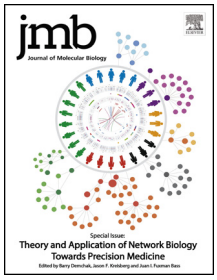
Review



Enabling Precision Medicine through Integrative Network Models

Victoria Yao[1,†], **Aaron K. Wong**[3,†] **and Olga G. Troyanskaya**[1,2,3]

*1 - **Department of Computer Science,** Princeton University, Princeton, NJ 08544, USA*
*2 - **Lewis-Sigler Institute for Integrative Genomics,** Princeton University, Princeton, NJ 08544, USA*
*3 - **Center for Computational Biology,** Flatiron Institute, Simons Foundation, New York, NY 10010, USA*

***Correspondence to Olga G. Troyanskaya:** Department of Computer Science, Princeton University, Princeton, NJ 08544, USA. ogt@cs.princeton.edu*
https://doi.org/10.1016/j.jmb.2018.07.004
Edited by Jason Francis Kreisberg

Abstract

A key challenge in precision medicine lies in understanding molecular-level underpinnings of complex human disease. Biological networks in multicellular organisms can generate hypotheses about disease genes, pathways, and their behavior in disease-related tissues. Diverse functional genomic data, including expression, protein–protein interaction, and relevant sequence and literature information, can be utilized to build integrative networks that provide both genome-wide coverage as well as contextual specificity and accuracy. By carefully extracting the relevant signal in thousands of heterogeneous functional genomics experiments through integrative analysis, these networks model how genes work together in specific contexts to carry out cellular processes, thereby contributing to a molecular-level understanding of complex human disease and paving the way toward better therapy and drug treatment. Here, we discuss current methods to build context-specific integrative networks, focusing on tissue-specific networks. We highlight applications of these networks in predicting tissue-specific molecular response, identifying candidate disease genes, and increasing power by amplifying the disease signal in quantitative genetics data. Altogether, these exciting developments enable biomedical scientists to characterize disease from pathophysiology to cellular system and, finally, to specific gene alterations—making significant strides toward the goal of precision medicine.

Networks as models of human biology

To realize the promise of precision medicine, we must elucidate the immense molecular complexity that forms the foundation of disease. Most human diseases are polygenic, perhaps even “omnigenic” [1]. While decades of targeted disease research and the rise of large-scale quantitative genetics studies such as genome-wide association studies (GWAS) have been valuable in identifying genes and genetic variants that may be linked to a wide range of diseases and phenotypes, it is increasingly clear that there is a “missing heritability” problem (i.e., even as the sample sizes in these studies continue to grow, only a small proportion of estimated heritability appears to be explained by the identified variants) [2, 3].

Understanding complex disease at the molecular level requires us to model specific molecular-level changes that lead to disease. These changes can happen through a variety of mechanisms, for example, regulatory abnormalities, modifications to protein interactions, or effects on signaling cascades. Modern genome-scale experimental techniques enable us to monitor and probe these molecular events by providing a wealth of data along multiple axes of cellular activity. Diverse high-throughput data vary in relevance depending on the biological process under study (e.g., a specific tissue, disease, or pathway), as both experimental technologies and perturbations capture different biological signal with varying degrees of success. Thus, integrative analysis of these data is paramount because (1) many diseases/tissues of interest are interrogated by multiple data sets; (2) each data set holds a complex mixture of signals relevant not only to the biological question or disease under study, but also to many other biological events (e.g., cancer data sets

have very strong immune signals, kidney disease data have strong inflammation signals); and (3) individual data sets are noisy, necessitating the identification of strong, recurring signals in relevant data sets.

A powerful set of approaches (Box 1) has emerged for integrating diverse data into functional maps of human cellular biology. These network approaches use a variety of machine learning and statistical algorithms to integrate very large collections of noisy and heterogeneous human "omics" data into functional maps, or networks [4–6]. Intuitively, an edge between two genes in these functional maps typically represents the probability that the genes are, directly or indirectly, participating in the same biological process or pathway (e.g., innate immune response, microtubule polymerization, axonogenesis). The genome-wide gene networks that result from integration of these data allow biologists to generate specific, experimentally testable hypotheses and provide a systems-level view of biological processes.

Biological network models have typically represented general views of organismal biology, not resolved to specific tissues or cell types. However, tissue and cellular context is critical for interpreting the behavior of genes and pathways, as gene function and interactions can vary greatly between tissues and cell types, and dysregulation of tissue- and cell-lineage-specific processes underlies many diseases. For example, selective neuronal vulnerability is a key characteristic of neurodegenerative diseases such as Parkinson's disease, and the neuronal subtypes as well as affected brain regions tend to be strong determinants of their corresponding clinical phenotypes [7]. In Parkinson's disease, the dopaminergic neurons in the substantia nigra pars compacta area of the brain are particularly susceptible to cell death, while highly similar dopaminergic neurons in the nearby ventral tegmental area are much less vulnerable. Thus, to fully capture the underlying biological processes relevant for a disease like Parkinson's disease, brain-region-specific networks are necessary.

Below, we discuss approaches to construct tissue- and cell-type-specific networks from integrations of large collections of public functional genomic data, and how such networks can be applied to study the molecular basis of human disease (Fig. 1). We begin with a discussion of methods that integrate heterogeneous data into networks and further technical innovations that effectively "summarize" these data into context-specific maps of the biological landscape of specific tissues and cell types. We then examine applications of these networks to the study of disease. Finally, we argue for the importance of making these networks and accompanying methods accessible to the wider community through user-friendly interactive public systems that are maintained over time.

Building tissue- and cell-type-specific functional networks

Methods that construct tissue-specific networks have historically been limited by the availability of experimental data for specific tissues and cell types, especially in humans. These direct approaches typically assemble available tissue-specific expression data into gene correlation networks [8–11] or overlay those expression data on global (non-tissue-specific) protein–protein interaction networks [12–14]. More sophisticated methods to construct context-specific regulatory networks by integrating (as opposed to simple overlaying) context-specific (e.g., tissue- or cell-type specific) expression data with a non-context-specific network have also been recently developed [15]. These approaches, while valuable, are applicable only to tissues and cell types which can be readily assayed [16] and depend almost entirely on the quality of the available tissue- or cell-type-specific data. For example, in cancer, where The Cancer Genome Atlas and other initiatives have amassed large, high-quality collections of diverse, genome-scale data to characterize specific cancer types, there has been significant progress in the development of network models, and they have yielded invaluable insights into the cancer landscape [17–21]. However, for the vast majority of normal human tissues, direct experimental assay remains infeasible (especially of living, and not postmortem, tissue), requiring computational methods that can infer tissue-specific interactions from large heterogeneous data compendia.

