

Online resources: history of variant consortia, repositories, resources

Tim Hubbard @timjph

King's College London, King's Health Partners
Genomics England

Bioinformatics, Interpretation and Data Quality in Genome Analysis
MSc in Genomics Medicine
15th February 2016

Questions about a variant

- Has it been seen before?
- How common is it?
- What is known about its effect on function?
- Rare diseases
 - Variant shouldn't be common
 - Might have already been linked to a disease

Resources

- Human or all species
- Small (SNP) or large variation (SV,CNV)
- Natural or diseased
- Raw or curated
- Open or closed

Repositories – dbSNP

<http://www.ncbi.nlm.nih.gov/SNP/>

Have a question about dbSNP? Try searching the SNP FAQ Archive!

Go

GENERAL

[RSS Feed](#)

[Contact Us](#)

[Site Map](#)

[dbSNP Homepage](#)

[NCBI Variation Resources](#)

[Announcements](#)

[dbSNP Summary](#)

[FTP Download](#)

SNP SUBMISSION

DOCUMENTATION

SEARCH

RELATED SITES

dbSNP Summary

RELEASE: NCBI dbSNP Build 146

dbSNP Component Availability Dates:

Component	Date available
dbSNP web query for build 146:	Nov 24, 2015
ftp data for build 146:	Nov 24, 2015
Entrez Indexing for build 146:	Nov 24, 2015

- The complete data for build 146 are available at <ftp://ftp.ncbi.nlm.nih.gov/snp/> in multiple formats.
- All formats and conventions are described in <ftp://ftp.ncbi.nlm.nih.gov/snp/00readme.txt>.
- Please address any questions or comments regarding the data to snp-admin@ncbi.nlm.nih.gov.

New Submission since previous build:

Organism	Current Build	New Submissions (ss#s)	New RefSNP Clusters (rs#s) (# validated)	New ss# with Genotype	New ss# with Frequency
Homo sapiens	146	32,024,450	985,775 (39,318)		
Bos taurus	146	58,867,577	4,365,917 (109,726)		
Mus musculus	146	27,951	9,369,418 (16)		

Repositories – dbVar

<http://www.ncbi.nlm.nih.gov/dbvar>



The screenshot shows the dbVar homepage. At the top, there's a navigation bar with links for NCBI, Resources, How To, and Sign in to NC. Below the navigation is a search bar with a dropdown set to "dbVar" and a "Search" button. A "Help" link is also visible. The main content area features a dark header with the "dbVar" logo and a descriptive text about the database. Below this is a grid of nine corn cobs of various colors (yellow, orange, red, purple) arranged horizontally. The page is divided into several sections: "Getting Started" (with links to Overview of Structural Variation, FAQ, Help, and Data Submission Sheet), "Accessing Data" (with links to Study Browser, Genome Browser, Organism List, and FTP Data Download), "Other NCBI Resources" (with links to dbSNP, ClinVar, Variation Portal, and Variation Tools), "Submitting Data" (with links to Submission Guidelines, Submission Templates, and CF Submissions), "dbVar News" (with a link to Announcements and an RSS icon), and "External Resources" (with links to Database of Genomic Variants archive (DGVa), Database of Genomic Variants (DGV), 1000 Genomes Project, and NHGRI Structural Variation Project).

NCBI Resources How To Sign in to NC

dbVar dbVar Search Help

Advanced

dbVar

dbVar is NCBI's database of genomic structural variation – it contains insertions, deletions, duplications, inversions, multinucleotide substitutions, mobile element insertions, translocations, and complex chromosomal rearrangements

