Published online 3 October 2017 *Nucleic Acids Research, 2018, Vol. 46, Database issue* ***D527–D534***
doi: 10.1093/nar/gkx876

DISNOR: a disease network open resource

Prisca Lo Surdo[1], Alberto Calderone[1], Marta Iannuccelli[1], Luana Licata[1], Daniele Peluso[1,2], Luisa Castagnoli[1], Gianni Cesareni[1,*] and Livia Perfetto[1,*]

[1]Bioinformatics and Computational Biology Unit, Department of Biology, University of Rome 'Tor Vergata', 00133 Rome, Italy and [2]Laboratory of Bioinformatic, IRCCS Fondazione Santa Lucia, 00143 Rome, Italy

Received August 07, 2017; Revised September 14, 2017; Editorial Decision September 18, 2017; Accepted September 25, 2017

ABSTRACT

DISNOR is a new resource that aims at exploiting the explosion of data on the identification of disease-associated genes to assemble inferred disease pathways. This may help dissecting the signaling events whose disruption causes the pathological phenotypes and may contribute to build a platform for precision medicine. To this end we combine the gene-disease association (GDA) data annotated in the DisGeNET resource with a new curation effort aimed at populating the SIGNOR database with causal interactions related to disease genes with the highest possible coverage. DISNOR can be freely accessed at http://DISNOR.uniroma2.it/ where >3700 disease-networks, linking ∼2600 disease genes, can be explored. For each disease curated in DisGeNET, DISNOR links disease genes by manually annotated causal relationships and offers an intuitive visualization of the inferred 'patho-pathways' at different complexity levels. User-defined gene lists are also accepted in the query pipeline. In addition, for each list of query genes—either annotated in DisGeNET or user-defined—DISNOR performs a gene set enrichment analysis on KEGG-defined pathways or on the lists of proteins associated with the inferred disease pathways. This function offers additional information on disease-associated cellular pathways and disease similarity.

INTRODUCTION

Advancing the understanding of the molecular basis underlying disorder aetiology is expected to further clinicians' ability to diagnose, treat, and manage patients. The recent explosion of information about genes that are found mutated in patients, made possible by the development of affordable sequencing technologies and genome-wide association studies (GWAS), has contributed to setting this goal within reach. The wealth of data associating gene defects to diseases, gene disease associations (GDAs), is collected, organized and integrated in a number of publicly available resources, such as DisGeNET (1), OMIM (2) and Orphanet (http://www.orpha.net/). These gene lists contain information about the biological processes whose perturbations cause a given genetic disorder and may offer hints about gene-products or pathways to target in order to reverse the disease phenotype.

However, in order to extract this 'hidden information', two main hurdles must be overcome. First, these gene lists are often long and contain genes that represent 'genetic noise'. In addition, one would need to link in a rational framework (pathways) gene mutations, which cause the same disease while sometimes appearing functionally unrelated (3).

Tools, such as Phenolyzer (http://phenolyzer.wglab.org) (4) and Phevor (5), have been developed to approach the first issue. Both tools, freely available, use prior information curated by other resources to implicate genes in diseases and output prioritized lists of disease-causing gene variants. However, they do not offer any information on the molecular mechanisms underlying the patho-phenotypes.

A second fundamental task is the development of computational models to use as frameworks to describe qualitatively, and possibly quantitatively, how the disease-associated biological entities work together to regulate the processes underlying the phenotypic outcome. Such a task requires an understanding of the organization of proteins and pathways in networks (6).

Network-based approaches, by systematically organizing literature information about biological pathways and interactions, may offer mechanistic insights and have already been proposed to help diagnosis, patient stratification and selection of therapeutic strategies, especially in oncology (7,8). These approaches are, however, based on information on physical interactions between gene products and are therefore missing mechanistic information on the directionality and the effect (activation/inactivation) of the observed functional relationships.

In addition, the lack of functional information on genes associated to less-studied diseases, has until now discour-

*To whom correspondence should be addressed. Tel: +39 06 7259 4305; Email: cesareni@uniroma2.it
Correspondence may also be addressed to Livia Perfetto. Email: livia.perfetto@live.it

aged the application of these approaches to other clinical contexts.

A number of commercially available software tools are currently available to address the need of linking disease genes by signaling relationships. Ingenuity Pathway Analysis (www.ingenuity.com), Pathway Studio (www.pathwaystudio.com) and Thomson Reuters MetaCore™ are three pathway analysis tools that, in addition to offering a pathway enrichment analysis, also generate network diagrams that illustrate how the genes in the list are connected by functional/causal relationships.

We introduce here a new resource: DISNOR, which builds on a new curation effort aimed at connecting poorly characterized disease genes to the integrated global signaling network in SIGNOR (9). SIGNOR organizes, in a public relational database, manually curated data on causal interactions (e.g. Entity A up/down-regulates Entity B) between biological entities, having a key role in signal transduction. The majority of the entities in SIGNOR are proteins. However, chemicals, small molecules, protein families, complexes, phenotypes and stimuli are also considered. In a recent independent survey of signalling databases (10), SIGNOR scored as the 'causal interaction' resource with the highest protein coverage and annotation detail. SIGNOR curators capture details describing, among other features, the sign of the interaction and the mechanism underlying the regulatory interaction (e.g. phosphorylation, ubiquitination, etc.), and always refer to the literature report describing the interaction.