To address this challenge, recent work by Greene *et al.* introduced a method that can simultaneously

Fig. 1. Tissue-specific functional interaction networks. Tissue-specific networks are constructed by integrating (a) hierarchy-aware tissue-specific knowledge and a large human data compendium using a tissue-specific regularized Bayesian integration method that (b) identifies and weights data sets based on their tissue-relevant signal. These integrative networks are used for downstream (c) tissue-specific disease analyses. More specifically, to construct the tissue-specific functional interaction standard (a), gene pairs are considered positive examples when they both participate in the same process (i.e., process co-membership) and are expressed in the tissue of interest. Negative examples include gene pairs that either do not participate in the same process or are expressed in other tissues (see Methods in Ref. [4] for more details). The tissue-specific regularized Bayesian integration method can then use this gold standard to mine the signal from a large data compendium to construct tissue-specific networks (b). As effective summaries of tissue-specific biology, the network can then be used as a representation of tissue-specific biology to help generate hypotheses relevant for human disease. For example, the network itself can be used as input to downstream machine learning methods to predict disease genes or reprioritize quantitative genetics data (see Fig. 2). The network itself can also provide functional interpretations for any gene sets of interest (see Fig. 3).

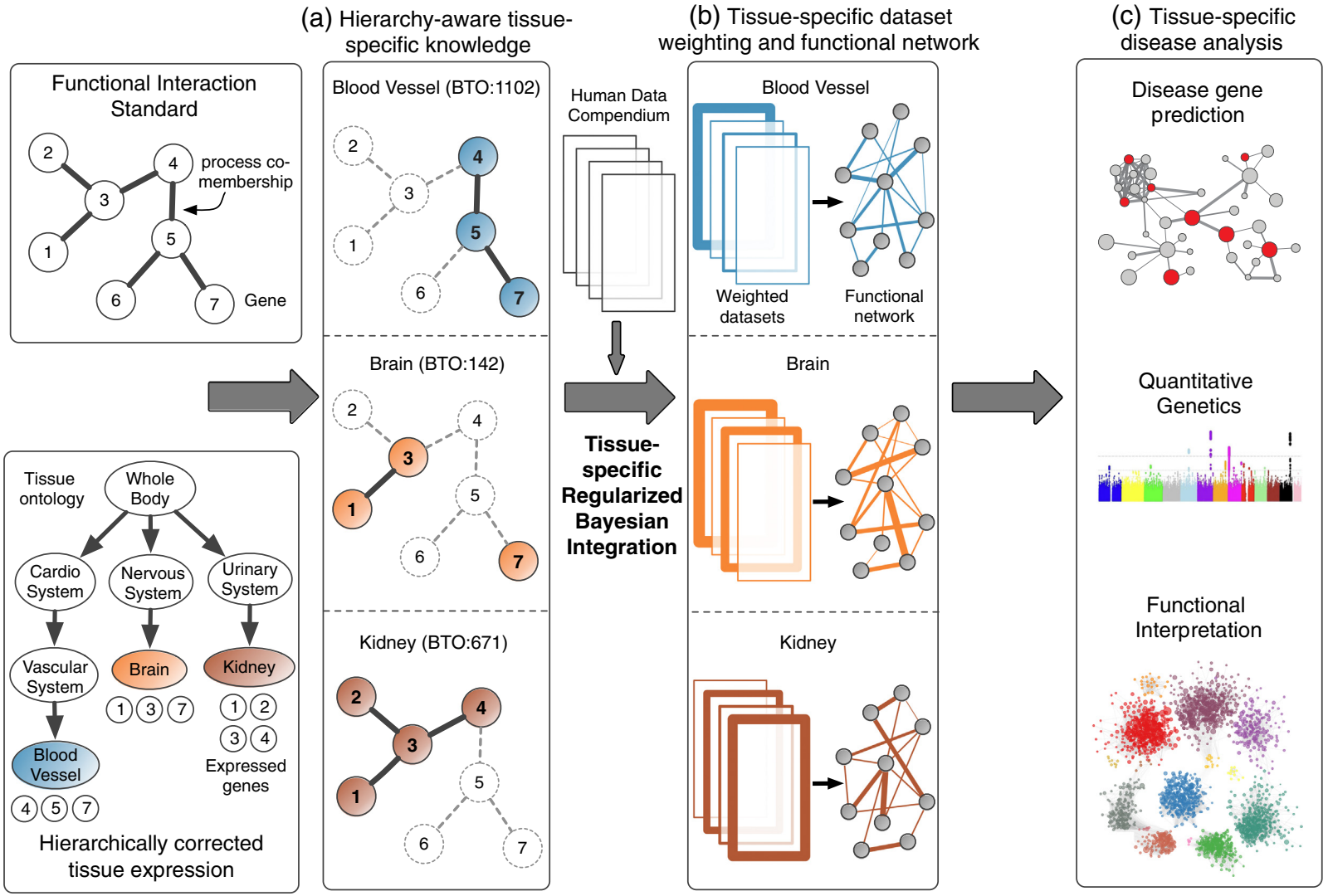


Fig. 1 (*legend on previous page*)

Box 1
Integrative network concepts

Term	Definition
Machine learning	The application of methods that learn patterns from data, usually with the goal of modeling the underlying structure and/or to make coherent predictions on new data
Supervised machine learning	Machine learning tasks that are given a "gold standard," which include examples of desired outputs given input data, and the objective of the method is to model this "training" data and make predictions for any new input data
Bayesian integration	Machine learning approaches that combine multiple sources of evidence (e.g., data sets) and make probabilistic predictions, for example, the likelihood that two genes are participating in the same pathway
Integrative network	A network that is constructed through the synthesis of heterogeneous data types
Functional network	An integrative network where the nodes are genes, and edge weights correspond to the posterior probability of a "functional relationship" (i.e., whether the genes perform similar functions, for example, participating in the same pathway or mediating the same interaction). These functional maps effectively summarize collections of genomic data in a biologically meaningful way.
Tissue-specific network	A functional network where edge weights represent the posterior probability of a tissue-specific functional relationship between a gene pair, thus representing how pathways and processes work in a particular tissue
Support vector machine (SVM)	A supervised machine learning method that projects input data into a high-dimensional space and identifies the hyperplane that best separates input data points based on the gold standard provided. Given a new data point, the model can then classify it accordingly.
Network-wide association study (NetWAS)	A supervised machine learning method that integrates the noisy disease signal in GWAS (as represented by genes with nominally significant *p*-values) together with the tissue biology signal in tissue-specific functional networks to reprioritize genes potentially associated with the disease/trait of interest.
Noise	Variation in the data that is spurious to the phenomenon actually being measured, in contrast to "signal" (i.e., the "true" underlying pattern in the data due to the phenomena being studied). For example, these variations could be a result of measurement errors introduced when capturing the data; they could also be due to general stochastic variation. One of the primary goals of machine learning is to separate signal from noise.

