DisGeNET Web Interface
USER GUIDE
DisGeNET Web Interface user guide

1. The web interface ....................................................................................................................... 2

1.1. What genes are associated to Alzheimer Disease? ................................................................. 3

1.2. What subtypes of Alzheimer Disease are in the database? .................................................. 7

1.3. How to retrieve the genes associated to several diseases at once? ....................................... 9

1.4. Which genes support the association between Chronic Kidney Disease and Diabetes Mellitus, Type 2? ........................................... 12

1.5. What diseases are associated to phosphatases? ..................................................................... 15

2. Citation........................................................................................................................................ 19

3. Contact......................................................................................................................................... 20

4. Acknowledgments........................................................................................................................ 20

5. License.......................................................................................................................................... 20
DisGeNET is a discovery platform to query and analyse human gene-disease associations. This platform can be accessed through a web interface, a Cytoscape plugin, a SPARQL endpoint, a faceted browser, and through scripts automatically generated by the queries to the web interface. In addition, the data can be downloaded as tab separated files, as a SQLite database and as an RDF data dump. The present tutorial is designed to describe the main functionalities associated to the web interface, providing examples of the type of questions that can be answered using the DisGeNET web interface.

1. THE WEB INTERFACE

The DisGeNET web interface has two entry points: the Search view and the Browser view. In both cases, the data can be explored in a “disease-centric” or a “gene-centric” way. In the Search view, the user can perform queries for individual genes or diseases (using the search box), or for lists of genes or diseases (using the multiple search button, Figure 1, #3, explained in section 1.3).
1.1. What genes are associated to Alzheimer Disease?

Select the type of search by clicking on the disease radio button (red box in Figure 1), and type the name of the disease in the search box to perform a free text search. You can also use a MeSH identifier (e.g. for “Alzheimer Disease”, D000544), OMIM identifier (omim:104300) or UMLS CUI (C0002395).

Figure 1: The search view, disease centric search.

Once you click on the magnifying glass icon, you will see a preview box summarizing the information on the entity of your query (“Alzheimer Disease”), followed by four hyperlinks that will lead you to different ways to visualize the results of your query. In the case of a disease search, the information provided in this preview is: the name of the disease, its UMLS concept unique identifier (CUI), the MeSH disease class and the UMLS semantic type.

Click on the link “Top 10 gene associations for this disease” (Figure 1, #1) to obtain a preview of the top-ten highest scoring associations.
Alzheimer Disease

**UMLS CUI**  umls:C0002395

**MeSH Class**  C10,F03

**MeSH Class Name**  Nervous System Diseases; Mental Disorders

**UMLS Semantic**  Disease or Syndrome

### Top 10 gene associations for this disease

Top scored gene associations for this disease.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene Name</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>APP</td>
<td>amyloid beta (A4) precursor protein</td>
<td>0.882</td>
</tr>
<tr>
<td>APOE</td>
<td>apolipoprotein E</td>
<td>0.6</td>
</tr>
<tr>
<td>PSEN1</td>
<td>presenilin 1</td>
<td>0.6</td>
</tr>
<tr>
<td>ACE</td>
<td>angiotensin I converting enzyme</td>
<td>0.521</td>
</tr>
<tr>
<td>BDNF</td>
<td>brain-derived neurotrophic factor</td>
<td>0.512</td>
</tr>
<tr>
<td>IL1B</td>
<td>interleukin 1, beta</td>
<td>0.498</td>
</tr>
<tr>
<td>MAPT</td>
<td>microtubule-associated protein tau</td>
<td>0.467</td>
</tr>
<tr>
<td>BACE1</td>
<td>beta-site APP-cleaving enzyme 1</td>
<td>0.466</td>
</tr>
<tr>
<td>NOS3</td>
<td>nitric oxide synthase 3 (endothelial cell)</td>
<td>0.446</td>
</tr>
<tr>
<td>PLAU</td>
<td>plasminogen activator, urokinase</td>
<td>0.433</td>
</tr>
</tbody>
</table>

Figure 2: The preview of the 10 highest scoring genes associated to Alzheimer Disease (C0002395)
Figure 2 shows a preview of the results of the query. Click on “Browse details...” (Figure 2, #1) to inspect all the results in more detail. This will open the “Summary of Associated Genes” tab in the Browser view (Figure 3).

