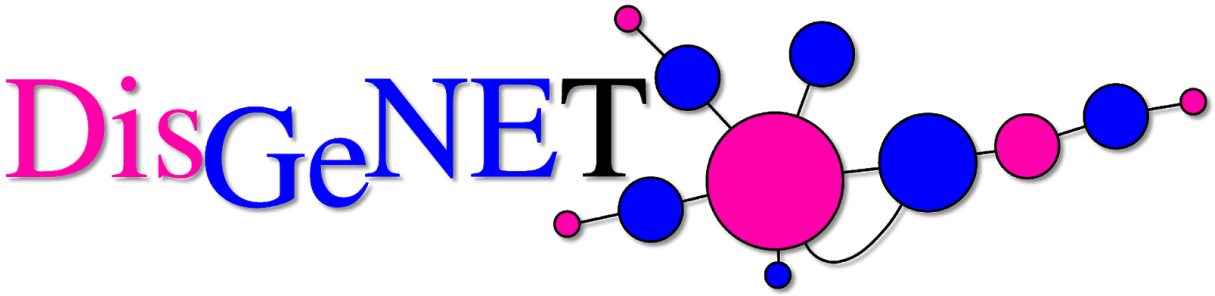


DisGeNET Web Interface

IBI Lab
[2020]

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)

1. The DisGeNET 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", a "gene-centric", or a "variant-centric" way. In the Search view, the user can perform queries for individual diseases, genes, or variants or several diseases, genes or variants (using the search box).
In the Browser view, the user can explore data from particular source databases (e.g. UniProt, Orphanet, among others).
In this section we show with examples the main functionalities of the web interface to explore DisGeNET data. Section 2 presents a brief outline of the database, detailed information can be found on the web (http://disgenet.org/dbinfo).

1.1. What Genes Are Associated with Alzheimer's Disease?

Go to the entry point of the DisGeNET search (http://disgenet.org/search) and select the type of search by clicking on the radio button (by default, diseases, Figure 1 #1), and type the name of the disease in the search box to perform a free text search. As you start typing, suggestions of results containing the typed word will be displayed. Select the term "Alzheimer's Disease", and click on the magnifying glass icon to launch the search. **It is important to wait for the autocomplete to launch. This applies to all search boxes in the new DisGeNET web interface.**

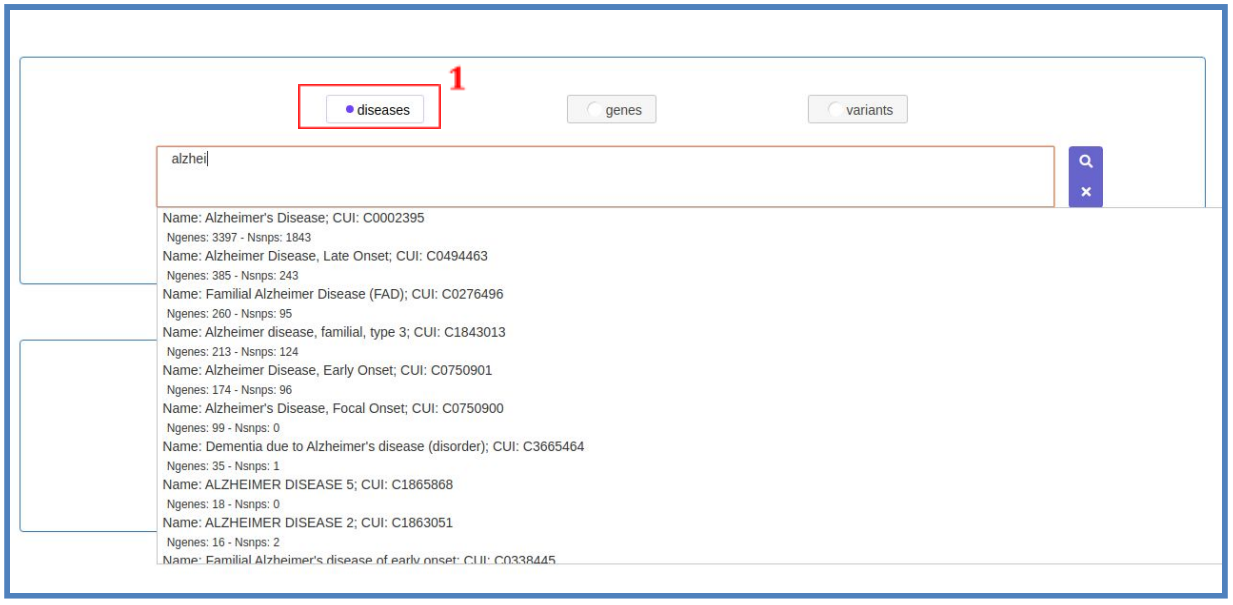


Figure 1: The search view, disease centric search.

Once you click on the magnifying glass icon, you will see a preview box summarizing the information on the entity of your query ("Alzheimer's Disease"), followed by five buttons

that will lead you to different ways to visualize the results of your query (Figure 2). 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, the UMLS semantic type, the type of Phenotypic Abnormality according to HPO, and the top level class from the Disease Ontology. The search by diseases can also be performed using MeSH identifiers, OMIM identifiers or UMLS concept unique identifiers (CUIs) (e.g. for "Alzheimer's Disease" D000544, 104300, and C0002395, respectively).

