



Published in final edited form as:

*Hum Genet.* 2012 October ; 131(10): 1677–1686. doi:10.1007/s00439-012-1198-7.

## Integration of biological networks and pathways with genetic association studies

**Yan V. Sun**

Department of Epidemiology, Emory University, Rollins School of Public Health, 1518 Clifton Road NE #3049, Atlanta 30322, GA, USA, yvsun@emory.edu

Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, GA, USA

### Abstract

Millions of genetic variants have been assessed for their effects on the trait of interest in genome-wide association studies (GWAS). The complex traits are affected by a set of inter-related genes. However, the typical GWAS only examine the association of a single genetic variant at a time. The individual effects of a complex trait are usually small, and the simple sum of these individual effects may not reflect the holistic effect of the genetic system. High-throughput methods enable genomic studies to produce a large amount of data to expand the knowledge base of the biological systems. Biological networks and pathways are built to represent the functional or physical connectivity among genes. Integrated with GWAS data, the network- and pathway-based methods complement the approach of single genetic variant analysis, and may improve the power to identify trait-associated genes. Taking advantage of the biological knowledge, these approaches are valuable to interpret the functional role of the genetic variants, and to further understand the molecular mechanism influencing the traits. The network- and pathway-based methods have demonstrated their utilities, and will be increasingly important to address a number of challenges facing the mainstream GWAS.

### Introduction

Genome-wide association studies (GWAS) have successfully identified over a thousand genetic associations with human traits and diseases. Although millions of single nucleotide polymorphisms (SNPs) are genotyped or imputed, GWAS typically investigate the genetic effect of a single SNP at a time. Most genetic associations have a small effect and require a rather large sample size (e.g., 10,000's) across several cohorts to be robustly identified. On the other hand, most common diseases are multigenic traits which involve a group of genes functioning at various stages of disease development. Given the complex genetic architecture and synergistic effects among these genes, the holistic effect of a gene network or a pathway is expected to have a larger effect than the sum of individual effect of each gene. In addition, it is usually challenging to interpret the genetic associations for their functional connection with the trait only based on the annotation of a single gene. Therefore, network- and pathway-based methods have been developed to boost the power to identify the candidate genes and to provide functional links to bridge the knowledge gap between the genetic variants and the phenotypes. Combining with the data and results from GWAS, these approaches can assess whether a group of genes with related functions are jointly associated with a trait of interest and generate specific hypothesis for follow-up experimental studies.

© Springer-Verlag 2012

Correspondence to: Yan V. Sun.

**Conflict of interest** The author declares that he has no competing interests.

In the following sections, I discuss the basic concepts of biological networks and pathways, and review the network- and pathway-based analyses for genetic association studies.

## Basic concepts of biological networks and pathways

Network is a collection of vertices (i.e., nodes) that are joined together in pairs by edges (Newman 2010). Figure 1a is an example with ten vertices (A to J) and 45 edges connecting all pairs of vertices. Networks have been widely used in many fields of biology to represent the relationships between biological entities. In molecular biology and genetics, networks are often used to represent the functional connections among large (e.g., protein, DNA) and small molecules (e.g., carbonates, lipids) within cells and organisms. Several types of biological networks, such as protein–protein interaction (PPI) networks, metabolic networks and gene regulatory networks, have been constructed to illustrate the complex relationship within the biological system. Rather than databases which capture a large number of individual studies, these networks present the accumulated knowledge as an interconnected illustration of all similar types of biological relationship. Some networks that contain the edges representing relationship with a specified direction are called directed network (Fig. 1b). An arrowed edge represents the direction between two vertices. For example,  $B \rightarrow A$  can represent a transcriptional regulation mechanism in which the protein product of gene B regulates the expression level of gene A. In another scenario, the relationship between two biological entities is always bidirectional (e.g., protein binding). A network representing these bidirectional relationships is called undirected network (Fig. 1a). A PPI network is an undirected network. The knowledge base of these biological networks is quickly expanding in the last decade with the new high-throughput technologies available in genomic studies. These networks produced by high-throughput experiments provide extensive functional information of genes.

A biological pathway, such as a metabolic pathway or a signaling pathway, involves a series of biochemical and molecular steps to achieve a specific function or to produce a certain product (e.g., a metabolite or a protein). Although “genes” are usually emphasized in biological pathways, they can also consist of other types of molecules in the cellular system, including metabolites and other inorganic molecules. The graphic presentation of a biological pathway is a special type of network that has a start and an end, and is usually directed. Several large databases have been created to curate these pathways in eukaryotic and prokaryotic organisms (Caspi et al. 2010; Kanehisa and Goto 2000). Pathguide (<http://www.pathguide.org>) contains over 300 resources of biological pathways. Most pathway-based genetic association studies treat these biological pathways as a set of related genes which jointly perform a biological function, ignoring the specific and usually directed relationship among the genes. For these studies, the pathway information is treated as a gene set rather than a typical network. The biological networks discussed in this article represent a global view of all connectivity among a large number of genes. Rather than defining a function-centered pathway, a biological network focuses on the partnership among molecules, regardless of the actual molecular functions they perform. These networks mostly are composed of homogeneous vertices—usually “genes” which may represent expression levels of genes (e.g., co-expression network) or proteins as gene products (e.g., PPI network). Because these networks are not limited to a certain functional outcome, they do not carry an obvious boundary. A biological network is defined by all connected vertices within a knowledge space. In other words, the collection of all known relationship among the “genes” determine the network. Due to the advancement of high-throughput technology, our understanding of these biological relationships is steadily growing. As a result, many of these biological networks are constantly expanding and changing, and are less stable than the well-curated pathways. Gene Ontology (GO) is another widely adopted source of functional annotation (<http://www.geneontology.org>). GO is a controlled vocabulary of gene

and gene product attributes including three primary categories, cellular component (CC), molecular function (MF) and biological process (BP). GO is designed to be species neutral. The relationship among GO terms are structured as a directed acyclic graph, with each term having defined relationships to one or more other terms in the same or sometimes other domains.

