

## Genetic Association–Guided Analysis of Gene Networks for the Study of Complex Traits

Casey S. Greene, PhD; Daniel S. Himmelstein, BS

The wealth of data measuring genetic variations and complex phenotypes in large cohorts provides a substantial resource, but it also raises significant challenges. The number of loci with strong statistical support for an association with one or more phenotypes has grown, but we are still not able to account for the genetic basis of many common phenotypes. New methods use gene networks to identify additional genes associated with phenotypes. Here we briefly review recently developed methods and categorize them by their analytic approach. We also highlight a factor that can confound network-based approaches: genes with more measured single-nucleotide polymorphisms (SNPs) tend to also be more connected across many types of networks. Finally, we highlight new techniques that use nominally significant, as opposed to genome-wide significant, associations to guide the analysis of functional relationship networks. In contrast with network-based methods that use documented associations to reallocate weight or to adjust  $P$  values, these methods do not require documented associations with the underlying phenotype. We provide an example of such techniques, the network-wide association study (NetWAS), and discuss how such methods complement the analytic toolkit available to modern geneticists. We conclude by discussing remaining challenges in the field and new application areas for network-based methods for analysis of association data.

### Challenges for Genome-Wide Association Studies

In a landmark study in 2005, Klein et al<sup>1</sup> reported the discovery of a variant in the complement factor H gene that was associated with age-related macular degeneration via a genome-wide association study (GWAS). Since that time, GWAS has been used to identify variants associated with numerous traits from complex diseases<sup>2</sup> to preferences for cilantro.<sup>3</sup> Recent work has demonstrated that GWAS discoveries provide a fruitful means to reposition drugs<sup>4</sup> and that the presence of a GWAS hit in a drug target is positively associated with the chance that a drug will progress through clinical trials and ultimately be approved.<sup>5</sup> Each GWAS produces a list of phenotype association  $P$  values for each variant, or gene for gene-based tests,<sup>6–10</sup> and variants or genes with  $P$  values below a certain threshold are considered associated with the phenotype.

Although GWASs have identified numerous genetic variants associated with diverse phenotypes, these variants generally do not explain the majority of phenotypic variance suspected to be genetic.<sup>11</sup> The unexplained genetic component of complex traits has been termed the missing heritability,<sup>12</sup> and there are numerous potential explanations for it. These explanations include gene–gene interactions,<sup>13,14</sup> heterogeneity (allelic<sup>15</sup> and phenotypic<sup>16</sup>), rare variants of strong effect,<sup>17</sup> parent of origin,<sup>18</sup> transgenerational,<sup>19</sup> and gene–environment effects,<sup>20,21</sup> in which an environmental trigger induces risk in specific individuals with certain genetic variants. These factors can contribute to make a study that is adequately powered for common variants of reasonable effect size underpowered, leading to an unexpected increase in the false-negative rate (ie, the proportion of true associations that are not discovered). The practice of requiring both a highly stringent significance threshold and replication for publication of association, while reducing false positives substantially, can further compromise the ability of GWASs to identify associated variants and genes.<sup>22</sup> This practice leads to many true associations falling into the nominally but not genome-wide significance range.<sup>22</sup> To extract these associations, we need methods capable of using orthogonal sources of information, for example, pathways or biological networks, to effectively reanalyze GWAS results.

### Pathway and Network Approaches Address the Gene-Prioritization Challenge by Overlaying Additional Data

Data mining approaches have been effectively applied to large-scale gene expression data using pathway and other expert knowledge to identify genes that participate in complex biological processes.<sup>23–26</sup> Given the context of missing heritability, the community has begun to develop and apply computational methods that use pathway or network information to extract relevant variants and genes that may be hidden within associations that are nominally significant but that do not reach a genome-wide significance threshold. These approaches can be divided into pathway-based approaches and network-based approaches.

Pathway-based methods analyze GWAS results in the context of curated gene sets made up of pathways from the Kyoto

Received September 11, 2015; accepted March 8, 2016.

From the Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia (C.S.G.); and Biological and Medical Informatics, University of California, San Francisco (D.S.H.).

Correspondence to Casey S. Greene, PhD, University of Pennsylvania, 10–131 SCTR, 3400 Civic Center Blvd, Philadelphia, PA 19104. E-mail csgreene@upenn.edu

(*Circ Cardiovasc Genet*. 2016;9:179–184. DOI: 10.1161/CIRCGENETICS.115.001181.)

© 2016 American Heart Association, Inc.

