Published online 25 May 2018 *Nucleic Acids Research, 2018, Vol. 46, Web Server issue* ***W65–W70***
doi: 10.1093/nar/gky408

GIANT 2.0: genome-scale integrated analysis of gene networks in tissues

Aaron K. Wong[1], Arjun Krishnan[2,3] and Olga G. Troyanskaya[1,4,5,*]

[1]Center for Computational Biology, Flatiron Institute, Simons Foundation, New York, NY 10010, USA, [2]Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI 48824, USA, [3]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA, [4]Department of Computer Science, Princeton University, Princeton, NJ 08544, USA and [5]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA

Received February 23, 2018; Revised April 20, 2018; Editorial Decision April 30, 2018; Accepted May 07, 2018

ABSTRACT

GIANT2 (Genome-wide Integrated Analysis of gene Networks in Tissues) is an interactive web server that enables biomedical researchers to analyze their proteins and pathways of interest and generate hypotheses in the context of genome-scale functional maps of human tissues. The precise actions of genes are frequently dependent on their tissue context, yet direct assay of tissue-specific protein function and interactions remains infeasible in many normal human tissues and cell-types. With GIANT2, researchers can explore predicted tissue-specific functional roles of genes and reveal changes in those roles across tissues, all through interactive multi-network visualizations and analyses. Additionally, the NetWAS approach available through the server uses tissue-specific/cell-type networks predicted by GIANT2 to re-prioritize statistical associations from GWAS studies and identify disease-associated genes. GIANT2 predicts tissue-specific interactions by integrating diverse functional genomics data from now over 61 400 experiments for 283 diverse tissues and cell-types. GIANT2 does not require any registration or installation and is freely available for use at http://giant-v2.princeton.edu.

INTRODUCTION

Tissue and cell type specificity are critical aspects of complex human disease. From impaired insulin signaling in diabetes (1,2) to neuronal loss in Parkinson's disease (3–5), understanding tissue- and cell-lineage specific processes is necessary in elucidating disease pathophysiology and disease-gene relationships. However, direct assay of tissue-specific function is highly challenging and in many human tissues and cell-types remains infeasible. Yet mapping these tissue-specific interactions is key to understanding pathway action in different tissues and their role in the manifestation of human disease.

Many resources collect and provide access to rich functional genomics data. For example, resources such as BioGRID (6) and Reactome (7) curate interaction data for querying and visualization. These data, however, represent global pathway function and cannot distinguish the tissue-specific actions of genes. Some resources such as Gene Expression Tissue Project (GTEx) (8) enable access to a rich collection of tissue expression profiles, and more broadly, NCBI GEO (9) provides search of thousands of gene expression experiments. Altogether, these resources provide measurements of genes' cellular activity, however they must be integrated to understand the precise functions of genes, particularly in a multi-cellular context. Successful methods by us (10–12) and others (13,14) can integrate these functional genomics data to predict functional interactions in human, many of which are accessible through a web server (13–15). However, these predictions lack tissue-specificity and none of them capture tissue and cell-type specific gene function, critical to understanding the complex and context-specific action of genes. Further, none of these resources can leverage tissue-specific interactions to aid researchers in the analysis of quantitative genetics data.

GIANT (Genome-wide Analysis of gene Networks in Tissues), introduced in 2015 (16), is a prediction server for human tissue-specific gene interactions that enables biomedical researchers to interrogate tissue-specific action through multi-network visualizations and analyses. Researchers can interact with GIANT by submitting individual genes or gene sets of interest for real-time integration of thousands of functional genomics experiments to predict tissue-specific interactions relevant to these genes and related processes. GIANT will return dynamic, interactive visualizations of predicted tissue-specific maps of the queried genes and tissues, and network-driven predictions of gene function and disease association. Additionally, GIANT al-

*To whom correspondence should be addressed. Tel: +1 609 258 1749; Fax: +1 609 258 1771; Email: ogt@genomics.princeton.edu

lows users to run NetWAS (16), a machine-learning based method that leverages tissue-specific interactions to reprioritize genome-wide association data and identify disease-associated genes. NetWAS analysis is performed entirely server-side, requiring no software installation or specific computational resources from the users. In addition to user-friendly, interactive visualizations, all predicted networks and user's NetWAS results are available for download.

The probabilistic model used in GIANT2 infers tissue-specific interactions from large data compendia by simultaneously extracting functional and tissue or cell-type specific signals, and has been extensively evaluated in our previous work (16). We showed that GIANT networks could predict the lineage-specific response to IL-1B stimulation in blood vessel, which was then experimentally confirmed. This result was not exclusive to blood vessel—we additionally showed that GIANT made accurate predictions for tissue- and cell-lineage-specific response post IL-1B simulation for all tissues and cell-types for which public data were available. Furthermore, GIANT could capture the changing functional roles of LEF1 across tissues, and map the disease-disease associations of Parkinson's disease. We introduced NetWAS, a method to effectively re-prioritize statistical associations from a GWAS study with predicted tissue-specific interactions. With this approach, GIANT reprioritized associations from a hypertension study, correctly identifying known hypertension genes, disease-related processes, and drug targets (without any prior knowledge of disease) and identified many candidate disease genes. GIANT has been continuously developed since original publication and here we describe the major updates to the server.

