

INVITED REVIEW

Pathway and network-based strategies to translate genetic discoveries into effective therapies

Casey S. Greene^{1,2,*} and Benjamin F. Voight^{1,2,3}¹Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine,²Institute for Translational Medicine and Therapeutics, Perelman School of Medicine and ³Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19103 USA

*To whom correspondence should be addressed at: Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, Philadelphia PA 19103 USA, Tel: +215-573-2991; Email: csgreene@upenn.edu

Abstract

One way to design a drug is to attempt to phenocopy a genetic variant that is known to have the desired effect. In general, drugs that are supported by genetic associations progress further in the development pipeline. However, the number of associations that are candidates for development into drugs is limited because many associations are in non-coding regions or difficult to target genes. Approaches that overlay information from pathway databases or biological networks can expand the potential target list. In cases where the initial variant is not targetable or there is no variant with the desired effect, this may reveal new means to target a disease. In this review, we discuss recent examples in the domain of pathway and network-based drug repositioning from genetic associations. We highlight important caveats and challenges for the field, and we discuss opportunities for further development.

Introduction

Human genetic association data has revealed how genes influence the susceptibility to rare and complex diseases. A central objective of drug development is to modify disease risk in predictable and beneficial ways, e.g. therapeutic lowering of blood pressure or cholesterol to prevent myocardial infarction. If the direction of perturbation can be inferred from human variation data—e.g. loss-of function etc.—these ‘natural’ perturbations could be harnessed to indicate a therapeutic hypothesis for rational drug design (1).

In addition, human genetic perturbations may also suggest targets that are the most likely to succeed through drug development. Nelson *et al.* (2) analyzed the evidence behind drug candidates that entered clinical trials. They found that the fraction of compounds with supporting genetic associations increases throughout the drug development pipeline, suggesting that drugs supported by associations are more likely to successfully

advance through the stages of clinical trials where drugs must show efficacy. This highlights the promise of genetic associations that can be mimicked by drugs. In the idealized case, they can take us from a new mechanism to an effective therapeutic.

The PCSK9 success story

An illustrative example for this human-genetics informed approach to drug design involves characterization of the Proprotein convertase subtilisin/kexin type 9 (PCSK9) connection to heart disease. In 2003, genetic variation at PCSK9 was determined to be a factor in autosomal dominant hypercholesterolemia (3). Specific variants within the gene—all missense mutations—segregated with the phenotype. However, no non-sense or frameshift mutations were identified, leading to the inference that these variants may be gain of function and that this gain results in higher levels of low-density lipoprotein cholesterol (LDL-C). Subsequently, PCSK9 was re-sequenced in the

Received: April 27, 2016. Revised: April 27, 2016. Accepted: May 19, 2016

© The Author (2016). Published by Oxford University Press.

All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

extremes of LDL-C, which uncovered nonsense mutations in the gene (4). The authors then designed an assay for two variants and observed that individuals with these alleles exhibited substantially lower LDL levels. In a follow-up study, nonsense mutations in the gene were revealed to be associated with both lower LDL levels as well as a substantial reduction in the lifetime risk of coronary heart disease (5).

Taken together, this genetic data strongly suggested that a therapeutic that mimics the effect of PCSK9 loss of function should not only lower LDL-C levels but also reduce risk of heart disease. Monoclonal antibodies against PCSK9 indeed demonstrate the LDL lowering effect in clinical trials (6), and biologics for PCSK9 are now available and are being tested in phase III clinical trials. Importantly, clinical trials for these agents have demonstrated their ability to lower LDL-C levels even without lifelong exposure to reduced PCSK9 activity. A recent meta-analysis of existing trials data suggests such treatments will be beneficial (7), and ongoing trials will assess whether or not this translates into a reduction in cardiovascular events after a heart attack (8).

