

DisGeNET Web Interface

IBI Lab
[2019]

USER GUIDE

© 2010-2019, Integrative Biomedical Informatics

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 plugin is distributed under the GNU GPL 3.0 license. More details about the GNU General Public License 3.0 is available [here](#).

1.	The DisGeNET Web Interface	5
1.1.	What Genes Are Associated To Alzheimer Disease?	5
1.2.	How To Retrieve The Genes Associated To Several Diseases At Once?	8
1.3.	What Diseases Are Associated To Transcription Factors?	9
1.4.	Are there any diseases other than cancer associated to BRCA1 gene?	11
1.5.	What are the GDAs classified as “definitive” by ClinGen?	11
1.6.	What are the variants associated to Long QT Syndrome?	12
1.7.	What congenital diseases are associated to variants producing a stop codon?	14
2.	The DisGeNET database	15
2.1.	Original data sources	15
2.2.	Generation of gene-disease networks	15
2.3.	Vocabulary mapping	16
2.4.	DisGeNET gene-disease association type ontology	16
3.	Citation	17
4.	References	18
5.	Contact	19
6.	Funding	20
7.	License	21
8.	About this document	21

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 To Alzheimer 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 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.**

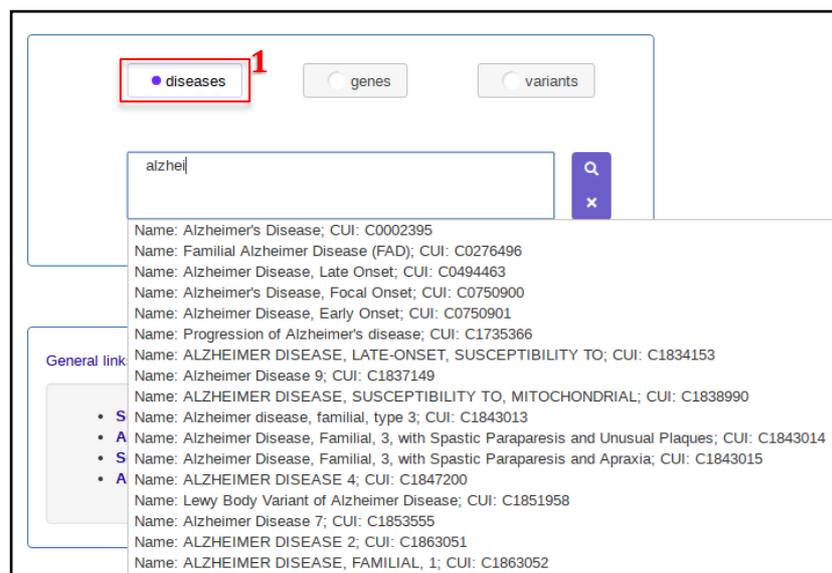


Figure 1: The search view, disease centric search.

Once you click on the magnifying glass icon, you will see a preview box summarizing the information on the entity of your query (“Alzheimer 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 Disease” D000544, 104300, and C0002395, respectively).

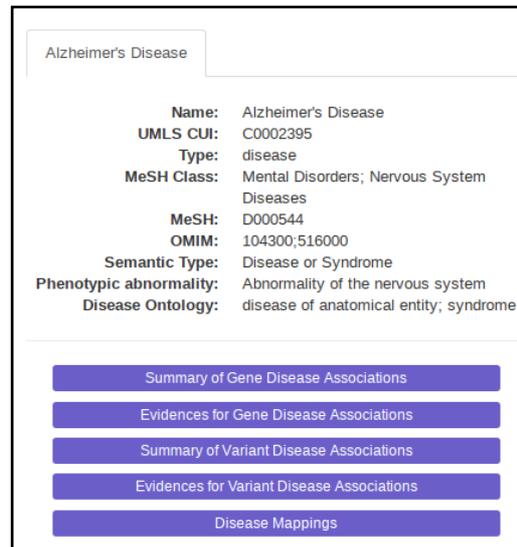


Figure 2: The preview of the search for Alzheimer Disease (C0002395)

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).

