Published online 5 October 2016

Nucleic Acids Research, 2016, Vol. 44, No. 20 ***9611–9623***
doi: 10.1093/nar/gkw897

Function-driven discovery of disease genes in zebrafish using an integrated genomics big data resource

Hongseok Shim[1,†], Ji Hyun Kim[2,†], Chan Yeong Kim[1,†], Sohyun Hwang[1], Hyojin Kim[1], Sunmo Yang[1], Ji Eun Lee[2,3,*] and Insuk Lee[1,*]

[1]Department of Biotechnology, College of Life Science and Biotechnology, Yonsei University, Seoul 03722, Korea, [2]Department of Health Sciences & Technology, SAIHST, Sungkyunkwan University, Seoul 06351, Korea and [3]Samsung Genome Institute, Samsung Medical Center, Seoul 06351, Korea

Received July 12, 2016; Revised September 23, 2016; Accepted September 29, 2016

ABSTRACT

Whole exome sequencing (WES) accelerates disease gene discovery using rare genetic variants, but further statistical and functional evidence is required to avoid false-discovery. To complement *variant-driven* disease gene discovery, here we present *function-driven* disease gene discovery in zebrafish (*Danio rerio*), a promising human disease model owing to its high anatomical and genomic similarity to humans. To facilitate zebrafish-based *function-driven* disease gene discovery, we developed a genome-scale co-functional network of zebrafish genes, DanioNet (www.inetbio.org/danionet), which was constructed by Bayesian integration of genomics big data. Rigorous statistical assessment confirmed the high prediction capacity of DanioNet for a wide variety of human diseases. To demonstrate the feasibility of the *function-driven* disease gene discovery using DanioNet, we predicted genes for ciliopathies and performed experimental validation for eight candidate genes. We also validated the existence of heterozygous rare variants in the candidate genes of individuals with ciliopathies yet not in controls derived from the UK10K consortium, suggesting that these variants are potentially involved in enhancing the risk of ciliopathies. These results showed that an integrated genomics big data for a model animal of diseases can expand our opportunity for harnessing WES data in disease gene discovery.

INTRODUCTION

Whole exome sequencing (WES) provides a major route for the identification of genes implicated in Mendelian diseases as well as complex diseases based on rare protein-coding variants in phenotypically extreme individuals (1). This *variant-driven* approach has helped identify numerous candidate genes involved in inherited disorders. However, false assignment of disease-associated variants has recently become a substantial issue. Multiple large-scale studies have shown that a significant portion of reported disease-associated variants are either common polymorphisms (2) or are prevalent in healthy individuals (3,4). Therefore, the community proposed a guideline for assessing the implication of sequence variants in diseases (5), which includes experimental validation of the predicted damaging impact of the candidate variants using functional assays in well-established animal models. In fact, *loss-of-function* studies for the genes harboring the predicted causal variants in animal models have often revealed no disease-relevant phenotype, indicating that variant-level evidence from individuals with diseases does not warrant causality in the pathogenesis.

In this study, we propose a *function-driven* approach for disease gene discovery, which takes the opposite direction of the *variant-driven* approach. In the *variant-driven* approach, the identification of implicated variants is generally followed by experimental validation using *loss-of-function* studies in animal models. Conversely, in the *function-driven* approach, the observation of disease-related mutant phenotypes in model animals leads to the identification of novel disease gene candidates, which is then followed by an examination of the 'qualifying variants' within the candidate genes in the individuals with the disease (Figure 1A). Securing functional evidence may reduce the burden of rigorousness in assessing genetic penetrance of the candidate

*To whom correspondence should be addressed. Tel: +82 2 2123 5559; Fax: +82 2 362 7265; Email: insuklee@yonsei.ac.kr
Correspondence may also be addressed to Ji Eun Lee. Tel: +82 2 3410 6129; Fax: +82 2 3410 0543; Email: jieun.lee@skku.edu
†These authors equally contributed to this work as the first authors.

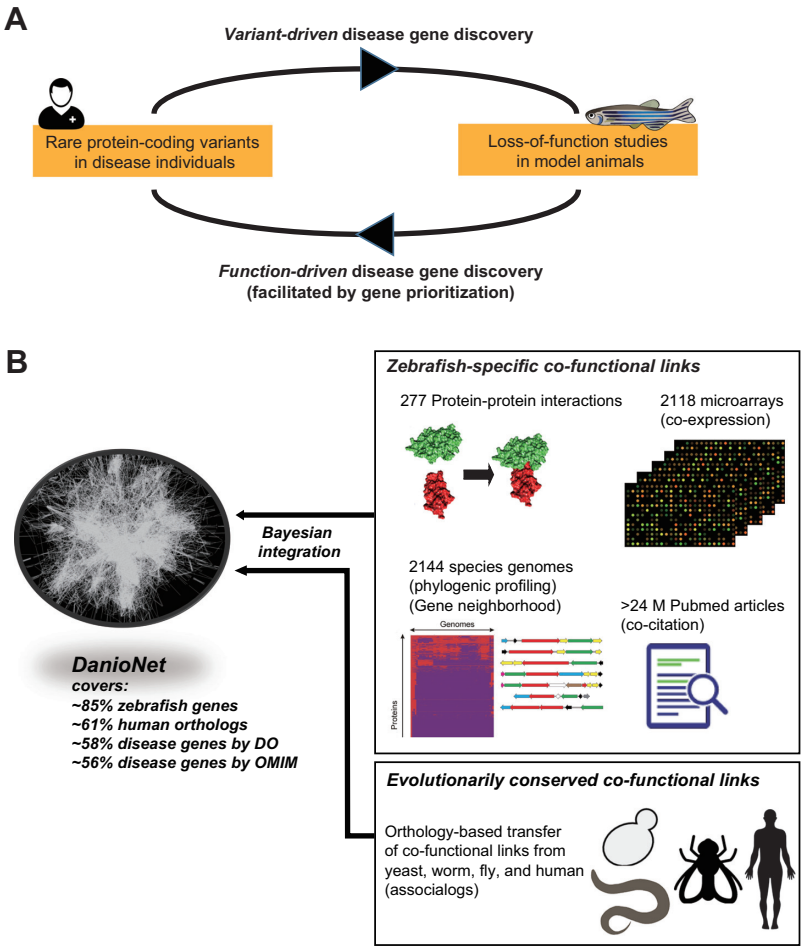


Figure 1. Function-driven disease gene discovery using DanioNet. (**A**) Relationship between *variant-driven* and *function-driven* disease gene discovery. (**B**) Overview of the construction of DanioNet. The functional associations between genes were inferred from 18 distinct data sets of various species—*Danio rerio*, *Caenorhabditis elegans, Drosophila melanogaster, Homo sapiens* and *Saccharomyces cerevisiae*—using various algorithms such as co-citation, co-expression, phylogenetic profiles, gene neighborhood and associalogs. The reliable associations that passed the benchmarking test were then integrated into the final network, DanioNet.

variants. In addition, animal models containing genetic defects with disease-relevant phenotypes enable the study of the molecular mechanisms underlying the progression of the diseases. Therefore, the *function-driven* approach can complement the *variant-driven* approach in the identification and characterization of novel disease genes.

Zebrafish, *Danio rerio*, is one of the most qualified laboratory animals for the *function-driven* discovery of human disease genes for several reasons. First, zebrafish has been a long-standing model of vertebrate developmental biology in the virtue of high fecundity, rapid development, easy laboratory maintenance, efficient gene perturbation using morpholino and allowance of non-invasive *in vivo* examination of organ development through the translucent embryo and body (6). In addition, the organ development in zebrafish is highly similar to that in humans (7,8) and the zebrafish genome contains more than 70% of human genes (9). Indeed, zebrafish has already become a common laboratory animal for *in vivo* validation of candidate disease genes identified from genome-wide association studies (GWAS) as well as WES (10–14). Furthermore, zebrafish embryos containing genetic defects of human diseases enable high throughput drug screening and toxicity tests *in vivo* (15) and facilitate the investigation of disease mechanisms through various molecular and genomics techniques suitable for the organism (7).

Function-driven disease gene discovery starts with the identification of animal orthologs whose genetic defects result in the disease-related phenotype. Searching for the animal genes with disease implications may be conducted through either forward or reverse genetics (16). The Ethyl methanesulfonate and the Sleeping Beauty transposon system have been used for mutagenesis in many forward genetics projects in zebrafish. However, although random mutagenesis approaches have identified many disease genes in zebrafish, the whole screening work suffers from laborious processes and low sensitivity for mutations of moderate or weak phenotypic impact. Reverse genetics can provide a feasible alternative if an effective gene prioritization method is available. Recently, many disease genes have been identified through network-based gene prioritization (17,18), in which the principle of guilt-by-association is ap-

plied to predict new genes for a disease based on their closeness to the genes implicated in the same disease in functional networks. The co-functional relationships between genes can be inferred from a large body of genomics data derived from diverse high-throughput experiments, which entails heterogeneity and variation in data quality. To manage interpretation and integration of the genomics big data, the Bayesian integration approach, which has proven suitable for inferring functional relationships between genes from heterogeneous genomics data, has been used (19,20). In this approach, each genomics data set is evaluated by gold-standard gene pairs to calculate the likelihood of co-functional relationships, and multiple likelihood scores from various data sets are then integrated with differential weights.