To conduct the pathway- or network-based analysis, a key step is to map the genotyped genetic variants to a known gene. Although the mapping of SNPs to genes may affect the analysis results, there is no consensus definition for such annotation. The membership is usually defined by the chromosomal distance between the variant and the gene boundary (i.e., transcriptional start site—TSS, and transcriptional end site—TES). However, the suggested sizes of the flanking region vary from study to study, ranging from 10 Kb to 2.5 Mb (Askland et al. 2012; Lee et al. 2011; Schaid et al. 2012; Wang et al. 2007, 2010; Weng et al. 2011). Although most known *cis*-regulatory elements are located within 5 kb around the genic region, there are known enhancers which can be recruited to the promoter but are located far from the transcriptional start site (measured by the base-pair distance). The various sizes reflect the different choices of being inclusive or exclusive of assigning SNPs to genes. Although a large size of flanking region can include more intergenic SNPs to the gene-based analysis, it can also include more irrelevant SNPs and may further complicate the interpretation in some gene-dense regions, when a single SNP can be mapped to multiple genes. Interestingly, Lee et al. (2011) recently showed that using the flanking region between 0 and 250 kb from the genes did not significantly change the boosting from the network. The intuitive choice of chromosomal distance to define a genic region can be limited when other functional data are available. For instance, both *cis*- and *trans*-eQTLs are associated with gene expression levels. These associations with gene functions can be more informative than the chromosomal distance to the genes. Adopting the new functional genomics annotations for mapping SNPs to genes can potentially improve the pathway and network analyses of GWAS data. A number of statistical methods have been developed to conduct gene-level associations with disease traits by collapsing multiple rare and common variants across a genic region. The statistical framework of these collapsing methods are not limited to gene-based association and can be expand to multiple genes in a biological pathway or a larger network. Given that these methods have been thoroughly reviewed recently (Bansal et al. 2010; Dering et al. 2011; Sun et al. 2011), they are not discussed in this article.

## Gene set enrichment methods

Gene set enrichment methods Gene set enrichment analysis (GSEA) was initially proposed to identify the enriched gene sets without preselecting significant genes in transcriptomic studies (Subramanian et al. 2005). When the set of genes are closely related and their expression levels are associated with the outcome, the statistical power of identifying the association can be potentially increased by borrowing the signals across multiple genes in a set. In practice, the gene set can be defined by known pathways, GO terms or any group of genes sharing similar or related biological characteristics. The original GSEA method implemented a variation of a Kolmogorov–Smirnov statistic to calculate an enrichment score (ES) for each pre-specified gene set. Considering a collection of gene sets  $S_1, S_2, \dots, S_k$ , we compute test statistic  $G_j$  for each gene in our data (total of  $N$  genes). Let  $G_k = (G_1, G_2, \dots, G_m)$  be the gene scores for the  $m$  genes in gene set  $S_k$ . The genes are ranked based on the correlation between their expression and the outcome. GSEA computes a gene set score  $ES_k(G_k)$  for each gene set  $S_k$ . ES reflects the degree to which a set  $S_k$  is overrepresented at the extremes (top or bottom) of the entire ranked list (Subramanian et al. 2005). The GSEA method performs permutations of the sample labels to compute the null distribution of the

ES statistic on each permuted data set. Finally, the significance level is corrected for multiple testing using the false discovery rate of the list of significant gene sets.

Although GSEA is an effective method to study gene sets, and has been widely adopted in the analyses of transcriptomics, other approaches of gene set enrichment have been proposed to address the limitations of the original GSEA (Efron and Tibshirani 2007; Newton et al. 2007). For example, the choice of Kolmogorov–Smirnov statistics in GSEA is reasonable but not necessary. Efron and Tibshirani developed a new procedure based on the “maxmean” statistic that has superior power characteristics (Efron and Tibshirani 2007). They showed that the new statistic has better power than GSEA and increased reproducibility. A “restandardization” procedure, which combines the permutation of sample labels and row randomization of gene sets, was also proposed to generate an unbiased null distribution of the “maxmean” statistic (Efron and Tibshirani 2007). To evaluate the performance of these gene set enrichment methods, Abatangelo et al. (2009) compared four of these approaches including GSEA using both simulated and real data. The performance of the naïve Fisher’s exact test was markedly worse than the other three methods. Although differences in the results from these enrichment methods are observed, the other three methods are all capable of identifying relevant gene sets known from the data sets. No method is uniformly dominant over all scenarios, and GSEA performs more consistently in finding enriched gene sets (Abatangelo et al. 2009).

### Gene set enrichment analysis for GWAS data

GWAS surveys millions of SNPs on human genome for disease-associated loci. Many of these genetic variants are located in or close to a genic region. Borrowing ideas of gene set enrichment from transcriptomic studies (Efron and Tibshirani 2007; Newton et al. 2007; Subramanian et al. 2005), several methods are implemented to conduct pathway-based analyses using the GWAS data and have been recently reviewed (Askland et al. 2009; Chen et al. 2010; Holmans 2010; Holmans et al. 2009; Kraft and Raychaudhuri 2009; Luo et al. 2010; Schaid et al. 2012; Wang et al. 2007, 2010; Weng et al. 2011; Zhang et al. 2010). These methods examine whether a set of related genes in the same biological pathway are jointly associated with a trait of interest and differ in the choice of summary statistics of genes and pathways. For example, Wang et al. (2007) assigned the highest statistic value among all SNPs mapped to a gene as the statistic value of the gene, and utilized the framework of GSEA to evaluate the statistical significance through permutation and correction for multiple testing. GSEA-SNP computes all SNPs annotated to a pathway without using a gene-level summary statistic (Holden et al. 2008). Chen et al. (2010) proposed a principal component approach to identifying “eigenSNPs” for each gene to assess their joint association. MAGENTA uses GWAS results to test for enrichment of genetic associations in predefined biological processes or sets of functionally related genes; it can be used for either hypothesis-testing or hypotheses-generating analyses (Segre et al. 2010). The in-depth comparison and application of these methods can be found in a recent review article by Wang et al. (2010). This article only discusses the most recent pathway-based methods for GWAS analysis that are not covered by previous reviews.