Summary of GDAs | Evidences for GDAs | Summary of VDAs | Evidences for VDAs | Disease Mappings

Alzheimer's Disease, C0002395

1 - 25 of 1981 results

Source: ALL

Results per page: 25

Gene	Uniprot	Gene Full Name	DPI _g	DSI _g	Protein Class	Score _{gda}	EL _{gda}	EI _{gda}	Num. PMIDs	Num. SNPs	Year initial	Year final
APP	P05067	amyloid beta precursor protein	0.862	0.430	enzyme modulator	0.900	no reported evidence	0.967	1764	11	1967	2018
APOE	P02649	apolipoprotein E	0.931	0.352		0.700	no reported evidence	0.922	2356	11	1991	2018
PSEN1	P49768	presenilin 1	0.690	0.490	calcium-binding prot...	0.700	no reported evidence	0.939	535	40	1990	2018
SORL1	Q92673	sortilin related receptor 1	0.483	0.619	receptor; transporter	0.700	no reported evidence	0.986	79	12	2004	2018
BCL2	P10415	BCL2, apoptosis regulator	0.862	0.312	signaling molecule	0.600	no reported evidence	1.000	15	0	1996	2016

Figure 3: “The Summary of GDAs” tab for the search using “Alzheimer Disease” (C0002395) in DisGeNET ALL

This view presents the results of the search from different perspectives (the tabs, Figure 3, #1), and allows to perform 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 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 to Alzheimer Disease. The counter indicates the total

number of genes associated to the disease (Figure 3, #3). There are 1,981 genes associated with “Alzheimer 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.

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 (#2), and by number of publications (#3).

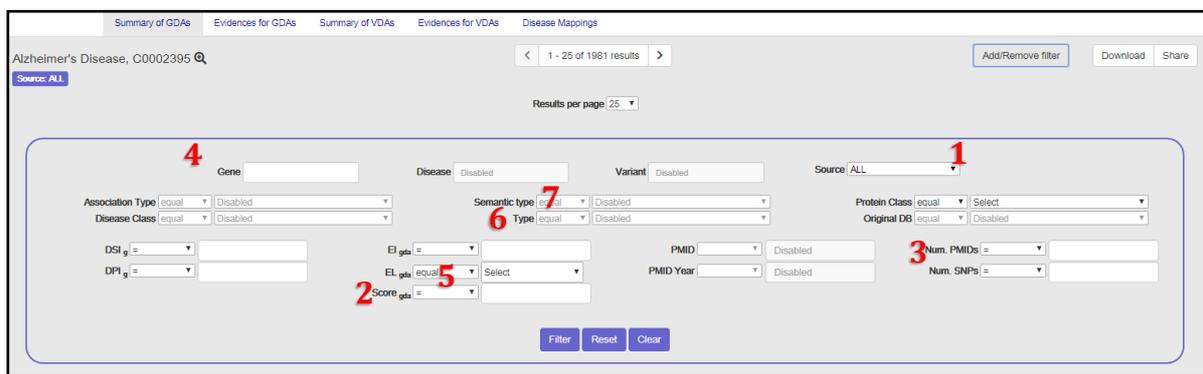


Figure 4: The **Add/Remove filter** panel

To explore in more detail the evidences supporting the association between two entities, for instance Alzheimer 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.

Gene	Score_gda	Association Type	Type	Original DB	Sentence supporting the association	PMID	PMID Year
APP	0.900	GeneticVariation	disease	BEFREE	APP transgenic mutants have also been investigated with respect to survival rates, neurologic functions, an...	23089603	2012
APP	0.900	AlteredExpression	disease	BEFREE	APP overexpression in the absence of NPC1 exacerbates metabolism of amyloidogenic proteins of Alzhei...	26433932	2016
APP	0.900	GeneticVariation	disease	BEFREE	The amyloid precursor protein (APP) locus on chromosome 21 influences the development of Alzheimer ds...	15184603	2004
APP	0.900	Biomarker	disease	BEFREE	Activated microglial cells are an integral component of fibrillar plaques in brains of subjects with Alzheimer's...	12624793	2003
APP	0.900	Biomarker	disease	BEFREE	Accordingly, transgenic mice expressing mutant human amyloid precursor protein (APP(SL)) serving as a ...	25742870	2015

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 of 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 to 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 to several of the Alzheimer 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.

