HumanNet v2: human gene networks for disease research

Sohyun Hwang[1,2,3,†], Chan Yeong Kim[1,†], Sunmo Yang[1], Eiru Kim[4], Traver Hart [4], Edward M. Marcotte[2,5] and Insuk Lee[1,*]

[1]Department of Biotechnology, College of Life Science and Biotechnology, Yonsei University, Seoul 03722, Korea, [2]Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, University of Texas, Austin, TX 78712, USA, [3]Department of Biomedical Science, College of Life Science, CHA University, Seongnam-si 13496, Korea, [4]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX USA and [5]Department of Molecular Biosciences, University of Texas at Austin, TX 78712, USA

ABSTRACT

Human gene networks have proven useful in many aspects of disease research, with numerous network-based strategies developed for generating hypotheses about gene-disease-drug associations. The ability to predict and organize genes most relevant to a specific disease has proven especially important. We previously developed a human functional gene network, HumanNet, by integrating diverse types of omics data using Bayesian statistics framework and demonstrated its ability to retrieve disease genes. Here, we present HumanNet v2 (http://www.inetbio. org/humannet), a database of human gene networks, which was updated by incorporating new data types, extending data sources and improving network inference algorithms. HumanNet now comprises a hierarchy of human gene networks, allowing for more flexible incorporation of network information into studies. HumanNet performs well in ranking disease-linked gene sets with minimal literature-dependent biases. We observe that incorporating model organisms' protein–protein interactions does not markedly improve disease gene predictions, suggesting that many of the disease gene associations are now captured directly in human-derived datasets. With an improved interactive user interface for disease network analysis, we expect HumanNet will be a useful resource for network medicine.

INTRODUCTION

Human gene networks have been widely used to investigate genetic factors of diseases and therapeutic targets (1). Gene networks can also augment disease genomics information derived from expression profiles (2–4), whole exome sequencing (5,6) and genome-wide association studies (GWAS) (7,8) for the discovery of disease-associated genes. Edges of the gene networks may represent diverse types of associations between genes which can be mapped by both experimental and computational methods. Because appropriately integrating interaction information from diverse sources can improve the breadth and accuracy of a network, many integrated human gene networks have been developed and a variety of topological analysis algorithms have been applied to generate new hypotheses about gene-disease-drug associations.

We previously developed an integrated human functional gene network, HumanNet, and demonstrated its capability of disease gene predictions (9). In order to construct the network, we inferred functional associations between human genes from protein–protein interactions (PPI), co-citation of human genes across PubMed abstracts, co-occurrence of protein domains, co-expression of genes across samples and genomic context associations. In addition, interactions between evolutionarily conserved proteins of model organisms were transferred to the human gene network. Those networks, inferred from different types of data, were evaluated and integrated using a Bayesian statistical framework. Since the first release of HumanNet, the amount of publicly available omics data has increased substantially and network inference algorithms have also improved significantly, and thus we expected that updating HumanNet could provide a greatly enhanced resource for network medicine.

In this report, we present HumanNet v2, which offers substantial performance improvements over v1, especially for the disease gene predictions. A new feature of the updated HumanNet is a four level inclusive hierarchy of the human gene networks: the first level has two networks, HumanNet-PI comprising human-derived PPIs and HumanNet-CF based on co-functional links inferred

*To whom correspondence should be addressed. Tel: +822 2123 5559; Fax: +822 362 7265; Email: insuklee@yonsei.ac.kr
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

from various types of genomics data; the integration of HumanNet-PI and HumanNet-CF produces the second level network HumanNet-FN which is an integrated functional gene network; the third level has two extended functional networks by either co-citation (HumanNet-XC) or interologs (10) from other species (HumanNet-XI); and the fourth level network is the fully extended network (HumanNet-XN) that contains all above functional links (Figure 1A).

We benchmarked each of the networks for their ability to prioritize disease-linked gene sets with two different network-based algorithms. We observed HumanNet-XC and HumanNet-XN to have equally good or better performance than STRING v10.5 (11) and significantly better performance than other integrated human gene networks such as ConsensusPathDB (CPDB) (12), GIANT (7), GeneMANIA (13) and FunCoup (14). Time-stamped benchmarking strategy demonstrated that the improvements in performance of HumanNet extended beyond the incorporation of literature-based information. Interestingly, while we offer networks extended by IL for completeness, we observed no gains in disease gene prediction quality by their incorporation, suggesting that data measured directly in humans has reached a high level of predictive power for the disease gene network. Users can download edge information of various human gene networks and perform disease gene predictions and disease network analysis via a highly interactive user interface on the HumanNet web server (www.inetbio.org/humannet).

NETWORK DATABASE IMPROVEMENT

Four-level inclusive hierarchy of human gene networks

To provide flexibility in utilizing the network's information for various purposes, we designed HumanNet v2 with a four-level inclusive hierarchy of human gene networks comprising networks based on 10 distinct types of data (Figure 1A and Supplemental Table 1). The previous version of HumanNet was constructed based on only functional associations between genes, which can be supported by various types of biological data. The PPI assay was a traditional approach for mapping the functional associations between genes. Human gene networks based on only PPIs generally have a limited network coverage, because there are many functional associations that are not mediated by physical interactions between proteins. However, PPI networks have advantages in terms of the mechanistic interpretation of disease-associated mutations (15). Therefore, we decided to maintain a human gene network based on only PPIs separately as one of the first-level networks, HumanNet-PI, which contains 158 499 links among 15 352 genes, based on PPIs by high-throughput assays (HT) and literature-curated PPIs (LC).

