

## Bladder Cancer Specific Pathway Interaction Networks

Qinxin Pan<sup>1</sup>, Ting Hu<sup>1,3</sup>, Angeline S. Andrew<sup>2</sup>, Margaret R. Karagas<sup>2</sup> and Jason H. Moore<sup>1,2,3</sup>

<sup>1</sup>Department of Genetics, Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA

<sup>2</sup>Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA

<sup>3</sup>Institute for Quantitative Biomedical Sciences, Dartmouth College, Hanover, NH 03755, USA

jason.h.moore@dartmouth.edu

### Abstract

With the development of high-throughput technologies, monitoring biological systems comprehensively has become feasible and affordable. However, the transition from high-throughput data to the underlying biology of various phenotypes remains challenging. Pathway analysis identifies biological processes that are associated with a particular phenotype, which provides insights into the underlying biological mechanisms. Therefore, pathway analysis has become a popular tool for analyzing high-throughput data. Most existing pathway analysis methods are based on a simple assumption that pathways act in isolation whereas they cooperate with each other in a complex manner. In this study, we focus on pathway interactions that are associated with bladder cancer risk. We identify disease-specific pathway-pathway interactions based on SNP-SNP interactions and gene-gene co-expression relationships. By analyzing the structure of pathway interaction networks, we highlight the “central” pathways that should be further studied.

### Introduction

Techniques such as high-throughput sequencing and gene/protein profiling have transformed biological research by enabling comprehensive monitoring of a biological system (Khatri et al., 2012). Conventional analyses of high-throughput data usually test the association of individual genes/proteins with a given phenotype. Although successfully applied in many studies, this approach fails to provide insights into the underlying biological mechanisms of the phenotype being studied. As an alternative approach, pathway analysis highlights the risk associated biological processes. The mechanisms of pathway associations can be used for developing strategies to diagnose, treat and prevent complex diseases (Ramanan et al., 2012), which makes high-throughput datasets more often viewed as a foundation to discover associated pathways (Hirschhorn, 2009).

In pathway analyses, gene sets corresponding to biological pathways are tested for significant associations with a phenotype. Multiple methods have been developed and among them pathway enrichment approach is the most popular one. Most pathway-enrichment-approach studies follow two categories: the threshold-based framework and the

rank-based framework. Threshold-based approaches usually statistically evaluate the fraction of genes in a particular pathway among all the significant markers (Boyle et al., 2004). Rank-based approaches rank all markers based on their significances and then look for pathways that have better rankings than the overall distribution (Subramanian et al., 2005). Although successfully applied in many studies, enrichment approaches depend on single-marker statistics and treat each gene independently. In reality, biological systems are driven by complex biomolecular interactions instead of individual genes (Schadt et al., 2009). Thus, methods that take biomolecular interactions into account remain needed.

Most pathway analysis approaches assume that each pathway is independent of the others, which could be problematic (Khatri et al., 2012). Pathways do not operate in isolation. Instead, they cooperate and work together as a unit. A recent study of yeast showed that rewiring of genetic interactions in response to DNA damage are more likely to occur among pairs of genes that belong to two different biological processes (Bandyopadhyay et al., 2010). Although whether interactions are more likely to occur within the same pathway or across different pathways in human is unknown, the significance of pathway interactions is in-negligible. There are several possible ways two pathways can interact: sharing components (Lu et al., 2007), components physically interacting with counterparts from the other pathway (Lu et al., 2007; Guo and Wang, 2009), components relating with components of the other pathway via transcription (Guo and Wang, 2009), etc. The different interacting ways can be reflected in different manners including pathway overlapping, direct or indirect protein protein interactions (PPIs), co-regulation of genes etc. So to fully understand pathway interactions, integration of knowledges from different levels is required.

Several studies have been done to address pathway interactions. Tong et al. carried out genetic screening and used synthetic lethality of two mutations to indicate interaction between two pathways (Tong et al., 2004). Later, Kelley et al. used such genetic interactions to link “pathways”, which is defined as sets of densely connected pro-

teins in the protein-protein-interaction (PPI) network (Kelley and Ideker, 2005). Meanwhile, overlapping in genes, proteins or metabolite contents between pathways has also been used to gain insights in possible relationships between pathways (Isserlin et al., 2010). Lastly, a function based approach is to find possible pathway crosstalks by looking at protein interactions between pathways (Li et al., 2008). Most previous methods highlight pathway crosstalks under general conditions. The crosstalks identified are not associated with any particular phenotype. In this study, we make use of the disease case-control data and search for the disease specific pathway interactions only, which distinguish our work from previous ones the most.