extract tissue or cell-type functional signals from large and diverse genomic data collections, including for tissues and cell types for which no tissue-specific experimental data exist. This approach generated genome-scale functional maps of 144 human tissues and cell-types by integrating a collection of data sets covering thousands of experiments from more than 14,000 distinct publications. Using a Bayesian integration technique, each data set was automatically assessed for its relevance to each of the tissue- and cell-lineage-specific functional contexts (Fig. 1a, b). With this approach, networks could be constructed for tissues with little or no directly assayed high-throughput data by automatically up-weighting data sets from related tissues and prioritizing these tissue-relevant signals over other data. For example, the method constructed a network for the dentate gyrus (a brain tissue with limited data) by extracting relevant signals from other (larger or related) tissues and cell types in the nervous system [4]. The resulting functional maps provided a detailed portrait of protein function and interactions in specific human tissues and cell lineages ranging in function (e.g., from B lymphocytes to the renal glomerulus) and in scale (e.g., from the substantia nigra to the whole brain).

These tissue-specific networks can be used to generate specific, testable hypotheses about gene functions, interactions, and disease association (Fig. 1c), which can then be experimentally validated. In contrast with general networks, which assume that the function of genes remains constant across tissues, these tissue-specific maps answer targeted, tissue-specific biological questions. For example, Greene *et al*. hypothesized that the cell-lineage-specific interactions of the gene *IL1B* in the blood vessel network (the tissue where it has a key role in inflammation) could accurately predict perturbation responses to IL-1β stimulation. To test this hypothesis, they profiled the gene expression of aortic smooth muscle cells with IL-1β stimulation and examined the genes whose expression was upregulated. Specifically, the top genes connected to *IL1B* in the blood vessel network responded to IL-1β stimulation in blood vessel cells; importantly, the blood vessel network was far more predictive of this response than a general, non-tissue-specific network or than networks from unrelated tissues.

Predicting disease-associated genes with integrative networks

The multifactorial nature of most human diseases renders them difficult to study. Targeted studies have

provided a useful foundation—in many diseases, some of the key genes and disease mechanisms are well studied. Large-scale quantitative genetics studies have provided further clues as to which genetic variants may affect disease risk. Yet together, these studies are still only able to explain at best a modest fraction of predicted genetic variance [2].

Molecular interaction networks provide effective summaries of cellular processes and thus represent powerful tools for investigating disease genes in a way that is complementary to experimental and quantitative genetic studies. Intuitively, network-based approaches for disease-associated gene prediction analyze network patterns that are associated with genes known to be involved in a specific disease (e.g., autism [22–25], cancer [26–28]) and then identify additional candidate genes based on shared interaction patterns. These candidate genes serve as hypotheses of disease association which can then be experimentally tested. The power of such approaches is in the data-driven nature of these predictions, which minimizes biases toward well-studied genes and processes.

Most biological/biomedical research is concentrated on genes and processes that already have some evidence of association with high impact outcomes (e.g., disease). Thus, genes that are already well studied continue to be studied even more, creating strong biases toward existing knowledge (known as *literature bias*, see Ref. [29]). Network-based disease prediction can mitigate literature bias by enabling researchers to find candidate genes which are not well represented in the current literature but exhibit strong support for pathogenic involvement in genomic data. In addition, the genome-wide ranking enables additional downstream functional interpretations for key dysregulated processes that would otherwise be intractable given only a handful of previously studied disease-associated genes.

These approaches initially used protein-protein physical interaction networks (PPI) [30–32]. PPI networks have not only the advantage of representing direct physical interactions, but also the limitation of relatively sparse coverage, especially in the context of tissue-specific networks. This can be addressed with tissue-specific functional networks [4] as described above. Another challenge for these approaches is that for many complex human diseases, only a limited number of disease-associated genes are well studied. To address this challenge, the general idea of predicting candidate disease-associated genes based on network interaction patterns similar to those of known disease-related genes has recently been extended to use an evidence-weighted machine learning approach [4, 22]. Specifically, in addition to using only high-confidence (e.g., experimentally verified) disease genes as training examples/gold standard for the machine learning classifier, genes with weak association (e.g., based on text mining) are also considered. The various gold standard gene sets are weighted according to their respective levels of confidence (e.g., experimentally verified genes receive much more weight than text-mining based associations) which the classifier takes into account during training. This approach was successfully used to predict hundreds of likely autism-associated genes using a brain-specific functional network [22].

Re-prioritizing quantitative genetic studies results with network models

Researchers have long recognized the need and power of distilling disease complexity by examining known disease genes and results from quantitative genetics studies in the context of biological networks. Nevertheless, it is becoming increasingly clear that the importance of considering molecular networks in interpreting the genetics of human disease may still have been underestimated. A recent study underscored this point: Yang *et al*. [33] addressed the problem of "missing heritability" in quantitative genetics studies by considering all common variants together (including those with effect sizes far below significance) and demonstrated that this explains most of the estimated heritability. This study and others have given rise to theories such as the omnigenic hypothesis [1]: genes can affect each other through their tightly interconnected networks, and as such, genes that have little direct bearing on a particular disease may, in aggregate, affect core disease pathways and influence disease risk. Thus, while quantitative genetics studies such as GWAS are valuable in providing unbiased glimpses into the genetic basis of disease, to fully realize their promise and identify core disease-associated genes, it is crucial to develop methods that analyze the results of these studies in the context of relevant biological networks.