To facilitate the identification of candidate disease genes in zebrafish by taking advantage of the massive amount of genomics big data, we developed a genome-scale co-functional network of zebrafish genes, DanioNet, which was constructed by Bayesian integration of 18 distinct types of genomics data. The constructed network covers ∼85% of the coding genome of zebrafish and ∼61% of human orthologs, thus enabling extensive search for genes for vertebrate developmental disorders as well as human diseases. Importantly, we proved that DanioNet is highly predictive for a wide spectrum of human diseases through rigorous statistical assessment. To demonstrate the feasibility of DanioNet in the identification of human disease genes, we predicted genes for ciliopathies, a group of disorders associated with genetic defects causing abnormal formation or function of the primary cilia, and performed experimental validation for eight candidate genes. We also validated the existence of rare coding variants in the candidate genes of individuals with ciliopathies. Candidate genes for the *function-driven* discovery of disease genes can be easily prioritized using a companion web server (www.inetbio.org/danionet).

MATERIALS AND METHODS

Sequences and functional annotation data for *D. rerio*

Sequences for 18 971 coding genes of *D. rerio* with reliable evidences from Ensembl (Zv9, 73 database version) (9) were used for network construction and analysis. Annotations for pathways and biological processes of the coding genes were downloaded from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (21) and Gene Ontology biological process (GO-BP) annotations (22). Anatomical annotation data for the same genes were downloaded from ZFIN (23).

Construction of DanioNet

Co-functional links were inferred from 18 distinct data types using supervised learning methods and were then integrated into DanioNet based on a Bayesian statistical framework. A detailed description of the network construction is provided in the Supplementary Data.

Zebrafish housing and manipulations

Adult zebrafish were maintained at 28.5°C and pH 7.0–7.9 with a cycle of light (13 h) and dark (11 h) per day in the automatic system (Constructed by Genomic-Design, South Korea). Zebrafish embryos were collected by natural breeding and incubated in E3 medium (297.7 mM NaCl, 10.7 mM KCl, 26.1 mM $CaCl_2$ and 24.1 mM $MgCl_2$) containing 1% methylene blue (Sigma-Aldrich, St. Louis, MO, USA) at 28.5°C. To inhibit the formation of melanin, we raised the zebrafish larvae (after 24 h post-fertilization (hpf)) in E3 medium containing 0.2 mM N-phenylthiourea (PTU; Sigma- Aldrich Chemistry, cat. # P7629).

Microinjection of morpholino oligos (MOs)

The translation and splicing-blocking MOs were designed and synthesized by Gene Tools (Philomath, OR, USA). Each MO was diluted in distilled water at a concentration of 2–4 μg/μl and then injected into the yolk of zebrafish embryos at one- to four-cell stages using a gas-used microinjection system (PV83 Pneumatic PicoPump, SYS-PV830, World Precision Instruments, USA). The morphology of the MO-injected embryos was observed under a dissecting microscope (SMZ1270, Nikon, Tokyo, Japan), and the images were captured using a mono-camera (DS-Qi2, Nikon, Tokyo, Japan) and analyzed using the NIS-Elements software (Nikon, Tokyo, Japan).

Whole-mount immunohistochemistry of the zebrafish larvae

The MO-injected embryos were raised at 28.5°C until 3 days post-fertilization (dpf) and fixed in Dent's fixative (80% MeOH, 20% DMSO) at room temperature (RT) for 3 h. The fixed zebrafish larvae were then immunostained with an anti-GT335 antibody (AdipoGen, cat.# A20631002; 1:400 dilution) as a primary antibody at 4°C overnight and with a goat anti-mouse Alexa Fluor® 488 or 594-conjugated antibody as a secondary antibody (Invitrogen, cat.# A11001 or A11005, 1:1000) at RT for 2 h. Both the antibodies were diluted in half-concentration blocking solution (10% normal goat serum, 0.5% Tween 20 in PBS). Then the zebrafish larvae were stained with 12 μM DAPI (Molecular probes, cat. #D3571) at room temperature for 15 min in PBST (0.5% Tween 20 in PBS). The fluorescent images were obtained using a LSM700 confocal microscope (Carl Zeiss, Germany).

Rare variant identification from the UK10K individuals

We obtained VCF files for 3781 healthy individuals from the TwinsUK cohort (EGAD00001000741) and the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort (EGAD00001000740) and 122 individuals with ciliopathies (EGAD00001000414) from the European Genome-phenome Archive (EGA) (24) under the UK10K data access agreement (25). We also obtained VCF files for human common variants from dbSNP (build146 GRCh37p13) (26). All the VCF files were regularized using VCFLib (https://github.com/vcflib/vcflib). To identify the rare variants, we annotated variants of the UK10K individuals with dbSNP common variants using GATK v3.5 VariantAnnotator (27). All the non-common variants are

considered as rare variants. We obtained the information of amino acid change and the functional impact of each variant using Variant Effect Predictor (VEP) (28).

Finding the ciliopathy network module with intermediate nodes

To find a well-connected subnetwork enriched for ciliopathy genes in DanioNet, we permitted the inclusion of intermediate nodes that passed a significance threshold as previously described (29). An intermediate node was a gene connected with at least two other genes, and the significance of an intermediate node was calculated using the z-score from a binomial proportion test as follows:

$$z = \frac{\left(\frac{a}{c} - \frac{b}{d}\right)}{\sqrt{\frac{\frac{b}{d}\cdot\left(1-\frac{b}{d}\right)}{c}}}$$

where a represents the links from the intermediate node to the input ciliopathy genes, b represents the total degree of the intermediate node in DanioNet, c represents the total functional links of the subnetwork to the ciliopathy genes and all the possible intermediates and d represents the total number of links in DanioNet. We used a z-score of 20 to extract a ciliopathy network.

RESULTS AND DISCUSSION

Co-functional network of zebrafish genes as an integrated genomics big data resource

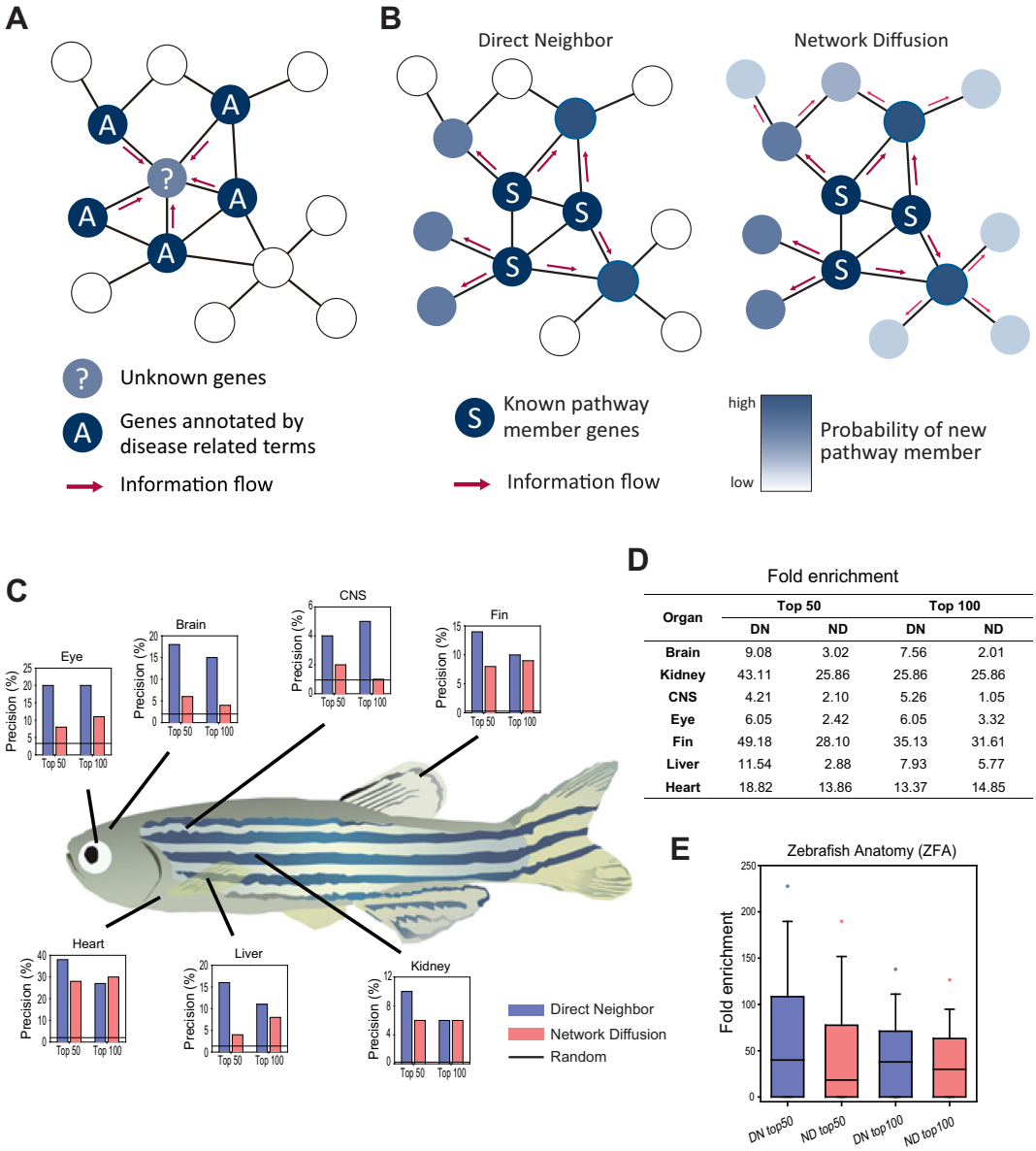
The construction of a functional network for zebrafish (*D. rerio*) is summarized in Figure 1B and has been described in detail in the Supplementary Data. A total of 18 971 coding genes were selected for the network modeling with reliable annotation using Ensembl (Zv9, 73 database version) (9). The networks inferred from 18 distinct types of genomics data using a Bayesian statistical framework (19,20) are summarized in Supplementary Table S1. Zebrafish-specific co-functional links were inferred from five distinct types of *D. rerio* data: co-citation of two zebrafish genes in Medline articles (DR-CC) (30), co-expression across microarray experiments (DR-CX), protein–protein interactions from literature curation (DR-LC), similarity of phylogenetic profiles (DR-PG) (31) and genomic neighborhoods of bacterial orthologs (DR-GN) (32). Particularly, we constructed a zebrafish co-expression network by integrating the links inferred from 21 microarray sample series of the Gene Expression Omnibus database (33), which comprised 2118 microarray samples (Supplementary Table S2). Evolutionarily conserved co-functional links were also inferred by orthology-based transfer of network links, associalogs (34), from four other species: co-citation and co-expression networks of *Caenorhabditis elegans* (CE-CC and CE-CX, respectively); co-expression network, protein–protein interactions derived from high-throughput assays and those from literature curation of *Drosophila melanogaster* (DM-CX, DM-HT and DM-LC, respectively); co-expression network, protein–protein interactions derived from high-throughput assays and those from literature curation of humans (HS-CX, HS-HT and HS-LC, respectively); and co-citation network, co-expression network, genetic interactions, protein–protein interactions derived from high-throughput assays and those from literature curation of *Saccharomyces cerevisiae* genes (SC-CC, SC-CX, SC-GT, SC-HT and SC-LC, respectively).

