

# Systems Biology Approaches and Applications in Obesity, Diabetes, and Cardiovascular Diseases

Qingying Meng · Ville-Petteri Mäkinen · Helen Luk · Xia Yang

Published online: 18 October 2012

© The Author(s) 2012. This article is published with open access at Springerlink.com

**Abstract** The metabolically connected triad of obesity, diabetes, and cardiovascular diseases is a major public health threat, and is expected to worsen due to the global shift toward energy-rich and sedentary living. Despite decades of intense research, a large part of the molecular pathogenesis behind complex metabolic diseases remains unknown. Recent advances in genetics, epigenomics, transcriptomics, proteomics and metabolomics enable us to obtain large-scale snapshots of the etiological processes in multiple disease-related cells, tissues and organs. These datasets provide us with an opportunity to go beyond conventional reductionist approaches and to pinpoint the specific perturbations in critical biological processes. In this review, we summarize systems biology methodologies such as functional genomics, causality inference, data-driven biological network construction, and higher-level integrative analyses that can produce novel mechanistic insights, identify disease biomarkers, and uncover potential therapeutic targets from a combination of omics datasets. Importantly, we also demonstrate the power of these approaches by application examples in obesity, diabetes, and cardiovascular diseases.

**Keywords** Metabolic disorders · Obesity · Diabetes · Cardiovascular diseases · Systems biology · Integrative genomics · Functional genomics · Causality inference · Network biology

## Introduction

Common metabolically connected diseases (MetDs) such as cardiovascular disease (CVD), type 2 diabetes (T2D), and obesity impose a substantial health burden worldwide, as demonstrated by the fact that both CVD and T2D are among the top ten leading causes of death in Europe and the United States. As obesity (defined as body mass index (BMI)  $>30 \text{ kg/m}^2$ ) is a key risk factor for both T2D and CVD, the rapidly growing obesity epidemic has further exacerbated the high morbidity and mortality, making an in-depth understanding of the mechanisms of MetDs and the development of novel therapeutic strategies more pressing.

Decades of research on MetDs show that obesity, T2D, and CVD are tightly linked and both genetic and environmental factors contribute to the susceptibility [1]. At the genetic level, hundreds of individual genetic loci are associated with MetDs as shown by recent genome-wide association studies (GWAS) [2•]. It is striking, however, that the genetic loci discovered together only explain a small proportion (typically  $<20 \%$ ) of disease heritability and a large proportion of the loci appear to act through unknown mechanisms [3]. Adding to the genetic complexity, environmental perturbations such as diet, lack of physical activity, and exposure to certain xenobiotic chemicals also increase susceptibility to MetDs but the exact mechanisms remain unclear [4]. Another layer of complexity lies in the intercellular interactions between different cell types, tissues, or organs relevant to MetDs. The above facts call for an improved systems framework to address the missing heritability and molecular mechanisms that underlie MetDs.

The past two decades have witnessed major advances in omics technologies and genome-scale molecular data can now be obtained in diverse experimental and epidemiological settings. A large number of genetic, epigenetic, transcriptomic, proteomic and metabolomic studies of MetDs have

Q. Meng · V.-P. Mäkinen · H. Luk · X. Yang (✉)  
Department of Integrative Biology and Physiology,  
University of California (UCLA),  
610 Charles E. Young Dr E.,  
Terasaki Life Sciences Building,  
Los Angeles, CA 90095, USA  
e-mail: xyang123@ucla.edu

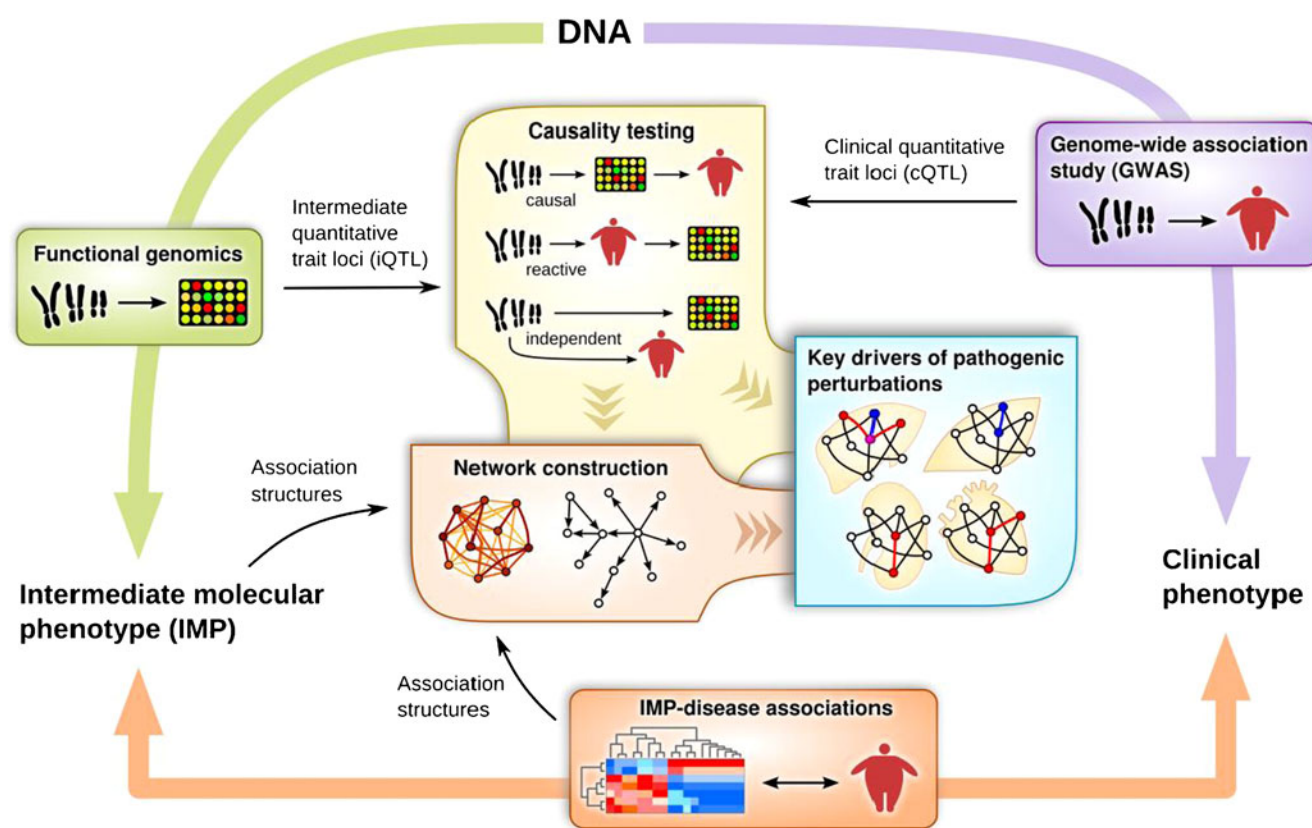
already been conducted and numerous datasets at various molecular levels have become available. But collecting information is not enough. A key challenge for the research community is to integrate the sea of data into useful knowledge and insights. To address this critical need, various systems biology approaches have been developed over the past few years to organize the fragmented but valuable information into structures that collectively explain the observed biological phenomena.