One such method is network-wide association study (NetWAS) [4], which integrates tissue-specific networks with the results from standard GWAS as a means to reprioritize every gene in the genome for potential disease association. The method is premised on the idea that the top GWAS associations are enriched for disease-relevant genes, even if they fall below statistical significance. NetWAS constructs a support vector machine (SVM) where the classifier features are edges from a tissue-specific network, and labeled examples are drawn from the GWAS result (Fig. 2). Briefly, nominally significant genes (e.g., p-value <0.01) from the GWAS are positive examples and random genes above the significance threshold negative examples in the SVM classification. The result is a genome-wide re-ranking of genes driven by their network similarity—in a relevant tissue—to the top GWAS-associated genes. The power of the NetWAS approach is derived from the fact that it is discovery-driven; only genes from the GWAS itself are used as training input, as

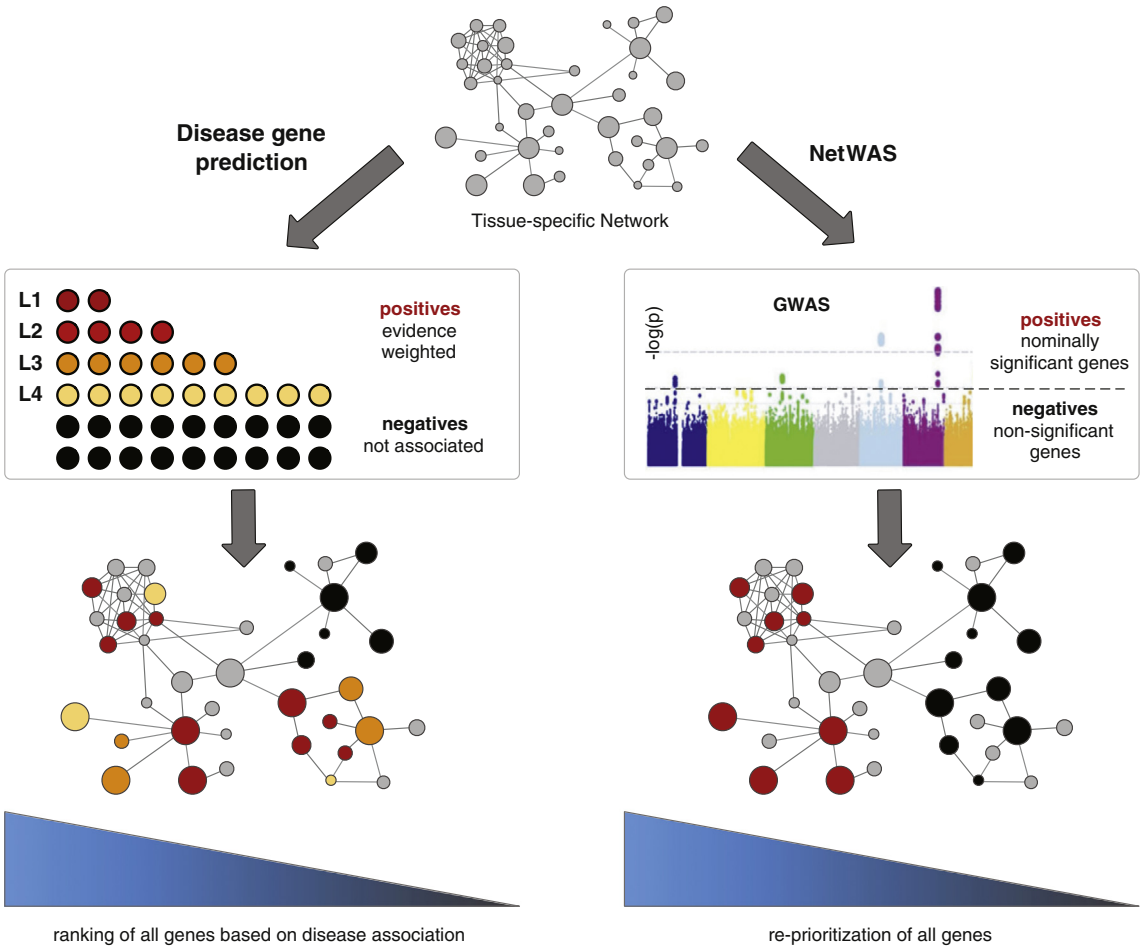


Fig. 2. Genome-wide candidate disease gene prediction using tissue-specific networks. Based on the shared network connectivity patterns of disease-associated genes, two approaches to disease gene prediction are as follows: (A) using an evidence-weighted machine learning approach that considers known disease genes with varying levels of confidence as positive examples and unrelated genes as negatives to rank all genes in the genome based on disease association, and (B) using nominally significant genes in a GWAS study as positive examples and non-significant genes as negatives to re-prioritize all genes in the genome. By using a tissue-specific network synthesized from a large collection of heterogeneous data to drive the prioritization of candidate disease genes, the problem of literature bias (where focus is concentrated on genes that have been previously characterized in a relevant disease context) is somewhat ameliorated.

opposed to genes from the potentially biased (and often sparse) prior disease knowledge. By fusing the biological models captured by the tissue-specific networks with the disease signal inherent in quantitative genetics studies, NetWAS has been successfully applied to generate hypotheses related to the molecular basis of a range of diseases, including obesity, type 2 diabetes, and systemic lupus erythematosus [4].

Other network approaches include methods based on PPI networks and prior disease knowledge [34–41]. These methods rank candidate genes from a genomic interval by their connectivity to known causal genes in protein–protein interaction networks. These approaches can be helpful in analyzing GWAS studies results, but their performance is highly dependent on the coverage and relevance of available protein–protein interactions and known disease gene associations. In general, when choosing an approach for generating candidate disease genes, scientists must balance the need for an accurate method (and well-conducted evaluations of predictions) with the challenge of avoiding serious biases toward well-studied disease genes and pathways.

Network-based functional interpretation of genes and gene sets

Networks are not only essential for the identification of candidate disease genes and pathways, but also for the functional interpretation of dysregulated processes

[28, 42]. By leveraging integrated networks that effectively summarize the functional landscape of a relevant context to study the corresponding disease, researchers can identify core disease pathways or processes on which a number of genetic variants converge. For example, in autism, likely disease-associated genes in some of the most common copy number variant regions appear connected to core autism genes by affecting cognition, brain development, and axonogenesis. Furthermore, a network-based view of disease-associated genes (both known and strongly predicted candidates) provides hypotheses of precisely how these genes are associated with the disease and assigns functional roles for previously uncharacterized candidate genes. For example, in the aforementioned autism study, by analyzing the functional connectivity of top autism-associated gene predictions in the brain-specific network, autism-associated brain-specific functional modules can be identified. One of the top candidates, *DIP2C*, was previously uncharacterized (Fig. 3) but functionally localized to the module representing Rho and insulin-like growth factor receptor pathways. Thus, network-based analyses can guide follow-up experiments on *DIP2C* to characterize its role in the cellular dysregulation in autism.

Integrated, tissue-specific networks can also be used to study the higher level topology of relationships between diseases, an important question both for differential diagnosis and treatment development. Greene *et al*. created data-driven maps quantifying the molecular relationships between diseases in the context of the tissue-specific functional networks. For example, using the substantia nigra network to chart out a disease map for Parkinson's disease highlighted both documented disease associations (e.g., basal ganglion disease, lysosomal storage disease) and more subtle connections (e.g., inherited metabolic disorder, thyroid cancer). This type of inter-disease association analysis provides an additional perspective on disease genetics and the crosstalk between their underlying pathways, which could ultimately lead to new insights into the molecular characterization of diseases.