For Bayesian integration along with supervised learning of the co-functional links, we used gold-standard *D. rerio* gene pairs that share annotations with the KEGG pathway database (21) or GO-BP (22) based on experimental or literature evidences, which cover 5506 zebrafish genes (∼29% of the 18 971 coding genes). However, the integrated network using supervised learning and Bayesian integration (19) of the genomics big data, DanioNet, contains 817 547 co-functional links between 16 063 genes, which covers ∼85% of the coding genome. Therefore, DanioNet provides new opportunities for functional predictions of many previously uncharacterized *D. rerio* genes. In addition, DanioNet contains ∼61% of human orthologs and ∼58% and ∼56% of disease genes annotated by Disease Ontology (DO) (35) and Online Mendelian Inheritance in Man (OMIM) (36), respectively, which potentiate the application of DanioNet for human disease research. DanioNet will continue to increase its coverage by incorporating various genomics data that continue to be accumulated in public databases. For example, we did not use gene expression data based on high throughput RNA sequencing (RNA-seq) for the construction of the current version of DanioNet, but plan to incorporate them into a future update. Recently, the amount of RNA-seq data in the public domain has grown rapidly. Accordingly, we anticipate that the incorporation of these data in the next version will significantly expand DanioNet.

Network algorithms for generating functional hypotheses

Co-functional networks usually contain some genes known for functional information, which can be propagated to other genes through the network, thus enabling their functional predictions. We can predict pathway/phenotype information for a gene (gene-centric prediction, Figure 2A) or new member genes for a pathway/phenotype (pathway-centric prediction, Figure 2B). Similar to reverse-genetics, which is gene-driven genetic search, gene-centric prediction begins with a gene of interest and progresses through the collection of information of pathway/phenotype annotations of all the network neighbors of the gene. Candidate pathways/phenotypes are ranked from the one with most edge weight to the neighbor genes annotated. DanioNet allows predictions for six different functional annotations: (i) GOBP (22) annotations of zebrafish genes, (ii) zebrafish anatomical (ZFA) (23) annotations, (iii) GOBP (22) annotations of human orthologs, (iv) OMIM (36) annotations of human orthologs, (v) DO (35) annotations of human orthologs and (vi) human phenotype ontology (37) annotations of human orthologs.

In contrast, pathway-centric prediction is driven by a pathway/phenotype, which is analogous to the forward-genetics approach. Occasionally, we have some known genes for the given phenotype of interest and we can then look for additional genes involved in the same pathway/phenotype. This can be done by propagating pathway/phenotype information through network neigh-



D Fold enrichment

Organ	Top 50		Top 100	
	DN	ND	DN	ND
Brain	9.08	3.02	7.56	2.01
Kidney	43.11	25.86	25.86	25.86
CNS	4.21	2.10	5.26	1.05
Eye	6.05	2.42	6.05	3.32
Fin	49.18	28.10	35.13	31.61
Liver	11.54	2.88	7.93	5.77
Heart	18.82	13.86	13.37	14.85

Figure 2. Predictions for zebrafish organ development genes using DanioNet. Schematic overview of (**A**) gene-centric prediction and (**B**) pathway-centric prediction, which can be further classified into a method using only direct neighbor (DN) and another using all the genes of the network by diffusing information throughout the network (ND). (**C**) DanioNet is highly predictive for genes involved in the development of each of the seven zebrafish organs. The precision of the determination of candidates by network prediction was substantially higher than that by random chance. (**D**) The same results are summarized as fold enrichment of network predictions compared to the predictions by random chance. (**E**) Distribution of fold enrichments for 254 zebrafish anatomical (ZFA) terms that contain at least four member genes. Fold enrichment is higher for the top 50 candidates and the method using only direct neighbors (DN) outperforms the method using all the genes of the network by network diffusion (ND).

bors *via* two categories of label propagation algorithms: direct neighborhood (DN) and network diffusion (ND) (38) (Figure 2B). The former transfers node information to only the directly connected neighbors, whereas the latter diffuses the information of nodes throughout the network. We previously demonstrated that neither of these algorithm are more effective than the other for all phenotypes, and the choice of network algorithm is largely determined by the connectivity of member genes for each phenotype in the network (39). Therefore, in this study, we implemented both network propagation algorithms in a companion web server (www.inetbio.org/danionet) of DanioNet.

DanioNet is highly predictive for vertebrate organ development genes

Zebrafish has been an important model organism for the study of vertebrate development (6). Forward and reverse genetics approaches have genetically dissected the developmental processes for each organ and anatomical structures of vertebrates. Reverse genetics in zebrafish is particularly useful for the discovery of novel genes associated with organ development because of the well-established experimental procedures of targeted gene-knockdown with highly specific morpholino and the ease of *in vivo* examination

of organ development owing to the transparent body of zebrafish. Hence, a bioinformatics platform for predicting highly probable candidate genes would facilitate the genetic dissection of organ development in vertebrates. We tested whether DanioNet can help predict associations between genes involved in the developmental process of the same organ or anatomical structures based on ZFA annotations (40). We purposely excluded the ZFA annotations from the data for the training network to avoid the circularity in evaluating the prediction capacity of DanioNet for zebrafish organ development. Given that only dozens of the most significant candidate genes were usually applied to the follow-up experimental analysis, we focused our evaluation on the top 50 or 100 candidates only. We first assessed the network predictions based on the precision of the top 50 or 100 candidates for genes involved in seven distinct organs—the brain, central nervous system, eye, fin, heart, kidney and liver. To conduct our assessment in a conservative manner, we used two independent gene sets for organ development, a set of seed genes from which network propagation of organ information starts, and a set of test genes to validate the new predictions. The seed gene sets were derived from GOBP terms relevant to development or morphogenesis of each organ (Supplementary Table S3) and the test gene sets were compiled from ZFA annotations. We observed substantially higher precisions by DanioNet, with both DN and ND algorithms, compared to those by random chance for all the seven organs (Figure 2C and D). Notably, the predictions based on only direct neighbors (DN) generally outperformed those based on all the network nodes (i.e. ND) for zebrafish organ development.

