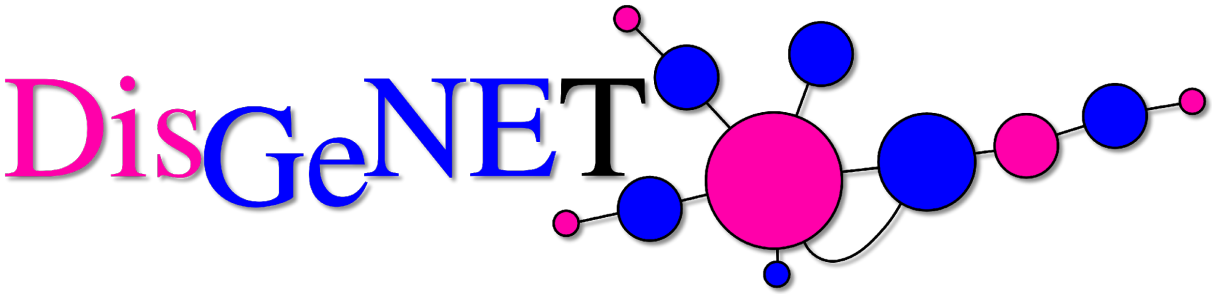

DisGeNET Cytoscape App

IBI Lab
[2021]

USER GUIDE

Integrative Biomedical Informatics Laboratory
Research Group on Biomedical Informatics (GRIB)
IMIM-Hospital del Mar Medical Research Institute
Universitat Pompeu Fabra (UPF)
PRBB Dr. Aiguader 88, E-08003 Barcelona (SPAIN)

The DisGeNET app is distributed under the GNU GPL 3.0 license. More details about the GNU General Public License 3.0 is available here.

DisGeNET Cytoscape App User Guide (version 7.0.x)

1. Installation guide

a. Download and install the DisGeNET Cytoscape App

Equipment: a personal computer with Internet access and an Internet browser.

Operating System: The DisGeNET App and Cytoscape are supported on Windows (Windows 7, Windows 8, Windows 10), Mac and Linux.

Java standard edition: Version 1.8 or higher is required (available at http://www.java.com/).

Cytoscape version DisGeNET is compatible with the Cytoscape 3.x versions.
We recommend Cytoscape versions (3.6.x) or later. The steps for downloading and installing the latest version of Cytoscape are described at http://www.cytoscape.org/.

App version: 7.x

The DisGeNET app needs to be installed from the Cytoscape App Store.

- Go to Apps in the Cytoscape menu
- Click on App Manager
- type *disgenet* on the search box
- click on the result, and then click install (Figure 1)
-

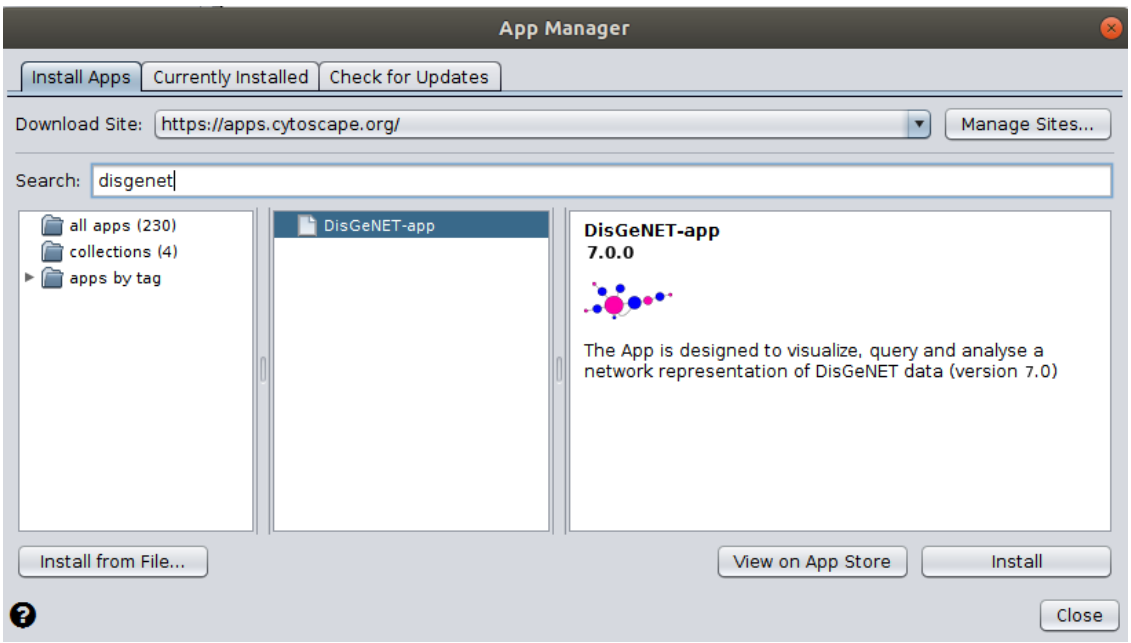

Figure 1: Installing the DisGeNET app from the Cytoscape App Store

Then, go again to Apps in the Cytoscape menu, and click on DisGeNET-> Start DisGeNET (Figure 2).

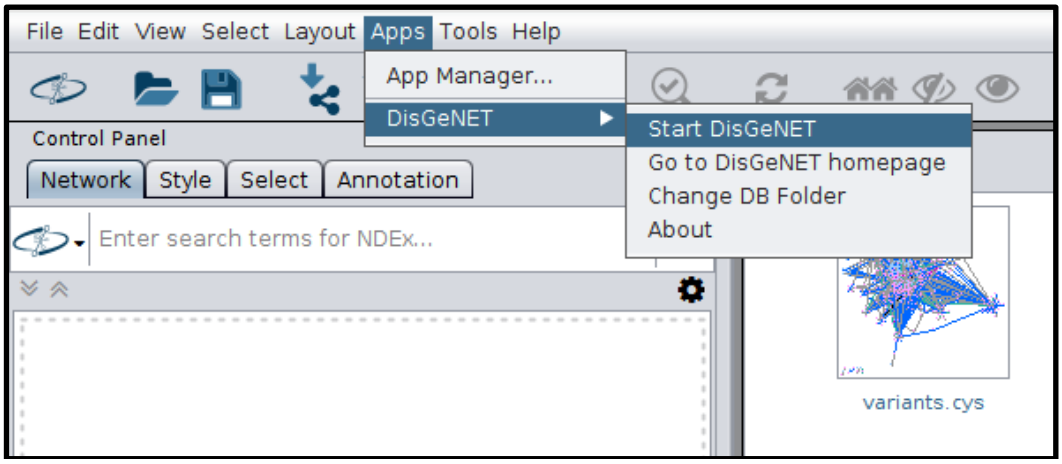

Figure 2: Starting the DisGeNET app

The first time that the DisGeNET app is started, it will ask for the directory of the database file. If the database file does not exist in the directory, it will proceed to automatically download it. Choose a directory where the database will be downloaded and unpacked (disgenet_2020.db ~1.3 Gb). For example, Downloads (Figure 3). For a detailed description of the database, see section 4.

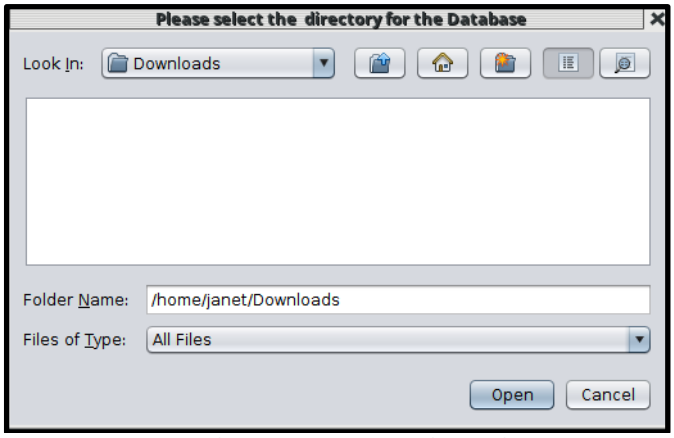

Figure 3: Configure DisGeNET app for the first run

A message of "Database file does not exist. Going to download it" will appear (Figure 4).

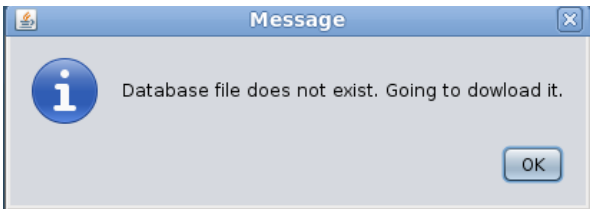

Figure 4: Configure DisGeNET database for the first run

The download might take several minutes. Please, be patient. When it finishes, a new message will appear “Database downloaded correctly.” The app will start afterwards. This also takes a couple of minutes. The app is ready to be used when the DisGeNET Control Panel appears (Figure 5).

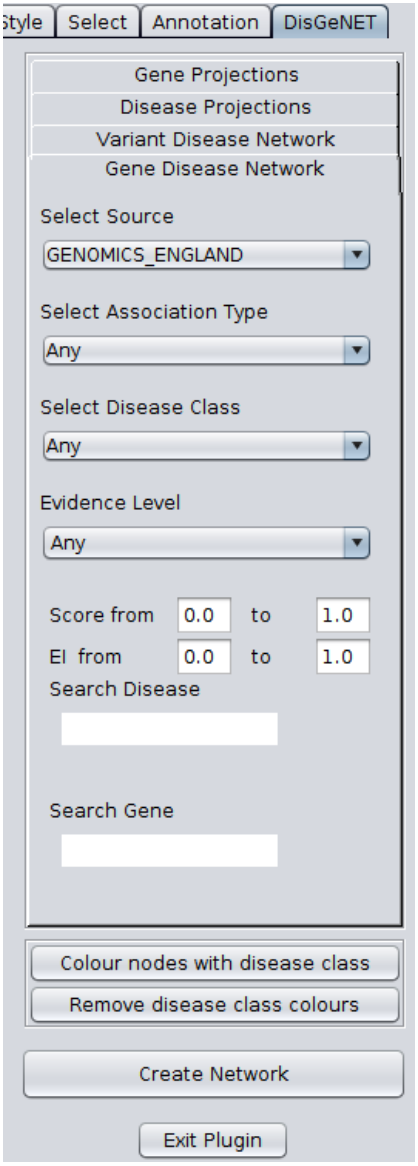

Figure 5: The DisGeNET app control panel

2. Brief description of DisGeNET

DisGeNET is a discovery platform that integrates human gene and variant-disease associations from various expert curated databases and the scientific literature, and includes Mendelian, rare, complex and environmental diseases, as well as abnormal phenotypes and traits (1–3).

a. Original data sources

In DisGeNET, the GDAs are grouped according to their type and level of curation: CURATED (containing gene-disease associations from human expert curated data sources), ANIMAL MODELS (containing gene-disease associations from animal model repositories), INFERRED (containing gene-disease associations from HPO, GWASDB and GWASCAT), and ALL (including CURATED, ANIMAL MODELS, INFERRED, and data derived from text mining the biomedical literature). DisGeNET VDAs are grouped according to their type and level of curation: CURATED (containing variant-disease associations from human expert curated data sources), and ALL (including CURATED data and data derived from text mining the biomedical literature). For the up-to-date list and description of data sources available in DisGeNET, please visit the DisGeNET Discovery Platform Website at http://disgenet.org/dbinfo, section "Original Data Sources".

b. Representing DisGeNET data using networks

Gene-disease associations (GDAs), and variant-disease associations (VDAs) are collected from several sources. The source databases use different vocabularies. In order to merge all GDA and VDAs and to present them in one comprehensive gene-disease, or variant-disease network, we (i) mapped gene identifiers to NCBI Entrez Gene identifiers if necessary, (ii) mapped disease vocabulary terms to the Unified Medical Language System® (UMLS®) Concept Unique Identifiers (CUIs), and (iii) integrated associations through the DisGeNET gene-disease association ontology (see section 2.4).
The data contained in DisGeNET is represented using bipartite graphs. The DisGeNET bipartite graph has two types of vertices (genes or variants and diseases) and the edges connect the vertices of different types (e.g. a gene with a disease). These bipartite graphs are multigraphs in which two vertices might be connected by more than one edge. These multiple edges represent the multiple evidences reporting the GDA or VDA.

c. Vocabulary mapping

For the up-to-date description of the disease and gene vocabulary mappings used

in DisGeNET please visit the DisGeNET Discovery Platform Website at: http://disgenet.org/dbinfo, section "Data Attributes".

d. DisGeNET gene-disease association type ontology

We have developed the DisGeNET gene-disease association type ontology to represent in a uniform and structured way the types of relations between genes and diseases found in the original data sources (Figure 6). For the details of the ontology used to describe gene-disease associations in DisGeNET please visit the DisGeNET Discovery Platform Website at: http://disgenet.org/dbinfo, section DisGeNET gene-disease association type ontology.