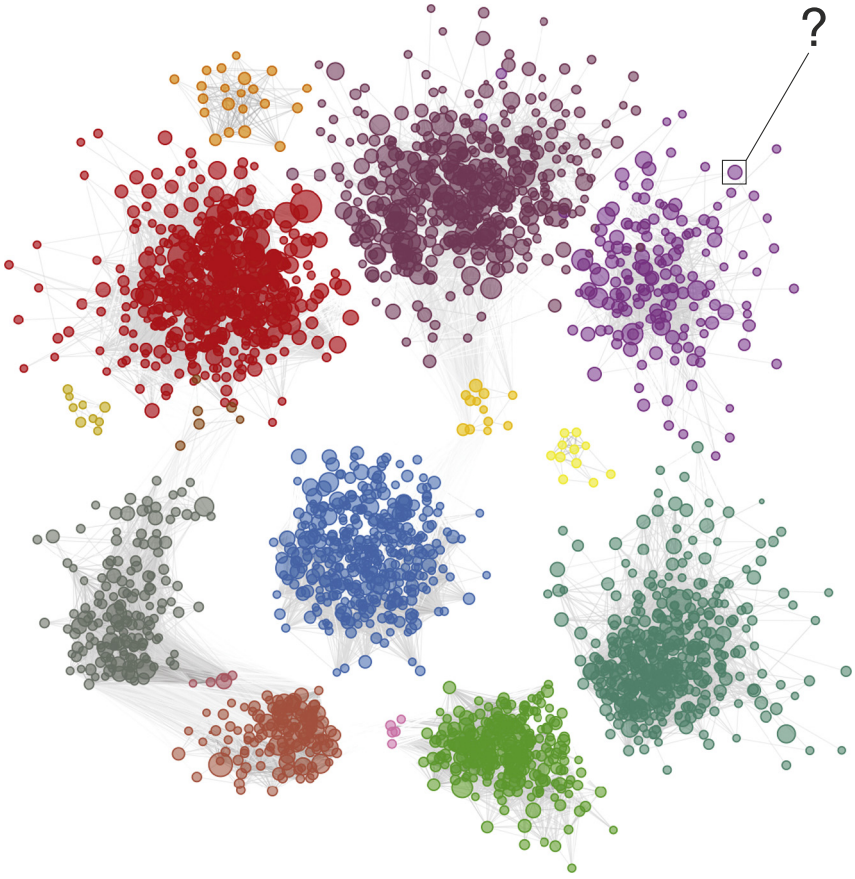


Fig. 3. Functional interpretations of disease genes using tissue-specific networks. Autism-associated functional modules identified by clustering of predicted autism-associated genes in a brain-specific functional network. The functional role of a previously uncharacterized gene can be hypothesized based on localization into the Rho and insulin-like growth factor receptor pathway module.

Making networks accessible to biomedical researchers through dynamic, interactive interfaces

Many of the integrated networks and associated analyses are available publicly through high-quality web interfaces, enabling broad access by biomedical researchers (Table 1). These resources are critical in making computational methods accessible to bench biologists, empowering them to generate testable hypotheses from data-driven, integrative analyses. Many feature dynamic, interactive visualizations, encouraging data exploration even for users without specific computational expertise. In addition, many resources enable researchers to apply integrated network analyses (e.g., GWAS reprioritization) to their own data [4]. These applications typically perform the analyses using server-side resources and, importantly, do not require users to install specialized software. By showcasing user-friendly interfaces and enabling access to sophisticated analyses, these integrated network resources can complement the tools of modern biologists to interpret and guide experiments.

Future directions

Here, we have primarily discussed approaches to leverage integrated networks for studying the genetic basis of disease. However, lifestyle and environmental factors (e.g., diet, exercise, sleep deprivation) also heavily contribute to the multifactorial nature of most complex human diseases. Unfortunately, the molecular-level effects of these non-genetic factors and potential gene-by-environment interaction effects are difficult to systematically probe and measure. The sheer amount of potential variation and confounding factors severely hampers the power of large-scale analysis efforts such as environment-wide association studies [43, 44]. Nevertheless, similar to the discovery that tissue-specific networks are powerful "signal filters" that can accentuate the signal captured in known disease gene associations and quantitative genetics studies, an important future research direction will be the effective development and application of integrative network methods for the study of environmental effects underlying disease development.

One line of research that can not only help uncover environmental effects on disease, but also further advance studies characterizing the genetic basis of disease, is the development of methods that can tightly integrate the precision and power of model organism studies with human data. A key methodological and conceptual challenge here is how to effectively fuse model organism studies with human disease information under a single framework, rather than making discoveries separately (e.g., genetic screens in model organisms, GWAS studies in humans) and using the other purely for verification. Network-based methods have already been shown to be highly beneficial for cross-organism gene annotation transfer by considering not only sequence similarity but also conservation of biological function between orthologs [45, 46]. They demonstrate dramatic improvement in predicting gene–pathway membership and highlight the synergistic advantage of cross-organism analyses. Developing further cross-organism network-based approaches is thus also likely to benefit the study of human disease.

As high-throughput omics technologies continue to improve and large-scale data collection becomes more prevalent, new opportunities and challenges in network-based analyses are also emerging. An important future direction will be fueled by the increase in higher-resolution longitudinal data—the development of dynamic network models to not only model specific biological contexts (e.g., brain, intestine, kidney), but how these contexts change over time (e.g., the development of the brain, age-associated degenerative disorders), in reaction to stimuli (e.g., molecular effects of traumatic brain injury), and potentially, crosstalk between contexts (e.g., the gut-brain connection and how it is affected by circadian rhythm). In addition, epigenetic factors, positioned at the interface between the environment and the genome, also have a profound impact on disease risk both within and across generations [47–50]. This realization, together with the development of innovative experimental methods, have brought forth several large-scale efforts to catalog the epigenome, including ENCODE [51], Roadmap [52], FANTOM [53], and BLUEPRINT [54], and these data have fueled significant progress in predicting the functional effects of noncoding variation using sophisticated computational methods [55–57]. These