We also tested the DanioNet-based predictions for 254 ZFA terms that contain at least four member genes. We observed a similar trend of improved predictions by DanioNet for the majority of the tested ZFA terms (Figure 2E). For example, median fold enrichment by DanioNet for the top 50 or top 100 candidates was over 45, which indicated that DanioNet is highly predictive for the genes involved in wide varieties of anatomical developmental processes. Recently, a classifier based on functional associations between zebrafish genes was used to predict the genes involved in each of the anatomical processes in zebrafish (41). However, this classifier was trained for only a fixed set of ZFA terms, whereas DanioNet-based prediction is not restricted to the predefined ZFA anatomical processes.

DanioNet is highly predictive for a wide spectrum of human diseases

Given that DanioNet is highly predictive for the anatomical development in zebrafish, we hypothesized that DanioNet could help reconstruct the functional modules associated with anatomical developmental processes. Presuming that functional modules for different anatomical development largely overlap due to pleiotropy, we searched for modules for anatomical development in DanioNet using CFinder (42), a software that identifies network communities with overlaps. A total of 314 communities were identified based on six clique-connected sub-graphs in the network of 100 000 most significant links of DanioNet (Supplementary Table S4). We then tested these communities for their association with ZFA gene sets using Fisher's exact test. Using a criterion of $P < 0.01$, we found that 197 (62.7%) of the 314 communities were significantly associated with at least one ZFA term. These results suggested that the DanioNet network communities tend to be associated with anatomical developmental processes in zebrafish. Notably, we also found that 185 (58.9%) of the 314 communities were significantly associated with at least one DO term using the same criterion, which also suggested that DanioNet has potential applications for human disease research.

Many of the network communities were associated with zebrafish anatomical terms and human disease terms that are related to each other (Figure 3A). For example, community C2 (i.e. the second largest community) was significantly associated with ZFA of microglial cells ($P = 3.90e\text{-}5$) and DO of multiple sclerosis ($P = 1.15e\text{-}10$), and the microglial cell is known to participate in all phases of multiple sclerosis (43). Similarly, community C5 is significantly associated with ZFA of retinal rod cells ($P = 2.30e\text{-}3$) and DO of retinal dystrophy ($P = 4.78e\text{-}4$), and community 13 is associated with ZFA of cardiac ventricle ($P = 6.66e\text{-}7$) and DO of aortic aneurysm ($P = 2.03e\text{-}4$). In addition to these, many other communities showed significant relationships with the associated ZFA and DO terms (see Figure 3A).

Given the associations between the network communities and specific human diseases in DanioNet, we hypothesized that DanioNet can help predict human disease genes using co-functional networks. We therefore measured the precision of the top 50 or 100 candidate genes determined using DanioNet predictions for 509 DO terms and 55 OMIM terms, which contain at least four member genes, and found a substantial increase in precision from that of random chance (Figure 3B and C). Notably, the predictions made using annotations of all the genes in the network (ND) generally outperformed those based on only DNs for human disease genes.

The disease genes predicted by DanioNet were also validated using candidate disease genes from GWAS. A major database of GWAS studies, GWASdb (44), assigns relevant DO annotations for individual studies. We performed network-based disease gene predictions using seed genes by DO annotation, then validated the new predictions by the GWAS candidate genes for the relevant DO annotations. We found that GWAS candidates are highly enriched among top 200 disease gene candidates by DanioNet with either DN or ND method for many DO disease terms (Figure 4A). Among 106 DO disease terms tested, 48 (45.3%) disease terms showed enrichment for the top 200 GWAS candidate genes by at least 2-fold (see Supplementary Table S5); the top 20 most predictive DO disease terms and their fold enrichment are presented in Figure 4B. These results clearly demonstrated the high predictive capacity of DanioNet for the wide spectrum of human diseases.

Function-driven discovery of novel genes for ciliopathies using DanioNet

To demonstrate the feasibility of *function-driven* disease gene discovery using DanioNet, we predicted novel genes for ciliopathies, a group of disorders associated with genetic defects causing abnormal formation or function of the pri-

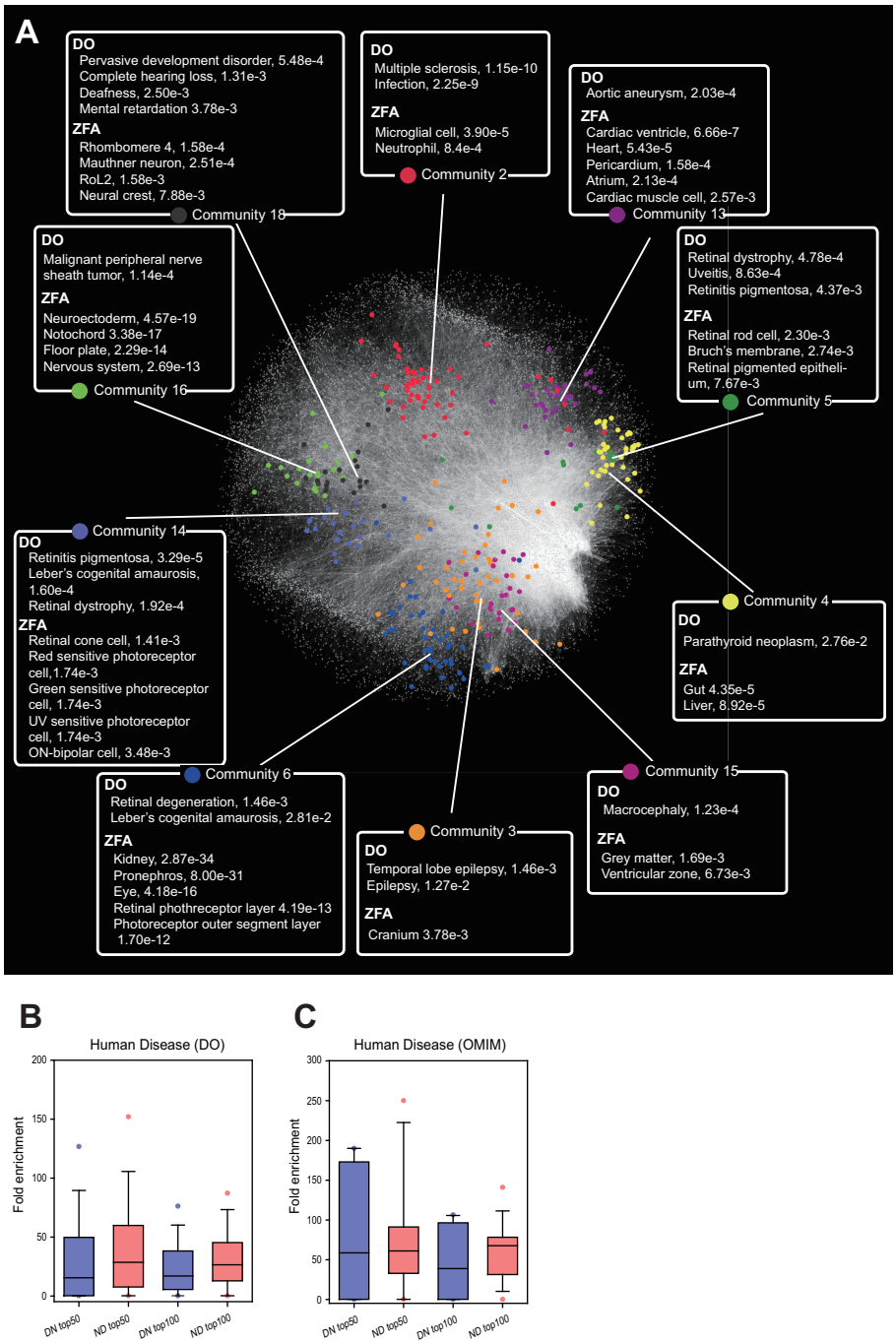


Figure 3. DanioNet is predictive for human disease genes. (**A**) Major communities of the largest component of the 100 000 most significant links of DanioNet were identified using CFinder analysis and are indicated by different color codes. A few significantly associated ZFA terms and DO terms for each community are listed along with their significance scores. Distribution of fold enrichment for (**B**) 509 DO terms and (**C**) 55 OMIM terms by DanioNet with either ND or DN network algorithm among the top predictions. We used human disease annotation terms with at least four member genes only.

mary cilia (45). Ciliopathies are Mendelian diseases with high genetic heterogeneity, in which mutations in many different genes can result in a wide spectrum of cilia-related symptoms. However, many individuals of the disease cohorts have no deleterious mutations in the known causative genes, thereby suggesting the potential association of many unknown genes with ciliopathies. In particular, in the case of diseases caused by genes with only rare mutation occurrence, cohort frequency-based disease gene discovery would be unsuitable. Therefore, we used a complementary approach based on functional evidences in model organisms, where network-based gene prioritization can greatly facilitate the identification of genes with disease-related mutant phenotype. The network propagation methods require seed genes for the target disease. We therefore compiled ciliopathy genes (74 human genes and 88 zebrafish orthologs)