SYSTEM DESCRIPTION AND UPDATES

A GIANT prediction starts with a set of genes and one or more tissues of interest specified by the user (Figure 1A). The server predicts the likelihood of functional relationships between these genes and to all other genes in the human genome, for each of the queried tissues, by probabilistically integrating thousands of genome-scale experiments in a tissue-specific manner (Figure 1B). The results are presented to the user as a gene network for each queried tissue, with posterior probabilities of functional relationships for the genes of interest specific to that tissue (Figure 1C). These predictions can reveal the tissue-specific pathway partners or functional roles of the genes of interest. GIANT server provides extensive user-friendly visualizations enabling the user to seamlessly explore these predicted networks. Users can adjust the visualization to suit their biological question by filtering interactions by confidence level, or limiting the network to the highest-connected genes. Additionally, GIANT provides dynamic gene enrichment analysis of the queried network (Figure 1D). Gene Ontology (GO) biological process (17), Kyoto Encyclopedia of Genes and Genomes (KEGG) (18) pathway and Online Mendelian Inheritance in Man (OMIM) (19) disease-gene enrichments are calculated in real-time as users adjust the visualized network. These version-controlled gene sets are downloaded from the Tribe web server (bioRxiv: https://doi.org/10.1101/055913) and made available on GIANT. These analyses aid interpretation of large gene sets, which are often the outcome of a high-throughput experiment, and help generate hypotheses for experimental follow-up.

A key feature of GIANT networks is the ability to delineate the tissue-specific changes of multifunctional genes. The GIANT web server enables this feature with advanced multi-network visualization. When users query GIANT with multiple tissues, gene interactions for each tissue are simultaneously predicted and displayed as separate networks with a coordinated layout. Genes that are shared across tissues are both highlighted visually and positioned similarly in their respective views. Interactions with a tissue-network are mirrored across network views. Altogether, these features are designed to aid interpretation of genes' changing interaction partners, and thus biological function, across tissues and cell-types.

GIANT also uses tissue-specific networks to make predictions that provide novel hypotheses related to human disease. NetWAS, in conjunction with tissue-specific/cell-type networks predicted by GIANT, effectively re-prioritizes statistical associations from distinct GWAS to identify disease-associated genes. Biologists can submit a GWAS result file, select a tissue relevant to the studied phenotype, and run NetWAS on the GIANT server.

The newest version of GIANT doubles the number of tissues and cell-types for which it can make on-the-fly functional network predictions to 283, including networks for 105 specific cell types (compared to 144 total networks and 23 specific cell types in the original GIANT release). Many of these cell types (and even tissues) are very challenging or impossible to assay experimentally in humans, with predictions from the GIANT server providing the only systems-level molecular coverage. We have carefully collected tissue-gene gold standards from established genomics resources (GTEx (8) and FANTOM5 (20)), improving both gene and tissue/cell-type coverage as compared to prior sources (21). These tissue-expression profiles are used to define tissue-gene relationships and to weight gene pairs by tissue-specificity during model training (See Supplemental Methods). We have also adopted a more uniform and well-maintained ontology of tissues and cell types (UBERON (22) and Cell Ontology (22)). This resulted in both a substantial increase in training data for each tissue, and in the total tissues and cell-types for which we could confidently predict interactions. Furthermore, the 283 GIANT2 network predictions are made based on over 61 400 experiments from 24 930 publications, spanning diverse data types (e.g. mRNA expression and protein-protein interaction data). This is a 60% increase in the experimental coverage compared to GIANT's original release. The updated web-server has been available and running for a year.

EXAMPLE USE CASE: MULTI-TISSUE ANALYSIS

With GIANT2, biologists can interrogate gene function in 283 diverse tissues and cell-types with multi-network visualizations and analyses. GIANT can reveal the changing roles of multifunctional genes by comparing the predicted tissue-specific interactions, the enriched biological processes, and gene-disease associations across tissues.

In Figure 2, the user queries the multifunctional gene PARK7 in two tissues: *brain* and *skeletal muscle tissue*.