*Circ Cardiovasc Genet* is available at <http://circgenetics.ahajournals.org>

DOI: 10.1161/CIRCGENETICS.115.001181

Encyclopedia of Genes and Genomes,<sup>27</sup> the Gene Ontology,<sup>28</sup> or other pathway resources.<sup>29,30</sup> Pathway-based methods,<sup>31–33</sup> which have been recently reviewed,<sup>34,35</sup> aim to identify pathways associated with the phenotype or disease of interest.<sup>36</sup> A primary strength of this class of methods is the interpretability of the results: most produce a list of pathways associated with a phenotype of interest. Although it may be challenging to identify a drug that will be efficacious directly from a single genetic variant or gene, the pathways identified by these approaches may suggest targeted agents or drug repositioning opportunities.<sup>37,38</sup> A primary limitation of these approaches is that, for most resources, pathways and gene sets are constructed based on curation of published literature. This places limitations on the breadth of curations and diversity of curations because annotations do not represent a random selection of true biological relationships.<sup>39</sup> In addition, annotation of genes to multiple related pathways can lead to a small set of genes driving an association for multiple distinct pathways that are difficult to untangle.<sup>40</sup> Finally, the pathway-based approaches use statistics based on gene sets and, hence, do not consider connections between genes, simply membership within the pathway.

In contrast with pathway-based approaches, network-based methods incorporate connectivity measures between genes, either from curated resources, interaction databases, or integrative resources that combine multiple data sources and types. The intuition behind the use of network-based approaches is simple: because genes do not act in isolation, we expect to observe associated genes either connected to or participating in common phenotype-specific subnetworks.<sup>41</sup> The means of encoding this expectation into an underlying analytical method varies by method.<sup>42,43</sup> In general, there are 2 types of approaches: (1) stringent evidence methods that identify genes or variants within neighborhoods with genes that possess a documented association to the phenotype; or (2) permissive evidence methods that identify network neighborhoods with an overabundance of associated variants or genes. Although these are the predominant groups that we focus on, some techniques generate scores independent of disease information using network topology, which are combined in a post hoc step with GWAS associations.<sup>44</sup> The stringent and permissive evidence methods both use disease or phenotype-specific associations to guide an analysis of networks to identify genetic factors underlying the phenotype of interest.

Network-based approaches can incorporate diverse types of networks. Example types of biological networks that are available include protein–protein interaction networks,<sup>45,46</sup> miRNA target networks,<sup>47,48</sup> transcription factor networks,<sup>49,50</sup> and shared function networks.<sup>51,52</sup> These networks can be either curated by experts from the literature or inferred based on available data. The benefit of curated networks is that they are expected to provide high fidelity for the edges that exist, whereas the benefit of inferred networks is that they are expected to provide a more complete representation of the underlying biological networks. A recent analysis revealed that individual network types exhibited varying performance for the prediction of gene–disease associations.<sup>53</sup> Networks that captured common directional responses to perturbations exhibited the highest mean performance across phenotypes

for single-type networks, but an integrated model provided the highest overall performance.<sup>53</sup>

### Stringent Evidence Methods

For stringent evidence methods, high-confidence gene–phenotype associations are used as a building block to identify promising network regions. In some cases, these associations are derived from literature or database support.<sup>54–56</sup> In other cases, these may be derived from a set of hits from one or more GWASs that passed a stringent statistical threshold, indicating genome-wide significance.<sup>51,53,57</sup> In each case, the objective of these algorithms is to identify network neighborhoods associated with a high density of prior disease annotations. Depending on the approach, an implicit or explicit assumption of guilt-by-association then allows new genes and variants to be prioritized based on their own connectivity patterns.

For example, a recently developed approach<sup>53</sup> uses stringent evidence of associations, collected from the GWAS catalog,<sup>58</sup> to integrate multiple types of networks. The intuition behind these approaches is that no single source of data is likely to contain sufficient information to accurately predict gene–disease relationships. In this scenario, integrating multiple networks with different types of information can improve performance beyond that achievable with single networks. Integrating and appropriately weighing these diverse evidence types presents a challenge. To address this, the authors used machine learning to automatically learn the relative importance of multiple biological networks. They demonstrate that this integration dramatically outperforms individual networks.

