

Yale University

From the Selected Works of Rolando Garcia-Milian

Spring April, 2016

Making Sense of Genomic Variation: Part 1 SNP Annotation

Rolando Garcia-Milian



This work is licensed under a [Creative Commons CC_BY-NC-ND International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).



Available at: https://works.bepress.com/rolando_garciamilian/11/

Making Sense of Genomic Variation: SNP Annotation -Part 1

Rolando Garcia-Milian
rolando.milian@yale.edu
Biomedical Sciences Research Support
Yale Cushing/Whitney Medical Library
Yale University.

Contents

Online Mendelian Inheritance in Man.....	3
dbSNP	5
SNPs – Trait Relationship: GWAS Catalog	7
Using Variant Effect Predictor (Ve!P) to Annotate a List of SNPs	9
Additional Annotation of Variants: Ensembl BioMart.....	14
References.....	20

Online Mendelian Inheritance in Man

For this example, we will find out the mode of inheritance of the genetic disorder ataxia telangiectasia and the gene associated with this disorder, genetic tests, and labs in US offering these tests. Go to the OMIM website <http://www.omim.org/> Enter “ataxia telangiectasia” in the search box and click on the “Search” button.

OMIM®

Online Mendelian Inheritance in Man®

An Online Catalog of Human Genes and Genetic Disorders

Updated 19 January 2016

ataxia telangiectasia

Search

Advanced Search : OMIM, Clinical Synopses, Gene Map | Search History

Need help? : Example Searches, OMIM Search Help, OMIM Tutorial

Mirror sites : us-east.omim.org, europe.omim.org

Look at the number of the results (1,233 entries). Use the Advance search to limit the number and specificity of your results. Click on the “Advance search” link located under the search box and click on the “OMIM” button. A new page will open showing all the limits available. Check the “Title” and the “#phenotype description, molecular basis known” limits and click on the “Search” button.

ataxia telangiectasia

Search

Advanced Search ▾ | Search History | Display Options ▾ | Retrieve Corresponding:

OMIM
Clinical Synopses
Gene Map

☐ atactic ☐ ataxic ☐ Add All
☐ ataxy ☐ dyssynergia

Search: 'ataxia telangiectasia'
Results: 1,233 entries.

ataxia telangiectasia Search

Sort by: ☒ Relevance ☐ Date updated ☐ Date created Entries per page: 10

Search in:
☐ MIM Number
☒ Title
☐ Text
☐ Allelic Variants

Only Records With:
☐ Allelic Variants
☐ Clinical Synopsis
☐ Gene Map Locus

MIM Number Prefix:
☐ * gene with known sequence
☐ + gene with known sequence and phenotype
☒ # phenotype description, molecular basis known
☐ % mendelian phenotype or locus, molecular basis unknown

The number of results will go down to 100. Click on the first record. “ # 208900. ATAXIA-TELANGIECTASIA; AT”

Note the table “Phenotype-Gene Relationship” showing the “Gene/Locus” associated with this disorder (ATM gene).

Under “TEXT”, it also says that “ataxia-telangiectasia (AT) is caused by homozygous or compound heterozygous mutation in the ATM gene (607585) on chromosome 11q22”

Under “Description”: is an autosomal recessive disorder

Phenotype-Gene Relationships

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key	Gene/Locus	Gene/Locus MIM number
11q22.3	Ataxia-telangiectasia	208900	AR	3	ATM	607585

[Clinical Synopsis](#)

TEXT

A number sign (#) is used with this entry because ataxia-telangiectasia (AT) is caused by homozygous or compound heterozygous mutation in the ATM gene (607585)

Description

Ataxia-telangiectasia (AT) is an autosomal recessive disorder characterized by cerebellar ataxia, telangiectases, immune defects, and a predisposition to malignancy. Ch cells are abnormally sensitive to killing by ionizing radiation (IR), and abnormally resistant to inhibition of DNA synthesis by ionizing radiation. The latter trait has been

Expand “Clinical Resources” tab- under the “External Links table located to the right of the page and click on the GTR (Genetic Testing Registry) link.

External Links
<ul style="list-style-type: none"> Protein Clinical Resources <ul style="list-style-type: none"> Clinical Trials EuroGentest Gene Reviews Genetic Alliance Genetics Home Reference GTR GARD OrphaNet POSSUM Animal Models Cell Lines

A new page will open showing the Genetic Testing Registry results for Ataxia-telangiectasia syndrome- 75 tests available in 29 labs. Scroll down the page to see the filtering options. Under “Lab location”, check United States: your results will narrow down to “52 tests for 1 condition in 17 labs”.