DISNOR accepts as input a list of genes, either user-defined or listed in a database annotating disease genes (i.e. DisGeNET), and searches in a database of causal relationships evidence for signaling connections between them. DISNOR is available at http://DISNOR.uniroma2.it/, where it is possible to recall and explore >3700 inferred disease networks, linking ~2600 genes associated to pathologies including rare diseases and cancers.

MATERIALS AND METHODS

Disease data

Protein-disease association data, as well as SNP-disease association data, are downloaded from the DisGeNET (http://www.disgenet.org) CURATED dataset (1). In DisGeNET every disease object is annotated to a 'group', 'phenotype' or 'disease', each defined according to the Unified Medical Language System (UMLS) semantic type (11) and to internal criteria. Data related to diseases annotated under the 'phenotype' (such as clinical and behavioural features) or 'group' (such as 'Neurodegenerative diseases') categories were not considered for inclusion in DISNOR. The disease dataset we currently consider includes 6,323 diseases and their associated genes. Disease terms have been annotated within our database with links to their Concept Unique Identifiers (CUIs) of the NCI Metathesaurus and MeSH Ids, and their category defined according to Disease Ontology (12) (http://www.disease-ontology.org).

Causal interaction data

Causal Interactions are extracted from SIGNOR (http://signor.uniroma2.it) (9). Every interaction in SIGNOR is annotated with details about: (i) the sign of the interaction: a regulator might up- or down-regulate the target by modulating its activity or quantity; (ii) the mechanism underlying the regulatory interaction (e.g. phosphorylation, ubiquitination, etc.) and (iii) a reference (usually the PubMedID) and a sentence (extracted from the cited article) supporting the interaction. Genes that are associated with a common disease in DisGeNET are used as seeds to query SIGNOR using different methods (see paragraph 'Querying Methods') and to extract information associated with the interaction.

Querying methods

Querying methods used to generate the result pages for a single disease are a combination of PostgreSQL database querying and PHP. It is possible to access the network information at varying levels of complexity (levels 1–3), each level being the result of a different search method accessing and/or filtering SIGNOR data. Furthermore, causal interaction networks at any level of complexity can be integrated with protein-protein interactions (PPI) extracted from *mentha* (13).

Level 1 (connect) searches for signaling interactions involving any two entities in the query list.

Level 2 (first neighbours) is a multi-step strategy that initially performs a search in SIGNOR for all interactions involving at least one of any of the seed entities, and then verifies which are occurring directly between two query entities or act as links between two searched entities. The retrieved interactions that do not fall into these categories are excluded by a pruning process. As a consequence the resulting network also includes nodes that link two query entities.

Level 3 (all) allows access to signaling interactions involving any of the seed entities and all the remaining proteins in the SIGNOR network without any further filtering.

The 'Add Physical Interaction from *mentha*' command allows integration of protein-protein interaction data from *mentha* (http://mentha.uniroma2.it/) (13). PPI relationships in *mentha* are assigned a reliability score ranging from 0 to 1. The score is calculated by taking into account the number of experimental evidence supporting the interaction, the experiment type, the interaction type and the experimental scale. By default, only PPIs with a score exceeding the 0.4 threshold are included. This threshold was shown to filter a PPI interaction list with the best F score (harmonic means of precision and recall) to approximate the functional relationships archived in Reactome (14,15).

External data

In addition to the graph representation of the retrieved disease network, the disease result page offers external links to related resources. With the aim of cross-integration, whenever possible, a link to the descriptive disease information contained within the NCI Metathesaurus (NCIm—https://ncim-stage.nci.nih.gov/ncimbrowser/) (16) has been established using disease CUIs and a link to MeSH anno-

tated information has been created using MeSH IDs (17) (https://www.nlm.nih.gov/mesh/). Furthermore, associated gene products were linked to their UniprotKB (http://www.uniprot.org) page (18).

Website

The web user interface was designed using Bootstrap (version 3.3.7). The functionality of the website was developed using JavaScript and PHP (version 7.0). The jQuery Select2 plugin (https://select2.github.io/) has been integrated in order to provide autocompleting support for disease terms, offering a more user-friendly querying experience.

Gene enrichment analysis

The results page offers the possibility of performing a Gene Enrichment Analysis (GEA). This analysis returns a list of the top 20 enriched diseases and pathways, ranked according to a *P*-value calculated with a randomization test. Furthermore, for each significantly enriched annotation, we also report the ratio between the number of genes that are in common with the query gene list and the total number of genes annotated to that disease or pathway. To limit server overload and ensure an acceptable response time, a maximum of 50 proteins are used as input for each query gene list in this analysis.

Gene-pathway annotations and gene-disease annotations were obtained from KEGG (19) and from DisGeNET respectively. The annotation dataset consists of 6935 proteins annotated with 316 different pathways and 6700 proteins annotated with 4530 different diseases.