In contrast to the PPI network, functional gene networks can be supported by diverse types of data (16), including PPIs. Despite lacking mechanistic information for the network links due to the broad edge definition, the typically high comprehensiveness of functional gene networks provides advantages in terms of generating functional hypotheses. We inferred co-functional associations between genes from six additional types of data: co-essentiality (CE) (17),

co-expression (CX) (18), associations by pathway database (DB), associations between protein domain profiles (DP) (19), associations by gene neighborhood (GN) (20) and associations between phylogenetic profiles (PG) (21). Network inference methods for each type of data are described in the Supplemental Methods. We integrated the six co-functional gene networks to generate another first-level network based on only inferred co-functional links from omics data, HumanNet-CF that contains 14 739 genes and 252 590 links. Integration of these two first-level networks produces the second-level network HumanNet-FN, an integrated functional gene network that contains 17 247 genes and 371 502 links.

Two networks at third-level were constructed based on the extended information of the functional associations by either co-citations (CC) across approximately 300 000 full-text articles of PubMed Central (HumanNet-XC) or interologs (IL) transferred from nine other species (HumanNet-XI). Co-citation made a significant contribution to the mapping of functional associations for several human gene networks, including HumanNet and STRING. However, the functional network by co-citation may cause biased benchmarking performance for disease gene discovery, because benchmarking data are also based on the literature. Some users may want to exclude the influence of co-citation during disease gene predictions. Therefore, we decided to maintain a human gene network extended by co-citation data separately. HumanNet-XC contains 17 790 genes and 424 501 links. In contrast to the HumanNet-XC, which contains only human-derived functional networks, HumanNet-XI includes interologs derived from five laboratory model organisms (*Caenorhabditis elegans, Drosophila melanogaster, Danio rerio, Mus musculus* and *Saccharomyces cerevisiae*) and four additional vertebrates: *Canis lupus familiaris* (dog), *Bos taurus* (cattle), *Rattus norvegicus* (Rat) and *Gallus gallus* (chicken). HumanNet-XI contains 17 303 genes and 418 525 links.

The fourth level network, HumanNet-XN, is a fully extended functional gene network by both co-citation and interologs. Interologs derived from non-human species provided 101 036 more links to HumanNet-XC, yet its genome coverage only increased from 94.6 to 95.3%. The most comprehensive network, HumanNet-XN, contains 17 929 genes and 525 537 links.

New types of data used for HumanNet v2

We incorporated functional associations inferred from two new types of data to the updated version of HumanNet. We inferred functional associations from co-annotations by pathway database. If a gene is involved in many different pathways, it may not belong to a specific pathway. Similarly, co-annotation involving such genes would be only weak indication of functional coupling. Thus, we measured the significance of functional association for given co-annotations by Fisher's exact test, giving more weight on gene pairs that share larger proportion of annotated pathways for each gene. We used pathway annotations by KEGG (22), BioCarta (23) and Recactome (24) databases. Network inference from pathway databases resulted in 125 550 links among 7512 human genes.
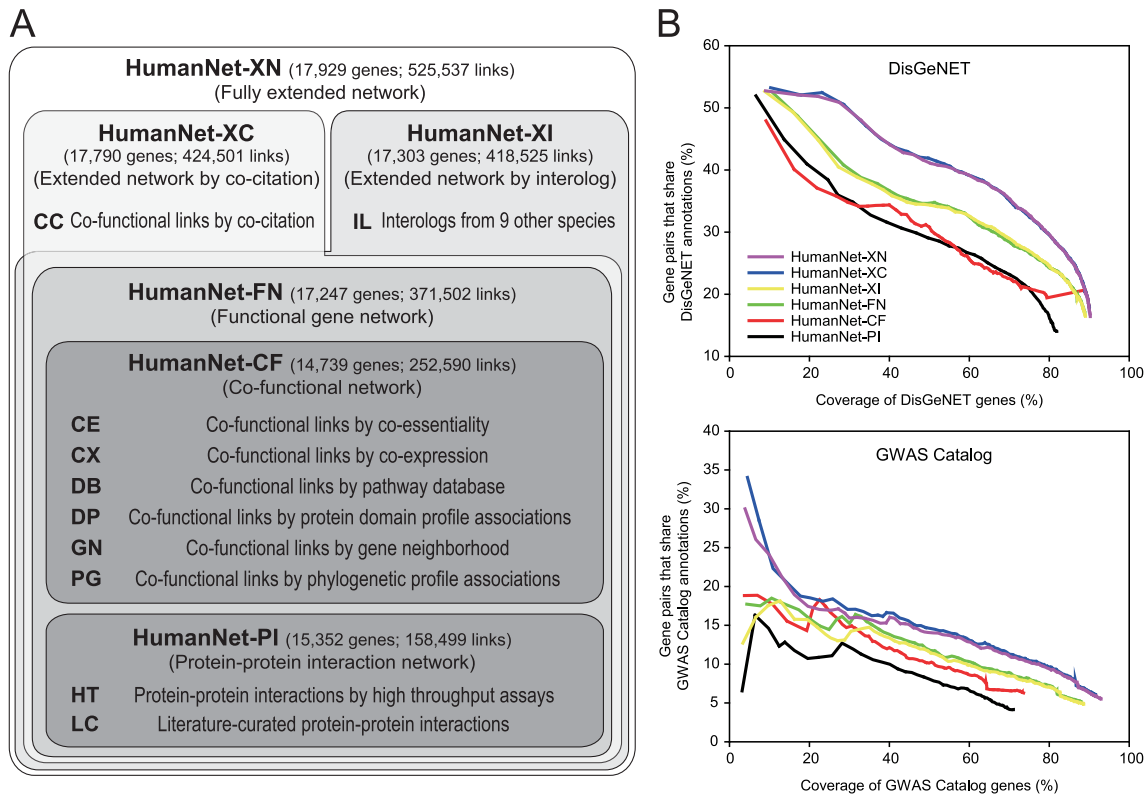
Figure 1. (**A**) Overview of the four level hierarchy of human gene networks in the HumanNet database. (**B**) Assessment of the six human gene networks at different levels of the hierarchy, based on measuring the precision of identifying gene pairs linked to the same human diseases (defined by DisGeNET or GWAS catalog with timestamp filtration) as a function of the coverage of the database genes.