Traditional approaches focus on direct correlations between the molecular traits and the clinical diagnosis under investigation, but more advanced integrative or systems biology methodologies can help infer causality, disseminate the regulatory relationships among molecular traits, identify those biological processes or networks that are perturbed, and ultimately pinpoint the key regulatory genes and mechanisms of disease pathophysiology (Fig. 1). In this article,

we systematically review the principles, advantages, and limitations of various systems biology approaches available and how these approaches have been applied to MetDs.

### Traditional Approaches-Association and Correlative Analyses

When a particular level of molecular data is generated, the most straightforward approach is to estimate the correlations between the molecular traits and clinical phenotypes. To simplify the discussion, we classify the molecular traits as either DNA, for which variation is typically captured by single nucleotide polymorphisms (SNPs), or intermediate molecular phenotypes (IMPs) that collectively represent the downstream layers such as the transcriptome, epigenome, proteome, or metabolome. Although studies on multiple types of IMPs have been fruitful in the past years [5–8],



**Fig. 1** Systems biology strategies that integrate large-scale genetic, intermediate molecular phenotypes (IMPs, primarily gene expression), and disease phenotypes. Traditional genetic association studies such as GWAS identify genetic loci associated with clinical disease phenotypes (cQTLs, right lavender edge), which provides causal information but lacks mechanistic insights. Molecular profiling experiments help identify IMPs associated or correlated with disease status (bottom orange edge) but the results are purely correlative with no causal information. More recent functional genomics efforts offer mechanistic insights on how DNA variations affect IMPs (primarily gene expression) via the identification of intermediate QTLs (iQTLs; left lime edge). By

leveraging both iQTL and cQTL and performing statistical testing to differentiate causal, reactive, and independent relationships between IMPs and disease, one can detect putative disease causal genes (*center yellow box*). IMPs, iQTLs, cQTLs, disease phenotypes, and genetic causality can all be fed into various network construction algorithms to reconstruct regulatory networks that inform on mechanisms of IMP and disease regulation (*center orange box*). Higher level integrative approaches that take advantage of multiple methodologies are used to derive key regulatory genes and subnetworks underlying disease development in a tissue-specific fashion (*center blue box*)

we mainly focus on gene expression profiling and genetic association studies due to their ubiquitous applications in biomedicine to illustrate the power and limitations of the traditional approaches.

#### *Identification of Genes Correlated with MetDs via Gene Expression Profiling*

The explosion of gene expression data has made central data repositories such as the NCBI Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) and EMBL-EBI ArrayExpress Archive (<http://www.ebi.ac.uk/arrayexpress/>) a power house for large-scale systematic meta-analysis, or expression GWAS (eGWAS). It is plausible that disease-related genes exhibit persistent differential expression patterns across multiple studies, and that by scanning through a large data repository related to a given disease condition, these genes can be identified [9, 10, 11•, 12–14]. In a series of studies, Butte and colleagues developed and applied various bioinformatics tools [15–17] to screen GEO datasets for a wide spectrum of human diseases or phenotypes including T2D [11•] and obesity [18]. When conducting an eGWAS across 130 independent experiments that included a total of 1175 T2D case–control microarrays, they identified CD44 as the top differentially expressed gene across studies and experimentally confirmed its role in modulating adipose tissue inflammation, insulin sensitivity, and glycemic control [11•]. Although proven informative and useful, the end products of such analysis are lists of genes correlated with disease status or phenotypes with limited ability to separate non-causal secondary effects from the causal perturbations that trigger and maintain the disease processes.

#### *Identification of Genetic Risks of MetDs by Linkage Studies and GWAS*

Genetic association studies between genetic markers and disease phenotypes could infer causality to a certain degree under the central dogma that heritable disease risks flow from DNA to other downstream molecular and physiological events. That said, the causality inference is not unambiguous due to potential confounding effects [19]. Before the arrival of comprehensive human genome map and affordable genotyping platforms, linkage studies were extensively used to identify genetic risk loci for MetDs by genotyping sparse genetic markers in human families or F2 crosses of animal models [20–28]. The resulting clinical quantitative trait loci (cQTLs) were then used for fine mapping and positional cloning. However, the linkage blocks represented by the genetic markers were typically large (covering tens or even hundreds of genes), making the identification of candidate genes a labor-intensive and costly process.

In the past few years, human GWAS have achieved great success in uncovering novel genetic risk loci for hundreds of diseases or traits including MetDs and their risk phenotypes [29]. GWAS examine the associations between common genetic variants represented by SNPs and clinical traits or diseases in large human cohorts. As of August 2012, the GWAS Catalog [2•] maintained by the National Human Genome Research Institute (<https://www.genome.gov/gwastudies>) has collected a total of 6598 SNPs for 677 unique human diseases or phenotypic traits from 1320 different publications. A number of loci have been associated with MetDs—56 for obesity, 48 for T2D, and 38 for CVD. Most of the genetic loci identified were never observed before the GWAS era and many of the candidate genes cannot be explained by current knowledge. For instance, from a recent GWAS of over 80,000 individuals, 13 out of the 23 significant genetic loci identified were novel and 17 appeared to contribute to CAD risk through mechanisms independent of traditional risk factors [3].

Despite the demonstrated power in uncovering novel genetic risk factors, GWAS have limitations. First, as SNP markers on GWAS genotyping panels are largely pre-selected tag SNPs rather than functional SNPs, the discovered loci may not directly represent the functional causal SNPs but may simply be proxies or in linkage disequilibrium with true causal genetic variations. Second, even if the SNP is causal itself, the exact functional consequences of the SNP and the genes affected are not directly identifiable, especially when the SNP is (as 40 % of the significant SNPs are) located in intergenic regions or within introns [30]. The third issue with GWAS is the missing heritability [31]. For example, the largest meta-analysis of GWAS for obesity where ~250 k individuals were surveyed identified a total of 32 significant loci for BMI but together they only contribute to 2–4 % of the genetic variance [32]. Similarly, all 56 established loci identified for T2D together only explain ~10 % of the heritability [33].

The lack of explanatory and predictive power has cast severe criticisms and doubts on GWAS. That being said, GWAS provide first-hand information about the putative causal relationships between genetic variants and clinical diseases traits. In addition, much of the hidden heritability that could not pass the stringent genome-wide significance threshold may lie within the lower tier GWAS signals and can be brought back to light using more advanced integrative methodologies, as discussed later in the review.

