

DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene–disease networks

Anna Bauer-Mehren, Michael Rautschka, Ferran Sanz and Laura I. Furlong*

Research Programme on Biomedical Informatics (GRIB) IMIM, DCEXS, Universitat Pompeu Fabra, C/Dr. Aiguader 88, 08003 Barcelona, Spain

Associate Editor: Olga Troyanskaya

ABSTRACT

Summary: DisGeNET is a plugin for Cytoscape to query and analyze human gene–disease networks. DisGeNET allows user-friendly access to a new gene–disease database that we have developed by integrating data from several public sources. DisGeNET permits queries restricted to (i) the original data source, (ii) the association type, (iii) the disease class or (iv) specific gene(s)/disease(s). It represents gene–disease associations in terms of bipartite graphs and provides gene centric and disease centric views of the data. It assists the user in the interpretation and exploration of the genetic basis of human diseases by a variety of built-in functions. Moreover, DisGeNET permits multicolouring of nodes (genes/diseases) according to standard disease classification for expedient visualization.

Availability: DisGeNET is compatible with Cytoscape 2.6.3 and 2.7.0, please visit <http://ibi.imim.es/DisGeNET/DisGeNETweb.html> for installation guide, user tutorial and download

Contact: lfurlong@imim.es

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on July 20, 2010; revised on August 19, 2010; accepted on September 16, 2010

1 INTRODUCTION

One of the most challenging problems in biomedical research is to understand the underlying mechanisms of human diseases. Great effort has been spent on determining genes associated to diseases (Botstein and Risch, 2003; Kann, 2010). However, there is more and more evidence that most human diseases cannot be attributed to single genes but arise due to complex interactions between multiple genetic variants and environmental risk factors (Hirschhorn and Daly, 2005). Several databases have been developed storing associations between genes and diseases such as Online Mendelian Inheritance in Man (OMIM; Hamosh *et al.*, 2005). Each of these databases focuses on different aspects of phenotype to genotype relationships. For instance, PharmGKB is specialized on how genetic variation is related to drug response (Altman, 2007), whereas the toxicogenomics database CTD stores information about the effect of environmental chemicals on human health (Mattingly *et al.*, 2006). Hence, integration of different databases is needed to allow a comprehensive view of the state of the art knowledge within this research field. It is widely established in bioinformatics to represent associations between biomedical entities as networks and to analyze their topology to get a global understanding of underlying

relationships (Butts, 2009; Goh *et al.*, 2007; Yildirim *et al.*, 2007). Cytoscape is a widely used Java-based, open-source software for networks visualization and analysis (Shannon *et al.*, 2003). The Cytoscape framework is extendable through the implementation of plugins. Up to now, a vast variety of plugins has been developed ranging from advanced network analysis tools to webservices. In the following, we present DisGeNET, a new Cytoscape plugin to query, integrate and visualize networks of human gene–disease associations.

2 OVERVIEW

2.1 The human gene–disease database

We compiled a comprehensive database of human gene–disease associations by integrating data from various expert curated databases and text-mining derived associations including Mendelian, complex and environmental diseases (Bauer-Mehren *et al.*, submitted for publication). We created bipartite graphs called OMIM, UNIPROT, PHARMGKB, CTD, CURATED (combining data from the curated databases), LHGDN (from the text-mining data only) and ALL (including all available gene–disease associations). Moreover, we calculated two network projections for each bipartite graph in order to generate a disease and gene centric data representations. These projections allow an enhanced view on the genetic basis of complex diseases. We furthermore classified all diseases into one of 26 possible disease classes following the MeSH hierarchy (Bauer-Mehren *et al.*, submitted for publication).

2.2 Gene–disease networks within Cytoscape

The gene–disease networks are bipartite graphs with two types of nodes (gene and disease) (Goh *et al.*, 2007; Newman, 2003). Gene and disease nodes are connected through edges if the according gene–disease association is covered in the gene–disease database. We allow displaying multiple edges between nodes, each representing a unique association found in the original data sources. Moreover, we colour the edges according to the association type following our gene–disease association ontology (Bauer-Mehren *et al.*, submitted for publication). The disease and gene projection networks are monopartite graphs only containing either gene or disease nodes. Nodes are connected through edges if the two genes (diseases) share a disease (gene) in the bipartite gene–disease network. Thus, this representation allows studying diseases with similar genetic origin or genes associated to similar diseases. DisGeNET can be started from the plugins menu in Cytoscape. The main panel consists of three tabs, one for the gene–disease networks called ‘Gene Disease Network’ and one for each projection, namely ‘Disease Projections’ and ‘Gene Projections’. The ‘Gene Disease

*To whom correspondence should be addressed.