### Permissive Evidence Methods

The second type of method analyzes association results, including those which do not meet genome-wide significance, from an individual study in the context of existing networks. Because the information used is from a single study, these methods are easier to evaluate because concordance with documented associations is unlikely to be driven by knowledge bias specific to the phenotype. To identify genes associated with evidence for an association with the disease, these methods have used permissive significance thresholds.<sup>59–62</sup>

An example of this type of approach is protein interaction network–based pathway analysis.<sup>59,60</sup> Protein interaction network–based pathway analysis identifies gene modules that have an enrichment of genes with nominally significant *P* values. Once these modules are identified, the biological significance of modules can be evaluated through gene set enrichment analysis, for example, by identifying Gene Ontology terms over-represented in each module. The recently developed integrative protein–interaction–network–based pathway analysis extension adds a step that allows signal to diffuse from enriched nodes along network edges before modules are identified.<sup>61</sup>

### New Methods Integrate Elements of Both Strategies

An example of a hybrid approach that incorporates elements from both types of approaches is the NetWAS.<sup>62</sup> NetWAS uses techniques common to stringent evidence algorithms, but performs an analysis that uses permissive evidence. Specifically,

NetWAS has 2 analytic similarities to stringent evidence methods: NetWAS uses tissue-specific networks in re-prioritization, which have demonstrated consistently better performance than tissue-naïve networks,<sup>56,62</sup> and NetWAS, like the heterogeneous network method,<sup>53</sup> uses a machine learning strategy to identify network characteristics predictive of disease relevance. NetWAS learns which genes in the network are connected to the associations identified by a study. Consistent with permissive evidence methods, NetWAS performs an analysis based on the patterns of association observed within an individual GWAS instead of a summarization of discovered associations. The final outcome of a NetWAS analysis is the generation of a network-based ranking of all genes to disease based on connectivity in the selected network. This hybrid approach can be used to reanalyze a GWAS in the context of a network specific to a tissue of interest to produce a prioritized list of candidates with both tissue and phenotype specificity.

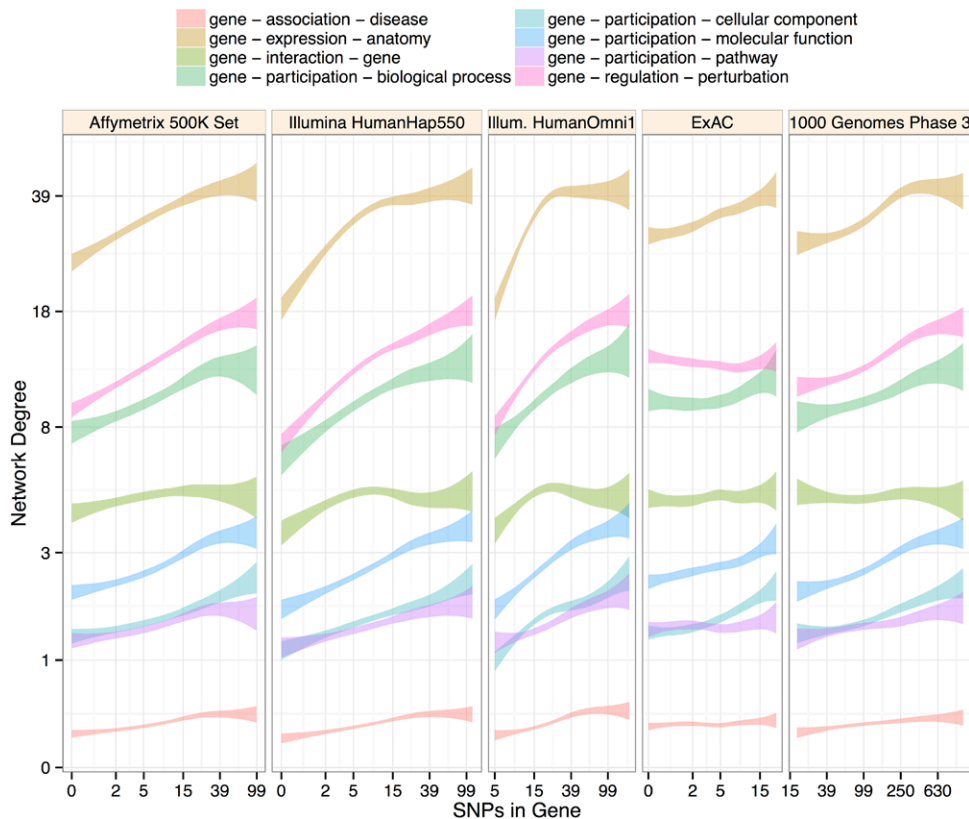
### Potential Confounding From Varying SNP Abundances

Network-based approaches provide an important new avenue to analyze genetic association data to reveal the genetic basis of common diseases. It is important for users and developers of these methods to be aware that there are factors that correspond between GWAS data and networks, which can act as confounding factors for some network-based approaches.