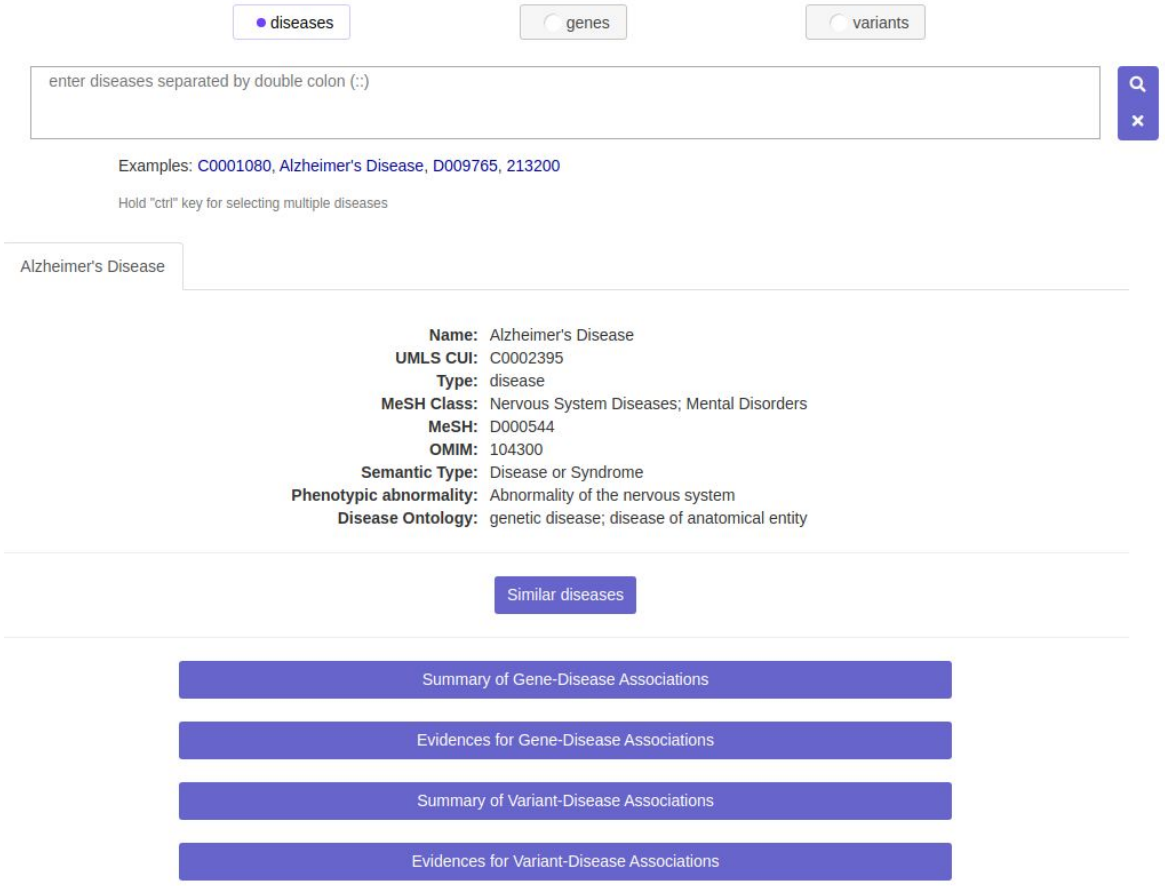


Figure 2: The preview of the search for Alzheimer's Disease (C0002395)

Click on the button "*Summary of Gene-Disease Associations*" to inspect all the results in more detail. This will open the "*Summary of GDAs*" tab in the Browser view (Figure 3).

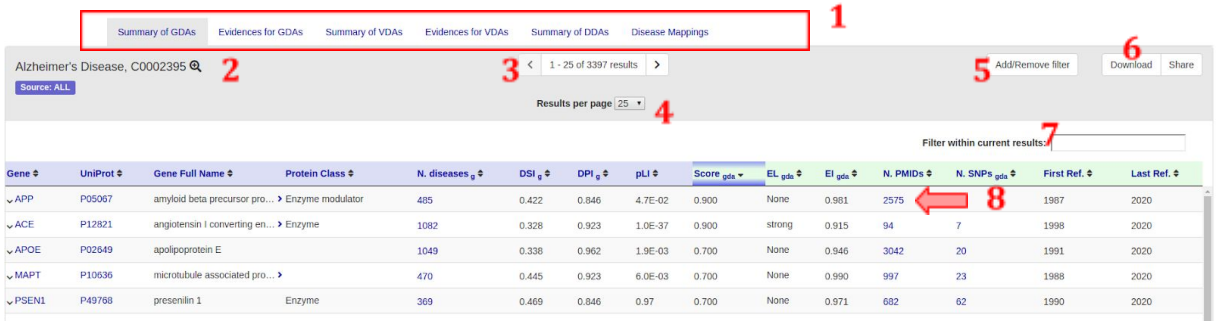


Figure 3: "***The Summary of GDAs***" tab for the search using "Alzheimer's Disease" (C0002395) in DisGeNET ALL

This view presents the results of the search from different perspectives (the tabs, Figure 3, #1), and allows performing additional searches within these results. They may be filtered according to different parameters, downloaded, and shared. Notice on top of the active tab (*Summary of GDAs*) the parameters used in your query: the disease ("Alzheimer's Disease", Figure 3, #2) and the data source (default value is ALL). Go to http://disgenet.org/dbinfo, section "Original Data Sources" for a thorough description of source databases). You can obtain more information about the disease by clicking on the magnifying glass next to the disease name (#2).

The "*Summary of GDAs*" tab presents one record per gene-disease association, in this case, for all the genes associated with Alzheimer's Disease. The counter indicates the total number of genes associated to the disease (Figure 3, #3). There are 3,397 genes associated with "Alzheimer's disease" (C0002395) in DisGeNET ALL). You can visualize more associations (up to 200) using the dropdown below the counter (Figure 3, #4), and navigate to the next set of associations through the ">" button at the right of the counter.

In the *Summary of GDAs* tab, the gene-disease associations are ranked by default according to the DisGeNET GDA score. The data can also be sorted by the number of supporting publications, the number of associated variants, or any other columns in the table.

Finally, by using the ***Filter within the current results*** box, the user can filter the data displayed in any of the columns**. Importantly, this box allows to filter only in the result set displayed in that view, and not within all the results of the query.** This means that if for example, we filter using the box by "Signalling molecule", seven records will be displayed, corresponding to the proteins annotated as signalling molecules within the first 25 proteins. Nevertheless, if we use the ***Add/Remove filter*** panel, 75 proteins will be displayed. See the example below to learn how to use this panel.

Click on the ***Add/Remove filter*** button (Figure 3, #5) to display all the possibilities to filter the results of the initial query. This will display the ***Add/Remove filter*** panel containing several filters that can be applied to the GDAs: Filtering by gene, protein class, score, DSI, DPI, EI, EL, and score. Additionally, the source database can be changed (Figure 4). For example, if you are interested in the genes associated to Alzheimer disease reported only by CTD human data, click on the dropdown menu next to "Source" to display the list of available sources (Figure 4, #1). Associations can also be filtered by score (Figure 4, #2), and by number of publications (Figure 4, #3).

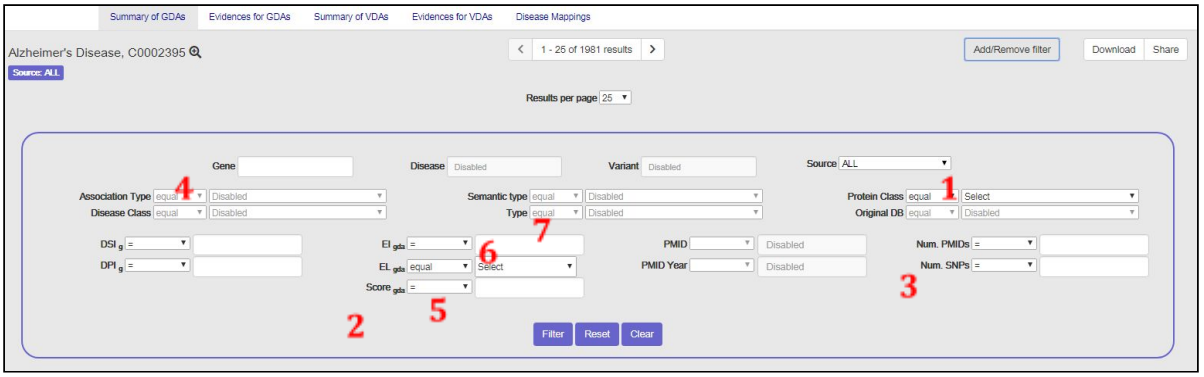


Figure 4: The ***Add/Remove*** filter panel

To explore in more detail the evidences supporting the association between two entities, for instance Alzheimer's Disease and the APP gene, click on the number of PMIDs supporting this association (Figure 3, #8) or use the Gene filter in the ***Add/Remove***

filter panel (Figure 4, #4), and then clicking on the "*Evidences of GDAs*" tab. The results are shown in Figure 5.