Another new type of data used for updating HumanNet was co-essentiality. Recently, several large-scale essential gene screens were conducted across hundreds of cancer cell lines using the shRNA and CRISPR-Cas9 systems. Functionally associated human genes tend to have correlations of essentiality profiles across many cancer cell lines (17). We obtained the functional links inferred from associations between essentiality profiles based on over 100 genome-scale pooled-library shRNA screens and over 400 CRISPR-Cas9 screens from cancer cell lines, which are downloadable from the PICKLES database (25). Network inference from co-essentiality resulted in 71 243 links among 4052 human genes.

Data source extensions

To improve HumanNet, we also extended the sources of each data type (summarized in Supplemental Table 1). The co-citation network of HumanNet v2 is based on ∼300 000 full-text articles from PubMed Central, whereas ∼750 000 Medline abstracts were used for the co-citation network of the previous version of HumanNet. Sources of PPI data were also substantially extended. The number of database and high-throughput assay sets (Supplemental Table 2) used for human-derived PPI networks increased from 5 to 14 and 3 to 12, respectively. As a result, the number of non-redundant PPIs of HumanNet v2 is 158 499 (connecting 15 352 genes), whereas the PPI network of HumanNet v1 has 60 287 links among 9428 genes. Given that PPIs gener-

ally provide high-quality functional associations, this substantially expanded PPI network will significantly improve the generation of functional hypotheses. To update the co-expression networks, we used 125 microarray-based and 33 RNA-seq-based gene expression omnibus (GEO) (26) series (GSEs) (16 220 samples in total) (Supplemental Table 3), whereas only 21 microarray-based GSEs (1603 samples in total) were used in the previous version. Thus, the amount of expression profile data for co-expression analysis has been increased by more than 10-fold. HumanNet includes networks based on genomic context associations (GN and PG). We utilized 1748 prokaryotic (1626 bacterial and 122 archaeal) genomes and 996 metagenomes (754 from human and 242 from ocean) (27,28) to analyze the genomic context associations for HumanNet v2, whereas only 432 prokaryotic (393 bacterial and 31 archaeal) genomes were used for HumanNet v1.

Network inference algorithm enhancement

Since the release of the first version of HumanNet, we have significantly improved the network inference algorithms for each data type. We found that associations between the phylogenetic profiles of proteins showed a higher correlation with functional association within each domain of life: Archaea, Bacteria, and Eukaryota (29). Thus, for HumanNet v2, we measured the associations between phylogenetic profiles that comprise reference genomes from each domain of

life, then integrated the networks based on domain-specific profiles into a single network (PG).

For the previous version of HumanNet, we inferred functional associations by gene neighborhood using only probability-based measures (30). We later found that probability-based and distance-based measures (31) of gene neighborhood are complementary and that their integration could significantly improve network quality (20). Thus, we generated two functional networks using probability- and distance-based measures of gene neighborhood. We also found that distance-based gene neighborhoods across metagenomes correlated with functional associations (32). We could infer two functional networks by gene neighborhood analysis using 754 human microbiomes (27) and 242 ocean metagenomes (28). The final gene neighborhood network (GN) was constructed by integrating the four networks.

The human gene network based on protein domain profiles for HumanNet v2 was improved by using a weighted mutual information (WMI) score that measured the mutual information (MI) between domain profiles of proteins by giving a higher weight to rarer protein domains (19).

Systematic network evaluation for disease gene discovery

Recently, a systematic network evaluation for their ability to retrieve disease gene sets was conducted for 21 human gene networks, including the previous version of HumanNet (33). The study reported that CPDB (12), GeneMania (13), GIANT (7) and STRING (11) had the best performance in terms of retrieval of literature-curated disease gene sets by DisGeNET (34) and sets of disease candidate genes mapped by $P < 5e-08$ from the GWAS catalog (35). To confirm these results and to evaluate the new human gene networks of HumanNet v2, we evaluated the four best performed gene networks reported by the aforementioned study, another large-scale human functional network, FunCoup (14), and HumanNet v2 for disease gene predictions. Importantly, we used 'time-stamped benchmarking' strategy (36) to avoid biased evaluation by co-citation links of HumanNet and STRING. Co-citation links of HumanNet v2 were captured from papers published until 2015. Thus, we used disease-associated genes identified via GWAS published only after 2016 for each trait of the GWAS catalog. With this timestamp filtration, we could obtain 231 traits that contain more than 10 genes mapped by $P < 5e-08$ from the GWAS catalog. Since the latest version of STRING was published in 2016, we expected that the same gene sets could be used for unbiased evaluation of STRING.