PCSK9 is not the only success

The translation of the PCSK9 genetic discovery to therapeutic is one example among an increasing number of such reports. Missense mutations in the islet-specific zinc transport (SLC30A8) are associated with increased risk to type 2 diabetes (9) and loss-of-function mutations are associated with 2-fold protection against diabetes (10). Loss-of-function mutations in the apolipoprotein C3 component (APOC3) have been shown to dramatically lower triglycerides, cholesterol levels, and risk of coronary heart disease (11). Loss of function mutations in Niemann-Pick C1-like 1 (NPC1L1) are associated with decreased LDL cholesterol and reduced risk of coronary heart disease (12). Although mutations in the amyloid precursor protein are known to be associated with early-onset Alzheimer's disease, a recent report identified key mutations which protect against AD and age-related cognitive decline (13). These recently identified examples, analogous to PCSK9, offer promising therapeutic hypotheses worthy of further examination.

From genes to pathways and networks

The PCSK9 and above examples represent best-case scenarios: a loss of function allele induced a desirable phenotype that could be mimicked by a treatment. This demonstrates the promise of association-based drug development; however, the number of such successes has been limited. The NHGRI GWAS catalog (14) was developed to curate genetic associations and now holds more than 15 000 SNP-trait associations, but we have not yet developed effective drugs based on all of these associations. There have been a limited number of cases where each required element falls cleanly into place. The success stories have occurred for variants in coding regions, but most associated variants in the NHGRI catalog map to the non-coding genome. For these associations identifying the causal variant and connecting it to the causal gene remains challenging.

Importantly, each variant-phenotype association may not directly suggest a druggable target. Pathway and network-based approaches aim to expand the number of associations that lead to effective treatments by expanding the set of potential targets. For example, a genetic association may suggest that altering the effect of a gene would be beneficial, but perhaps that gene's product can't be perturbed using biologics or small molecules. A

pathway-based strategy could seek to push towards the desired state by targeting genes upstream or downstream in a signaling pathway, while a network-based strategy for example Guney et al. (15) may seek to target interacting partners to achieve the desired outcome (Fig. 1).

Pathway and network-based drug identification from genetic associations

Researchers are now developing approaches that analyze genetic associations in the context of networks and pathways to prioritize potential drugs (16–21). There is not yet a standard approach or set of techniques that are widely applied, so we highlight examples of the types of techniques being employed.

Cordell et al. (18) performed a meta-analysis of genetic association data for primary biliary cirrhosis and identified both known and new variants associated with the disease. Pathway-based analysis revealed several associated with the disease, and intersecting these pathways with annotations from the Drug Gene Interaction database (22) highlighted candidates that could modulate pathways associated with the disease. Such analyses lay the groundwork for studies that aim to evaluate the repurposed treatments in models of the disease.

Okada et al. (19) integrated many types of data, in addition to pathway annotations, to identify potential treatments for rheumatoid arthritis (RA). The authors started with a large meta-analysis of genome-wide association data, which provided a set of associated variants. The authors then identified genes that were in associated pathways, showed relevant mouse phenotypes, had protein-protein interaction network support, and many other factors of potential relevance. This allowed the authors to define a score for each gene based on the number of resources supporting its association. They then identified drugs that targeted the top genes. Overlaying drug-target information from DrugBank (23) and the Therapeutic Targets Database (23) revealed that a number of top candidates from the analysis were already targets of existing drugs for RA, suggesting that drugs targeting other top candidates might also be repurposed to treat RA. A similar approach has also been used to identify anti-cancer drugs (21).

Both of the above techniques use curated knowledge to link associations to potential drugs. In contrast with approaches that overlay curated knowledge to calculate a score based on summed sources of evidence, Greene et al. (17) used machine learning methods to identify regions of networks that were informative of genetic associations through a technique termed NetWAS. The NetWAS method uses the results of a genome-wide association study to find these regions of the networks. The analysis described in the article used genes that had a nominally significant (uncorrected $P < 0.01$) association with hypertension endpoints. Genes identified by the network analysis were more likely to be associated with hypertension-related gene ontology processes (25) and more likely to have been annotated as hypertension genes in the OMIM database (26). These genes were also more likely to be the target of antihypertensive therapies in the DrugBank database (23), suggesting that compounds that target other highly-ranked genes would be candidate antihypertensive treatments.