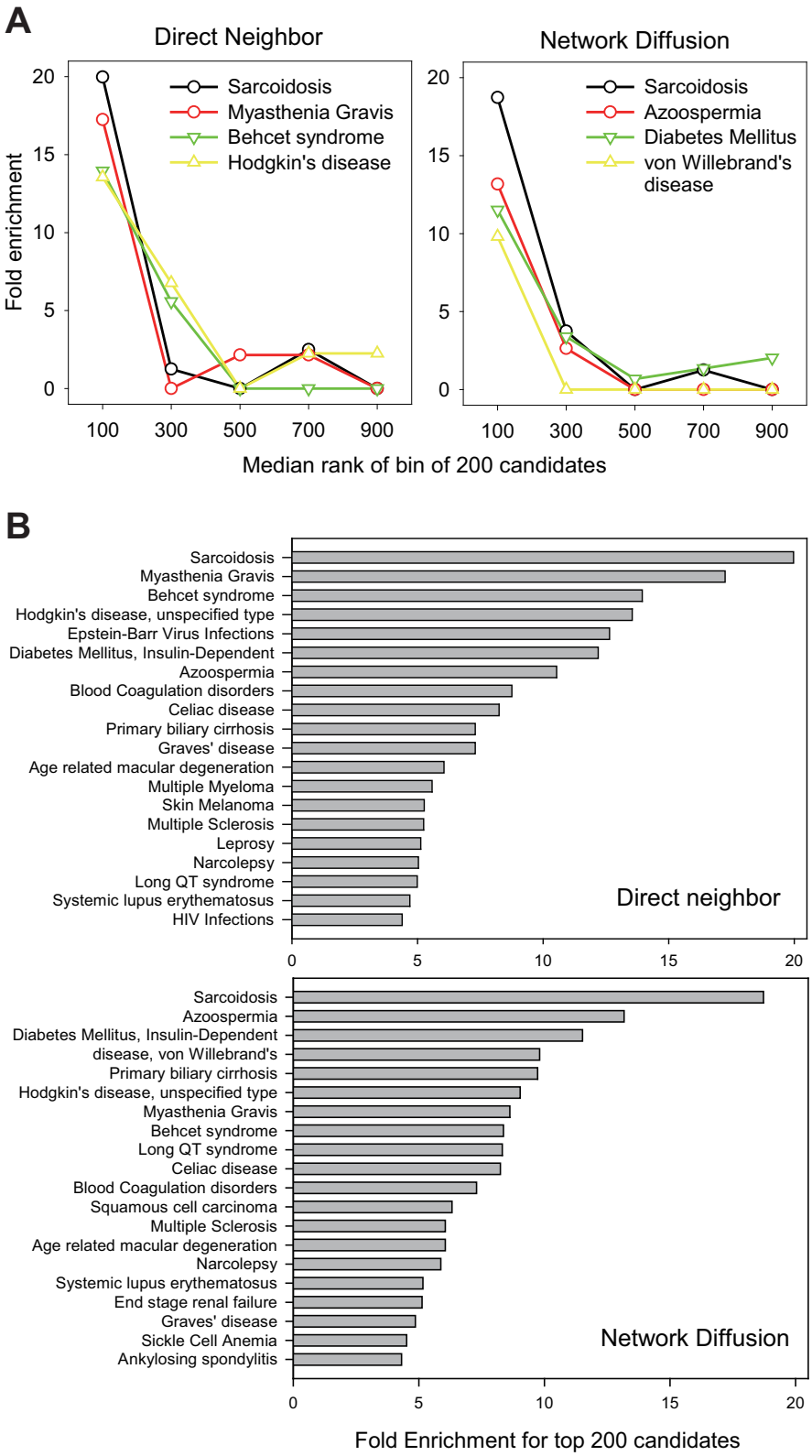


Figure 4. DanioNet is predictive for candidate disease genes from GWAS. (**A**) Fold enrichment of GWAS candidates among the top predictions by DanioNet using the DN or ND method. Candidate disease genes from GWAS were highly enriched within the top 200 predictions made using the network method for various DO diseases. (**B**) Fold enrichment of the top 20 most predictive diseases by DO annotation using DanioNet with DN or ND.

from several review papers published before the year 2013 (46–50) (Supplementary Table S6). Using these as the seed information, we prioritized zebrafish genes for ciliopathies using direct neighbors in DanioNet. To validate our predictions, we collected more recently reported ciliopathy genes (33 human genes and 39 zebrafish orthologs) from literatures published after 2013 (Supplementary Table S7) and known ciliary genes from the SYSCILIA gold standard (SCGSv1) database (51) (268 human genes and 333 zebrafish orthologs) (Supplementary Table S8). Among the top 100 predictions with corresponding human orthologs (Supplementary Table S9), we found seven recently reported ciliopathy genes (7% of precision) and 35 known ciliary genes (35% of precision). Therefore, compared to the precisions expected by random chance (39/18 971 = 0.21% for the recently reported ciliopathy genes and 333/18 971 = 1.75% for known ciliary genes), ∼33-fold enrichment (7% versus 0.21%) and ∼20-fold enrichment (35% versus 1.75%) of precisions among top 100 predictions by DanioNet were observed for the validation gene sets, respectively. These results strongly suggest that other candidates are also likely to be associated with ciliopathies.

We also experimentally validated some of the top predictions by conducting *loss-of-function* studies in zebrafish. We selected eight candidate genes—*rpe*, *ak5*, *tctex1d1*, *slc12a8*, *tekt3*, *tubb4b* and *pax2a*, which are neither ciliopathy genes nor ciliary genes, and *cnga5* was suggested as a ciliary gene that in yet unknown for ciliopathies. These candidate ciliopathy genes were knocked down by two different morpholinos targeting non-overlapping regions of unprocessed mRNA (MO1 and MO2, whose sequence information is available from Supplementary Table S10) and then about 100 morpholino-injected embryos (morphants) for each gene were analyzed for various ciliopathy-relevant phenotypes such as body curvature, hydrocephalus, malformed otolith and heart edema (Figure 5A and Supplementary Figure S1A). We analyzed each morphological phenotype in detail and found that body curvature and heart edema were common defects for most of the candidate genes but that hydrocephalus and otolith malformation were significant for some of the candidate genes, suggesting possible involvements of the candidate genes in ciliopathies (Figure 5B and Supplementary Figure S1B). Moreover, to clarify their cilia-related functions, the morphants were analyzed for cilia-specific phenotypes, such as heart asymmetry and olfactory ciliogenesis in zebrafish. Absent or reversed asymmetries of the hearts were displayed in certain genes, including *ak5*, *tekt3* and *tubb4b*, thereby indicating the association of these genes with the function of nodal cilia (Figure 5C–E). Remarkably, depletions of the majority of the candidates affected the formation of olfactory cilia, which suggested that these candidates play an important role in ciliogenesis (Figure 5F–H). To rule out the possibility that the reduced olfactory cilia might be due to malformation of the olfactory placode, its cell numbers were counted and compared within the areas of observation. The results revealed little difference in the number of cells generating olfactory cilia between control and candidate gene morphants (Figure 5I).

The potential involvement of the validated candidates in ciliopathies was further tested by gene-centric prediction of their OMIM disease terms. Based on the enriched OMIM terms among network neighbors, the probable OMIM terms for each of the eight candidate ciliopathy genes were prioritized. Notably, we observed many ciliopathies such as Bardet-Biedl syndrome (BBS), Joubert syndrome (JBTS), Mechkel–Gruber syndrome (MKS) and nephronophthisis among the top ten predicted diseases (Supplementary Table S11).

Novel genes for ciliopathies have rare coding variants in individuals with ciliopathies

The eight candidate genes associated with cilia formation/function have not been reported as ciliopathy genes to date. Therefore, we investigated whether individuals with ciliopathies have potential causal variants for the eight novel ciliopathy genes. Recently, the UK10K consortium project released genomic variant information for nearly 10 000 individuals including approximately 1000 with rare diseases (25). We obtained the exome sequencing data for 122 ciliopathy individuals and 3781 healthy controls from the UK10K consortium and searched for rare variants (MAF < 0.01%) of the eight novel ciliopathy genes. We could not find any homozygous rare variant, but found 29 heterozygous rare variants within *RPE, PAX2, SLC12A8, TEKT3* and *TUBB4A* in 18 individuals (Supplementary Table S12). Among them, 19 variants were missense or frameshift variants, and many of them were evaluated as deleterious or damaging variants using SIFT (52), PolyPhen2 (53) and VEP (28). Notably, a missense variant of the human *PAX2* gene, which changes alanine from 406 amino acid position to proline (A406P), was observed in 6 of the 122 UK10K individuals with ciliopathies (highlighted in Supplementary Table S12), but was not found in any of the 3781 controls from the ALSPAC or the TwinsUK cohorts. High enrichment of the A406P mutations in *PAX2* among patients suggests that this variant is a potential genetic modifier that enhances the risk of ciliopathies via epistatic interactions with other causal genes (54). WES of a large number of patients has been proposed as a promising approach to dissect the genetics of ciliopathies due to their oligogenicity and attributes of complex disease (55). The enrichment of rare variants among candidate disease genes with functional evidences demonstrates how integrated genomics big data resources can facilitate the functional interpretation of patient-derived genomic variants in disease progression.