#### Functional Genomics

Once a genetic locus has been linked to a disease phenotype, the most intuitive step is to search for candidate genes in the neighborhood of the locus. However, tens to hundreds of genes may underlie each cQTL and a large proportion of the

GWAS loci lie within intergenic regions, making it difficult to pinpoint the true underlying causal gene and mechanisms. Functional genomics aims to bridge the gap. In principle, genetic loci can impose disease risk via the products of the genome, that is, IMPs such as gene expression, microRNA, DNA methylation, protein levels and downstream metabolites. By linking genetic markers or SNPs to IMPs, functional consequences of genetic perturbations and mechanistic insights can be inferred.

As summarized in our recent review articles [34, 35•, 36], various types of IMPs have been screened in terms of their relationships with genetic variations to produce a set of intermediate molecular QTLs (iQTLs). In particular, the subset of iQTLs that are related to gene expression, denoted by expression QTLs (eQTLs), have been identified in a spectrum of MetD-related tissues or cell types including whole blood, monocyte, lymphocyte, liver, adipose, and muscle in humans [35•] and in a wider coverage of cell/tissue types in animal models or other model organisms [37–39]. Furthermore, large-scale projects such as The Encyclopedia of DNA Elements (ENCODE; <http://www.encodeproject.org/ENCODE/index.html>) and Genotype-Tissue Expression (GTEx; <http://commonfund.nih.gov/GTEx/>) are likely to provide a comprehensive characterization of the full functional genome in the future. These iQTL/eQTL studies are able to establish the fundamental genetic regulatory architecture of IMPs, and thus serve as the basis for causality inference between genes and disease phenotypes.

#### Causality Inference via Integration of Genetic, Gene Expression, and Disease Phenotypes

Based on the central dogma, heritable information flows from DNA to RNA, to proteins, and then to phenotypic traits. When cQTL and eQTL overlap, that is, a genetic locus or multiple genetic loci are in association with both disease phenotype and the expression levels of a gene, there is a higher probability that the gene associated with the disease-linked genetic loci is also the causal gene. Based on this principle, Meng et al. identified a significant overlap between the cQTL of aortic calcification and the eQTL for gene *ABCC6* in a mouse F2 population and experimentally validated *ABCC6* as the causal gene for the phenotype [40]. The same principle has been applied to human GWAS studies where disease-associated SNPs are intersected with eSNPs (SNPs associated with gene expression levels under eQTLs) databases to derive candidate genes for MetDs and related phenotypic traits [3, 35•, 41, 42•, 43, 44, 45•, 46•, 47, 48]. The identification of *SORT1* as the gene behind the chromosome 1p13.3 locus for LDL cholesterol and CVD is another example of successful causal inference via functional genomics. Altogether three genes – *PSRC1*, *CELSR2*, and *SORT1* – were located within the adjacency of the 1p13.3

risk locus, but *SORT1* was identified as the strongest candidate due to the most significant association between *SORT1* expression and the 1p13.3 GWAS SNP in a liver eQTL study [44]. Tested in both transgenic and knockdown mouse models, *SORT1* was successfully validated as the causal gene for LDL and CVD via modulation of hepatic lipoprotein export by two groups [49, 50]. In cases where the GWAS locus is within an intergenic region, eSNPs can reflect genetic control of gene expression regardless of the genomic position. For instance, expression levels of *ULK3* are consistently associated with the intergenic CVD locus 15q24 across adipose, monocyte, liver, and blood [35•], thus representing a plausible candidate gene for this locus.

As IMPs can be upstream (i.e., causal), downstream (i.e., reactive), or independent of a disease phenotype, eQTL and cQTL overlap does not directly implicate causality. The three relationships can be formally tested via statistical and mathematical inference by using DNA or genetic variation information as the anchor (Fig. 1). To this end, a likelihood-based causality model selection (LCMS) procedure was developed by Schadt et al. [20]. The LCMS causality test was applied to obesity and atherosclerosis phenotypes in mouse F2 populations and successfully identified hundreds of candidate causal genes for adiposity and aortic lesions [20, 51, 52•]. In the obesity study, perturbation of eight top obesity candidate causal genes for obesity – *Zfp90*, *Lpl*, *Tgfb2*, *C3ar1*, *Gpx3*, *Gas7*, *Lactb*, and *Gyk* – was found to alter adiposity or fat pad mass in knockout or transgenic mouse models via modulation of genes involved in metabolic pathways and a liver network of genes involved in lipid metabolism [51]. In the atherosclerosis study, the top causal genes for aortic lesions were enriched for those involved in inflammatory processes such as lymphocyte activation and B cell receptor signaling [52•]. Knockout mouse model of a candidate causal gene, *C3ar1*, was found to reduce aortic lesions. In addition, the expression levels of most causal genes in the aortic arch altered accompanying lesion progression in two independent atherosclerosis mouse models. Furthermore, several causal genes overlapped with candidate genes from CVD-related human GWAS. These validation experiments strongly support the validity and power of LCMS in predicting reliable causal genes for complex MetDs.

#### Construction of Molecular Networks via Integration of Genetics, IMPs, and Disease Phenotypes

All the methodologies outlined above yield lists of molecular markers that are linked to disease development, but they offer little information on how genes and other IMPs are organized and how they operate together in complex biological systems. Networks have emerged as appealing tools to address this complexity; they depict the active

agents in the systems as nodes, and their interactions as edges that connect the nodes. Notably, the edges can represent different types of relationships such as correlation, physical binding, biochemical reactions or transcriptional regulation, thereby transcending the boundaries of conventional statistics.

Networks can be constructed based on curated knowledge (knowledge-driven) or computational modeling of large-scale genomic data (data-driven). Examples of knowledge-driven networks include protein-protein interaction (PPI) networks in the Human Protein Reference Database (<http://www.hprd.org/>), Biological General Repository for Interaction Datasets (BioGRID, <http://thebiogrid.org/>), and Ingenuity networks ([www.ingenuity.com](http://www.ingenuity.com)). These networks can capture literature-supported relationships but are far from being comprehensive and novel relationships or insights will not be covered. On the other hand, data-driven network reconstruction or reverse-engineering approaches systematically and objectively scan through and integrate all data points to uncover novel relationships among IMPs within a cell or a tissue, or even across tissues [53]. Information obtained from correlations, cQTLs, eQTLs, and causality inference discussed above can all be efficiently incorporated and utilized in various network reconstruction approaches (Fig. 1).

Examples of data-driven networks include weighted gene co-expression network analysis (WGCNA) [54, 55], Bayesian network (BN) [56–60], graphical Gaussian models (GGMs) [61–63], and algorithm for the reconstruction of accurate cellular networks (ARACNE) [64]. Although these different methodologies can be applied to various types of common complex diseases, studies on MetDs have primarily employed WGCNA and BN to identify disease-associated key regulatory genes and gene sub-networks [65–69, 70•, 71, 72•].