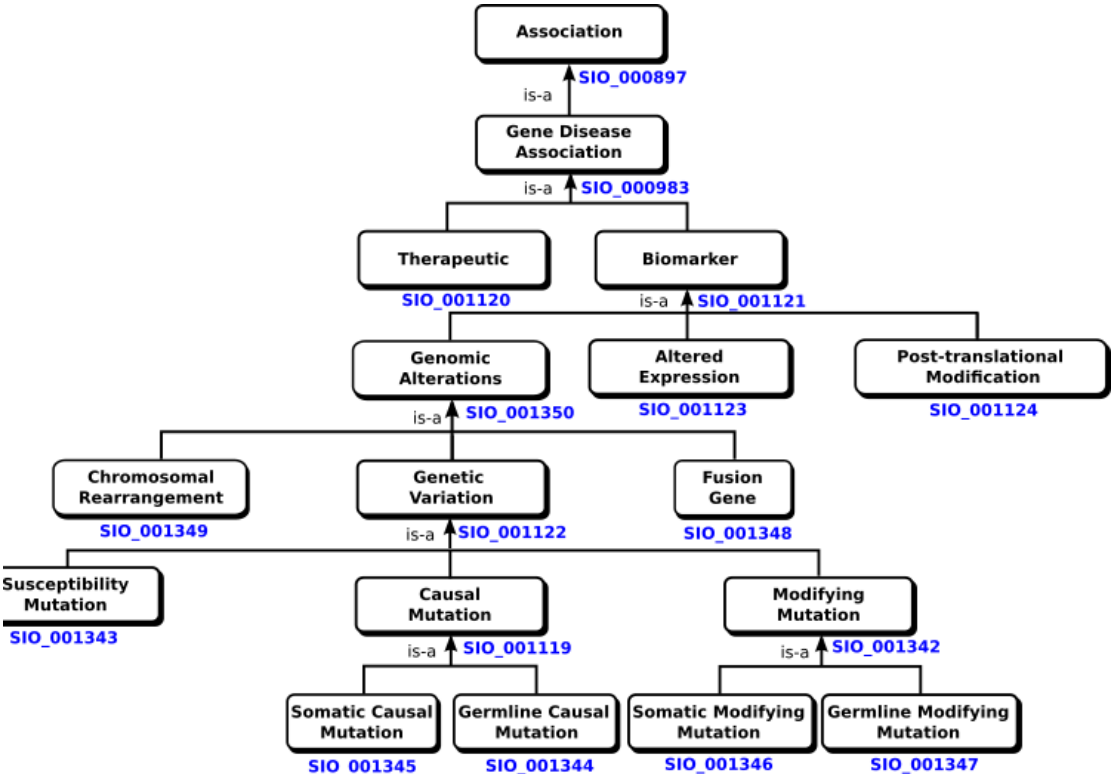

Figure 6: DisGeNET gene-disease association type ontology

3. Tutorial

The DisGeNET Cytoscape app is designed to visualize, query and analyse a network representation of the gene-disease and the variant-disease associations contained in DisGeNET. The app uses the current version (v7.0) of the DisGeNET that contains 1,134,942 gene-disease associations (GDAs), between 21,671 genes and 30,170 diseases, disorders, traits, and clinical or abnormal human phenotypes, and 369,554 variant-disease associations (VDAs), between 194,515 variants and 14,155 diseases, traits, and phenotypes.

The DisGeNET app (versions 5.1 or later) includes the DisGeNET Cytoscape style by default. Additionally, the style can be downloaded from http://disgenet.org/app, section "Additional files".

a. Basic functions

The DisGeNET App Control Panel (Figure 5) allows adjusting the parameters of the queries in order to create different types of networks. The panel contains two tabs to generate the *Gene Disease Network,* and the *Variant Disease Network*.
The *Gene Disease Network* tab is displayed by default. In this tab, different GDA networks can be generated by selecting different data *Sources*, *Association Types* and/or *Disease Classes* from their respective drop-down menus. The GDA networks may be also filtered using a cut-off value of the DisGeNET score, and/or Evidence Index (EI) and Evidence Level (EL). In addition, GDA networks can be built around specific disease(s) or gene(s) of interest using the *Search* boxes provided in the panel.
Using the *Variant-Disease Network*, different VDA networks can be generated by selecting different data *Sources*, *Association Types* and/or *Disease Classes* from their respective drop-down menus. The VDA networks may be also filtered using a cut-off value of the DisGeNET score, and/or Evidence Index (EI). In addition, VDA networks can be built around specific disease(s), gene(s), or variant(s) of interest using the *Search* boxes provided in the panel.

i. Generate gene-disease networks

In order to obtain a GDA network containing data from one specific source, for example, CURATED data, which includes information from all expert curated databases in our database (CGI, ClinGen, Genomics England, UniProt, CTD_human, PsyGeNET, and Orphanet), select the Source of interest (*CURATED*), and press the button *Create Network.*
The GDA network contains 20,884 nodes and 151,277 edges. Apply a Cytoscape layout algorithm to generate the view of choice, e.g. select the layout **Organic**. For more information on the layout styles, see http://apps.cytoscape.org/apps/yfileslayoutalgorithms. Once the network is obtained, specific information on the nodes and their relationships can be explored

using the Cytoscape Table Panel (at the bottom, right) (Figure 7).

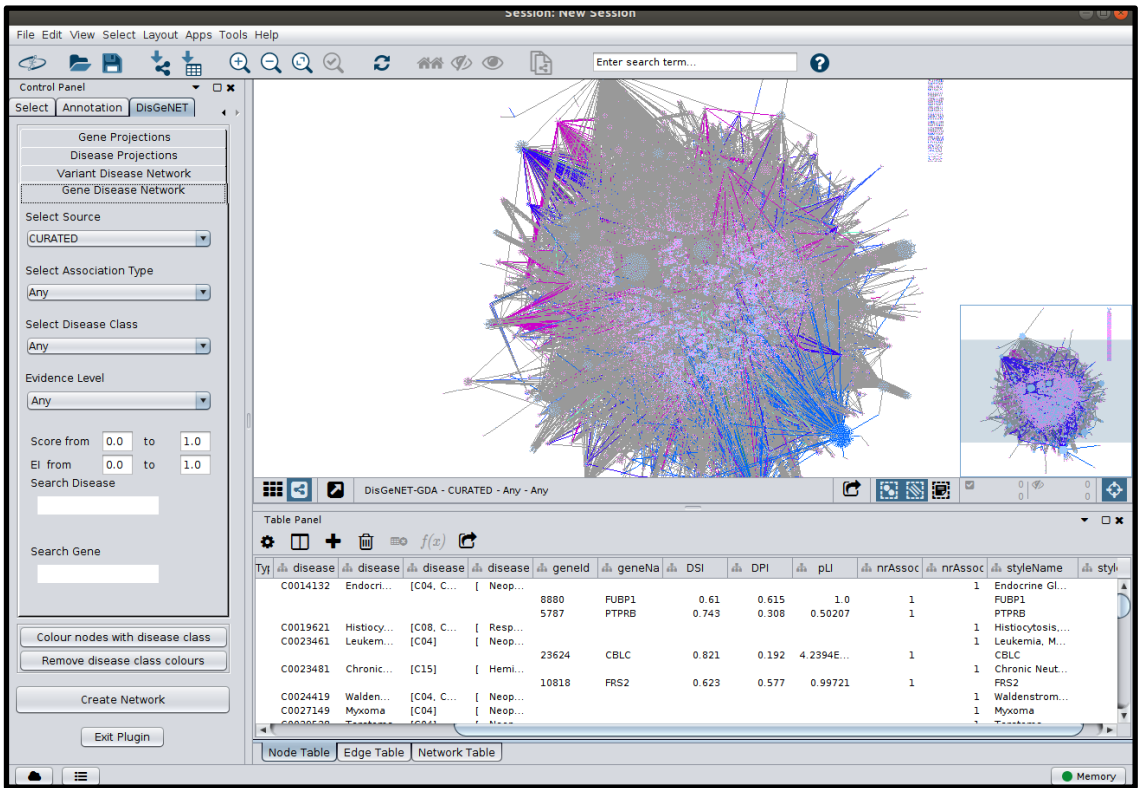

Figure 7: The CURATED GDA network

ii. Generate variant-disease networks

To obtain a VDA network containing data from one specific source, for example, the Genetic Association Database, select the *Variant Disease Network* tab, the source of interest (*UNIPROT*), and press the button *Create Network.* The results are shown inFigure 8. The network is composed of 26,145 nodes and 182,304 edges.

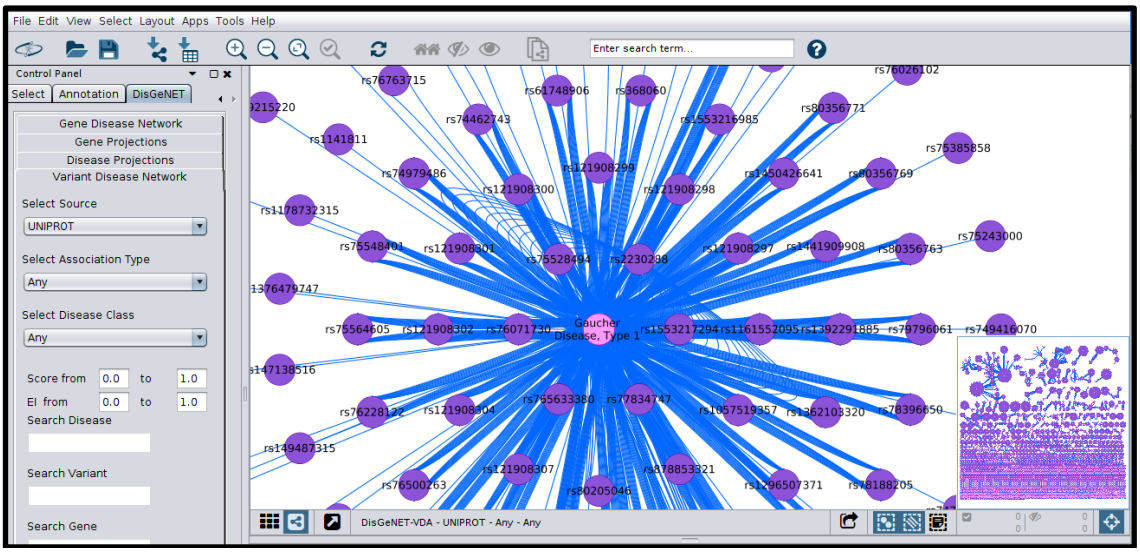

Figure 8: The UNIPROT VDA network

iii. Create networks by DisGeNET association type

The DisGeNET App allows searching by different categories of gene-disease association types, as described by the DisGeNET association type ontology (Figure 6).

To create a GDA network from CURATED data restricted to association type "Causal Mutation", select the Source, (for instance *CURATED*). Choose *Causal Mutation*, from the Association Type dropdown menu, and press *Create Network*.
The GDA network obtained contains 5876 nodes and 6560 edges (Figure 9).

