction barrier function and its role in cancer metastasis. *Biochim Biophys Acta* 1788:872–891.
22. Sakamoto S, McCann RO, Dhir R, Kyprianou N (2010) Talin1 promotes tumor invasion and metastasis via focal adhesion signaling and anoikis resistance. *Cancer Res* 70: 1885–1895.
23. Cooper CR, Pienta KJ (2000) Cell adhesion and chemotaxis in prostate cancer metastasis to bone: A minireview. *Prostate Cancer Prostatic Dis* 3:6–12.
24. Zhou Z, et al. (2006) Synergy of p53 and Rb deficiency in a conditional mouse model for metastatic prostate cancer. *Cancer Res* 66:7889–7898.
25. Ni Z, Lou W, Leman ES, Gao AC (2000) Inhibition of constitutively activated Stat3 signaling pathway suppresses growth of prostate cancer cells. *Cancer Res* 60: 1225–1228.
26. Neckers L, Schulte TW, Mimnaugh E (1999) Geldanamycin as a potential anti-cancer agent: Its molecular target and biochemical activity. *Invest New Drugs* 17:361–373.
27. Grenert JP, et al. (1997) The amino-terminal domain of heat shock protein 90 (hsp90) that binds geldanamycin is an ATP/ADP switch domain that regulates hsp90 conformation. *J Biol Chem* 272:23843–23850.
28. Whitesell L, Lindquist SL (2005) HSP90 and the chaperoning of cancer. *Nat Rev Cancer* 5:761–772.
29. McIlwrath AJ, Brunton VG, Brown R (1996) Cell-cycle arrest and p53 accumulation induced by geldanamycin in human ovarian tumour cells. *Cancer Chemother Pharmacol* 37:423–428.
30. Kim HR, Lee CH, Choi YH, Kang HS, Kim HD (1999) Geldanamycin induces cell cycle arrest in K562 erythroleukemic cells. *IUBMB Life* 48:425–428.
31. Plate JM, Iyengar RM, Sutton P, Bonomi P (2007) Nrdp1 and ErbB3 expression in non-small cell lung cancer lines: C7-05. Proffered Paper Abstracts: Session C7: Tumor and Cell Biology C7-05. *J Thorac Oncol* 2:S381.
32. Keshava Prasad TS, et al. (2009) Human protein reference database—2009 update. *Nucleic Acids Res* 37(Database issue):D767–D772.
33. Gallagher IJ, et al. (2010) Integration of microRNA changes in vivo identifies novel molecular features of muscle insulin resistance in type 2 diabetes. *Genome Med* 2:9.
34. Mootha VK, et al. (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34: 267–273.
35. Misu H, et al. (2007) Genes involved in oxidative phosphorylation are coordinately upregulated with fasting hyperglycaemia in livers of patients with type 2 diabetes. *Diabetologia* 50:268–277.
36. Vondra K, et al. (1977) Enzyme activities in quadriceps femoris muscle of obese diabetic male patients. *Diabetologia* 13:527–529.
37. Petersen KF, Dufour S, Befroy D, Garcia R, Shulman GI (2004) Impaired mitochondrial activity in the insulin-resistant offspring of patients with type 2 diabetes. *N Engl J Med* 350:664–671.
38. Takamura T, et al. (2008) Obesity upregulates genes involved in oxidative phosphorylation in livers of diabetic patients. *Obesity* 16:2601–2609.
39. Feng Y, Mitchison TJ, Bender A, Young DW, Tallarico JA (2009) Multi-parameter phenotypic profiling: Using cellular effects to characterize small-molecule compounds. *Nat Rev Drug Discov* 8:567–578.
40. Terstappen GC, Schlüpen C, Raggiacchi R, Gaviraghi G (2007) Target deconvolution strategies in drug discovery. *Nat Rev Drug Discov* 6:891–903.
41. Varambally S, et al. (2005) Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell* 8:393–406.