We first assessed network accuracy for identifying two genes involved in the same human diseases. We found that two first level networks, HumanNet-PI and HumanNet-CF, to have the worse accuracy than the integrated functional network, HumanNet-FN in terms of connecting gene pairs linked to the same diseases annotated by DisGeNET or GWAS catalog with timestamp filtration as a function of the coverage of the database genes (Figure 1B). This result is consistent with the observation that all of the best performing human gene networks reported by the aforementioned study were functional networks rather than PPI networks (33). We found HumanNet-XC to have the best

performance in identifying gene pairs for the same diseases. Notably, incorporating interologs into HumanNet-FN and HumanNet-XC did not notably improve network precision compared with HumanNet-XI and HumanNet-XN, respectively. To evaluate contribution of each evidence to the integrated gene network, accuracy and genome coverage of networks by each data type were also assessed based on the same disease annotations (Supplemental Figure 1).

Next, we compared the best performing HumanNet-XC with the previous HumanNet (v1) as well as five other human gene networks, and found that HumanNet-XC outperformed all the other human gene networks (Figure 2A). In addition, we observed that HumanNet-PI has overall higher accuracy than another scored human PPI network, InWeb (Figure 2B). These results indicate that HumanNet v2 might provide the most appropriate networks for disease research by utilizing protein physical interactions as well as functional associations.

Next, we evaluated the networks for their ability to retrieve disease gene sets. The network performance for disease gene recovery correlates with the efficiency of disease gene discovery by network-based gene prioritization. Network-based gene prioritization for diseases can use two alternative strategies: direct neighborhood and network diffusion (37). Direct neighborhood methods prioritize genes using the disease information of their directly connected network neighbors only (38,39). In contrast, network diffusion methods prioritize genes by propagating disease information throughout the entire network (40). Recently, network diffusion methods have increased in popularity, and the web server of the previous HumanNet version also employed network diffusion for disease gene prioritization. However, more recently, multiple studies have shown that direct neighborhood is generally more efficient than network diffusion in obtaining disease genes in the top predictions (41,42). Because typically only a few hundred candidates at most are considered for the follow-up functional analysis, we benchmarked the retrieval efficiency of disease genes by the area under the receiver operating characteristic curve (AUROC) until a false positive rate of 1% (FPR < 0.01). With this benchmarking analysis, we found HumanNet-XC and HumanNet-XN to have significantly better performance than all other networks by direct neighborhood with the unbiased disease gene sets (Figure 3A). We observed similar results for AUROC until FPR of 2% and 5% (Supplemental Figure 2). In consistent with the results of previous systematic evaluation, HumanNet v1 showed worse performance than STRING, GeneMania, and GIANT with the time-stamped benchmarking, indicating large influence of co-citation information on the earlier version of HumanNet (33).

It is also possible to prioritize disease genes with network diffusion techniques such as random walk with the restart model (40). For benchmarking the retrieval efficiency of disease gene sets by network diffusion, we used 'performance gain' scores based on the area under the precision recall curve (AUPRC) as described in a previous study on systematic network evaluations (33). With this benchmarking analysis, we found HumanNet-XC, HumanNet-XN, and STRING to have significantly better performance than other networks (Figure 3B).
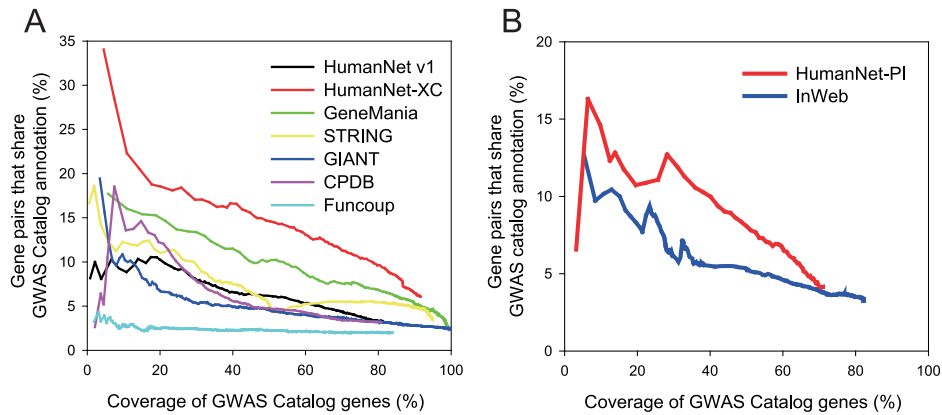
Figure 2. Assessment of human functional gene networks (**A**) and PPI networks (**B**) for genes linked to the same human diseases (defined by GWAS catalog with timestamp filtration) as a function of the coverage of the database genes.
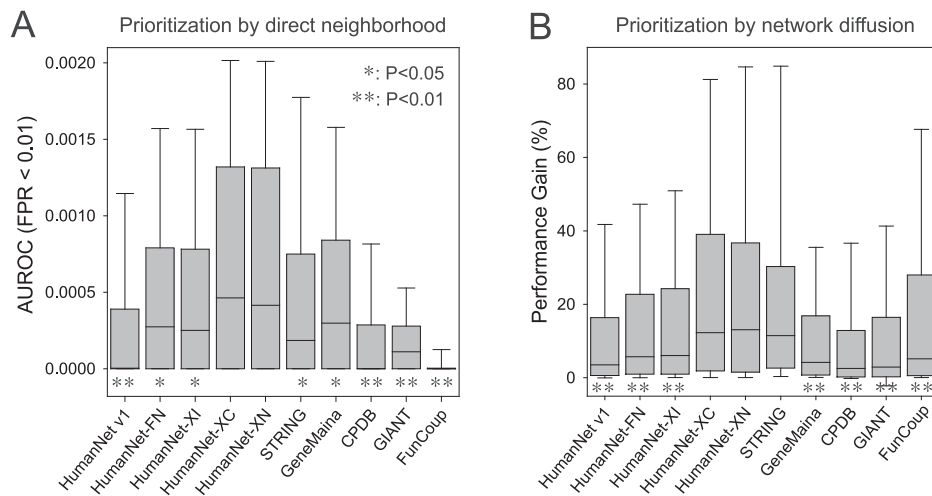