The highlighted techniques focus on drug repurposing for a single disease starting from a single genome-wide study or meta-analysis. In contrast with these, project Repheto (20) is an effort to perform large-scale drug repurposing from a catalog of genetic associations. They use heterogeneous networks that connect genes, diseases, tissues, and other factors (27). They

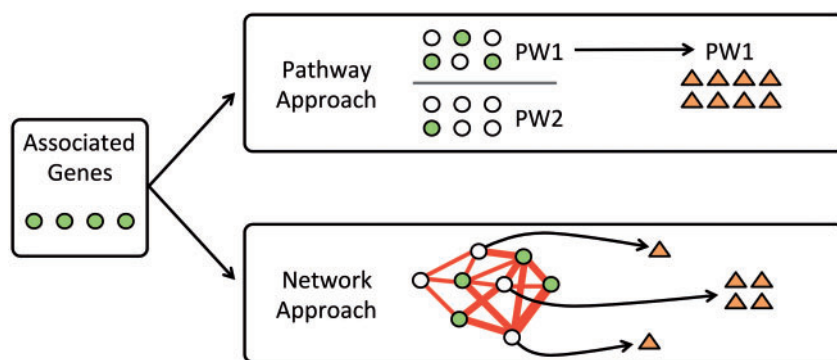


Figure 1. A genetic association study identifies genes associated with a phenotype. If those genes do not directly suggest effective strategies for therapeutics, researchers can use pathway or network-based approaches to expand the potential compound set. In the pathway-based approach, one or more associated pathways are identified and drugs that target those pathways are prioritized. In the network approach, genes are mapped onto an interaction network and genes that are highly connected to associated genes are identified. Drugs that target those genes represent candidates for further study.

then apply machine learning methods to predict effective drug-disease pairs. The entire process has been visible on an open science platform called *Thinklab*. The open platform reveals the extensive sets of analyses that lay the groundwork for machine learning applications. Though work is ongoing the initial results are highly promising.

Challenges for network and pathway-based approaches

Although promising results of network and pathway-based drug discovery approaches are emerging, methods must still overcome some challenges. A key step for all network and pathway-based approaches is to connect associated variants to the relevant gene or genes that are causal for the phenotype of interest. Methods developers need to take into account potential sources of bias in how the underlying SNPs are mapped to a candidate gene. For example, genes with more SNPs are often more central in many networks and more annotated in pathway resources (28,29). Network and pathway-based approaches need to use analytical strategies that map SNP-based P-values to gene-based P-values without suffering inflation from large genes (30,31). Approaches that fail to account for this structure may simply identify generic, as opposed to phenotype-specific, associations.

In addition to such biases, there are important biological challenges to mapping variants to genes. A variant may affect a distal, as opposed to the closest, gene. A recent example which highlights this issue is intronic variation near an obesity-associated locus (32–34). It is now clear that this is due, at least in part, to long range interactions between these variants and the genes *IRX3* and *IRX5* (35,36). The variant to gene mapping problem remains a challenging area of work. Techniques to predict enhancer-gene interactions (37) or those that use expression quantitative trait loci (38) may help to improve accurate variant to gene mappings. This is an active area of research, and we expect improvements in this area to immediately improve the sensitivity and specificity of pathway-based methods.

An often-overlooked aspect of studies that incorporate multiple data types and sources for drug repurposing is licensing and resource accessibility. The manner in which restrictive database licenses have hampered progress by the Himmelstein *et al.* team (20) has been instructive (39). For example, the authors have been unable to use the MSigDB database of chemical and genetic perturbations (40). Although many resources do

allow reuse by academic users, unclear and restrictive licenses have the potential to sap time from researchers engaged in network and pathway-based drug repurposing studies. This demonstrates that practitioners need to consider not only whether or not resources may provide useful data but also whether or not using such resources in a robust and reproducible way is compatible with their licenses.

Opportunities on the horizon

Network and pathway approaches benefit from the ability to use multiple distinct associations to identify a relevant pathway or region of a biological network. Although individual phenotypes have provided a fruitful starting point, collections of phenotypes may be an even more promising avenue for such approaches. We might expect that related phenotypes may not be influenced by the same variant and genetic mechanism, but the same pathway could drive them. An early report suggests that the use of such multi-phenotype or phenome-wide data for drug repurposing may be fruitful (41).