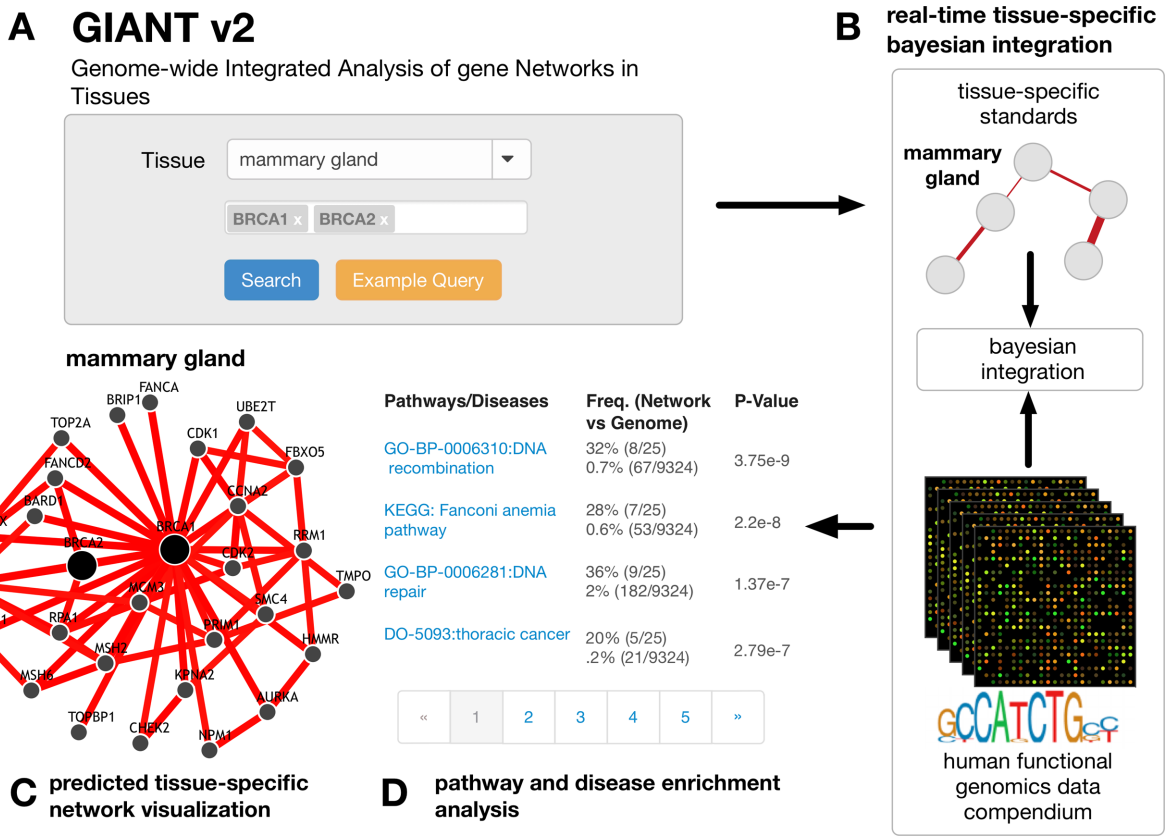


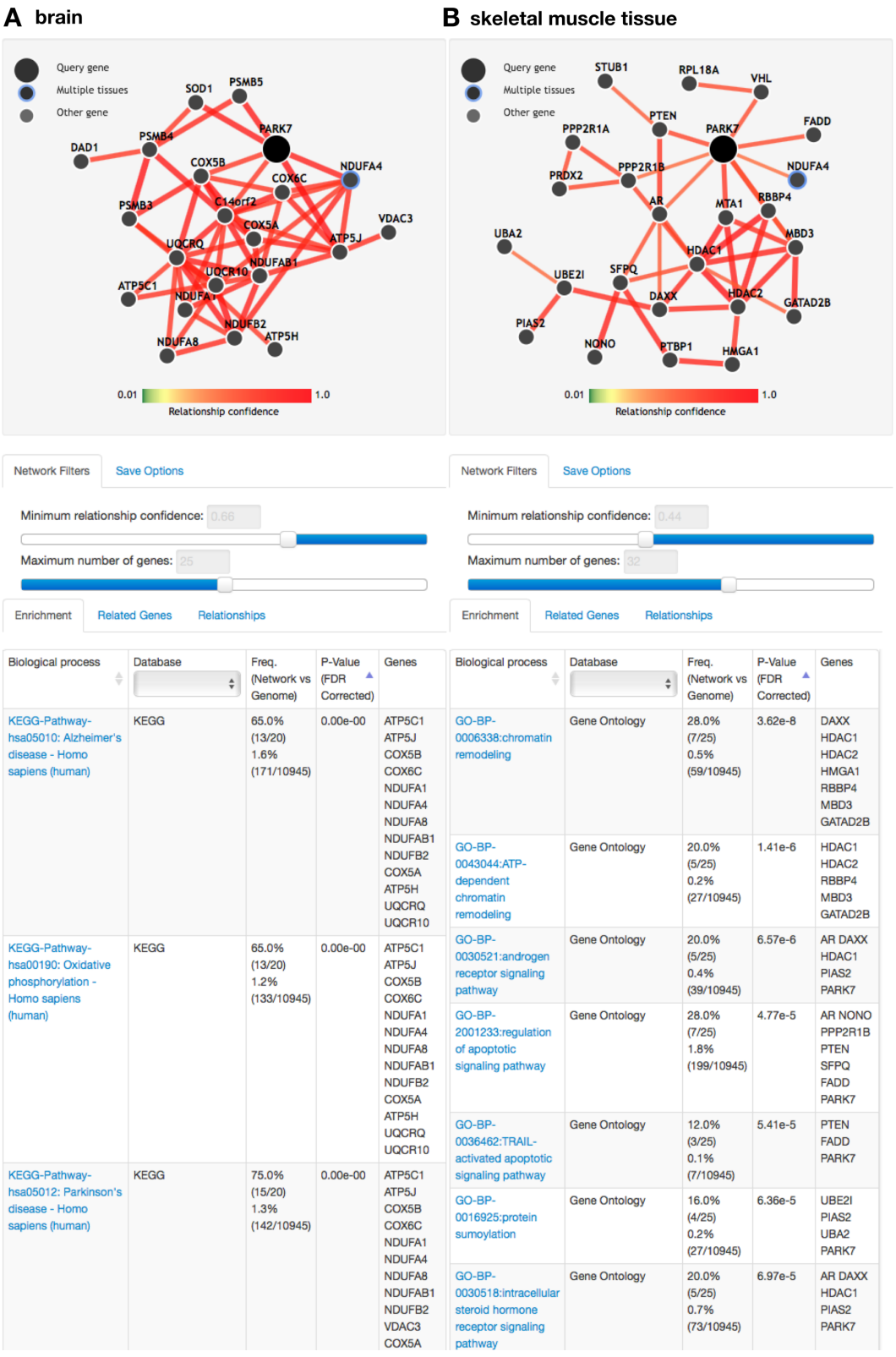
Figure 1. A schematic of the GIANT tissue-specific interaction prediction server. (**A**) GIANT is queried with two genes BRCA1 and BRCA2 in the *mammary gland* tissue. For optimal performance, we suggest that users query less than ten genes, in at most four tissues. The server response time increases with the number of queried genes and tissues (typically 2 seconds per gene/tissue). (**B**) GIANT integrates thousands of datasets from the human data compendium on-the-fly and predicts interactions to BRCA1 and BRCA2 with pre-computed tissue-specific Bayesian models. (**C**) The predicted interactions to BRCA1 and BRCA2 are shown as a network visualization where edges are predicted posterior probabilities of two genes functionally interacting in mammary gland. (**D**) Additional pathway and disease enrichment analysis of the displayed network is available to the user.

