Published in final edited form as:
*Semin Nephrol*. 2010 March ; 30(2): 177–184. doi:10.1016/j.semnephrol.2010.01.008.

# - Linking variants from genome wide association analysis to function via transcriptional network analysis

**Benjamin Keller, PhD [Associate Professor]**[1], **Sebastian Martini, MD [Postdoctoral Fellow]**[2], **John Sedor, MD [MetroHealth Research Endowment Professor]**[3], and **Matthias Kretzler, MD [Professor]**[4,*]

[1] Computer Science, Eastern Michigan University

[2] Internal Medicine, University of Michigan

[3] Kidney Disease Research Center, Departments of Medicine and Physiology and Biophysics, Case Western Reserve University, Rammelkamp Center for Research and Education, MetroHealth System Campus

[4] Department of Internal Medicine, Center for Computational Medicine and Biology, University of Michigan

## Abstract

A current challenge in interpretation of Genome Wide Association studies (GWAS) is to establish the mechanistic links between the measured genotype and observed phenotype. The integration of gene expression with disease GWAS is emerging as an important strategy for deciphering these regulatory mechanisms. For renal disease, the availability of both tissue- and disease-specific expression data makes the strategy a compelling option. In this review, three approaches of integrating Single Nucleotide Polymorphism (SNP) genotypes with transcriptional regulation are discussed: (1) interpreting the functional role of transcripts affected by a SNP, (2) identifying the mechanistic role of non-coding SNPs in regulation, and (3) identifying regulatory candidate SNPs with expression associations. Combining these strategies in an integrative manner should allow the discovery of more extensive regulatory information. Linking genetics to systems biology more directly promises the opportunity to explain how genetic variants contribute to disease in a truly holistic manner.

## 1. Introduction

Renal disease presents a unique opportunity to study complex human disease in an integrative manner. With renal biopsy tissue available for molecular analysis, tissue and disease-specific gene expression can be used for the analysis and interpretation of kidney-disease-associated genetic variants, such as single nucleotide polymorphisms (SNPs). In general, a challenge of Genome Wide Association studies (GWAS) is linking the observed genotype to phenotype associations with the molecular mechanisms that drive the associations (see Sedor and Freedman, introduction to this seminar in nephrology issue [1,2]). Measuring changes in genome-

*Corresponding author: Matthias Kretzler MD, Division of Nephrology, Department of Internal Medicine, University of Michigan, 1570 MSRB II, 1150 W. Medical Center Drive, Ann Arbor, Mi 48109-0676, Phone: (734) 764-3157, FAX: (734) 763-0982, Kretzler@med.umich.edu.

wide transcript abundance in glomeruli and tubulointerstitial compartments of kidney biopsies can define the effect(s) of CKD-associated SNPs on transcriptional regulatory mechanisms. Importantly, coupled analyses of gene expression and genetic variants can define the functional pathways disrupted in the disease, addressing a major weakness with many of the GWAS analyses reported to date. GWAS analyses are powered to identify common variants with small effect in patients with common diseases, and the functional consequences of the identified variants have remained unclear.

This review describes strategies by which the genetic etiology of nephropathy could be approached by integrating GWAS and genome-wide expression data. The use of gene expression profiling in the study of renal disease is an area of active investigation [3-5]. However only one publication, a study using kidney function as an indicator of aging [6], integrates renal expression analyses with genetic association analysis. We anticipate such studies to become an integral aspect of the efforts to define the impact of kidney disease risk SNPs and/or other genetic variants on nephropathy phenotypes in the near future. We present three examples in which the regulatory role of nephropathy-associated genetic variations can be determined: (1) using pathway analysis to define the functional role of transcript(s), whose expression level is altered by a SNP, (2) identifying the mechanistic role for non-coding SNPs associated with nephropathy phenotypes, and (3) identifying candidate regulatory SNPs using genome expression data from kidney biopsies as a phenotype(s). We focus on renal disease data, where possible, but draw broadly from many diseases to illustrate opportunities available to researchers in the field. Finally, we discuss how these strategies can be used in an integrative manner to generate more extensive and definitive mechanistic information.