SNP set enrichment analysis (SSEA) implemented an adaptive rank truncated product method to select at least one representative SNP for each gene (Weng et al. 2011). For each gene, the number of representative SNPs,  $K$ , is determined by minimizing the gene-level  $p$  value with all possible  $K$  (in practice,  $K = 1$  to 10 to reduce computational burden). To correct for the various sizes of genes (measured by the number of tested SNPs), SSEA forces an equal number of representative SNPs (i.e.,  $K$ ) for genes within a pathway by using the average  $K$  of all genes in a pathway. Similar to the original GSEA, a weighted Kolmogorov–Smirnov test was used to assess the enrichment of significance on the pathway level. Applied to GWAS of schizophrenia, SSEA is able to identify statistically significant

pathways that replicate in Caucasians and African Americans, two ethnic groups with distinct genetic background (Weng et al. 2011). Askland et al. (2012) explored three methods to adjust for various gene sizes and identified that ion channel genes were consistently enriched in GWAS data sets of schizophrenia. Schaid et al. (2012) developed a new score on the SNP-level test in the framework of generalized linear models. At the gene level, the average of SNP-level score squared is used to summarize the joint association of a gene. To score a gene set, the weighted average of the gene-level score is used, where the weight is the inverse variance of the gene-level score. This general approach combines SNPs to genes and to gene sets to assess the enrichment of a given gene set, and avoids the cancelation of positive and negative effects of multiple trait-associated genes. It allows SNPs to be mapped to multiple genes or gene sets based on chromosomal locations, and directly adjust for linkage disequilibrium (LD) among SNPs within the same genic region and between genes in the same gene set (Schaid et al. 2012).

These pathway- and gene set-based approaches consider the joint contribution of multiple SNPs and genes involved in a biological pathway and highlight the potential functional relationship between pathways and diseases. Taking advantage of the enrichment of multiple associations which may not pass the genome-wide significance threshold individually, the pathway-based SNP analysis could have better statistical power for detecting disease-associated pathways than the single SNP analysis, with a smaller number of statistical tests. At the pathway level, the statistical significance can more robustly reflect the functional relationship and may be less affected by genetic structure than the single SNP test. Therefore, the significantly associated pathways can be more consistently replicated across ethnic groups compared to the SNP-level association. With the knowledge of these predefined pathways, the findings of disease-associated pathways can be complementary to the single SNP analysis in understanding the molecular mechanism of human diseases.

## Network analysis

Network analysis has been extensively applied to study biological networks including genetic networks (Newman 2010). The high-throughput technologies of genomic research have produced a deluge of data to enable network studies. Large databases such as BioGRID, DIP, HPRD, IntAct, IMID, and MIPS (Aranda et al. 2010; Keshava Prasad et al. 2009; Pagel et al. 2005; Stark et al. 2006; Warde-Farley et al. 2010; Xenarios et al. 2002) document hundreds of thousands of physical and genetic interactions from a number of organisms including humans. In this article, I focus on two types of molecular networks, co-expression networks and PPI networks, which have been relatively well understood and utilized in statistical genetic studies to potentially fill the gap of functional relationship between the genetic variants and the trait of interests.

### Co-expression network

Co-expression networks are weighted, undirected gene networks that represent the correlation among gene expression levels. The vertices represent the genes or the probes measuring the expression levels of gene transcripts. An edge between a pair of vertices represents statistically significant correlation and is weighted by the quantitative measurements of the correlation (e.g., correlation coefficient). Genes regulated by the same transcriptional mechanism may have similar profile of gene expression levels. Therefore, the co-expression network has been extensively studied to identify putative co-regulated genes in multiple organisms (Ghazalpour et al. 2006; Ideker et al. 2001; Schadt et al. 2008; Zhu et al. 2008). The network analyses of co-expression network may also suggest functional domains in the molecular system, and can be particularly powerful when the perturbation instruments are applied (Ideker et al. 2001). Over 20,000 functional genomic studies (e.g., gene expression microarray data sets) have been stored in the Gene Expression Omnibus

(Barrett et al. 2011). A large number of co-expression networks have been generated to infer genes with related functions and used to study the functional enrichment (Subramanian et al. 2005). The genetic association of the gene expression levels has also been extensively studied to identify expression-associated SNPs (eSNPs) with *cis* and *trans* genetic effects (Emilsson et al. 2008; Schadt et al. 2008; Stranger et al. 2005) and stored in databases (Gamazon et al. 2010). These eSNPs are often known as expression quantitative trait loci (eQTL). Several online databases, such as Genevar (<http://www.sanger.ac.uk/resources/software/genevar/>), the eQTL Browser (<http://www.eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>) and seeQTL ([http://www.bios.unc.edu/research/genomic\\_software/seeQTL/help.htm](http://www.bios.unc.edu/research/genomic_software/seeQTL/help.htm)), have been built and maintained based on the genetic associations of gene expression levels in multiple tissues. GWAS hits overlapping with the eSNPs can suggest putative function of transcriptional regulation. Several GWAS studies have observed that the significant associations in GWAS overlap are enriched with eSNPs (International Consortium for Blood Pressure Genome-Wide Association Studies et al. 2011; Teslovich et al. 2010; Zhong et al. 2010). The suggested role in transcriptional regulation can be followed up with functional analysis to verify the impact on transcription. Schadt et al. developed a new model to capture regulatory causality that relates the genetic variants to gene expression networks and disease traits (Chen et al. 2008; Dobrin et al. 2009; Emilsson et al. 2008; Sieberts and Schadt 2007). This integrative model infers the putative functional association with transcriptional regulation, as well as the putative causal relationship, which better facilitates the design of follow-up study to empirically confirm the molecular function.

### PPI network and interactome

The physical interactions between proteins are important for a wide range of biological functions including biochemical reactions, cytoskeletal structures, signal transduction systems and transcriptional regulation. Several high-throughput experimental methods (e.g., two-hybrid system and affinity purification-mass spectrometry) can measure the binding affinity to identify the physical interactions among proteins. The yeast two-hybrid system was used to screen over 4,000 human proteins and identified over 3,000 mostly novel PPI interactions (Stelzl et al. 2005). Other high-throughput studies of human interactome also identified a large number of PPI pairs (Gandhi et al. 2006; Sowa et al. 2009). Large databases such as BioGRID (Stark et al. 2006) and MIPS (Pagel et al. 2005) have been developed to document all known PPI information from multiple organisms including humans. As a result, we are able to query thousands of identified interactions among human proteins using these databases, to define the hypothesis-testing space for a new type of evaluation of gene-gene interactions that integrates aspects of statistical and functional evidence (Ma et al. 2012; Sun and Kardia 2010). Constructed PPI networks for genes within associated loci for complex diseases also showed abundant physical interactions between protein products of associated genes (Rossin et al. 2011). Integrated with the GWAS analyses, the information of PPI can help to interpret the genetic associations with human diseases, and provide hints to plausible functions of the genetic variants (Hannum et al. 2009; Jia et al. 2011).