DanioNet elucidates a functional module for ciliopathies

Several diseases belonging to ciliopathies genetically and phenotypically overlap each other (56). Therefore, we hypothesized that the 81 seed genes and eight novel genes for ciliopathies could be interconnected to reconstruct a functional module for ciliopathies in DanioNet. To avoid potential circularity in the functional interpretation of an elucidated module, we excluded network links based on co-citation of zebrafish genes (DR-CC) that may be derived from the studies of ciliopathies. To find a more extended module with high enrichment for ciliopathy genes, we permitted the inclusion of intermediate nodes that are signif-

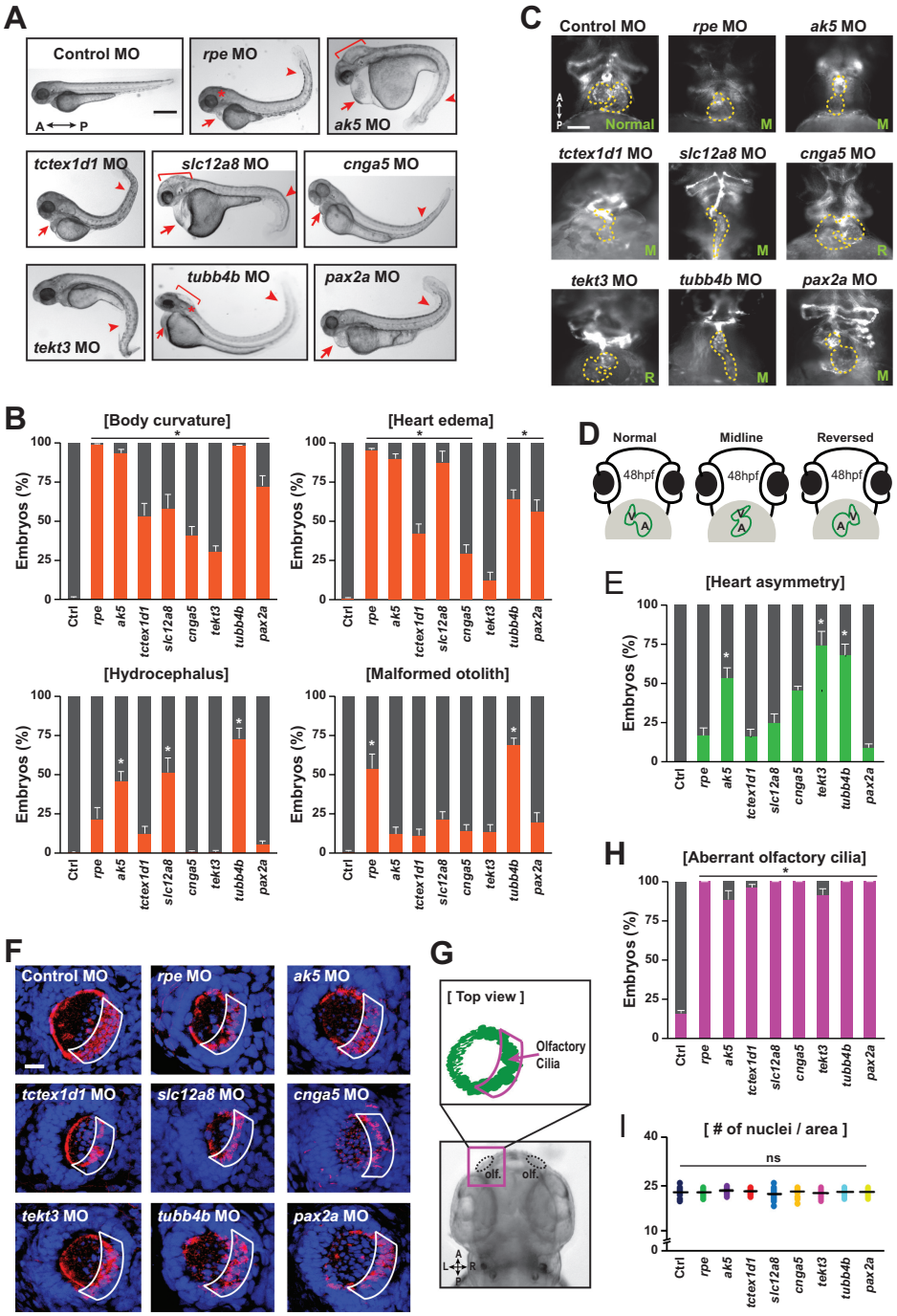


Figure 5. Validation of novel ciliopathy genes in zebrafish. (**A**) The MO-injected zebrafish larvae were observed at 3 days post-fertilization (dpf) under a bright-field microscope. The depletion of majority of the candidate genes resulted in ciliopathy-associated morphological abnormalities such as curved tail (arrowhead), heart edema (arrow) and abnormal otolith (asterisk) and hydrocephalus (bracket). Scale bar is 500 μm. A, anterior; P, posterior. (**B**) The graphs summarize the quantified data of each ciliopathy-related phenotype analyzed in the candidate genes-silenced zebrafish larvae. The data are shown as the mean ± s.d.: $^{*}P < 0.01$ (Student's *t*-test). All data were obtained from at least three independent experiments. (**C**) The asymmetry of the heart, controlled by nodal cilia, was observed in the Tg(*flk1:eGFP*) zebrafish injected with morpholinos at 48 hpf. The representative images of each morphant indicate midlined (M) or reversed (R) asymmetry of the hearts. Scale bar is 125 μm. A, anterior; P, posterior. (**D**) The diagrams show simplified normal and abnormal (M and R) heart asymmetry defects in the zebrafish at 48 hpf. V, ventricle; A, atrium. (**E**) The Tg(*flk1:eGFP*) zebrafish larvae with impaired heart asymmetry were counted and the quantified data were presented in the graph. The data are shown as the mean ± s.d.: $^{*}P < 0.01$ (Student's *t*-test). All data were obtained from at least three independent experiments. (**F**) The MO-injected zebrafish larvae were fixed and immunostained with anti-GT335 antibody at 3 dpf. The control MO-injected zebrafish larvae showed nicely organized olfactory cilia, while most of the candidate gene-silenced larvae showed abrogation of the cilia. Scale bar is 10 μm. (**G**) The diagram shows the head area of the zebrafish larvae containing olfactory organs at 72 h post-fertilization (hpf). The rectangle indicates the photographed region after immunostaining with anti-GT335 antibodies and DAPI, and the inset shows the GT335-marked olfactory cilia. Olf: Olfactory placodes. (**H**) The graph presents the quantified data of the analysis of olfactory ciliogenesis defects. The data are shown as the mean ± s.d.: $^{*}P < 0.01$ (Student's *t*-test). (**I**) The graph quantifies the data by counting DAPI-labeled cells within the area of observation in each morphant.