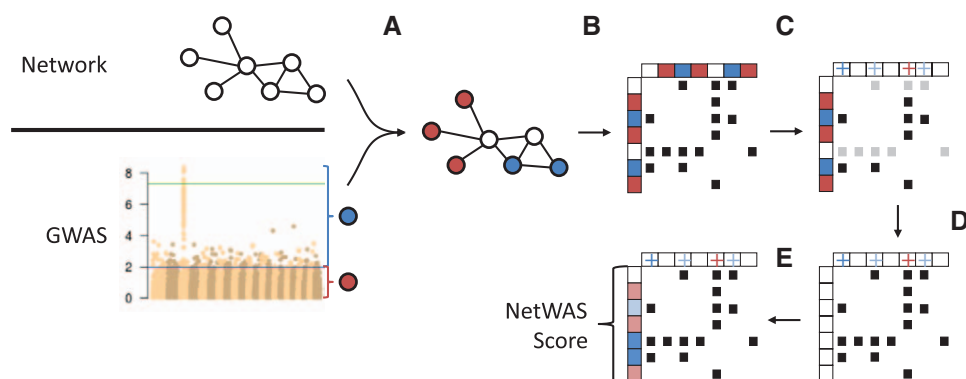
Pioneering approaches collapsed variant-level  $P$  values to gene-level  $P$  values by using only the minimum  $P$  value for variants within each gene.<sup>51,59</sup> This process leads to genes with more measured variants receiving lower  $P$  values, potentially for no reason other than the gene's size.<sup>63</sup> This confounding effect is compounded because genes with more measured variants also tend to be more connected in networks that are commonly used by these methods (Figure 1; code released into the public domain at <https://github.com/dhimmel/snplentiful><sup>65</sup>). Fortunately, the development of gene-based<sup>6-10</sup> methods that consider the nonuniform mapping of measured variants to genes has alleviated this challenge. When applying network and pathway-based methods, it is important to use algorithms that either use gene- or pathway-based tests or permutation of case-control status to account for this feature of the data. The example approach that we examine in depth in the next section, NetWAS, uses results from gene-based association tests to address this issue.

### An Illustration of Network-Based Methods Using NetWAS As an Example

NetWAS<sup>62</sup> operates by identifying phenotype-associated patterns in biological networks. The uncovered patterns are then used to rank each gene for association with the phenotype. The specific steps of NetWAS are to identify genes with a nominal association; to use those genes to guide a machine



**Figure 1.** Genes with more SNPs tend to have higher network degree. The number of SNPs per gene was calculated for 3 genotyping arrays (Affymetrix 500K Set, Illumina HumanHap550, Illumina HumanOmni1), exome sequencing (ExAC), and whole genome sequencing (1000 Genomes Phase 3). The network degree (number of edges) for each gene was calculated on hetio-ind,<sup>64</sup> a network containing multiple types of nodes and edges. Models drawn as 95% confidence bands show the relationship between SNP abundance and network connectivity for 8 types of edges. For most edge types and platforms, genes with more measured SNPs tend to be more connected.



**Figure 2.** Schematic diagram of the Network-wide Association Study (NetWAS) procedure. In step A, nominally significant genome-wide association study (GWAS) hits (above blue line) are combined with a tissue-specific network to create a labeled network (blue, nominally significant; red, not nominally significant; white, not measured). In step B, this network is converted to a gene–gene adjacency matrix (dark squares indicate edge from the network). The adjacency matrix is symmetrical. In step C, a weight is calculated for each gene related to the extent to which it interacts with positive genes using a machine learning algorithm, for example, support vector machines. Scores can range from specifically interacting with positives (high positive weight, dark blue plus) to specifically interacting with negatives (high negative weight, dark red plus). In step D, labels are removed and only weights associated with columns are considered. In step E, genes represented as rows are scored based on the weights of the genes that they are connected to. This generates degrees of positive scores (blue) or negative scores (red). These scores are the output of the NetWAS procedure.

learning analysis of tissue-specific networks; and to use the results of the network analysis to rank all genes in the network based on the model constructed by machine learning (Figure 2 and discussed in detail below).

In the first step of NetWAS, a gene-based test is applied to convert SNP association  $P$  values into scores for each gene. The goal of this step is to generate a positive set of genes that are enriched for true associations. In Greene et al,<sup>62</sup> the versatile gene-based association study<sup>6</sup> method was used, but any gene-based test that effectively controls for the number of variants in genes can be applied. Genes are then selected based on a lenient statistical threshold, for example,  $P < 0.01$  in Greene et al, but genes could alternatively be selected based on a permissive false discovery rate threshold. A negative set is constructed consisting of genes that show no evidence of association. The goal of the negative set is to define the universe of genes that were measured but not identified as significant, and so in practice, this could be set to encompass only genes with little evidence for association (eg,  $P \geq 0.2$ ). In Greene et al, this was constructed as simply the complement of the positive set (genes having  $P \geq 0.01$ ). The negative set allows NetWAS to be readily applied to platforms that are not genome-wide, for example, the Immunochip<sup>66</sup> or MetaboChip<sup>67</sup> platforms. Once the positive and negative sets are constructed, they are overlaid on a selected network (Figure 2A).