The Alzheimer's Disease Neuroimaging Initiative has collected and shared genotype, fluid biomarker, and imaging data on a set of individuals (42). These data can be used for network and pathway-based association analysis of individual phenotypes (43). Multi-phenotype analysis may identify consistent pathways or network regions associated with certain imaging phenotypes, which could suggest potential interventions.

Another major source of phenotyping data is the electronic health record (44,45). The breadth of these data presents both challenges and opportunities (46). Because these records contain numerous phenotypes with treatments available, existing drug-disease pairs can be used to evaluate methods.

In a few cases, a genetic association may directly suggest a drug-development strategy. For most cases, translating genetic associations into treatments is likely to be complex. In these cases, network and pathway-based approaches can integrate information from multiple types of variants with distinct effects to reveal pathways that can be modulated to affect a phenotype. The expanding availability of sequence measurements with extensive phenotyping makes this a promising area for the development of new methods. Systems capable of converting such data into drug candidates may help researchers repurpose existing therapies for other uses and target drugs for development more efficiently.

Conflict of Interest statement. None declared.

Funding

Support to CSG was provided via the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative via GBMF 4552. Support to BFV is provided by the American Heart Association (13SDG14330006) and the US National Institutes of Health/ National Institute of Diabetes and Digestive and Kidney Disorders (R01DK101478).