GTR: GENETIC TESTING REGISTRY

208900[mim]

[GTR Home](#) > [Tests](#) > [Search results - 208900\[mim\]](#) > **Filter applied** ([Remove all](#))

Apply filters

▼ **Condition/Phenotype**

Showing test for 1 condition

Enter text to filter the conditions

Select a condition [reset](#)

Ataxia-telangiectasia syndrome (75)

Showing 1 to 20 of 75 tests for 1 condition in 29 labs

C Clinical test, **R** Research test

C Syndromes with immunodeficiency Panel

Lab: [CeGaT GmbH](#) Tuebingen, Baden-Wuerttemberg, Germany

Conditions	Test targ
DiGeorge sequence	ADAR
Aicardi Goutieres syndrome 1	ATM

Apply filters

▼ **Condition/Phenotype**

Showing test for 1 condition

Enter text to filter the conditions

Select a condition [reset](#)

Ataxia-telangiectasia syndrome (52)

Showing 1 to 20 of 52 tests for 1 condition in 17 labs

C Clinical test, **R** Research test

C Tier 2: Familial Myelodysplastic Syndrome/Acute Leu

Lab: [Genetic Services Laboratory University of Chicago](#) Chicago, Illinois

Conditions	Test targ
Ataxia-telangiectasia syndrome	ATM

dbSNP

The ATM gene has been associated with Ataxia-telangiectasia, B-cell non-Hodgkin lymphoma, T-cell prolymphocytic leukemia, and susceptibility to Breast cancer. Find all missense variants reported for this gene. How many of these are pathogenic? Go to the NCBI database main page: <http://www.ncbi.nlm.nih.gov/> From the drop-down menu select the SNP database. On the search box type ATM[gene], and click on the search button.

NCBI Resources ☒ How To ☒

NCBI
National Center for Biotechnology Information

NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine

All Databases
Nucleotide
OMIM
PMC
PopSet
Probe
Protein
Protein Clusters
PubChem BioAssay
PubChem Compound
PubChem Substance
PubMed
PubMed Health
SNP
SRA
Structure

to NCBI
enter for Biotechnolog:
ation.
BI | [Mission](#) | [Organiz](#)
Submit
or manuscripts into
ses

How To ☒ rgmilia

SNP

Click on “Human” located on the right side of the page under the filter “Organism”. Click on “Pathogenic” under the Clinical significance” filter, and click on “missense” under the “Function class” filter.

dbSNP [Save search](#) [Advanced](#)

Organism: Homo sapiens
Variation Class: mnp, snp
Clinical Significance: ☒ pathogenic
Annotation: Cited in PubMed, OMIM, PubMed, nucleotide, protein
Function Class: 3' splice site, 3' utr, 5' splice site, 5' utr, coding synonymous, frame shift, intron, ☒ missense

Display Settings: ☒ Summary, 20 per page, Sort by

Results: 1 to 20 of 28

Filters activated: pathogenic, missense. [Clear all](#)

☐ rs28942103 [Homo sapiens]

1. TATTAGGTGGACCACACAGGAGAAT [A/G] TGGAA
Chromosome: 11:108334988
Gene: ATM (GeneView) C
Functional Consequence: intron variant, missense
Allele Origin: G(germline)/A(germ)
Clinical significance: Pathogenic
Validated: by cluster
HGVS: NC_000011.10:g.111111111G>A, NG_009830.1:g.111111111G>A, XM_005271414.1:c.111111111G>A, XM_005271415.1:c.111111111G>A, XM_005271416.1:c.111111111G>A, XM_005271561.3:c.111111111G>A, XM_005271562.3:c.111111111G>A, XM_005271564.1:c.111111111G>A, XM_006718845.1:c.111111111G>A, XM_011542841.1:c.111111111G>A, XM_011542843.1:c.111111111G>A, XM_011542845.1:c.111111111G>A, XP_005271618.1:c.111111111G>A

The format, number of results per page, and organization of results can be changed by clicking on “Display setting” menu. Results can be downloaded

SNP [Save search](#) [Advanced](#)

Display Settings: ☒ Summary, 20 per page, Sorted by SNP_ID

Format	Items per page	Sort by
<input checked="" type="radio"/> Summary	<input type="radio"/> 5	<input type="radio"/> Default order
<input type="radio"/> Graphic Summary	<input type="radio"/> 10	<input type="radio"/> Organism
<input type="radio"/> FASTA	<input checked="" type="radio"/> 20	<input checked="" type="radio"/> SNP_ID
<input type="radio"/> FlatFile	<input type="radio"/> 50	<input type="radio"/> Success Rate
<input type="radio"/> Chromosome Report	<input type="radio"/> 100	<input type="radio"/> Heterozygosity
<input type="radio"/> Old Summary	<input type="radio"/> 200	<input type="radio"/> Chromosome Base Position
<input type="radio"/> dbSNP Batch Report		