### Integration of network analysis with GWAS

Several network-based methods have been developed to boost the power of gene identification in GWAS and to bridge the functional connection between GWAS findings and the pathophysiology of human diseases. Built upon the guilt-by-association (GBA) algorithm, Lee et al. (2011) developed a naïve Bayes framework to identify disease genes using a large human gene network and GWAS results of human diseases. The strength of the genetic association is measured as additive Bayes factor to infer the disease genes in addition to the prediction network. In Network Interface Miner for Multigenic Interactions

(NIMMI), the disease- or trait-associated sub-networks are identified by combining the weighted PPI network with the statistical significance from GWAS. The known trait-associated genes are mostly found in the top-ranked sub-networks with better sensitivity than the single locus analysis (Akula et al. 2011). A dense module searching method is implemented in dmGWAS to identify sub-networks for complex diseases by integrating the association signal from GWAS data sets into the human PPI network. This method can utilize the local PPI module to define the gene set for testing the enrichment of genetic associations (Jia et al. 2011).

Another integrative application of biological networks and GWAS data is to study epistasis (i.e., gene-gene interaction), which may partially explain the large proportion of missing heritability in the current GWAS study (Moore and Williams 2009; Zuk et al. 2012). In the context of the network analysis, integrating the rich information available from PPI databases and the high-throughput measurements of genetic variants, we have a unique opportunity to relate the functional interaction between gene products (e.g., proteins), and the statistical interactions associated with human traits. Several methods have used network analysis to study epistasis (Davis et al. 2010; Hu et al. 2011; Sun and Kardia 2010). Sun and Kardia (2010) demonstrated in their recent study that the functional epistasis and statistical epistasis can be integrated to identify new interacting genetic loci. The candidate pairs of interacting genetic loci were filtered by the PPI pairs among human genes, and the statistical interaction models were tested to identify epistatic effects, which may function through a protein complex with multiple peptides bound together. Although this study only examined copy number variation (CNV), this approach can be extended to analyze genome-wide SNPs by grouping the SNPs into genes.

## Visualization of biological networks

Graphic visualization is a key component to analyze the networks and present the results. To study large biological networks composed of thousands of vertices and several times more edges, a specially designed computational tool is necessary. Cytoscape is a Java-based, open source bioinformatics software platform for visualizing complex interaction networks and biological pathways (Cline et al. 2007; Shannon et al. 2003). Through powerful and flexible data integration, filtering, browsing and interactive visualization techniques, this software allows the user to explore and analyze the relationship between disease phenotypes and molecular networks and/or particular components of molecular networks (genes, proteins and their interactions). The software facilitates the integration of information from a variety of data sources (such as functional annotation, transcriptomic profiles, genetic association tests, etc.) through user-configurable visualization schemes that display attributes of network components using labels, colors, line shapes and physical proximity. Several network analysis and visualization tools, such as statnet, sna and igraph, are also available as R packages, which are convenient for users who need to integrate the network analysis with other statistical analyses in a single programming environment.

## Discussions

The network- and pathway-based methods essentially represent a framework to incorporate biological knowledge into the statistical genetic studies of complex traits. These approaches can integrate the biological networks and pathways at three major stages of genetic association study (illustrated in Fig. 2). Current methods use networks and pathways to preselect genetic variants for targeted analysis, to enrich the statistical associations and to identify functional modules based on statistical significance, but mostly focus on a single type of integration using one source of biological networks. As discussed in previous sections, there are several types of networks providing functional information at different

molecular levels (e.g., transcriptional regulation, gene expression and PPI). For all these functional information of biological pathways and networks, centralized databases have been built and constantly updated to track the information based on literatures and large-scale studies. However, each type of functional information has its strengths and weaknesses. On one hand, GO does not include the biological context, such as species, tissue specificity and cellular environment, in the annotation. Compared to co-expression and PPI networks, pathway or network analysis based on GO will only lead to the “domain” knowledge, rather than the specific functional role underlying the pathophysiology of human disease. With the proper design and careful selection of the research question, integration of co-expression and PPI data to network analysis may generate new hypotheses of specific molecular function that can be experimentally examined. On the other hand, limited by the technology of measurements and the complexity of the functional relationship, the completeness of these databases varies substantially, which may introduce bias into the statistical analysis if not appropriately considered. For example, GO provides relatively stable functional annotation for genes across species. Therefore, the results of the inferred biological attributes from the statistical analysis are relatively consistent using different versions. For less complete databases such as PPI networks and co-expression networks, the network topology may change substantially after inclusion of a large-scale experimental study. Thus, the statistical results based on such incomplete information may be biased, especially when multiple pathways and networks are compared.

Although over a thousand genetic loci have been associated with human traits and diseases by the GWAS approach, the functional variants within these loci are mostly unknown. The next-generation sequencing (NGS) enables high-throughput survey of all genetic variants, including functional variants of human genome. For instance, thousands of exomes have been sequenced to identify variants in the coding regions, especially those variants with lower frequency (Tennessen et al. 2012). Because of the low allele frequency, the power to detect the genetic association for a single variant or even for a gene is limited. Pathway and network analysis may play an important role in increasing the power for analyzing the exomic data by combining multiple related genes in a pathway or a network, and to infer the biological function underlying the disease traits. Furthermore, given the putative function of these variants identified by NGS, and the knowledge of biological networks, we are able to generate more specific hypothesis to guide statistical testing, such as variants affecting PPI and transcriptional factor—regulatory element binding. These well-defined hypothesis testing can be fruitful using the functional variants from whole exome sequencing data and enhance our understanding of the causal variants and the molecular system of complex human traits.