The *P*-value is calculated as follows: (i) we analyze the queried genes and, for each gene, we extract terms (e.g. KEGG pathways or DisGeNET diseases) from the pathway/disease annotation datasets; (ii) for each pathway or disease term we count how many genes annotated with the term occur in the query gene list; (iii) we randomly select 10 000 sets of the same size of the query gene list and repeat the count and (iv) we count how many times we observe in the random set a match greater than or equal to the one obtained from the starting set and we divide this number by the total number of iterations.

RESULTS

DISNOR scope and novelty

DISNOR is a knowledge base that supports the assembly of disease-related signaling networks to help infer, on a large scale, the pathways that are disrupted in any specific disease. To this end, we integrate the gene-disease annotation in DisGeNET with a new curation effort aimed at capturing signaling information involving disease genes. In addition, the resource offers several querying strategies to extract from the database of signaling relationships proteins that interact with the user-defined proteins and to assemble disease specific signaling networks.

As shown in Figure 1B, DISNOR is a resource that links disease-associated genes (extracted from the DisGeNET resource) via causal relationships annotated in the SIGNOR database. The premise behind the approach is that genes associated with the same or similar diseases commonly reside in the same neighborhood of molecular networks (20) (Figure 1A and B) suggesting that the same biological function or pathway is perturbed in different ways in different patients (21).

DISNOR uses the information about genes associated to diseases, as annotated in DisGeNET (http://www.DisGeNET.org) (1), a resource integrating information on gene-disease association (GDAs) from several public data sources. The data in DisGeNET are organized according to type and level of curation and to reliability distinguishing CURATED from PREDICTED data. DISNOR only considers the curated dataset.

As a source of causal relationships to link disease genes, we choose SIGNOR a database that focuses on the capture, organization and display of signaling interactions between biological entities. SIGNOR uses a logic-based model to present causal signaling interaction data as a signed directed graph, including ∼18 300 entries (9) (Figure 2A).

To summarize, the scope of DISNOR is that of assembling, on a large scale, disease networks defined as the set of causal interactions among genes that have been observed to be mutated in the same pathology, according to the GDA data annotated in DisGeNET.

Curation and data integration

Figure 1 summarizes the integration procedure:

Step 1) For every disease DISNOR extracts from DisGeNET a list of observed genetic variants. As shown in Figure 2A, DisGeNET integrates GDAs data from different resources and contains information for >13 000 diseases. However, only 6300 are annotated with at least one gene variant.

Step 2) For each disease, DISNOR queries SIGNOR, using as an input the corresponding disease associated genes annotated in DisGeNET ('seed' genes). It retrieves cause-effect relationships between 'seed genes' and their next neighbors in the signaling network.

Step 3) DISNOR filters the causal network to generate the disease networks at different levels of complexity (see also Figure 3A).

When we started this project, only 2204 of the 7241 genes associated to diseases in DisGeNET had at least one interaction annotated in SIGNOR. To increase the coverage and the connectivity of disease genes, we carried out a manual curation effort. Essentially, we first created a list of genes that, despite being functionally associated to diseases, were not yet linked to the large causal interaction network in SIGNOR. Next, we used automatic text mining approaches to select articles mentioning the query gene/protein in their abstracts. Expert curators reviewed the automatically retrieved articles to confirm their relevance. Finally, genes were prioritized for curation according to their frequency in the GDA dataset and causal interactions were manually annotated in the DISNOR database, according to SIGNOR curation policy (9,22). As a result, >2500 causal interactions between proteins that have relevance in disease pathways have been curated into the SIGNOR database, allowing approximately 400 new gene products to be integrated