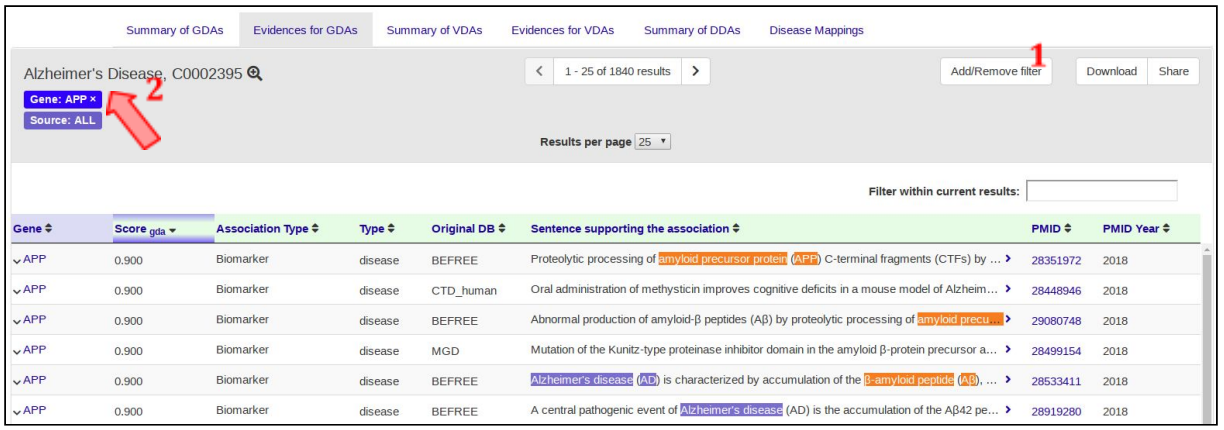


Figure 5: "***Evidences of GDAs***" tab for the APP gene and Alzheimer Disease

In the "*Evidences of GDAs*" tab, there is one line per evidence supporting the association between the gene and the disease. One "evidence" is a publication from a database source with a particular DisGeNET Association Type. 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.
By clicking on the ***Add/Remove filter*** button (Figure 5, #1), several filters can be applied, for example, by source database (e.g. CTD) and association type (e.g. Genetic Variation, indicating that variants in this gene have been found to be associated with Alzheimer's Disease). You can also filter the results to keep only associations published during the last two years using the PMID year filter.
Closing the "Gene: APP" button (Figure 5, #2) will remove the selection on the gene, and will show the evidences linking Alzheimer Disease to all its associated genes for DisGeNET ALL. This action leads to the same results view as clicking on the button "*Evidences for Gene-Disease Associations*" in Figure 1.

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 5, #2).

Finally, click on the ***Download*** button (Figure 3, #6) to download the results of the analysis as a tabulated file or as an excel file.

1.2. How To Retrieve The Genes Associated To Several Diseases At Once?

The search button allows searching for multiple diseases in a single query. For example, Figure 1 shows a preview of the terms that contain the string "Alzheimer" in the database. To search for genes associated with several of the Alzheimer's subtypes at once, select the terms of interest by clicking on them in the dropdown menu (Figure 6).

Once all the terms of interest have been selected, click on the magnifying glass next to the search box to submit the query.

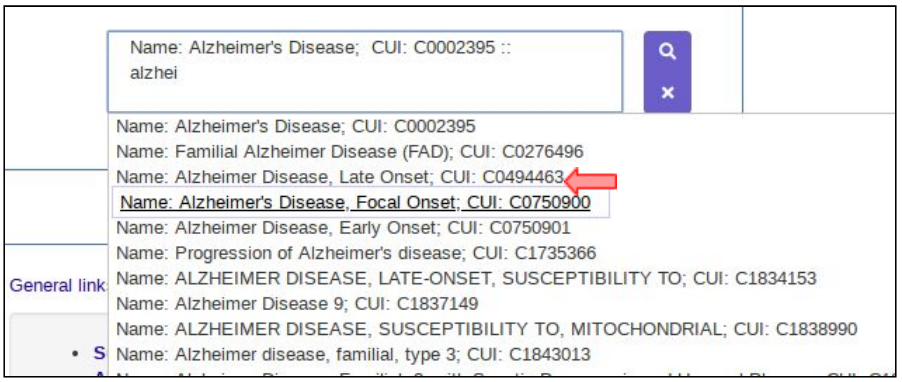


Figure 6: Searching multiple diseases

The results of searching Alzheimer's Disease (CUI: C0002395), Familial Alzheimer Disease (CUI: C0276496), and Alzheimer Disease, Late Onset (CUI: C0494463) are shown in Figure 7. Click on the button "*Summary of Gene-Disease Associations*" to inspect all the results.

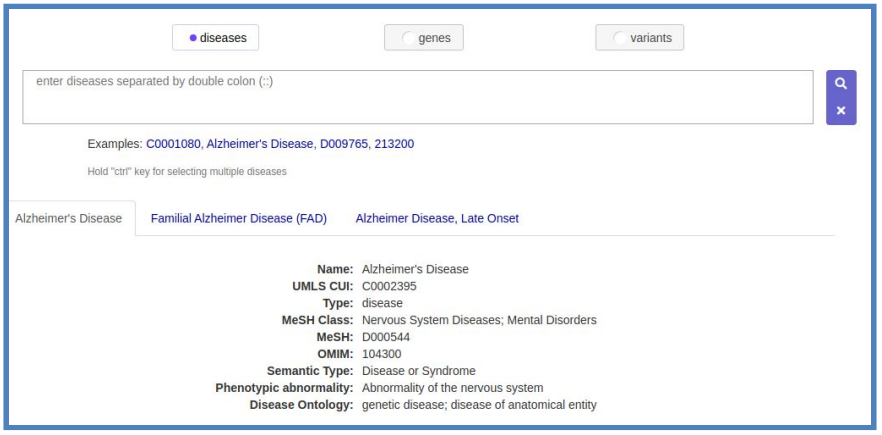


Figure 7: Preview of the results of searching *multiple diseases*

Figure 8 illustrates the "*Summary for the GDAs*" tab for the multiple diseases search. Notice the names of the diseases (1) and the specific Alzheimer type (2) and the associated gene in the first and second column, respectively. Go to "*Evidences for the GDAs*" tab to explore the details of each association. You may then filter, explore, or download your results, as explained above for the individual search.

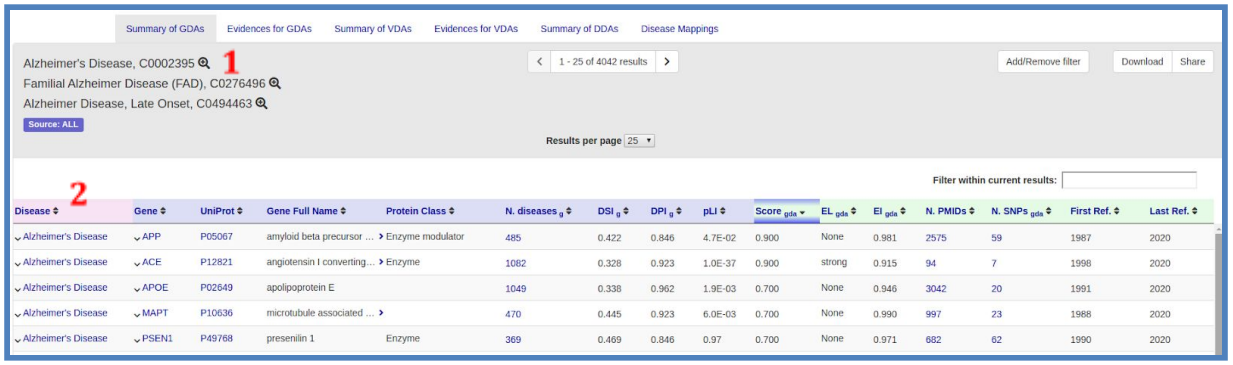


Figure 8: The "***Summary of GDAs***" tab showing several types of Alzheimer's Disease in DisGeNET ALL

1.3. How to find diseases similar to a disease of interest

It is possible to expand the search for a disease with semantically similar concepts. To do this, click on the button “Similar diseases” after the results of searching for a disease are displayed (red box in Figure 9).