References

- Sanseau, P., Agarwal, P., Barnes, M.R., Pastinen, T., Richards, J.B., Cardon, L.R. and Mooser, V. (2012) Use of genome-wide association studies for drug repositioning. *Nat. Biotechnol.*, **30**, 317–320.
- Nelson, M.R., Tipney, H., Painter, J.L., Shen, J., Nicoletti, P., Shen, Y., Floratos, A., Sham, P.C., Li, M.J., Wang, J. et al. (2015) The support of human genetic evidence for approved drug indications. *Nat. Genet.*, **47**, 856–860.
- Abifadel, M., Varret, M., Rabès, J.P., Allard, D., Ouguerram, K., Devillers, M., Cruaud, C., Benjannet, S., Wickham, L., Erlich, D. et al. (2003) Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat. Genet.*, **34**, 154–156.
- Cohen, J., Pertsemlidis, A., Kotowski, I.K., Graham, R., Garcia, C.K. and Hobbs, H.H. (2005) Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat. Genet.*, **37**, 161–165.
- Cohen, J.C., Boerwinkle, E., Mosley, T.H. and Hobbs, H.H. (2006) Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.*, **354**, 1264–1272.
- Stein, E.A., Gipe, D., Bergeron, J., Gaudet, D., Weiss, R., Dufour, R., Wu, R. and Porfy, R. (2012) Effect of a monoclonal antibody to PCSK9, REGN727/SAR236553, to reduce low-density lipoprotein cholesterol in patients with heterozygous familial hypercholesterolaemia on stable statin dose with or without ezetimibe therapy: a phase 2 randomised controlled trial. *Lancet (London, England)*, **380**, 29–36.
- Navarese, E.P., Kolodziejczak, M., Schulze, V., Gurbel, P.A., Tantry, U., Lin, Y., Brockmeyer, M., Kandzari, D.E., Kubica, J.M., D'Agostino, R.B. et al. (2015) Effects of proprotein convertase subtilisin/kexin type 9 antibodies in adults with hypercholesterolemia: a systematic review and meta-analysis. *Ann. Intern. Med.*, **163**, 40–51.
- Schwartz, G.G., Bessac, L., Berdan, L.G., Bhatt, D.L., Bittner, V., Diaz, R., Goodman, S.G., Hanotin, C., Harrington, R.A., Jukema, J.W. et al. (2014) Effect of alirocumab, a monoclonal antibody to PCSK9, on long-term cardiovascular outcomes following acute coronary syndromes: rationale and design of the ODYSSEY outcomes trial. *Am. Heart J.*, **168**, 682–689.
- Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S. et al. (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*, **445**, 881–885.
- Flannick, J., Thorleifsson, G., Beer, N.L., Jacobs, S.B.R., Grarup, N., Burt, N.P., Mahajan, A., Fuchsberger, C., Atzmon, G., Benediktsson, R. et al. (2014) Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat. Genet.*, **46**, 357–363.
- Crosby, J., Peloso, G.M., Auer, P.L., Crosslin, D.R., Stitzel, N.O., Lange, L.A., Lu, Y., Tang, Z., Zhang, H., Hindy, G. et al. (2014) Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N. Engl. J. Med.*, **371**, 22–31.
- The Myocardial Infarction Genetics Consortium Investigators (2014) Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N. Engl. J. Med.*, **371**, 2072–2082.
- Jonsson, T., Atwal, J.K., Steinberg, S., Snaedal, J., Jonsson, P.V., Bjornsson, S., Stefansson, H., Sulem, P., Gudbjartsson, D., Maloney, J. et al. (2012) A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature*, **488**, 96–99.
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorf, L. et al. (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001–6.
- Guney, E., Menche, J., Vidal, M., Barabasi, A. (2016) Network-based in silico drug efficacy screening. *Nature Communications*, **7**, 10331.
- Hu, Z., Chang, Y.C., Wang, Y., Huang, C.L., Liu, Y., Tian, F., Granger, B. and Delisi, C. (2013) VisANT 4.0: Integrative network platform to connect genes, drugs, diseases and therapies. *Nucleic Acids Res.*, **41**, W225–W231.
- Greene, C.S., Krishnan, A., Wong, A.K., Ricciotti, E., Zelaya, R.A., Himmelstein, D.S., Zhang, R., Hartmann, B.M., Zaslavsky, E., Sealfon, S.C. et al. (2015) Understanding multi-cellular function and disease with human tissue-specific networks. *Nat. Genet.*, **47**, 569–576.
- Cordell, H.J., Han, Y., Mells, G.F., Li, Y., Hirschfeld, G.M., Greene, C.S., Xie, G., Juran, B.D., Zhu, D., Qian, D.C. et al. (2015) International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat. Commun.*, **6**, 8019.
- Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S. et al. (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, **506**, 376–381.
- Himmelstein, D., Lizee, A., Khankhanian, P., Brueggeman, L., Chen, S., Hessler, C., Hadley, D. and Baranzini, S. (2015) Rephetio: repurposing drugs on a hetnet. *Thinklab*, 10.15363/thinklab.4.
- Zhang, J., Jiang, K., Lv, L., Wang, H., Shen, Z., Gao, Z., Wang, B., Yang, Y., Ye, Y., Wang, S. et al. (2015) Use of Genome-Wide Association Studies for Cancer Research and Drug Repositioning. *PLoS One*, **10**, e0116477.
- Griffith, M., Griffith, O.L., Coffman, A.C., Weible, J.V., McMichael, J.F., Spies, N.C., Koval, J., Das, I., Callaway, M.B., Eldred, J.M. et al. (2013) DGIdb: mining the druggable genome. *Nat. Methods*, **10**, 1209–1210.
- Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z. and Woolsey, J. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, **34**, D668–D672.
- Zhu, F., Shi, Z., Qin, C., Tao, L., Liu, X., Xu, F., Zhang, L., Song, Y., Liu, X., Zhang, J. et al. (2012) Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Res.*, **40**, D1128–D1136.
- The Gene Ontology Consortium. (2012) The Gene Ontology: enhancements for 2011. *Nucleic Acids Res.*, **40**, D559–D564.
- Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A. and McKusick, V.A. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–7.