Meanwhile, network science has emerged as a useful tool for modeling biological interactions and dependencies (Ideker and Sharan, 2008). Compared to analysis of groups of distinct objects, networks provide more information on the relatedness and interconnectivity of them, which makes network a suitable framework for investigating pathway interactions.

In this study, we focus on pathway interactions that are associated with bladder cancer risk. We infer pathway interaction networks from two different levels: SNP-SNP interactions and gene-gene co-expression relationships. By analyzing the structure of pathway interaction networks, we identify essential pathways that hold great potentials for future studies.

## Methods

### SNP interaction network

A recent study characterized the space of pairwise interactions in a population-based bladder cancer association study (Hu et al., 2011). In their SNP interaction network, each vertex corresponds to a single nucleotide polymorphism (SNP). An edge linking a pair of vertices corresponds to an interaction between two SNPs. Weights assigned to each SNP and each pair of SNPs quantify how much of the disease status the corresponding SNP and SNP pair can explain. The significance of the SNP interaction network is not limited to single main or interaction effects. Instead, it describes the overall significance of the global interaction structure, which makes this approach systematic. To identify bladder cancer specific pathway interactions, we adopt the SNP interaction network from this previous study and the network is shown in Figure 1.

### Dataset

The microarray dataset used in this study is publicly available from Gene Expression Omnibus (GEO) website (dataset ID: GDS1479) and more details about this dataset can be obtained at Dyrskjot et al. (2003). Briefly, the original study profile about 22,000 genes to analyze bladder biopsies of superficial transitional cell carcinomas with or without surrounding carcinoma in situ (CIS) lesions and muscle

invasive carcinomas (mTCC). To match disease stages of individuals in the bladder cancer SNP dataset, we only use two out of the five groups in the original study. One group contains 15 tumor biopsies from superficial transitional cell carcinoma (sTCC) without surrounding CIS, which serves as cases. The other group contains 9 biopsies of normal bladder mucosa from patients without a bladder cancer history, which serves as controls. The cRNA from different samples are hybridized to Affymetrix U133A GeneChips. After data processing as described by Dyrskjot et al. (2003), there are about 22,000 genes in the dataset.

### Differential co-expression network

Genes with similar expression patterns may form complexes, pathways, or participate in regulatory and signaling circuits (Eisen et al., 1998; Ideker et al., 2002; Huang et al., 2007). Therefore, gene co-expression networks, which describe the pairwise relations among gene expression profiles, have become a popular tool for microarray analysis. A gene co-expression network is an undirected graph, where the vertices correspond to genes, and edges between genes represent significant co-expression relationships (Stuart et al., 2003). The weight of an edge can be computed using different correlation measures and an edge is included in the network if its weight pass a certain pre-specified threshold.

Gene co-expression networks have been successfully applied in many studies mostly to identify functional gene modules (Stuart et al., 2003; Presson et al., 2008; Weston et al., 2008). However, to study a particular disease, it is the difference between cases and controls that provides the most information about the underlying mechanism. In other words, rather than asking ‘what parts of the system are the most abundant or dominant’, we should ask ‘What parts of the system are most distinctive between different conditions’ (Ideker and Krogan, 2012). Therefore, to gain insights into the transition from healthy individuals to bladder cancer patients, we adapt the framework of the differential network previously proposed (Bandyopadhyay et al., 2010). Instead of constructing two static co-expression networks, one for cases and one for controls, we build differential co-expression networks to describe the changes of pairwise relations among genes from controls to cases. In this way, co-expressions present in both conditions are downplayed or removed from the differential network. The co-expressions that reflect the changes from controls to cases are distinguished from those that support the housekeeping functions.

Specifically, we filter the probes in the microarray dataset and only 308 of them, which also exist in the SNP dataset, are considered in this section. We compute Spearman’s correlation for all  $\binom{308}{2} = 47,278$  pairs of transcripts separately in cases ( $C_{case}$ ) and controls ( $C_{controls}$ ). Pairwise differential co-expression is calculated as  $C_{case} - C_{control}$  and assigned to the corresponding edge as its weight. The negative

differential co-expression network is constructed to include only edges with significant negative weights and the positive differential co-expression network only include edges with significant positive weights. In other words, the negative differential co-expression network describes the gene pairs which lose correlations from controls to cases, while the positive differential co-expression network describes those which gain correlations.

**Pathway interaction network**

We construct pathway interaction network  $G_1$  from SNP interaction network,  $G_2^-$  from negative differential co-expression network and  $G_2^+$  from positive differential co-expression network. In all pathway interaction networks, a vertex represents a pathway. Two pathways are connected if there are more edges than by chance between the two corresponding pathways in the underlying biomarker network.