GIANT returns predicted tissue-specific interactions to PARK7 in the two tissues and displays them as separate network views. The interactions are visualized with a coordinated layout where common genes have the same position in their respective network visualizations. The PARK7 interaction partners in brain and skeletal muscle tissue are considerably different, reflecting the different functional roles of PARK7. Notably, in the brain network (Figure 2A), PARK7 and its partners are significantly enriched for genes involved in Parkinson's disease (PD), consistent with PARK7's known role in familial PD (23). In skeletal muscle tissue (Figure 2B), PARK7 interaction partners are highly enriched for 'androgen receptor signaling pathway'. Human PARK7 has been previously established as a regulator of androgen receptor (24,25), whose signaling contributes to muscle mass maintenance (26), and PARK7 orthologs have been specifically linked to muscle hypertrophy (27). Thus, as shown with this example, GIANT-predicted tissue-specific networks are able to distinguish the distinct functional roles of PARK7, revealed through differences in predicted interactions. These predictions might help biomedical researchers studying PARK7 as a therapeutic target understand its tissue-specific pleiotropic effects.

EXAMPLE USE CASE: NETWORK-GUIDED GWAS

Most complex diseases have tissue-specific origins and manifestations. With NetWAS, the tissue-specific/cell-type interactions captured in GIANT networks are used to re-prioritize results from a genome-wide association study of interest to the user. NetWAS is premised on the idea that top GWAS associations are enriched with disease-relevant genes, even if they fall below statistical significance (16). By learning the connectivity patterns of these top genes in relevant tissue networks, NetWAS can further enrich for phenotype-associated genes in a genome-wide re-ranking of the GWAS.

NetWAS trains a support-vector-machine (SVM), where the features of the SVM are interactions between genes in the selected tissue networks, positive labels are genes whose *P*-value fall below a selected cutoff, and negative labels are random genes above the cutoff. The SVM classifies—with five-fold cross-validation—all genes in the genome based on the tissue-specific interactions of the top GWAS genes. Note that no prior disease knowledge is used in this process - all disease signal is extracted from the GWAS study. Thus, NetWAS is discovery driven, where the GWAS itself is used to identify connectivity patterns rather than limited and potentially biased prior disease knowledge.



Biological process	Database	Freq. (Network vs Genome)	P-Value (FDR Corrected)	Genes
KEGG-Pathway-hsa05010: Alzheimer's disease - Homo sapiens (human)	KEGG	65.0% (13/20) 1.6% (171/10945)	0.00e-00	ATP5C1 ATP5J COX5B COX6C NDUFA1 NDUFA4 NDUFA8 NDUFAB1 NDUFB2 COX5A ATP5H UQCRQ UQCR10
KEGG-Pathway-hsa00190: Oxidative phosphorylation - Homo sapiens (human)	KEGG	65.0% (13/20) 1.2% (133/10945)	0.00e-00	ATP5C1 ATP5J COX5B COX6C NDUFA1 NDUFA4 NDUFA8 NDUFAB1 NDUFB2 COX5A ATP5H UQCRQ UQCR10
KEGG-Pathway-hsa05012: Parkinson's disease - Homo sapiens (human)	KEGG	75.0% (15/20) 1.3% (142/10945)	0.00e-00	ATP5C1 ATP5J COX5B COX6C NDUFA1 NDUFA4 NDUFA8 NDUFAB1 NDUFB2 VDAC3 COX5A

Biological process	Database	Freq. (Network vs Genome)	P-Value (FDR Corrected)	Genes
GO-BP-0006338:chromatin remodeling	Gene Ontology	28.0% (7/25) 0.5% (59/10945)	3.62e-8	DAXX HDAC1 HDAC2 HMGA1 RBBP4 MBD3 GATAD2B
GO-BP-0043044:ATP-dependent chromatin remodeling	Gene Ontology	20.0% (5/25) 0.2% (27/10945)	1.41e-6	HDAC1 HDAC2 RBBP4 MBD3 GATAD2B
GO-BP-0030521:androgen receptor signaling pathway	Gene Ontology	20.0% (5/25) 0.4% (39/10945)	6.57e-6	AR DAXX HDAC1 PIAS2 PARK7
GO-BP-2001233:regulation of apoptotic signaling pathway	Gene Ontology	28.0% (7/25) 1.8% (199/10945)	4.77e-5	AR NONO PPP2R1B PTEN SFPQ FADD PARK7
GO-BP-0036462:TRAIL-activated apoptotic signaling pathway	Gene Ontology	12.0% (3/25) 0.1% (7/10945)	5.41e-5	PTEN FADD PARK7
GO-BP-0016925:protein sumoylation	Gene Ontology	16.0% (4/25) 0.2% (27/10945)	6.36e-5	UBE2I PIAS2 UBA2 PARK7
GO-BP-0030518:intracellular steroid hormone receptor signaling pathway	Gene Ontology	20.0% (5/25) 0.7% (73/10945)	6.97e-5	AR DAXX HDAC1 PIAS2 PARK7

Figure 2. GIANT multi-network visualization. (**A**) Network showing PARK7 interactions in *brain*. The genes with the highest confidence interactions to PARK7 in *brain* are highly enriched for Parkinson's disease associated genes, among others. (**B**) In the skeletal muscle tissue network, PARK7 and its neighbors are enriched for androgen receptor signaling pathways.