Getting Started

- Overview of Structural Variation
- FAQ
- Help
- Data Submission Sheet

Accessing Data

- Study Browser
- Genome Browser
- Organism List
- FTP Data Download

Other NCBI Resources

- dbSNP
- ClinVar
- Variation Portal
- Variation Tools

Submitting Data

- Submission Guidelines
- Submission Templates
- CF Submissions

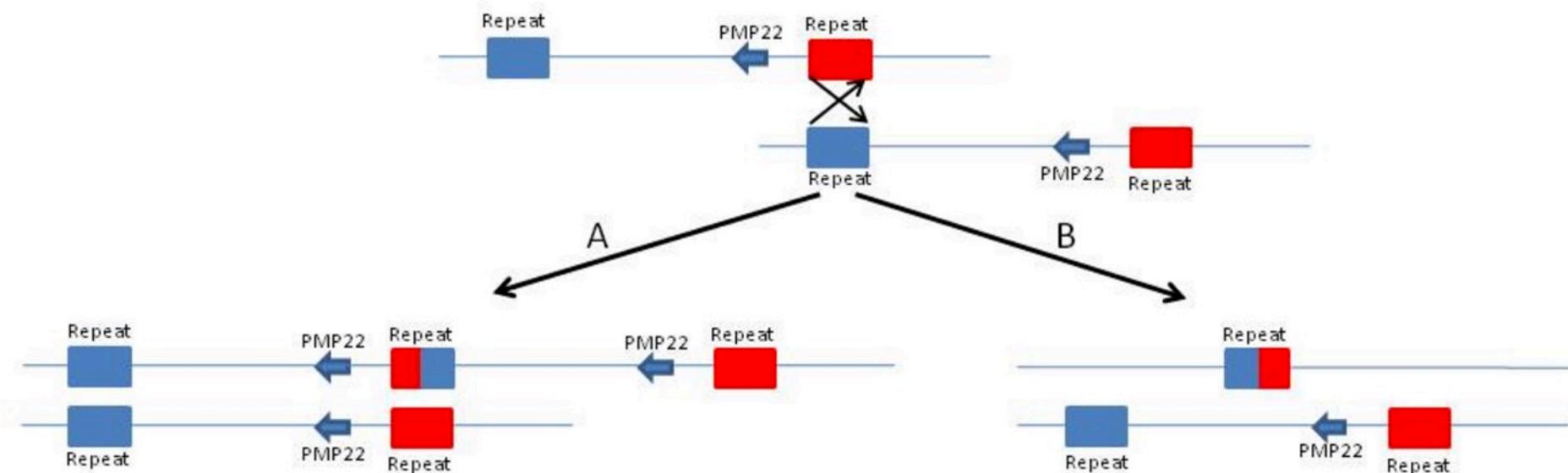
dbVar News

- Announcements

External Resources

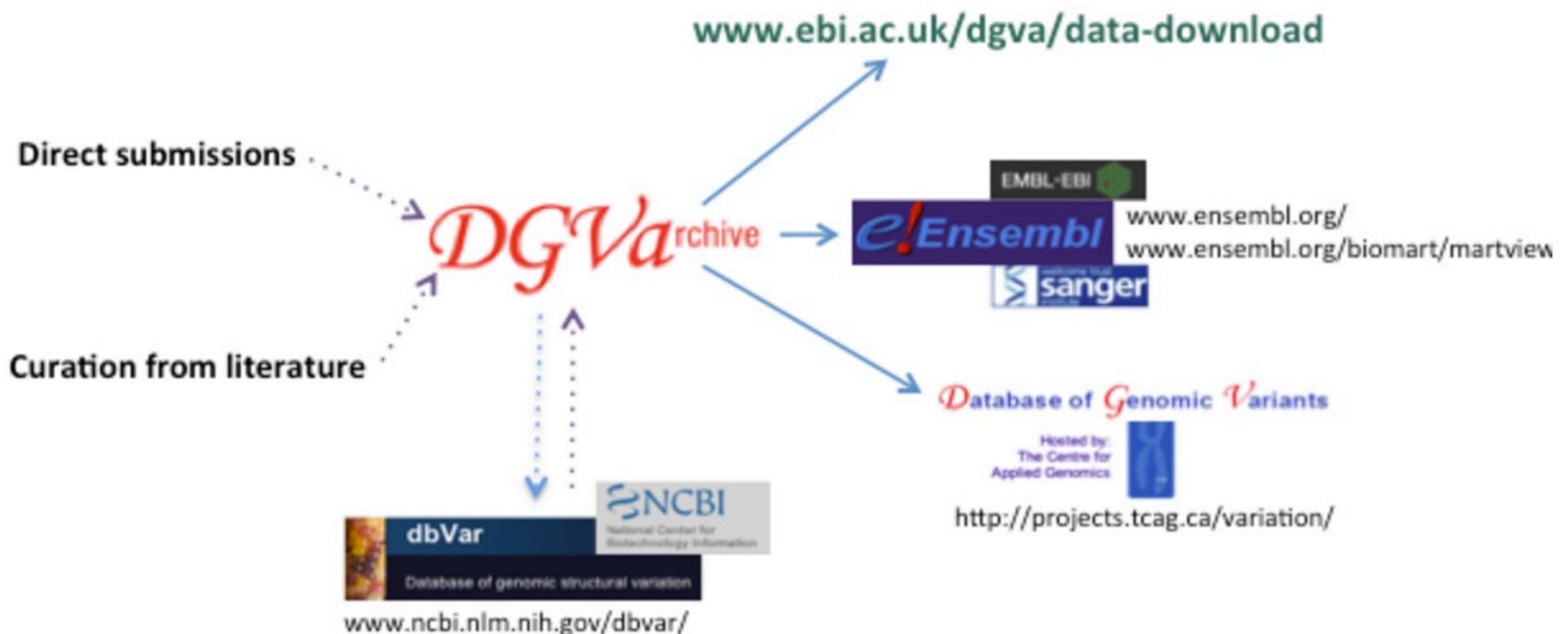
- Database of Genomic Variants archive (DGVa)
- Database of Genomic Variants (DGV)
- 1000 Genomes Project
- NHGRI Structural Variation Project