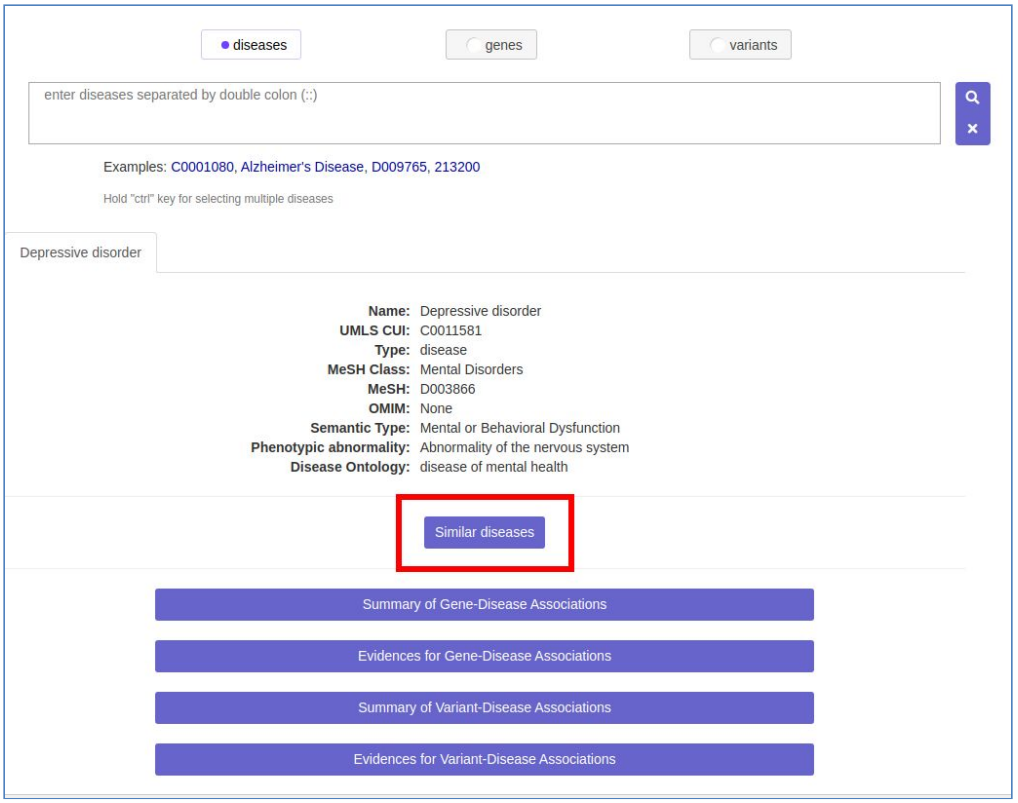


Figure 9: Expanding the search with similar diseases

After clicking the button, the 10 most similar diseases to the query disease (in the example, Depressive Disorder) will be displayed (Figure 10).

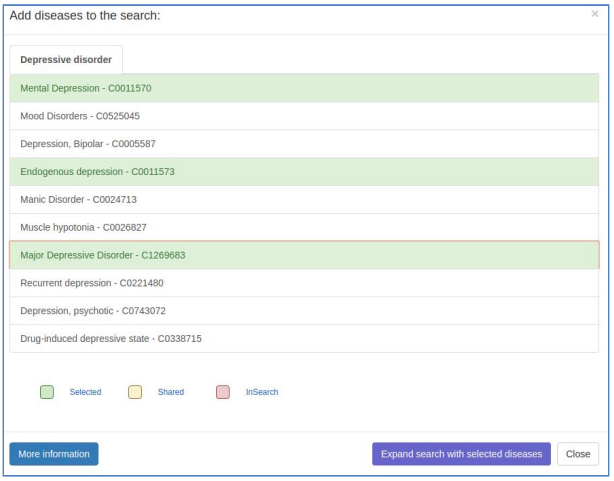


Figure 10: Top 10 diseases similar to Depressive Disorder

You can then select one or more concepts and include them in the search by clicking on the "Expand your search using the selected diseases" button. In this way, the results of your search will include those of your original concept as well as the added concepts. This expanded search applies to all the results (GDAs, VDAs and DDAs). The disease similarity between concepts is computed using the Sokal-Sneath semantic similarity distance (Sánchez et al., 2011) on the taxonomic relations provided by the Unified Medical Language System Metathesaurus (Bodenreider, 2004). Only the relationships of type is-a (which describe the taxonomy in any ontology) are taken into account.

1.4. What diseases are associated with Transcription Factors alterations?

Click on the button "*Summary of All Gene-Disease Associations*" (Figure 9, #1). This will take you to the "*Summary of GDAs*" tab (Figure 10).

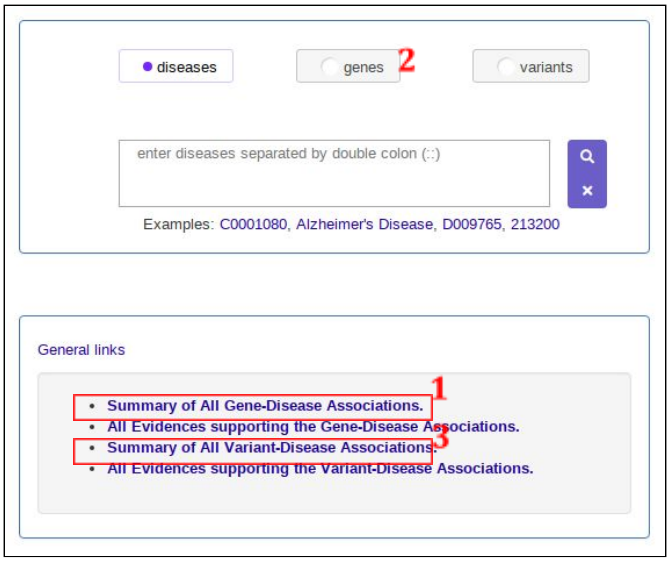


Figure 9: The entry point of the Search

The "*Summary of GDAs*" tab for all data (Figure 10) presents all the gene-disease associations in DisGeNET, ordered by score and number of publications. Notice that the columns are coloured according to the type of entity: attributes of the genes are blue, of the diseases, pink, and of the GDAs are green. Click on the ***Add/Remove filter*** button (Figure 10, #1) to filter the GDAs by protein class. Figure 11 shows all gene-disease associations in DisGeNET ALL involving a transcription factor (50,460 gene-disease associations).