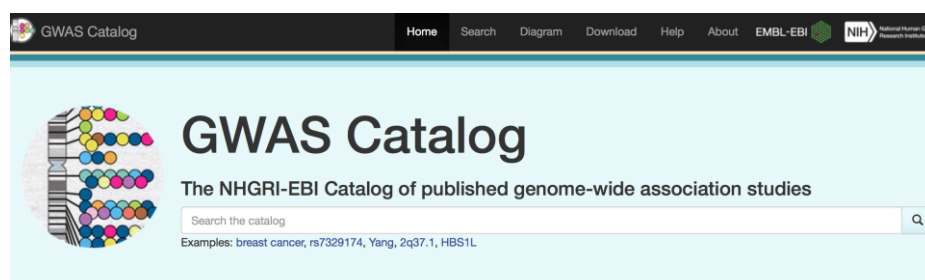
[Apply](#)

The same results can be obtained by searching the NCBI Gene database: <http://www.ncbi.nlm.nih.gov/gene> or from the Variation Viewer tool <http://www.ncbi.nlm.nih.gov/variation/view/>

SNPs – Trait Relationship: GWAS Catalog

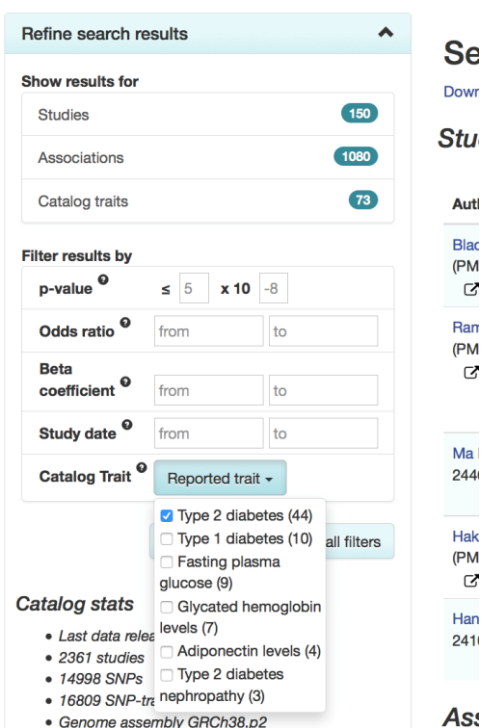
In this exercise we will find SNPs associated with “type 2 diabetes” by using the National Human Genome Research Institute (NHGRI) Catalog of Published Genome-Wide Association Studies (GWAS) (Welter et al., 2014)

Go to the main page of the GWAS Catalog: <https://www.ebi.ac.uk/gwas/home>
Type “diabetes” in the search box and click on the search icon.



Click on the “Reported trait” drop-down menu button next to Catalog Trait located in the “Refine search results” –left of the page. Select “Type 2 diabetes”¹. Note the number of studies in the catalog (44). Click on the “Apply filter” button located right below the “Reported trait” button.

The page will update to show the Studies listed in the Catalog. From the left bar, click on “Associations” (340) located under the “Show results for” section.



¹ The GWAS Catalog use Experimental Factor Ontology to map phenotypes/reported traits. In this case, “type 2 diabetes” is mapped to the term “type II diabetes mellitus”. This page http://www.ebi.ac.uk/efo/EFO_0001360 provides more information on this term as well as synonyms, parental term, etc.

Refine search results

Show results for

Studies44

Associations340

Catalog traits1

Filter results by

p-value

≤

5

x 10

-8

Odds ratio

from

to

Beta coefficient

from

to

Study date

from

to

Catalog Trait

Reported trait

Apply filters

Clear all filters

The page will update to show the 340 SNP associated with “type 2 diabetes”. Please notice that the resulting table of associations can be downloaded from the “Download search results” link at the top of the table. However for this exercise, we will filter and download only those associations with a p value < 5 x 10e-8 (default). Click on the “Apply filters” button again.

Show results for

Studies44

Associations340

Catalog traits1

Filter results by

p-value

≤

5

x 10

-8

Odds ratio

from

to

Beta coefficient

from

to

Study date

from

to

Catalog Trait

Reported trait

Apply filters

Clear all filters

Download search results

Associations

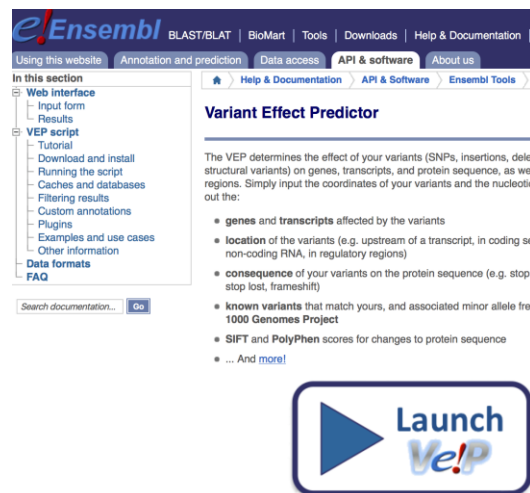
SNP	R ²
rs3916765-A	0.0
rs8090011-G	0.0
rs7178572-G	0.0