Pathogenic structural variation example



Charcot-Marie Tooth (CMT) disease. Unequal crossing over between two highly homologous repeats on chromosome 17p12 can result in (A) 3 copies of the PMP22 gene with the CMT1A phenotype or the reciprocal (B) and 1 copy of the PMP22 gene with the HNPP phenotype

Multiple related SV databases



Pathogenic variation: HGMD

<http://www.hgmd.cf.ac.uk/ac/>

The Human Gene Mutation Database
at the Institute of Medical Genetics in Cardiff

Home Search help Statistics New genes What is new Background Publications Contact Register Login LSDBs Other links

Symbol: Missense/nonsense Go!

The Human Gene Mutation Database (HGMD®) represents an attempt to collate known (published) gene lesions responsible for human inherited disease. and is maintained in Cardiff by D.N. Cooper, E.V. Ball, P.D. Stenson, A.D. Phillips, K. Howells, S. Heywood, M.J. Hayden, M.E. Mort and M.P. Horan.

*Please note that this less up-to-date public version of our database is freely available only to [registered](#) users from academic institutions/non-profit organisations. All commercial users are required to purchase a license from QIAGEN®, our commercial partner. A license to [HGMD Professional](#) is available to both commercial and academic/non-profit users wishing to access the most up-to-date version of the database (visit QIAGEN® to request a [free trial](#) of HGMD Professional). Read more about how HGMD is [funded](#). BBC reports recent study utilising 1000 Genomes and HGMD data ([BBC news](#)). You may not copy, store or re-distribute HGMD data without express written permission (i) from the curators or (ii) via your license agreement. Copyright © Cardiff University 2015.

Register for Public Version

Table:	Description:	Public entries: This site. Academic/non-profit users only	Total entries: HGMD Professional 2015.4
	Mutation totals (as of 2016-02-16)	127836	179235
Gene symbol	The gene description, gene symbol (as recommended by the HUGO Nomenclature Committee) and chromosomal location is recorded for each gene. In cases where a gene symbol has not yet been made official, a provisional symbol has been adopted which is denoted by lower-case letters.	4860	7189
cDNA sequence	cDNA reference sequences are provided, numbered by codon.	4788	7421
Genomic coordinates	Genomic (chromosomal) coordinates have been calculated for missense/nonsense, splicing, regulatory, small deletions, small insertions and small indels.	0	157581
HGVS nomenclature	Standard HGVS nomenclature has been obtained for missense/nonsense, splicing, regulatory, small deletions, small insertions and small indels.	0	158606
Missense/nonsense	Single base-pair substitutions in coding regions are presented in terms of a triplet change with an additional flanking base included if the mutated base lies in either the first or third position in the triplet.	71196	100090
Splicing	Mutations with consequences for mRNA splicing are presented in brief with information specifying the relative position of the lesion with respect to a numbered intron donor or acceptor splice site. Positions given as positive integers refer to a 3' (downstream) location, negative integers refer to a 5' (upstream) location.	11875	16356
Regulatory	Substitutions causing regulatory abnormalities are logged in with thirty nucleotides flanking the site of the mutation on both sides. The location of the mutation relative to the transcriptional initiation site, initiation codon, polyadenylation site or termination codon is given.	2458	3414
Small deletions	Micro-deletions (20 bp or less) are presented in terms of the deleted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character (^).	19803	26623
Small insertions	Micro-insertions (20 bp or less) are presented in terms of the inserted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character (^).	8136	11184
Small indels	Micro-indels (20 bp or less) are presented in terms of the deleted/inserted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character (^).	1904	2553
Gross deletions	Information regarding the nature and location of each lesion is logged in narrative form because of the extremely variable quality of the original data reported.	8835	13556
Gross insertions	Information regarding the nature and location of each lesion is logged in narrative form because of the extremely variable quality of the original data reported.	1973	3272