For example, to construct pathway interaction network  $G_1$ , we use permutation test to determine whether edges are more likely to occur between two pathways than by chance in the SNP interaction network. First, we annotate all the vertices in the SNP interaction network with their functions using canonical pathway annotations from MsigDB collection 2. All SNPs in this dataset are within coding region, so a SNP-to-gene-to-pathway mapping is straight forward. To investigate the enrichment of pathway-pathway interactions, we count the frequency of edges in the SNP interaction network between two particular pathways. Note that pathways with more SNPs might display more edges in the SNP interaction network. To control the bias introduced by pathway sizes, we use permutation test to generate a null distribution. Specifically, we randomly shuffle the SNP-pathway(s) annotations for 1,000 times and assess the frequency of edges between two particular pathways in the original SNP interaction network at each time. The  $P$ -value of a pathway-pathway interaction enrichment is computed as the fraction of corresponding edge frequencies on permuted data which are no smaller than that of the real data. Similarly, pathway interaction networks  $G_2^-$  and  $G_2^+$  are generated based on the negative and positive differential co-expression networks accordingly. In all three cases, a pathway pair is included in the pathway interaction network if the  $P$ -value of the pathway-pathway interaction enrichment is smaller than 0.001.

**Essential vertices identification**

To identify key vertices in the networks, we use bottleneck vertices to survey independent regulation as it has been shown that bottleneck vertices are regulated in a condition-dependent manner in biological networks (Yu et al., 2007). We define bottleneck vertices as vertices with the 10 highest betweenness centrality scores. The betweenness centrality score is based on the number of shortest paths that cross a given vertex, and thus reflects how embedded (i.e. central) a vertex is in the network (Freeman, 1977).

Hubs, highly connected vertices in other words, play an integral role in maintaining network integrity and in mass information transfer. Therefore, we also identify hubs of the networks. In this study, we define hub vertices as vertices with the 10 highest degree scores.

Both hub and bottleneck vertices have been associated with functional essentiality (Yu et al., 2007; Zotenko et al., 2008). We are interested in their ability to influence the propagation of signals across the network and their importance to maintain the integrity of the network. Particularly, we define hub-bottleneck vertices as the overlapping vertices between hub vertices and bottleneck vertices.

**Results**

**Pathway interaction network derived from SNP interaction network**

We study whether SNP-SNP interactions are more likely to occur between certain pathway pairs or not. As described previously, we construct pathway interaction network  $G_1$  by only including pathway pairs which have more edges between them in the SNP interaction network than by chance ( $P < 0.001$ ). Figure 2 shows the structure of pathway interaction network  $G_1$ .  $G_1$  has 386 vertices and 635 edges. There are two connected components in the network and the vertices cluster into several communities. Moreover, there is no self-loop in  $G_1$ , which indicates that SNP-SNP interactions mostly happen between different pathways instead of within the same pathway .

A heavy-tail degree distribution of  $G_1$  indicates the existence of hub vertices. The hub-bottleneck vertices (pathways) are reported in Table 1. Among all the hub vertices, KEGG\_FOCAL\_ADHESION is connected to 269 neighbors and displays the highest betweenness centrality.

Pathway	Betweenness Centrality	Degree
KEGG_FOCAL_ADHESION	0.93	268
BIOCARTA_VITCB_PATHWAY	0.11	32
KEGG_STEROID_HORMONE_BIOSYNTHESIS	0.09	13
REACTOME_BASIGIN_INTERACTIONS	0.07	53
BIOCARTA_INTEGRIN_PATHWAY	0.06	51
BIOCARTA_NO1_PATHWAY	0.06	51
KEGG_VIRAL_MYOCARDITIS	0.05	44
REACTOME_PACKAGING_OF_TELOMERE_ENDS	0.02	10
REACTOME_TELOMERE_MAINTENANCE	0.02	10

Table 1: Hub-bottleneck vertices (pathways) in pathway interaction network  $G_1$  derived from SNP interaction network. Pathways are ranked in descending order of betweenness centrality.

**Differential co-expression network**

To study the bladder cancer specific co-expression patterns, we construct differential co-expression networks from the microarray dataset. To match the SNP dataset, only 308 transcripts that also exist in the SNP dataset are used in this section. The changes of Spearman correlation between all

$\binom{308}{2} = 47,278$  pairs of transcripts from controls to cases are shown in Figure 3. The differential co-expression correlation,  $(C_{\text{case}} - C_{\text{control}})$ , follows a normal distribution with  $mean \pm stdv = 0.0006 \pm 0.51214$ . The 95% confidence interval is  $[-1.003, 1.004]$ . Transcript pairs with  $C_{\text{case}} - C_{\text{control}} < -1.003$  are included in the negative differential co-expression network, which means that there are co-expressions in controls but not in cases. Similarly, transcript pairs with  $C_{\text{case}} - C_{\text{control}} > 1.004$  are included in the positive differential co-expression network, which indicates that there are co-expressions in cases but not in controls.

