disgenet2r: An R package to explore the molecular underpinnings of human diseases

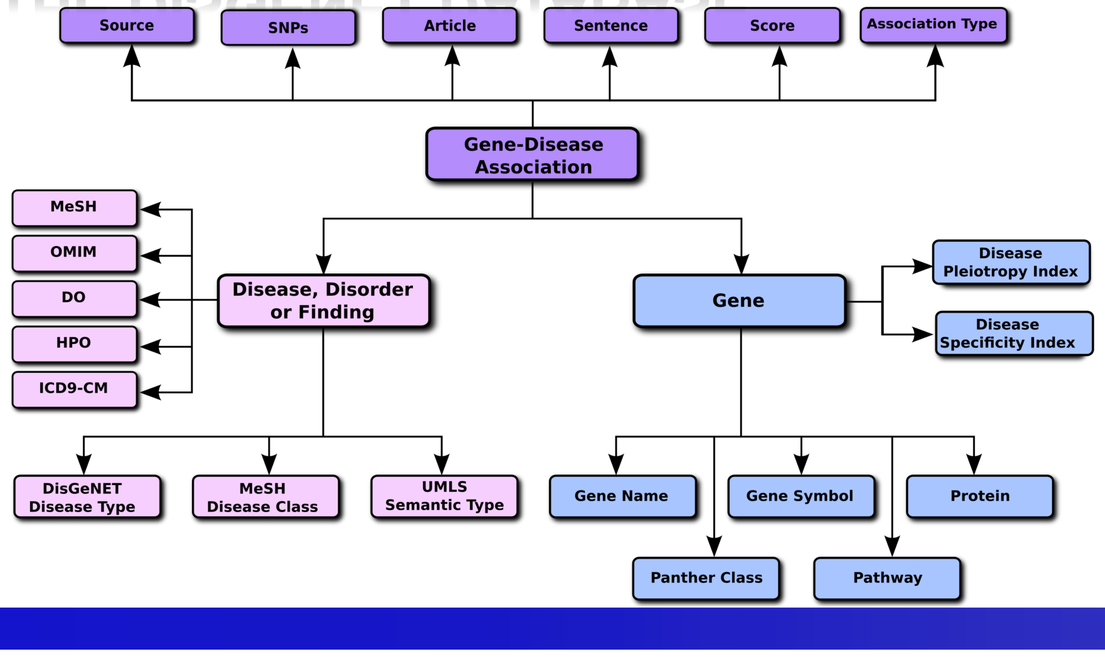
Alba Gutiérrez-Sacristán, **Janet Piñero**, Núria Queralt-Rosinach, Laura I. Furlong

Research Programme on Biomedical Informatics (GRIB) , Hospital del Mar Medical Research Institute (IMIM), Pompeu Fabra University (UPF)

ABSTRACT

DisGeNET [1] is a discovery platform designed to answer questions concerning the molecular mechanisms underlying human diseases. DisGeNET data can be explored using a suite of tools which includes a web interface, a Cytoscape plugin [2], and a SPARQL endpoint [3]. In this contribution, we present **disgenet2r**, an R package for exploring DisGeNET. **disgenet2r** contains a variety of functions for leveraging DisGeNET using the powerful visualization and statistical capabilities of the R environment. disgenet2r is specially designed to harness the large amount of information contained in DisGeNET, facilitating its analysis and interpretation. The package offers different types of visualization of DisGeNET data, such as heatmaps and networks, and it is especially well suited to explore the genetic basis of diseases as well as disease comorbidity. Furthermore, to allow answering more sophisticated research questions that need the interrogation of multiple, heterogeneous and disparate resources, the disgenet2r package permits benefiting of the potential of the Semantic Web technologies, without the need of special expertise in this area. This is achieved through a set of functions that connect DisGeNET with other resources present in the Linked Open Data, covering different information such as gene expression, drug activity, and biological pathways, just to mention a few examples. The disgenet2r package also expedites the integration of DisGeNET data with other R/Bioconductor packages, and allows the construction of complex bioinformatic workflows. We illustrate the functionality of disgenet2r through several use cases to show how the package can be applied to aid particular user's needs. The source code and documentation of disgenet2r package are available at **https://bitbucket.org/albags/disgenet2r**.