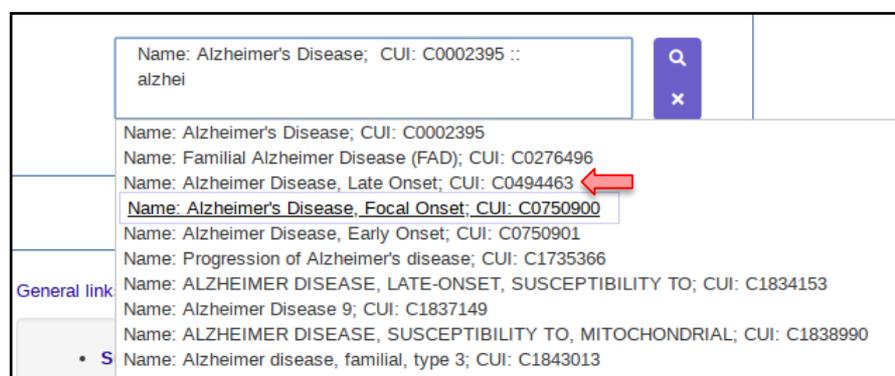


Figure 6: Searching multiple diseases

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.

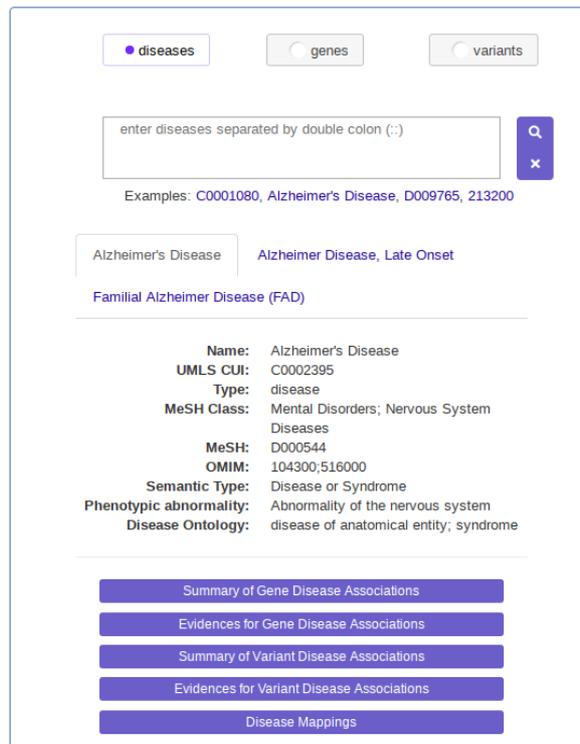


Figure 7: Preview of the results of searching *multiple diseases*

Figure 8 illustrates the results of 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 (3) to explore the details of each association. You may then filter, explore, or download your results, as explained above for the individual search.

Disease	Gene	Uniprot	Gene Full Name	DPI	DSI	Protein Class	Score	EL	Num. PMIDs	Num. SNPs	Year added	Year final	
Alzheimer's Disease	APP	P05067	amyloid beta precursor p...	0.862	0.430	enzyme modulator	0.900	no reported evid...	0.967	1764	39	1987	2018
Alzheimer's Disease	APOE	P02649	apolipoprotein E	0.931	0.352		0.700	no reported evid...	0.922	2356	11	1991	2018
Alzheimer's Disease	PSEN1	P49768	presenilin 1	0.690	0.490	calcium-binding p...	0.700	no reported evid...	0.939	535	40	1990	2018
Alzheimer's Disease	SCRL1	Q82673	sortilin related receptor 1	0.483	0.619	receptor; transpo...	0.700	no reported evid...	0.886	79	12	2004	2018
Alzheimer's Disease	BCL2	P10415	BCL2, apoptosis regulator	0.862	0.312	signaling molecule	0.600	no reported evid...	1.000	15	0	1996	2016

Figure 8: The *“Summary of GDAs”* tab showing several types of Alzheimer Disease in DisGeNET ALL

1.3. What Diseases Are Associated To Transcription Factors?

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).