The table will update to show only those SNPs with association value p< 5 x 10e-8. The total number of associations will go down from 340 to 134. There are other relevant filters that you would like to apply depending on your specific case (e.g. Odds ratio, Study date, and Beta coefficient)

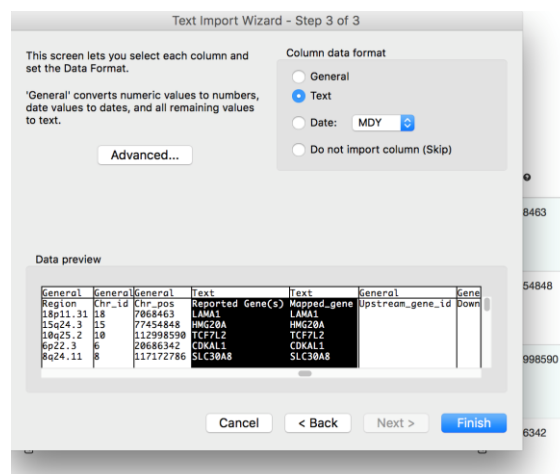
Click on the “Download search result” to save the result. We will use these in the next section to annotate a list of SNPs.

Using Variant Effect Predictor (Ve!P) to Annotate a List of SNPs

In this exercise we will annotate a list of SNPs associated with type 2 diabetes- obtained from the previous section by using the Variant Effect Predictor (McLaren et al., 2010) Open the Variant Effect Predictor (part of the Ensembl genome browser tools) main page <http://www.ensembl.org/info/docs/tools/vep/index.html> and click on the “Launch Ve!P” button. A new page will open




Use Microsoft Excel to open the table downloaded in the previous section containing the SNPs associated with type 2 diabetes. Define the column data format as “Text” for those columns containing the gene symbols. Excel automatically converts some gene symbols (MARCH3, SEPT1, DEC1) into dates (Zeeberg et al., 2004) From the Excel sheet- SNPs column-, copy the 184 SNPs “rs” identifiers, as shown in the figure below.



	U	V	W	X	Y
eam	Strongest SNP-Risk Allel	SNPs	Merged	Snps_id_curre	Context
	rs8090011-G	rs8090011		8090011	intron
	rs7178572-G	rs7178572		7178572	intron
	rs7903146-T	rs7903146		7903146	intron
	rs7766070-A	rs7766070		7766070	intron
	rs3802177-G	rs3802177		3802177	UTR-3
	rs7903146-T	rs7903146		7903146	intron
	rs9939609-A	rs9939609		9939609	intron
	rs7766070-A	rs7766070		7766070	intron
029	rs5015480-C	rs5015480		5015480	
	rs4402960-T	rs4402960		4402960	intron
	rs864745-T	rs864745		864745	intron
502	rs12779790-G	rs12779790		12779790	


Paste these identifiers into the VEP box next to "Either paste data". Please note that one can change the current GRCh38.p5 to the previous assembly GRCh37 or the species. For this example, we will use the GRCh38 assembly and Human (Homo sapiens) as species.

Variant Effect Predictor

 **VEP for Human GRCh37**

If you are looking for VEP for Human GRCh37, please go to [GRCh37 website](#).

Species:

 Human (Homo sapiens)

Assembly: GRCh38.p5

Name for this data (optional):

Either paste data:

rs8090011
rs7178572
rs7903146
rs7766070
rs3802177
rs7903146
rs9939609
rs7766070

Provide a name for your work in the box next to "Name for this data (optional)". Select "the Transcript database to use" as "Ensembl and RefSeq transcripts". Under the "Identifiers" tab, select the "Gene symbol" (default). Please note that there are many other identifiers available (e.g. UniProt, etc.).

Transcript database to use:	<input type="radio"/> Ensembl transcripts <input type="radio"/> Gencode basic transcripts <input type="radio"/> RefSeq transcripts <input checked="" type="radio"/> Ensembl and RefSeq transcripts
Include additional EST and CCDS transcripts:	<input type="checkbox"/>
Identifiers and frequency data <small>Additional identifiers for genes, transcripts and variants; frequency data</small>	
Identifiers	
Gene symbol:	<input checked="" type="checkbox"/>

Under “Frequency data” section. Select “Yes” from the drop-down menu next to “Find co-located known variants” and check the “1000 Genomes global minor allele frequency” (default).