The 2 sets of genes, nominally associated positives and unassociated negatives, are used to guide an analysis of tissue-specific networks. To perform this analysis, the network is converted into an adjacency matrix, as depicted in Figure 2B. This analysis uses weighted networks, but for illustrative purposes, binary networks (connected, unconnected) are depicted in Figure 2. A machine learning algorithm is then applied to derive weights associated with connectivity to each gene (Figure 2C). In Greene et al, a support vector machine algorithm was applied, though the framework is amenable to any machine learning–based classifier. The algorithm identifies nodes in the network that tend to be either more or less

connected to genes with a nominally significant association than to unassociated genes. From this point forward, the status of significant/nonsignificant in the GWAS is no longer considered. A predictor is constructed by using the weights assigned to each node by the machine learning algorithm (Figure 2D).

In the final stage of NetWAS, the predictor trained in the second step is applied to all genes (Figure 2E). The predictor contains a weight for each gene that reflects the extent to which it is connected primarily to nominally significant, as opposed to unassociated, genes. The values of each gene's edges to each other gene are then multiplied by these weights and summed to produce a prediction for that gene (Figure 2, NetWAS Score). This prediction captures the extent to which the gene's network connectivity indicates consistency with the nominally significant set.

Focusing specifically on cardiovascular phenotypes, a NetWAS-predicted gene set outperformed the underlying GWAS on key measures in an analysis of multiple phenotypes related to hypertension.<sup>40</sup> Specifically, for each phenotype, NetWAS ranked genes with a documented role in hypertension more highly than the corresponding GWAS; it ranked genes annotated to hypertension-specific Gene Ontology<sup>28</sup> processes more highly than the GWAS; and it ranked genes targeted by antihypertensive drugs more highly than the GWAS. In addition, the NetWAS top-ranked genes exhibited literature support for involvement in hypertension. Although our discussion here focuses on cardiovascular genetics, an analysis of publicly available data revealed strong performance across GWASs of multiple phenotypes.<sup>62</sup>

## Conclusions

Although GWAS has not fully revealed the genetic basis of common human disease, it is now becoming clear that these data can be useful in a large-scale data mining framework. Network-based methods provide a powerful means to identify the mechanistic basis of complex phenotypes from GWAS results. To take advantage of this resource, we will need to

continue to develop, evaluate, and apply algorithms that take into account the complexities of the underlying phenotypes in one or more tissues.

In addition to analytical methods that leverage the multi-tissue to phenotype mapping, methods that incorporate related phenotypes also represent an area of considerable potential. Phenome-wide association studies<sup>68</sup> have complemented GWAS by mapping the associations of variants across multiple phenotypes. Results with NetWAS for distinct hypertension-related phenotypes revealed that performance aggregated across multiple phenotypes outperformed single NetWASs. This suggests that phenome-wide application of network-based methods represent a promising area for new algorithm development and applications. Methods capable of integrating multiple related phenotypes simultaneously to identify a common genetic basis could improve power and help to disentangle the genetic basis of complex diseases.

Although approaches developed to date have delivered promising results that have added value to existing GWAS, we anticipate that this active area of research will continue to advance. We expect that methods that improve our ability to mine these networks, methods that integrate across multiple tissues or cell lineages, and methods that integrate across multiple phenotypes will each contribute to advances in our understanding of the genetic basis of complex cardiovascular phenotypes. We expect that combined approaches that integrate across multiple aspects simultaneously will also provide new opportunities to discover the genetic basis of complex cardiovascular traits.

### Acknowledgments

We are strong supporters of CC-BY licensing; unfortunately, this option was not available for this article, but we appreciate the journal for making the article free for readers.

### Sources of Funding

This work was funded by the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative through Grant GBMF4552 to Dr Greene. This material is based on work supported by the National Science Foundation under Grant No 1144247 to D.S. Himmelstein. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article.

### Disclosures

None.