Could you find the gene in the network carrying Causal Mutations for the largest number of diseases? Hint: order genes by the column nrAssociatedDiseases

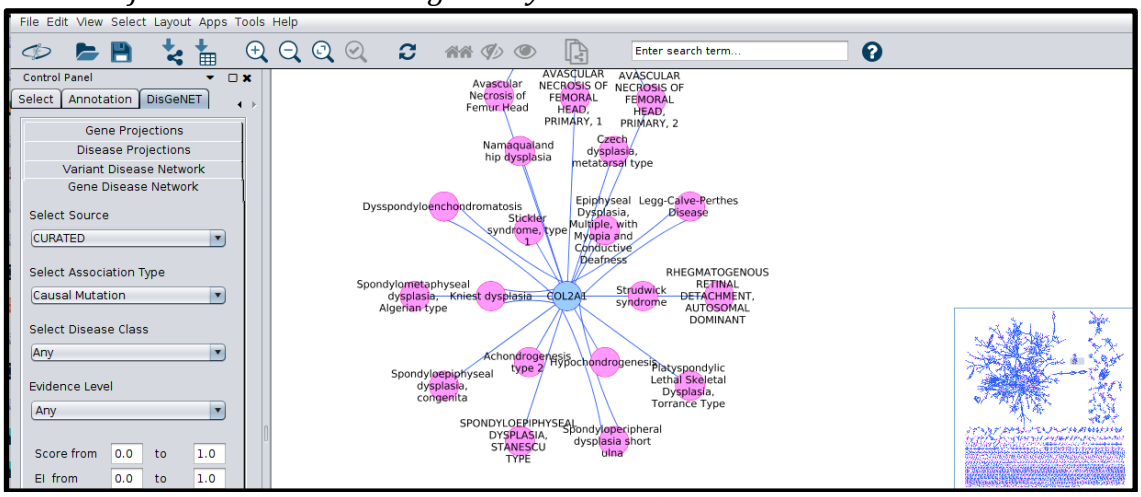

Figure 9: The CURATED GDA network for Causal Mutations

iv. Create networks by disease class

Networks can also be created by restricting to a specific type of MeSH disease class. The disease classification is based on the *Diseases* branch (C) and three categories (F01, F02, and F03) of the *Psychiatry and Psychology* Branch (F) of the MeSH hierarchy.
To generate a network of ANIMAL MODELS data, containing only *Nutritional and Metabolic Diseases*, Select the Source (*ANIMAL_MODELS*), and choose the Disease Class (*Nutritional and Metabolic Diseases*) from the Disease Class dropdown menu. Then, press the button *Create Network* (Figure 10). This GDA network has 1510 nodes and 3087 edges.
Which is the disease with the largest number of associated genes in this network? (Hint: use the network analyzer utility from Cytoscape Tools Main Menu)

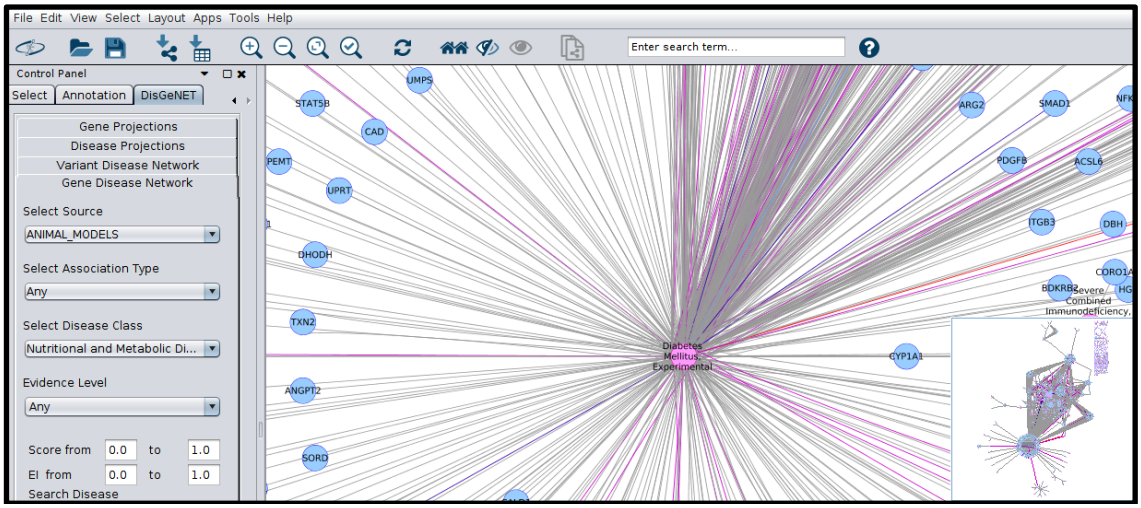

Figure 10: The GDA network for Nutritional and Metabolic Diseases in animal models of disease

v. Create networks by gene, disease, or variant

The *Search* option included in the DisGeNET control panel can also be used to generate different types of networks for a single disease, gene, or variant. This search may also be filtered by *Source, Association Type, Disease Class*, and *Score*. For example, the figure 15 shows the results for the search by a single gene, the ***methyl-CpG binding protein 2*** gene (MECP), filtering by disease class *Mental Disorders* in the curated data in DisGeNET (Figure 11).

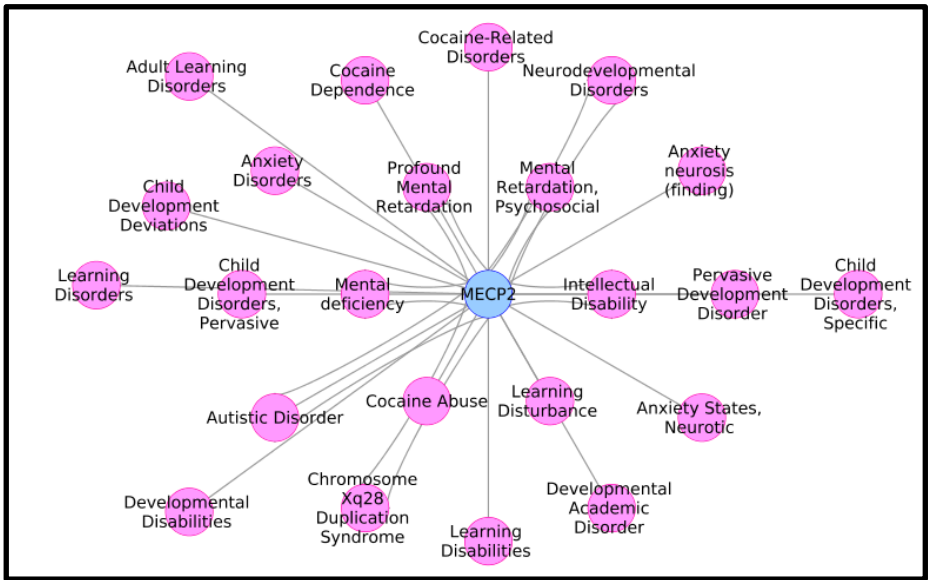

Figure 11: GDA network containing the metabolic diseases associated with the methyl-CpG binding protein 2 gene (MECP2) in DisGeNET CURATED.

vi. Multiple entity search in the DisGeNET App

The *Search* option can also be used to:

- Generate a network around a group of diseases or genes, matching a keyword
- Generate a network for a list of genes, diseases or variants, and their combinations

    a. Search by a disease matching a keyword

To build a GDA network containing DisGeNET data from CTD (human data) for all the types of Alzheimer Disease in this database, select the Source, *CTD_human* and write in the disease Search box '**alzheimer**' to create a network based on this keyword in the disease name. Then, press *Create Network* (Figure 12).
How many different subtypes of Alzheimer's Disease are included in DisGeNET (CTD human data)?
How many genes are associated with Alzheimer's Disease (C0002395)?

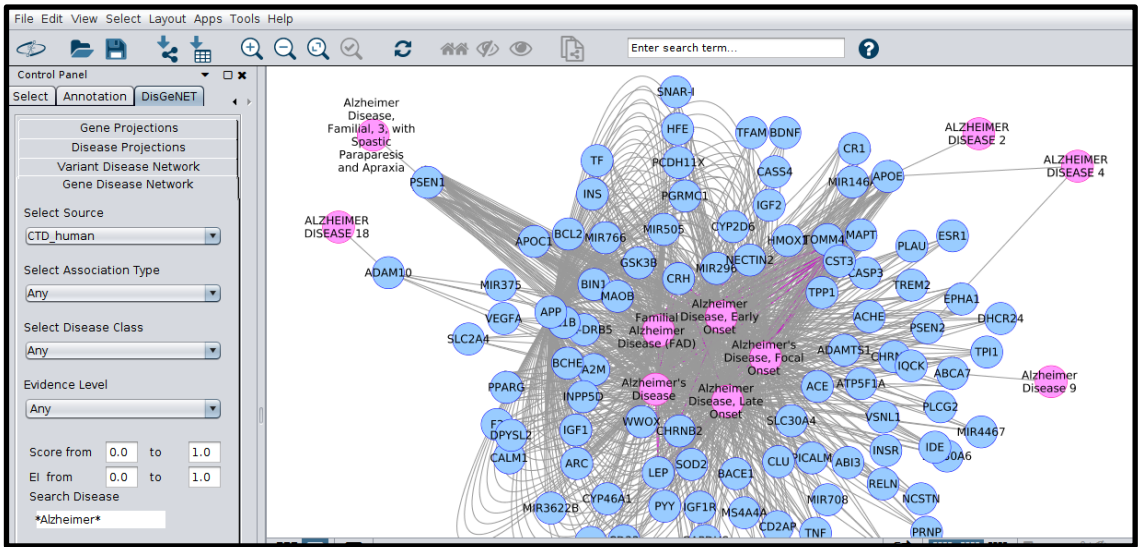

Figure 12: GDA network containing all subtypes of Alzheimer Disease in CTD, human data

This network can be further filtered by gene. For example, to generate the network of all the Alzheimer subtypes, and the amyloid beta precursor protein (APP) gene, type *APP* in the Gene Search Box, and press *Create Network* (Figure 13). A network with all subtypes of Alzheimer associated to the APP gene is created.

Each edge in the GDA Network represents the supporting evidence for a gene-disease association uniquely defined by the source, one association type, and one publication. The colour of each edge distinguishes the association type. Use the Edge Table in the Table Panel to explore the evidence for each association.

What is the score of the APP-Alzheimer's Disease association?

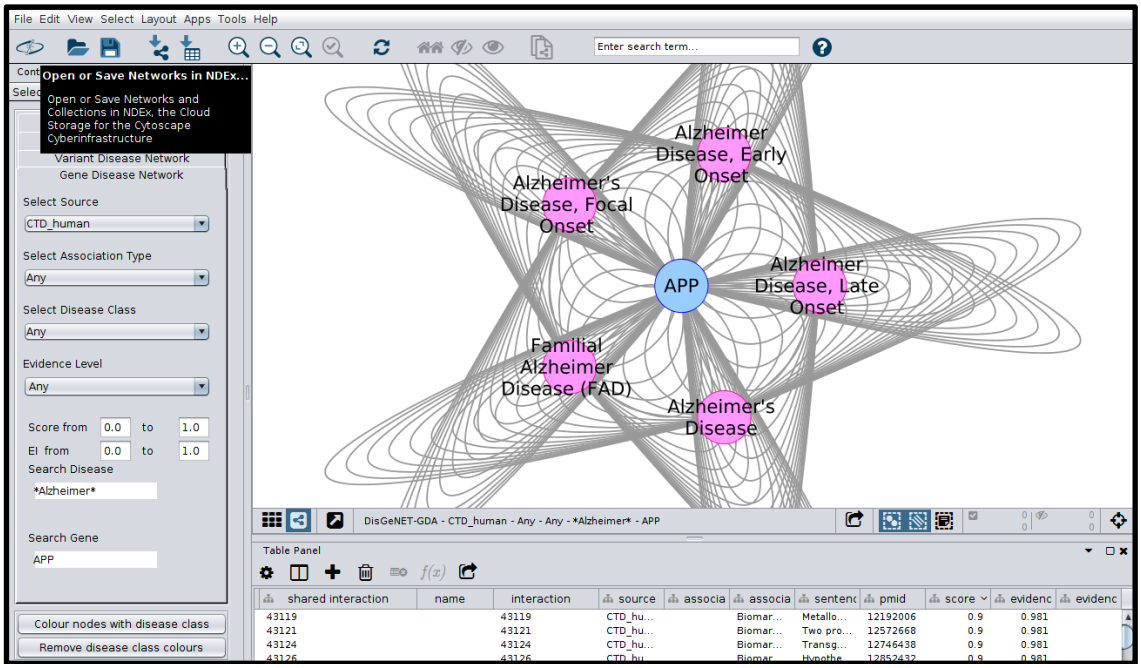

Figure 13: The GDA network for the amyloid beta precursor protein (APP) and all subtypes of Alzheimer Disease

b. Search by a list of genes

To build a GDA network associated with a list of genes, enter in the *Gene Search* Box the list of genes separated by ";" and press *Create Network*. Figure 14 shows the results of querying DisGeNET CURATED data for a list of potassium channels: *KCNE1; KCNE2; KCNH2; KCNG1*. *How many genes appear in the network?*