Frequency data	
Find co-located known variants:	<input type="text" value="Yes"/>
Frequency data for co-located variants:	<input checked="" type="checkbox"/> 1000 Genomes global minor allele frequency <input type="checkbox"/> 1000 Genomes continental allele frequencies <input type="checkbox"/> ESP allele frequencies <input type="checkbox"/> ExAC allele frequencies
PubMed IDs for citations of co-located variants:	<input checked="" type="checkbox"/>
Include flagged variants:	<input type="checkbox"/>

(p) = functionality from [VEP plugin](#)

Under the “Extra options” tab select the annotations of your interest. For this example: “Transcript biotype”, “Protein domain”, “Transcript support level”, and “miRNA structure”.

Miscellaneous	
Transcript biotype:	<input checked="" type="checkbox"/>
Protein domains:	<input checked="" type="checkbox"/>
Exon and intron numbers:	<input type="checkbox"/>
Transcript support level:	<input checked="" type="checkbox"/>
APPRIS:	<input type="checkbox"/>
Identify canonical transcripts:	<input type="checkbox"/>
miRNA structure^(p):	<input checked="" type="checkbox"/>
Upstream/Downstream distance^(p):	<input checked="" type="radio"/> Disabled <input type="radio"/> Enabled

Under the “Pathogenicity predictions” tab, select “SIFT: Prediction and score”, “PolyPhen: Prediction and score”, “Condel: Enabled – Prediction and score”.

Pathogenicity predictions

SIFT:	Prediction and score
PolyPhen:	Prediction and score
dbNSFP^(p):	<input checked="" type="radio"/> Disabled <input type="radio"/> Enabled
Condel^(p):	<input type="radio"/> Disabled <input checked="" type="radio"/> Enabled
Score/prediction:	Prediction and score
LoFtool^(p):	<input type="checkbox"/>

Under “Regulatory data”, select “dbscSNV”. Under “Conservation”, select BLOSUM62 for this example.

Regulatory data

Get regulatory region consequences:	Yes
--------------------------------------------	-----

Splicing predictions

dbscSNV^(p):	<input checked="" type="checkbox"/>
MaxEntScan^(p):	<input type="checkbox"/>

Conservation

BLOSUM62^(p):	<input checked="" type="checkbox"/>
--------------------------------	-------------------------------------

(p) = functionality from [VEP plugin](#)

We will not filter the SNP list for any of the options available under “Filters”. Select “No filtering” (default) and click on the “Run” button.

Filters

Filter by frequency:

☒ No filtering

☐ Exclude common variants

☐ Advanced filtering

Return results for variants in coding regions only:

☐

Restrict results:

Show all results


NB: Restricting results may exclude some variants

[Run >](#) [Clear](#) [Close form](#)

Your job will be queued until done. Once it is done, click on the “View results” link.

Recent jobs



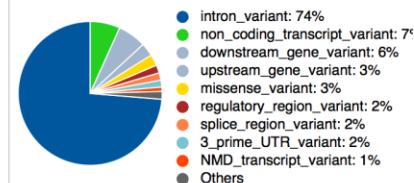
Show/hide columns (1 hidden)		
Analysis	Jobs	
Variant Effect Predictor	 VEP analysis of SNPs associated to type 2 diabetes from GWAS in Homo_sapiens	Done View results

A new page will open showing the annotation results of the SNP list consisting of a summary statistics, variant functional and coding consequences followed by a table.

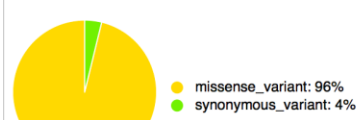
Summary statistics

Category	Count
Variants processed	207
Variants remaining after filtering	207
Novel / existing variants	19 (9.2%) / 188 (90.8%)
Overlapped genes	211
Overlapped transcripts	957
Overlapped regulatory features	27

Consequences (all)



Coding consequences



The results in the table can be further filtered by location, allele, consequence, allele frequency, clinical significance, etc. This table can be downloaded as VCF, VEP format, or TEX (best for Excel) for further analysis. Help on Ensembl prediction and variant classification can be found here:

http://www.ensembl.org/info/genome/variation/predicted_data.html

Results preview


Navigation				Filters	
Page: 1 of 42 Show: 1 5 10 50 All variants				Uploaded variant	
Show/hide columns				Location	
Uploaded variant	Location	Allele	Consequence	Allele	
rs17106184	1:50444313-50444313	A	upstream c	Consequence	
rs17106184	1:50444313-50444313	A	upstream c	Impact	
rs17106184	1:50444313-50444313	A	intron_v	Symbol	
rs17106184	1:50444313-50444313	A	intron_v	Gene	
rs17106184	1:50444313-50444313	A	intron_v	Feature type	
rs17106184	1:50444313-50444313	A	intron_v	Feature	
rs17106184	1:50444313-50444313	A	intron_v	Biotype	
rs17106184	1:50444313-50444313	A	intron_v	Exon	
rs17106184	1:50444313-50444313	A	intron_v	Intron	
rs17106184	1:50444313-50444313	A	intron_v	HGVSc	
rs17106184	1:50444313-50444313	A	intron_v	HGVSp	
rs17106184	1:50444313-50444313	A	intron_v	cDNA position	
rs17106184	1:50444313-50444313	A	intron_v	CDS position	
rs17106184	1:50444313-50444313	A	intron_v	Protein position	
rs17106184	1:50444313-50444313	A	intron_v	Amino acids	
rs17106184	1:50444313-50444313	A	intron_v	Codons	
rs17106184	1:50444313-50444313	A	intron_v	Existing variant	