It is important to note that we focus on SNPs, but strategies may vary for other types of genetic variations, such as copy number variants. Also, our focus is on non-coding SNPs because, first, these variants have been the most common output from GWAS analysis, and, second, because their mechanism and function are less well defined than alterations in protein coding regions.

## 2. Possible Relationships between Transcripts and SNPs

The first two strategies we discuss uses sets of putative risk genes, each of which with some relationship to a SNP with a disease-association. These putative risk genes or transcripts, which we will refer to as *candidate* genes or transcripts, have traditionally been identified from a disease GWAS SNP using a strategy of *guilt-by-proximity*, which, in its simplest form, identifies genes within some base-pair distance of the SNP on the chromosome. Ioannidis *et al.*[7] give more detail on candidate selection, but the critical point is that genotyped SNPs are members of a linkage disequilibrium (LD) block, which is defined by the non-random association of genetic variants at two or more loci. A SNP showing a strong association with the disease phenotype is merely a representative of a potentially large group of SNPs and every member of the haplotype block may be the causal variant. Therefore, sets of putative risk genes identified by proximity can potentially be large. However, while LD blocks may be large on certain chromosomes for some populations [8], they are formed by the stochastic properties of recombination and so tend to be bounded by loci where recombination frequently occurs. In addition, these LD relationships are used for genotype imputation [8] prior to association, an analytic tool that reduces the chance of missing the causal SNP but is limited by the known SNPs in the mapping population. Another complication is that guilt-by-proximity is based on the premise that the causal SNP id physically close to the associated variant. However, proximity may also be based three-dimensional DNA structures, such as transient looping due to binding of an enhancer [9].

Recent studies have demonstrated that mRNA expression levels vary in both natural and experimental populations. To explain the genetic basis for the observed variation, Expression

Quantitative Trait Locus (eQTL) mapping characterizes the genomic variation that regulates transcript abundance. eQTLs identify putative regulatory associations that help clarify the relationship between SNPs and gene transcripts (see below for further discussion). However, they are subject to both LD and correlation among expression probes. We use the terms *proximal* and *distal* when referring to SNP-transcript relationships based on sequence distance, and *cis*- and *trans*-associations to refer to proximal and distal relationships with a confirmed regulatory effect. Note that a *cis*- or *trans*-association is not necessarily causal for the disease phenotype. A SNP-gene relationship may be determined either using candidate genes identified via guilt-by-proximity or an eQTL analysis. However, the identification of SNPs affecting distal regulatory regions with eQTL has yet to be addressed in human disease populations in a comprehensive manner.

## 3. Interpreting Impact of Variation on Cellular Function and Disease

The mechanistic effect(s) of a disease-associated SNP may be deduced by understanding the cellular or tissue function of the transcript(s) it regulates. One approach to this task is to characterize a group of disease-associated SNPs with functional and/or pathways analysis to define the mechanisms underlying the associations and link putative risk genes to the disease phenotype. The premise underlying analysis of gene expression and GWAS datasets is that renal disease phenotypes manifest renal tissue-specific changes in transcript regulation, and so the integration of mRNA expression from renal tissue will help focus the analysis. Therefore, in this setting, we can use differential expression between diseased and unaffected tissue to identify regulatory pathways and transcriptional mechanisms in a disease- and tissue specific manner.

We currently envision a sequential strategy (Figure 1) to prioritize candidate transcripts for further validation. First, disease candidate transcripts are screened for their expression in renal tissue compartments. Next, differential regulation of transcripts in tissue compartments involved in disease manifestation could be assessed. After disease regulation is established, two further strategies to integrate these transcripts in a disease context are possible.

The first strategy uses co-regulation to infer functional roles of the candidate transcript(s). Following the strategy of *guilt-by-association*, transcripts strongly co-regulated in both a disease- and tissue compartment-specific manner with the candidate transcript could be used to identify putative functional roles for the transcript. An advantage to this approach is that it does not require *a priori* functional characterization of the candidate transcript, and has the potential to provide initial clues on the regulatory pathways associated with a novel gene. However, the functional mechanism responsible for co-regulation may range from a direct interaction at the molecular level to alteration in the cellular fabric of the tissues studied. Therefore, defining experimental validation for these dependencies can be challenging.