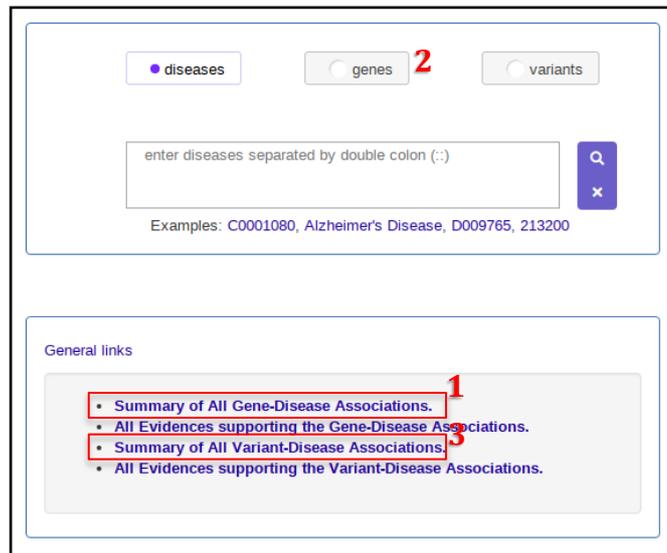


Figure 9: The entry point of the Search

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. 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 (48653 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.

Gene	Gene Full Name	DPI _g	DSI _g	Disease	Type	Score _g	EL _{gda}	EL _{gda}	Num. PMIDs	Num. SNPs	Year initial	Year final
WRN	Werner syndrome RecQ like ...	0.586	0.545	Werner Syndrome	disease	1.000	no reported evidence	0.915	140	32	1993	2017
SLC39A4	solute carrier family 39 memb...	0.517	0.652	Acrodermatitis enteropathica	disease	1.000	no reported evidence	1.000	19	9	2000	2014
RUNX2	runt related transcription facto...	0.621	0.490	Cleidocranial Dysplasia	disease	1.000	no reported evidence	0.975	96	15	1993	2018
FBN1	fibrillin 1	0.690	0.469	Marfan Syndrome	disease	1.000	definitive	0.957	412	375	1973	2018
MYO5B	myosin VB	0.552	0.642	Microvillus inclusion disease	disease	1.000	no reported evidence	1.000	16	8	2008	2018
COL7A1	collagen type VII alpha 1 chain	0.586	0.552	Hallopeau-Siemens Disease	disease	1.000	no reported evidence	0.961	95	35	1993	2018

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

Gene	Gene Full Name	DPI _g	DSI _g	Disease	Type	Score _g	EL _{gda}	EL _{gda}	Num. PMIDs	Num. SNPs	Year initial	Year final
RUNX2	runt related transcription facto...	0.621	0.490	Cleidocranial Dysplasia	disease	1.000	no reported evidence	0.975	96	15	1993	2018
FOXL2	forkhead box L2	0.448	0.561	BLEPHAROPHIMOSIS, PTOS...	disease	1.000	no reported evidence	1.000	42	56	1993	2016
THRB	thyroid hormone receptor beta	0.690	0.584	Thyroid Hormone Resistance ...	disease	1.000	strong	0.989	98	10	1989	2017
WT1	Wilms tumor 1	0.690	0.444	Denys-Drash Syndrome	disease	1.000	no reported evidence	0.960	57	13	1983	2017
CEBPA	CCAAT enhancer binding pro...	0.586	0.533	Leukemia, Myelocytic, Acute	disease	1.000	no reported evidence	0.987	160	11	1993	2018

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.