The negative differential co-expression network has 297 vertices and 1,114 edges. Among two connected components, the largest one has 293 vertices. Similarly, the positive differential co-expression network possesses 287 vertices and 1,114 edges. There are no isolated islands in the network. Both differential co-expression networks display a heavy-tail degree distribution, which indicates the existence of hub genes.

To identify the key players for maintaining network integrity and signal propagation across the network, we identify the hub-bottleneck vertices in both negative and positive differential co-expression networks (Table 2). Interestingly, gene GATA3, ERBB2 and HFE are identified in both differential co-expression networks.

Gene	Betweenness Centrality	Degree
GATA3	0.21	67
MSH2	0.13	47
ERBB2	0.07	31
HFE	0.04	23
HADHA	0.04	21
LIG1	0.03	26

Gene	Betweenness Centrality	Degree
HFE	0.13	37
FZD7	0.09	26
FOXC1	0.06	53
ERBB2	0.06	27
GATA3	0.06	28

Table 2: Hub-bottleneck vertices (genes) in negative (top) and positive (bottom) differential co-expression networks. Genes are ranked in descending order of betweenness centrality.

### Pathway interaction network derived from differential co-expression network

We investigate whether edges in the differential co-expression networks are more likely to occur between particular pathway pairs than by chance. As described previously, we construct pathway interaction network  $G_2^-$  by only including pathway pairs which have more edges between them in the negative differential network than by chance ( $P < 0.001$ ). Similarly, pathway interaction network  $G_2^+$  is obtained from the positive differential co-expression network.

As shown in Figure 4, pathway interaction network  $G_2^-$  has 78 vertices and 100 edges. There are three connected

components and the entire network divide into several communities. The hub-bottleneck vertices (pathways) are reported in Table 3. Meanwhile, pathway interaction network  $G_2^+$  has 70 vertices and 83 edges. The network has six connected components (Figure 5). Surprisingly, no common edge is shared by  $G_2^-$  and  $G_2^+$ . This means that when comparing cases with controls, the gaining of co-expression and the loss of co-expression happen between different pathway pairs.

Although  $G_2^-$  and  $G_2^+$  do not share any common edge, the essential vertices in the two networks overlap with differences (Table 3). Two pathways, BIOCARTA\_GATA3\_PATHWAY and KEGG\_MELANOGENESIS, are identified as hub-bottleneck vertices in both networks.

It is also an interesting thing that there is no self-loop in  $G_2^-$  (Figure 4) and only two self-loops are observed in  $G_2^+$  (Figure 5). In other words, no pathway possesses more edges within itself than by chance in the negative differential co-expression network. Only two pathways, REACTOME\_GPCR\_LIGAND\_BINDING and SA\_CASPASE\_CASCADE, have more edges within themselves than by chance in the positive differential co-expression network. This indicates that differential co-expressions seem to happen mostly between different pathways rather than within the same pathway.

Pathways	Betweenness Centrality	Degree
KEGG_MISMATCH_REPAIR	0.72	20
BIOCARTA_GATA3_PATHWAY	0.58	24
KEGG_ERBB_SIGNALING_PATHWAY	0.55	12
KEGG_MELANOGENESIS	0.34	13
REACTOME_TRANSCRIPTION	0.19	5
REACTOME_RNA_POLYMERASE_I_III_AND_MITOCHONDRIAL_TRANSCRIPTION	0.19	5
REACTOME_SEMA4D_INDUCED_CELL_MIGRATION_AND_GROWTH_CONE_COLLAPSE	0.15	9

Pathways	Betweenness Centrality	Degree
BIOCARTA_GATA3_PATHWAY	1.0	13
KEGG_MELANOGENESIS	0.90	31
REACTOME_RNA_POLYMERASE_I_PROMOTER_OPENING	0.83	5
REACTOME_CLASS_B2_SECRETIN_FAMILY_RECEPTORS	0.08	8
REACTOME_CLASS_A1_RHODOPSIN_LIKE_RECEPTORS	0.06	5
REACTOME_PEPTIDE_LIGAND_BINDING_RECEPTORS	0.06	5

Table 3: Hub-bottleneck vertices (pathways) in pathway interaction networks  $G_2^-$  (top) and  $G_2^+$  (bottom). The two pathway interaction networks are derived from negative (top) and positive (bottom) differential co-expression networks accordingly. Pathways are ranked in descending order of betweenness centrality.