The second strategy is to identify the regulation of transcript dependent pathways. If a functional role of a candidate transcript can be identified from prior biological knowledge, the presence of the transcript in a functional pathway with disease-associated regulation would link this pathway to the disease phenotype. This strategy can be particularly powerful if a transcription factor is targeted by a SNP(s). In this case, promoter regions of differentially regulated mRNAs can be screened for the presence of putative or experimentally validated transcription factor binding sites, which are altered by a GWAS associated SNP. RNAs with the putative binding site and altered expression in disease can help to further define the downstream functional consequences of the genetic variant. Differentially regulated mRNAs targeted by the SNP-associated transcription factor can be evaluated for enrichment in functional categories, which can establish a link to the phenotype studied in the GWAS (for description of promoter modeling approaches employed in these studies see [3,5]).

## 4. Identifying Mechanistic Role of Variations

As discussed earlier, a candidate causal relationship between a variation and a gene is generally defined by sequence proximity. The underlying assumption for this approach is if a variant allele occurs more frequently in (disease) cases than in controls and, at the same time, if changes in gene expression co-occur in the nearby genes, then there is likelihood for causality. The subsequent research required to determine the putative functional roles a given candidate susceptibility gene may have in the onset or progression of, or the protection from a disease is time consuming but necessary for our understanding of the disease pathophysiology. Several caveats make this a difficult task: candidate SNPs might show similar association with expression differences in several nearby genes, making the pre-selection of candidate susceptibility genes difficult. In addition, a variation in a (nearby) gene (partially) responsible for a disease might only affect protein function and not gene expression itself, or it might affect gene expression in a different cell-type/organ or under different (patho)-physiological conditions.

Our focus here is on a strategy to prioritize candidate SNPs proximal to a gene for their potential to have a *cis*-effect. Numerous approaches have been conducted to test whether alleles at polymorphic sites in the 5′ region are able to alter the level of promoter activity. For example, Giacopelli *et al.*[10] characterized three polymorphic variants of the OPN promoter, and showed that the haplotype determined the promoter activity and different alleles led to changes in the transcription factor binding affinity. Predicted transcription factor binding sites were obtained through *in silico* methods and confirmed using Electrophoretic Mobility Shift Assays (EMSA). Consistent with these findings, over-expression of transcription factors with binding capabilities at the SNP-induced alterations in the promoter activity consistent with others, providing evidence that allelic variants alter transcriptional regulation through modification of transcription factor binding sites (recent studies:[11-14]). We illustrate this approach using a GWAS analysis of the Genetics of Kidneys in Diabetes (GoKinD) dataset to define susceptibility genes for diabetic nephropathy (DN) in type 1 diabetes [15] (also see article by Pezzolesi *et al* in this issue). Thirteen SNPs were associated with diabetic nephropathy and were replicated in the Diabetes Control and Complications Trial (DCCT)/Epidemiology of Diabetes Interventions and Complications (EDIC) cohort. Six of those 13 SNPs were located in relative proximity to four genes: *FRMD3, CARS, CPVL/CHN2*. An imputed SNP rs1888747 had the strongest statistical association and is located 1745 basepairs proximal to the first transcription start site (TSR 340242) of the gene *FRMD3*. Current knowledge about FRMD3 is limited, confirming the agnostic approach of GWAS; to date, a PubMed search using FRMD3 as keyword only identifies three papers (search date: November, 2009). FRMD3 belongs to a protein superfamily that shares the highly conserved membrane-association domain FERM and is defined by the erythrocyte membrane-cytoskeleton linker protein band 4.1. FRMD3 has been identified as a putative tumor suppressor gene but has not been reported in a diabetes or renal disease context. With *a priori* information limited, a strategy of integrating both (a) a functional context for FRMD3 in DN, and (b) a theoretical model for transcriptional regulation altered by the SNP rs1888747 could yield sufficient functional insight into the disease-specific regulation of FRMD3 to allow the generation of testable hypothesis for further experimental validation.