### References

- Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005;308:385–389. doi: 10.1126/science.1109557.
- Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661–678.
- Eriksson N, Wu S, Do CB, Kiefer AK, Tung JY, Mountain JL, et al. A genetic variant near olfactory receptor genes influences cilantro preference. *Flavour*. 2012;1:22. doi: 10.1186/2044-7248-1-22.
- Sanseau P, Agarwal P, Barnes MR, Pastinen T, Richards JB, Cardon LR, et al. Use of genome-wide association studies for drug repositioning. *Nat Biotechnol*. 2012;30:317–320. doi: 10.1038/nbt.2151.
- Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47:856–860. doi: 10.1038/ng.3314.
- Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, et al; AMFS Investigators. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet*. 2010;87:139–145. doi: 10.1016/j.ajhg.2010.06.009.
- Huang H, Chanda P, Alonso A, Bader JS, Arking DE. Gene-based tests of association. *PLoS Genet*. 2011;7:e1002177. doi: 10.1371/journal.pgen.1002177.
- Mishra A, Macgregor S. VEGAS2: Software for more flexible gene-based testing. *Twin Res Hum Genet*. 2015;18:86–91. doi: 10.1017/thg.2014.79.
- Pedroso II, Barnes MR, Lourdasamy A, Al-Chalabi A, Breen G. FORGE: multivariate calculation of gene-wide p-values from genome-wide association studies. *bioRxiv*. 2015. doi: 10.1101/023648.
- Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X. Sequence kernel association tests for the combined effect of rare and common variants. *Am J Hum Genet*. 2013;92:841–853. doi: 10.1016/j.ajhg.2013.04.015.
- Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet*. 2012;90:7–24. doi: 10.1016/j.ajhg.2011.11.029.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747–753. doi: 10.1038/nature08494.
- Phillips PC. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet*. 2008;9:855–867. doi: 10.1038/nrg2452.
- Greene CS, Penrod NM, Williams SM, Moore JH. Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS One*. 2009;4:e5639. doi: 10.1371/journal.pone.0005639.
- Zhang G, Karns R, Sun G, Indugula SR, Cheng H, Havas-Augustin D, et al. Finding missing heritability in less significant Loci and allelic heterogeneity: genetic variation in human height. *PLoS One*. 2012;7:e51211. doi: 10.1371/journal.pone.0051211.
- Manchia M, Cullis J, Turecki G, Rouleau GA, Uher R, Alda M. The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. *PLoS One*. 2013;8:e76295. doi: 10.1371/journal.pone.0076295.
- Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol*. 2010;8:e1000294. doi: 10.1371/journal.pbio.1000294.
- Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, et al; DIAGRAM Consortium. Parental origin of sequence variants associated with complex diseases. *Nature*. 2009;462:868–874. doi: 10.1038/nature08625.
- Nadeau JH. Transgenerational genetic effects on phenotypic variation and disease risk. *Hum Mol Genet*. 2009;18(R2):R202–R210. doi: 10.1093/hmg/ddp366.
- Thomas D. Gene–environment-wide association studies: emerging approaches. *Nat Rev Genet*. 2010;11:259–272. doi: 10.1038/nrg2764.
- Zhang R, Chu M, Zhao Y, Wu C, Guo H, Shi Y, et al. A genome-wide gene–environment interaction analysis for tobacco smoke and lung cancer susceptibility. *Carcinogenesis*. 2014;35:1528–1535. doi: 10.1093/carcin/bgu076.
- Williams SM, Haines JL. Correcting away the hidden heritability. *Ann Hum Genet*. 2011;75:348–350. doi: 10.1111/j.1469-1809.2011.00640.x.
- Hibbs MA, Myers CL, Huttenhower C, Hess DC, Li K, Caudy AA, et al. Directing experimental biology: a case study in mitochondrial biogenesis. *PLoS Comput Biol*. 2009;5:e1000322. doi: 10.1371/journal.pcbi.1000322.
- Ju W, Greene CS, Eichinger F, Nair V, Hodgkin JB, Bitzer M, et al. Defining cell-type specificity at the transcriptional level in human disease. *Genome Res*. 2013;23:1862–1873. doi: 10.1101/gr.155697.113.
- Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, Sokolov A, et al. A large-scale evaluation of computational protein function prediction. *Nat Methods*. 2013;10:221–227. doi: 10.1038/nmeth.2340.
- Guan Y, Ackert-Bicknell CL, Kell B, Troyanskaya OG, Hibbs MA. Functional genomics complements quantitative genetics in identifying disease-gene associations. *PLoS Comput Biol*. 2010;6:e1000991. doi: 10.1371/journal.pcbi.1000991.
- Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 1999;27:29–34.
- The Gene Ontology Consortium. The gene ontology: enhancements for 2011. *Nucleic Acids Res*. 2012;40:D559–D564.
- Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, et al. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res*. 2011;39(database issue):D685–D690. doi: 10.1093/nar/gkq1039.
- Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes.