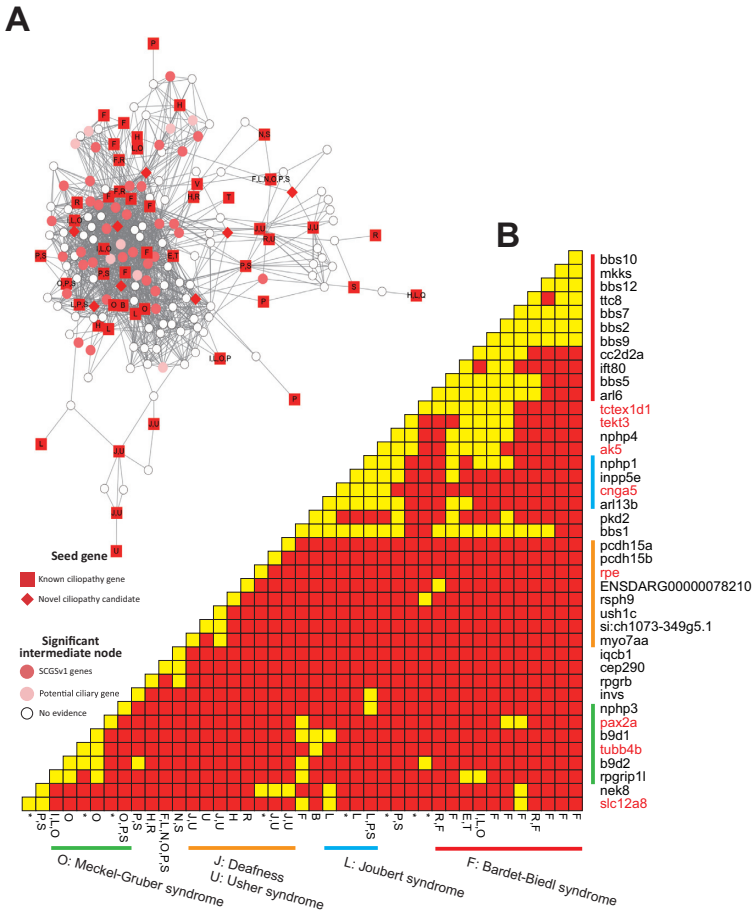


Figure 6. Network modules for ciliopathies. (**A**) A subnetwork of 82 known and eight novel ciliopathy genes in DanioNet. Among the 90 ciliopathy genes, 41 were connected in the largest subnetwork in DanioNet. The network also includes significant intermediate nodes to increase network modularity. Notably, some intermediate nodes are in fact gold standards or potential ciliary genes annotated by the SCGSv1 database (see Supplementary Tables S8 and S13). (**B**) Clusters of ciliopathy genes based on their overlap of network neighbors. Yellow color indicates the presence of a network connection between the two genes crossed. The 82 genes and eight novel candidates (* or red gene name) from this study were clustered by their adjacency matrix based on DanioNet. We observed four clusters enriched for distinct disease phenotypes. Disease codes: [B] ADPKD, Autosomal Dominant Polycystic Kidney Disease; [E] ATD, asphyxiating thoracic dystrophy; [F] BBS, Bardet–Biedl syndrome; [H] CILD, primary ciliary dyskinesia; [I] COACH, Cerebellar vermis hypo/aplasia, Oligophrenia, Ataxia, Coloboma and Hepatic fibrosis; [J] Deafness; [L] JBTS, Joubert syndrome; [N] LCA, Leber Congenital Amaurosis; [O] MKS, Meckel–Gruber syndrome; [P] NPHP, nephronophthisis; [R] RP, Retinitis Pigmentosa; [S] SLSN, Senior Løken syndrome; [T] SRP, Short-Rib-Polydactyly; [U] Usher, Usher syndrome.

icantly connected to other genes of the module (see Materials and Methods). The largest network enriched for the ciliopathy genes included 41 of the 90 ciliopathy genes (Figure 6A). We found that the known genes as well as the novel genes for ciliopathies were well connected with each other, although they were involved in diseases with diverse phenotypes, which is consistent with their overlap in causal genes and phenotypes. Interestingly, many of the intermediated nodes were also found to be implicated in ciliopathies when annotated using gold standards (Supplementary Table S8) or potential ciliary genes (Supplementary Table S13) from the SCGSv1 database (51).

Although diverse ciliopathies are genetically interwoven in the functional module, clusters for individual diseases with distinct phenotypes may also be observable. Hence, we conducted clustering analysis for the ciliopathy genes based on network context similarity, where genes with similar network neighbors are clustered together. We found that the ciliopathy genes were roughly clustered into four groups enriched for genes for MKS, Usher syndrome (Usher), JBTS and BBS (Figure 6B). Interestingly, there was a rough correlation between the relative phenotypic effect of the novel ciliopathy genes in zebrafish and the severity of the ciliopathies to which they were closely related in the clusters. MKS generally shows more severe disease phenotypes than other ciliopathies. We observed a high phenotypic effect during *loss-of-function* assays for *pax2a*, *tubb4b* and *slc12a8* in zebrafish (see Figure 5C), and these genes were clustered with genes for MKS. In contrast, the knockdown of *tctex1d1*, *cnga5a* and *tekt3*, which were clustered with genes for BBS and JBTS, showed a relatively low phenotypic effect in zebrafish. This correlation between the phenotypic effect in zebrafish and the disease severity in humans further supports the usefulness of *function-driven* disease gene discovery using zebrafish model.

CONCLUSIONS

The main purpose of causative gene identification is the treatment of the diseases. Thus, disease models are ultimately necessary for validating the phenotypic effect of causal gene defects and for screening therapeutic agents. In this respect, a disease gene discovery pipeline that starts with obtaining functional evidence of candidate disease genes, followed by the analysis of genetic variants in patients, could be beneficial. Among laboratory animals, zebrafish has several practical advantages for genetic analysis in addition to its genomic and anatomical similarity to humans. Therefore, we developed DanioNet, an integrated genomics big data resource to facilitate zebrafish-based *function-driven* disease gene discovery. Rigorous statistical assessment confirmed the high prediction capacity of DanioNet not only for anatomical development in zebrafish, but also for a wide variety of human diseases. To demonstrate the feasibility of DanioNet-assisted *function-driven* disease gene discovery, we predicted and experimentally validated eight novel genes for ciliopathies, a class of multi-organ diseases caused by disorders in the primary cilium. We also found rare coding variants of the novel ciliopathy genes in individuals with ciliopathies from the UK10K consortium. Furthermore, we used DanioNet to reconstruct a functional module of ciliopathies that is highly enriched for the known and novel ciliopathy genes and shows clusters of ciliopathies for distinct phenotypes. Given that ciliopathies have attributes of complex diseases, our proposed method should be useful for the genetic dissection of human complex diseases as well.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This study makes use of data generated by the UK10K Consortium, which is derived from samples from the TwinsUK Cohort, the ALSPAC Cohort, and Cilia in Disease and Development study (CINDAD). A full list of the investigators who contributed to the generation of the data is available from www.UK10K.org. Funding for UK10K was provided by the Wellcome Trust under award **WT091310**.
Author contributions: H.S. constructed the integrated functional gene network of *Danio rerio* and performed bioinformatics analysis. J.K. selected the candidate genes for ciliopathy and performed experimental validation in zebrafish. C.K. performed network analysis and sequence variant data analysis. S.H., H.K. and S.Y. assisted with network modeling and web-server construction. J.E.L. and I.L. supervised and coordinated the study. H.S., J.K., C.K., J.E.L. and I.L. wrote and edited the manuscript.

FUNDING

National Research Foundation of Korea [2012M3A9B4028641, 2012M3A9C7050151, 2015R1A2A1A15055859 to I.L. and 2016R1C1B2008930 to J.E.L.]. Funding for open access charge: National Reseach Foundation of Korea [2012M3A9B4028641, 2012M3A9C7050151, 2015R1A2A1A15055859 to I.L. and 2016R1C1B2008930 to J.E.L.].
Conflict of interest statement. None declared.

REFERENCES

1. Bamshad,M.J., Ng,S.B., Bigham,A.W., Tabor,H.K., Emond,M.J., Nickerson,D.A. and Shendure,J. (2011) Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.*, **12**, 745–755.
2. Bell,C.J., Dinwiddie,D.L., Miller,N.A., Hateley,S.L., Ganusova,E.E., Mudge,J., Langley,R.J., Zhang,L., Lee,C.C., Schilkey,F.D. *et al.* (2011) Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci. Transl. Med.*, **3**, 65ra64.
3. Xue,Y., Chen,Y., Ayub,Q., Huang,N., Ball,E.V., Mort,M., Phillips,A.D., Shaw,K., Stenson,P.D., Cooper,D.N. *et al.* (2012) Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. *Am. J. Hum. Genet.*, **91**, 1022–1032.
4. Norton,N., Robertson,P.D., Rieder,M.J., Zuchner,S., Rampersaud,E., Martin,E., Li,D., Nickerson,D.A., Hershberger,R.E., National Heart, L *et al.*2012) Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. *Circ. Cardiovasc. Genet.*, **5**, 167–174.
5. MacArthur,D.G., Manolio,T.A., Dimmock,D.P., Rehm,H.L., Shendure,J., Abecasis,G.R., Adams,D.R., Altman,R.B., Antonarakis,S.E., Ashley,E.A. *et al.* (2014) Guidelines for investigating causality of sequence variants in human disease. *Nature*, **508**, 469–476.
6. Lawson,N.D. and Wolfe,S.A. (2011) Forward and reverse genetic approaches for the analysis of vertebrate development in the zebrafish. *Dev. Cell*, **21**, 48–64.
7. Santoriello,C. and Zon,L.I. (2012) Hooked! Modeling human disease in zebrafish. *J. Clin. Invest.*, **122**, 2337–2343.
8. Davis,E.E., Frangakis,S. and Katsanis,N. (2014) Interpreting human genetic variation with in vivo zebrafish assays. *Biochim. Biophys. Acta.*, **1842**, 1960–1970.
9. Howe,K., Clark,M.D., Torroja,C.F., Torrance,J., Berthelot,C., Muffato,M., Collins,J.E., Humphray,S., McLaren,K., Matthews,L. *et al.* (2013) The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, **496**, 498–503.
10. Gieger,C., Radhakrishnan,A., Cvejic,A., Tang,W.H., Porcu,E., Pistis,G., Serbanovic-Canic,J., Elling,U., Goodall,A.H., Labrune,Y. *et al.* (2011) New gene functions in megakaryopoiesis and platelet formation. *Nature*, **480**, 201–208.
11. Liu,L.Y., Fox,C.S., North,T.E. and Goessling,W. (2013) Functional validation of GWAS gene candidates for abnormal liver function during zebrafish liver development. *Dis. Model Mech.*, **6**, 1271–1278.
12. Xiao,S.M., Kung,A.W.C., Gao,Y., Lau,K.S., Ma,A., Zhang,Z.L., Liu,J.M., Xia,W.B., He,J.W., Zhao,L. *et al.* (2012) Post-genome wide association studies and functional analyses identify association of MPP7 gene variants with site-specific bone mineral density. *Hum. Mol. Genet.*, **21**, 1648–1657.
13. Manzini,M.C., Tambunan,D.E., Hill,R.S., Yu,T.W., Maynard,T.M., Heinzen,E.L., Shianna,K.V., Stevens,C.R., Partlow,J.N., Barry,B.J. *et al.* (2012) Exome sequencing and functional validation in zebrafish identify GTDC2 mutations as a cause of Walker-Warburg syndrome. *Am. J. Hum. Genet.*, **91**, 541–547.
14. Mirabello,L., Macari,E.R., Jessop,L., Ellis,S.R., Myers,T., Giri,N., Taylor,A.M., McGrath,K.E., Humphries,J.M., Ballew,B.J. *et al.* (2014) Whole-exome sequencing and functional studies identify RPS29 as a novel gene mutated in multicase Diamond-Blackfan anemia families. *Blood*, **124**, 24–32.
15. MacRae,C.A. and Peterson,R.T. (2015) Zebrafish as tools for drug discovery. *Nat. Rev. Drug Discov.*, **14**, 721–731.
16. Lieschke,G.J. and Currie,P.D. (2007) Animal models of human disease: Zebrafish swim into view. *Nat. Rev. Genet.*, **8**, 353–367.
17. Carter,H., Hofree,M. and Ideker,T. (2013) Genotype to phenotype via network analysis. *Curr. Opin. Genet. Dev.*, **23**, 611–621.
18. Lee,I. (2013) Network approaches to the genetic dissection of phenotypes in animals and humans. *Anim. Cells Syst*, **17**, 75–79.
19. Lee,I., Date,S.V., Adai,A.T. and Marcotte,E.M. (2004) A probabilistic functional network of yeast genes. *Science*, **306**, 1555–1558.