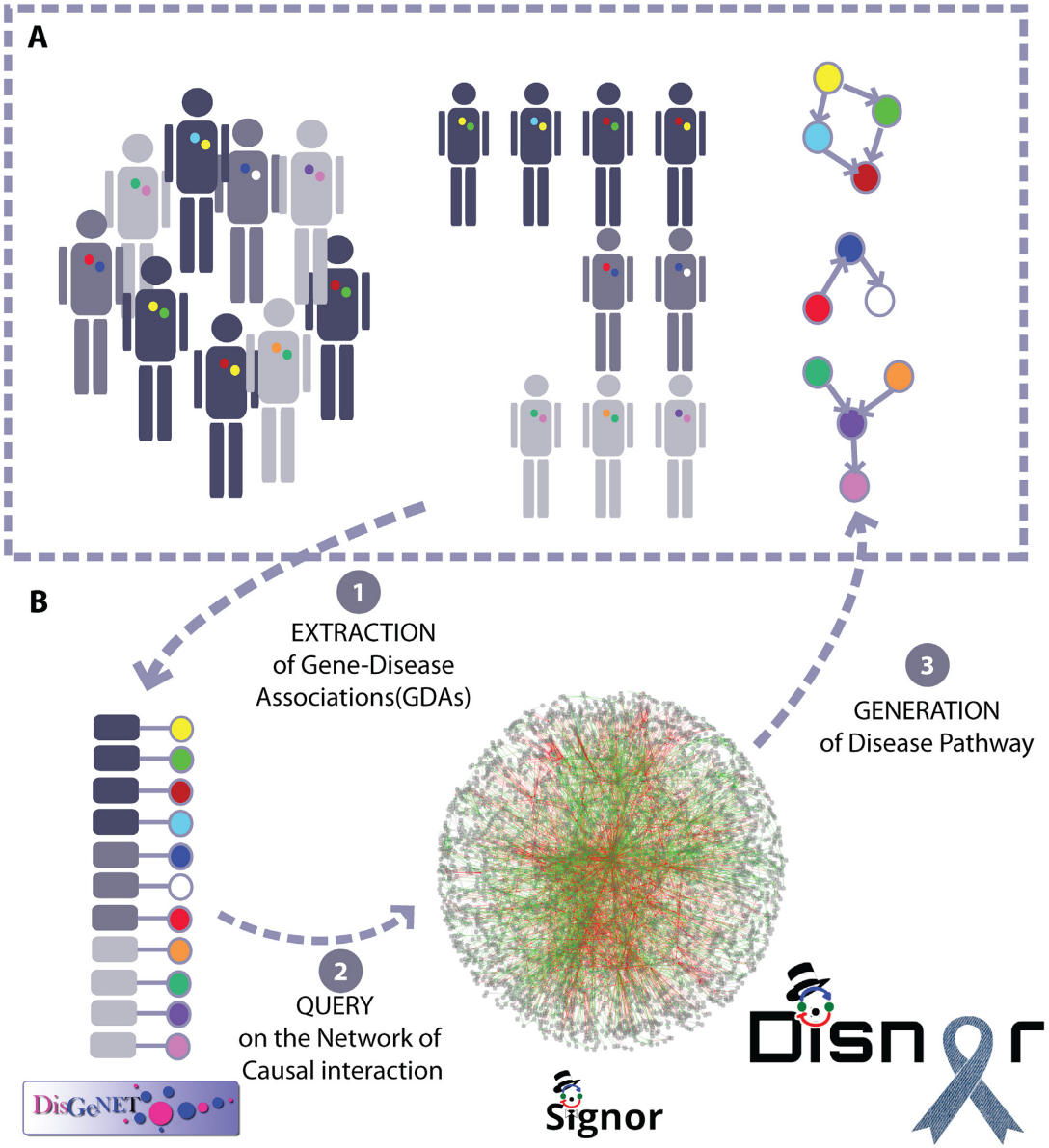

Figure 1. Integration strategy. (**A**) Schematic illustration of the rationale behind DISNOR and its application to network medicine: a population of patients can be stratified according to the pathogenic phenotype. Patients affected by the same disease carry mutations in a heterogeneous set of genes, which, conversely, appear to affect common pathways or biological processes (21). Gray-scale individuals represent different diseases, while differentially coloured circles represent genetic variants. (**B**) A drawing that illustrates the DISNOR integration strategy. *Step 1*: for each disease, DISNOR extracts 'seed' genes (from DisGeNET); *step 2*: seed genes are used to query the SIGNOR knowledge base to extract causal interactions involving 'seed' genes; *step 3*: DISNOR generates a graphic representation of the disease network. Gene-disease associations (GDAs) in DisGeNET are depicted as a link between grey-scale rectangles (representing different diseases) and colored circles (representing genetic variants).

into the network. Most of the remaining proteins are poorly studied from a molecular/mechanistic perspective and their connection to the global causal network is not yet possible. DISNOR also offers the possibility to use physical protein interaction data to link disease genes (see below), thereby allowing the assembly of more connected networks. For comparison, protein interaction databases integrating different resources, such as *mentha*, annotate interactions for approximately 90% of the disease gene-products.

Statistics

At the time of this writing, DISNOR offers more than 3,700 inferred signaling networks that are suggested to play a role in patho-phenotypes. As shown in Figure 2B, the disease class with the highest coverage in DISNOR, using 'first-neighbours query strategy', is represented by 'diseases of cellular proliferation'. For >80% of the malignancies, DISNOR suggests a level 2 graph with an average network size of 20 nodes and 280 edges, covering more than 75% of cancer-associated genes. Conversely, the two less-covered disease classes in DISNOR are 'rare diseases' and 'diseases

A

DisGeNET (v5.0) CURATED	
Total Diseases	13,418
Diseases*	6,323
Genes	7,241
SNPs	46,589
GDAs	23,821