With continuing effort to improve the quality and the completeness of the biological networks and pathways, new initiatives try to expand the knowledge space of biological function and interaction (Ravasi et al. 2010). Developing databases that curate information about all the molecular interactions within a cell (e.g., DNA–DNA, DNA–RNA, DNA–protein, RNA–RNA, RNA–protein and protein–protein) would greatly assist in providing an a priori biological hypothesis space for association testing (Pattin and Moore 2008, 2009). Additionally, most studies of biological networks focus on the static relationship among genes and proteins under a single experimental condition. From the vast amount of documented linkages within the molecular systems, the knowledge base of the complex genetic and physical architecture has been built to understand the biological functions and processes. However, the biological networks are highly dynamic to adapt to the changing environment and to maintain critical cellular functions. Differential network studies analyze the dynamics of biological networks under different conditions and help to provide a more complete understanding of the complexity of biological systems (Ideker and Krogan 2012).



New approaches to integrating the multiple biological networks with genetic association study may further bridge the gap between the genetic variants and complex traits.

For both hypothesis-testing and hypothesis-generating methods,  $p$  values are calculated under a statistical null hypothesis. Depending on the different hypotheses and designs, such as self-contained or competitive tests, the interpretation of the  $p$  values from the pathway analyses can be very different from the single SNP association tests. For example, an over-represented GO term with a significant  $p$  value indicated its enrichment among a large number of GO terms given the input genes, which does not directly assess the association between the genes and the traits. Therefore, it is critical to understand the test statistics of these pathway analysis methods, to appropriately interpret and compare the  $p$  values from various studies. In the context of gene expression analysis, Goeman and Bühlmann discussed this methodological issue in great detail (Goeman and Bühlmann 2007).

Thomas has pointed out and re-emphasized that “incorporation of the underlying biology of the disease” is one unique characteristic of genetic epidemiology (Thomas 2000; Thomas 2012). The network- and pathway-based methods provide an integrative framework to enable such incorporation of biological knowledge. With the advancement of high-throughput technology (Yu et al. 2011), and the availability of high-quality measurement of the biological system, the integrative approach to studying the genetics of complex traits and diseases will be more fruitful. These approaches will produce more interpretable results which may lead to a better understanding of the diseases and better prevention, diagnosis and treatment eventually.

## Acknowledgments

This study was partly supported by the National Institute of Health grant HL100245.

## References

- Abatangelo L, Maglietta R, Distaso A, D'Addabbo A, Creanza T, Mukherjee S, Ancona N. Comparative study of gene set enrichment methods. *BMC Bioinformatics*. 2009; 10:275. [PubMed: 19725948]
- Akula N, Baranova A, Seto D, Solka J, Nalls MA, Singleton A, Ferrucci L, Tanaka T, Bandinelli S, Cho YS, Kim YJ, Lee JY, Han BG, McMahon FJ. Bipolar Disorder Genome Study Consortium; Wellcome Trust Case-Control Consortium. A network-based approach to prioritize results from genome-wide association studies. *PLoS One*. 2011; 6:e24220. [PubMed: 21915301]
- Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, Derow C, Feuerhann M, Ghanbarian AT, Kerrien S, Khadake J, Kerssemakers J, Leroy C, Menden M, Michaut M, Montecchi-Palazzi L, Neuhauser SN, Orchard S, Perreau V, Roechert B, van Eijk K, Hermjakob H. The IntAct molecular interaction database in 2010. *Nucleic Acids Res*. 2010; 38:D525–D531. [PubMed: 19850723]
- Askland K, Read C, Moore J. Pathways-based analyses of whole-genome association study data in bipolar disorder reveal genes mediating ion channel activity and synaptic neurotransmission. *Hum Genet*. 2009; 125:63–79. [PubMed: 19052778]
- Askland K, Read C, O'Connell C, Moore JH. Ion channels and schizophrenia: a gene set-based analytic approach to GWAS data for biological hypothesis testing. *Hum Genet*. 2012; 131:373–391. [PubMed: 21866342]
- Bansal V, Libiger O, Torkamani A, Schork NJ. Statistical analysis strategies for association studies involving rare variants. *Nat Rev Genet*. 2010; 11:773–785. [PubMed: 20940738]
- Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muetter RN, Holko M, Ayanbule O, Yefanov A, Soboleva A. NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res*. 2011; 39:D1005–D1010. [PubMed: 21097893]