#### *Weighted Gene Co-expression Network Analysis (WGCNA)*

WGCNA aims to identify the correlation patterns among IMPs (primarily gene expression traits) across samples involved in a study. The construction of co-expression network starts with a Pearson correlation matrix between all gene pairs, followed by transformation of the correlation matrix into an adjacency matrix [54, 55]. The adjacency matrix is further transformed into a topological overlap matrix based on the direct interactions between genes as well as the indirect interactions with all the other genes [73]. An average linkage hierarchical clustering algorithm is then applied to the topological overlap matrix, which is followed by a dynamic cut-tree algorithm to identify gene modules [74]. Correlations between the principal components of each module and phenotypic traits measured in the same individuals can be calculated to derive informative modules that

link to the disease of interest. Alternatively, co-expression networks can be constructed separately for disease cases and controls, and network modules that demonstrate differential network topology and connectivity between cases and controls can be identified [75].

In contrast to simple clustering algorithms where genes are grouped based on the strength of pair-wise correlations, WGCNA searches for higher-level co-regulation structures. Importantly, the gene memberships of a module are determined not only by their direct correlations but also by the similarity in their relationships with the other genes [54, 73]. The network structure derived is hence comprised of more cohesive and biologically more meaningful modules that contain genes with shared regulatory mechanisms, involved in similar biological functions or pathways, or enriched for disease associated genes [44, 54, 65–69, 76–78].

Numerous studies have applied WGCNA to study the molecular mechanisms underlying MetDs or related phenotypes [67, 68, 71, 72•, 76, 79–86]. As exemplified in two parallel studies, Chen et al. and Emilsson et al. identified a co-expression module that is conserved between liver and adipose tissues, conserved between human and mouse, highly enriched for macrophage- and spleen-related inflammatory genes, and linked to various metabolic phenotypes including adiposity, atherosclerosis, and plasma lipids, insulin and glucose levels [65, 66]. This module is termed macrophage-enriched metabolic network (MEMN). Several novel genes in MEMN, including *Lactb*, *Lpl*, and *Ppm1l*, were experimentally confirmed to affect adiposity in knockout and transgenic mouse models [65]. Ghazalpour et al. and Lum et al. identified gene modules related to obesity and diabetes by using gene expression data from liver and whole brain tissues from a mouse cross segregating on metabolic phenotypes. Genes in the obesity-linked liver module are involved in adipogenesis and fatty acid metabolism whereas genes in the obesity- and T2D-associated brain subnetworks are involved in diverse processes including RNA splicing, circadian rhythm, and lipid metabolism. By constructing co-expression networks in six metabolically related tissues in a mouse population with varying T2D susceptibility, Keller et al. identified a cell cycle regulatory module in islets that predicts islet replication and diabetes development [84]. In two studies on a Finnish cohort, Inouye et al. constructed co-expression networks using blood transcriptomic data and identified a lipid-leukocyte module that was highly enriched for inflammatory genes and significantly linked to over 80 serum metabolites including lipoprotein subclasses, lipids, and amino acids, thereby playing an important role in connecting inflammation, metabolism, adiposity, and atherogenesis [71, 72•]. All these examples substantiate the power of WGCNA in identifying novel genes and mechanisms that contribute to MetDs.

### *Bayesian Network (BN)*

Although WGCNA is highly informative for deriving the overall organization of genes or other IMPs and for linking particular co-expression modules to disease phenotypes, the detailed relationships among genes within a module or between modules can be less descriptive. Graphical network modules such as BNs can provide more granular views of the interactions and directionalities between genes. BNs define a partitioned joint conditional probability distribution over all nodes (genes or other IMPs) in a network where the probability distribution of states of a node depends only on the states of its parent nodes [87]. Therefore, BNs are probability-based directed acyclic graphs. The conditional probabilities reflect not only relationships between genes, but also the stochastic nature of these relationships. Due to computational constraints, thousands of plausible BNs can be generated using Monte Carlo Markov chain (MCMC) simulations [88] rather than an exhaustive search for all possible network structures. The posterior probability of each BN model given observed data can be calculated using the Bayes formula. A consensus BN that contain nodes and edges appearing in a large proportion of all plausible network models is then derived. As probability distributions are bi-directional and can lead to mathematically equivalent structures, it is not possible to infer causal directions between nodes. Fortunately, BN framework can incorporate a variety of prior information, ranging from literature, genetic, transcription factor binding, metabolomics, to proteomic data, to break the symmetry among nodes and infer causal directions [56, 57]. As the BN algorithm imposes heavy computing burden and only conserved nodes and edges across plausible networks are kept, BNs are sparser than co-expression networks and not all genes profiled are included in the BN model.

A number of studies in a variety of species have demonstrated that BNs can capture fundamental properties of molecular interactions in complex systems and can infer mechanisms [44, 56, 57, 78, 89, 90]. In searching for the mechanisms underlying the previously discussed lipid and CVD locus 1p13.3, Schadt et al. found that the three candidate genes adjacent to the locus—SORT1, CELSR2, and PSRC1—are highly connected in liver BNs. In addition, the neighborhood subnetworks of the three genes, particularly that of SORT1, are enriched for genes involved in multiple biological processes relevant to lipid regulation and CVD development, thus providing mechanistic support on the involvement of the candidate genes in CVD [44]. To illustrate how candidate causal genes identified via the LCMS causality test described above interact and affect obesity, causal genes were mapped to a liver BN and they were found to be highly connected in a subnetwork, with the top causal gene *Zfp90* being upstream of the other causal

genes [20]. In a follow-up validation study, by mapping the liver genes perturbed by the overexpression or knockout of top obesity candidate causal genes to liver BNs, a liver core subnetwork that is highly enriched for genes involved in lipid metabolism and fat cell differentiation pathways was identified, further elucidating the mechanisms underlying obesity development [51].

### High Level Integrative Approaches

As shown in Table 1, each methodology described above provides different levels of information with gradual increase in the power to inform on causal genes, biological processes and mechanisms involved in disease pathogenesis. However, each methodology also carries intrinsic limitations. To maximize our ability to discover novel insights, higher level integrative approaches that take advantage of different combinations of the above-mentioned methods have been recently explored and we highlight two such methodologies.