Figure 3. Assessment of predictive ability of networks for unbiased GWAS catalog disease gene sets based on the distribution of (**A**) the area under receiver operating characteristic curve (AUROC) until 1% of false positive rate (FPR < 0.01) and (**B**) performance gain scores based on area under precision recall curve (AUPRC). For each box-and-whisker plot, the boundaries of the box represent the first and third quartiles and the whiskers represent the 10th and 90th percentiles. Significance of performance difference from that of HumanNet-XC is indicated by asterisk (*: $P < 0.05$, **: $P < 0.01$, Wilcoxon rank sum test).

Notably, in all of the above benchmarking analysis, we did not observe a significant increase in performance by incorporating interologs ($P > 0.05$ for HumanNet-FN versus HumanNet-XI and for HumanNet-XC versus HumanNet-XN, Wilcoxon rank sum test). These results suggest that many of the evolutionarily conserved gene links for the same diseases are now captured directly in human-derived data. However, we cannot exclude the possibility that intrologs can improve gene prioritization for non-pathogenic cellular processes such as core metabolic pathways.

WEB INTERFACE IMPROVEMENT

Implementation of a new user interface

We implemented back-end and front-end servers for HumanNet v2 to facilitate effective interactions with users. For the back-end server implementation, we used Redis (https://redis.io), an in-memory DB which reduces the data loading time significantly compared with that

from a hard drive. We designed the back-end interface as an Application Programming Interface (API) to communicate with the front-end server and also job requests from users. We employed several open-sourced Cascading Style Sheet components and JavaScript libraries for front-end server implementation. We designed the website layout using Bootstrap4 and its components (https://getbootstrap.com). Cytoscape.js (43) and its extensions, 'cytoscape.js-cose-bilkent' (from https://doi.org/10.5281/zenodo.1098231) and 'cytoscape.js-panzoom' (from http://doi.org/10.5281/zenodo.835037) were employed to provide the graph and network visualization.