Additional Annotation of Variants: Ensembl BioMart

In this example, we will use BioMart (Kinsella et al., 2011) to find whether these variants have a other phenotype associated and a Mendelian Inheritance in Man entry and description. In addition, BioMart can be use to add additional annotations related not only to the variant but also to associated genes and regulatory regions.

Open Ensembl BioMart <http://www.ensembl.org/biomart>

Select “Ensembl Variation 83” from the CHOOSE DATABASE drop-down menu. Select “Homo sapiens Short Variants (SNPs and indels excluding flagged variants) GRCh38.p5” from the CHOOSE DATASET drop-down menu.


[BLAST/BLAT](#) | [BioMart](#) | [Tools](#) | [Download](#)

[New](#) | [Count](#) | [Results](#)

Dataset
 [None selected]

- CHOOSE DATABASE -
 - CHOOSE DATABASE -
 Ensembl Genes 83
Ensembl Variation 83
 Ensembl Regulation 83
 Vega 63

NewCountResults

Dataset

[None selected]

Ensembl Variation 83

- CHOOSE DATASET -

- CHOOSE DATASET -

Homo sapiens Short Variants (SNPs and indels excluding flagged variants) (GRCh38.p5)

Homo sapiens Somatic Short Variants (SNPs and indels excluding flagged variants) (GRCh38.p5)

Homo sapiens Somatic Structural Variants (GRCh38.p5)

Homo sapiens Structural Variants (GRCh38.p5)

Bos taurus Short Variants (SNPs and indels excluding flagged variants) (UMD3.1)

Bos taurus Structural Variants (UMD3.1)

Canis familiaris Short Variants (SNPs and indels excluding flagged variants) (CanFam3.1)

Canis familiaris Structural Variants (CanFam3.1)

Danio rerio Short Variants (SNPs and indels excluding flagged variants) (GRCz10)

Danio rerio Structural Variants (GRCz10)

Drosophila melanogaster Short Variants (SNPs and indels excluding flagged variants) (B10000)

Equus caballus Short Variants (SNPs and indels excluding flagged variants) (EquCab2)

Equus caballus Structural Variants (EquCab2)

Felis catus Short Variants (SNPs and indels excluding flagged variants) (Felis_catus_6.2)

Gallus gallus Short Variants (SNPs and indels excluding flagged variants) (Galgal4)

Macaca mulatta Short Variants (SNPs and indels excluding flagged variants) (MMUL_1)

Macaca mulatta Structural Variants (MMUL_1)

Meleagris gallopavo Short Variants (SNPs and indels excluding flagged variants) (UMD2)

Click on the “Filter” link located on the left side of the page. A new menu will appear in the right pane. Expand the “General Variant Filters” heading by clicking on the plus sign next to it. Check the “Filter by Variant Name” option and paste your 124 variants associated with type 2 diabetes found in the GWAS Catalog - listed below:

NewCountResults

Dataset

Homo sapiens Short Variants (SNPs and indels excluding flagged variants) (GRCh38.p5)

Filters

[None selected]

Attributes

Variant Name

Variant source

Chromosome name

Chromosome position start (bp)

Chromosome position end (bp)

Dataset

[None Selected]

(If filter values)

NOTE: Due to the increase in data, it is no longer feasible to display all results. Use filters when querying the variation map.

REGION:

GENERAL VARIANT FILTERS:

GENE ASSOCIATED VARIANT FILTERS:

REGULATORY REGION ASSOCIATED INFORMATION FILTERS:

rs8090011	rs12779790	rs10923931	rs7178572
rs7178572	rs7961581	rs6931514	rs2028299
rs7903146	rs7578597	rs7903146	rs4812829
rs7766070	rs4607103	rs5215	rs7593730
rs3802177	rs7754840	rs1048886	rs243021
rs7903146	rs7756992	rs1333051	rs4457053
rs9939609	rs7903146	rs7305618	rs972283
rs7766070	rs7903146	rs163182	rs896854
rs5015480	rs1111875	rs3923113	rs13292136
rs4402960	rs13266634	rs16861329	rs231362
rs864745	rs7903146	rs1802295	rs1552224