20. Troyanskaya,O.G., Dolinski,K., Owen,A.B., Altman,R.B. and Botstein,D. (2003) A Bayesian framework for combining heterogeneous data sources for gene function prediction (in Saccharomyces cerevisiae). *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 8348–8353.
21. Kanehisa,M., Goto,S., Sato,Y., Furumichi,M. and Tanabe,M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.*, **40**, D109–D114.
22. Gene Ontology,C., Blake,J.A., Dolan,M., Drabkin,H., Hill,D.P., Li,N., Sitnikov,D., Bridges,S., Burgess,S., Buza,T. *et al.* (2013) Gene Ontology annotations and resources. *Nucleic Acids Res.*, **41**, D530–D535.
23. Howe,D.G., Bradford,Y.M., Conlin,T., Eagle,A.E., Fashena,D., Frazer,K., Knight,J., Mani,P., Martin,R., Moxon,S.A.T. *et al.* (2013) ZFIN, the Zebrafish Model Organism Database: Increased support for mutants and transgenics. *Nucleic Acids Res.*, **41**, D854–D860.
24. Lappalainen,I., Almeida-King,J., Kumanduri,V., Senf,A., Spalding,J.D., Ur-Rehman,S., Saunders,G., Kandasamy,J., Caccamo,M., Leinonen,R. *et al.* (2015) The European Genome-phenome Archive of human data consented for biomedical research. *Nat. Genet.*, **47**, 692–695.
25. Consortium,U.K., Walter,K., Min,J.L., Huang,J., Crooks,L., Memari,Y., McCarthy,S., Perry,J.R., Xu,C., Futema,M. *et al.* (2015) The UK10K project identifies rare variants in health and disease. *Nature*, **526**, 82–90.
26. Sherry,S.T., Ward,M.H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
27. McKenna,A., Hanna,M., Banks,E., Sivachenko,A., Cibulskis,K., Kernytsky,A., Garimella,K., Altshuler,D., Gabriel,S., Daly,M. *et al.* (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, **20**, 1297–1303.
28. McLaren,W., Pritchard,B., Rios,D., Chen,Y.A., Flicek,P. and Cunningham,F. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*, **26**, 2069–2070.
29. Berger,S.I., Posner,J.M. and Ma'ayan,A. (2007) Genes2Networks: Connecting lists of gene symbols using mammalian protein interactions databases. *BMC Bioinformatics*, **8**, 372.
30. Jenssen,T.K., Laegreid,A., Komorowski,J. and Hovig,E. (2001) A literature network of human genes for high-throughput analysis of gene expression. *Nat. Genet.*, **28**, 21–28.
31. 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.
32. 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.
33. 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.
34. Kim,E., Kim,H. and Lee,I. (2013) JiffyNet: a web-based instant protein network modeler for newly sequenced species. *Nucleic Acids Res.*, **41**, W192–W197.
35. Kibbe,W.A., Arze,C., Felix,V., Mitraka,E., Bolton,E., Fu,G., Mungall,C.J., Binder,J.X., Malone,J., Vasant,D. *et al.* (2015) Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res.*, **43**, D1071–D1078.
36. 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.
37. Kohler,S., Doelken,S.C., Mungall,C.J., Bauer,S., Firth,H.V., Bailleul-Forestier,I., Black,G.C.M., Brown,D.L., Brudno,M., Campbell,J. *et al.* (2014) The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.*, **42**, D966–D974.
38. 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.
39. Shim,J.E., Hwang,S. and Lee,I. (2015) Pathway-dependent effectiveness of network algorithms for gene prioritization. *Plos One*, **10**, e0130589.
40. Bradford,Y., Conlin,T., Dunn,N., Fashena,D., Frazer,K., Howe,D.G., Knight,J., Mani,P., Martin,R., Moxon,S.A. *et al.* (2011) ZFIN: Enhancements and updates to the Zebrafish Model Organism Database. *Nucleic Acids Res.*, **39**, D822–D829.
41. Musso,G., Tasan,M., Mosimann,C., Beaver,J.E., Plovie,E., Carr,L.A., Chua,H.N., Dunham,J., Zuberi,K., Rodriguez,H. *et al.* (2014) Novel cardiovascular gene functions revealed via systematic phenotype prediction in zebrafish. *Development*, **141**, 224–235.
42. Derenyi,I., Palla,G. and Vicsek,T. (2005) Clique percolation in random networks. *Phys. Rev. Lett.*, **94**, 160202.
43. Jack,C., Ruffini,F., Bar-Or,A. and Antel,J.P. (2005) Microglia and multiple sclerosis. *J. Neurosci. Res.*, **81**, 363–373.
44. Li,M.J., Liu,Z., Wang,P., Wong,M.P., Nelson,M.R., Kocher,J.P., Yeager,M., Sham,P.C., Chanock,S.J., Xia,Z. *et al.* (2016) GWASdb v2: An update database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.*, **44**, D869–D876.
45. Hildebrandt,F., Benzing,T. and Katsanis,N. (2011) Ciliopathies. *N. Engl. J. Med.*, **364**, 1533–1543.
46. van Reeuwijk,J., Arts,H.H. and Roepman,R. (2011) Scrutinizing ciliopathies by unraveling ciliary interaction networks. *Hum. Mol. Genet.*, **20**, R149–R157.
47. Forsythe,E. and Beales,P.L. (2013) Bardet-Biedl syndrome. *Eur. J. Hum. Genet.*, **21**, 8–13.
48. Romani,M., Micalizzi,A. and Valente,E.M. (2013) Joubert syndrome: congenital cerebellar ataxia with the molar tooth. *Lancet Neurol.*, **12**, 894–905.
49. Logan,C.V., Abdel-Hamed,Z. and Johnson,C.A. (2011) Molecular genetics and pathogenic mechanisms for the severe ciliopathies: Insights into neurodevelopment and pathogenesis of neural tube defects. *Mol. Neurobiol.*, **43**, 12–26.
50. Wolf,M.T. and Hildebrandt,F. (2011) Nephronophthisis. *Pediatr. Nephrol.*, **26**, 181–194.
51. van Dam,T.J., Wheway,G., Slaats,G.G., Group,S.S., Huynen,M.A. and Giles,R.H. (2013) The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. *Cilia*, **2**, 7.
52. Vaser,R., Adusumalli,S., Leng,S.N., Sikic,M. and Ng,P.C. (2016) SIFT missense predictions for genomes. *Nat. Protoc.*, **11**, 1–9.
53. Adzhubei,I.A., Schmidt,S., Peshkin,L., Ramensky,V.E., Gerasimova,A., Bork,P., Kondrashov,A.S. and Sunyaev,S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
54. Badano,J.L., Leitch,C.C., Ansley,S.J., May-Simera,H., Lawson,S., Lewis,R.A., Beales,P.L., Dietz,H.C., Fisher,S. and Katsanis,N. (2006) Dissection of epistasis in oligogenic Bardet-Biedl syndrome. *Nature*, **439**, 326–330.
55. Novarino,G., Akizu,N. and Gleeson,J.G. (2011) Modeling human disease in humans: The Ciliopathies. *Cell*, **147**, 70–79.
56. Davis,E.E. and Katsanis,N. (2012) The ciliopathies: A transitional model into systems biology of human genetic disease. *Curr. Opin. Genet. Dev.*, **22**, 290–303.