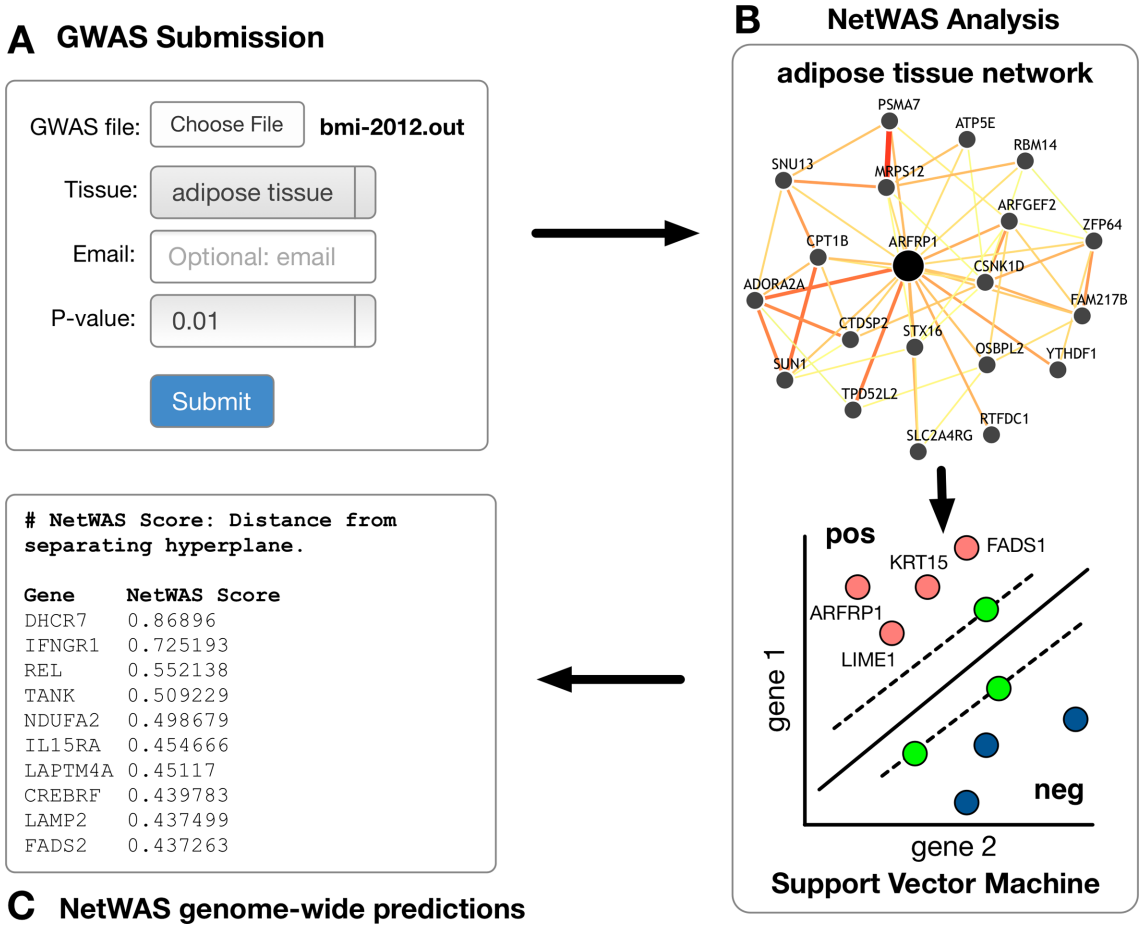


Figure 3. A schematic of a NetWAS analysis. (**A**) A user submits a BMI GWAS result file consisting of gene-wise *P*-values and selects 'adipose tissue' as a relevant tissue and a *P*-value cutoff of 0.01. (**B**) NetWAS builds an SVM where features are the predicted tissue-specific interactions in adipose tissue, positive labels are genes in the BMI GWAS whose *P*-value is less than 0.01 and negative labels are random genes whose *P*-value is above 0.01. (**C**) NetWAS results are a re-ranking of all genes in the genome. The NetWAS score is the direct output of the SVM (i.e. distance to the separating hyperplane). Higher (positive) scores indicate that a gene is more likely to be associated with the studied trait. The results can be emailed to the user and are available directly through GIANT through a unique result-specific URL.

Figure 3 shows the NetWAS workflow for a GWAS of Body Mass Index (BMI) (28) with adipose tissue. A researcher with a GWAS result (bmi-2012.out) uploads her result file of gene association *P*-values using the GIANT (Figure 3A) web form. GIANT supports many file formats of commonly used tools that pool SNP associations to gene-wise *P*-values (29)—a required step before running NetWAS. The user selects two options taking into account her particular GWAS result: (i) a *P*-value cutoff used to select 'top' genes for training (the default is 0.01, which has been successfully applied in many NetWAS analyses (16)) and (ii) a tissue/cell-type relevant to the studied phenotype (adipose tissue). Upon submission, NetWAS is run on GIANT servers (Figure 3B) and does not require software installation or dedicated computational resources by the user. The result is a genome-wide re-ranking of genes driven by their network similarity to the top GWAS genes (Figure 3C). This re-ranking has been shown (16) to improve disease association signal over the original GWAS in this BMI study (28), and many others (30,31).

SUMMARY

GIANT is a dynamic, interactive web server that offers biologists a diverse collection of tools to answer experimental questions in the context of human tissue-specific functional maps. GIANT integrates thousands of genomics datasets to predict gene interactions in 283 tissues and cell-types, and enables re-analysis of quantitative genetics data through NetWAS. These tools are accessible to biomedical researchers through a user-friendly interface with flexible visualizations. Importantly, the tools and analyses in GIANT are data-driven and reach beyond existing, curated biological knowledge. Thus, GIANT can complement the tools of modern biologists to interpret and guide experiments involving tissue-specific gene action.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Science Foundation (NSF) career award [DBI-0546275]; National Institutes of Health [R01 GM071966]; O.G.T. is a senior fellow of the Genetic Networks program of the Canadian Institute for Advanced Research (CIFAR). Funding for open access charge: Flatiron Institute.
Conflict of interest statement. None declared.