Pathogenic variation: ClinVar

<http://www.ncbi.nlm.nih.gov/clinvar/>

The screenshot shows the ClinVar website homepage. At the top, there is a navigation bar with links for NCBI, Resources, How To, and Sign in to NCBI. Below the navigation bar is a search bar with the placeholder "Search ClinVar for gene symbols, HGVS expressions, conditions, and more". There is also an "Advanced" search link and a "Search" button. A "Help" link is located on the right side of the search bar. Below the search bar is a horizontal menu with links for Home, About, Access, Using the website, How to submit, Statistics, and FTP site. The main content area features a dark blue header with the word "ClinVar" in white. Below the header, a text box contains a DNA sequence: ACTGATGGTATGGGCCAAGAGATATATCT CAGGTACGGCTGTCACTAGACCTCAC CAGGGCTGGGCATAAAAGTCAGGGCAGAGC CCATGGTGCATCTGACTCCTGAGGAGAAGT GCAGGTTGGTATCAAGGTTACAAGACAGGT GGCACTGACTCTCTGCCTATTGGTCTAT. To the right of the sequence, a text box says "ClinVar aggregates information about genomic variation and its relationship to human health." On the left side, there is a sidebar titled "Using ClinVar" with links for About ClinVar, Data Dictionary, Downloads/FTP site, FAQ, Contact Us, RSS feed/What's new?, and Factsheet. In the center, there is a section titled "Tools" with links for ACMG Recommendations for Reporting of Incidental Findings, Clinical Remapping - Between assemblies and RefSeqGenes, RefSeqGene/LRG, Submissions, Variation Reporter, and Variation Viewer. On the right side, there is a section titled "Related Sites" with links for ClinGen, GeneReviews®, GTR®, MedGen, OMIM®, and Variation.

Using ClinVar

- [About ClinVar](#)
- [Data Dictionary](#)
- [Downloads/FTP site](#)
- [FAQ](#)
- [Contact Us](#)
- [RSS feed/What's new?](#)
- [Factsheet](#)

ClinVar

ClinVar aggregates information about genomic variation and its relationship to human health.

Tools

- [ACMG Recommendations for Reporting of Incidental Findings](#)
- [Clinical Remapping - Between assemblies and RefSeqGenes](#)
- [RefSeqGene/LRG](#)
- [Submissions](#)
- [Variation Reporter](#)
- [Variation Viewer](#)

Related Sites

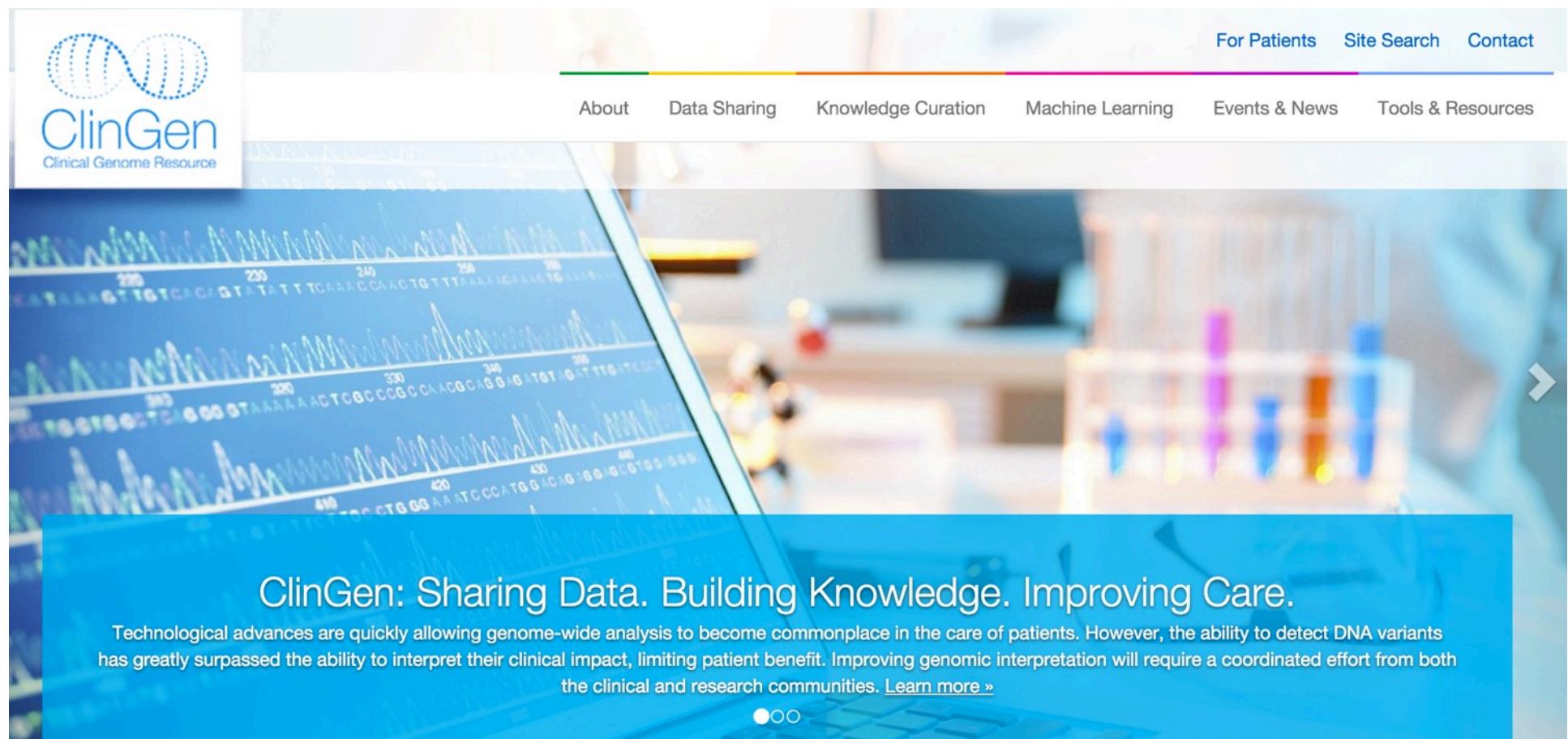
- [ClinGen](#)
- [GeneReviews®](#)
- [GTR®](#)
- [MedGen](#)
- [OMIM®](#)
- [Variation](#)