15

rs1531343	rs10946398	rs10811661	rs7903146
rs7957197	rs10811661	rs7903146	rs9552911
rs11634397	rs4402960	rs1111875	rs7903146
rs8042680	rs7903146	rs2237892	rs7903146
rs5945326	rs4402960	rs5945326	rs1470579
rs7578326	rs7754840	rs7754840	rs7903146
rs1387153	rs10811661	rs7903146	rs1470579
rs1470579	rs8050136	rs7903146	rs7903146
rs1801214	rs13266634	rs3842770	rs1470579
rs10440833	rs1111875	rs2283228	rs7903146
rs849134	rs5219	rs343092	rs7756992
rs3802177	rs4402960	rs2244020	rs17791513
rs10965250	rs13266634	rs7903146	rs1111875
rs5015480	rs7901695	rs2283228	rs11257655
rs7903146	rs391300	rs343092	rs163184
rs11642841	rs17584499	rs231356	rs849135
rs7172432	rs2237895	rs6815464	rs5215
rs10906115	rs515071	rs7041847	rs4812829
rs1359790	rs7656416	rs6017317	rs2261181
rs8050136	rs7903146	rs6467136	rs9936385
rs5219	rs7560163	rs831571	rs4402960
rs7903146	rs10886471	rs9470794	rs1801282
rs7903146	rs7403531	rs3786897	rs4430796
rs2237892	rs10814916	rs1535500	rs10811661
rs4712524	rs2383208	rs10229583	rs3802177
rs6769511	rs11257655	rs75493593	rs4458523
rs2237897	rs4430796	rs6813195	rs2943640
rs4712523	rs12010175	rs9502570	rs7612463
rs2383208	rs5945326	rs17106184	rs7178572
rs2237892	rs791595	rs3132524	rs12571751
rs13266634	rs11787792	rs6808574	rs11717195
rs1111875	rs312457	rs702634	rs12970134
rs7903146	rs8181588	rs1727313	rs4506565
rs7903146	rs1470579	rs7903146	rs7018475
rs8050136	rs7754840	rs2237897	rs7766070

GENERAL VARIANT FILTERS:

☐ Variant source

ClinVar
dbSNP
ESP
HGMD-PUBLIC
HumanCoreExome-12

☒ Filter by Variant Name (e.g. rs123, CM000001) [Max 500 advised]

rs4402960
rs864745
rs12779790
rs7961581
rs7578597
rs4607103
rs7754840
rs7756992
rs7903146
rs7903146
rs1111875
rs13266634

In order to find the MIM annotations for these variants we need to cross the Ensembl Variation 83 with the Ensembl Genes 83 database (containing the MIM annotations). For this, click on the “Dataset” link located at the bottom of the left pane under “Attributes”. Select “(Ensembl Genes 83) Homo sapiens genes (GRCh38.p5)”

Dataset Homo sapiens Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5)	- CHOOSE ADDITIONAL DATASET - - CHOOSE ADDITIONAL DATASET - (Ensembl Genes 83) Homo sapiens genes (GRCh38.p5)
Filters Filter by Variant Name (e.g. rs123, CM000001) [Max 500 advised]: [ID-list specified]	[Ensembl Variation 83] Bos taurus Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5) [Ensembl Variation 83] Canis familiaris Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5) [Ensembl Variation 83] Danio rerio Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5) [Ensembl Variation 83] Drosophila melanogaster Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5) [Ensembl Variation 83] Equus caballus Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5) [Ensembl Variation 83] Felis catus Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5) [Ensembl Variation 83] Gallus gallus Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5) [Ensembl Variation 83] Homo sapiens Somatic Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5) [Ensembl Variation 83] Homo sapiens Somatic Structural Variants (SVs) (GRCh38.p5) [Ensembl Variation 83] Homo sapiens Structural Variants (SVs) (GRCh38.p5) [Ensembl Variation 83] Macaca mulatta Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5) [Ensembl Variation 83] Meleagris gallopavo Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5) [Ensembl Variation 83] Monodelphis domestica Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5) [Ensembl Variation 83] Mus musculus Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5) [Ensembl Variation 83] Nomascus leucogenys Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5) [Ensembl Variation 83] Ornithorhynchus anatinus Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5) [Ensembl Variation 83] Ovis Aries Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5)
Attributes Variant Name Variant source Chromosome name Chromosome position start (bp) Chromosome position end (bp)	
Dataset None Selected]	

Dataset Homo sapiens Short Variants (SNPs and indels excluding lagged variants) (GRCh38.p5)	<input checked="" type="radio"/> Features <input type="radio"/> Variant (Germline) <input type="radio"/> Structures <input type="radio"/> Variant (Somatic) <input type="radio"/> Homologs <input type="radio"/> Sequences
Filters Filter by Variant Name (e.g. rs123, CM000001) [Max 500 advised]: [ID-list specified]	<input type="checkbox"/> GENE:
Attributes Variant Name Variant source Chromosome name Chromosome position start (bp) Chromosome position end (bp)	<input type="checkbox"/> EXTERNAL:
Dataset Homo sapiens genes (GRCh38.p5)	<input type="checkbox"/> PROTEIN DOMAINS AND FAMILIES
Filters [None selected]	
Attributes Ensembl Gene ID Ensembl Transcript ID	

A new menu will appear on the right pane. Expand the “EXTERNAL” heading by clicking on the plus sign next to it. You might want to uncheck the “Ensembl Gene ID” and “Ensembl Transcript ID” located under the “GENE” heading if not needed since these will add additional columns to the table.