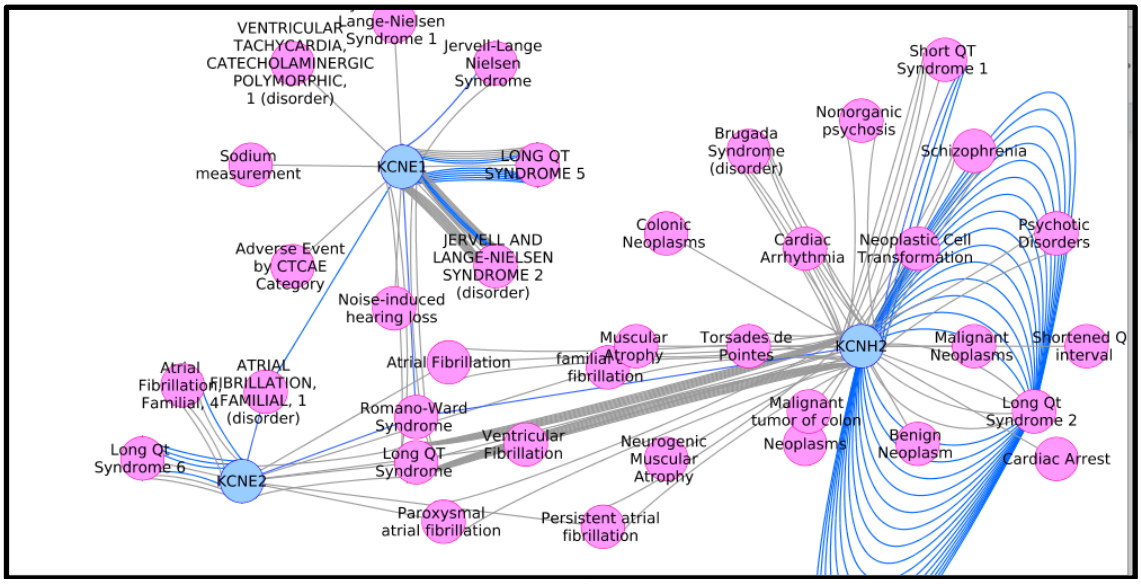

Figure 14: The CURATED GDA network for the genes KCNE1; KCNE2; KCNH2

c. Search by a list of variants

A similar procedure can be followed to query DisGeNET for the diseases associated with a list of variants. Go to the *Variant Disease Network* tab, select source "*CURATED*" and type the variants separated by ";" in the *Search Variant* and then press *Create Network*. Figure 15 shows the results of querying the following list of variants:
rs121907927;rs373661718;rs121907927;rs780356070;rs200015827;rs121907928;rs757259413;rs121907919;rs121907928

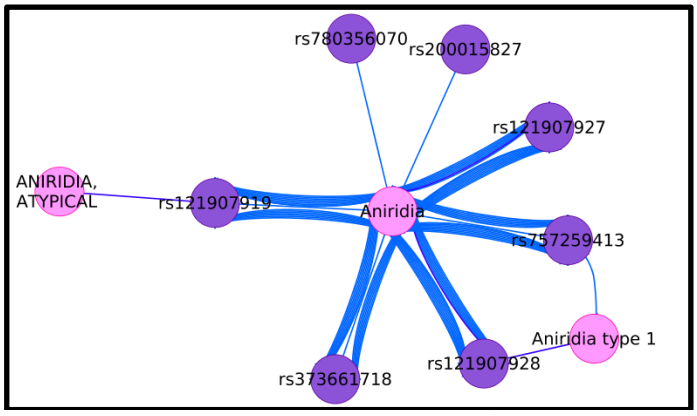

Figure 15: The CURATED VDA network for variants rs121907927;rs373661718;rs121907927;rs780356070;rs200015827;rs121907928;rs757259413;rs121907919; rs121907928

Notice that in this network, the genes associated with the variants can be displayed

by clicking on the box "***Show associated genes***" in the DisGeNET control panel (Figure 16). Apply the radial layout to obtain a similar visualization.

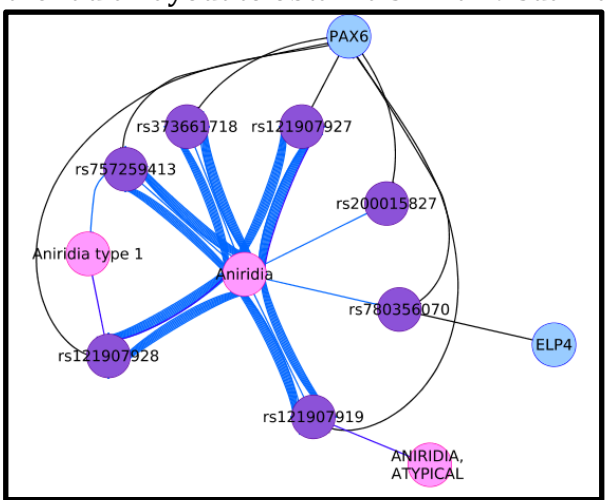

Figure 16: The CURATED VDA network for variants rs121907927;rs373661718;rs121907927;rs780356070;rs200015827;rs121907928;rs757259413;rs121907919; rs121907928 and their associated genes

d. Search by a list of diseases

Go to the *Gene Disease Network* tab and in the *Search Disease* box, enter the list following list of diseases:
C0268337;C0268342;C0268336;C0268335;C0268338;C0013720.
Filter by score > 0.3, and then press *Create Network* (Figure 17).

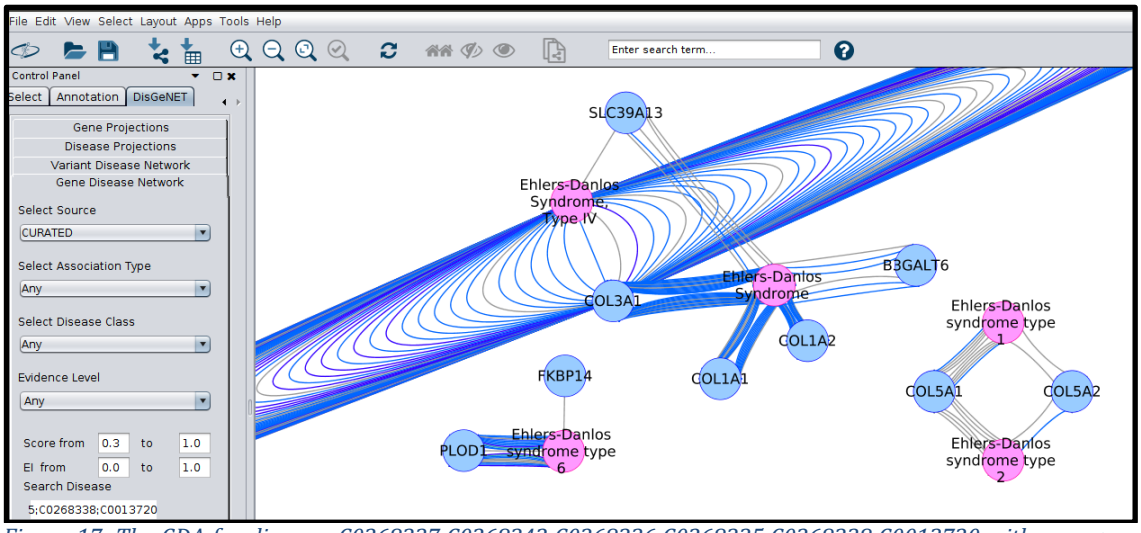

Figure 17: The GDA for diseases C0268337;C0268342;C0268336;C0268335;C0268338;C0013720 with score >= 0.4

The input terms in the *Search Disease* box can be combined, using UMLS CUIs, disease names, or a regular expression. For example, the following list C0002395;Schizophrenia;*Parkinson*, will retrieve the GDA network for Alzheimer's Disease, Schizophrenia, and all subtypes of Parkinson disease in DisGeNET. See the results of the search in DisGeNET CURATED data in Figure 18.

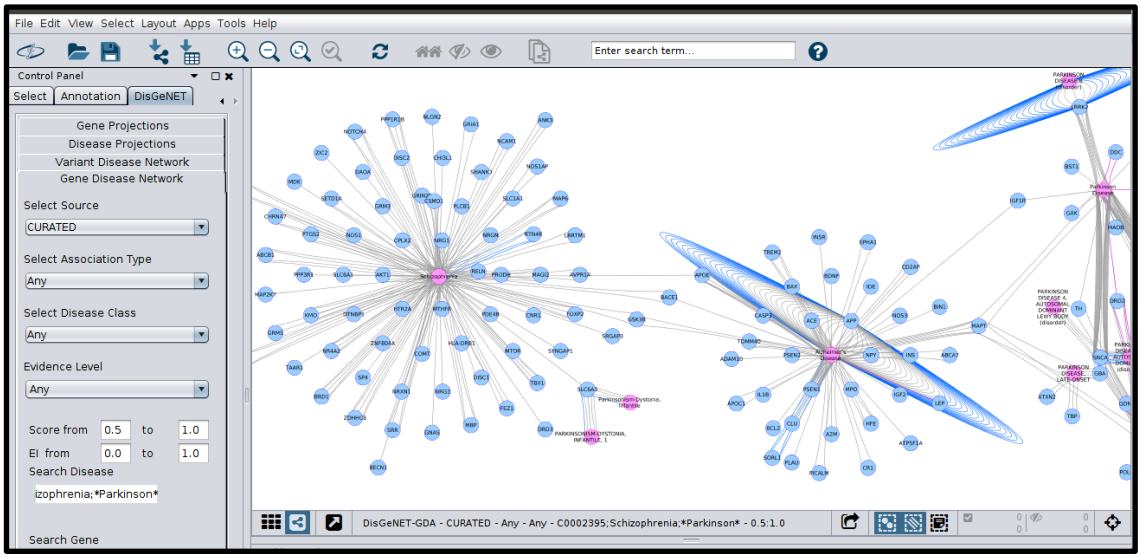

*Figure 18: The CURATED GDA network for diseases: C0002395; Schizophrenia;*Parkinson* with score >= 0.5*

Notice that the app returns a network if at least one of the entities in the query is included in DisGeNET data.

b. Advanced functions

i. Colouring nodes by disease class

The DisGeNET app allows colouring the nodes of a network according to the MeSH disease classification. In the case of diseases, the colouring is based on the disease class annotation of the disease node. Diseases belonging to more than one disease class will appear with more than one colour. In the case of genes and variants, the MeSH disease class is assigned from the disease(s) associated to them.

For example, in the *Gene Disease Network* tab, search for *INSR;INS;LEP;LEPR* in the gene search box, using the source *ANIMAL MODELS, Any* association type and *Any* Disease Class, and score >=0. Press "Create Network" (Figure 19).

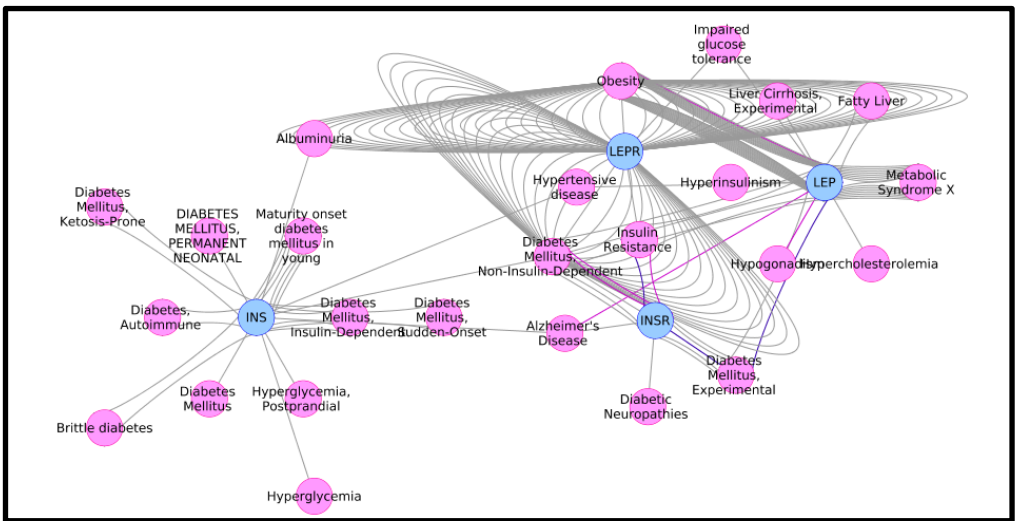

Figure 19: GDA network for INSR;INS;LEP;LEPR in animal models (score ≥0).

Once the network is created, click on *Colour nodes with disease class*, and all the nodes of the diseases in the network will be coloured according to their MeSH disease classes, while the genes will be coloured according to the classes of their associated diseases. The colour for each disease class is displayed in the Disease Class Legend shown at the right side. The border of the nodes keeps the original colour indicating if it represents a disease, a gene or a variant. Notice that if some diseases do not have MeSH disease class, they will be coloured in gray (Figure 20).