## Discussion

We have identified pathway-pathway interactions that are associated with bladder cancer risk. Specifically, we investigate pathway interaction at two levels: genetic inter-

actions and co-expression relations. We construct one pathway interaction network from the SNP interaction network and two from the differential co-expression networks. There are some limitations to our approach that are worth highlighting. First, there are multiple other ways pathways can interact with each other, i.e. through sharing common components. Examining pathway relationships more thoroughly at different levels would be helpful. Second, co-expression relationships are transferable. In other words, if two genes are both correlated with a third gene, a correlation will be observed between those two genes. Using partial correlation instead of correlation itself might be helpful to identify direct gene-gene co-expression relationships (de la Fuente et al., 2004).

Despite these limitations, our analyses have helped to identify the pathway interactions that are associated with bladder cancer risk and highlight the importance of pathway interactions. In our study, we observe several interesting aspects and discuss them as follows.

First, we characterize the differential co-expression relationships between cases and controls. The hub-bottleneck genes are identified in the differential co-expression networks. Three genes, GATA3, ERBB2 and HFE, are recognized in both positive and negative differential co-expression networks. Gene GATA3 encodes trans-acting T-cell-specific transcription factor GATA-3, which is an important regulator of T-cell development and plays an important role in endothelial dysplasia. Previously, GATA3 has been suggested as a marker of urothelial differentiation (Higgins et al., 2007; Miyamoto et al., 2012). The fact that GATA3 appears as a hub-bottleneck gene in both positive and negative differential co-expression networks (Table 2) suggests that GATA3 could be actively turning on or off its neighbor genes' expression, which might be highly associated with bladder cancer risk. Whether GATA3 is causing the expression level changes of its neighbor genes and how its neighbor genes contribute to the disease risk should be further studied. Gene ERBB2 encodes a member of the epidermal growth factor (EGF) receptor family of receptor tyrosine kinases. Tumor-specific overexpression of ERBB receptors or their isoforms has been reported in a previous study (Junttila et al., 2003). Whether and how ERBB2 influences its neighbor genes' expressions in the context of bladder cancer risk should be explored. Gene HFE encodes a membrane protein HLA-h, which binds to transferrin receptor (TFR) and reduces its affinity for iron-loaded transferrin (Feder et al., 1998). Whether and how HFE is associated with bladder cancer risk remains unknown. The central role of HFE in the differential co-expression networks might reflect its association with bladder cancer risk.

Second, traditional pathway analyses focus on identifying pathways that are enriched for significant biomarkers. We took a different route and look for pathway pairs that are enriched for between-pathway SNP-SNP interac-

tions or between-pathway differential co-expressions. Consequently, we identify pathways that are "central" in the whole system instead of pathways that are individually associated with the disease risk. We run Gene Set Enrichment Analysis (Subramanian et al., 2005) on both the SNP dataset and the microarray dataset. Most of the pathways reported in Table 1 and 3 are not enriched in the significant biomarkers. However, existing knowledge indicates that some of them could be highly involved in the underlying mechanisms of bladder cancer. For instance, two telomere related pathways are identified as hub-bottleneck vertices in  $G_1$  (Table 1). It is well known that telomere dysfunction or loss can cause sister-chromatid fusions that is associated with oncogene amplification (Campbell et al., 2010; Murnane, 2012).

Third, it is interesting that all networks constructed in this study possess a heavy-tail degree distribution. In other words, most vertices in the networks have only few neighbors whereas a few vertices have many neighbors. This structure makes the network robust to random removal of vertices and vigorous to external perturbations (Barabasi and Bonabeau, 2003; Wang and Chen, 2002). Scale-free structures have been observed in various biological networks (Jeong et al., 2001; Li et al., 2004). It is interesting that we observe a similar structure at the pathway level.

Fourth, we find that the pathway interaction network  $G_1$  does not share any common edge with  $G_2^+$  or  $G_2^-$ . The hub-bottleneck vertices in  $G_1$  are not as "central" in  $G_2^-$  or  $G_2^+$ . Although genetic interactions are very different with co-expression patterns, this result is still surprising. This means that the interactions across pathways are of distinguishing patterns at different levels.

Last, we find that both SNP-SNP interactions and differential co-expressions mostly happen between different pathways rather than within the same pathway. This result suggests that SNP interactions and co-expression patterns in known pathways stay stable across cases and controls. It is the SNP interactions and co-expression patterns between these pathways that are reprogrammed across different disease conditions. Previous studies in yeast have shown that static genetic interactions are enriched within known pathways (Kelley and Ideker, 2005), whereas differential genetic interactions are much more likely to occur among pairs of genes connecting two different pathways than among pairs of genes within the same pathway (Bandyopadhyay et al., 2010). Although co-expressions do not necessarily indicate genetic interactions, it is still encouraging to observe similar results in human SNP and microarray data in our study. Also, this observation further emphasizes the important role of pathway interactions in disease association studies, which highlights the significance of our work.

