DisGeNET: a discovery platform to support translational research and drug discovery

Janet Piñero, Núria Queralt-Rosinach, Àlex Bravo, Ferran Sanz and Laura I. Furlong

Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain

In the last two decades, biomedical research has experienced a revolution driven by technological advances that have resulted in the dramatic growth of the volume and variety of data. The fragmented nature of this process is one of the major current bottlenecks in the analysis and extraction of knowledge from this sea of information. In addition, this information is represented and annotated using different data models, vocabularies and standards, which hampers its integration and poses extra hurdles to its interpretation. This is the case of our current knowledge on disease genetics, scattered across several resources and human variation catalogs each focusing on one aspect of the gene-disease relationship. The creation of tools that collect and homogeneously annotate our current knowledge on the genetic basis of diseases is key to the development of translational bioinformatics and drug discovery.

To target this need we have developed DisGeNET[a] [1], a discovery platform that collects and integrates the available information on gene-disease associations (GDAs), covering the whole landscape of human diseases, and using standards for their annotation and representation. DisGeNET possesses several unique features that make it a platform of choice for biomedical translational researchers and drug discovery applications. First, it contains one of the largest collections of GDAs arising from both expert-curated knowledge, and information extracted from the scientific literature using NLP-based text mining techniques, with special attention paid to the explicit provenance of the association. Second, mapping to different biomedical vocabularies annotating diseases are provided for gene-disease associations, thus facilitating the work of clinical and biomedical researchers. Third, it features a score developed to rate the confidence of each GDA. Finally, several ways to access the data are available, to serve better the purposes of different types of users.

DisGeNET is also available as a Linked Open Data (LOD) data set through DisGeNET-RDF. DisGeNET-RDF is interlinked to other biomedical databases to support the development of bioinformatics tools for translational research through evidence-based exploitation of a rich and fully interconnected LOD cloud. DisGeNET-RDF has been developed in the Open PHACTS project context and implemented in the Open PHACTS Discovery Platform. This allows the formulation of sophisticated queries that need the interrogation of different resources to be answered. More concretely, our SPARQL endpoint allows query federation to interrogate DisGeNET with several linked open data

[a] http://www.disgenet.org

resources with a single query. These include data on gene expression, drugs and other chemicals, biological pathways and networks, kinetic models, to just mention some examples.

In this poster we will illustrate different ways to use DisGeNET to answer complex scientific questions in translational research and drug discovery and development using the SPARQL endpoint and Open PHACTS API calls.

[1] Piñero, J., Queralt-Rosinach, N., Bravo, A., Deu-Pons, J., Bauer-Mehren, A., Baron, M., Sanz, F., Furlong, L. I. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database, 2015(0), bav028–bav028. doi:10.1093/database/bav028

Funding

The research leading to these results has received support from Instituto de Salud Carlos III-Fondo Europeo de Desarollo Regional (PI13/00082 and CP10/00524), the Innovative Medicines Initiative Joint Undertaking under grants agreements n° 115002 (eTOX) and nº 115191 (Open PHACTS)], resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution. This project has also received funding from the European Union's Horizon 2020 research and innovation programme 2014-2020 under Grant Agreement No 634143. Laura I. Furlong received support from Instituto de Salud Carlos III Fondo Europeo de Desarollo Regional (CP10/00524). The Research Programme on Biomedical Informatics (GRIB) is a node of the Spanish National Institute of Bioinformatics (INB).

[a] http://www.disgenet.org