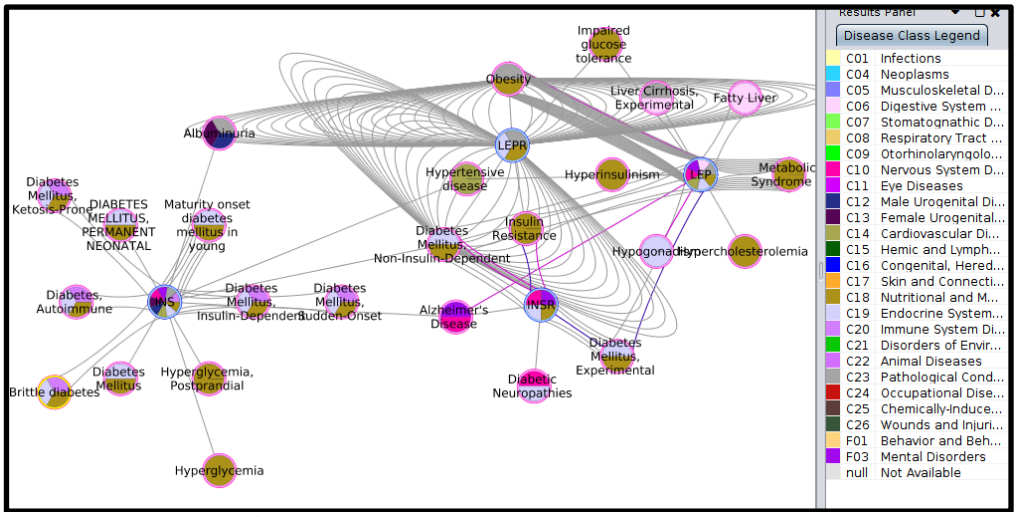

Figure 20: GDA network for INSR;INS;LEP;LEPR, in animal models, coloured according to MeSH disease classes.

ii. DisGeNET expand

The DisGeNET *expand* function allows to retrieve all the data available for a specific node in a network. The expand function is particularly useful when an initial search on a single database (e.g. UniProt) is performed, and once the network is obtained, you want to know if there are other associations in the whole DisGeNET database (namely ALL data set).
The function can be used to create a new DisGeNET network using the selected node(s) for the query or to expand the existing network with nodes and edges found in DisGeNET ALL. Currently, the function offers the following options:

- On a GDA network, applied to a gene, it will retrieve the diseases associated with the gene from ALL data.
- On a GDA network, applied to a disease, it will retrieve the genes associated with the disease from ALL data.
- On a VDA network, applied to a variant, it will retrieve the diseases associated with the variant from ALL data.
- On a VDA network, applied to a disease, it will retrieve the variants associated with the disease from ALL data.
- On a VDA network, applied to a gene associated with a variant, it will create a new GDA network with the diseases associated with the gene from ALL data.

For an example on how to apply the DisGeNET expand function to the GDA network for gene Alzheimer's Disease and APOE from CURATED data, see Figure 25. Right-click on the node of interest, in this case the APOE node, and go to **"Apps->DisGeNET->Expand"**. You have then to select either "Expand current net" or "Create new net". The ***"Set Parameters"*** box will appear to select the database (Figure 21).

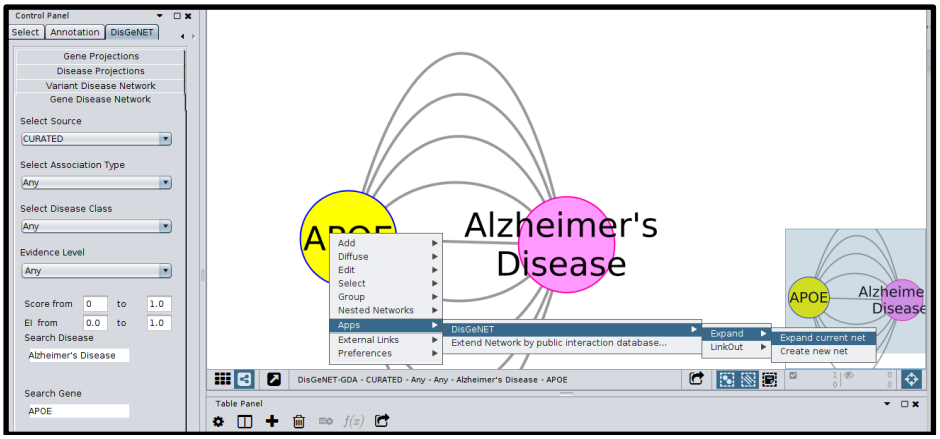

Figure 21: The expand function menu, on the node view context menu for the gene APOE and Alzheimer's Disease.

The ***"Set Parameters"*** box allows filtering the expanded network by data source, by score, and by attributes such as the disease class, and the association type. Select the database "*CURATED*", for example, and Disease class equal to "Nervous System Diseases". The results are shown in Figure 22.

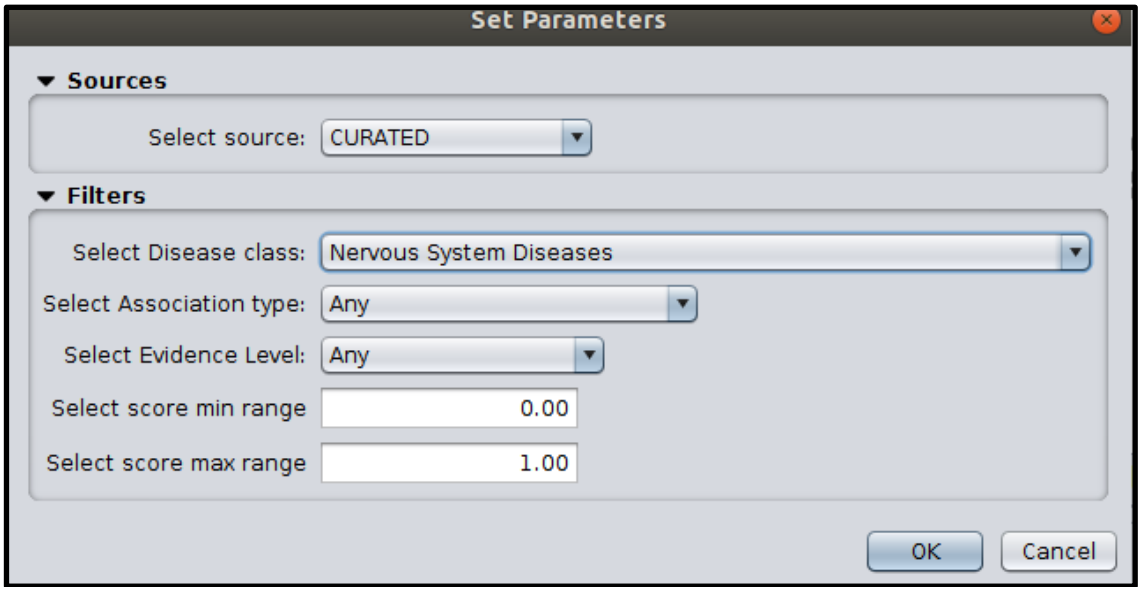

Figure 22: The Set Parameters Box associated to the expand function

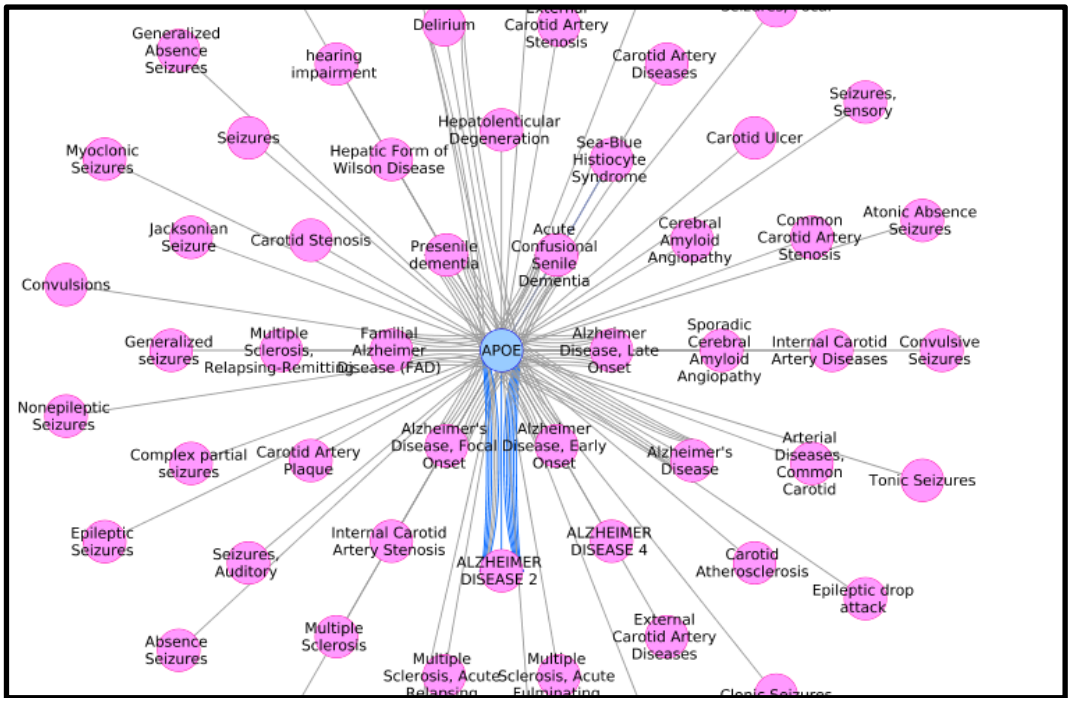

Figure 23: The expanded network for gene APOE (CURATED data), and disease class "Nervous System Diseases", displayed using the organic layout.

iii. DisGeNET Linkouts

1. LinkOuts for the nodes

In order to get more information about a specific gene, disease or variant from a DisGeNET network, you can use the *LinkOut* function to the reference databases (NCBI for Genes, LinkedLifeData for UMLS CUIs and dbSNP for SNPs). The LinkOut function is available in the node context menu, which can be accessed by right-clicking a selected node (Figure 24Figure 23). The available linkouts are specific to the node type.

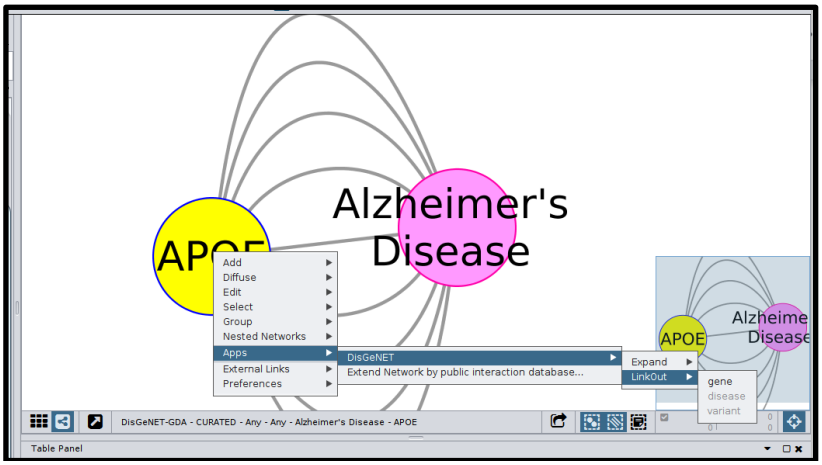

Figure 24: The LinkOut function menu, on the node view context menu, for the gene APOE.

2. LinkOuts for the edges

The user can also explore the publication supporting the GDA, or VDA, by clicking on the edge connecting the pair of interest, and accessing the LinkOut to NCBI Pubmed (Figure 25)

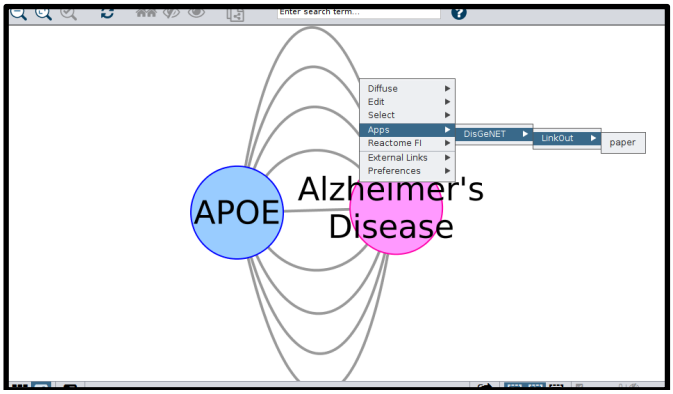

Figure 25: The LinkOut function menu, on the edge view context menu, for the association between Alzheimer's Disease and the gene APOE.

iv. Annotating foreign networks with DisGeNET data

The DisGeNET Cytoscape App also allows annotating networks generated by other applications or uploaded by the user. To access the function, right click on an empty space of the network view, to open the network context menu. Go to "**App->DisGeNET**" and select the desired function:

Annotate foreign protein networks with DisGeNET diseases

It is possible to annotate a foreign network containing genes using one of the following gene identifiers:

- Gene NCBI identifier
- Gene symbol

To annotate an external network, we will use as an example one of the networks in the Cytoscape Starter Panel. To display the Starter Panel, go to the main Cytoscape menu, and click "**View->show Starter Panel**". Click on the network "TCGA Colorectal Cancer" to open the session. Choose a set of nodes of the network (in the example we selected the first neighbors of the gene PLK1) and generate a new network (**"File->New->Network->From Selected nodes, all edges"**). This is done to reduce the size of the resulting network. After the new network is created, right click on an empty space of the network view, to open the network context menu. Go to "**App->DisGeNET->Annotate genes with DisGeNET->Genes->Annotate genes with diseases from the selected source"** (Figure 26). That will display the ***"Set Parameters"*** Box (Figure 27).