- Nucleic Acids Res.* 2011;39(database issue):D691–D697. doi: 10.1093/nar/gkq1018.
31. Holden M, Deng S, Wojnowski L, Kulle B. GSEA-SNP: applying gene set enrichment analysis to SNP data from genome-wide association studies. *Bioinformatics.* 2008;24:2784–2785. doi: 10.1093/bioinformatics/btn516.
  32. O'Dushlaine C, Kenny E, Heron EA, Segurado R, Gill M, Morris DW, et al. The SNP ratio test: pathway analysis of genome-wide association datasets. *Bioinformatics.* 2009;25:2762–2763. doi: 10.1093/bioinformatics/btp448.
  33. Zhang K, Cui S, Chang S, Zhang L, Wang J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res.* 2010;38(web server issue):W90–W95. doi: 10.1093/nar/gkq324.
  34. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet.* 2010;11:843–854. doi: 10.1038/nrg2884.
  35. Mooney MA, Nigg JT, McWeeney SK, Wilmot B. Functional and genomic context in pathway analysis of GWAS data. *Trends Genet.* 2014;30:390–400. doi: 10.1016/j.tig.2014.07.004.
  36. Kar SP, Seldin MF, Chen W, Lu E, Hirschfield GM, Invernizzi P, et al; Italian PBC Genetics Study Group. Pathway-based analysis of primary biliary cirrhosis genome-wide association studies. *Genes Immun.* 2013;14:179–186. doi: 10.1038/gene.2013.1.
  37. Papassotiropoulos A, Gerhards C, Heck A, Ackermann S, Aerni A, Schick-tanz N, et al. Human genome-guided identification of memory-modulating drugs. *Proc Natl Acad Sci U S A.* 2013;110:E4369–E4374. doi: 10.1073/pnas.1314478110.
  38. Cordell HJ, Han Y, Mells GF, Li Y, Hirschfield GM, Greene CS, et al; Canadian-US PBC Consortium; Italian PBC Genetics Study Group; UK-PBC Consortium. International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat Commun.* 2015;6:8019. doi: 10.1038/ncomms9019.
  39. Greene CS, Troyanskaya OG. Accurate evaluation and analysis of functional genomics data and methods. *Ann NY Acad Sci.* 2012;1260:95–100. doi: 10.1111/j.1749-6632.2011.06383.x.
  40. Donato M, Xu Z, Tomoiaga A, Granneman JG, Mackenzie RG, Bao R, et al. Analysis and correction of crosstalk effects in pathway analysis. *Genome Res.* 2013;23:1885–1893. doi: 10.1101/gr.153551.112.
  41. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, et al. Disease networks. Uncovering disease–disease relationships through the incomplete interactome. *Science.* 2015;347:1257601. doi: 10.1126/science.1257601.
  42. Leiserson MD, Eldridge JV, Ramachandran S, Raphael BJ. Network analysis of GWAS data. *Curr Opin Genet Dev.* 2013;23:602–610. doi: 10.1016/j.gde.2013.09.003.
  43. Jia P, Zhao Z. Network-assisted analysis to prioritize GWAS results: principles, methods and perspectives. *Hum Genet.* 2014;133:125–138. doi: 10.1007/s00439-013-1377-1.
  44. Akula N, Baranova A, Seto D, Solka J, Nalls MA, Singleton A, et al; Bipolar Disorder Genome Study (BiGS) Consortium; Wellcome Trust Case-Control Consortium. A network-based approach to prioritize results from genome-wide association studies. *PLoS One.* 2011;6:e24220. doi: 10.1371/journal.pone.0024220.
  45. Chatr-Aryamontri A, Breitkreutz B-J, Heinicke S, Boucher L, Winter A, Stark C, et al. The BioGRID interaction database: 2013 update. *Nucleic Acids Res.* 2013;41(database issue):D816–D823. doi: 10.1093/nar/gks1158.
  46. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, et al. Human protein reference database–2009 update. *Nucleic Acids Res.* 2009;37(Database issue):D767–D772. doi: 10.1093/nar/gkn892.
  47. Cheng C, Yan KK, Hwang W, Qian J, Bhardwaj N, Rozowsky J, et al. Construction and analysis of an integrated regulatory network derived from high-throughput sequencing data. *PLoS Comput Biol.* 2011;7:e1002190. doi: 10.1371/journal.pcbi.1002190.
  48. Gennarino VA, D'Angelo G, Dharmalingam G, Fernandez S, Russolillo G, Sanges R, et al. Identification of microRNA-regulated gene networks by expression analysis of target genes. *Genome Res.* 2012;22:1163–1172. doi: 10.1101/gr.130435.111.
  49. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, et al. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet.* 2003;34:166–176. doi: 10.1038/ng1165.
  50. Fuxman Bass JI, Sahni N, Shrestha S, Garcia-Gonzalez A, Mori A, Bhat N, et al. Human gene-centered transcription factor networks for enhancers and disease variants. *Cell.* 2015;161:661–673. doi: 10.1016/j.cell.2015.03.003.