An alternative way to obtain the same results is from the Browse view, selecting as source database ALL, and then on the "*Summary of GDAs*" tab for all data (Figure 10) apply the same filter as mentioned before to select GDAs for genes encoding transcription factors. The results are shown in Figure 11.

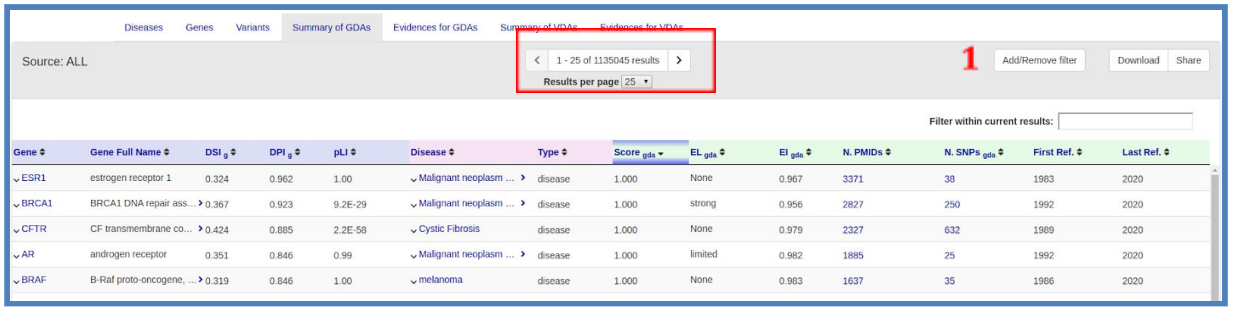


Figure 10: "***The Summary of GDAs***" tab for all gene-disease associations

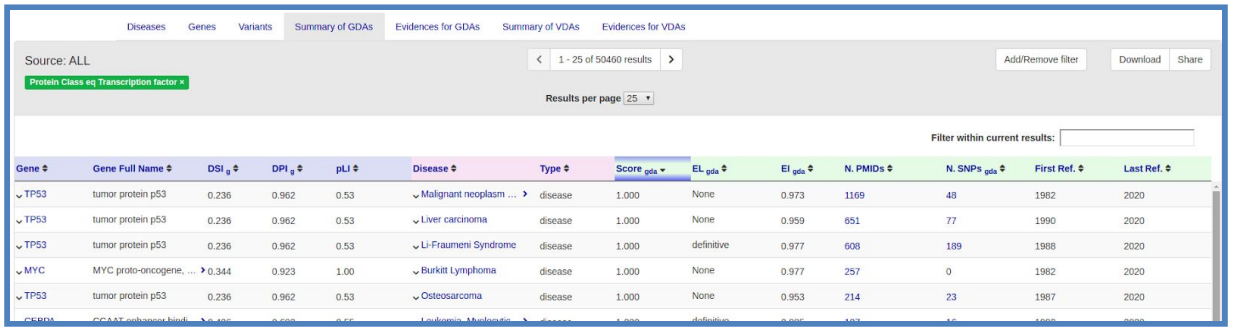


Figure 11: The GDAs involving transcription factors.

You can further filter the results of the query by applying other filters in the ***Add/Remove filter*** panel. For example, to investigate the transcription factors that are involved in diseases of the Endocrine System, filter the results using the dropdown menu corresponding to disease class. The results are displayed in Figure 12.

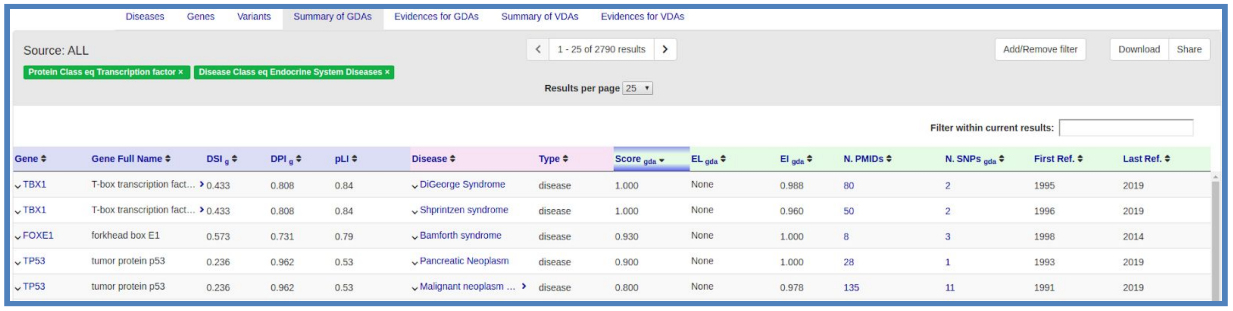


Figure 12: The GDAs involving transcription factors associated with diseases of the endocrine system.

1.5. Are there any diseases other than cancer associated with the BRCA1 gene?

Click on the radio button corresponding to gene (Figure 9, #2) to activate the search by gene. Search the gene BRCA1 and go to the "*Summary of GDAs*" tab. Display the ***Add/Remove filter*** panel (Figure 4) and set the filter type equal "disease", and the filter Semantic type does not contain "Neoplastic Process". The results are shown in Figure 13.

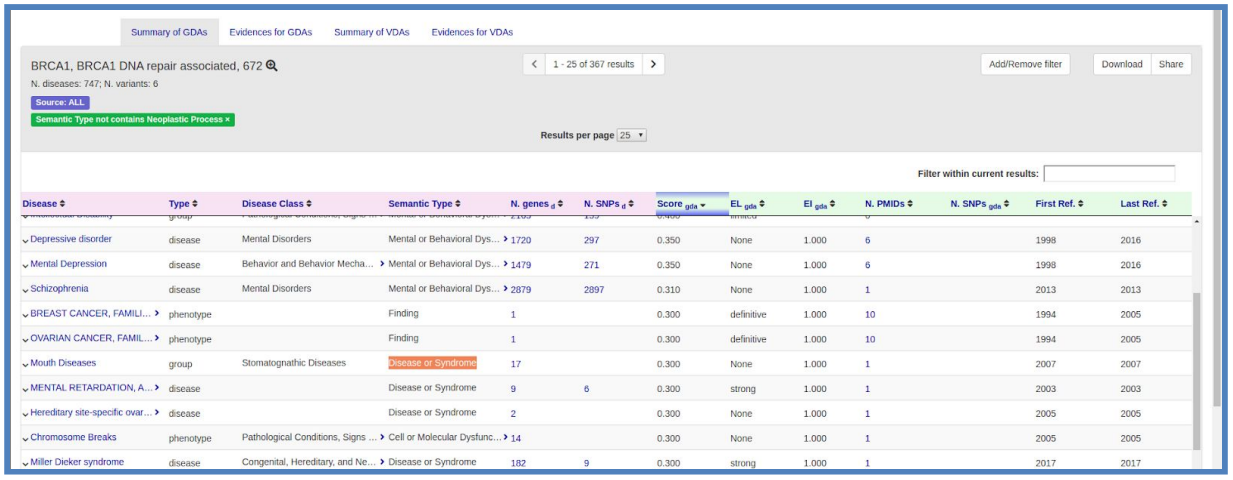
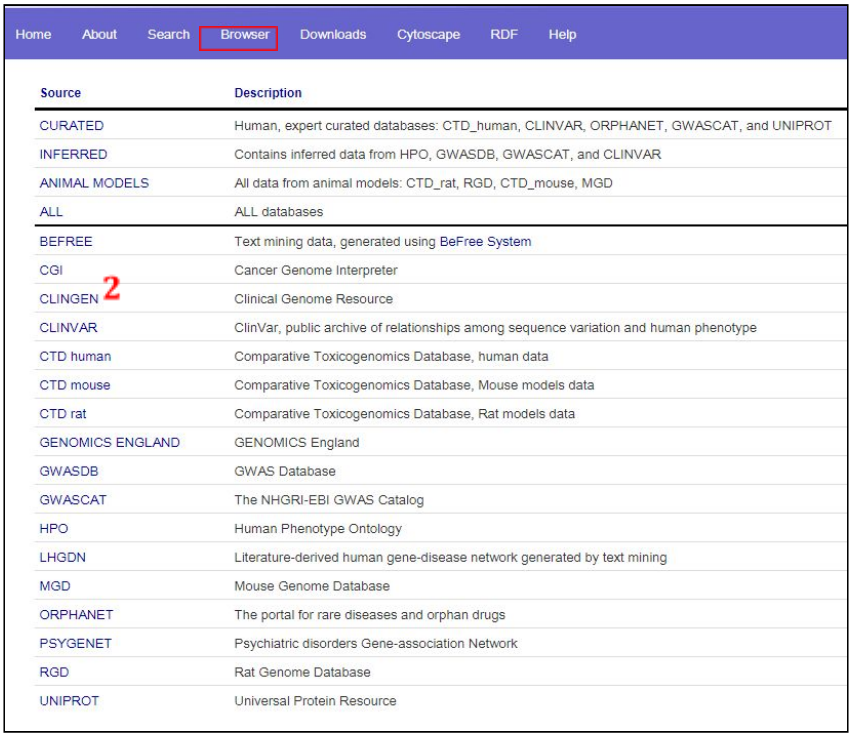