REFERENCES

1. Smith,U. (2002) Impaired ('diabetic') insulin signaling and action occur in fat cells long before glucose intolerance–is insulin resistance initiated in the adipose tissue? *Int. J. Obes. Relat. Metab. Disord.*, **26**, 897–904.
2. Kubota,T., Kubota,N., Kumagai,H., Yamaguchi,S., Kozono,H., Takahashi,T., Inoue,M., Itoh,S., Takamoto,I., Sasako,T. *et al.* (2011) Impaired insulin signaling in endothelial cells reduces insulin-induced glucose uptake by skeletal muscle. *Cell Metab.*, **13**, 294–307.
3. Levy,O.A., Malagelada,C. and Greene,L.A. (2009) Cell death pathways in Parkinson's disease: proximal triggers, distal effectors, and final steps. *Apoptosis*, **14**, 478–500.
4. Polymeropoulos,M.H., Lavedan,C., Leroy,E., Ide,S.E., Dehejia,A., Dutra,A., Pike,B., Root,H., Rubenstein,J., Boyer,R. *et al.* (1997) Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science*, **276**, 2045–2047.
5. Michel,P.P., Hirsch,E.C. and Hunot,S. (2016) Understanding dopaminergic cell death pathways in parkinson disease. *Neuron*, **90**, 675–691.
6. Chatr-Aryamontri,A., Oughtred,R., Boucher,L., Rust,J., Chang,C., Kolas,N.K., O'Donnell,L., Oster,S., Theesfeld,C., Sellam,A. *et al.* (2017) The BioGRID interaction database: 2017 update. *Nucleic Acids Res.*, **45**, D369–D379.
7. Fabregat,A., Sidiropoulos,K., Garapati,P., Gillespie,M., Hausmann,K., Haw,R., Jassal,B., Jupe,S., Korninger,F., McKay,S. *et al.* (2016) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **44**, D481–D487.
8. Consortium,G.T. (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
9. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.*, **41**, D991–D995.
10. Park,C.Y., Wong,A.K., Greene,C.S., Rowland,J., Guan,Y., Bongo,L.A., Burdine,R.D. and Troyanskaya,O.G. (2013) Functional knowledge transfer for high-accuracy prediction of under-studied biological processes. *PLoS Comput. Biol.*, **9**, e1002957.
11. Huttenhower,C., Haley,E.M., Hibbs,M.A., Dumeaux,V., Barrett,D.R., Coller,H.A and Troyanskaya,O.G. (2009) Exploring the human genome with functional maps. *Genome Res.*, **19**, 1093–1106.
12. Myers,C.L. and Troyanskaya,O.G. (2007) Context-sensitive data integration and prediction of biological networks. *Bioinformatics*, **23**, 2322–2330.
13. Zuberi,K., Franz,M., Rodriguez,H., Montojo,J., Lopes,C.T., Bader,G.D. and Morris,Q. (2013) GeneMANIA prediction server 2013 update. *Nucleic Acids Res.*, **41**, W115–W122.
14. Ogris,C., Guala,D. and Sonnhammer,E.L.L. (2018) FunCoup 4: new species, data, and visualization. *Nucleic Acids Res.*, **46**, D601–D607.
15. Wong,A.K., Krishnan,A., Yao,V., Tadych,A. and Troyanskaya,O.G. (2015) IMP 2.0: a multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks. *Nucleic Acids Res.*, **43**, W128–W133.
16. 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 multicellular function and disease with human tissue-specific networks. *Nat. Genet.*, **47**, 569–576.
17. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.*, **25**, 25–29.
18. Kanehisa,M., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, **44**, D457–D462.
19. Amberger,J.S., Bocchini,C.A., Schiettecatte,F., Scott,A.F. and Hamosh,A. (2015) OMIM.org: online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.*, **43**, D789–D798.
20. Lizio,M., Harshbarger,J., Shimoji,H., Severin,J., Kasukawa,T., Sahin,S., Abugessaisa,I., Fukuda,S., Hori,F., Ishikawa-Kato,S. *et al.* (2015) Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.*, **16**, 22.
21. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human protein reference Database–2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
22. Mungall,C.J., Torniai,C., Gkoutos,G.V., Lewis,S.E. and Haendel,M.A. (2012) Uberon, an integrative multi-species anatomy ontology. *Genome Biol.*, **13**, R5.
23. Bonifati,V., Rizzu,P., van Baren,M.J., Schaap,O., Breedveld,G.J., Krieger,E., Dekker,M.C., Squitieri,F., Ibanez,P., Joosse,M. *et al.* (2003) Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. *Science*, **299**, 256–259.
24. Takahashi,K., Taira,T., Niki,T., Seino,C., Iguchi-Ariga,S.M. and Ariga,H. (2001) DJ-1 positively regulates the androgen receptor by impairing the binding of PIASx alpha to the receptor. *J. Biol. Chem.*, **276**, 37556–37563.
25. Niki,T., Takahashi-Niki,K., Taira,T., Iguchi-Ariga,S.M. and Ariga,H. (2003) DJBP: a novel DJ-1-binding protein, negatively regulates the androgen receptor by recruiting histone deacetylase complex, and DJ-1 antagonizes this inhibition by abrogation of this complex. *Mol. Cancer Res.*, **1**, 247–261.
26. Ophoff,J., Van Proeyen,K., Callewaert,F., De Gendt,K., De Bock,K., Vanden Bosch,A., Verhoeven,G., Hespel,P. and Vanderschueren,D. (2009) Androgen signaling in myocytes contributes to the maintenance of muscle mass and fiber type regulation but not to muscle strength or fatigue. *Endocrinology*, **150**, 3558–3566.
27. Yu,H., Waddell,J.N., Kuang,S. and Bidwell,C.A. (2014) Park7 expression influences myotube size and myosin expression in muscle. *PLoS One*, **9**, e92030.
28. Randall,J.C., Winkler,T.W., Kutalik,Z., Berndt,S.I., Jackson,A.U., Monda,K.L., Kilpelainen,T.O., Esko,T., Magi,R., Li,S. *et al.* (2013) Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.*, **9**, e1003500.
29. Mishra,A. and Macgregor,S. (2015) VEGAS2: Software for more flexible Gene-Based testing. *Twin Res. Hum. Genet.*, **18**, 86–91.
30. Ridker,P.M., Chasman,D.I., Zee,R.Y., Parker,A., Rose,L., Cook,N.R., Buring,J.E. and Women's Genome Health Study Working, G. (2008) Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women. *Clin. Chem.*, **54**, 249–255.
31. Fritsche,L.G., Chen,W., Schu,M., Yaspan,B.L., Yu,Y., Thorleifsson,G., Zack,D.J., Arakawa,S., Cipriani,V., Ripke,S. *et al.* (2013) Seven new loci associated with age-related macular degeneration. *Nat. Genet.*, **45**, 433–439.