#### *Integration of Disease-Related Gene Sets and Networks*

To harness the strengths of data-driven regulatory networks, the information from gene expression profiling, causality testing and GWAS can be overlaid onto co-expression networks and BNs to infer disease mechanisms and key regulatory genes. In a recent study, we identified a consistent differential gene signature comprised of ~1500 inflammation-related genes, termed inflammatome, from 12 tissue-specific gene expression profiling data of 11 different mouse and rat disease models of obesity, diabetes, atherosclerosis, respiratory diseases, autoimmune diseases, inflammation, aging, and sarcopenia [70]. This inflammatome gene set was integrated with the GWAS catalog and the metabolic disease-related MEMN to confirm the causal nature of the gene signature. By intersecting the inflammatome signature with co-expression networks and BNs constructed from multiple independent datasets, we not only extracted a consistent subnetwork representing the inflammatome signature and detailing gene-gene relationships, but also identified top key regulatory genes based on the BN network topology [70, 78, 91]. The identification of the common inflammatome signature, its network architecture, and key drivers via this highly integrative approach sheds light on the shared etiology and potential therapeutic targets between MetDs and other common diseases or pathophysiological conditions, a level of mechanistic insights far beyond that of what IMP profiling could offer by itself. Other types of networks such as PPI networks can certainly also be integrated, as exemplified by Mori et al. in a study comparing two mouse strains with different susceptibility to diabetes [92]. By leveraging tissue-specific gene expression with PPI networks, they identified an inflammation- and

**Table 1** Comparison of integrative methodologies discussed in the manuscript

Methodology	Brief description	Information derived	Advantages	Limitations
IMP-disease association or correlation analysis	Association or correlation analysis between IMPs and disease phenotypes	List of differential IMPs between cases and controls or IMPs correlated with quantitative phenotypes	Informative on IMPs co-segregating with disease	No causality information
Linkage studies or GWAS	Association between genetic markers or dense SNPs with disease phenotypes	List of genetic loci associated with disease (cQTLs)	Implicates potential causal role of genetic loci	Confers little information on underlying genes and mechanisms
Functional genomics	Association between genetic markers or dense SNPs with IMPs	IMPs that are associated with genetic loci	Infers functional consequences of genetic loci on IMPs; inform on molecular mechanisms	No information on disease relevance
Causality test	Testing causal, reactive, and independent relationships between IMPs and disease by anchoring at shared genetic loci (cQTL/eQTL overlap)	List of genes tested causal for the disease	Inform on candidate causal genes for disease	Statistical inference only and validation needed; little mechanisms
WGCNA network modeling	Organizing IMPs into co-regulated network modules based on correlations between IMPs	Global overview of co-regulation or co-expression structure of IMPs and modules associated with disease phenotypes	Inform on disease mechanisms	Mainly a co-regulation structure but with little regulatory mechanisms
BN modeling	Integrating multiple levels of IMPs to define regulatory relationships	Graphical model depicting detailed interactions and relationships between IMPs	Inform on regulatory mechanisms between IMPs	Computationally intensive, sparse, no feedback loops
Network-driven higher level integration	Integrate network models with GWAS, functional genomics, causality, and IMP profiling to identify key driver genes and subnetworks associated with disease	Key driver genes and subnetworks associated with disease	Prioritize genes and provide mechanisms	Although most informative given higher amount of data incorporated, still hypothesis generating in nature and warrants experimental validation

immune system-related adipose subnetwork that contributes to the differences in diabetes risk.

#### *Integration of GWAS, Functional Genomics, Biological Pathways, and Data-Driven Networks*

In order to explore the candidate causal genes and mechanisms behind GWAS, several novel functional genomics and network-driven methodologies have been recently developed. By integrating GWAS with functional genomics (primarily eQTLs or eSNPs), Zhong et al. found that eSNPs from T2D-related tissues such as liver and adipose tissues are more enriched for T2D risk SNPs [45•]. Furthermore, by coupling T2D GWAS and eSNPs with biological pathways, candidate causal genes, co-expression networks, and BNs, multiple subnetworks and biological processes such as lipid and fatty acid metabolism, calcium signaling, PPAR

signaling, TGF-beta signaling, tight junction, complement and coagulation, antigen processing and presentation, and fat cell differentiation were found to be enriched for T2D genetic risks [45•, 46•, 93]. It is of note that most genes in the significant pathways or subnetworks only showed modest association in GWAS and therefore were missed by the traditional GWAS analysis. These results support the hypothesis that a large number of genes in relevant biological processes with modest effect sizes, rather than only a handful of individual genes with strong effects, collectively contribute to disease development. In another method, Kang et al. screened hundreds of co-expression networks for genes that were consistently co-expressed with known top T2D GWAS candidate genes as a means to identify novel T2D genes [94•]. These new methodologies not only provide mechanistic explanations for GWAS findings but also demystify a significant amount of the missing heritability.

## Conclusion

Systems biology approaches that leverage genetic, tissue-specific IMP profiling data, and disease phenotypes have evolved rapidly in the past decade. Through their applications in various MetDs in both animal models and human populations, these highly integrative systems biology approaches have unveiled unprecedented insights into disease etiology and uncovered a large number of candidate novel genes, pathways, and subnetworks associated with MetDs. By far, inflammation and immune response related genes and processes have been the most consistent signal across tissue types, across studies, and across MetDs, and thus convincingly represent a key shared component of MetDs. The systems integration of tissue-specific molecular data also revealed many tissue- and disease-specific processes, such as liver-centric lipid metabolism and transport pathways for obesity and CVD; liver- and adipose-specific oxidative phosphorylation, fatty acid oxidation, PPAR signaling, fat cell differentiation for obesity, insulin resistance, and T2D; liver-specific glucagon signaling and islet-specific cell cycle regulation for T2D; and circadian rhythm and RNA splicing processes in brain for obesity and T2D.

These findings highlight how tissue-specific gene networks and their cross-tissue interactions, rather than individual genes, mediate MetD etiology. It is therefore critical to shift from a traditional view of disease mechanisms as independent actions of individual genes to a network view, where a large number of genes coordinately define a particular network state in individual tissues and the interactions of gene networks in multiple tissues ultimately lead to MetD onset.

## Future Directions

Although proven predictive and informative, the existing systems biology methodologies are far from being comprehensive and accurate. Further refinement of existing methods and development of more advanced approaches are thus warranted. First of all, incorporation of next generation sequencing, DNA methylome, microRNA, metabolomics, and other types of data into the systems biology framework has become more pressing than ever, as such data are being rapidly generated and poured into data depositories in the past couple of years. Although some of the existing methodologies can be easily adapted for additional data types, innovative approaches guided by biological insights are still in great need. For instance, based on the regulatory relationships across data types, it is necessary to develop methodologies that leverage multiple levels of IMPs simultaneously to construct more sophisticated networks such as co-regulatory microRNA-gene-metabolite networks.

Second, the involvement of multiple cell types, tissues, and organs in MetDs demands methodologies that explore cross-

tissue interactions. As demonstrated elegantly in a recent study by Dutta et al., communications and signaling transductions across multiple critical tissues including heart muscle, sympathetic nervous system, bone marrow niches, and spleen are all involved in the enhanced inflammatory response after MI to induce atherosclerosis acceleration [95]. This type of relationship can only be revealed when data integration reaches organism-wise systems level. Although the construction of cross-tissue networks has been sporadically attempted [84, 96], such efforts have to be further expanded to increase tissue coverage and to develop more efficient methodologies.