Gene	Gene Full Name	DPI	DSI	Disease	Type	Score	EL	EI	Num. PMIDs	Num. SNPs	Year initial	Year final
THRB	thyroid hormone receptor beta	0.690	0.584	Thyroid Hormone Resistance Syndrome	disease	1.000	strong	0.989	98	10	1989	2017
WT1	Wilms tumor 1	0.690	0.444	Denys-Drash Syndrome	disease	1.000	no reported evidence	0.960	57	13	1983	2017
TBX1	T-box 1	0.724	0.479	Shprintzen syndrome	disease	1.000	no reported evidence	0.977	49	2	1996	2017
AR	androgen receptor	0.793	0.380	Androgen-Insensitivity Syndrome	disease	1.000	no reported evidence	0.940	235	50	1970	2018
GATA3	GATA binding protein 3	0.793	0.477	Barakat syndrome	disease	1.000	no reported evidence	0.970	39	8	2000	2017

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

1.4. Are there any diseases other than cancer associated to 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 to not equal to “Neoplastic Process”. The results are shown in Figure 13.

Disease	Type	Disease Class	Semantic Type	Score _{gda}	EL _{gda}	EI _{gda}	Num. PMIDs	Num. SNPs	Year initial	Year final
Depressive disorder	disease	Mental Disorders	Mental or Behavioral Dy...	0.340	no reported evidence	1.000	5		1998	2013
Mental Depression	disease	Behavior and Behavior Mechanisms	Mental or Behavioral Dy...	0.340	no reported evidence	1.000	5		1998	2013
Schizophrenia	disease	Mental Disorders	Mental or Behavioral Dy...	0.300	no reported evidence	1.000	1		2014	2014
Congenital anemia	disease		Disease or Syndrome	0.300	moderate		3		2015	2017
FANCONI ANEMIA, COMPLEMENTATION GR...	disease		Disease or Syndrome	0.200	no reported evidence	0.941	34	2	2003	2017

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

1.5. 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 (518 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 2	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 (234 GDAs) are shown in Figure 15.

Gene	Gene Full Name	DP1 _g	DSI _g	Disease	Type	Score _g	EL _{gda}	EI _{gda}	Num. PMIDs	Num. SNPs	Year _{first}	Year _{last}
TP53	tumor protein p53	0.897	0.251	Li-Fraumeni Syndrome	disease	1.000	definitive	0.919	8	0	1989	2018
PTPN11	protein tyrosine phosphatase...	0.828	0.429	Noonan Syndrome	disease	1.000	definitive	0.957	12	0	2001	2018
EXT1	exostosin glycosyltransferase 1	0.655	0.577	Hereditary Multiple Exostoses	disease	1.000	definitive	0.969	7	0	1995	2018
BRAF	B-Raf proto-oncogene, serine...	0.793	0.352	Cardio-facio-cutaneous syndro...	disease	1.000	definitive	0.957	6	0	1993	2017
PKD1	polycystin 1, transient recepto...	0.621	0.537	Polycystic Kidney, Autosomal ...	disease	1.000	definitive	0.972	7	0	1989	2018

Figure 15: ClinGen GDAs with EL definitive

1.6. What are the variants associated to 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 to the disease (Figure 17).