Figure 13: The GDAs involving the gene BRCA1, and diseases that are not neoplastic processes

1.6. What are the GDAs classified as “definitive” by ClinGen?

To explore data from DisGeNET from a specific source, go to the entry point of the Browser (http://disgenet.org/browser/0/0/0/0/_a/_b./) or click on the Browser button (Figure 14, #1). This will show all the sources in DisGeNET. Click on ClinGen to display the data from this database (1260 GDAs). In ClinGen, the GDAs are classified according to the supporting evidences in *definitive, strong, moderate, limited, and disputed.* In DisGeNET we have imported this label for GDAs reported by ClinGen, and from Genomics England panel app, which also annotates associations using similar criteria and we refer to it as the Evidence Level (EL). To display the GDAs from ClinGen, go to the “*Summary of GDAs*” tab and in the ***Add/Remove filter*** panel select *definitive* in the filter EL (Figure 4, #5).



Source	Description
CURATED	Human, expert curated databases: CTD_human, CLINVAR, ORPHANET, GWASCAT, and UNIPROT
INFERRED	Contains inferred data from HPO, GWASDB, GWASCAT, and CLINVAR
ANIMAL MODELS	All data from animal models: CTD_rat, RGD, CTD_mouse, MGD
ALL	ALL databases
BEFREE	Text mining data, generated using BeFree System
CGI	Cancer Genome Interpreter
CLINGEN	Clinical Genome Resource
CLINVAR	ClinVar, public archive of relationships among sequence variation and human phenotype
CTD human	Comparative Toxicogenomics Database, human data
CTD mouse	Comparative Toxicogenomics Database, Mouse models data
CTD rat	Comparative Toxicogenomics Database, Rat models data
GENOMICS ENGLAND	GENOMICS England
GWASDB	GWAS Database
GWASCAT	The NHGRI-EBI GWAS Catalog
HPO	Human Phenotype Ontology
LHGDN	Literature-derived human gene-disease network generated by text mining
MGD	Mouse Genome Database
ORPHANET	The portal for rare diseases and orphan drugs
PSYGENET	Psychiatric disorders Gene-association Network
RGD	Rat Genome Database
UNIPROT	Universal Protein Resource

Figure 14: The Browser view

The results (571 GDAs) are shown in Figure 15.

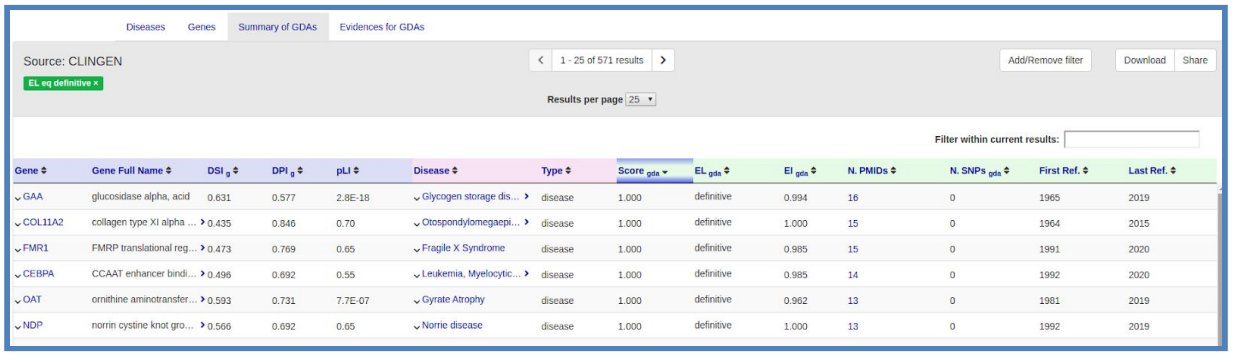


Figure 15: ClinGen GDAs with EL definitive

1.7. What are the variants associated with Long QT Syndrome?

Type "Long QT Syndrome" in the search box, using the disease centric search (Figure 16). Click on the button "*Summary of Variant-Disease Associations*". This will lead to the "*Summary of VDAs*" tab, with the information of the SNPs associated with the disease (Figure 17).