Third, most of the current methodologies capture static information that only represents snapshots of disease status at a given time. Dynamic models that take IMP data generated from time-course experiments are therefore needed to capture the dynamic nature of disease progression. All these different levels of technical and biological challenges have to be properly addressed in the future to allow a full dissemination of MetD etiology. Only when a comprehensive understanding is achieved, can effective diagnostic, preventative, and therapeutic strategies toward these disabling and deadly diseases become a reality.

**Disclosure** No potential conflicts of interest relevant to this article were reported.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

Papers of particular interest, published recently, have been highlighted as:

- Of importance

1. Lusis AJ, Attie AD, Reue K. Metabolic syndrome: from epidemiology to systems biology. *Nat Rev Genet.* 2008;9(11):819–30. doi:10.1038/nrg2468.
2. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A.* 2009;106(23):9362–7. *Described GWAS catalog and its utilities.*
3. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* 2011;43(4):333–8. doi:10.1038/ng.784.
4. Ordovas JM, Smith CE. Epigenetics and cardiovascular disease. *Nat Rev Cardiol.* 2010;7(9):510–9. doi:10.1038/nrcardio.2010.104.
5. Ala-Korpela M, Kangas AJ, Inouye M. Genome-wide association studies and systems biology: together at last. *Trends Genet: TIG.* 2011;27(12):493–8. doi:10.1016/j.tig.2011.09.002.



6. Bensimon A, Heck AJ, Aebersold R. Mass spectrometry-based proteomics and network biology. *Annu Rev Biochem.* 2012;81:379–405. doi:10.1146/annurev-biochem-072909-100424.
7. Sarwal MM, Sigdel TK, Salomon DR. Functional proteogenomics—embracing complexity. *Semin Immunol.* 2011;23(4):235–51. doi:10.1016/j.smim.2011.08.002.
8. Grassmann J, Scheerle RK, Letzel T. Functional proteomics: application of mass spectrometry to the study of enzymology in complex mixtures. *Anal Bioanal Chem.* 2012;402(2):625–45. doi:10.1007/s00216-011-5236-4.
9. Morgan AA, Dudley JT, Deshpande T, Butte AJ. Dynamism in gene expression across multiple studies. *Physiol Genomics.* 2010;40(3):128–40. doi:10.1152/physiolgenomics.90403.2008.
10. Dudley JT, Tibshirani R, Deshpande T, Butte AJ. Disease signatures are robust across tissues and experiments. *Mol Syst Biol.* 2009;5:307. doi:10.1038/msb.2009.66.
11. • Kodama K, Horikoshi M, Toda K, Yamada S, Hara K, Irie J, et al. Expression-based genome-wide association study links the receptor CD44 in adipose tissue with type 2 diabetes. *Proc Natl Acad Sci U S A.* 2012;109(18):7049–54. doi:10.1073/pnas.1114513109. *Demonstrated the use of meta-analysis of gene expression data for identification of novel disease-associated genes.*
12. Schulze R. *Meta-analysis: a comparison of approaches.* Toronto: Hogrefe & Huber; 2004.
13. Hong F, Breitling R. A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. *Bioinformatics.* 2008;24(3):374–82. doi:10.1093/bioinformatics/btm620.
14. Moreau Y, Aerts S, De Moor B, De Strooper B, Dabrowski M. Comparison and meta-analysis of microarray data: from the bench to the computer desk. *Trends Genet: TIG.* 2003;19(10):570–7.
15. Chen R, Li L, Butte AJ. AILUN: reannotating gene expression data automatically. *Nat methods.* 2007;4(11):879. doi:10.1038/nmeth1107-879.
16. Butte AJ, Kohane IS. Creation and implications of a phenome-genome network. *Nat Biotechnol.* 2006;24(1):55–62. doi:10.1038/nbt1150.
17. Butte AJ, Chen R. Finding disease-related genomic experiments within an international repository: first steps in translational bioinformatics. *AMIA Annual Symposium proceedings/AMIA Symposium AMIA Symposium.* 2006:106–10.
18. English SB, Butte AJ. Evaluation and integration of 49 genome-wide experiments and the prediction of previously unknown obesity-related genes. *Bioinformatics.* 2007;23(21):2910–7. doi:10.1093/bioinformatics/btm483.
19. Davey Smith G, Ebrahim S. ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol.* 2003;32(1):1–22.
20. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet.* 2005;37(7):710–7. doi:10.1038/ng1589.
21. Wang S, Yehya N, Schadt EE, Wang H, Drake TA, Lusis AJ. Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. *PLoS Genet.* 2006;2(2):e15. doi:10.1371/journal.pgen.0020015.
22. Wang SS, Martin LJ, Schadt EE, Meng H, Wang X, Zhao W, et al. Disruption of the aortic elastic lamina and medial calcification share genetic determinants in mice. *Circ Cardiovasc Genet.* 2009;2(6):573–82. doi:10.1161/CIRCGENETICS.109.860270.
23. Wang SS, Shi W, Wang X, Velky L, Greenlee S, Wang MT, et al. Mapping, genetic isolation, and characterization of genetic loci that determine resistance to atherosclerosis in C3H mice. *Arterioscler Thromb Vasc Biol.* 2007;27(12):2671–6. doi:10.1161/ATVBAHA.107.148106.
24. Clee SM, Yandell BS, Schueler KM, Rabaglia ME, Richards OC, Raines SM, et al. Positional cloning of Sorcs1, a type 2 diabetes quantitative trait locus. *Nat Genet.* 2006;38(6):688–93. doi:10.1038/ng1796.
25. Lee JH, Reed DR, Li WD, Xu W, Joo EJ, Kilker RL, et al. Genome scan for human obesity and linkage to markers in 20q13. *Am J Hum Genet.* 1999;64(1):196–209. doi:10.1086/302195.
26. Lembertas AV, Perusse L, Chagnon YC, Fislis JS, Warden CH, Purcell-Huynh DA, et al. Identification of an obesity quantitative trait locus on mouse chromosome 2 and evidence of linkage to body fat and insulin on the human homologous region 20q. *J Clin Invest.* 1997;100(5):1240–7. doi:10.1172/JCI119637.
27. Kalidas K, Wasson J, Glaser B, Meyer JM, Duprat LJ, White MF, et al. Mapping of the human insulin receptor substrate-2 gene, identification of a linked polymorphic marker and linkage analysis in families with Type II diabetes: no evidence for a major susceptibility role. *Diabetologia.* 1998;41(11):1389–91. doi:10.1007/s001250051081.
28. Chiodini BD, Lewis CM. Meta-analysis of 4 coronary heart disease genome-wide linkage studies confirms a susceptibility locus on chromosome 3q. *Arterioscler Thromb Vasc Biol.* 2003;23(10):1863–8. doi:10.1161/01.ATV.0000093281.10213.DB.
29. Loos RJ. Genetic determinants of common obesity and their value in prediction. *Best Pract Res Clin Endocrinol Metabol.* 2012;26(2):211–26. doi:10.1016/j.beem.2011.11.003.
30. Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med.* 2010;363(2):166–76. doi:10.1056/NEJMr0905980.
31. Maher B. Personal genomes: The case of the missing heritability. *Nature.* 2008;456(7218):18–21. doi:10.1038/456018a.
32. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet.* 2010;42(11):937–48. doi:10.1038/ng.686.
33. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet.* 2010;42(7):579–89. doi:10.1038/ng.609.
34. Yang X, Zhang B, Zhu J. Functional genomics- and network-driven systems biology approaches for pharmacogenomics and toxicogenomics. *Curr Drug Metabol.* 2012;13(7):952–67.
35. • Yang X. Use of functional genomics to identify candidate genes underlying human genetic association studies of vascular diseases. *Arterioscler Thromb Vasc Biol.* 2012;32(2):216–22. doi:10.1161/ATVBAHA.111.232702. *This review summarizes functional genomics resources for vascular diseases.*
36. Kasarskis A, Yang X, Schadt E. Integrative genomics strategies to elucidate the complexity of drug response. *Pharmacogenomics.* 2011;12(12):1695–715. doi:10.2217/pgs.11.115.
37. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, et al. Genetics of gene expression surveyed in maize, mouse and man. *Nature.* 2003;422(6929):297–302. doi:10.1038/nature01434.
38. Tesson BM, Jansen RC. eQTL analysis in mice and rats. *Methods Mol Biol.* 2009;573:285–309. doi:10.1007/978-1-60761-247-6\_16.
39. van Nas A, Ingram-Drake L, Sinsheimer JS, Wang SS, Schadt EE, Drake T, et al. Expression quantitative trait loci: replication, tissue- and sex-specificity in mice. *Genetics.* 2010;185(3):1059–68. doi:10.1534/genetics.110.116087.
40. Meng H, Vera I, Che N, Wang X, Wang SS, Ingram-Drake L, et al. Identification of Abcc6 as the major causal gene for dystrophic cardiac calcification in mice through integrative genomics. *Proc Natl Acad Sci U S A.* 2007;104(11):4530–5. doi:10.1073/pnas.0607620104.
41. Cotsapas C, Speliotes EK, Hatoum IJ, Greenawalt DM, Dobrin R, Lum PY, et al. Common body mass index-associated variants confer risk of extreme obesity. *Hum Mol Genet.* 2009;18(18):3502–7. doi:10.1093/hmg/ddp292.