Fig. 1. (A) Cytoscape screenshot of DisGeNET. The diseases Alzheimer Disease and Myocardial Infarction and their shared genes are displayed (in yellow). (B) The same network is shown with nodes coloured according to the disease classes of the nodes.

Network' tab contains three drop-down menus allowing a restriction to (i) source, (ii) association type and (iii) disease class. The two projection panels only contain two drop-down menu options to restrict the query to source and disease class. DisGeNET incorporates an advanced search function for each of the three network types. The user can search for a gene or a disease of interest and even for any set of diseases or genes by using the wild card symbol (asterisk). The search box can be either used to create new networks or to select nodes of already generated networks. DisGeNET makes use of the Cytoscape VizMapper to create visual styles for the networks. Gene nodes are coloured in blue and disease nodes in magenta (Fig. 1A). The node size increases with increasing number of associated diseases, respectively genes. Edges are coloured corresponding to the association type. Moreover, disease and gene nodes can be coloured according to the disease class by using the 'Colour nodes with disease class' button. Since it is possible to have diseases and genes assigned to more than one disease class, multicolour pie charts can be overlaid onto (and removed from) nodes (Fig. 1B).

2.3 Use cases

Some exemplary use cases showing the utility of DisGeNET are available at <http://ibi.imim.es/DisGeNET/DisGeNETweb.html#UserGuide> and in the Supplementary Material section.

They address problems such as: (i) which are the genes annotated to breast neoplasm in expert-curated databases?; (ii) do comorbidities of Alzheimer disease and myocardial infarction observed in patients reflect in a common genetic origin?; or (iii) which are the diseases that are associated to post-translational modifications such as phosphorylation?

3 CONCLUSION

DisGeNET is a coherent tool for easy analysis and interpretation of human gene–disease networks. It allows user-friendly access to

a comprehensive database comprising gene–disease associations for Mendelian, complex and environmental diseases. DisGeNET displays gene–disease association networks as bipartite graphs and provides gene centric and disease centric views of the data. It assists the interpretation and exploration of human diseases with respect to their genetic origin. Diverse options for generating subnetworks, as well as an advanced search tool, facilitate not only the analysis of single diseases but also the study of sets of diseases or certain disease classes specified through their associated genes. Herein, the multicolouring of gene and disease nodes offers a convenient visualization of disease classifications in the networks. We plan regular updates of the underlying gene–disease association database as well as the integration of further data sources.

Funding: This work was supported by the European Commission [EU-ADR, ICT-215847]; Innovative Medicines Initiative [eTOX, 115002]; and the AGAUR [to A.B.M.]. The GRIB is a node of the Spanish National Institute of Bioinformatics and the COMBIOMED network.

Conflict of Interest: none declared.

REFERENCES

- Altman, R.B. (2007) PharmGKB: a logical home for knowledge relating genotype to drug response phenotype. *Nat. Genet.*, **39**, 426.
- Botstein, D. and Risch, N. (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.*, **33**, 228–237.
- Butts, C.T. (2009) Revisiting the foundations of network analysis. *Science*, **325**, 414–416.
- Goh, K.-I. *et al.* (2007) The human disease network. *Proc. Natl Acad. Sci. USA*, **104**, 8685–8690.
- Hamosh, A. *et al.* (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–D517.
- Hirschhorn, J.N. and Daly, M.J. (2005) Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.*, **6**, 95–108.

- Kann,M.G. (2010) Advances in translational bioinformatics: computational approaches for the hunting of disease genes. *Brief. Bioinform.*, **11**, 96–110.
- Mattingly,C.J. et al. (2006) The comparative toxicogenomics database: a cross-species resource for building chemical-gene interaction networks. *Toxicol. Sci.*, **92**, 587–595.
- Newman,M.E.J. (2003) The structure and function of complex networks. *SIAM Rev.*, **45**, 167–256.
- Shannon,P. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
- Yildirim,M.A. et al. (2007) Drug-target network. *Nat. Biotech.*, **25**, 1119–1126.