Genome wide mRNA expression signatures were obtained from 20 protocol biopsy samples of micro-dissected glomeruli from Pima Indians with type 2 diabetes with and without clinical signs of DN [16]. This population reveals similar minor allele frequencies to the initial GoKinD population used to establish the association (rs1888747 has a MAF in Pima Indians of 0.26, and 0.28 in the GoKinD population (personal communication with A.S. Krolewski and M.G. Pezzolesi)). Using a gene expression data driven approach, transcripts with significant co-regulation with FRMD3 were identified in the DN glomeruli. FRMD3 co-regulated transcripts

in DN showed significant enrichment in canonical pathways with established functions in DN (top scoring pathways in Ingenuity Pathway Analysis Software: BMP, IGF-1 and Insulin).

Next, we performed *in silico* studies to test for the presence of a common, regulatory transcriptional mechanism shared between the BMP-pathway associated and in DN co-regulated genes and *FRMD3*. For this, we used the concept that promoter modules are responsible for tissue- and disease- specific differential regulation. These modules are characterized by a set of two or more transcription factors in the same relative order and spacing working in concert in a defined functional context. This approach has been used to define co-regulated podocyte slit diaphragm specific transcripts and NF-kB dependent networks in DN (for a detailed discussion of this approach see [3,5]).

A systematic computational search of transcriptional frameworks among the promoter regions of the 8 co-regulated BMP pathway members and FRMD3 identified a common module of four transcription factor binding sites. To test the model for enrichment in BMP pathway members, the search window in the promoter region was extended to cover the same distance as seen in the DN associated SNP 1888747 in the FRMD3 promoter (1745 base pairs proximal of the first transcription start site). With this approach, the conserved transcriptional regulatory module was found in the promoter sequence of four additional BMP pathway members (BMPR2, CREBBP, PITX2, RAF1). This example illustrates that the definition of transcriptional regulatory patterns can help prioritize allelic variants for experimental validation of their influence on transcriptional regulation. Based on these *in silico* predictions, strategies for experimental validation can be developed to expand these associations towards conditional dependencies.

## 5. Associating Variation with Regulation

The final strategy we discuss is the association of genetic variation and expression levels, otherwise known as expression Quantitative Trait Loci (eQTL). These associations are computed using standard QTL approaches [17] generally using normalized transcript expression levels, derived from tissue-specific or disease-specific microarray experiments using biologic samples from the genotyped population. Quantitative traits have a long history in plant genetics going back to Mendel, with recent initial efforts in “systems genetics” focused on *Arabidopsis*[18-22]. However, eQTL analysis has been increasingly applied in the interpretation of human disease GWAS (see Cookson *et al.*[17] for examples).

While disease GWAS with a thousand to two thousand subjects tend to be underpowered, tissue- or disease-specific eQTL studies tend to be severely underpowered, sometimes with subjects numbering fewer than one hundred and analysis requiring huge numbers of hypothesis tests that require correction for multiple comparisons [17]. (For instance, an all-against-all association of Illumina RefSeq-8 probes with 2 million imputed SNPs, yields around 46 billion association pairs.) While eQTL power can be improved by considering non-genetic factors affecting expression [23], the power issue is unlikely to be resolved by increasing the number of subjects due to issues with the collection of tissue for expression, and so care is needed in formulating hypotheses. (See [23] for a discussion of power in QTL studies.) One way this challenge has been addressed to perform eQTL analysis over a set of transcripts prioritized by some other approach – the recent publication on kidney aging by Wheeler *et al.*[6] is an example. Another approach is to consider proximal and distal associations separately, and to recognize that *trans*-acting variations will tend to regulate specific biological pathways.