SIGNOR	
Genes/Proteins	3,829
Causal Interactions	18,344
Publications	6,879

*Diseases considered in DISNOR, with at least a SNP (gene) associated

B

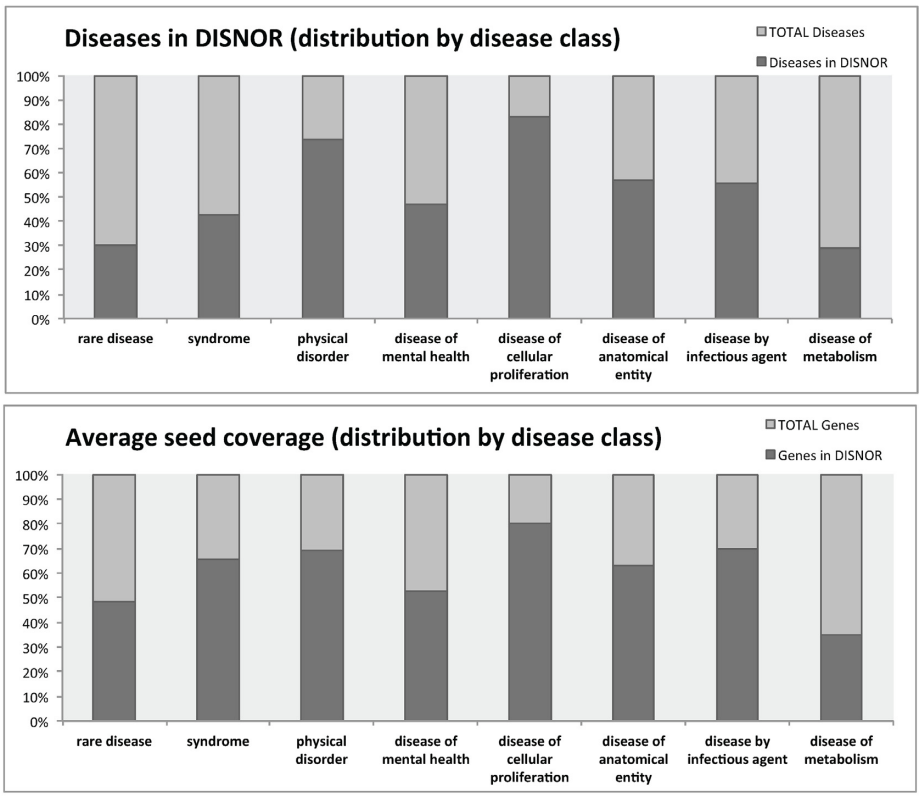

Figure 2. Statistics. The two tables in (**A**) summarize the content of the curated dataset of DisGeNET V5.0 (top) and the content of the SIGNOR knowledge base (bottom). The plots in (**B**) display the percentage of disease-pathway coverage (top) and the distribution of average percentage of 'seed' gene coverage (bottom) in DISNOR across different 'disease-ontology' classes (12) and using 'First Neighbours' querying method.

of metabolism', displaying 30% and 29% of Level 2 pathway coverage, respectively. However, by using the 'All query strategy', the coverage increases to 54% and 47% also for these disease classes (supplementary Figure S1). To date, 1383 'rare diseases' have a Level 2 (or Level 3) causal interaction diagram in DISNOR, and 45% of ∼2500 rare-disease-associated genes are represented.

Web interface

As a result of the curation effort and the integration with DisGeNET annotation, DISNOR offers disease networks, for approximately 60% of diseases (Figure 2B), connecting more than 2,600 gene products. This wealth of data can be accessed at http://DISNOR.uniroma2.it. The homepage summarizes the purpose of the project and offers links to external resources.

As previously discussed, the main source of disease information is DisGeNET, associating each disease to one or more gene variants. Conversely, mutations in one gene can be observed in patients with different diseases. Finally, different SNPs of the same gene might be linked to different sets of diseases. Accordingly, DISNOR data are accessible through three different entry points: 'Disease Browser', 'Search by SNP or GENE ID' and 'Multiple Protein Search'. For each search type, a step-by-step tutorial can be accessed via a link in the DISNOR home page.

Disease browser. The 'Disease Browser' option is the simplest way to query DISNOR and to explore disease networks. It allows browsing through a list of 3,700 automatically generated disease networks. Users can type, in the search box, the name of the disease. The field offers autocomplete functionality in order to facilitate searching by name. Alternatively, a drop-down menu allows the user to select one disease. In addition, the disease list can be filtered by selecting one or more categories (rare disease; syndrome; genetic disease; physical disorder; disease of mental health; disease of cellular proliferation; disease of anatomical entity; disease by infectious agent; or disease of metabolism), which are defined according to the classification of Disease Ontology (12).

As shown in Figure 3C, after each query, DISNOR returns an interactive viewer window presenting the graphical representation of the interactions linking the seed entities. The graphic visualizer offers a schematic and detail-rich representation of the retrieved interactions. Nodes are gene products or other biologically relevant entities such as small molecules, phenotypes etc., and edges represent the causal relationships between them. The graphic visualizer is an intuitive, customizable and dynamic display of the disease pathway. The interpretation of the signaling cascade is facilitated by the adoption of symbols and colour codes for nodes and edges, whose explanation is provided in the graph legend and further discussed by Lo Surdo *et al.* (22).