THE DISGENET DATABASE



Summary of DisGeNET statistics

Data Set	Genes	Diseases/Phenotypes	SNPs	GDAs
CURATED	9362	7607	35094	32834
PREDICTED	2743	2064	0	10264
ALL	17381	15093	46589	429036

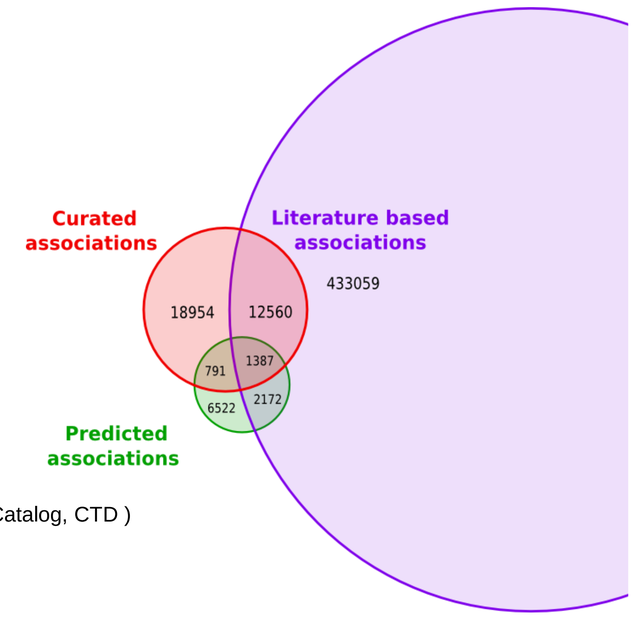
The DisGeNET score

$$S = \sum_{i=1}^{5} C_i + \sum_{j=1}^{2} M_j + \sum_{k=1}^{3} L_k$$

C_i ⇨ Curated Sources (UniProt, ClinVar, Orphanet, the GWAS Catalog, CTD)

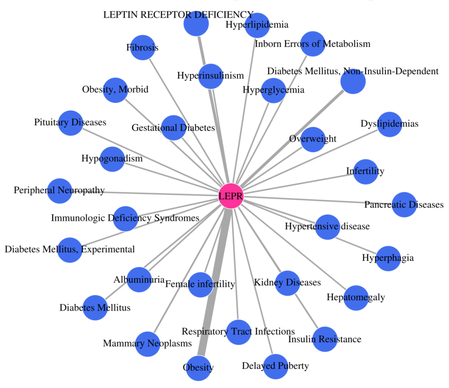
M_j ⇨ Predicted Sources (MGD, RGD, CTD(animal models))

L_k ⇨ Text Mining Data (GAD, LHGDN, BeFree)

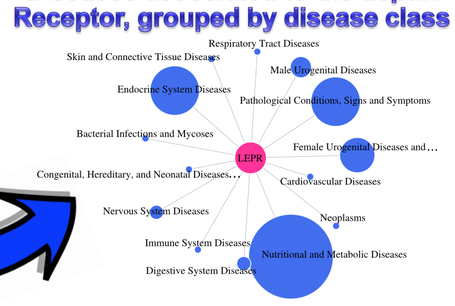


THE GENES

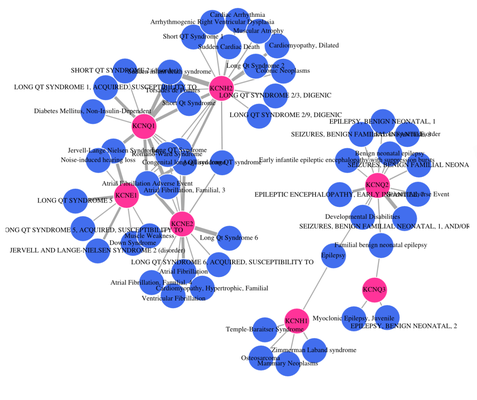
What are the diseases associated to the Leptin Receptor?