Submitter highlights

We gratefully acknowledge those who have submitted data and provided advice during the development of ClinVar. Subscribe to our [RSS feed](#) and follow us on [Twitter](#) to receive announcements of the release of new datasets. More [information about our submitters](#) is available, as well as a list of submitters with [the number of records each has submitted](#).

ClinGen:

<https://www.clinicalgenome.org/>



Pathogenic variation: Decipher

<https://decipher.sanger.ac.uk>

The screenshot shows the DECIPHER website interface. At the top, there is a navigation bar with links for 'About', 'Browse', 'DDD(UK)', a search bar containing 'Search DECIPHER' with a magnifying glass icon, and user account links for 'Join' and 'Login'. Below the navigation bar, the main content area displays 'Search results for 'gene:BRCA2'' with a link to 'Refine Search'. There are three tabs: 'Consented patients' (11), 'Syndromes' (0), and 'DDD Research Variants' (0). The 'Results' tab is selected. A blue header bar below these tabs contains 'Results' and 'Browser' buttons.

Search results: 1 to 11 of 11

Show: All Filter...

Decipher ID	Variant	Sex	Interval	Phenotype(s)	Patient Public Variants	Contact
1604	13 32169305 32972598 loss	46XY	803.29 kb		1	<input type="checkbox"/>
1605	13 32169305 32972598 loss	46XY	803.29 kb		1	<input type="checkbox"/>
1606	13 32169305 32972598 loss	46XX	803.29 kb		1	<input type="checkbox"/>
250663	13 30002277 33859688 deletion	46XX	3.86 Mb	Athetosis, Intrauterine growth retardation, Microcephaly	1	<input type="checkbox"/>
252508	13 28868848 34102234 loss	46XY	5.23 Mb	Abnormality of the vertebrae, Constipation, Hypertelorism, Intellectual disability, Pectus excavatum, Proportionate short stature, Webbed neck	1	<input type="checkbox"/>

Other repositories

- Restricted access archives
 - dbGaP - Database of Genotypes and Phenotypes
 - EGA - European Genome-phenome Archive

Locus Specific Databases

The screenshot shows the HGVS (Human Genome Variation Society) website. The header features the HGVS logo and the text "HUMAN GENOME VARIATION SOCIETY". A search bar is located in the top right corner. Below the header is a blue banner with a 3D molecular structure. The main navigation menu includes links for HOME, ABOUT HGVS, GUIDELINES, MEMBERSHIP, DATABASES, MEETINGS, and CONTACT US.

HUMAN MUTATION

Members who choose to subscribe to *Human Mutation* will receive a substantial discount.

"In This Issue"

Free-access essays

[February 2015](#)

[October 2014](#)

IMPORTANT NOTE:

Genes are in order of **HUGO APPROVED GENE DESIGNATION** not alias. e.g. "p53" will be found under "TP53" while "CD40L" or "TNFSF5" will be found under "CD40LG" and so on.

If you wish to find an Approved gene symbol please select [HGNC Search](#).

If your gene is not in these lists, you may like to check the "Disease Centered", "Mitochondrial Mutations" or "Other mutation Databases" database links as it may be in one of those.

If you wish to add an LSDB please go to the [LSDB Submission Page](#)

Please note that some LSDBs have a *Curator vacancy* - if you would like to serve please contact the acting curator via that database.

Locus Specific Databases

Specify the HGNC Gene Symbol:

Go to Gene!