  51. Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res.* 2011;21:1109–1121. doi: 10.1101/gr.118992.110.
  52. Wong AK, Park CY, Greene CS, Bongo LA, Guan Y, Troyanskaya OG. IMP: a multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks. *Nucleic Acids Res.* 2012;40(web server issue):W484–W490. doi: 10.1093/nar/gks458.
  53. Himmelstein DS, Baranzini SE. Heterogeneous network edge prediction: a data integration approach to prioritize disease-associated genes. *PLoS Comput Biol.* 2015;11:e1004259. doi: 10.1371/journal.pcbi.1004259.
  54. Köhler S, Bauer S, Horn D, Robinson PN. Walking the interactome for prioritization of candidate disease genes. *Am J Hum Genet.* 2008;82:949–958. doi: 10.1016/j.ajhg.2008.02.013.
  55. Vanunu O, Magger O, Ruppin E, Shlomi T, Sharan R. Associating genes and protein complexes with disease via network propagation. *PLoS Comput Biol.* 2010;6:e1000641. doi: 10.1371/journal.pcbi.1000641.
  56. Magger O, Waldman YY, Ruppin E, Sharan R. Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks. *PLoS Comput Biol.* 2012;8:e1002690. doi: 10.1371/journal.pcbi.1002690.
  57. Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y, et al; International Inflammatory Bowel Disease Genetics Consortium. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.* 2011;7:e1001273. doi: 10.1371/journal.pgen.1001273.
  58. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 2014;42(database issue):D1001–D1006. doi: 10.1093/nar/gkt1229.
  59. Baranzini SE, Galwey NW, Wang J, Khankhanian P, Lindberg R, Pelletier D, et al; GeneMSA Consortium. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum Mol Genet.* 2009;18:2078–2090. doi: 10.1093/hmg/ddp120.
  60. Baranzini SE, Khankhanian P, Patsopoulos NA, Li M, Stankovich J, Cot-sapas C, et al. Network-based multiple sclerosis pathway analysis with GWAS data from 15,000 cases and 30,000 controls. *Am J Hum Genet.* 2013;92:854–865.
  61. Wang L, Mousavi P, Baranzini SE. iPINBPA: an integrative network-based functional module discovery tool for genome-wide association studies. *Pac Symp Biocomput.* 2015:255–266.
  62. Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, et al. Understanding multicellular function and disease with human tissue-specific networks. *Nat Genet.* 2015;47:569–576. doi: 10.1038/ng.3259.
  63. Mirina A, Atzmon G, Ye K, Bergman A. Gene size matters. *PLoS One.* 2012;7:e49093. doi: 10.1371/journal.pone.0049093.
  64. Himmelstein DS, Jensen LJ. One network to rule them all. *Thinklab.* 2015. doi: 10.15363/thinklab.d102.
  65. Himmelstein D, Greene C. Highly connected genes contain more SNPs. *Zenodo.* 2015. doi: 10.5281/zenodo.30105.
  66. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. *Arthritis Res Ther.* 2011;13:101. doi: 10.1186/ar3204.
  67. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* 2012;8:e1002793. doi: 10.1371/journal.pgen.1002793.
  68. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics.* 2010;26:1205–1210. doi: 10.1093/bioinformatics/btq126.

KEY WORDS: association studies ■ bioinformatics ■ genetics ■ genomics ■ network biology

## Genetic Association–Guided Analysis of Gene Networks for the Study of Complex Traits

Casey S. Greene and Daniel S. Himmelstein

*Circ Cardiovasc Genet.* 2016;9:179-184  
doi: 10.1161/CIRCGENETICS.115.001181

*Circulation: Cardiovascular Genetics* is published by the American Heart Association, 7272 Greenville Avenue,  
Dallas, TX 75231

Copyright © 2016 American Heart Association, Inc. All rights reserved.  
Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the  
World Wide Web at:

<http://circgenetics.ahajournals.org/content/9/2/179>

**Permissions:** Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Circulation: Cardiovascular Genetics* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the [Permissions and Rights Question and Answer](#) document.

**Reprints:** Information about reprints can be found online at:  
<http://www.lww.com/reprints>

**Subscriptions:** Information about subscribing to *Circulation: Cardiovascular Genetics* is online at:  
<http://circgenetics.ahajournals.org/subscriptions/>