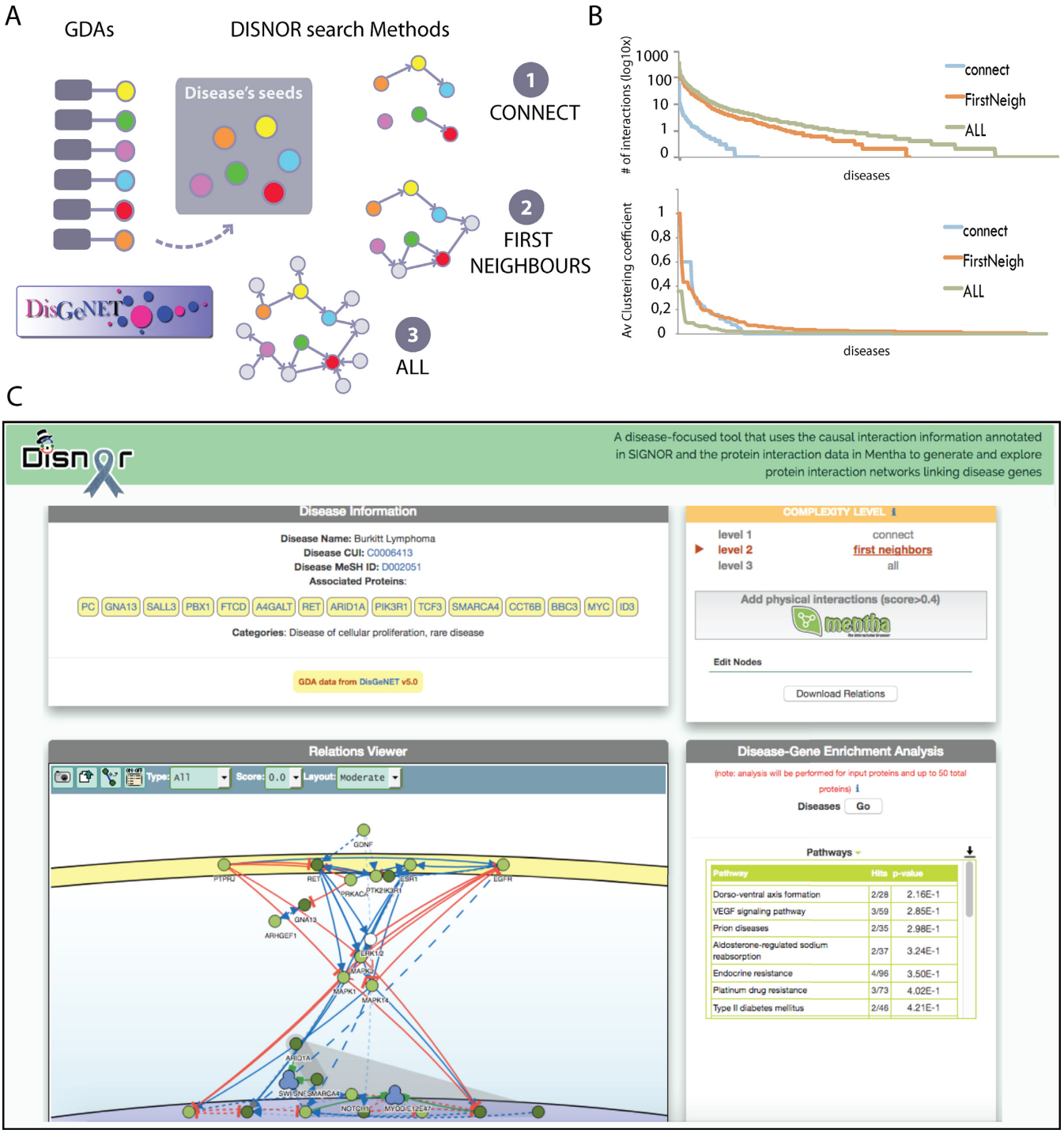

Figure 3. Query Strategies. The image in (**A**) illustrates the three main query strategies used in DISNOR. They all accept as input a list of 'seed' genes extracted from DisGeNET, as described in Figure 1B. *connect*: the system retrieves cause-effect relationships between the 'seed' genes; *first neighbours*: the system (i) retrieves causal interactions between 'seed' genes and their neighbors; (ii) degree-1 nodes are pruned; *all*: the system retrieves any causal interactions between 'seed' genes and their neighbours. The plot in (**B**) describes the network size distribution (top) and the clustering coefficient distribution (bottom) of disease networks in DISNOR, considering the search strategy described in (A): connect (light blue), first neighbors (orange) and all (light green). The screenshot in (**C**) is an example of the result page of a 'Disease Browser search' ('Burkitt Lymphoma'). The web page is organized in four parts: the disease information summary; the interactive graphic viewer; the editable list of pathway 'seeds'; and the box to browse the network at different complexity levels (using strategies described in (A). In the graphics viewer, attributes of nodes and edges are represented with different colors and symbols as described by Lo Surdo *et al.* (22).