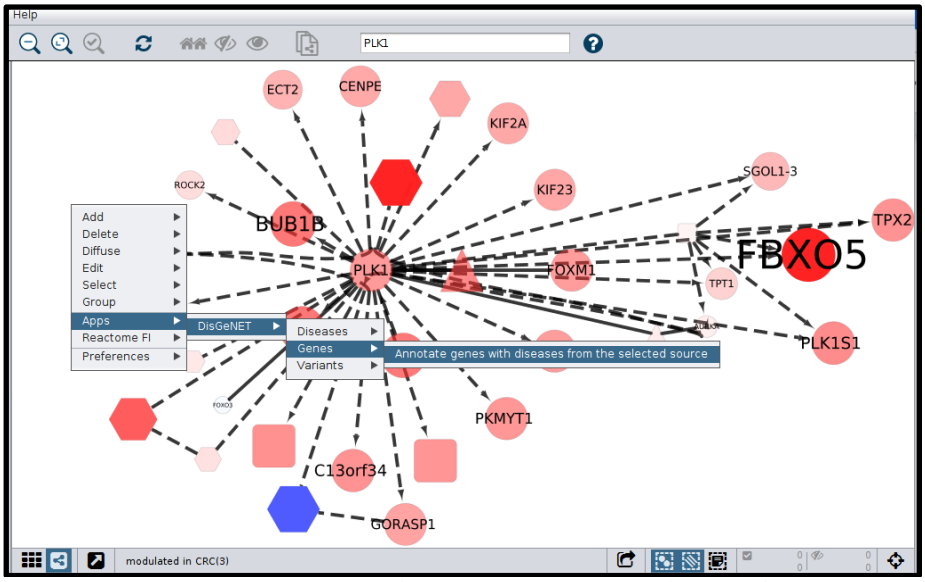

Figure 26: SubNetwork extracted from the example session AfinityPurification.cys

Notice that in this case, ***"Set Parameters"*** Box includes a new field, to select the name of the column containing the gene identifier (NCBI identifier or Symbol). Additionally, you might select the source of the annotation, a range of score, or the disease class. In the example, we used *CURATED* and "*Neoplasms*".

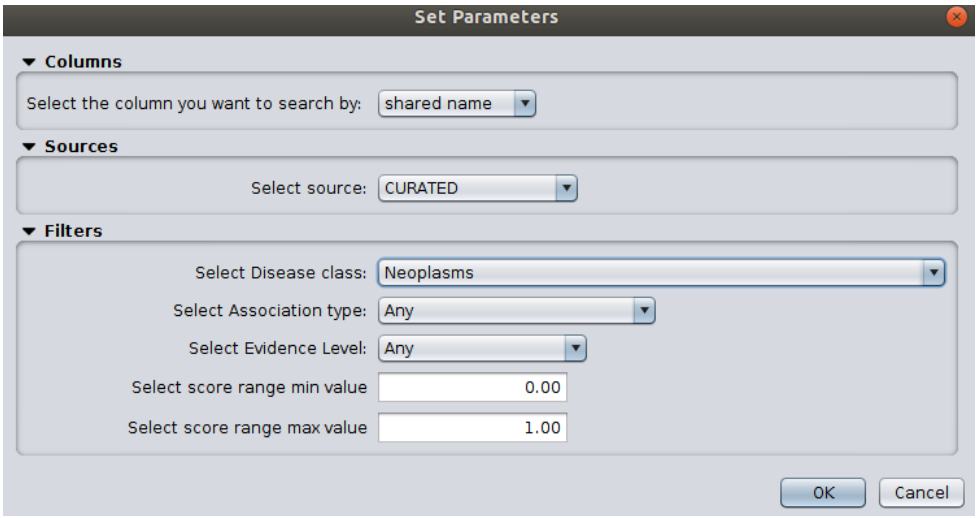

Figure 27: The Set Parameters Box associated to the Annotate external networks function

A new GDA network will be created with the diseases associated with the genes in the network according to the selected source (in the example, data from CURATED sources). The results are shown in Figure 28.

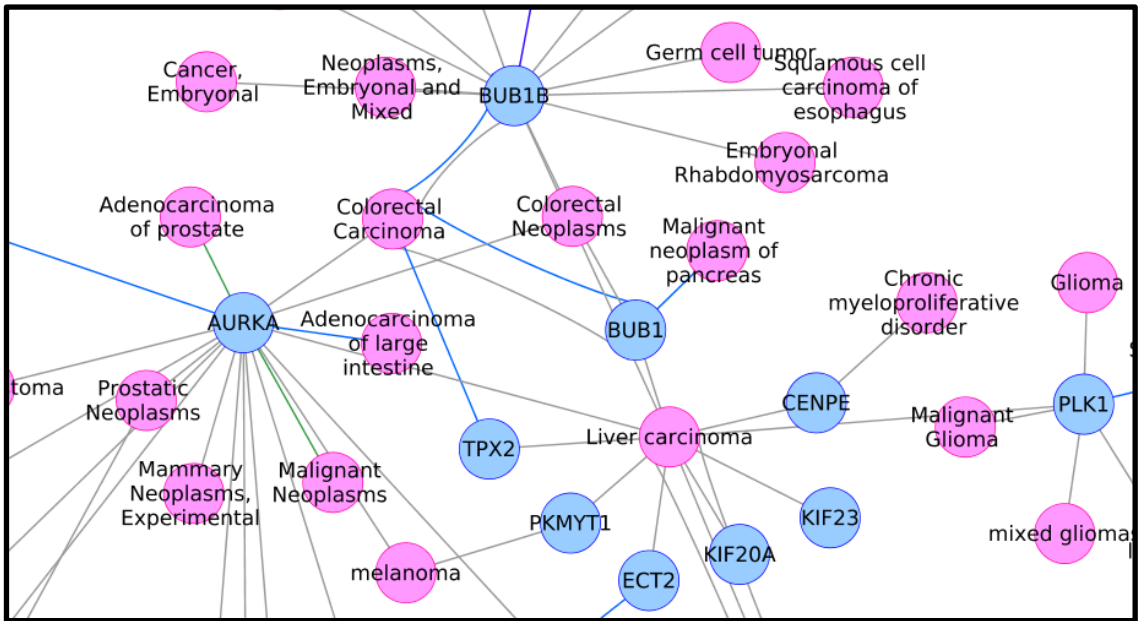

Figure 28: The CURATED GDA network for the genes from the external network

Annotate foreign variant networks with DisGeNET diseases

Similarly, a network containing variants identified by the dbSNP identifiers can be annotated. For example, copy the table in Annexes, containing the variants with their chromosomes, taken from the Supplementary table 4 from reference (5), a publication describing a GWAS that identified 143 risk variants for type 2 diabetes. Save it as a txt file in a document. In order to uncover which of the reported variants have already been associated to other traits, for example in the GWAS Catalog, follow these steps:

Create a new network from the file by clicking in the main menu of Cytoscape **"File->Import->Network->file"**. The result should be a network like the one in Figure 29. Then, right click on an empty space of the network view, to open the network context menu.

Go to "**App->DisGeNET->Variants->Annotate variants with diseases from the selected source"**. The "Set parameters" Box will be shown, to select the name of the column with the dbSNP identifier, and the source of the data. The results of the annotation using GWAS catalog data are shown in Figure 16.

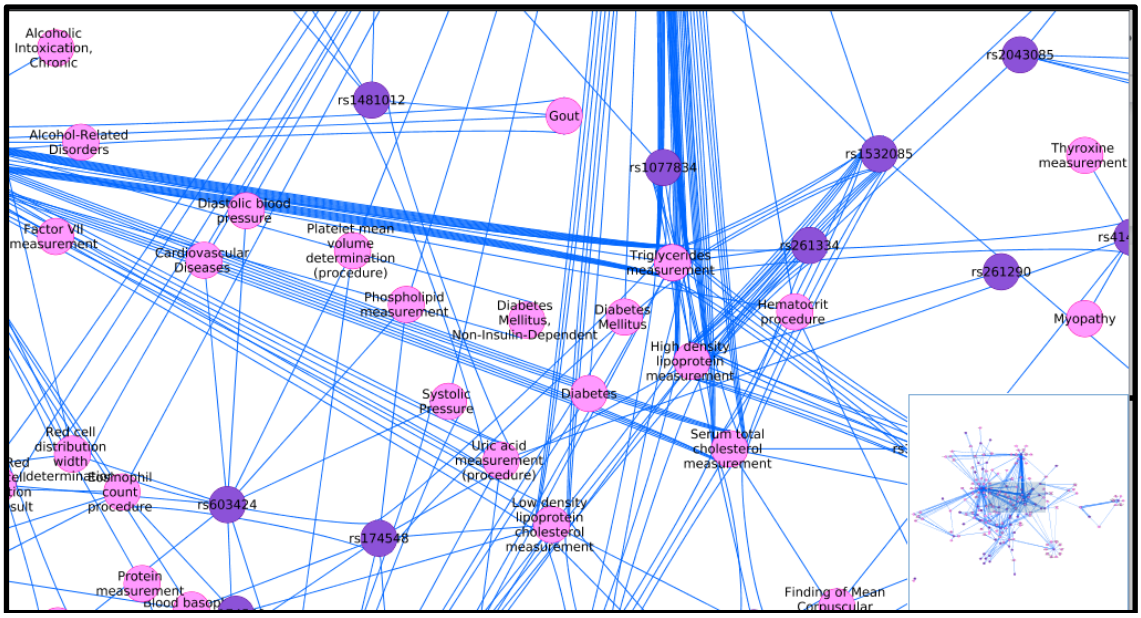

Figure 29: The annotated VDA network from DisGeNET All sources

Annotating a drug-target network with DisGeNET diseases

The DisGeNET Cytoscape App also allows annotating networks generated by other applications or uploaded by the user.

We will generate a drug-target network for the drug 5-fluorouracil from the STITCH database and annotate the targets with disease information.

1. Go to the STITCH database (http://stitch.embl.de/) and search the targets of 5-fluorouracil in humans.
2. Retrieve the network of drug targets selecting the sources: "Experiments", "Databases". This should result in a network of 5-FU and 10 proteins.
3. Download the network file in TSV format and import it in Cytoscape as a drug-target network. The network file should look like this, with one interaction per row of the table:

node1	node2	combined_score
CASP8	CASP3	0.998
DPYD	5-fluorouracil	0.995
BAX	TP53	0.993
UPP1	5-fluorouracil	0.992
UPP2	5-fluorouracil	0.963
TYMS	5-fluorouracil	0.960
BAX	5-fluorouracil	0.900
CYP2A6	5-fluorouracil	0.900
TP53	5-fluorouracil	0.900
CASP8	5-fluorouracil	0.900
CASP3	5-fluorouracil	0.900
DHFR	5-fluorouracil	0.900
UPP1	CYP2A6	0.899
UPP2	DPYD	0.899
DPYD	UPP1	0.899
DHFR	TYMS	0.899
DPYD	CYP2A6	0.899
UPP2	CYP2A6	0.899
DHFR	TP53	0.575
CASP3	BAX	0.569
CASP8	TP53	0.569
CASP8	BAX	0.569

Import the network by clicking on **File→ Import→Network→File**. *Note:* be careful to select the proper columns as nodes and edges in your network. Once the data has been imported in Cytoscape, a drug-target network for 5-FU will be obtained and displayed in Cytoscape (Figure 30):

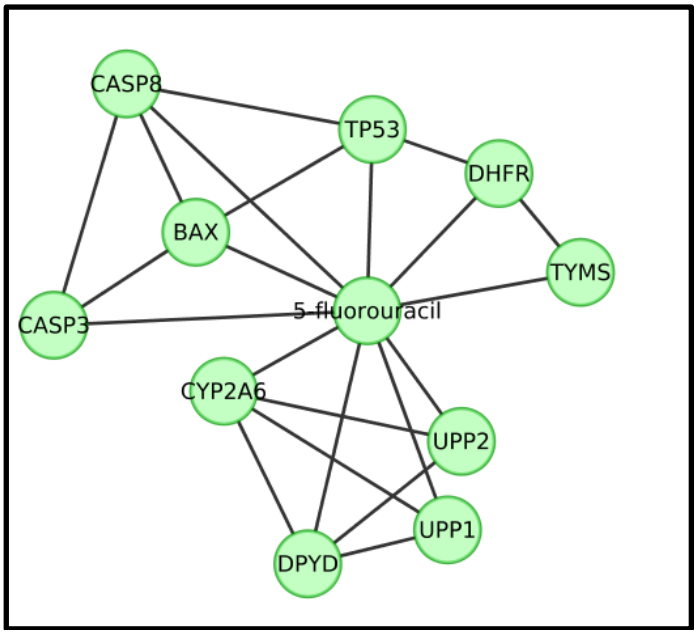

Figure 30: Imported network representing the drug targets of 5-FU

Questions:

- How many targets are in the 5-FU network?
- What is the degree of the 5-FU node?