Proximal associations are candidates for *cis*-acting SNPs, though many also affect intronic or downstream enhancers or repressors. Proximal associations are computed using standard QTL algorithms, and proximity is determined by distance from the expression probe, though care

is required in the case of large genes [24]. For instance, we might call proximal any SNP within one kilobase of the probe, which might miss the promoter region for a gene that is longer than a kilobase. For these genes, we need to extend the definition based on the gene (or, better, transcript), and add associations for all probes for that gene. Veyreyas *et al.*[25] also consider a novel strategy of splitting the proximal region of a gene into buckets.

Distal (called *trans* in the literature) associations are those not covered by the definition of proximal. However, trans-SNPs identified by an eQTL often associate with several genes (called a *trans*-hotspot)[19,26]. These hotspots presumably correspond to associations with co-regulated transcripts, suggesting these SNPs affect a, so-called, master regulator of a regulatory subsystem. Therefore, a network-based analysis is appropriate, and approaches that build networks before, during and after the eQTL analysis have all been proposed.

Associations with prior networks are computed by first building a collection of moderately small networks of transcripts, and then computing the association either with an aggregate expression level for each network, or some form of multi-trait analysis [19,20,27]. For instance, we might base our networks on metabolic pathways, in which case we would partition the transcripts first by pathway, and then keep edges that correspond to high positive correlations of expression intensity. We can also build networks based solely on the correlation between expression probes, and use a sufficiently high threshold on the correlation to form subnetworks for the association.

An alternative is for the association algorithm to compute the subnetworks during the computation. For instance, in the algorithm by Kim and Xing [28], the correlation network of expression probes is constructed, given to the algorithm, and the algorithm computes the subnetworks with the best association with the genotype. Lee *et al.*[29] use a similar strategy, but also integrate evidence of functional relationships of transcripts. These approaches are less dependent on the definition of the prior network, and may be more likely to find new relationships.

Posterior networks are computed after the eQTL analysis, and so may suffer from issues of statistical power that do not apply to analyses based on prior networks. Early approaches to *a posteriori* network construction [18,30] use functional annotations to construct putative regulatory networks from the eQTL associations. More recent work (e.g., Aten *et al.*[31] and Kang *et al.*[32]) construct probabilistic causal networks by inferring regulatory relationships among transcripts from expression probe correlation and the strength of SNP-probe associations.

These approaches give us putative regulatory associations that could potentially be used to explain disease phenotype associations with genetic variants identified by GWAS [24,33]. However, naive matching of associations from eQTL analysis and GWAS may not properly capture regulatory causality. Schadt [34] uses a model of causality that relates genotype to transcript *and* disease phenotype in one analysis, and has applied this approach in several studies [35-38]. This model allows the analysis to identify regulatory associations that also are putatively causal, which can better focus follow-up analysis and confirmatory experimentation. The approach of systems genetics would similarly allow identification of regulatory associations that also explain QTL for protein abundance or metabolite levels [22] as a further step toward understanding associations with the molecular mechanism of the disease phenotype.

## 6. Conclusions

The integration of gene expression with disease GWAS is an important strategy to deciphering the regulatory mechanisms leading to disease phenotypes. For renal disease, the availability of both tissue- and disease-specific expression makes the strategy even stronger.

Though we have described three distinct strategies, we expect the application of a holistic strategy that captures aspects of all approaches discussed depending on the sources of the different kinds of data. A situation in which disease phenotype, expression and genotype are all available from the same individuals presents the optimal opportunity to use a computational strategy as described by Schadt [34], followed by analysis of proximal associations to prioritize those more likely to have a *cis*-effect, and by pathway analysis of affected transcripts to identify putative functional roles. In the case of the integration of data obtained from different populations (e.g., expression data obtained from one population being analyzed with genotypes obtained from an independent population), this strategy will need to be relaxed and, for instance, combine tissue-specific eQTL analysis with GWAS with suitable statistics from different populations.

In addition, more sophisticated models will need to be developed that are open to the integration of additional measures of regulatory activity, such as miRNA levels, post-transcriptional regulation, as well as epigenetic modifications. Beyond capturing regulation better, we can also look forward to the use of additional quantitative measurements of cellular, tissue or disease systems such as protein abundance and metabolites. Linking genetics to systems biology more directly promises to help explain how genetic variants contribute to disease in a truly holistic manner.