[A](#) (132) | [B](#) (54) | [C](#) (205) | [D](#) (45) | [E](#) (30) | [F](#) (80) | [G](#) (88) | [H](#) (109) | [I](#) (76) | [J](#) (9) | [K](#) (51)
| [L](#) (35) | [M](#) (124) | [N](#) (53) | [O](#) (17) | [P](#) (143) | [Q](#) (1) | [R](#) (69) | [S](#) (130) | [T](#) (112) | [U](#) (26) | [V](#)
(12) | [W](#) (12) | [X](#) (7) | [Y](#) (1) | [Z](#) (25)

HGNC GENE SYMBOL	HGNC gene description	OMIM NO	DATABASE NAME/INTERNET ADDRESS	CURATORS
ABCA12	ATP-binding cassette, sub-family A (ABC1), member 12	607800	Mutations of the ATP-binding Cassette Transporter A 12 (ABCA12) associated with harlequin ichthyosis, congenital ichthyosiform erythroderma and lamellar ichthyosis	Masashi Akiyama and Kaori Sakai
ABCA13	ATP-binding cassette, sub-family A (ABC1), member 13	607807	LOVD - Leiden Open Variation Database	Johan T. den Dunnen and Ben Pickard
ABCA4	ATP-binding cassette, sub-family A (ABC1), member 4	601691	Mutations of the ATP-binding Cassette Transporter Retina	Retina International
ABCB11	ATP-binding cassette, sub-family B (MDR/TAP), member 11	603201	CCHMC - Human Genetics Mutation Database	Ammar Husami
ABCB4	ATP-binding cassette, sub-family B (MDR/TAP), member 4	171060	CCHMC - Human Genetics Mutation Database	Ammar Husami, Brian Richardson, Edita Freeman, Kerry Shooner, Thedia Jacobs and Theru A. Sivakumaran

Locus Specific Database for ABCA13

LOVD
Leiden Open Variation Database

LOVD - Leiden Open Variation Database
ATP-binding cassette, sub-family A, member 13 (ABCA13) 

Curators: [Johan den Dunnen](#) and [Ben Pickard](#)

[Home](#) [Variants](#) [Submitters](#) [Submit](#) [Documentation](#)

ABCA13 homepage [Switch gene](#)

LOVD Gene homepage

General information	
Gene name	ATP-binding cassette, sub-family A, member 13
Gene symbol	ABCA13
Chromosome Location	7p12.3
Database location	chromium.lovd.nl
Curator	Johan den Dunnen and Ben Pickard
PubMed references	View all (unique) PubMed references in the ABCA13 database
Date of creation	April 25, 2007
Last update	January 05, 2010
Version	ABCA13 100105
Add sequence variant	Submit a sequence variant
First time submitters	Register here
Total number of unique DNA variants reported	10
Total number of individuals with variant(s)	107
Total number of variants reported	107
Subscribe to updates of this gene	

Graphical displays and utilities	
Summary tables	Summary of all sequence variants in the ABCA13 database, sorted by type of variant (with graphical displays and statistics)
UCSC Genome Browser	Show variants in the UCSC Genome Browser (compact view)
Ensembl Genome Browser	Show variants in the Ensembl Genome Browser (compact view)
NCBI Sequence Viewer	Show distribution histogram of variants in the NCBI Sequence Viewer

<http://www.lrg-sequence.org/>

Locus•Reference•Genomic

All Find LRGs
e.g. LRG_1, COL1A1, NM_000088.3
View all | Search help

Home About Download FAQ

View LRGs
View the list of existing LRGs and the status of LRGs in progress.

Download LRG data
Download LRG records in XML and FASTA format from the [FTP site]. See all the downloads.

Request an LRG
If no LRG record exists for your gene of interest [View all LRGs], you can request one to be created for you.

Remap to LRG
RefSeqGene to LRG | GRCh38 to LRG | GRCh37 to LRG

Submit variants
You may submit information about variants to dbSNP or dbVar/DGVA and obtain an accession number for each variant. The results will be returned to you as outlined in this document.

LRG

LRG sequences provide a stable genomic DNA framework for reporting mutations with a permanent ID and core content that never changes.

- ▶ News
- ▶ Collaborators
- ▶ View the list of LRG web services
- ▶ Read more about LRGs
- ▶ FAQ
- ▶ Contact

Projects

- Small scale clinical sequencing projects recording variants suspected to be causing disease
 - Locus specific databases
- Large scale sequencing projects cataloging variation
 - 1000 genomes project
 - Exome sequencing projects