42. • Greenawalt DM, Dobrin R, Chudin E, Hatoum IJ, Suver C, Beaulaurier J, et al. A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. *Genome Res.* 2011;21(7):1008–16. doi:10.1101/gr.112821.110. *Comprehensive functional genomics study of multiple tissues in a large human cohort.*
43. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet.* 2012;44(9):981–90. doi:10.1038/ng.2383.
44. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, et al. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* 2008;6(5):e107.
45. • Zhong H, Beaulaurier J, Lum PY, Molony C, Yang X, Macneil DJ, et al. Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet.* 2010;6:e1000932. *The first implementation of high level integration of GWAS, functional genomics, and networks to identify T2D genes.*
46. • Zhong H, Yang X, Kaplan LM, Molony C, Schadt EE. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am J Hum Genet.* 2010;86(4):581–91. *A novel methodology for functional genomics-driven pathway analysis of GWAS.*
47. Small KS, Hedman AK, Grundberg E, Nica AC, Thorleifsson G, Kong A, et al. Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat Genet.* 2011;43(6):561–4. doi:10.1038/ng.833.
48. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature.* 2010;466(7307):707–13. doi:10.1038/nature09270.
49. Kjolby M, Andersen OM, Breiderhoff T, Fjorback AW, Pedersen KM, Madsen P, et al. SORT1, encoded by the cardiovascular risk locus 1p13.3, is a regulator of hepatic lipoprotein export. *Cell Metabol.* 2010;12(3):213–23.
50. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature.* 2010;466(7307):714–9.
51. Yang X, Deignan JL, Qi H, Zhu J, Qian S, Zhong J, et al. Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. *Nat Genet.* 2009;41(4):415–23. doi:10.1038/ng.325.
52. • Yang X, Peterson L, Thieringer R, Deignan JL, Wang X, Zhu J, et al. Identification and validation of genes affecting aortic lesions in mice. *J Clin Invest.* 2010;120(7):2414–22. doi:10.1172/JCI42742. *Demonstrated the use of causality test to identify tissue-specific causal genes for atherosclerosis in mouse models.*
53. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet.* 2004;5(2):101–13.
54. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol.* 2005;4:Article17.
55. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinforma.* 2008;9:559. doi:10.1186/1471-2105-9-559.
56. Zhu J, Lum PY, Lamb J, GuhaThakurta D, Edwards SW, Thieringer R, et al. An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet Genome Res.* 2004;105(2–4):363–74.
57. Zhu J, Wiener MC, Zhang C, Fridman A, Minch E, Lum PY, et al. Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. *PLoS Comput Biol.* 2007;3(4):e69. doi:10.1371/journal.pcbi.0030069.
58. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. *J Comput Biol: J Comput Mol Cell Biol.* 2000;7(3–4):601–20. doi:10.1089/106652700750050961.
59. Pe'er D. Bayesian network analysis of signaling networks: a primer. *Sci STKE: Signal Transduct Knowl Environ.* 2005;2005(281):pl4. doi:10.1126/stke.2812005pl4.
60. Pe'er D, Regev A, Elidan G, Friedman N. Inferring subnetworks from perturbed expression profiles. *Bioinformatics.* 2001;17 Suppl 1:S215–24.
61. Schafer J, Strimmer K. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics.* 2005;21(6):754–64. doi:10.1093/bioinformatics/bti062.
62. Shimamura T, Imoto S, Yamaguchi R, Miyano S. Weighted lasso in graphical Gaussian modeling for large gene network estimation based on microarray data. *Genome Info Int Conf Genome Info.* 2007;19:142–53.
63. Toh H, Horimoto K. Inference of a genetic network by a combined approach of cluster analysis and graphical Gaussian modeling. *Bioinformatics.* 2002;18(2):287–97.
64. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinforma.* 2006;7 Suppl 1:S7.
65. Chen Y, Zhu J, Lum PY, Pinto S, MacNeil DJ, et al. Variations in DNA elucidate molecular networks that cause disease. *Nature.* 2008;452(7186):429–35.
66. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, et al. Genetics of gene expression and its effect on disease. *Nature.* 2008;452(7186):423–8.
67. Gargalovic PS, Imura M, Zhang B, Gharavi NM, Clark MJ, Pagnon J, et al. Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. *Proc Natl Acad Sci U S A.* 2006;103(34):12741–6.
68. Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, et al. Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet.* 2006;2(8):e130.
69. Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, et al. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc Natl Acad Sci U S A.* 2006;103(46):17402–7.
70. • Wang IM, Zhang B, Yang X, Zhu J, Stepaniants S, Zhang C, et al. Systems analysis of eleven rodent disease models reveals an inflammatome signature and key drivers. *Mol Syst Biol.* 2012;8:594. doi:10.1038/msb.2012.24. *Demonstrated the power of integrating tissue-specific gene expression profiling and network models to identify key regulators and processes shared across multiple diseases.*
71. Inouye M, Kettunen J, Soinen P, Silander K, Ripatti S, Kumpula LS, et al. Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol Syst Biol.* 2010;6:441. doi:10.1038/msb.2010.93.
72. • Inouye M, Silander K, Hamalainen E, Salomaa V, Harald K, Jousilahti P, et al. An immune response network associated with blood lipid levels. *PLoS Genet.* 2010;6(9). doi:10.1371/journal.pgen.1001113. *Integration of transcriptome, metalome, and network modeling to identify subnetworks that associate with lipid regulation.*
73. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi A-L. Hierarchical Organization of Modularity in Metabolic Networks. *Science.* 2002;297(5586):1551–5. doi:10.1126/science.1073374.
74. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol.* 2007;1:54.
75. Fuller TF, Ghazalpour A, Aten JE, Drake TA, Lusia AJ, Horvath S. Weighted gene coexpression network analysis strategies applied to mouse weight. *Mamm Genome: Off J Int Mamm Genome Soc.* 2007;18(6–7):463–72. doi:10.1007/s00335-007-9043-3.
76. Lum PY, Chen Y, Zhu J, Lamb J, Melmed S, Wang S, et al. Elucidating the murine brain transcriptional network in a segregating