Diseases associated to the Leptin Receptor, grouped by disease class



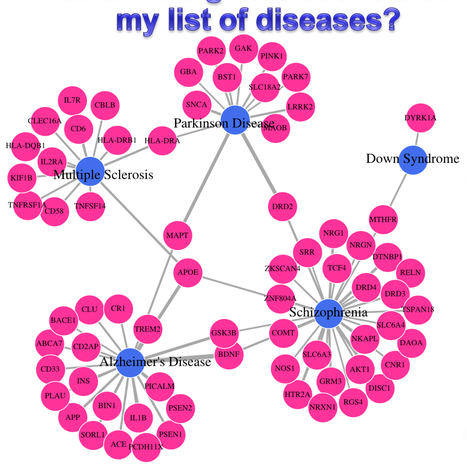
What are the diseases associated to the my list of genes?



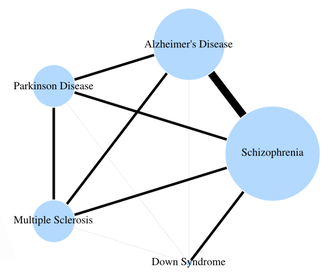
Gene-Disease Association Heatmap

THE DISEASES

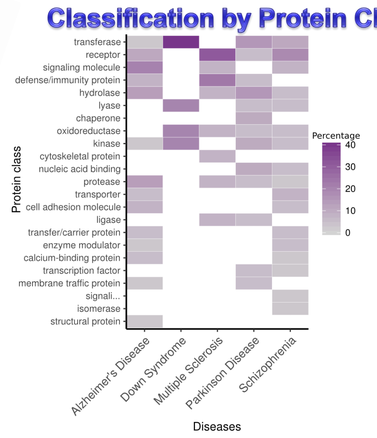
What are the genes associated to my list of diseases?



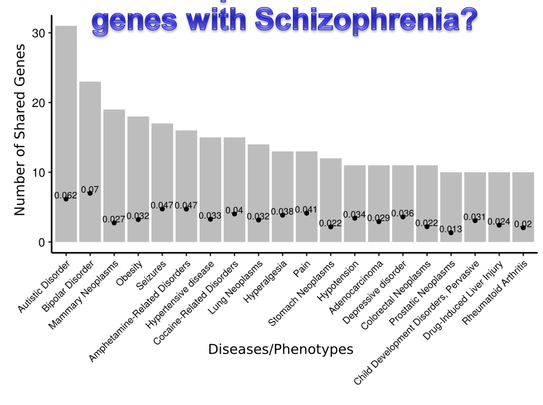
Disease Relationships Network



Classification by Protein Class

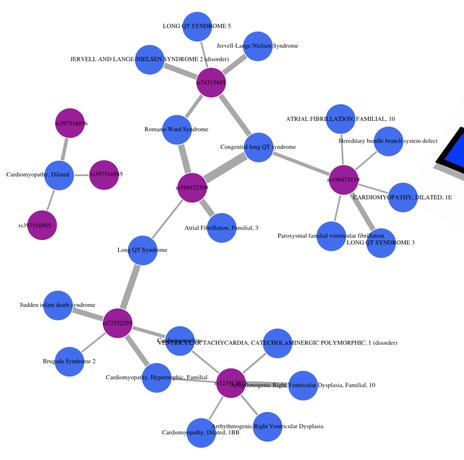


What are the top 20 diseases that share genes with Schizophrenia?