Table 1. Public resources of integrated networks

		Type	Network analyses
HumanBase [4]	hb.flatironinstitute.org	144 integrated human tissue networks	NetWAS for re-prioritizing GWAS data Candidate disease gene prediction Multi-gene query
STRING [58]	string-db.org	Multi-organism networks	Multi-gene query
FunCoup [59]	funcoup.sbc.su.se	Multi-organism networks	Multi-gene query
IMP [60]	imp.princeton.edu	Multi-organism networks	Custom function prediction analysis Multi-gene query
GeneMania [61]	genemania.org	Multi-organism networks	Multi-gene query

advances in elucidating the impact of genetic variation on epigenetic factors are the first steps toward unraveling the complex interplay between the genome, epigenome, and environment, especially in the context of disease. Important next steps include better understanding the effects of environment on epigenetic variation, of epigenetic variation on gene expression, and of epistatic interactions. By modeling this interplay from a systems perspective in a manner that incorporates genetic variation, researchers could capture individual differences in their network representations, thus addressing an underlying assumption of all current network-based approaches—that single networks (even context-specific ones) are identical across individuals. In the long run, such approaches will be critical in integrating molecular-level models with whole-organism physiology and eventually are likely to become an integrated part of not only biomedical research, but also the development and application of precision medicine diagnoses and treatments.

Acknowledgments

V.Y. was supported in part by US NIH grant T32 HG003284. This work was supported by the NIH (R01 GM071966). O.G.T. is a senior fellow of the Genetic Networks program of the Canadian Institute for Advanced Research (CIFAR).

Received 1 April 2018;
Received in revised form 15 June 2018;
Accepted 3 July 2018
Available online 9 July 2018

Keywords:
integrative networks;
quantitative genetics data;
tissue specificity

Equal contributions.***Abbreviations used:***
GWAS, genome-wide association studies; PPI, protein-protein interaction; NetWAS, network-wide association study; SVM, support vector machine.

References

[1] E.A. Boyle, Y.I. Li, J.K. Pritchard, An expanded view of complex traits: from polygenic to omnigenic, Cell 169 (2017) 1177–1186.
[2] E.E. Eichler, J. Flint, G. Gibson, A. Kong, S.M. Leal, J.H. Moore, J.H. Nadeau, Missing heritability and strategies for finding the underlying causes of complex disease, Nat. Rev. Genet. 11 (2010) 446–450.
[3] T.A. Manolio, F.S. Collins, N.J. Cox, D.B. Goldstein, L.A. Hindorff, D.J. Hunter, M.I. McCarthy, E.M. Ramos, L.R. Cardon, A. Chakravarti, J.H. Cho, A.E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C.N. Rotimi, M. Slatkin, D. Valle, A.S. Whittemore, M. Boehnke, A.G. Clark, E.E. Eichler, G. Gibson, J.L. Haines, T.F.C. MacKay, S.A. McCarroll, P.M. Visscher, Finding the missing heritability of complex diseases, Nature 461 (2009) 747–753.
[4] C.S. Greene, A. Krishnan, A.K. Wong, E. Ricciotti, R.A. Zelaya, D.S. Himmelstein, R. Zhang, B.M. Hartmann, E. Zaslavsky, S.C. Sealfon, D.I. Chasman, G.A. Fitzgerald, K. Dolinski, T. Grosser, O.G. Troyanskaya, Understanding multicellular function and disease with human tissue-specific networks, Nat. Genet. 47 (2015) 569–576.
[5] S. Mostafavi, D. Ray, D. Warde-Farley, C. Grouios, Q. Morris, GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function, Genome Biol. 9 (2008) S4.
[6] C. Huttenhower, E.M. Haley, M.A. Hibbs, V. Dumeaux, D.R. Barrett, H.A. Coller, O.G. Troyanskaya, Exploring the human genome with functional maps, Genome Res. 19 (2009) 1093–1106.
[7] S. Saxena, P. Caroni, Selective neuronal vulnerability in neurodegenerative diseases: from stressor thresholds to degeneration, Neuron 71 (2011) 35–48.
[8] E. Pierson, Gtex Consortium, D. Koller, A. Battle, S. Mostafavi, K.G. Ardlie, G. Getz, F.A. Wright, M. Kellis, S. Volpi, E.T. Dermitzakis, Sharing and specificity of co-expression networks across 35 human tissues, PLoS Comput. Biol. 11 (2015), e1004220.
[9] J.L. Min, G. Nicholson, I. Halgrimsdottir, K. Almstrup, A. Petri, A. Barrett, M. Travers, N.W. Rayner, R. Mägi, F.H. Pettersson, J. Broxholme, M.J. Neville, Q.F. Wills, J. Cheeseman, GIANT Consortium, MolPAGE Consortium, M. Allen, C.C. Holmes, T.D. Spector, J. Fleckner, M.I. McCarthy, F. Karpe, C.M. Lindgren, K.T. Zondervan, Coexpression network analysis in abdominal and gluteal adipose tissue reveals regulatory genetic loci for metabolic syndrome and related phenotypes, PLoS Genet. 8 (2012), e1002505.
[10] M.P. Keller, Y. Choi, P. Wang, D.B. Davis, M.E. Rabaglia, A.T. Oler, D.S. Stapleton, C. Argmann, K.L. Schueler, S. Edwards, H.A. Steinberg, E. Chaibub Neto, R. Kleinhanz, S. Turner, M.K. Hellerstein, E.E. Schadt, B.S. Yandell, C. Kendziorski, A.D. Attie, A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility, Genome Res. 18 (2008) 706–716.
[11] R. Dobrin, J. Zhu, C. Molony, C. Argman, M.L. Parrish, S. Carlson, M.F. Allan, D. Pomp, E.E. Schadt, Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease, Genome Biol. 10 (2009) R55.
[12] A.J. Cornish, I. Filippis, A. David, M.J.E. Sternberg, Exploring the cellular basis of human disease through a large-scale mapping of deleterious genes to cell types, Genome Med. 7 (2015) 95.
[13] O. Magger, Y.Y. Waldman, E. Ruppin, R. Sharan, Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks, PLoS Comput. Biol. 8 (2012), e1002690.
[14] A. Bossi, B. Lehner, Tissue specificity and the human protein interaction network, Mol. Syst. Biol. 5 (2009) 260.
[15] Y. Wang, D.-Y. Cho, H. Lee, J. Fear, B. Oliver, T.M. Przytycka, NetREX: Network Rewiring using EXpression—Towards Context Specific Regulatory Networks, 2017, https://doi.org/10.1101/126664.
[16] GTEx Consortium, The Genotype-Tissue Expression (GTEx) project, Nat. Genet. 45 (2013) 580–585.
[17] P. Dao, Y.-A. Kim, D. Wojtowicz, S. Madan, R. Sharan, T.M. Przytycka, BeWith: a between-within method to discover