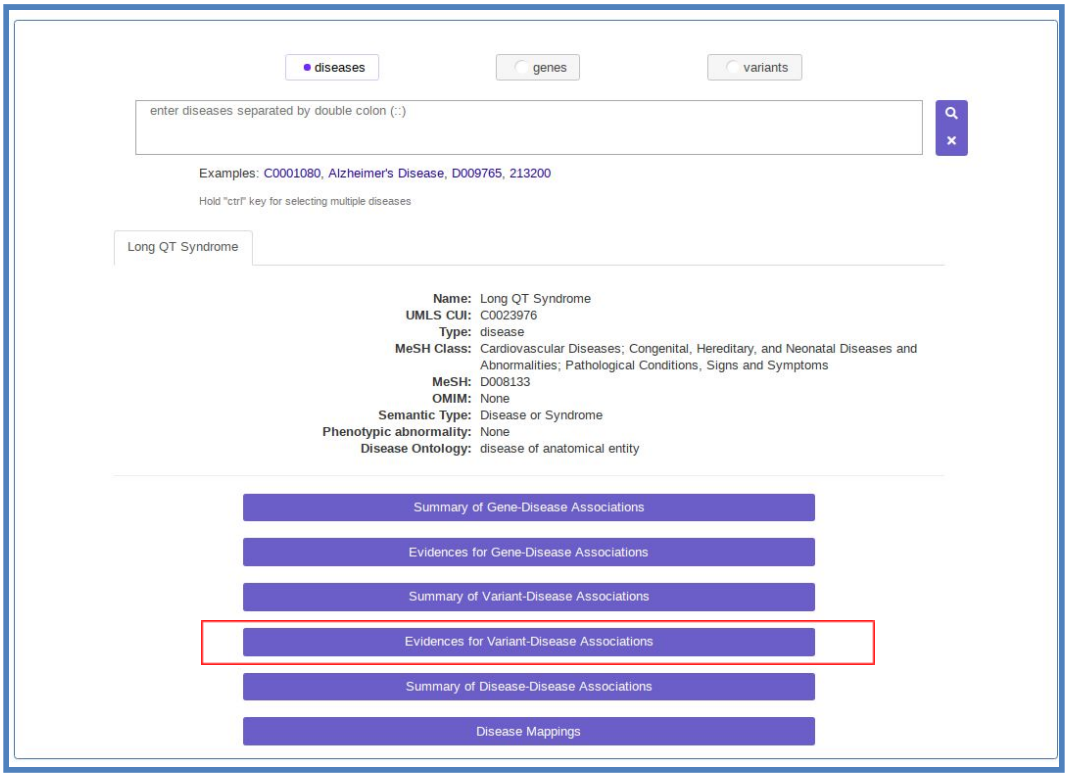


Figure 16: The disease centric view: how to retrieve the variants associated to Long QT Syndrome

The "*Summary of VDAs*" tab also includes information about the gene, the associated gene(s), its position, the most severe consequence type according to the Variant Effect Predictor, the reference and alternative alleles. Also, it will show the variant class: SNP, deletion, insertion, indel, somatic SNV, substitution, sequence alteration, or tandem repeat. We also show the allelic frequencies in the gnomAD exomes and genomes. Additionally, the gene, the score of the VDA, and the number of papers reporting the association are also included in this tab (Figure 17).

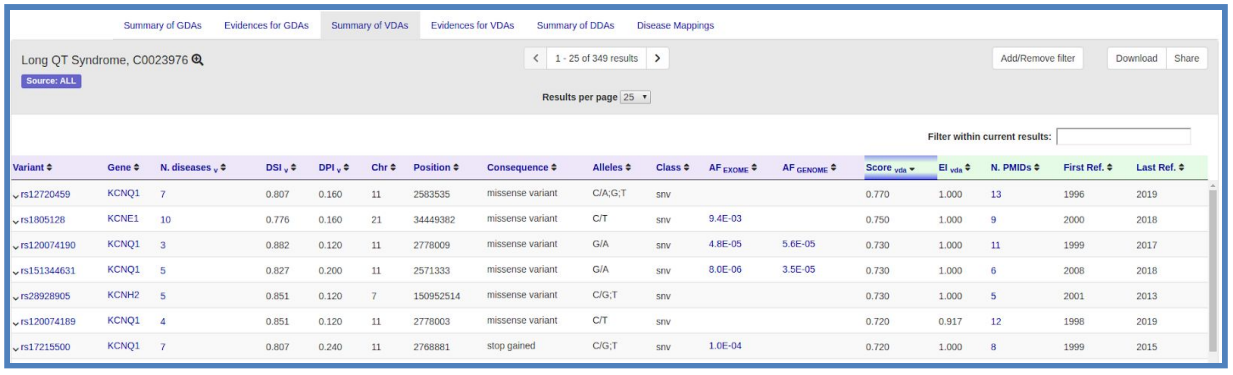
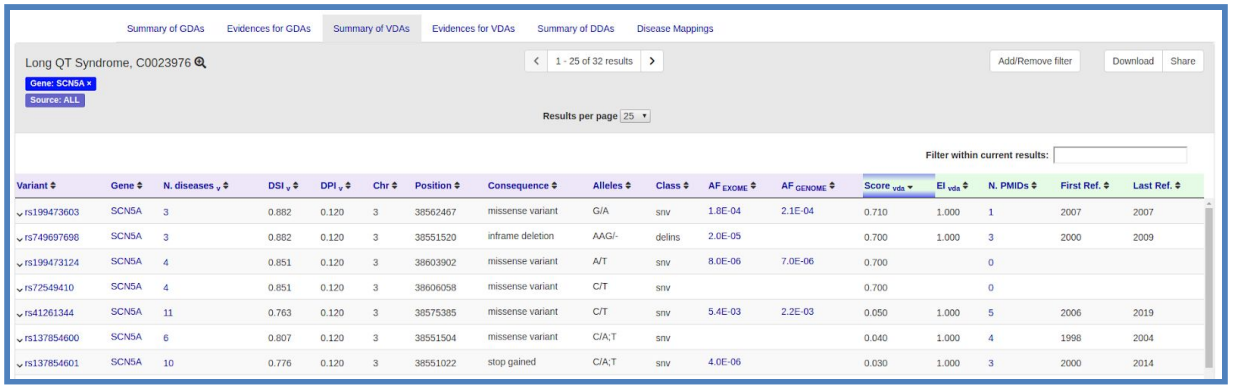


Figure 17: The "***Summary of VDAs***" tab for Long QT Syndrome

You can filter to keep VDAs associated to a specific gene. Go to the ***Add/Remove filter*** button and click on the Gene filter (Figure 4, #4). Filter by the gene SCN5A. The results are shown in Figure 18.



Variant	Gene	N. diseases	DSI	DPI	Chr	Position	Consequence	Alleles	Class	AF EXOME	AF GENOME	Score vda	EI vda	N. PMIDs	First Ref.	Last Ref.
rs199473603	SCN5A	3	0.882	0.120	3	38562467	missense variant	G/A	snv	1.8E-04	2.1E-04	0.710	1.000	1	2007	2007
rs749697698	SCN5A	3	0.882	0.120	3	38551520	inframe deletion	AAG/-	delins	2.0E-05		0.700	1.000	3	2000	2009
rs199473124	SCN5A	4	0.851	0.120	3	38603902	missense variant	A/T	snv	8.0E-06	7.0E-06	0.700		0		
rs72549410	SCN5A	4	0.851	0.120	3	38606058	missense variant	C/T	snv			0.700		0		
rs41261344	SCN5A	11	0.763	0.120	3	38575385	missense variant	C/T	snv	5.4E-03	2.2E-03	0.050	1.000	5	2006	2019
rs137854600	SCN5A	6	0.807	0.120	3	38551504	missense variant	C/A;T	snv			0.040	1.000	4	1998	2004
rs137854601	SCN5A	10	0.776	0.120	3	38551022	stop gained	C/A;T	snv	4.0E-06		0.030	1.000	3	2000	2014

Figure 18: The “Summary of VDAs” tab for Long QT Syndrome and gene SCN5A

1.8. What congenital diseases are associated with variants producing a stop codon?

From the entry point of the search (Figure 9, #2), click on the link "*Summary of All Variant-Disease Associations*". Click on the ***Add/Remove filter*** button (Figure 3, #5) to display the ***Add/Remove filter*** panel. The panel is presented in Figure 19. Notice that this panel is very similar to the one corresponding to the *Summary of GDAs* tab (Figure 4), but there are some fields active that were disabled in the previous one. Select the Consequence equal "stop gained" and Disease Class equal "Congenital, Hereditary, and Neonatal Diseases and Abnormalities" (Figure 19, #1 and #2 respectively). The results are shown in Figure 20.