The Browser view presents the results of your search from different perspectives (the tabs), and allows to perform additional searches within these results. They may be filtered according to pre-defined parameters, downloaded, and shared. Furthermore, scripts to retrieve the results of the searches are automatically generated. Notice on top of the tabs (violet buttons) the parameters used in your query: the disease (“Alzheimer Disease”, Figure 3, #1) and the data source (default value is ALL, Figure 3, #2). Go to http://www.disgenet.org/web/DisGeNET/v2.1/dbinfo#sources to obtain a more thorough description of source databases). You can change these parameters by closing the corresponding buttons.

The “Summary of associated genes” tab presents one record per gene-disease association. Clicking on the counter at the bottom of the page (Figure 3, #3), you can find out the number of genes associated to your disease. In this case, 1991 genes are associated with “Alzheimer disease” (C0002395) in DisGENET ALL subset. You can visualize more associations (up to 100) using the dropdown menu at the bottom-left of the page (Figure 3, #4), and navigate to the next set of associations through the “Next” button at the right hand corner (Figure 3, #5).

The search box (Figure 3, #6) allows performing free text searches within the results of each tab. For instance, you may search for a particular protein class within the genes associated to Alzheimer disease. To do so, type the name of a protein class, for example “signaling molecule”, and only the Alzheimer genes corresponding to this protein class will be displayed. To find out what other protein classes can be queried, see http://disgenet.org/web/DisGeNET/v2.1/help#panther.

The gene-disease associations can be ranked according to the number of supporting publications (Figure 3, #7) or the DisGeNET score (Figure 3, #8). By default, the associations are ranked by the DisGeNET score. In addition, you can also visualize only a subset of the results by applying a filter based on the DisGeNET score or the number of publications. Click on the filter icon and choose a cut-off value, for example greater than or equal to 0.3. This will indicate that the association is supported by at least one curated database (CTD human or Uniprot). For more information on the score see http://disgenet.org/web/DisGeNET/v2.1/dbinfo#score). If you want to see in more detail the evidences supporting the association between two entities, for instance Alzheimer Disease and the APP gene, click on the score corresponding to this association (Figure 4).
In the “All association evidences” tab, there is one line per evidence supporting the association between the gene and the disease (72 in this example). The number of evidences depends on the original source(s) reporting the association, the association type and the number of publications supporting each disease association. For each publication, we show an exemplary sentence where the gene and the disease are highlighted. In order to make visualization friendlier, we restrict the number of sentences to 10 for each association. Filters based on source database (e.g. CTD) and association type (e.g. Biomarker) might be applied on this tab. Closing the “APP” button (red arrow) will lead to the “All association evidences” tab that shows the evidences linking Alzheimer Disease to its associated genes for DisGeNET ALL (6991 evidences). This action leads to the same results view as clicking on the hyperlink “All evidences for this disease” of Figure 1, #2.

Notice again that in the Browser view you can change any of the parameters of your original query by closing the buttons that appear at the top of the page (In this example, by closing the “APP” button, Figure 4, red arrow). This might be done, for example, to change the source of the evidences. If you are interested in the genes associated to Alzheimer disease reported only by UniProt, by closing the button representing the source database (in Figure 3, #2, ALL), the list of available sources will be displayed (Figure 5).

Finally, using the “download” button (Figure 4, #1) you can download the results of this query as a tab separated file. More interestingly, each step of your query and filters is recorded in the system and can be executed again by R, Perl, Python or bash scripts that are generated automatically. These scripts are also available from the download button, and can be incorporated to your own data analysis workflow. Alternatively, you can share your results by sending a link by e-mail or embedding an HTML view of the results on a web site (available from the “share” button, Figure 3 #2).
1.2. WHAT SUBTYPES OF ALZHEIMER DISEASE ARE IN THE DATABASE?

Go to the Browser view, and click on the icon for ALL sources (Figure 5, red arrow).

![Browser view](image)

You will see the Browser view, with four tabs: “Genes”, “Diseases”, “Summary of All Associations” and “All Data”. Click on the “Diseases” tab and type “alzheimer” on the box on the top left (red arrow, Figure 6). There are 25 different diseases containing the word alzheimer in DisGeNET. Download the information using the “download” button at the top right (Figure 6, # 1).
Open the downloaded file in Excel or Open Office, and copy the first column (The list of CUIs for diseases that contain the word “alzheimer in the name”). It should look like this:

<table>
<thead>
<tr>
<th>CUI</th>
<th>Disease</th>
<th>Mesh Disease Class</th>
<th># of Genes</th>
<th># of Diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td>umls:C0002395</td>
<td>Alzheimer Disease</td>
<td>Nervous System Disease</td>
<td>1991</td>
<td>9187</td>
</tr>
<tr>
<td>umls:C0494463</td>
<td>Alzheimer's disease with late onset</td>
<td>Nervous System Disease</td>
<td>184</td>
<td>4642</td>
</tr>
<tr>
<td>umls:C0276496</td>
<td>Alzheimer disease, familial</td>
<td>Nervous System Disease</td>
<td>118</td>
<td>4283</td>
</tr>
<tr>
<td>umls:C0750901</td>
<td>Alzheimer's disease with early onset</td>
<td>Nervous System Disease</td>
<td>32</td>
<td>2102</td>
</tr>
<tr>
<td>umls:C0338448</td>
<td>Alzheimer disease type 2</td>
<td>Nervous System Disease</td>
<td>7</td>
<td>781</td>
</tr>
<tr>
<td>umls:C2967191</td>
<td>Alzheimer's disease</td>
<td>Nervous System Disease</td>
<td>6</td>
<td>927</td>
</tr>
<tr>
<td>umls:C1970209</td>
<td>Alzheimer Disease 12</td>
<td>Nervous System Disease</td>
<td>6</td>
<td>1875</td>
</tr>
<tr>
<td>umls:C1851958</td>
<td>Lewy Body Variant of AD</td>
<td>Nervous System Disease</td>
<td>5</td>
<td>831</td>
</tr>
<tr>
<td>umls:C1863051</td>
<td>Alzheimer Disease 2</td>
<td>Nervous System Disease</td>
<td>3</td>
<td>258</td>
</tr>
<tr>
<td>umls:C1853360</td>
<td>Alzheimer Disease 11</td>
<td>Nervous System Disease</td>
<td>3</td>
<td>798</td>
</tr>
</tbody>
</table>

Figure 6: Fragment of the “alzheimer” subset
1.3. HOW TO RETRIEVE THE GENES ASSOCIATED TO SEVERAL DISEASES AT ONCE?

Copy the list of CUIs from the above example, or save it as a file. Click on the "Multiple selections" button (Figure 1, #3), and follow the "Create custom filter" link. This will display the multiple search box (Figure 7).

Choose the column “UMLS CUI in the format: umls:C0002395, ...” (Figure 7, #1). Type a filter name (Figure 7, #2), ex: “Alzheimer List” and paste the contents of the clipboard in the box “Values”. Click on the “Create filter” button (Figure 7, #4). Depending on the number of diseases, this could take a couple of minutes.

![Figure 7: The multiple search box](image_url)

The results are shown in Figure 8. Notice the differences with Figure 1: the disease box now shows the filter name “Alzheimer List” (Figure 8, #1) and some of the names of the diseases in the list are shown (red arrows). Click on the “Browse details...” link (Figure 8, #2) to see all results. The same query could have been carried out uploading a text file with the identifiers, separated by a comma or by a new line, and using the “Browse” button in the multiple search box (Figure 7, #5).
Figure 8: Preview of the results of the multiple disease search

Figure 9 illustrates the results of the multiple diseases search. Notice now the name of the filter in the disease name button (#1) and the specific Alzheimer type (#2) and the associated gene in the first and second column, respectively. Go to “All Association Evidences” tab (#3) to explore the details of each association. You may then filter, explore, or download your results, as explained above for the individual search.
This search for multiple diseases may be performed using a list of MeSH or OMIM identifiers, provided that the list is composed of only one type of disease identifiers. Similar multiple search operations can be performed for genes, using Entrez Identifiers, Hugo Gene Symbols or UniProt accessions.
1.4. Which genes support the association between chronic kidney disease and diabetes mellitus, type 2?

Go to the search tab, disease view and type “kidney disease” (Figure 10)

As there is no exact match in the database for “kidney disease”, the results of the search displayed under the box (in the pink rectangle) are the names of the diseases containing the query phrase. Click on the link “Chronic Kidney Disease” (uims:C1561643) to get the report of the disease of interest.
Now, click on the Top 10 diseases that share genes with this disease (Figure 11, #1) and then, on “Browse details…” link. The results are displayed in Figure 12.
The “Summary of Diseases that share genes” tab (Figure 12) contains a list of diseases sharing at least one gene with Chronic Kidney Disease, and the number of genes that they share. The number of shared genes varies according to the source that is selected (in this case, ALL). Click on the “Number of Shared Genes” link, in the row corresponding to Diabetes Mellitus, Type 2 (C0011860), to inspect the genes in common between the two diseases. There are 134 common genes between the two diseases. You can also explore which of them are transporters, or are related to a transporter. Figure 13 shows the results of such analysis.
1.5. WHAT DISEASES ARE ASSOCIATED TO PHOSPHATASES?