relationships between cancer modules via integrated analysis of mutual exclusivity, co-occurrence and functional interactions, PLoS Comput. Biol. 13 (2017), e1005695.
[18] B. Wang, A.M. Mezlini, F. Demir, M. Fiume, Z. Tu, M. Brudno, B. Haibe-Kains, A. Goldenberg, Similarity network fusion for aggregating data types on a genomic scale, Nat. Methods 11 (2014) 333–337.
[19] D. Silverbush, S. Cristea, G. Yanovich, T. Geiger, N. Beerenwinkel, R. Sharan, ModulOmics: Integrating Multi-Omics Data to Identify Cancer Driver Modules, 2018, https://doi.org/10.1101/288399.
[20] D.-Y. Cho, T.M. Przytycka, Dissecting cancer heterogeneity with a probabilistic genotype–phenotype model, Nucleic Acids Res. 41 (2013) 8011–8020.
[21] R. Shen, Q. Mo, N. Schultz, V.E. Seshan, A.B. Olshen, J. Huse, M. Ladanyi, C. Sander, Integrative subtype discovery in glioblastoma using iCluster, PLoS One 7 (2012), e35236.
[22] A. Krishnan, R. Zhang, V. Yao, C.L. Theesfeld, A.K. Wong, A. Tadych, N. Volfovsky, A. Packer, A. Lash, O.G. Troyanskaya, Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder, Nat. Neurosci. 19 (2016) 1454–1462.
[23] T.-L. Lee, M.J. Raygada, O.M. Rennert, Integrative gene network analysis provides novel regulatory relationships, genetic contributions and susceptible targets in autism spectrum disorders, Gene 496 (2012) 88–96.
[24] N.N. Parikshak, R. Luo, A. Zhang, H. Won, J.K. Lowe, V. Chandran, S. Horvath, D.H. Geschwind, Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism, Cell 155 (2013) 1008–1021.
[25] F. Hormozdiari, O. Penn, E. Borenstein, E.E. Eichler, The discovery of integrated gene networks for autism and related disorders, Genome Res. 25 (2015) 142–154.
[26] Y.-A. Kim, D.-Y. Cho, T.M. Przytycka, Understanding genotype-phenotype effects in cancer via network approaches, PLoS Comput. Biol. 12 (2016), e1004747.
[27] H.-Y. Chuang, E. Lee, Y.-T. Liu, D. Lee, T. Ideker, Network-based classification of breast cancer metastasis, Mol. Syst. Biol. 3 (2007) 140.
[28] M.D.M. Leiserson, F. Vandin, H.-T. Wu, J.R. Dobson, J.V. Eldridge, J.L. Thomas, A. Papoutsaki, Y. Kim, B. Niu, M. McLellan, M.S. Lawrence, A. Gonzalez-Perez, D. Tamborero, Y. Cheng, G.A. Ryslik, N. Lopez-Bigas, G. Getz, L. Ding, B.J. Raphael, Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes, Nat. Genet. 47 (2015) 106–114.
[29] C.S. Greene, O.G. Troyanskaya, Accurate evaluation and analysis of functional genomics data and methods, Ann. N. Y. Acad. Sci. 1260 (2012) 95–100.
[30] J. Xu, Y. Li, Discovering disease-genes by topological features in human protein-protein interaction network, Bioinformatics 22 (2006) 2800–2805.
[31] J. Chen, B.J. Aronow, A.G. Jegga, Disease candidate gene identification and prioritization using protein interaction networks, BMC Bioinf. 10 (2009) 73.
[32] S. Navlakha, C. Kingsford, The power of protein interaction networks for associating genes with diseases, Bioinformatics 26 (2010) 1057–1063.
[33] J. Yang, B. Benyamin, B.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, P.A. Madden, A.C. Heath, N.G. Martin, G.W. Montgomery, M.E. Goddard, P.M. Visscher, Common SNPs explain a large proportion of the heritability for human height, Nat. Genet. 42 (2010) 565–569.
[34] S. Köhler, S. Bauer, D. Horn, P.N. Robinson, Walking the interactome for prioritization of candidate disease genes, Am. J. Hum. Genet. 82 (2008) 949–958.
[35] O. Vanunu, O. Magger, E. Ruppin, T. Shlomi, R. Sharan, Associating genes and protein complexes with disease via network propagation, PLoS Comput. Biol. 6 (2010), e1000641.
[36] J. Zhu, Y. Qin, T. Liu, J. Wang, X. Zheng, Prioritization of candidate disease genes by topological similarity between disease and protein diffusion profiles, BMC Bioinf. 14 (Suppl. 5) (2013) S5.
[37] E.J. Rossin, K. Lage, S. Raychaudhuri, R.J. Xavier, D. Tatar, Y. Benita, International Inflammatory Bowel Disease Genetics Constortium, C. Cotsapas, M.J. Daly, Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology, PLoS Genet. 7 (2011), e1001273.
[38] P. Jia, S. Zheng, J. Long, W. Zheng, Z. Zhao, dmGWAS: dense module searching for genome-wide association studies in protein–protein interaction networks, Bioinformatics 27 (2010) 95–102.
[39] P. Jia, L. Wang, A.H. Fanous, C.N. Pato, T.L. Edwards, International Schizophrenia Consortium, Z. Zhao, Network-assisted investigation of combined causal signals from genome-wide association studies in schizophrenia, PLoS Comput. Biol. 8 (2012), e1002587.
[40] Network-based multiple sclerosis pathway analysis with GWAS data from 15,000 cases and 30,000 controls, Am. J. Hum. Genet. 92 (2013) 854–865.
[41] Y. Liu, M. Brossard, C. Sarnowski, A. Vaysse, M. Moffatt, P. Margaritte-Jeannin, F. Llinares-López, M.H. Dizier, M. Lathrop, W. Cookson, E. Bouzigon, F. Demenais, Network-assisted analysis of GWAS data identifies a functionally-relevant gene module for childhood-onset asthma, Sci. Rep. 7 (2017) 938.
[42] E. Cerami, E. Demir, N. Schultz, B.S. Taylor, C. Sander, Automated network analysis identifies core pathways in glioblastoma, PLoS One 5 (2010), e8918.
[43] M.A. Hall, S.M. Dudek, R. Goodloe, D.C. Crawford, S.A. Pendergrass, P. Peissig, M. Brilliant, C.A. McCarty, M.D. Ritchie, Environment-wide association study (EWAS) for type 2 diabetes in the Marshfield Personalized Medicine Research Project Biobank, Pac. Symp. Biocomput. (2014) 200–211.
[44] D.P. McGinnis, J.S. Brownstein, C.J. Patel, Environment-wide association study of blood pressure in the National Health and Nutrition Examination Survey (1999–2012), Sci. Rep. 6 (2016), 30373.
[45] C.Y. Park, A.K. Wong, C.S. Greene, J. Rowland, Y. Guan, L.A. Bongo, R.D. Burdine, O.G. Troyanskaya, Functional knowledge transfer for high-accuracy prediction of under-studied biological processes, PLoS Comput. Biol. 9 (2013), e1002957.
[46] M.D. Chikina, O.G. Troyanskaya, Accurate quantification of functional analogy among close homologs, PLoS Comput. Biol. 7 (2011), e1001074.
[47] M.K. Skinner, Environmental stress and epigenetic transgenerational inheritance, BMC Med. 12 (2014), https://doi.org/10.1186/s12916-014-0153-y.
[48] S. Sharma, T.K. Kelly, P.A. Jones, Epigenetics in cancer, Carcinogenesis 31 (2009) 27–36.
[49] J.-Y. Hwang, K.A. Aromolaran, R.S. Zukin, The emerging field of epigenetics in neurodegeneration and neuroprotection, Nat. Rev. Neurosci. 18 (2017) 347–361.