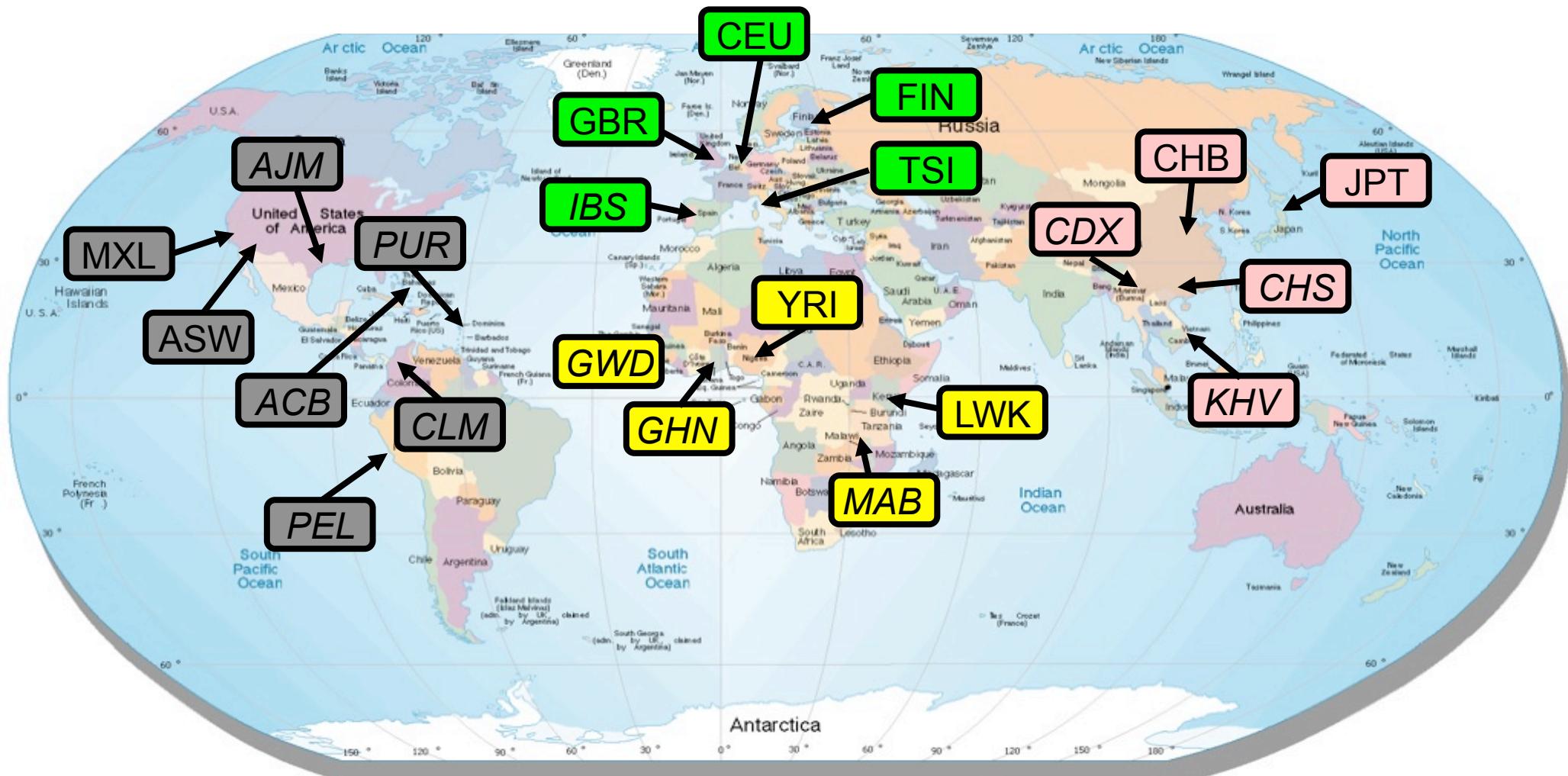
Consortia

- Discovering common variants
- Whole Genome
 - 1000 genomes project (2,500 genomes)
- Whole Exome
 - Exome Variant Server (EVS)
Aggregating large scale exome sequencing projects