For each disease, the result page of DISNOR offers the following additional functionalities:

- The box at the top summarizes a disease's details, such as its name, its classification, the list of 'seed' genes, and links to NCI Metathesaurus (16) and to MeSH terms (17) associated to the pathology.
- The top-right box allows the user to display interactions at different levels of complexity using three different query strategies (described in material and methods and in Figure 3A). By default the visualizer displays Level 2 interactions ('First Neighbors'), since this level maximizes the number of retrievable disease networks while maintaining the connectivity of the subgraph (Figure 3B); if no interaction is retrieved at level 2, level 3 is shown.
- By clicking on 'Add physical interactions from *mentha*', the user can also include in the graph protein–protein interactions from the mentha database (13) (http://mentha.uniroma2.it/) involving any seed entities and filtered by score. Only interactions whose reliability score is >0.4 are shown.
- To the right of the viewer it is also possible to consult and edit the list of the seed entities, by removing or adding nodes.
- On the right, the user can visualize KEGG pathways that are enriched in the 'seed' proteins and their neighbors (19). The number of proteins in the pathway that are in common with the query list is shown together with the corresponding *P*-value.
- Similarly, the system also provides a list of the top 20 enriched disease annotations, thus allowing users to infer diseases that are related to the the query protein list.

Search by SNP or GENE. An alternative way to query DISNOR is by single gene (or SNP). This search strategy allows the user to identify diseases and disease pathways that are associated to the query genetic variant.

The gene name, the UniprotID or the SNP ID (23) of a gene can be typed in the search box. A popup window appears showing a table summarizing the relationships between each gene, its SNPs and the list of associated candidate diseases. The table also provides links to dbSNP (which contains information on the SNP) (23) and to SIGNOR. By selecting one disease, DISNOR returns the result page of the specific disease, as described in the previous section.

Multiple protein search. DISNOR search strategies retrieve causal connections between 'seed' genes. In the 'Disease Brower', the lists of 'seed' genes are automatically generated according to GDAs. Alternatively, DISNOR allows users to select and display a set of causal interactions between a custom entity-list. Users can type the UniprotIDs or the gene names of the query proteins/genes in the search box. The system returns the interactive viewer box presenting the graphical representation of the interactions involving the seed entities. The result page offers all the functionality already described for the result page of the disease browser: (i) graph visualization at different levels of complexity; (ii) integration with the PPI network in the mentha database; (iii) possibility to edit the 'seed' entity-list and (iv) identification of disease and pathway annotations that are enriched in the query protein-list.

In conclusion, DISNOR allows the integration of a list of proteins (or genes) into a network of causal relationships and identifies pathways and pathologies that are related to the query list of proteins (or genes). Network-based approaches based on physical and causal interaction networks have already been established to have a significant role in patient stratification and in tailored treatment decisions (7,8,24). As such, DISNOR might be regarded as a tool to support individualized diagnostic, prognostic and therapeutic approaches.

DISCUSSION

DISNOR connects ~2600 gene-products, whose mutations are associated to a specific pathological condition, through cause-effect relationships extracted from the SIGNOR knowledge base. We briefly discuss here two main applications of DISNOR in network medicine:

i) Assembly and visualization of functional networks, often revealing the molecular mechanisms underlying 3700 pathological phenotypes. The 'Disease Browser' offers the possibility to disclose implicit functional information contained in the list of the causal interactors of disease genes. This eventually extends the list of candidate targets that might be considered in therapeutic strategies.
   Moreover, any resulting patho-network might be employed in graph topology approaches, to identify gene/protein modules or communities allowing for the identification of new candidate disease genes (25).
   Finally, the disease gene enrichment analysis tool permits identification of similar diseases. Disease similarity has relevance in the understanding of the molecular mechanisms underlying a pathological condition, in the identification of appropriate therapies and in comorbidity studies (26). Common approaches to measuring disease similarity are mostly based on the comparison of phenotypic features or symptoms (27), of shared disease genes (28) or, alternatively, of the network of PPIs (26). Here, we propose a new strategy, in which similarity between two diseases is assessed considering not only the disease-associated genes, but also their first neighbours in a causal interaction network.
ii) Analysis of the functional and disease connections between genetic variants. The 'multi-protein search' and the 'search by SNP or GENE' enable assembling the network of causal interactions linking one or more user-defined genetic variants and displaying them in an intuitive graph representation. The possibility to link a personalized set of gene variants (for instance from a patient genotype) through cause-effect relationships may allow an understanding of the molecular events occurring in a individualized context and, as a consequence, offers a rationalized framework for the development of tailored treatments.

The present version of DISNOR is the result of our latest curation effort aimed at increasing the coverage of disease

genes in the SIGNOR database. This is far from being exhaustive and fully accurate. Our curation team will continue this effort, but we encourage our disease-expert colleagues who value this resource to come up with suggestions and criticisms and possibly help, in their domain of expertise, to increase the coverage and accuracy of DISNOR.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

FUNDING

DEPTH project of the European Research Council [32274]; Italian Association for Cancer Research [triennial fellowship Starwood Hotels & Resorts id. 18137]; ELIXIR-ITA, Italian Ministry of Education (to L.L.). Funding for open access charge: DEPTH project of the European Research Council [32274].
Conflict of interest statement. None declared.