- mouse population to identify core functional modules for obesity and diabetes. *J Neurochem*. 2006;97 Suppl 1:50–62.
77. Zhu J, Zhang B, Schadt EE. A systems biology approach to drug discovery. *Adv Genet*. 2008;60:603–35.
  78. Zhu J, Zhang B, Smith EN, Drees B, Brem RB, Kruglyak L, et al. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet*. 2008;40(7):854–61.
  79. Morine MJ, Toomey S, McGillicuddy FC, Reynolds CM, Power KA, Browne JA, et al. Network analysis of adipose tissue gene expression highlights altered metabolic and regulatory transcriptional activity in high-fat-diet-fed IL-1RI knockout mice. *J Nutr Biochem*. 2012. doi:10.1016/j.jnutbio.2012.04.012.
  80. Jesmin J, Rashid MS, Jamil H, Hontecillas R, Bassaganya-Riera J. Gene regulatory network reveals oxidative stress as the underlying molecular mechanism of type 2 diabetes and hypertension. *BMC Med Genomics*. 2010;3(1):45. doi:10.1186/1755-8794-3-45.
  81. Attie AD, Keller MP. Gene co-expression modules and type 2 diabetes. *Results Probl Cell Differ*. 2010;52:47–56. doi:10.1007/978-3-642-14426-4\_5.
  82. Wang K, Narayanan M, Zhong H, Tompa M, Schadt EE, Zhu J. Meta-analysis of inter-species liver co-expression networks elucidates traits associated with common human diseases. *PLoS Comput Biol*. 2009;5(12):e1000616. doi:10.1371/journal.pcbi.1000616.
  83. Sengupta U, Ukil S, Dimitrova N, Agrawal S. Expression-based network biology identifies alteration in key regulatory pathways of type 2 diabetes and associated risk/complications. *PLoS One*. 2009;4(12):e8100. doi:10.1371/journal.pone.0008100.
  84. Keller MP, Choi Y, Wang P, Davis DB, Rabaglia ME, Oler AT, et al. A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility. *Genome Res*. 2008;18(5):706–16. doi:10.1101/gr.074914.107.
  85. Plaimas K, Eils R, Konig R. Identifying essential genes in bacterial metabolic networks with machine learning methods. *BMC Syst Biol*. 2010;4:56. doi:10.1186/1752-0509-4-56.
  86. Plaisier CL, Horvath S, Huertas-Vazquez A, Cruz-Bautista I, Herrera MF, Tusie-Luna T, et al. A systems genetics approach implicates USF1, FADS3, and other causal candidate genes for familial combined hyperlipidemia. *PLoS Genet*. 2009;5(9):e1000642. doi:10.1371/journal.pgen.1000642.
  87. Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. The Morgan Kaufmann series in representation and reasoning. San Mateo, Calif: Morgan Kaufmann Publishers; 1988.
  88. Madigan DaY J. Bayesian graphical models for discrete data. *Int Stat Rev*. 1995;63:215–32.
  89. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, et al. A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science*. 2003;302(5644):449–53.
  90. Lee I, Date SV, Adai AT, Marcotte EM. A probabilistic functional network of yeast genes. *Science*. 2004;306(5701):1555–8.
  91. Yang X, Zhang B, Molony C, Chudin E, Hao K, Zhu J, et al. Systematic genetic and genomic analysis of cytochrome P450 enzyme activities in human liver. *Genome Res*. 2010;20(8):1020–36. doi:10.1101/gr.103341.109.
  92. • Mori MA, Liu M, Bezy O, Almind K, Shapiro H, Kasif S, et al. A systems biology approach identifies inflammatory abnormalities between mouse strains prior to development of metabolic disease. *Diabetes*. 2010;59(11):2960–71. doi:10.2337/db10-0367. *Integration of tissue-specific gene expression with protein-protein networks to identify key processes that associate with metabolic diseases*.
  93. Greenawalt DM, Sieberts SK, Cornelis MC, Girman CJ, Zhong H, Yang X, et al. Integrating genetic association, genetics of gene expression, and single nucleotide polymorphism set analysis to identify susceptibility Loci for type 2 diabetes mellitus. *Am J Epidemiol*. 2012;176(5):423–30. doi:10.1093/aje/kws123.
  94. • Kang HP, Yang X, Chen R, Zhang B, Corona E, Schadt EE, et al. Integration of disease-specific single nucleotide polymorphisms, expression quantitative trait loci and coexpression networks reveal novel candidate genes for type 2 diabetes. *Diabetologia*. 2012;55(8):2205–13. doi:10.1007/s00125-012-2568-3. *A novel method that integrates top GWAS candidate genes with functional genomics and network models to identify novel disease genes*.
  95. • Dutta P, Courties G, Wei Y, Leuschner F, Gorbato R, Robbins CS, et al. Myocardial infarction accelerates atherosclerosis. *Nature*. 2012;487(7407):325–9. doi:10.1038/nature11260. *Excellent demonstration of the power of systems-level investigation that involves the study of multiple tissues in understanding disease mechanisms*.
  96. Dobrin R, Zhu J, Molony C, Argman C, Parrish ML, Carlson S, et al. Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. *Genome Biol*. 2009;10(5):R55. doi:10.1186/gb-2009-10-5-r55.