Next, annotate the targets (genes) with **curated disease information** from DisGeNET. To access the *annotate function*, right click on an empty space of the network view, to open the network context menu. Go to "**App->DisGeNET-> Genes-> Annotate genes with diseases from the selected source"** (Figure 31).

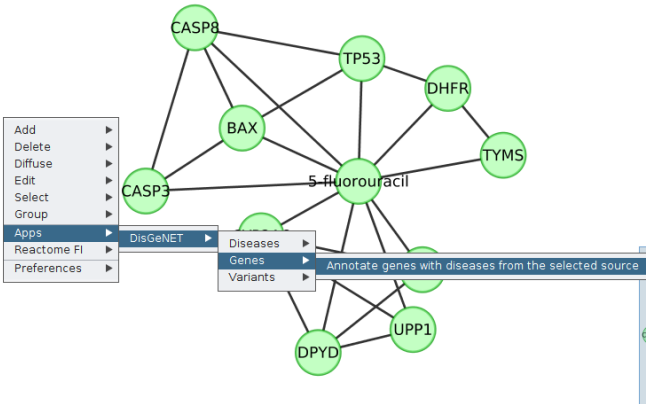

Figure 31: Annotating external networks with diseases

Then, in the "Set Parameters" Box choose the column “name”. As a result, a new GDA network for the targets of 5-FU will be generated (Figure 32).

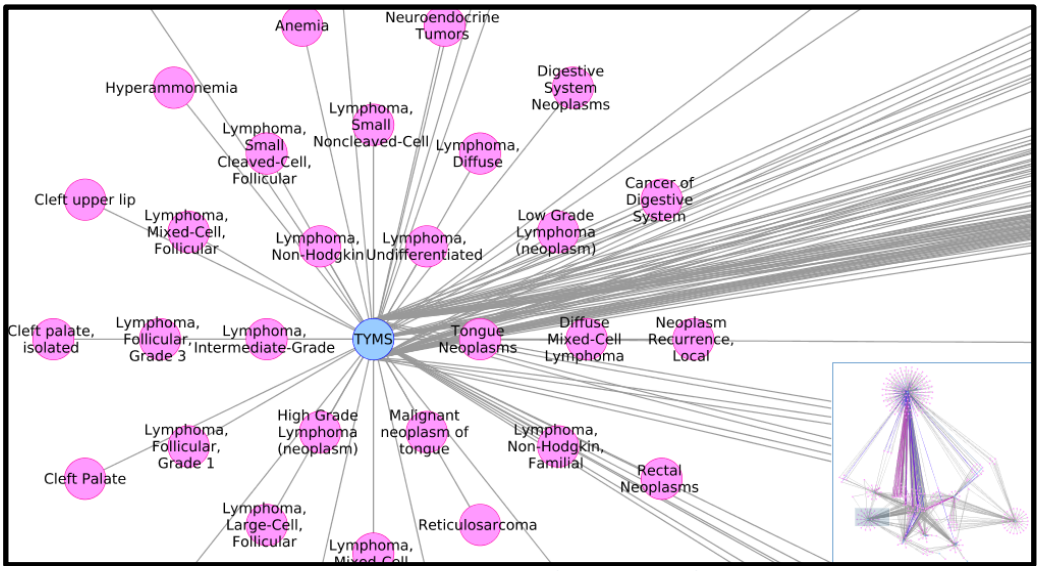

Figure 32: GDA network of the targets of 5-FU

Generate a sub-network of GDAs for TYMS, one the 5-FU targets.

Questions:

- What is the degree of TYMS in the GDA network?
- What classes of diseases are represented in the TYMS GDA network? *Tip:* inspect the MeSH disease classes in the Node Table or by colouring the network by disease class

c. DisGeNET automation

The Cytoscape Automation is a set of tools that allows users to create workflows executed entirely within Cytoscape or by external tools (such as RStudio or Jupyter). The DisGeNET Automation API allows querying the Cytoscape DisGeNET app from an external environment such as R, and Python, using REST calls.
The DisGeNET app includes an automation module with a set of REST endpoints. The documentation of the endpoints is available at the Swagger page of Cytoscape that can be accessed by going to the Cytoscape menu and clicking "**Help->Automation->CyRest API**". The API is accessible directly through the Swagger user interface within Cytoscape or by using any REST-enabled client. Figure 33 shows all the available endpoints, including those available by default through Cytoscape, and the one provided by the Cytoscape apps in the user installation of Cytoscape. We provide examples of scripts at http://disgenet.org/app.

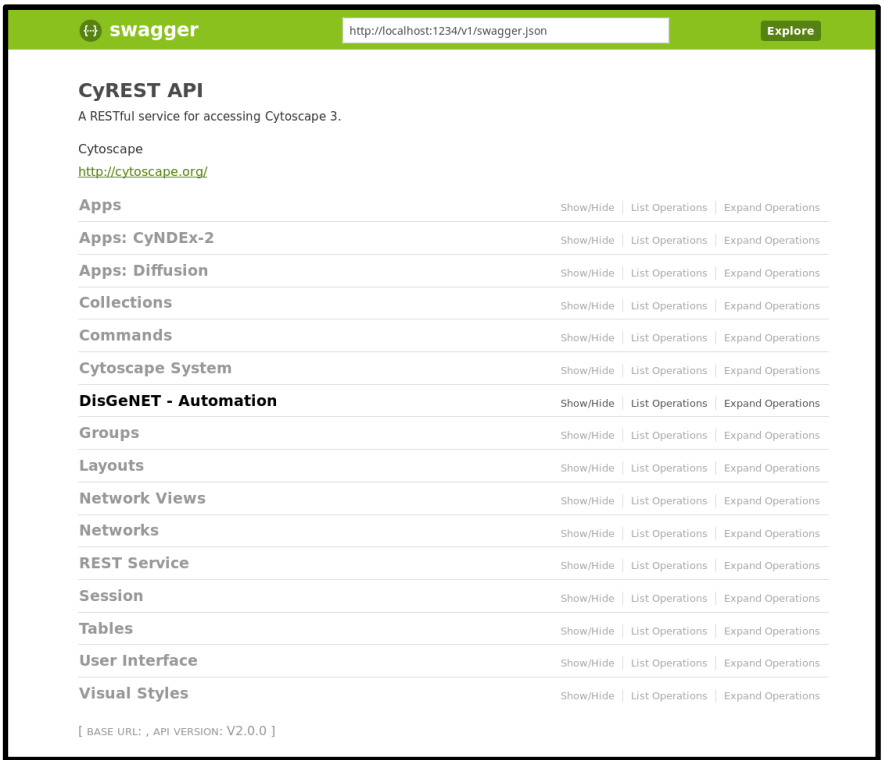

Figure 33: The swagger interface page for the CyRest API.

Expanding the *DisGeNET - Automation* menu will show the REST endpoints corresponding to DisGeNET (**¡Error! No se encuentra el origen de la referencia.**)

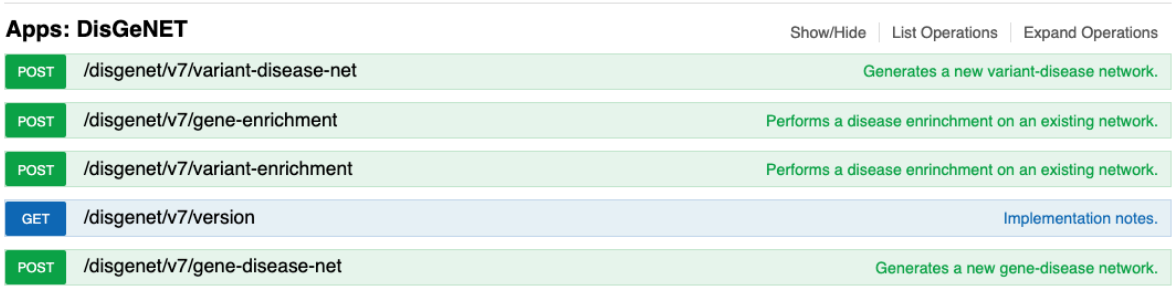

Figure 34: The available DisGeNET REST endpoints.

Important: In order to run the automation scripts, you first need to launch Cytoscape.

i. Using the DisGeNET automation in R

To use the DisGeNET automation in R, we provide an R script that can be found at http://www.disgenet.org/static/disgenet_ap1/files/current/disGeNETAutomation.R

To generate a VDA network, you can use the following R line:

variantDisResult <- disgenetRestCall("variant-disease-net",variantDisParams)

Previously, you need to define the parameters of your search, for example:

```
variantDisParams <- list(
        source= "UNIPROT",
        assocType= "Genetic Variation",
        diseaseClass= "Neoplasms",
        diseaseSearch= " ",
        geneSearch= " ",
        variantSearch= " ",
        initialScoreValue= "0.0",
        finalScoreValue = "1.0",
        showGenes= "true"
)
```

By executing the disgenetRestCall function, the results of the query will be displayed in Cytoscape.

The function *disgenetRestCall(netType, params)* creates the url using the function *disgenetRestUrl* and executes the REST call to the desired REST point and with the network parameters provided by the user, the function returns the results in a list containing a message with the result of the operation, and a list containing information of the network.

- *netType:* one of the following: gene-disease-net or variant-disease-net
- *netParams:* the only required field is the source. (see example below)

Example of *netParams* list for the gene-disease network.

```
geneDisParams <- list(
        source = "UNIPROT",
        assocType = "Genetic Variation",
        diseaseClass = "Neoplasms",
        diseaseSearch = " ",
        geneSearch = " ",
        initialScoreValue = "0.0",
        finalScoreValue = "1.0"
)
```

The function *disgenetRestUrl(netType, host, port, version)* creates the REST url using the following parameters:

- *netType:* one of the following: gene-disease-net or variant-disease-net
- *host:* the host/ip where the REST point is placed (by default, localhost).
- *port:* the port in which the REST point is listening (by default, 1234).
- *version:* the version of the REST point, should match the version of your DisGeNET App (by default, the latest version).

This function is ready to be used with the default Cytoscape automation setup

inside the *disgenetRestCall* function.

ii. Using the DisGeNET automation in python

To use the DisGeNET automation in Python, an example of script can be found at http://www.disgenet.org/static/disgenet_ap1/files/current/disgenet-automation.py

The script contains four functions that allow calling the automation module, and simplify the access of the data.

The function *disgenetRestUrl(netType, host, port, version)* creates the REST url using the following parameters:

- *netType:* one of the following: gene-disease-net or variant-disease-net
- *host:* the host/ip where the REST point is placed (by default, localhost).
- *port:* the port in which the REST point is listening (by default, 1234).
- *version:* the version of the REST point, should match the version of your DisGeNET App (by default, the latest version).

This function is ready to be used with the default Cytoscape automation setup inside the *disgenetRestCall* function.

The function *disgenetRestCall(netType, params)* creates the url using the function above and executes the REST call to the desired REST point and with the given net parameters, the function returns the results in a list containing a message with the result of the operation, and a list containing information of the network.

- *netType:* one of the following: gene-disease-net or variant-disease-net
- *netParams:* the only required field is the source. (see example below)

Example of *netParams* for the gene-disease network.

```
geneDisParams = {
        "source" : "UNIPROT",
        "assocType" : "Genetic Variation",
        "diseaseClass" : "Neoplasms",
        "diseaseSearch" : "",
        "geneSearch" : "",
        "initialScoreValue" : "0.0",
        "finalScoreValue" : "1.0"
}
```

For functions printHash and printOperationResult see the function documentation to get more information.

4. The SQLite database

The DisGeNET app queries a local version of DisGeNET data that is downloaded as an SQLite database. Each version of the App corresponds to a specific version of the SQLite database. The diagram of the data contained in the SQLite database (version 7.0) corresponding to the current version of the App (7.x), can be explored in Figure 35.