Supplemental Methods

The methods in this GIANT update follow as previously described in the 2015 paper (1), except where noted.

Data download and processing

We collected 1540 genome-scale data sets including 61,400 conditions from an estimated 24,900 publications. Interaction data were downloaded from BioGRID (2), IntAct (3), MINT (4) and MIPs (5). Shared transcription factor regulation was estimated from JASPAR (6) binding motifs. Chemical and genetic perturbation (c2:CGP) and microRNA target (c3:MIR) profiles were downloaded from the Molecular Signatures Database (MSigDB) (7). All datasets were processed as previously described (1) and those methods are reproduced here with updated summary counts:

BioGRID edges were discretized into five bins, labeled 0 to 4, where the bin number reflected the number of experiments supporting the interaction. For the remaining databases, edges were discretized into the presence or absence of an interaction.

To estimate shared transcription factor regulation, binding motifs were downloaded from JASPAR. Genes were scored for the presence of transcription factor binding sites using the MEME software suite (8). FIMO (9) was used to scan for each transcription factor profile within the 1-kb sequence upstream of each gene. Motif matches were treated as binary scores (present if $P < 0.001$). The final score for each gene pair was obtained by calculating the Pearson correlation between the motif association vectors for the genes.

Chemical and genetic perturbation (c2:CGP) and microRNA target (c3:MIR) profiles were downloaded from the Molecular Signatures Database (MSigDB) (7). Each gene pair's score was the sum of shared profiles weighted by the specificity of each profile (1/len(genes)). The resulting scores were converted to *z* scores and discretized into bins ((−infinity, −1.5), [−1.5, −0.5), [–0.5, 0.5), [0.5, 1.5), [1.5, 2.5), [2.5, 3.5), [3.5, 4.5), [4.5, infinity)).

We downloaded all gene expression data sets from NCBI's Gene Expression Omnibus (GEO) (10) and collapsed duplicate samples. GEO contains 1533 human data sets. Genes with more than 30% of values missing were removed, and remaining missing values were imputed using ten neighbors. Non-log-transformed data sets were log transformed. Expression measurements were summarized to Entrez identifiers, and duplicate identifiers were merged. The Pearson correlation was calculated for each gene pair, normalized with Fisher's z transform, mean subtracted and divided by the standard deviation. The resulting z scores were discretized into bins ((−infinity, −1.5), [–1.5, −0.5), [–0.5, 0.5), [0.5, 1.5), [1.5, 2.5), [2.5, 3.5), [3.5, infinity)).

Hierarchically aware gold standard construction

***Functional knowledge extraction*.** We constructed a tissue-naive functional relationship gold standard as described previously but with updated GO annotations (11). We processed experimentally derived gene annotations (GO evidence codes: EXP, IDA, IPI, IMP, IGI and IEP) from a set of 618 expert-selected GO biological process terms. To increase the coverage of functional interactions, we transferred experimentally confirmed mouse GO annotations to human functional analogs identified by FKT (12), a high-specificity annotation transfer method, for the 592 GO terms with mouse annotations. This resulted in a tissue-naive gold standard of 1,393,224 functionally related gene pairs (positive examples) and 16,281,559 potentially unrelated pairs (negative examples).

Ontology-aware gene-tissue annotations. Gene-to-tissue annotations were derived from GTEx (13) and FANTOM5 (14) RNA-seq data. Expression data, in transcripts per million (TPM) units, from both resources were quantile normalized to enable cross-sample comparisons. Samples were mapped to tissue and cell-type terms in UBERON (15) and Cell Ontology and propagated along a shared hierarchy. Genes were assigned to tissues (designated as 'tissue-expressed') given the following rules:

1. A gene was declared in a *sample* as:
    - 'ON' if its TPM ≥ 6 and TPM ≥ *gene_median*
    - 'OFF' if its TPM ≤ 1 and TPM < *gene_median*
2. A gene was declared in a *tissue* as:
    - 'not-expressed' if the gene is 'OFF' in ≥ 3 samples and 'ON' in ≤ 1 sample
    - 'tissue-expressed' if 'ON' in ≥ 3 samples and 'not-expressed' ≥ 2 unrelated tissues (based on the tissue/cell-type ontology)

where *gene_median* is the median TPM of a gene across all samples.

***Integration of tissue-specific and functional knowledge*.** We combined the above gene-to-tissue annotations with the tissue-naive functional gold standard to construct a hierarchical tissue-specific knowledgebase. We labeled each gene pair (positive or negative) in the tissue-naive functional relationship standard as specifically coexpressed in a tissue if both genes were designated as tissue-expressed (T, T).

After labeling specifically coexpressed gene pairs (edges) across all tissues, we considered four classes of edges—C1, C2, C3 and C4—to constitute each tissue standard.