## Acknowledgments

We would like to thank Viji Nair for the generation of the DN associated transcriptional pathways, Markus Pezzolezi and Andrzej Krolewski for provision of the GoKind candidate SNPs and Sudha Iyengar for the provision of the FIND candidate transcriptional targets.

Supported in part by NIH grants U54 DA021519, P30 DK081943, R21 DK079441 and DK-064719.

## References

1. Sedor JR, Freedman BI. Introduction: Genetics and the Kidney: Promise, Potential and Challenges. Seminars in Nephrology. 2010 in press.
2. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. Nat Rev Genet 2009;10(4):241–251. [PubMed: 19293820]
3. Schmid H, Boucherot A, Yasuda Y, et al. Modular activation of nuclear factor-kappaB transcriptional programs in human diabetic nephropathy. Diabetes 2006;55(11):2993–3003. [PubMed: 17065335]
4. Yasuda Y, Cohen CD, Henger A, Kretzler M. Gene expression profiling analysis in nephrology: towards molecular definition of renal disease. Clin Exp Nephrol 2006;10(2):91–98. [PubMed: 16791393]
5. Cohen CD, Klingenhoff A, Boucherot A, et al. Comparative promoter analysis allows de novo identification of specialized cell junction-associated proteins. Proc Natl Acad Sci U S A 2006;103(15): 5682–5687. [PubMed: 16581909]
6. Wheeler HE, Metter EJ, Tanaka T, et al. Sequential use of transcriptional profiling, expression quantitative trait mapping, and gene association implicates MMP20 in human kidney aging. PLoS Genetics 2009;5(10):e1000685. [PubMed: 19834535]
7. Ioannidis J, Thomas G, Daly M. Validating, augmenting and refining genome-wide association signals. Nature Reviews Genetics 2009;10(5):318–329.
8. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet 2009;10:387–406. [PubMed: 19715440]

9. Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. Nature 2009;461 (7261):199. [PubMed: 19741700]
10. Giacopelli F, Marciano R, Pistorio A, et al. Polymorphisms in the osteopontin promoter affect its transcriptional activity. Physiol Genomics 2004;20(1):87–96. [PubMed: 15479859]
11. Zhong L, Dong-Hai Q, Hong-Ying L, Qing-Feng L. Analysis of the nicastrin promoter rs10752637 polymorphism and its association with Alzheimer's disease. Eur J Neurosci. 2009
12. Moore LE, Brennan P, Karami S, et al. Apolipoprotein E/C1 locus variants modify renal cell carcinoma risk. Cancer Res 2009;69(20):8001–8008. [PubMed: 19808960]
13. Wu X, Ye Y, Kiemeney LA, et al. Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer. Nat Genet 2009;41(9):991–995. [PubMed: 19648920]
14. Wang L, Lu X, Li Y, et al. Functional analysis of the C-reactive protein (CRP) gene - 717A>G polymorphism associated with coronary heart disease. BMC Med Genet 2009;10:73. [PubMed: 19624831]
15. Pezzolesi MG, Poznik GD, Mychaleckyj JC, et al. Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes. Diabetes 2009;58(6):1403–1410. [PubMed: 19252134]
16. Berthier CC, Zhang H, Schin M, et al. Enhanced expression of Janus kinase-signal transducer and activator of transcription pathway members in human diabetic nephropathy. Diabetes 2009;58(2): 469–477. [PubMed: 19017763]
17. Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. Nature Reviews Genetics 2009;10(3):184–194.
18. Keurentjes JJ, Fu J, Terpstra IR, et al. Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proceedings of the National Academy of Sciences of the United States of America 2007;104(5):1708–1713. [PubMed: 17237218]
19. Kliebenstein D. Quantitative genomics: analyzing intraspecific variation using global gene expression polymorphisms or eQTLs. Annual Review of Plant Biology 2009;60:93–114.
20. Kliebenstein D. Quantification of variation in expression networks. Methods in molecular biology (Clifton, NJ) 2009;553:227–245.
21. Kliebenstein D, West MA, van Leeuwen H, et al. Identification of QTLs controlling gene expression networks defined a priori. BMC Bioinformatics 2006;7:308. [PubMed: 16780591]
22. Rowe H, Hansen B, Halkier B, Kliebenstein D. Biochemical networks and epistasis shape the Arabidopsis thaliana metabolome. The Plant cell 2008;20(5):1199–1216. [PubMed: 18515501]
23. Stegle O, Kannan A, Durbin R, Winn J. Accounting for non-genetic factors improves the power of eqtl studies. Lecture Notes in Computer Science 2008;4955:411.
24. Dimas A, Deutsch S, Stranger B, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 2009;325(5945):1246–1250. [PubMed: 19644074]
25. Veyrieras J, Kudaravalli S, Kim S, et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genetics 2008;4(10):e1000214. [PubMed: 18846210]
26. Ye C, Galbraith S, Liao J, Eskin E. Using network component analysis to dissect regulatory networks mediated by transcription factors in yeast. PLoS Computational Biology 2009;5(3):e1000311. [PubMed: 19300475]
27. Wentzell A, Rowe H, Hansen B, et al. Linking Metabolic QTLs with Network and cis-eQTLs Controlling Biosynthetic Pathways. PLoS Genetics 2007;3(9):e162.
28. Kim S, Xing E. Statistical estimation of correlated genome associations to a quantitative trait network. PLoS Genetics 2009;5(8):e1000587. [PubMed: 19680538]
29. Lee S, Dudley A, Drubin D, et al. Learning a prior on regulatory potential from eQTL data. PLoS Genetics 2009;5(1):e1000358. [PubMed: 19180192]
30. Lan H, Chen M, Flowers JB, et al. Combined expression trait correlations and expression quantitative trait locus mapping. PLoS Genetics 2006;2(1):e6. [PubMed: 16424919]
31. Aten J, Fuller T, Lusis A, Horvath S. Using genetic markers to orient the edges in quantitative trait networks: the NEO software. BMC Systems Biology 2008;2:34. [PubMed: 18412962]

32. Kang EY, Shpitser I, Ye C, Eskin E. Detecting the Presence and Absence of Causal Relationships between Expression of Yeast Genes with Very Few Samples. RECOMB 2009, LNCS 5541 2009:466–481.

33. Dimas A, Stranger B, Beazley C, et al. Modifier effects between regulatory and protein-coding variation. PLoS Genetics 2008;4(10):e1000244. [PubMed: 18974877]

34. Sieberts S, Schadt E. Moving toward a system genetics view of disease. Mammalian genome : official journal of the International Mammalian Genome Society 2007;18(67):389–401. [PubMed: 17653589]

35. Chen Y, Zhu J, Lum P, et al. Variations in DNA elucidate molecular networks that cause disease. Nature 2008;452(7186):429–435. [PubMed: 18344982]

36. Dobrin R, Zhu J, Molony C, et al. Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. Genome biology 2009;10(5):R55. [PubMed: 19463160]

37. Emilsson V, Thorleifsson G, Zhang B, et al. Genetics of gene expression and its effect on disease. Nature 2008;452(7186):423–428. [PubMed: 18344981]

38. Schadt E, Molony C, Chudin E, et al. Mapping the genetic architecture of gene expression in human liver. PLoS Biology 2008;6(5):e107. [PubMed: 18462017]