[50] A.P. Feinberg, The key role of epigenetics in human disease prevention and mitigation, N. Engl. J. Med. 378 (2018) 1323–1334.

[51] The ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome, Nature 489 (2012) 57–74.

[52] Roadmap Epigenomics Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M.J. Ziller, V. Amin, J.W. Whitaker, M.D. Schultz, L.D. Ward, A. Sarkar, G. Quon, R.S. Sandstrom, M.L. Eaton, Y.-C. Wu, A.R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R.A. Harris, N. Shoresh, C.B. Epstein, E. Gjoneska, D. Leung, W. Xie, R.D. Hawkins, R. Lister, C. Hong, P. Gascard, A.J. Mungall, R. Moore, E. Chuah, A. Tam, T.K. Canfield, R.S. Hansen, R. Kaul, P.J. Sabo, M.S. Bansal, A. Carles, J.R. Dixon, K.-H. Farh, S. Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T.R. Mercer, S.J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R.C. Sallari, K.T. Siebenthall, N.A. Sinnott-Armstrong, M. Stevens, R.E. Thurman, J. Wu, B. Zhang, X. Zhou, A.E. Beaudet, L.A. Boyer, P.L. De Jager, P.J. Farnham, S.J. Fisher, D. Haussler, S.J.M. Jones, W. Li, M.A. Marra, M.T. McManus, S. Sunyaev, J.A. Thomson, T.D. Tlsty, L.-H. Tsai, W. Wang, R.A. Waterland, M.Q. Zhang, L.H. Chadwick, B.E. Bernstein, J.F. Costello, J.R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J.A. Stamatoyannopoulos, T. Wang, M. Kellis, Integrative analysis of 111 reference human epigenomes, Nature 518 (2015) 317–330.

[53] R. Andersson, C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithel, B. Lilje, N. Rapin, F.O. Bagger, M. Jørgensen, P.R. Andersen, N. Bertin, O. Rackham, A.M. Burroughs, J.K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhata, S. Maeda, Y. Negishi, C.J. Mungall, T.F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C.O. Daub, P. Heutink, D.A. Hume, T.H. Jensen, H. Suzuki, Y. Hayashizaki, F. Müller, A.R.R. Forrest, P. Carninci, M. Rehli, A. Sandelin, An atlas of active enhancers across human cell types and tissues, Nature 507 (2014) 455–461.

[54] L. Chen, B. Ge, F.P. Casale, L. Vasquez, T. Kwan, D. Garrido-Martín, S. Watt, Y. Yan, K. Kundu, S. Ecker, A. Datta, D. Richardson, F. Burden, D. Mead, A.L. Mann, J.M. Fernandez, S. Rowlston, S.P. Wilder, S. Farrow, X. Shao, J.J. Lambourne, A. Redensek, C.A. Albers, V. Amstislavskiy, S. Ashford, K. Berentsen, L. Bomba, G. Bourque, D. Bujold, S. Busche, M. Caron, S.-H. Chen, W. Cheung, O. Delaneau, E.T. Dermitzakis, H. Elding, I. Colgiu, F.O. Bagger, P. Flicek, E. Habibi, V. Iotchkova, E. Janssen-Megens, B. Kim, H. Lehrach, E. Lowy, A. Mandoli, F. Matarese, M.T. Maurano, J.A. Morris, V. Pancaldi, F. Pourfarzad, K. Rehnstrom, A. Rendon, T. Risch, N. Sharifi, M.-M. Simon, M. Sultan, A. Valencia, K. Walter, S.-Y. Wang, M. Frontini, S.E. Antonarakis, L. Clarke, M.-L. Yaspo, S. Beck, R. Guigo, D. Rico, J.H.A. Martens, W.H. Ouwehand, T.W. Kuijpers, D.S. Paul, H.G. Stunnenberg, O. Stegle, K. Downes, T. Pastinen, N. Soranzo, Genetic drivers of epigenetic and transcriptional variation in human immune cells, Cell 167 (2016) 1398–1414. e24.

[55] J. Zhou, O.G. Troyanskaya, Predicting effects of noncoding variants with deep learning-based sequence model, Nat. Methods 12 (2015) 931–934.

[56] G.R.S. Ritchie, I. Dunham, E. Zeggini, P. Flicek, Functional annotation of noncoding sequence variants, Nat. Methods 11 (2014) 294–296.

[57] D. Lee, D.U. Gorkin, M. Baker, B.J. Strober, A.L. Asoni, A.S. McCallion, M.A. Beer, A method to predict the impact of regulatory variants from DNA sequence, Nat. Genet. 47 (2015) 955–961.

[58] D. Szklarczyk, J.H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N.T. Doncheva, A. Roth, P. Bork, L.J. Jensen, C. von Mering, The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible, Nucleic Acids Res. 45 (2017) D362–D368.

[59] C. Ogris, D. Guala, E.L.L. Sonnhammer, FunCoup 4: new species, data, and visualization, Nucleic Acids Res. 46 (2018) D601–D607.

[60] A.K. Wong, A. Krishnan, V. Yao, A. Tadych, O.G. Troyanskaya, IMP 2.0: a multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks, Nucleic Acids Res. 43 (2015) W128–W133.

[61] K. Zuberi, M. Franz, H. Rodriguez, J. Montojo, C.T. Lopes, G.D. Bader, Q. Morris, GeneMANIA prediction server 2013 update, Nucleic Acids Res. 41 (2013) W115–W122.