In summary, we construct bladder cancer specific pathway interaction networks from both SNP-SNP interactions and gene-gene co-expression patterns. Our study highlights key pathway interactions that should be further investigated

and emphasizes the importance of disease-specific pathway interactions.

### Acknowledgements

This work was supported by NIH grants R01-LM009012, R01-LM010098, and R01-AI59694.

### References

- Bandyopadhyay, S., Mehta, M., Kuo, D., and etc. (2010). Rewiring of genetic networks in response to dna damage. *Science*, 330(6009):1385–1389.
- Barabasi, A. L. and Bonabeau, E. (2003). Scale-free networks. *Scientific American*, 288:60–69.
- Boyle, E. I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J. M., and Sherlock, G. (2004). Go::termfinder—open source software for accessing gene ontology information and finding significantly enriched gene ontology terms associated with a list of genes. *Bioinformatics*, 20(18):3710–3715.
- Campbell, P. J., Yachida, S., Mudie, L. J., and etc. (2010). The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature*, 467(7319):1109–1113.
- de la Fuente, A., Bing, N., Hoeschele, I., and Mendes, P. (2004). Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics*, 20(18):3565–3574.
- Dyrskjot, L., Thykjaer, T., Kruhoffer, M., Jensen, J. L., Marcussen, N., Hamilton-Dutoit, S., Wolf, H., and Orntoft, T. F. (2003). Identifying distinct classes of bladder carcinoma using microarrays. *Nat Genet*, 33(1):90–96.
- Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*, 95(25):14863–14868.
- Feder, J. N., Penny, D. M., Irrinki, A., Lee, V. K., Lebron, J. A., Watson, N., Tsuchihashi, Z., Sigal, E., Bjorkman, P. J., and Schatzman, R. C. (1998). The hemochromatosis gene product complexes with the transferrin receptor and lowers its affinity for ligand binding. *Proc Natl Acad Sci U S A*, 95(4):1472–1477.
- Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41.
- Guo, X. and Wang, X.-F. (2009). Signaling cross-talk between tgfbeta/bmp and other pathways. *Cell Res*, 19(1):71–88.
- Higgins, J. P. T., Kaygusuz, G., Wang, L., and etc. (2007). Placental s100 (s100p) and gata3: markers for transitional epithelium and urothelial carcinoma discovered by complementary dna microarray. *Am J Surg Pathol*, 31(5):673–680.
- Hirschhorn, J. N. (2009). Genomewide association studies—illuminating biologic pathways. *N Engl J Med*, 360(17):1699–1701.
- Hu, T., Sinnott-Armstrong, N. A., Kiralis, J. W., Andrew, A. S., Karagas, M. R., and Moore, J. H. (2011). Characterizing genetic interactions in human disease association studies using statistical epistasis networks. *BMC Bioinformatics*, 12:364.
- Huang, Y., Li, H., Hu, H., Yan, X., Waterman, M. S., Huang, H., and Zhou, X. J. (2007). Systematic discovery of functional modules and context-specific functional annotation of human genome. *Bioinformatics*, 23(13):i222–9.
- Ideker, T. and Krogan, N. J. (2012). Differential network biology. *Mol Syst Biol*, 8:565.
- Ideker, T., Ozier, O., Schwikowski, B., and Siegel, A. F. (2002). Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, 18 Suppl 1:S233–40.
- Ideker, T. and Sharan, R. (2008). Protein networks in disease. *Genome Res*, 18(4):644–652.
- Isserlin, R., Merico, D., Alikhani-Koupaei, R., Gramolini, A., Bader, G. D., and Emili, A. (2010). Pathway analysis of dilated cardiomyopathy using global proteomic profiling and enrichment maps. *Proteomics*, 10(6):1316–1327.
- Jeong, H., Mason, S. P., Barabasi, A. L., and Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, 411(6833):41–42.
- Junttila, T. T., Laato, M., Vahlberg, T., and etc. (2003). Identification of patients with transitional cell carcinoma of the bladder overexpressing erbb2, erbb3, or specific erbb4 isoforms: real-time reverse transcription-pcr analysis in estimation of erbb receptor status from cancer patients. *Clin Cancer Res*, 9(14):5346–5357.
- Kelley, R. and Ideker, T. (2005). Systematic interpretation of genetic interactions using protein networks. *Nat Biotechnol*, 23(5):561–566.
- Khatri, P., Sirota, M., and Butte, A. J. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol*, 8(2):e1002375.
- Li, S., Armstrong, C. M., Bertin, N., and etc. (2004). A map of the interactome network of the metazoan c. elegans. *Science*, 303(5657):540–543.
- Li, Y., Agarwal, P., and Rajagopalan, D. (2008). A global pathway crosstalk network. *Bioinformatics*, 24(12):1442–1447.
- Lu, L. J., Sboner, A., Huang, Y. J., Lu, H. X., Gianoulis, T. A., Yip, K. Y., Kim, P. M., Montelione, G. T., and Gerstein, M. B. (2007). Comparing classical pathways and modern networks: towards the development of an edge ontology. *Trends Biochem Sci*, 32(7):320–331.
- Miyamoto, H., Izumi, K., Yao, J. L., Li, Y., Yang, Q., McMahon, L. A., Gonzalez-Roibon, N., Hicks, D. G., Tacha, D., and Netto, G. J. (2012). Gata binding protein 3 is down-regulated in bladder cancer yet strong expression is an independent predictor of poor prognosis in invasive tumor. *Hum Pathol*, 43(11):2033–2040.
- Murnane, J. P. (2012). Telomere dysfunction and chromosome instability. *Mutat Res*, 730(1-2):28–36.
- Presson, A. P., Sobel, E. M., Papp, J. C., Suarez, C. J., Whistler, T., Rajeevan, M. S., Vernon, S. D., and Horvath, S. (2008). Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome. *BMC Syst Biol*, 2:95.