REFERENCES

1. 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.
2. Hamosh,A., Scott,A.F., Amberger,J.S., Bocchini,C.A. and McKusick,V.A. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–D517.
3. Biesecker,L.G. (2012) Opportunities and challenges for the integration of massively parallel genomic sequencing into clinical practice: lessons from the ClinSeq project. *Genet. Med.*, **14**, 393–398.
4. Yang,H., Robinson,P.N. and Wang,K. (2015) Phenolyzer: phenotype-based prioritization of candidate genes for human diseases. *Nat. Methods*, **12**, 841–843.
5. Singleton,M.V., Guthery,S.L., Voelkerding,K.V., Chen,K., Kennedy,B., Margraf,R.L., Durtschi,J., Eilbeck,K., Reese,M.G., Jorde,L.B. *et al.* (2014) Phevor combines multiple biomedical ontologies for accurate identification of disease-causing alleles in single individuals and small nuclear families. *Am. J. Hum. Genet.*, **94**, 599–610.
6. Kholodenko,B., Yaffe,M.B. and Kolch,W. (2012) Computational approaches for analyzing information flow in biological networks. *Sci. Signal.*, **5**, re1.
7. Shi,M., Beauchamp,R.D. and Zhang,B. (2012) A network-based gene expression signature informs prognosis and treatment for colorectal cancer patients. *PLoS One*, **7**, e41292.
8. Lee,M.J., Ye,A.S., Gardino,A.K., Heijink,A.M., Sorger,P.K., MacBeath,G. and Yaffe,M.B. (2012) Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell*, **149**, 780–794.
9. Perfetto,L., Briganti,L., Calderone,A., Perpetuini,A.C., Iannuccelli,M., Langone,F., Licata,L., Marinkovic,M., Mattioni,A., Pavlidou,T. *et al.* (2016) SIGNOR: a database of causal relationships between biological entities. *Nucleic Acids Res.*, **44**, D548–D554.
10. Turei,D., Korcsmaros,T. and Saez-Rodriguez,J. (2016) OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods*, **13**, 966–967.
11. Bodenreider,O. (2004) The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res.*, **32**, D267–D270.
12. Schriml,L.M. and Mitraka,E. (2015) The Disease Ontology: fostering interoperability between biological and clinical human disease-related data. *Mamm. Genome*, **26**, 584–589.
13. Calderone,A., Castagnoli,L. and Cesareni,G. (2013) mentha: a resource for browsing integrated protein-interaction networks. *Nat. Methods*, **10**, 690–691.
14. Calderone,A., Formenti,M., Aprea,F., Papa,M., Alberghina,L., Colangelo,A.M. and Bertolazzi,P. (2016) Comparing Alzheimer's and Parkinson's diseases networks using graph communities structure. *BMC Syst. Biol.*, **10**, 25.
15. Fabregat,A., Sidiropoulos,K., Garapati,P., Gillespie,M., Hausmann,K., Haw,R., Jassal,B., Jupe,S., Korninger,F., McKay,S. *et al.* (2016) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **44**, D481–D487.
16. Tuttle,M.S., Olson,N.E., Campbell,K.E., Sherertz,D.D., Nelson,S.J. and Cole,W.G. (1994) Formal properties of the metathesaurus. *Proc. Symp. Comput. Applic. Med. Care*, 145–149.
17. Rogers,F.B. (1963) Medical subject headings. *Bull. Med. Libr. Assoc.*, **51**, 114–116.
18. UniProt,C. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
19. Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
20. Wang,X., Gulbahce,N. and Yu,H. (2011) Network-based methods for human disease gene prediction. *Brief. Funct. Genomics*, **10**, 280–293.
21. Menche,J., Guney,E., Sharma,A., Branigan,P.J., Loza,M.J., Baribaud,F., Dobrin,R. and Barabasi,A.L. (2017) Integrating personalized gene expression profiles into predictive disease-associated gene pools. *NPJ Syst. Biol. Applic.*, **3**, 10.
22. Lo Surdo,P., Calderone,A., Cesareni,G. and Perfetto,L. (2017) SIGNOR: a database of causal relationships between biological entities—a short guide to searching and browsing. *Curr. Protoc. Bioinformatics*, **58**, doi:10.1002/cpbi.28.
23. 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.
24. Barabasi,A.L., Gulbahce,N. and Loscalzo,J. (2011) Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.*, **12**, 56–68.
25. Oti,M. and Brunner,H.G. (2007) The modular nature of genetic diseases. *Clin. Genet.*, **71**, 1–11.
26. Menche,J., Sharma,A., Kitsak,M., Ghiassian,S.D., Vidal,M., Loscalzo,J. and Barabasi,A.L. (2015) Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science*, **347**, 1257601.
27. Zhou,X., Menche,J., Barabasi,A.L. and Sharma,A. (2014) Human symptoms-disease network. *Nat. Commun.*, **5**, 4212.
28. Goh,K.I., Cusick,M.E., Valle,D., Childs,B., Vidal,M. and Barabasi,A.L. (2007) The human disease network. *PNAS*, **104**, 8685–8690.