Disease-focused hypothesis generation

The HumanNet v2 web server facilitates human disease research by predicting disease genes or disease annotations. Network-based disease gene predictions are generally based on the network connections to the genes known to be involved in the disease. We dubbed these known disease genes
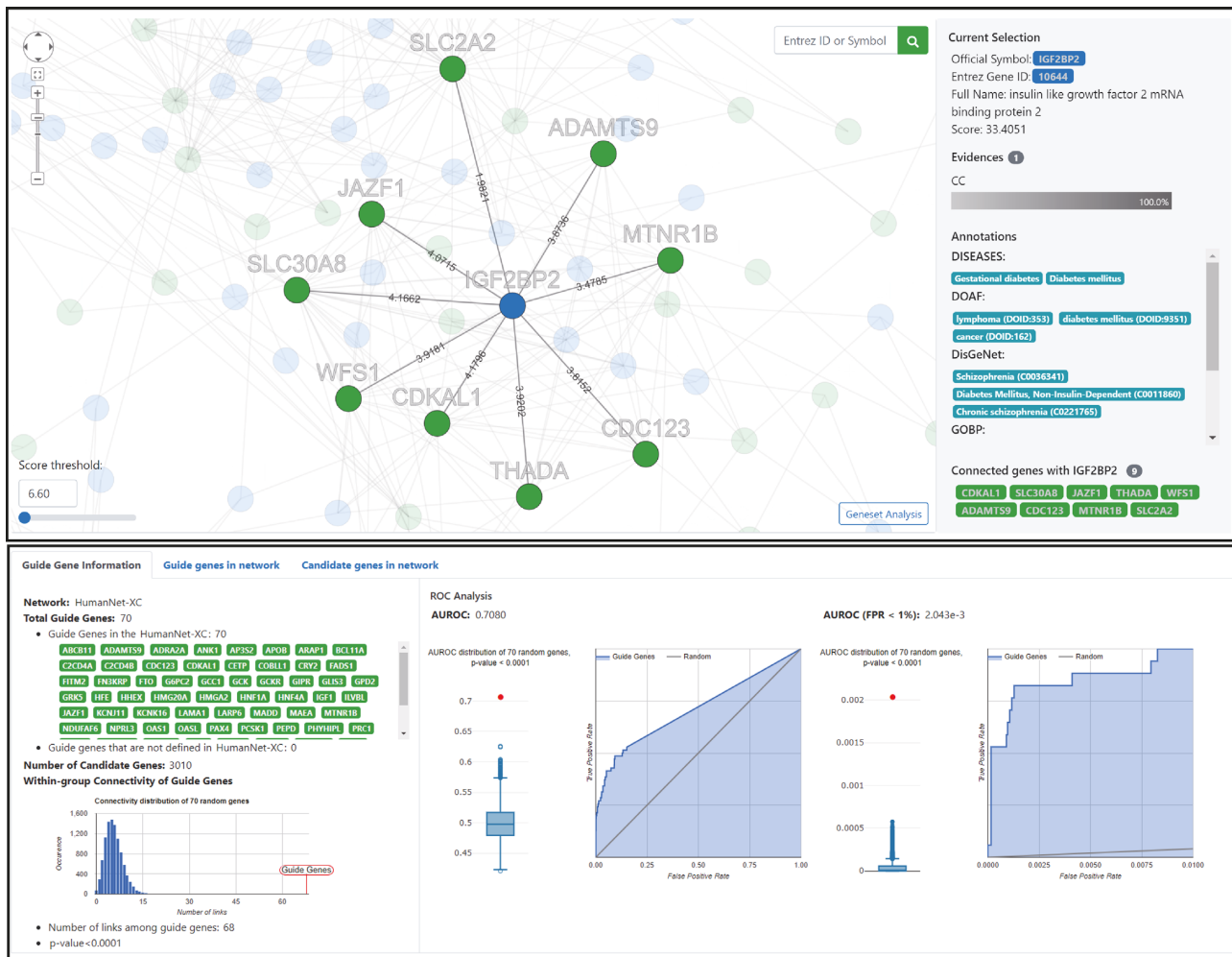
Figure 4. Screenshots of the HumanNet reports page for the network-based disease gene prediction using HumanNet-XC based on submission of 70 genes for type 2 diabetes mellitus (defined by DISEASES) as guide (query) genes. The upper panel shows the interactive network viewer, visualizing a network of guide genes (green nodes) and their top 100 direct neighbors, which can be interpreted as putative candidate genes (blue nodes). Here, the local subnetwork of the third ranked candidate, IGF2BP2 and its neighbors is highlighted. The retrieved gene IGF2BP2 is already annotated for diabetes mellitus by DISEASES, DOAF and DisGeNET, serving to validate the specific prediction result. The lower panel reports data on the guide genes, including the statistical significance of within group connectivity of guide genes, and the observed network performance for guide gene recovery reported as ROC curves.

'guide genes' because they guide the network-based predictions of new disease gene candidates. We can estimate the predictive performance of networks based on the efficiency of guide gene recovery. The HumanNet v2 server uses a direct neighborhood approach rather than network diffusion for network-based gene prioritization, because at most a few hundred candidates are considered for follow-up functional analysis and direct neighborhood generally outperforms network diffusion methods for the early retrieval of guide genes (41,42). The HumanNet v2 server uses HumanNet-XC as a default network, because it showed the best performance for disease gene recovery in our benchmarking analyses.

Using multiple guide genes for network-based predictions is desirable, because predictions based on multiple network connections are more confident due to the ensemble effect. The functional coherence of guide genes would be a meaningful indicator of their effectiveness. Therefore, the HumanNet server reports on the significance of within-group connectivity of guide genes using 10 000 random gene sets of the same size (Figure 4, lower panel). The Human-Net server also reports on ROC plots, which indicate the predictive performance of networks for a disease based on the efficiency of guide gene recovery. To evaluate statistical significance of the observed AUROC score, HumanNet v2 server generates null models using 10 000 random gene sets of the same size. Users can submit pre-defined disease gene sets from DisGeNET (34), DISEASES (44), Disease Ontology Annotation Framework (DOAF) (45), GWAS catalog (35) and Human Phenotype Ontology (HPO) (46). Users can also submit a set of genes targeted by a drug based on DiSigDB (47). Thus, predictions guided by the DiSigDB gene set are likely candidates of novel targets for the same drug. HumanNet users can also predict disease annotations of a gene based on the network-neighbors. For a query gene, the HumanNet server collects disease annotations from its

network neighbors and lists them starting from the most enriched one.

Interactive network viewer

Network-based disease gene prediction generates a network of guide genes and new candidate genes for disease. Further investigation of the disease gene network would provide functional insights which might be useful for narrowing down final candidates and for mode-of-action studies. Therefore, we designed a network viewer enabling users to conduct interactive analyses of the disease gene network. The HumanNet v2 server generates a network of guide genes and the top 100 candidate genes for the disease. Initially, the entire network appears in the viewer to give a brief idea of the disease gene network, but soon after, all the candidate genes disappear. Then, users can select different numbers of top candidate genes for a new disease network by thresholding the prediction score (Figure 4, upper panel). Users can select a particular gene of the disease network not only from the network viewer but also from the table of candidate genes. The network viewer highlights a local subnetwork of the choice of gene and its network neighbors. Users can also see additional information such as annotations of the GO biological process and diseases for the chosen gene and supporting evidence for the local network connections. The interactive thresholding for candidate gene selection allows users to consider various disease gene networks with different trade-offs between degree of confidence and coverage. Disease-association for the selected group of candidate genes can be summarized by gene-set analysis (GSA). Users can select the top N candidate genes and run GSA with not only GO biological processes but also annotated disease genes from DisGeNET (34), DISEASES (44), DOAF (45) and HOP (46).

