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

Janet Piñero, Nuria Queralt-Rosinaich, Álex Bravo, Ferran Sanz and Laura I. Furlong

Integrative Biomedical Informatics Group, Research Programme on Biomedical Informatics, Hospital del Mar Medical Research Institute, Pompeu Fabra University

Motivation: Better understanding of human gene component and disease mechanisms for translational research and drug discovery and development.

Challenge: One of the major current bottlenecks for knowledge on the genetic component of diseases is that the information is fragmented. The vast amount of biomedical information about genotype-phenotype relations is distributed in several databases, represented and annotated using different data models, vocabularies and standards, and it is dynamic and technology-specific, which hampers their access, integration, analysis, and interpretation.

Approach: DisGeNET Discovery Platform collects and integrates the available information on gene-disease associations (GDAs), covering the whole spectrum of human diseases, and using standards for their annotation and representation.

Implementation: The platform is composed of a knowledge base and a set of tools for domain analysis and interpretation.

KNOWLEDGE BASE

- LARGE-SCALE EXTRACTION AND INTEGRATION

- NORMALIZATION
  - NCBi Gene ID
  - UMLS CUIs

HARMONIZATION

DisGeNET association type ontology

TOOLS FOR EXPLORATION AND ANALYSIS

DIFFERENT USER PROFILES

- REFLEXIVITY
  - Meta: data-item description
  - dataset description
- Transparency
  - Advanced analysis
- Validation
  - Higher speed
  - Sharer results
- Embed in workflows

ACCESS

- Web: http://www.disgenet.org/
- RDF: http://dev.disgenet.org/

INTEROPERABILITY

- COMMON IDs and ONTOLOGIES
  - GENE: NCBi Gene ID
  - DISEASE: PANTHER Classification
  - Vocabularies: UMLS CUIs, SNOMED CT

- SYNTACTIC
  - RDF: RDF

- SEMANTIC
  - 11 common ontologies

DATA

- Data providers
  - Disease annotation in the Open PHACTS Discovery Platform

- OMIM included
  - > 10,000 000 of triples

DISCOVERY EXPLORER

- API

References


Acknowledgements

The authors thank the Open PHACTS partners, Michel Dumontier and the OpenLink staff for their input, collaboration and help. Funding: the work is supported through EU FP709619 (Open PHACTS), an EU FP7 project, and the HBP (Human Brain Project) under grant agreement 604102 (HBP). The work is performed within the frame of the Semantic Web Research Institute funded by the EU FP7 under grant agreement 248910 (Open PHACTS), and at University of Cergy-Pontoise under grant agreement 2008-2010 under grant agreements nos. 464491 (Medicinfo) and no. 463389 (Sirte-Genacom). The Research Programme on Biomedical Informatics is supported by a grant from the Spanish government to the University of Cergy-Pontoise and funded by the European Regional Development Fund. The project is supported by the Spanish Ministries of Education, Culture and Sport and Health, and the EU European Social Fund.