External References (max 3)

- ☐ ArrayExpress
- ☐ ChEMBL ID(s)
- ☐ Clone based Ensembl gene name
- ☐ Clone based Ensembl transcript name
- ☐ Clone based VEGA gene name
- ☐ Clone based VEGA transcript name
- ☐ CCDS ID
- ☐ Database of Aberrant 3' Splice Sites (DBASS3) |
- ☐ DBASS3 Gene Name
- ☐ Database of Aberrant 5' Splice Sites (DBASS5) |
- ☐ DBASS5 Gene Name
- ☐ EMBL (Genbank) ID
- ☐ Ensembl Human Transcript IDs
- ☐ Ensembl Human Translation IDs
- ☐ LRG to Ensembl link gene
- ☐ LRG to Ensembl link transcript
- ☐ EntrezGene ID
- ☐ EntrezGene transcript name ID
- ☐ Human Protein Atlas Antibody ID
- ☐ VEGA gene ID(s) (OTTG)
- ☐ VEGA transcript ID(s) (OTTT)
- ☐ VEGA protein ID(s) (OTTP)
- ☐ HGNC ID(s)
- ☐ HGNC symbol
- ☐ HGNC transcript name
- ☐ MEROPS ID
- ☒ MIM Morbid Accession
- ☒ MIM Morbid Description

Click on the “Results” button located on top of the left pane. The first 10 rows of results will appear on the right pane. You can select “All” from the drop-down menu to see all results (will open on a new tab). Save the the web page as text.

Results can also be saved by using BioMart export function located on top of the right pane: “Export all results to: File”. There is an option to export as XLS file format. Click on the “Go” button.

View

10

rows as

HTML

☐ Unique results only

Variant Name	Variant source	Chromosome name	Chromosome start (bp)	Chromosome position	Chromosome position end (bp)	MIM Morbid Accession	MIM Morbid Description
rs1801214	dbSNP	4	6301295	6301295	6301295	116400	CATARACT 41; CTRCT41;C
rs1801214	dbSNP	4	6301295	6301295	6301295	222300	WOLFRAM SYNDROME 1; V AND DEAFNESS; DIDMOAD
rs1801214	dbSNP	4	6301295	6301295	6301295	600965	DEAFNESS, AUTOSOMAL C DFNA14;DEAFNESS, AUTO
rs1801214	dbSNP	4	6301295	6301295	6301295	614296	WOLFRAM-LIKE SYNDROM OPTIC ATROPHY AND/OR II
rs1801214	dbSNP	4	6301295	6301295	6301295	116400	CATARACT 41; CTRCT41;C
rs1801214	dbSNP	4	6301295	6301295	6301295	222300	WOLFRAM SYNDROME 1; V AND DEAFNESS; DIDMOAD
rs1801214	dbSNP	4	6301295	6301295	6301295	600965	DEAFNESS, AUTOSOMAL C DFNA14;DEAFNESS, AUTO
rs1801214	dbSNP	4	6301295	6301295	6301295	614296	WOLFRAM-LIKE SYNDROM OPTIC ATROPHY AND/OR II
rs1801214	dbSNP	4	6301295	6301295	6301295	116400	CATARACT 41; CTRCT41;C
rs1801214	dbSNP	4	6301295	6301295	6301295	222300	WOLFRAM SYNDROME 1; V AND DEAFNESS; DIDMOAD

★ URL

XML

Perl

Help

Export all results to

File

XLS

HTML

CSV

TSV

XLS

☒ Unique results only

Go

Email notification to

View

10

rows as

HTML

☒ Unique results only

Variant	Variant	Chromosome	Chromosome position	MIM Morbid Description	MIM Morbid
---------	---------	------------	---------------------	------------------------	------------

References

- Kinsella, R. J., Kahari, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., . . . Flicek, P. (2011). Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database (Oxford)*, 2011, bar030. doi:10.1093/database/bar030
- McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., & Cunningham, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*, 26(16), 2069-2070. doi:10.1093/bioinformatics/btq330
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., . . . Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*, 42(Database issue), D1001-1006. doi:10.1093/nar/gkt1229
- Zeeberg, B. R., Riss, J., Kane, D. W., Bussey, K. J., Uchio, E., Linehan, W. M., . . . Weinstein, J. N. (2004). Mistaken identifiers: gene name errors can be introduced inadvertently when using Excel in bioinformatics. *BMC Bioinformatics*, 5, 80. doi:10.1186/1471-2105-5-80