CONCLUSION

In this report, we present an updated HumanNet by incorporating new types of data, extending data sources and improving network inference algorithms. The new Human-Net was designed to have an inclusive, four level hierarchy of human gene networks. Based on our benchmarking results for their performance of disease gene recovery, we conclude that HumanNet serves as one of the better human gene networks for prioritizing disease-linked genes and reconstructing disease-relevant gene modules. We recommend HumanNet-XC for most network-based disease research, but other networks will be useful for other purposes. For example, HumanNet-PI is recommended for the mode-of-action studies of disease mutation, HumanNet-FN for more conservative predictions of disease genes and HumanNet-XN for studies requiring the most comprehensive networks. Due to the continuous growth of omics data repositories and the advent of new types of functional genomics data such as single cell transcriptome profiles, we might be able to keep improving HumanNet in the future. With a highly interactive web server for generating hypotheses, we expect HumanNet to be a highly useful *in silico* resource for the study of human diseases.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Barabasi,A.L., Gulbahce,N. and Loscalzo,J. (2011) Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.*, **12**, 56–68.
2. Jiang,P., Wang,H., Li,W., Zang,C., Li,B., Wong,Y.J., Meyer,C., Liu,J.S., Aster,J.C. and Liu,X.S. (2015) Network analysis of gene essentiality in functional genomics experiments. *Genome Biol.*, **16**, 239.
3. Nitsch,D., Tranchevent,L.C., Thienpont,B., Thorrez,L., Van Esch,H., Devriendt,K. and Moreau,Y. (2009) Network analysis of differential expression for the identification of disease-causing genes. *PLoS One*, **4**, e5526.
4. Gwinner,F., Boulday,G., Vandiedonck,C., Arnould,M., Cardoso,C., Nikolayeva,I., Guitart-Pla,O., Denis,C.V., Christophe,O.D., Beghain,J. *et al.* (2017) Network-based analysis of omics data: the LEAN method. *Bioinformatics*, **33**, 701–709.
5. Cho,A., Shim,J.E., Kim,E., Supek,F., Lehner,B. and Lee,I. (2016) MUFFINN: cancer gene discovery via network analysis of somatic mutation data. *Genome Biol.*, **17**, 129.
6. Horn,H., Lawrence,M.S., Chouinard,C.R., Shrestha,Y., Hu,J.X., Worstell,E., Shea,E., Ilic,N., Kim,E., Kamburov,A. *et al.* (2018) NetSig: network-based discovery from cancer genomes. *Nat. Methods*, **15**, 61–66.
7. 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.
8. Shim,J.E., Bang,C., Yang,S., Lee,T., Hwang,S., Kim,C.Y., Singh-Blom,U.M., Marcotte,E.M. and Lee,I. (2017) GWAB: a web server for the network-based boosting of human genome-wide association data. *Nucleic Acids Res.*, **45**, W154–W161.
9. Lee,I., Blom,U.M., Wang,P.I., Shim,J.E. and Marcotte,E.M. (2011) Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res.*, **21**, 1109–1121.
10. Yu,H., Luscombe,N.M., Lu,H.X., Zhu,X., Xia,Y., Han,J.D., Bertin,N., Chung,S., Vidal,M. and Gerstein,M. (2004) Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res.*, **14**, 1107–1118.
11. Szklarczyk,D., Morris,J.H., Cook,H., Kuhn,M., Wyder,S., Simonovic,M., Santos,A., Doncheva,N.T., Roth,A., Bork,P. *et al.* (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.*, **45**, D362–D368.
12. Herwig,R., Hardt,C., Lienhard,M. and Kamburov,A. (2016) Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nat. Protoc.*, **11**, 1889–1907.
13. Franz,M., Rodriguez,H., Lopes,C., Zuberi,K., Montojo,J., Bader,G.D. and Morris,Q. (2018) GeneMANIA update 2018. *Nucleic Acids Res.*, **46**, W60–W64.
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. Sahni,N., Yi,S., Taipale,M., Fuxman Bass,J.I., Coulombe-Huntington,J., Yang,F., Peng,J., Weile,J., Karras,G.I., Wang,Y. *et al.* (2015) Widespread macromolecular interaction perturbations in human genetic disorders. *Cell*, **161**, 647–660.

16. Shim,J.E., Lee,T. and Lee,I. (2017) From sequencing data to gene functions: co-functional network approaches. *Anim. Cells Syst.*, **21**, 77–83.

17. Wang,T., Yu,H., Hughes,N.W., Liu,B., Kendirli,A., Klein,K., Chen,W.W., Lander,E.S. and Sabatini,D.M. (2017) Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic ras. *Cell*, **168**, 890–903.

18. Yang,S., Kim,C.Y., Hwang,S., Kim,E., Kim,H., Shim,H. and Lee,I. (2017) COEXPEDIA: exploring biomedical hypotheses via co-expressions associated with medical subject headings (MeSH). *Nucleic Acids Res.*, **45**, D389–D396.

19. Shim,J.E. and Lee,I. (2016) Weighted mutual information analysis substantially improves domain-based functional network models. *Bioinformatics*, **32**, 2824–2830.

20. Shin,J., Lee,T., Kim,H. and Lee,I. (2014) Complementarity between distance- and probability-based methods of gene neighbourhood identification for pathway reconstruction. *Mol. Biosyst.*, **10**, 24–29.