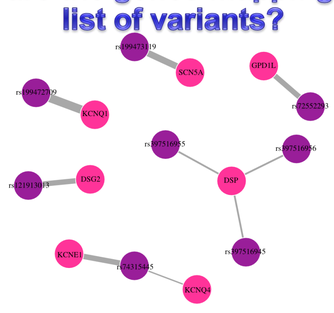


THE VARIANTS

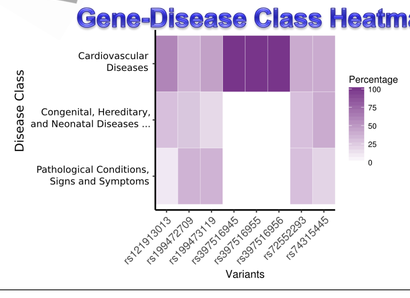
What are the diseases associated to my list of variants?



What are the genes mapping to my list of variants?



Gene-Disease Class Heatmap

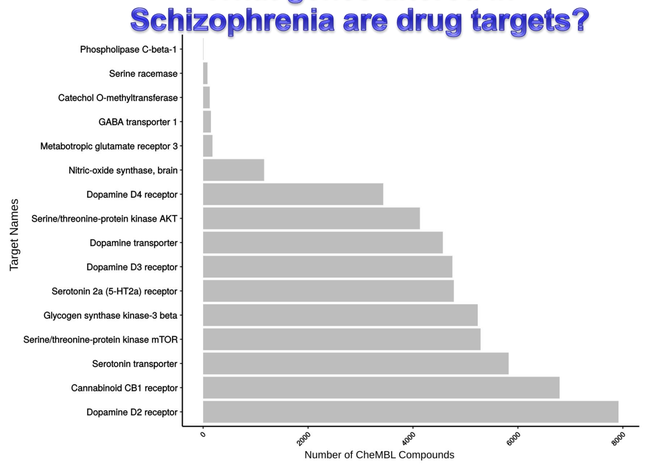


LINKING TO OTHER RESOURCES

What are the pathways associated to Familial dilated cardiomyopathy?

Pathway Name (Wikipathway)	N. of Genes
Striated Muscle Contraction	7
Arrhythmogenic Right Ventricular Cardiomyopathy	5
Adipogenesis	3
Calcium Regulation in the Cardiac Cell	2
MicroRNAs in cardiomyocyte hypertrophy	2
Physiological and Pathological Hypertrophy of the Heart	2
SIDS Susceptibility Pathways	2
ACE Inhibitor Pathway	1
BDNF signaling pathway	1
Cardiac Hypertrophic Response	1
EGF/EGFR Signaling Pathway	1
GPCRs, Class A Rhodopsin-like	1
Monoamine GPCRs	1
Heart Development	1
TFs Regulate miRNAs related to cardiac hypertrophy	1
Hypertrophy Model	1
Hedgehog Signaling Pathway	1
TGF beta Signaling Pathway	1
Regulation of Actin Cytoskeleton	1
Serotonin Receptor 2 and ELK-SRF/GATA4 signaling	1

What genes altered in Schizophrenia are drug targets?



Funding:
The research leading to these results has received support from Instituto de Salud Carlos III-Fondo Europeo de Desarrollo Regional (CP10/00524 and PI13/00082), 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, and the European Union Horizon 2020 Programme 2014–2020 (MedBioinformatics No. 634143 and Elixir-Excelerate No. 676559. L.I.F received support from Instituto de Salud Carlos III Fondo Europeo de Desarrollo Regional (CP10/00524). The Research Programme on Biomedical Informatics (GRIB) is a node of the Spanish National Institute of Bioinformatics (INB).

References:
[1] J. Piñero, *et al.* Database (2015) 2015:bav028–bav028.
[2] A. Bauer-Mehren *et al.* Bioinformatics 26 (2010) 2924–6.
[3] N. Queralt-Rosinach *et al.* Bioinformatics (2016) doi: 10.1093/bioinformatics/btw214.