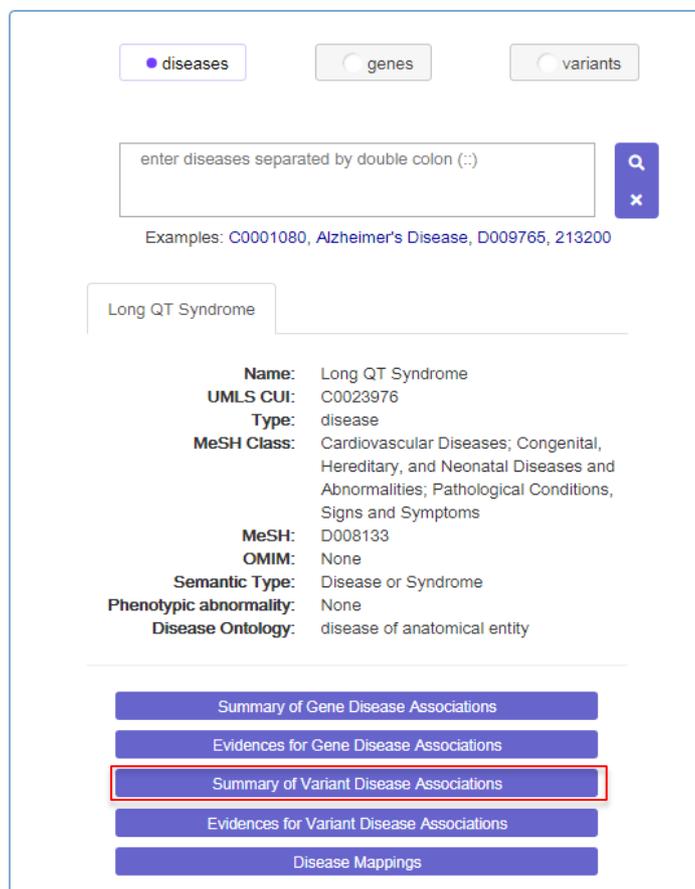


Figure 16: The disease centric view: how to retrieve the variants associated to Long QT Syndrome

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).

Variant	Gene	DPI _v	DSI _v	Chr	Position	Consequence	Alleles	Class	AF _{EXOME}	AF _{GENOME}	SCORE _{vda}	EI _{vda}	Num. PMIDs	Year _{initial}	Year _{final}
rs12720459	KCNQ1	0.107	0.821	11	2583535	missense variant	C/A,G,T	snp			0.730	1.000	7	1999	2017
rs151344631	KCNQ1	0.179	0.821	11	2571333	missense variant	G/A	snp	8.0E-06	3.2E-05	0.720	1.000	2	2008	2015
rs17215500	KCNQ1	0.286	0.801	11	2768881	stop gained	C/G,T	snp	1.0E-04		0.720	1.000	6	2000	2017
rs199472918	KCNH2	0.107	0.769	7	150951738	missense variant	A/G	snp	7.2E-05	6.4E-05	0.720	1.000	9	2000	2017
rs9333849	KCNH2	0.179	0.769	7	150951679	missense variant	C/A,G,T	snp			0.720	1.000	7	2000	2017

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	DPI	DSI	Chr	Position	Consequence	Alleles	Class	AF EXOME	AF GENOME	Score	EI	Num. PMIDs	Year initial	Year final
rs199473133	SCN5A	0.107	0.801	3	38603747	missense variant	G/A	snp	3.1E-05	9.6E-05	0.710	1.000	1	2003	2003
rs72549410	SCN5A	0.107	0.784	3	38606058	missense variant	C/T	snp			0.700		1	2016	2016
rs749697698	SCN5A	0.107	0.878	3	38551520	inframe deletion	GAGAG	in-del	2.0E-05		0.700		3	2000	2010
rs137654600	SCN5A	0.107	0.769	3	38551504	missense variant	C/A,T	snp			0.040	1.000	4	1998	2004
rs137854601	SCN5A	0.107	0.756	3	38551022	missense variant	C/A,T	snp	4.0E-06		0.020	1.000	2	2008	2015