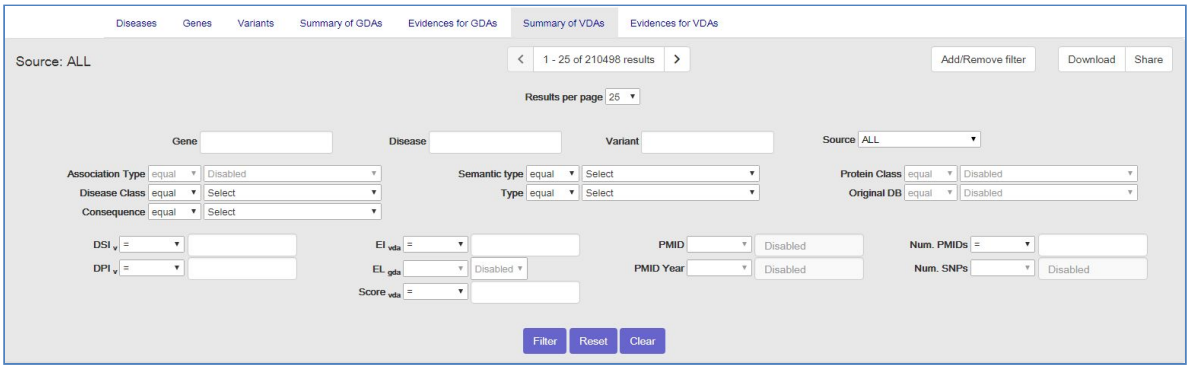


Figure 19: The ***Add/Remove*** filter panel for the "***Summary of VDAs***" tab

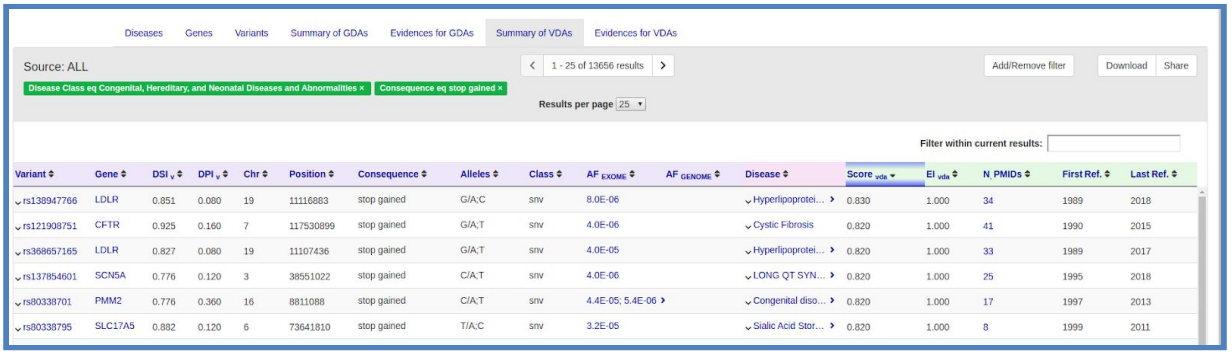


Figure 20: The Congenital, Hereditary, and Neonatal Diseases and Abnormalities caused by variants producing a stop codon

1.9. What are the diseases associated to Autistic Disorder in DisGeNET CURATED?

Search *Autistic Disorder* in DisGeNET. Click on the button "*Summary of Disease-Disease Associations*". Once the results are displayed, click on the ***Add/Remove filter*** button to display the panel and select source equal "CURATED". The results are shown in Figure 21.

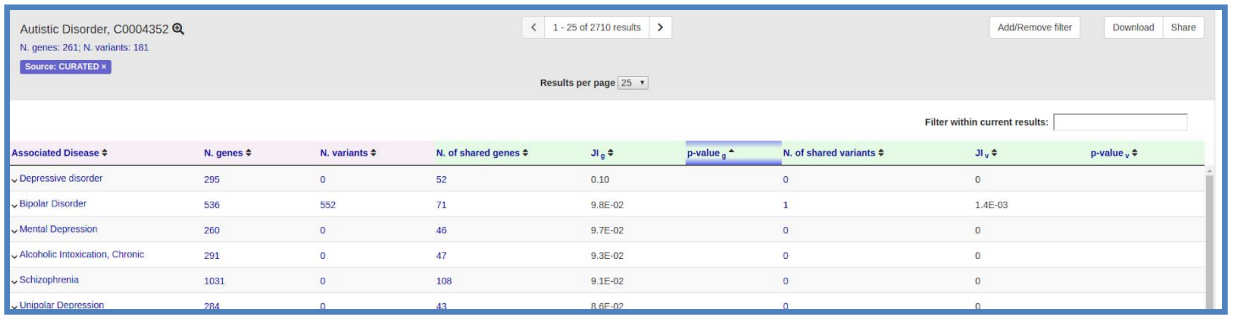


Figure 21: The diseases that share genes with Autistic Disorder in DisGeNET CURATED

2. The DisGeNET database

DisGeNET is a discovery platform containing one of the largest publicly available collections of genes and variants associated with human diseases [1–2]. DisGeNET integrates data from expert curated repositories, GWAS catalogues, animal models and the scientific literature. DisGeNET data are homogeneously annotated with controlled vocabularies and community-driven ontologies. Additionally, several original metrics are provided to assist the prioritization of genotype–phenotype relationships. Data integration in DisGeNET is performed by means of gene and disease vocabulary mapping and by using the DisGeNET gene-disease association ontology as described below (2.3).

2.1. Original data sources

In DisGeNET, the data is grouped according to their type and level of curation: CURATED (containing gene-disease associations from human expert curated data sources), PREDICTED (containing gene-disease associations from animal model repositories), INFERRED (GDAs inferred from HPO and from VDAs), and ALL (containing CURATED, PREDICTED, INFERRED, 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

2.2. 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”.

2.3. DisGeNET gene-disease association type ontology

We developed the DisGeNET gene-disease association ontology to represent in a uniform and structured way the types of relations between genes and diseases found in the original data sources. For the up-to-date version of the ontology used to describe gene-disease associations in DisGeNET please visit the DisGeNET Discovery Platform Website at: http://disgenet.org/dbinfo, section “The DisGeNET Association Type Ontology”.

3. 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:

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

- 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].

4. 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. Bauer-Mehren, A., Rautschka, M., Sanz, F., Furlong, L.I.: DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. Bioinformatics. 26, 2924–6 (2010).
3. Sánchez D1, Batet M.J: Semantic similarity estimation in the biomedical domain: an ontology-based information-theoretic perspective.Biomed Inform. 2011 Oct;44(5):749-59. doi: 10.1016/j.jbi.2011.03.013.
4. Bodenreider O: The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D267-70

5. 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

6. Funding

We received support from 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]; Agència de Gestió d'Ajuts Universitaris i de Recerca Generalitat de Catalunya [2017SGR00519]; Generalitat de Catalunya. Convocatoria Emergents FEDER. 2019-2021[Proyecto VEIS: Valorización de EGA para la Industria y la Sociedad]. 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].

7. 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://grib.imim.es/research/integrative-biomedical-informatics/.

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/.
For further details on DisGeNET license, please consult the page:
http://disgenet.org/legal.

8. About this document