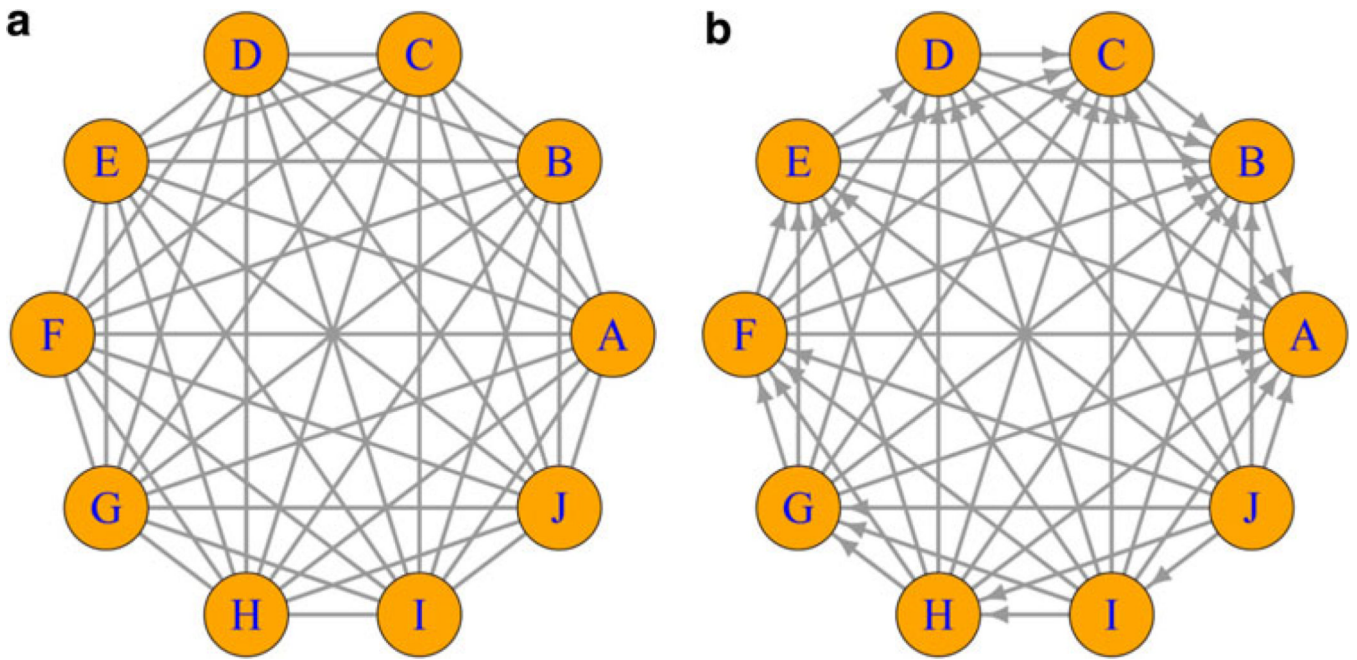
- Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, Kaipa P, Karthikeyan AS, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Paley S, Popescu L, Pujar A, Shearer AG, Zhang P, Karp PD. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 2010; 38:D473–D479. [PubMed: 19850718]
- Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK, Leonardson A, Castellini LW, Wang S, Champy MF, Zhang B, Emilsson V, Doss S, Ghazalpour A, Horvath S, Drake TA, Lusk AJ, Schadt EE. Variations in DNA elucidate molecular networks that cause disease. *Nature.* 2008; 452:429–435. [PubMed: 18344982]
- Chen LS, Hutter CM, Potter JD, Liu Y, Prentice RL, Peters U, Hsu L. Insights into colon cancer etiology via a regularized approach to gene set analysis of GWAS data. *Am J Hum Genet.* 2010; 86:860–871. [PubMed: 20560206]
- Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc.* 2007; 2:2366–2382. [PubMed: 17947979]
- Davis NA, Crowe JE Jr, Pajewski NM, McKinney BA. Surfing a genetic association interaction network to identify modulators of antibody response to smallpox vaccine. *Genes Immun.* 2010; 11:630–636. [PubMed: 20613780]
- Dering C, Hemmelmann C, Pugh E, Ziegler A. Statistical analysis of rare sequence variants: an overview of collapsing methods. *Genet Epidemiol.* 2011; 35(Suppl 1):S12–S17. [PubMed: 22128052]
- Dobrin R, Zhu J, Molony C, Argman C, Parrish ML, Carlson S, Allan MF, Pomp D, Schadt EE. Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. *Genome Biol.* 2009; 10:R55. [PubMed: 19463160]
- Efron B, Tibshirani R. On testing the significance of sets of genes. *Ann Appl Stat.* 2007; 1:107–129.
- Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S, Mouy M, Steinthorsdottir V, Eiriksdottir GH, Bjornsdottir G, Reynisdottir I, Gudbjartsson D, Helgadottir A, Jonasdottir A, Styrkarsdottir U, Gretarsdottir S, Magnusson KP, Stefansson H, Fossdal R, Kristjansson K, Gislason HG, Stefansson T, Leifsson BG, Thorsteinsdottir U, Lamb JR, Gulcher JR, Reitman ML, Kong A, Schadt EE, Stefansson K. Genetics of gene expression and its effect on disease. *Nature.* 2008; 452:423–428. [PubMed: 18344981]
- Gamazon ER, Zhang W, Konkashbaev A, Duan S, Kistner EO, Nicolae DL, Dolan ME, Cox NJ. SCAN: SNP and copy number annotation. *Bioinformatics.* 2010; 26:259–262. [PubMed: 19933162]
- Gandhi TK, Zhong J, Mathivanan S, Karthick L, Chandrika KN, Mohan SS, Sharma S, Pinkert S, Nagaraju S, Periaswamy B, Mishra G, Nandakumar K, Shen B, Deshpande N, Nayak R, Sarker M, Boeke JD, Parmigiani G, Schultz J, Bader JS, Pandey A. Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat Genet.* 2006; 38:285–293. [PubMed: 16501559]
- Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, Brozell A, Schadt EE, Drake TA, Lusk AJ, Horvath S. Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet.* 2006; 2:e130. [PubMed: 16934000]
- Goeman JJ, Buhlmann P. Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics.* 2007; 23:980–987. [PubMed: 17303618]
- Hannum G, Srivas R, Guenole A, van Attikum H, Krogan NJ, Karp RM, Ideker T. Genome-wide association data reveal a global map of genetic interactions among protein complexes. *PLoS Genet.* 2009; 5:e1000782.
- Holden M, Deng S, Wojnowski L, Kulle B. GSEA-SNP: applying gene set enrichment analysis to SNP data from genome-wide association studies. *Bioinformatics.* 2008; 24:2784. [PubMed: 18854360]
- Holmans P. Statistical methods for pathway analysis of genome-wide data for association with complex genetic traits. *Adv Genet.* 2010; 72:141–179. [PubMed: 21029852]