C1: positive functional edges between genes specifically co-expressed in the tissue [T–T].
C2: positive functional edges between a gene expressed in the tissue and another specifically expressed in an unrelated tissue [T–T′].
C3: negative functional edges between genes specifically co-expressed in the tissue [T–T].
C4: negative functional edges between one gene expressed in the tissue and another specifically expressed in an unrelated tissue [T–T′].

Among the four tissue classes, C1 represented tissue-specific functional relationships. To identify tissue-specific relationships, we constructed a specific gold standard for each tissue by labeling edges in C1 as positives and edges in the other classes as negatives. Because C3 is defined on the basis of tissue-expressed genes and C2 and C4 are defined on the basis of non-expressed genes, the number of edges in these classes varied across tissues according to how specific (cell type, tissue, organ or system), well studied (or easily studied) and well curated (literature bias) they are. To construct comparable networks across tissues, we used a negative set composed of equal proportions of edges from C2, C3 and C4.

***Tissue-specific weighting*.** We calculated a weight for each C1 edge corresponding to its tissue-specificity. For every gene represented in FANTOM5 (14) and GTEx (13) (as processed above), we calculated tissue expression as the median TPM of the gene across all samples corresponding to the tissue. We calculated tissue-specificity as the z-score of a gene's expression in a tissue (x_i) compared to its mean (u) and standard deviation expression across all non-related tissues - defined as tissues that do not share the same tissue system (e.g. for brain, non-related tissue were all tissues not part of the nervous system).

$$z = \frac{x_i - \mu}{\sigma}$$

$$c_{g1g2} = \max(z_{g1}, z_{g2}, 0)$$

For a positive edge with incident genes g_1 and g_2, c_{g1g2} is the number of counts the edge will contribute to the conditional probability table during the learning phase of the naive bayes classifier. Note that gene pairs whose constituent genes are not specifically expressed (both gene z-scores are less than 0) will effectively be excluded during learning.

Data Integration. We constructed functional networks from genome-scale data by performing a weighted tissue-specific Bayesian integration. We trained one naive Bayesian classifier for each tissue using the tissue-specific standards described above, where each positive edge was additionally weighted by the tissue-specific expression of the incident genes, as described above.

In each case, we constructed a class node, i.e., the presence or absence of a functional relationship between a pair of genes that is conditioned on nodes for each data set. For large-scale genomics data sets, the assumption of conditional independence required for a naive Bayes classifier is often not met, so we calculated and corrected for non-biological conditional dependency (12).

Each tissue model trained on the hierarchy-aware tissue-specific knowledge was used to make genome-wide predictions by estimating the probability of tissue-specific functional interaction between all pairs of genes. We also estimated the probability of global functional interactions for the tissue-naive network. We assigned a prior probability of a functional relationship of 0.1 for all models, allowing edge probabilities to be compared across tissues.

Network-based reprioritization of genome-wide association study.

NetWAS was implemented as previously described (1). We trained a support vector machine classifier using nominally significant ($P <$ user defined cutoff) genes as positive examples and 10,000 randomly selected non-significant ($P \geq$ user defined cutoff) genes as negatives. The classifier was constructed using the chosen tissue network, where the features of the classifier were the edge weights of the labeled examples to all the genes in the network. Genes were re-ranked using their distance from the hyperplane.

Bayesian integration and NetWAS analysis was performed with the open-source C++ software, Sleipnir Library for Computational Functional Genomics (16), available at: http://libsleipnir.bitbucket.io/

1. 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 multicellular function and disease with human tissue-specific networks. *Nature genetics*, **47**, 569-576.
2. Chatr-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N.K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A. *et al.* (2017) The BioGRID interaction database: 2017 update. *Nucleic acids research*, **45**, D369-D379.
3. Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U. *et al.* (2012) The IntAct molecular interaction database in 2012. *Nucleic acids research*, **40**, D841-846.
4. Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardozza, A.P., Santonico, E. *et al.* (2012) MINT, the molecular interaction database: 2012 update. *Nucleic acids research*, **40**, D857-861.
5. Mewes, H.W., Frishman, D., Guldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Morgenstern, B., Munsterkotter, M., Rudd, S. and Weil, B. (2002) MIPS: a database for genomes and protein sequences. *Nucleic acids research*, **30**, 31-34.
6. Portales-Casamar, E., Thongjuea, S., Kwon, A.T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W.W. and Sandelin, A. (2010) JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic acids research*, **38**, D105-110.
7. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, **102**, 15545-15550.
8. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic acids research*, **37**, W202-208.
9. Grant, C.E., Bailey, T.L. and Noble, W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics (Oxford, England)*, **27**, 1017-1018.
10. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets--update. *Nucleic acids research*, **41**, D991-995.
11. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*, **25**, 25-29.
12. Park, C.Y., Wong, A.K., Greene, C.S., Rowland, J., Guan, Y., Bongo, L.A., Burdine, R.D. and Troyanskaya, O.G. (2013) Functional knowledge transfer for high-accuracy prediction of under-studied biological processes. *PLoS computational biology*, **9**, e1002957.
13. Consortium, G.T. (2013) The Genotype-Tissue Expression (GTEx) project. *Nature genetics*, **45**, 580-585.
14. Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., Sahin, S., Abugessaisa, I., Fukuda, S., Hori, F., Ishikawa-Kato, S. *et al.* (2015) Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol*, **16**, 22.

15. Mungall, C.J., Torniai, C., Gkoutos, G.V., Lewis, S.E. and Haendel, M.A. (2012) Uberon, an integrative multi-species anatomy ontology. *Genome Biol*, **13**, R5.
16. Huttenhower, C., Schroeder, M., Chikina, M.D. and Troyanskaya, O.G. (2008) The Sleipnir library for computational functional genomics. *Bioinformatics (Oxford, England)*, **24**, 1559-1561.