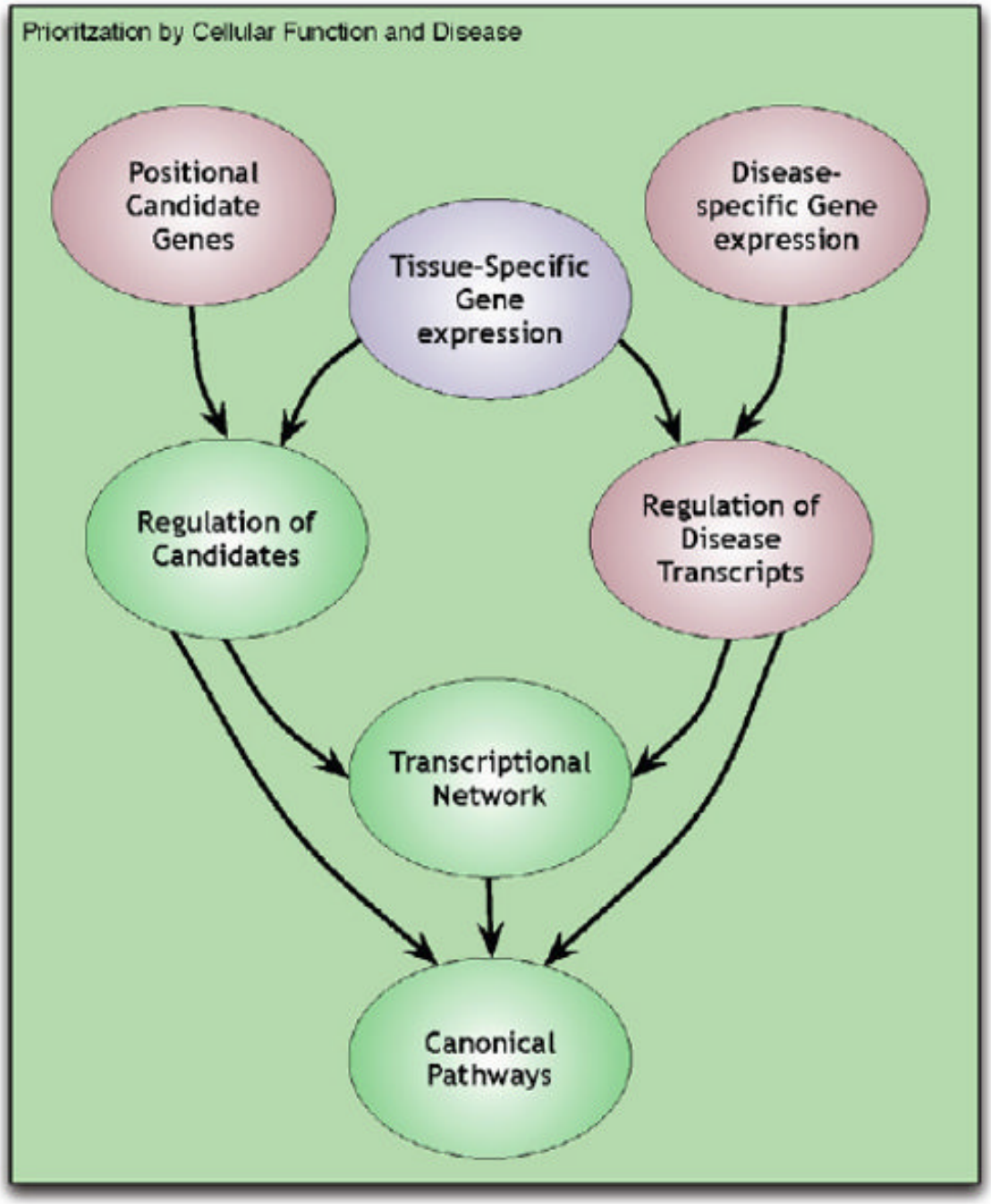


**Figure 1. Linking Genome Wide Association Studies to phenotypes via transcriptional network analysis**

Positional candidate genes are identified in GWAS. Disease specific gene expression alterations are defined by genome wide mRNA expression profiling of the tissue manifesting the GWAS phenotype.

GWAS candidates can be prioritized for further analysis using disease specific regulation of the candidate gene.

The functional impact of the candidate gene on the disease phenotype can be assessed by defining the disease-associated regulation of canonical pathways with established relevance of the candidate gene (*a priori* network). Co-regulation and transcriptional network modeling

approaches can be used to established *de novo* disease specific networks potentially impacted by the candidate gene.