- Holmans P, Green EK, Pahwa JS, Ferreira MA, Purcell SM, Sklar P, Owen MJ, O'Donovan MC, Craddock N. Wellcome Trust Case-Control, C. Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. *Am J Hum Genet.* 2009; 85:13–24. [PubMed: 19539887]
- Hu T, Sinnott-Armstrong NA, Kiralis JW, Andrew AS, Karagas MR, Moore JH. Characterizing genetic interactions in human disease association studies using statistical epistasis networks. *BMC Bioinformatics.* 2011; 12:364. [PubMed: 21910885]
- Ideker T, Krogan NJ. Differential network biology. *Mol Syst Biol.* 2012; 8:565. [PubMed: 22252388]
- Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science.* 2001; 292:929. [PubMed: 11340206]
- Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, Smith AV, Tobin MD, Verwoert GC, Hwang SJ, Pihur V, Vollenweider P, O'Reilly PF, Amin N, Bragg-Gresham JL, Teumer A, Glazer NL, Launer L, Zhao JH, Aulchenko Y, Heath S, Sober S, Parsa A, Luan J, Arora P, Dehghan A, Zhang F, Lucas G, Hicks AA, Jackson AU, Peden JF, Tanaka T, Wild SH, Rudan I, Igl W, Milaneschi Y, Parker AN, Fava C, Chambers JC, Fox ER, Kumari M, Go MJ, van der Harst P, Kao WH, Sjogren M, Vinay DG, Alexander M, Tabara Y, Shaw-Hawkins S, Whincup PH, Liu Y, Shi G, Kuusisto J, Tayo B, Seielstad M, Sim X, Nguyen KD, Lehtimaki T, Matullo G, Wu Y, Gaunt TR, Onland-Moret NC, Cooper MN, Platou CG, Org E, Hardy R, Dahgam S, Palmen J, Vitart V, Braund PS, Kuznetsova T, Uitterwaal CS, Adeyemo A, Palmas W, Campbell H, Ludwig B, Tomaszewski M, Tzoulaki I, Palmer ND, Aspelund T, Garcia M, Chang YP, O'Connell JR, Steinle NI, Grobbee DE, Arking DE, Kardina SL, Morrison AC, Hernandez D, Najjar S, McArdle WL, Hadley D, Brown MJ, Connell JM, et al. International Consortium for Blood Pressure Genome-Wide Association Studies; CARDIoGRAM consortium; CKDGen Consortium; KidneyGen Consortium; EchoGen consortium; CHARGE-HF consortium. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature.* 2011; 478:103–109. [PubMed: 21909115]
- Jia P, Zheng S, Long J, Zheng W, Zhao Z. dmGWAS: dense module searching for genome-wide association studies in protein–protein interaction networks. *Bioinformatics.* 2011; 27:95–102. [PubMed: 21045073]
- Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000; 28:27–30. [PubMed: 10592173]
- Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadran S, Chaerkady R, Pandey A. Human protein reference database–2009 update. *Nucleic Acids Res.* 2009; 37:D767–D772. [PubMed: 18988627]
- Kraft P, Raychaudhuri S. Complex diseases, complex genes: keeping pathways on the right track. *Epidemiology.* 2009; 20:508–511. [PubMed: 19525687]
- Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res.* 2011; 21:1109–1121. [PubMed: 21536720]
- Luo L, Peng G, Zhu Y, Dong H, Amos CI, Xiong M. Genome-wide gene and pathway analysis. *Eur J Hum Genet.* 2010; 18:1045–1053. [PubMed: 20442747]
- Ma L, Brautbar A, Boerwinkle E, Sing CF, Clark AG, Keinan A. Knowledge-driven analysis identifies a gene–gene interaction affecting high-density lipoprotein cholesterol levels in multi-ethnic populations. *PLoS Genet.* 2012; 8:e1002714.
- Moore JH, Williams SM. Epistasis and its implications for personal genetics. *Am J Hum Genet.* 2009; 85:309–320. [PubMed: 19733727]
- Newman, MEJ. *Networks: an introduction.* Oxford, New York: Oxford University Press; 2010.
- Newton MA, Quintana FA, den Boon JA, Sengupta S, Ahlquist P. Random-set methods identify distinct aspects of the enrichment signal in gene-set analysis. *Ann Appl Stat.* 2007; 1:85–106.

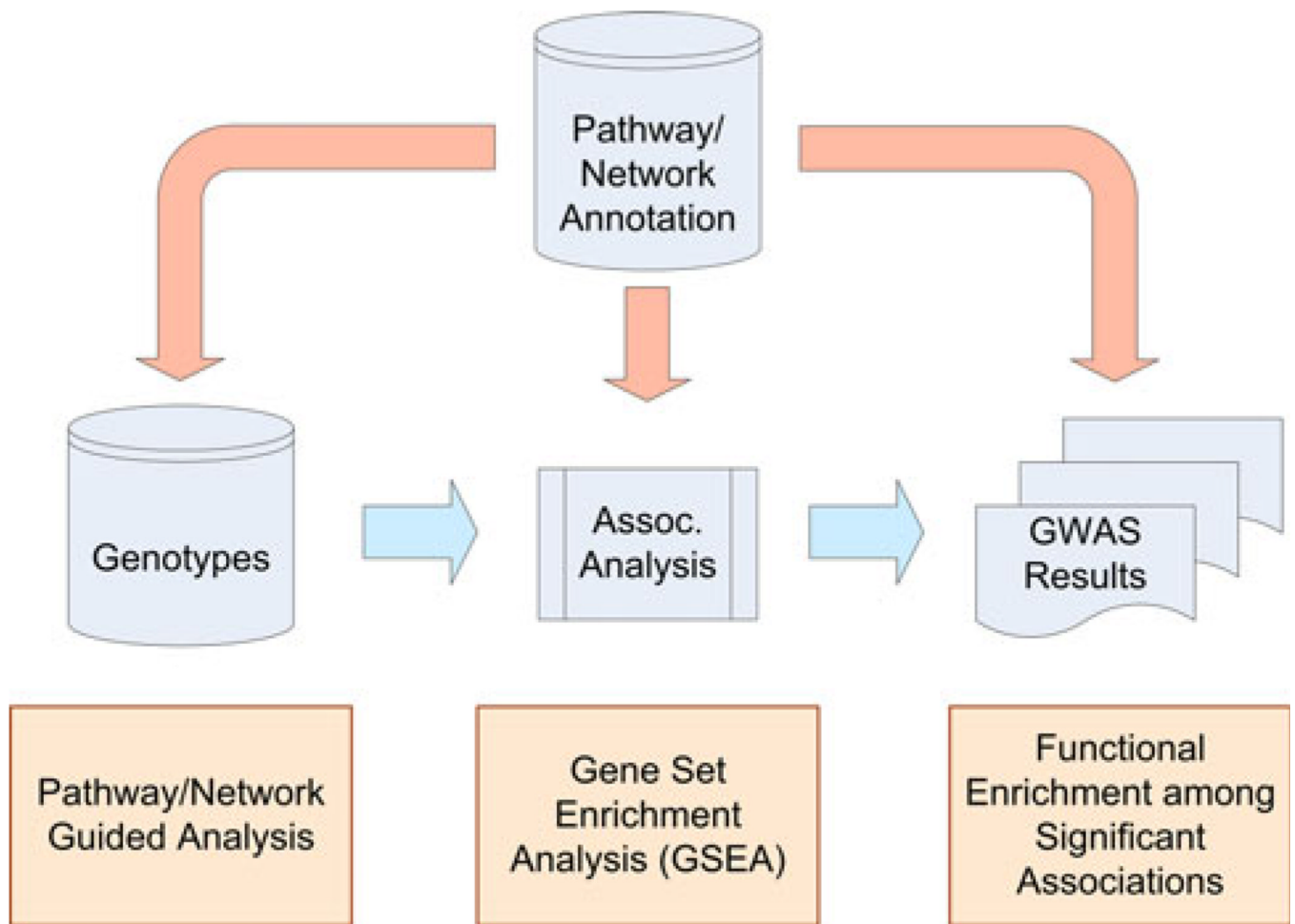
- Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Mark P, Stumpflen V, Mewes HW, Ruepp A, Frishman D. The MIPS mammalian protein–protein interaction database. *Bioinformatics* (Oxford, England). 2005; 21:832–834.
- Pattin KA, Moore JH. Exploiting the proteome to improve the genome-wide genetic analysis of epistasis in common human diseases. *Hum Genet.* 2008; 124:19–29. [PubMed: 18551320]
- Pattin KA, Moore JH. Role for protein–protein interaction databases in human genetics. *Expert Rev Proteomics.* 2009; 6:647–659. [PubMed: 19929610]
- Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, Akalin A, Schmeier S, Kanamori-Katayama M, Bertin N, Carninci P, Daub CO, Forrest AR, Gough J, Grimmond S, Han JH, Hashimoto T, Hide W, Hofmann O, Kamburov A, Kaur M, Kawaji H, Kubosaki A, Lassmann T, van Nimwegen E, MacPherson CR, Ogawa C, Radovanovic A, Schwartz A, Teasdale RD, Tegner J, Lenhard B, Teichmann SA, Arakawa T, Ninomiya N, Murakami K, Tagami M, Fukuda S, Imamura K, Kai C, Ishihara R, Kitazume Y, Kawai J, Hume DA, Ideker T, Hayashizaki Y. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell.* 2010; 140:744–752. [PubMed: 20211142]
- Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y, Cotsapas C, Daly MJ. International Inflammatory Bowel Disease Genetics Consortium. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.* 2011; 7:e1001273.
- Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, GuhaThakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl CA, van Nas A, Mehrabian M, Drake TA, Lusk AJ, Smith RC, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, Ulrich R. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* 2008; 6:e107. [PubMed: 18462017]
- Schaid DJ, Sinnwell JP, Jenkins GD, McDonnell SK, Ingle JN, Kubo M, Goss PE, Costantino JP, Wickerham DL, Weinshilboum RM. Using the gene ontology to scan multilevel gene sets for associations in genome wide association studies. *Genet Epidemiol.* 2012; 36:3–16. [PubMed: 22161999]
- Segre AV, Consortium D, investigators M, Groop L, Mootha VK, Daly MJ, Altshuler D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* 2010; 6
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 13:2498–2504. [PubMed: 14597658]
- Sieberts SK, Schadt EE. Moving toward a system genetics view of disease. *Mamm Genome.* 2007; 18:389–401. [PubMed: 17653589]
- Sowa ME, Bennett EJ, Gygi SP, Harper JW. Defining the human deubiquitinating enzyme interaction landscape. *Cell.* 2009; 138:389–403. [PubMed: 19615732]
- Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 2006; 34:D535–D539. [PubMed: 16381927]
- Stelzl U, Worm U, Lalowski M, Haenic C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE. A human protein–protein interaction network: a resource for annotating the proteome. *Cell.* 2005; 122:957–968. [PubMed: 16169070]
- Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavaré S, Deloukas P, Dermitzakis ET. Genome-wide associations of gene expression variation in humans. *PLoS Genet.* 2005; 1:e78. [PubMed: 16362079]
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005; 102:15545–15550. [PubMed: 16199517]
- Sun YV, Kardia SL. Identification of epistatic effects using a protein–protein interaction database. *Hum Mol Genet.* 2010; 19:4345–4352. [PubMed: 20736252]