Click on the “Summary of all Gene-Disease Associations” link (Figure 14, #1). This will take you to the “Summary of All Associations” tab (Figure 15).

- genes  ○ diseases

Examples: APP, P04637, histone deacetylase

Figure 13: Genes in common between Chronic Kidney Disease (C1561643) and Diabetes Mellitus, Type 2 (C0011860), transporter subset

Figure 14: “Summary of all Gene-Disease Associations” link
The “Summary of all Gene-Disease Associations” tab (Figure 15) presents all the gene-disease associations in DisGeNET, with the score of each association and the number of publications. In the search box (Figure 15, #1) of this tab, you can perform free text searches to find genes, diseases, protein classes, disease classes. Figure 15 shows a set of all gene-disease associations in DisGeNET ALL involving a phosphatase protein (7269 gene-disease associations).

Nevertheless, this subset of associations not only contains all genes annotated to the phosphatase protein class, but also includes records containing the word “phosphatase” on the gene or the disease name fields. To retrieve only the genes that belong to the phosphatase protein class, it is necessary to access programmatically the web interface.

By using the “download” button, on the top left of the browser (Figure 6, #1), it is possible to display the “Download” box (Figure 16).
Click on the “Use in Bash script” link (Figure 16, #1), to obtain the code corresponding to a bash script that will execute the query in a programmatic manner. If you execute this script as it is generated by the system, it will retrieve the 7269 gene-disease associations between 333 genes and 295 diseases. If you modify it to retrieve only the gene-disease associations involving proteins that belong to the protein class “phosphatase”, like it is shown in Box 1, it will retrieve 6561 associations, between 257 genes and 290 diseases. These are the associations for the genes classified as “phosphatase” in DisGeNET ALL.

Note that once you download the script, you can modify it to perform different queries. This script can be executed at any time to reproduce the results obtained using the web platform, and can be integrated into complex bioinformatic workflows to annotate genes and diseases with DisGeNET data.
Box 1: Modified bash script having the *where* statement modified to retrieve only the genes belonging to the protein class “phosphatase”.

```bash
#!/bin/bash
query="DEFINE
    c0='/data/gene_disease_score_onexus',
    c1='/data/sources',
    c2='/data/genes',
    c3='/data/diseases'
ON
    'http://bitbucket.org/janis_pi/disgenet_onexus.git'
SELECT
    c0 (geneId, score, diseaseId, geneId, sourceId, pmids),
    c2 (geneId, name, description),
    c3 (diseaseId, externalLink, name)
FROM
    c0
WHERE
    (
        c2.pantherName CONTAINS 'phosphatase'
    )
    AND
    c1 = 'ALL'
) ORDER BY
    c0.score DESC"

echo $query | curl -X POST -d @: http://www.disgenet.org/oql
```
2. CITATION

- The Browser, and the current version of the data:
  Janet Piñero, Núria Queralt, Àlex Bravo, Jordi Deu-Pons, Anna Bauer-Mehren, Martin Baron, Ferran Sanz, Laura I Furlong. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Submitted

- The plugin:

- To cite specific data:
  Gene-disease association data were retrieved from the DisGeNET Database, GRIB/IMIM/UPF Integrative Biomedical Informatics Group, Barcelona. (http://www.disgenet.org). [Month, year of data retrieval].
3. CONTACT

© 2010-2014, Integrative Biomedical Informatics

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://ibi.imim.es

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

4. ACKNOWLEDGEMENTS

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 n° [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. The Research Programme on Biomedical Informatics (GRIB) is a node of the Spanish National Institute of Bioinformatics (INB).

5. LICENSE

The DisGeNET database and Knowledge discovery platform are provided to advance the knowledge about human diseases and their associated genes, and are intended to be used only for research and education. Any reproduction or use for commercial purpose is prohibited without the prior express written permission of the Integrative Biomedical Informatics group http://ibi.imim.es/.

The DisGeNET database is made available under the Open Database License whose full text can be found at http://opendatacommons.org/licenses/odbl/1.0/. Any rights in individual contents of the database are licensed under the Database Contents License whose text can be found at http://opendatacommons.org/licenses/odbl/1.0/.