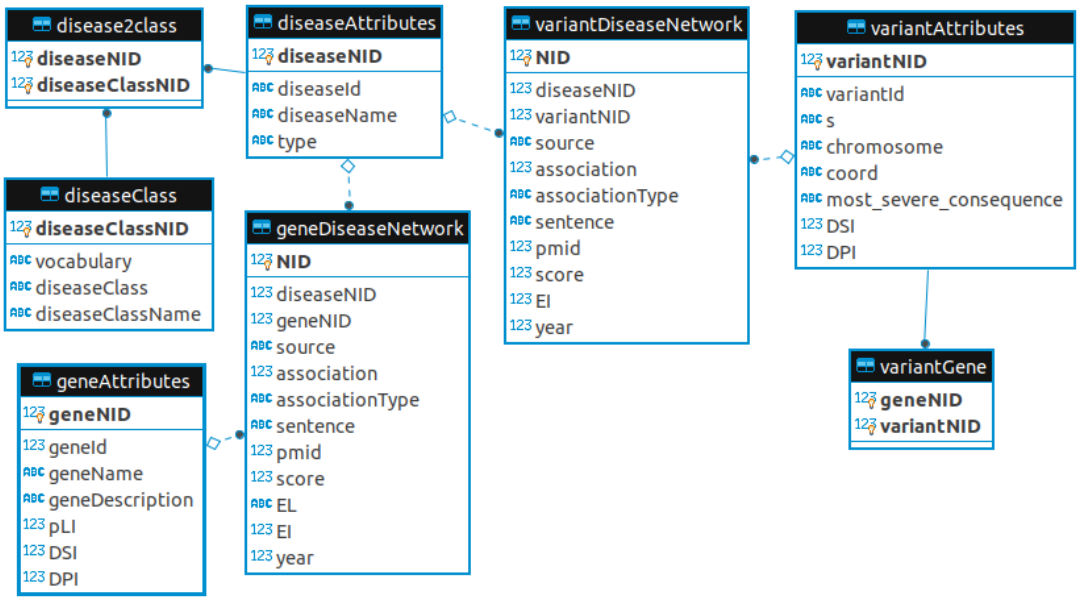

Figure 35: The relational schema of the DisGeNET SQLite database (version 7.0)

5. The node and edge tables in the Cytoscape Table Panel

In tables 1-3, we show a brief description of each field in the node Table, and edge Table, in the "Table Panel" of Cytoscape for the different networks generated by the DisGeNET app. For more information, visit http://www.disgenet.org/dbinfo.

Table 1: Edge attributes in the gene-disease network

Name	Description
interaction	Unique identifier for this association.
source	Original database in which this gene-disease association is reported.
associationType	Association type of the GDA according to the DisGeNET association type ontology (see section 2.4).
sentence	A representative sentence from the publication describing the association between the gene and the disease (If a representative sentence is not found, we provide the title of the paper).
pmid	PubMed identifier of the publication supporting the reported gene-disease association, if available.
score	The DisGeNET GDA score ranges from 0 to 1, and takes into account the number and type of sources (level of curation, model organisms), and the number of publications supporting the association.
EI	The Evidence Index for the GDA, that indicates the existence of contradictory results in publication.
EL	The Evidence Level measures the strength of evidence of a gene-disease relationship

Table 2: Node attributes in the gene-disease network

Name	Description
name	Name of the node, corresponding to NCBI identifier for genes, and UMLS CUIs for diseases
nodeType	The type of node (gene or disease).
diseaseId	UMLS® CUI of the disease.
diseaseName	Name of the disease.
diseaseClass	List of disease class identifiers according to the MeSH hierarchy.
diseaseClassName	List of disease classes according to MeSH hierarchy.
geneId	NCBI identifier of the gene.

geneName	Official Symbol of the gene.
Gene DSI	The Disease Specificity Index of the gene
Gene DPI	The Gene Pleiotropy Specificity Index of the gene
Gene pLI	The probability of a gene of being loss-of-function intolerant
nrAssociatedDiseases nrAssociatedGenes	Number of associated diseases or genes (number of first neighbours of the node).
styleName	Name of gene or disease, needed for the DisGeNET visual style.
styleSize	Number of first neighbours of the node, needed for the DisGeNET visual style.

Note: *ID* and *canonicalName* are internal unique identifiers used by the system.

Table 3: Edge attributes in the variant-disease network

Name	Description
interaction	Unique identifier for this association.
source	Original database in which this gene-disease association is reported.
associationType	Association type of the VDA according to the DisGeNET association type ontology (see section 2.4).
pmid	PubMed identifier of the publication supporting the reported variant-disease association, if available.
sentence	A representative sentence from the publication describing the association between the gene and the disease (If a representative sentence is not found, we provide the title of the paper).
score	The DisGeNET VDA score ranges from 0 to 1, and takes into account the number of sources, and the number of publications supporting the association.
EI	The Evidence Index for the VDA indicates the existence of contradictory results in publication.

6. Annexes

Table with the example variants to annotate external network

rsid	Chromosome
rs9429103	1
rs10890427	1
rs111790575	1
rs1126742	1
rs13375749	1
rs211723	1
rs211710	1
rs7552404	1
rs11161430	1
rs478093	1
rs61817724	1
rs6684114	1
rs1260326	2
rs6719753	2
rs3738848	2
rs10197755	2
rs796419162	2
rs13409366	2
rs10201159	2
rs13384756	2
rs13410232	2
rs13431529	2
rs10189885	2
rs6546869	2
rs6744398	2
rs7601356	2
rs1047891	2
rs715	2
rs887829	2

rs6742078	2
rs6804368	3
rs6800284	3
rs10010582	4
rs358236	4
rs1843481	4
rs1481012	4
rs141471965	4
rs4253255	4
rs4253328	4
rs37369	5
rs248386	5
rs27044	5
rs11386832	5
rs10075801	5
rs2405522	5
rs1801020	5
rs75077631	5
rs2545801	5
rs1165213	6
rs2817188	6
rs1165153	6
rs1165192	6
rs72939920	6
rs12208357	6
rs662138	6
rs316019	6
rs10242455	7
rs4921913	8
rs35570672	8
rs4873099	8
rs2759009	9
rs36034585	9
rs1171616	10
rs1171615	10

rs61886778	10
rs603424	10
rs10766469	11
rs174528	11
rs7943728	11
rs174533	11
rs174535	11
rs174536	11
rs61896141	11
rs102274	11
rs174545	11
rs174548	11
rs174549	11
rs174554	11
rs174555	11
rs174560	11
rs174561	11
rs174564	11
rs174566	11
rs5792235	11
rs99780	11
rs174580	11
rs151042642	11
rs113570042	11
rs11820589	11
rs964184	11
rs34265203	12
rs10774021	12
rs11613331	12
rs17329885	12
rs4149056	12
rs11045832	12
rs1871395	12
rs58310495	12
rs59205959	12

rs2939302	12
rs1799958	12
rs3916	12
rs7141433	14
rs17101394	14
rs8008068	14
rs11158671	14
rs8014023	14
rs16952714	15
rs80123226	15
rs261290	15
rs2043085	15
rs1532085	15
rs1077835	15
rs1077834	15
rs2070895	15
rs261334	15
rs4775633	15
rs28582913	16
rs7208714	17
rs4330	17
rs4335	17
rs4343	17
rs4362	17
rs1799763	17
rs922442	19
rs8012	19
rs7247977	19
rs62128825	19
rs2547239	19
rs2547237	19
rs212099	19
rs641738	19
rs8736	19
rs2540645	22

rs131813	22
rs131793	22

7. Citation

If you are using DisGeNET for your own research, please cite:

- The Browser, and the current version of the data:

Janet Piñero, Juan Manuel Ramírez-Anguita, Josep Saüch-Pitarch, Francesco Ronzano, Emilio Centeno, Ferran Sanz, Laura I Furlong. The DisGeNET knowledge platform for disease genomics: 2019 update (2019)
https://doi.org/10.1093/nar/gkz1021

- DisGeNET-RDF:

Queralt-Rosinach N, Piñero J, Bravo À, Sanz F, Furlong LI. **DisGeNET-RDF: Harnessing the Innovative Power of the Semantic Web to Explore the Genetic Basis of Diseases.** *Bioinformatics*. Bioinformatics (2016) doi: 10.1093/bioinformatics/btw214

- The Cytoscape App:

Piñero J, Saüch J, Sanz F, Furlong LI. **The DisGeNET Cytoscape App: exploring and visualizing disease genomics data.** *Under review*

Bauer-Mehren A, Rautschka M, Sanz F, Furlong LI. **DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks.** *Bioinformatics*. (2010) doi: 10.1093/bioinformatics/btq538

Bauer-Mehren A, Bundschus M, Rautschka M, Mayer MA, Sanz F, Furlong LI: **Gene-disease network analysis reveals functional modules in Mendelian, complex and environmental diseases.** *PLoS ONE* (2011)
doi:10.1371/journal.pone.0020284.

- To cite specific data:

Gene-disease association data retrieved from DisGeNET v7.0 (http://www.disgenet.org/), Integrative Biomedical Informatics Group, GRIB/IMIM/UPF . [Month, year of data retrieval].

8. References

1. Janet Piñero, Juan Manuel Ramírez-Anguita, Josep Saüch-Pitarch, Francesco Ronzano, Emilio Centeno, Ferran Sanz, Laura I Furlong. The DisGeNET knowledge platform for disease genomics: 2019 update (2019) https://doi.org/10.1093/nar/gkz1021
2. Janet Piñero, Àlex Bravo, Núria Queralt-Rosinach, Alba Gutiérrez-Sacristán, Jordi Deu-Pons, Emilio Centeno, Javier García-García, Ferran Sanz, and Laura I. Furlong. (2016) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucl. Acids Res.* doi:10.1093/nar/gkw943
3. Piñero,J., Queralt-Rosinach,N., Bravo,A., Deu-Pons,J., Bauer-Mehren,A., Baron,M., Sanz,F. and Furlong,L.I. (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database*, 2015, bav028–bav028
4. Bauer-Mehren,A., Rautschka,M., Sanz,F. and Furlong,L.I. (2010) DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. *Bioinformatics*, **26**, 2924–6
5. Bauer-Mehren,A., Bundschus,M., Rautschka,M., Mayer,M.A., Sanz,F. and Furlong,L.I. (2011) Gene-Disease Network Analysis Reveals Functional Modules in Mendelian, Complex and Environmental Diseases. *PLoS One*, **6**, 13
6. Xue A Wu Y Zhu Z Zhang F Kemper K et. al. (2018) Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nature Communications, vol: 9 (1) pp: 2941

9. Contact

Integrative Biomedical Informatics Group
Research Unit on Biomedical Informatics - GRIB
IMIM-Hospital del Mar Medical Research Institute
Universitat Pompeu Fabra (UPF)
Dr. Aiguader 88
08003 Barcelona, Spain

phone: +34 93 316 0521
fax: +34 93 316 0550
web: http://grib.imim.es/research/integrative-biomedical-informatics/

If you have questions or comments about DisGeNET data, the database, the website, the plugin, the browser, the RDF representation or the downloads, please contact us at: support(at)disgenet(dot)org

10. Funding

IMI-JU resources of which are composed of financial contribution from the EU-FP7 [FP7/2007–2013] and EFPIA companies in kind contribution [116030 to TransQST, 777365 to eTRANSAFE], and the EU H2020 Programme 2014–2020 [676559 to Elixir-Excelerate.]; Project 001-P-001647 - Valorisation of EGA for Industry and Society funded by the European Regional Development Fund (ERDF) and Generalitat de Catalunya; Agència de Gestió d'Ajuts Universitaris i de Recerca Generalitat de Catalunya [2017SGR00519]. The Research Programme on Biomedical Informatics (GRIB) is a member of the Spanish National Bioinformatics Institute (INB), funded by ISCIII and FEDER (PRB2-ISCIII [PT13/0001/0023, of the PE I+D+i 2013–2016]). The DCEXS is a 'Unidad de Excelencia María de Maeztu', funded by the MINECO [MDM-2014-0370].

11. License

The DisGeNET app is distributed under the GNU GPL 3.0 license. More details about the GNU General Public License 3.0 is available here.

The DisGeNET database is made available under the Attribution-NonCommercial-ShareAlike 4.0 International License whose text can be found here: https://creativecommons.org/licenses/by-nc-sa/4.0/.

If DisGeNET is incorporated into other works, we ask that DisGeNET is properly cited (see the citation guidelines), and that the version number of DisGeNET is clearly displayed.
http://disgenet.org/legal

12. About this document