Ramanan, V. K., Shen, L., Moore, J. H., and Saykin, A. J. (2012). Pathway analysis of genomic data: concepts, methods, and prospects for future development. *Trends Genet*, 28(7):323–332.

Schadt, E. E., Friend, S. H., and Shaywitz, D. A. (2009). A network view of disease and compound screening. *Nat Rev Drug Discov*, 8(4):286–295.

Stuart, J. M., Segal, E., Koller, D., and Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science*, 302(5643):249–255.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102(43):15545–15550.

Tong, A. H. Y., Lesage, G., Bader, G. D., and etc. (2004). Global mapping of the yeast genetic interaction network. *Science*, 303(5659):808–813.

Wang, X. F. and Chen, G. (2002). Synchronization in scale-free dynamical networks: robustness and fragility. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, 49(1):54–62.

Weston, D. J., Gunter, L. E., Rogers, A., and Wulschleger, S. D. (2008). Connecting genes, coexpression modules, and molecular signatures to environmental stress phenotypes in plants. *BMC Syst Biol*, 2:16.

Yu, H., Kim, P. M., Sprecher, E., Trifonov, V., and Gerstein, M. (2007). The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol*, 3(4):e59.

Zotenko, E., Mestre, J., O’Leary, D. P., and Przytycka, T. M. (2008). Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. *PLoS Comput Biol*, 4(8):e1000140.

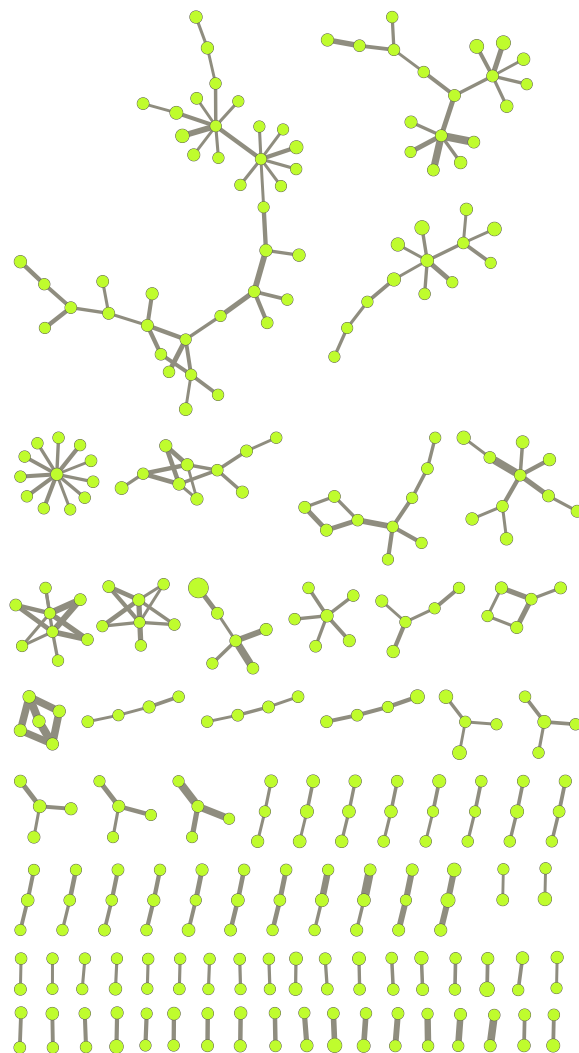


Figure 1: SNP interaction network. A vertex represents a SNP and an edge represents the interaction between the two vertices it connects. The weight of an edge reflects the pairwise interaction strength. There are 319 vertices and 255 edges in the SNP interaction network. The 319 SNPs cover 185 genes. The width of an edge and the size of a vertex are proportional to their weights. More details about this network are available at Hu et al. (2011)