- Sun YV, Sung YJ, Tintle N, Ziegler A. Identification of genetic association of multiple rare variants using collapsing methods. *Genet Epidemiol.* 2011; 35(Suppl 1):S101–S106. [PubMed: 22128049]
- Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, Akey JM, Broad GO, Seattle GO. on behalf of the NESP. Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. *Science.* 2012
- Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, Lee JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RY, Wright AF, Witterman JC, Wilson JF, Willemssen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJ, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruukonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BW, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature.* 2010; 466:707–713. [PubMed: 20686565]
- Thomas DC. Genetic epidemiology with a capital“E”. *Genet Epidemiol.* 2000; 19:289–300. [PubMed: 11108640]
- Thomas DC. Genetic epidemiology with a capital E: where will we be in another 10 years? *Genetic Epidemiology.* 2012
- Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet.* 2007; 81:1278–1283. [PubMed: 17966091]
- Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet.* 2010; 11:843–854. [PubMed: 21085203]
- Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 2010; 38:W214–W220. [PubMed: 20576703]
- Weng L, Macciardi F, Subramanian A, Guffanti G, Potkin SG, Yu Z, Xie X. SNP-based pathway enrichment analysis for genome-wide association studies. *BMC Bioinformatics.* 2011; 12:99. [PubMed: 21496265]
- Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* 2002; 30:303–305. [PubMed: 11752321]
- Hu H, Tardivo L, Tam S, Weiner E, Gebreab F, Fan C, Svrikapa N, Hirozane-Kishikawa T, Rietman E, Yang X, Sahalie J, Salehi-Ashtiani K, Hao T, Cusick ME, Hill DE, Roth FP, Braun P, Vidal M. Next-generation sequencing to generate interactome datasets. *Nat Methods.* 2011; 8:478–480. [PubMed: 21516116]
- Zhang K, Cui S, Chang S, Zhang L, Wang J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res.* 2010; 38:W90–W95. [PubMed: 20435672]
- Zhong H, Beaulaurier J, Lum PY, Molony C, Yang X, Macneil DJ, Weingarh DT, Zhang B, Greenawalt D, Dobrin R, Hao K, Woo S, Fabre-Suver C, Qian S, Tota MR, Keller MP, Kendziorzski CM, Yandell BS, Castro V, Attie AD, Kaplan LM, Schadt EE. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet.* 2010; 6:e1000932.
- Zhu J, Zhang B, Smith EN, Drees B, Brem RB, Kruglyak L, Bumgarner RE, Schadt EE. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet.* 2008; 40:854–861. [PubMed: 18552845]

Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc Natl Acad Sci USA*. 2012



**Fig. 1.**  
Examples of undirected and directed networks



**Fig. 2.** Integration of pathway and network information at different stages of genetic association studies