27. Himmelstein, D.S. and Baranzini, S.E. (2015) Heterogeneous Network Edge Prediction: A Data Integration Approach to Prioritize Disease-Associated Genes. *PLOS Comput. Biol.*, **11**, e1004259.
28. Himmelstein, D.S. and Greene, C.S. (2015) Highly connected genes contain more SNPs. [10.5281/zenodo.30105](https://doi.org/10.5281/zenodo.30105).
29. Greene, C.S. and Himmelstein, D.S. (2016) Genetic association-guided analysis of gene networks for the study of complex traits. *Circ. Cardiovasc. Genet.*, **9**, 179–184.
30. Mishra, A. and Macgregor, S. (2015) VEGAS2: Software for more flexible gene-based testing. *Twin Res. Hum. Genet.*, **18**, 86–91.
31. Pedroso, I.I., Barnes, M.R., Lourdosamy, A., Al-Chalabi, A. and Breen, G. (2015) FORGE: multivariate calculation of genome-wide p-values from Genome-Wide Association Studies Authors and Affiliations Cold Spring Harbor Labs Journals [bioRxiv](https://doi.org/10.1101/023648), doi: 10.1101/023648.
32. Dina, C., Meyre, D., Gallina, S., Durand, E., Körner, A., Jacobson, P., Carlsson, L.M., Kiess, W., Vatin, V., Lecoecur, C., et al. (2007) Variation in *FTO* contributes to childhood obesity and severe adult obesity. *Nat. Genet.*, **39**, 724–6.
33. Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R.B., Elliott, K.S., Lango, H., Rayner, N.W. et al. (2007) A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, **316**, 889–894.
34. Scuteri, A., Sanna, S., Chen, W.M., Uda, M., Albai, G., Strait, J., Najjar, S., Nagaraja, R., Orrù, M., Usala, G. et al. (2007) Genome-wide association scan shows genetic variants in the *FTO* gene are associated with obesity-related traits. *PLoS Genet.*, **3**, e115.
35. Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F. et al. (2014) Obesity-associated variants within *FTO* form long-range functional connections with *IRX3*. *Nature*, **507**, 371–375.
36. Claussnitzer, M., Dankel, S.N., Kim, K.H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I.S., Beaudry, J.L., Puviondran, V. et al. (2015) *FTO* Obesity Variant Circuitry and Adipocyte Browning in Humans. *N. Engl. J. Med.*, **373**, 895–907.
37. Whalen, S., Truty, R.M. and Pollard, K.S. (2016) Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat. Genet.*, **48**, 488–96.
38. Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Nicolae, D.L., Cox, N.J. et al. (2015) A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.*, **47**, 1091–1098.
39. Himmelstein, D., Jensen, L.J., Smith, M., Fortney, K. and Chung, C. (2015) Integrating resources with disparate licensing into an open network. [Thinklab](https://doi.org/10.15363/THINKLAB.D107), 10.15363/THINKLAB.D107.
40. Himmelstein, D. MSigDB licensing. [Thinklab](https://doi.org/10.15363/THINKLAB.D108), 10.15363/THINKLAB.D108.
41. Rastegar-Mojarad, M., Ye, Z., Kolesar, J.M., Hebring, S.J. and Lin, S.M. (2015) Opportunities for drug repositioning from phenome-wide association studies. *Nat. Biotechnol.*, **33**, 342–345.
42. Saykin, A.J., Shen, L., Foroud, T.M., Potkin, S.G., Swaminathan, S., Kim, S., Risacher, S.L., Nho, K., Huentelman, M.J., Craig, D.W. et al. (2010) Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. *Alzheimer's Dement.*, **6**, 265–273.
43. Song, A., Yan, J., Kim, S., Risacher, S.L., Wong, A.K., Saykin, A.J., Shen, L. and Greene, C.S. (2016) Network-based analysis of genetic variants associated with hippocampal volume in Alzheimer's disease: a study of ADNI cohorts. *BioData Min.*, **9**, 3.
44. Denny, J.C., Ritchie, M.D., Basford, M.A., Pulley, J.M., Bastarache, L., Brown-Gentry, K., Wang, D., Masys, D.R., Roden, D.M. and Crawford, D.C. (2010) PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*, **26**, 1205–1210.
45. Pendergrass, S.A., Brown-Gentry, K., Dudek, S., Frase, A., Torstenson, E.S., Goodloe, R., Ambite, J.L., Avery, C.L., Buyske, S., Bůžková, P. et al. (2013) Phenome-wide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. *PLoS Genet.*, **9**, e1003087.
46. Denny, J.C., Bastarache, L., Ritchie, M.D., Carroll, R.J., Zink, R., Mosley, J.D., Field, J.R., Pulley, J.M., Ramirez, A.H., Bowton, E. et al. (2013) Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.*, **31**, 1102–1110.