21. Shin,J. and Lee,I. (2017) Construction of functional gene networks using phylogenetic profiles. *Methods Mol. Biol.*, **1526**, 87–98.

22. Kanehisa,M., Furumichi,M., Tanabe,M., Sato,Y. and Morishima,K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.

23. Nishimura,D. (2001) BioCarta. *Biotech Softw. Int. Rep.*, **2**, 117–120.

24. Fabregat,A., Jupe,S., Matthews,L., Sidiropoulos,K., Gillespie,M., Garapati,P., Haw,R., Jassal,B., Korninger,F., May,B. *et al.* (2018) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **46**, D649–D655.

25. Lenoir,W.F., Lim,T.L. and Hart,T. (2018) PICKLES: the database of pooled in-vitro CRISPR knockout library essentiality screens. *Nucleic Acids Res.*, **46**, D776–D780.

26. 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.

27. Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature*, **486**, 207–214.

28. Sunagawa,S., Coelho,L.P., Chaffron,S., Kultima,J.R., Labadie,K., Salazar,G., Djahanschiri,B., Zeller,G., Mende,D.R., Alberti,A. *et al.* (2015) Ocean plankton. Structure and function of the global ocean microbiome. *Science*, **348**, 1261359.

29. Shin,J. and Lee,I. (2015) Co-Inheritance analysis within the domains of life substantially improves network inference by phylogenetic profiling. *PLoS One*, **10**, e0139006.

30. Bowers,P.M., Pellegrini,M., Thompson,M.J., Fierro,J., Yeates,T.O. and Eisenberg,D. (2004) Prolinks: a database of protein functional linkages derived from coevolution. *Genome Biol.*, **5**, R35.

31. Korbel,J.O., Jensen,L.J., von Mering,C. and Bork,P. (2004) Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs. *Nat. Biotechnol.*, **22**, 911–917.

32. Kim,C.Y. and Lee,I. (2017) Functional gene networks based on the gene neighborhood in metagenomes. *Anim. Cells Syst.*, **21**, 301–306.

33. Huang,J.K., Carlin,D.E., Yu,M.K., Zhang,W., Kreisberg,J.F., Tamayo,P. and Ideker,T. (2018) Systematic evaluation of molecular networks for discovery of disease genes. *Cell Syst.*, **6**, 484–495.

34. Pinero,J., Bravo,A., Queralt-Rosinach,N., Gutierrez-Sacristan,A., Deu-Pons,J., Centeno,E., Garcia-Garcia,J., Sanz,F. and Furlong,L.I. (2017) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.*, **45**, D833–D839.

35. MacArthur,J., Bowler,E., Cerezo,M., Gil,L., Hall,P., Hastings,E., Junkins,H., McMahon,A., Milano,A., Morales,J. *et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.*, **45**, D896–D901.

36. Bornigen,D., Tranchevent,L.C., Bonachela-Capdevila,F., Devriendt,K., De Moor,B., De Causmaecker,P. and Moreau,Y. (2012) An unbiased evaluation of gene prioritization tools. *Bioinformatics*, **28**, 3081–3088.

37. Wang,P.I. and Marcotte,E.M. (2010) It's the machine that matters: Predicting gene function and phenotype from protein networks. *J. Proteomics*, **73**, 2277–2289.

38. Guala,D., Sjolund,E. and Sonnhammer,E.L. (2014) MaxLink: network-based prioritization of genes tightly linked to a disease seed set. *Bioinformatics*, **30**, 2689–2690.

39. Lee,I., Li,Z. and Marcotte,E.M. (2007) An improved, bias-reduced probabilistic functional gene network of baker's yeast, Saccharomyces cerevisiae. *PLoS One*, **2**, e988.

40. Kohler,S., Bauer,S., Horn,D. and Robinson,P.N. (2008) Walking the interactome for prioritization of candidate disease genes. *Am. J. Hum. Genet.*, **82**, 949–958.

41. Shim,J.E., Hwang,S. and Lee,I. (2015) Pathway-dependent effectiveness of network algorithms for gene prioritization. *PLoS One*, **10**, e0130589.

42. Guala,D. and Sonnhammer,E.L.L. (2017) A large-scale benchmark of gene prioritization methods. *Sci. Rep.*, **7**, 46598.

43. Franz,M., Lopes,C.T., Huck,G., Dong,Y., Sumer,O. and Bader,G.D. (2016) Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*, **32**, 309–311.

44. Pletscher-Frankild,S., Palleja,A., Tsafou,K., Binder,J.X. and Jensen,L.J. (2015) DISEASES: text mining and data integration of disease-gene associations. *Methods*, **74**, 83–89.

45. Xu,W., Wang,H., Cheng,W., Fu,D., Xia,T., Kibbe,W.A. and Lin,S.M. (2012) A framework for annotating human genome in disease context. *PLoS One*, **7**, e49686.

46. Kohler,S., Vasilevsky,N.A., Engelstad,M., Foster,E., McMurry,J., Ayme,S., Baynam,G., Bello,S.M., Boerkoel,C.F., Boycott,K.M. *et al.* (2017) The human phenotype ontology in 2017. *Nucleic Acids Res.*, **45**, D865–D876.

47. Yoo,M., Shin,J., Kim,J., Ryall,K.A., Lee,K., Lee,S., Jeon,M., Kang,J. and Tan,A.C. (2015) DSigDB: drug signatures database for gene set analysis. *Bioinformatics*, **31**, 3069–3071.