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

1.7. What congenital diseases are associated to 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.

Source: ALL

Results per page: 25

Gene: Disease: Variant: Source: ALL

Association Type: equal | Disabled

Disease Class: equal | Select

Consequence: equal | Select

Semantic type: equal | Select

Type: equal | Select

Protein Class: equal | Disabled

Original DB: equal | Disabled

DSI: EI: PMID: Disabled

DPI: EL: Disabled

Score: PMID Year: Disabled

Num. PMIDs: Num. SNPs: Disabled

Filter Reset Clear

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

Variant	Gene	DPI	DSI	Chr	Position	Consequence	Alleles	Class	AF EXOME	AF GENOME	Disease	Score	EI	Num. PMIDs	Year initial	Year final
rs28933068	FOF3	0.107	0.784	4	1805644	C/A,G,T	snp	1.6E-05			Hypochondroplasia (disorder)	0.900	0.952	29	1973	2016
rs78655421	CFTR	0.214	0.801	7	117530975	G/A,C,T	snp	1.5E-03; 1.2... >	1.1E-03; 3.2... >		Cystic Fibrosis	0.900	0.882	69	1990	2017
rs77932196	CFTR	0.107	0.878	7	117540270	G/A,C,T	snp	2.4E-05; 2.4... >	3.2E-05		Cystic Fibrosis	0.860	0.833	15	1991	2017
rs104894834	GLA; RPL8A...	0.143	0.923	X	101403846	G/A	snp				Fabry Disease	0.820	1.000	8	1992	2014
rs121434286	CLN3	0.107	1.000	16	28482500	C/A,T	snp	2.4E-05	3.2E-05		Juvenile Neuronal Ceroid Lipo... >	0.820	1.000	6	1993	2012

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

2. The DisGeNET database

DisGeNET is a discovery platform containing one of the largest publicly available collections of genes and variants associated to human diseases [1–3]. 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.

The current version of DisGeNET (v6.0) contains 628,685 gene-disease associations (GDAs), between 17,549 genes and 24,166 diseases, disorders, traits, and clinical or abnormal human phenotypes, and 210,498 variant-disease associations (VDAs), between 117,337 variants and 10,358 diseases, traits, and phenotypes. The integration is performed by means of gene and disease vocabulary mapping and by using the DisGeNET gene-disease association ontology as described below (2.4).

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. Generation of gene-disease networks

Gene-disease associations were collected from several sources. The source databases use different vocabularies. In order to merge all gene-disease associations and to present them in one comprehensive gene-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 our gene-disease association ontology (see section 2.4). We also constructed different gene-disease networks depending on the original data source (only containing original data), and according how the data is organized in DisGeNET: CURATED, PREDICTED, and ALL (containing all gene-disease associations). All gene-disease networks are represented as bipartite graphs. A bipartite graph has two types of vertices and the edges run only between vertices of un-like types (Newman, 2003). The bipartite graphs are multigraphs in which two vertices can be connected by more than one edge. In our networks, the multiple edges represent the multiple evidences reporting the gene-disease association.

2.3. 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”.

2.4. 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:

Piñero J, Bravo A, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, Furlong LI. **DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants** *Nucleic Acids Research* (2017) doi:10.1093/nar/gkw943

Piñero J, Queralt-Rosinach N, Bravo À, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F, Furlong LI. **DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes.** *Database* (2015) doi:10.1093/database/bav028

- ❖ 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 plugin:

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, Bundschuh 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 v6.0 (<http://www.disgenet.org/>), Integrative Biomedical Informatics Group, GRIB/IMIM/UPF . [Month, year of data retrieval].

4. References

1. Piñero, J., Queralt-Rosinach, N., Bravo, A., Deu-Pons, J., Bauer-Mehren, A., Baron, M., Sanz, F., Furlong, L.I.: DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database. 2015, bav028–bav028 (2015).
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. Bauer-Mehren, A., Bundschuh, M., Rautschka, M., Mayer, M.A., Sanz, F., Furlong, L.I.: Gene-Disease Network Analysis Reveals Functional Modules in Mendelian, Complex and Environmental Diseases. PLoS One. 6, 13 (2011).

5. Contact

© 2010-2019, 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://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](mailto:support@disgenet.org)

6. Funding

We received support from ISCIII-FEDER (PI13/00082, PI17/00230, CPII16/00026), IMI-JU under grants agreements no. 116030 (TransQST) and no. 777365 (eTRANSAFE) resources of which are composed of financial contribution from the EU-FP7 (FP7/2007-2013) and EFPIA companies in kind contribution, and the EU H2020 Programme 2014-2020 under grant agreements no. 634143 (MedBioinformatics) and no. 676559 (Elixir-Excelerate). The Research Programme on Biomedical Informatics (GRIB) is a member of the Spanish National Bioinformatics Institute (INB), PRB2-ISCIII and is supported by grant PT13/0001/0023, of the PE I+D+i 2013-2016, funded by ISCIII and FEDER. The DCEXS is a “Unidad de Excelencia María de Maeztu”, funded by the MINECO (ref: 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

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/> or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.