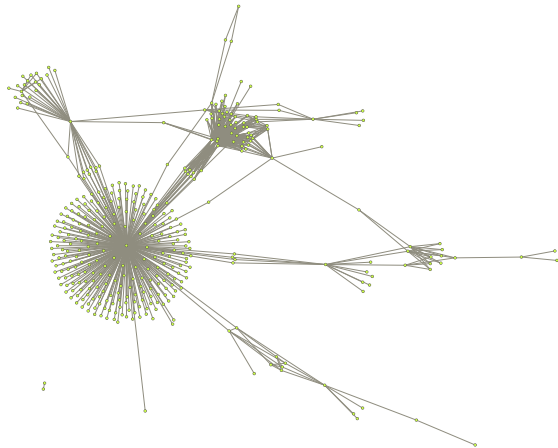


Figure 2: Pathway interaction network  $G_1$  revealed by SNP interaction network. A vertex represents a pathway and an edge exists if the number of SNP interactions between the two corresponding pathways is significantly larger than by chance ( $P < 0.001$ ). There are 386 vertices and 635 edges in the network. The largest connected components has 384 vertices. The network shows a heavy tail degree distribution.

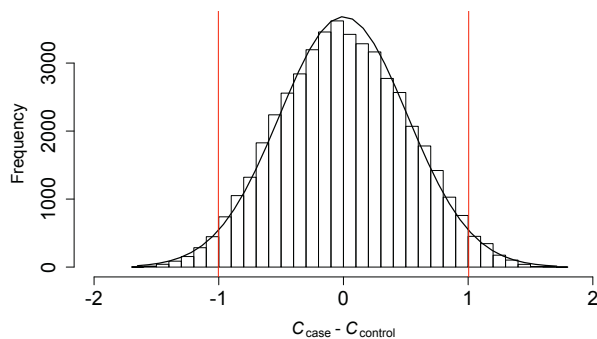


Figure 3: Frequency distribution of pairwise differential co-expression,  $C_{case} - C_{control}$ . Spearman correlation for all 47,278 pairs of genes are calculated for cases ( $C_{case}$ ) and controls ( $C_{control}$ ) separately and the difference ( $C_{case} - C_{control}$ ) are presented.  $C_{case} - C_{control}$  ranges from -1.652 to 1.711 and follows a normal distribution. Red lines represent the 95% confidence interval ( $C_{case} - C_{control} = -1.003$  and 1.004 respectively). There are 1,011 pairs of transcripts on the left 2.5% tail and 1,114 on the right 2.5% tail.

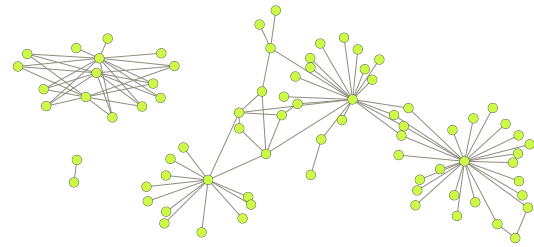


Figure 4: Pathway interaction network  $G_2^-$  revealed by negative differential co-expression network. A vertex represents a pathway and an edge exists if differential co-expression are more likely to occur between the two pathways it connects than by chance in the negative differential co-expression network. The network  $G_2^-$  has 78 vertices and 100 edges.

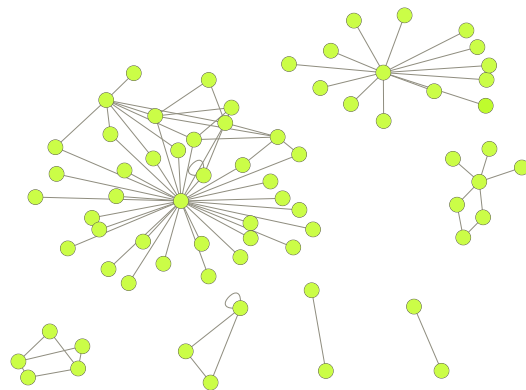


Figure 5: Pathway interaction network  $G_2^+$  revealed by positive differential co-expression network. A vertex represents a pathway. An edge indicates that there are more edges between the two pathways it connects than by chance in the positive differential co-expression network. There are 70 vertices and 83 edges in the network. The vertices fall into six connected components.