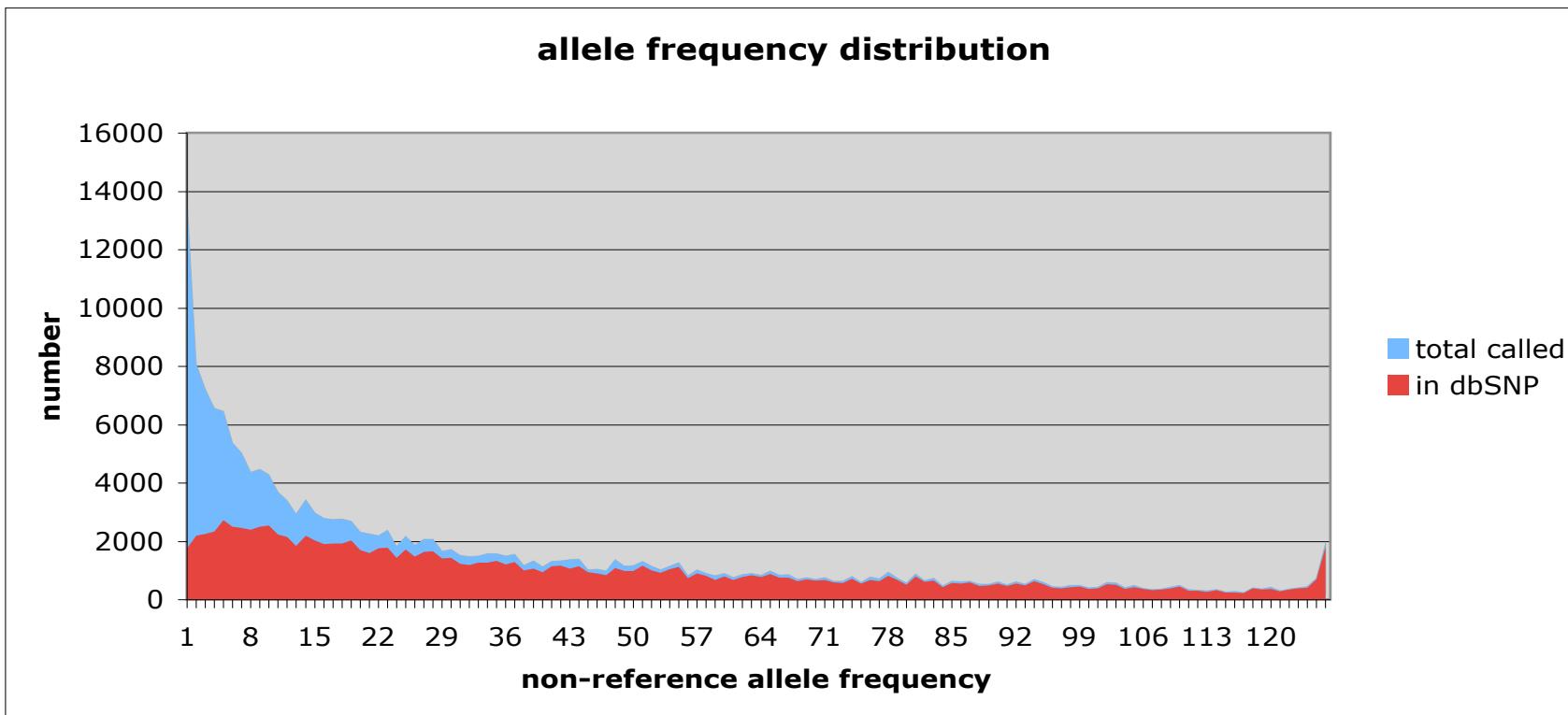
- International Consortium to generate a high resolution catalogue of human variation to support medical genetics
 - Baylor, BGI, Broad, Sanger, WashU, Berlin, AB, Illumina, 454
 - EBI/NCBI, many analysts, Samples/ELSI, NHGRI/WT
- Aim to find SNPs, indels, structural variants down to 1% minor allele frequency with >90% power
 - Call and phase on the sampled individuals to support imputation
- Produce an open resource building on HGP, HapMap etc.
 - Anonymised samples without phenotypes
 - Data publicly available, cell lines available
 - Do this in multiple populations

Main project: 3 × 400 samples at 4x in 2009
extend to 2,000 samples in 2010



Major population groups comprised of subpopulations of ~100 each

Allele frequency distribution



Overall 66% CEU calls were in dbSNP 129
(compare ~90% calls in one individual)

<http://evs.gs.washington.edu/EVS/>

 NHLBI Exome Sequencing Project (ESP)
Exome Variant Server

[Home](#) [Data Browser](#) [Data Usage and Release](#) [How to Use](#) [What's New](#) [Contact and FAQ](#) [Downloads](#)

The goal of the **NHLBI GO Exome Sequencing Project (ESP)** is to discover novel genes and mechanisms contributing to heart, lung and blood disorders by pioneering the application of next-generation sequencing of the protein coding regions of the human genome across diverse, richly-phenotyped populations and to share these datasets and findings with the scientific community to extend and enrich the diagnosis, management and treatment of heart, lung and blood disorders.

The groups participating and collaborating in the NHLBI GO ESP include:

- Seattle GO - University of Washington, Seattle, WA
- Broad GO - Broad Institute of MIT and Harvard, Cambridge, MA
- WHISP GO - Ohio State University Medical Center, Columbus, OH
- Lung GO - University of Washington, Seattle, WA
- WashU GO - Washington University, St. Louis, MO
- Heart GO - University of Virginia Health System, Charlottesville, VA
- ChargeS GO - University of Texas Health Sciences Center at Houston

The group includes some of the largest well-phenotyped populations in the United States, representing more than 200,000 individuals altogether from the:

- Women's Health Initiative ([WHI](#))
- Framingham Heart Study ([FHS](#))
- Jackson Heart Study ([JHS](#))
- Multi-Ethnic Study of Atherosclerosis ([MESA](#))
- Atherosclerosis Risk in Communities ([ARIC](#))
- Coronary Artery Risk Development in Young Adults ([CARDIA](#))
- Cardiovascular Health Study ([CHS](#))
- Lung Health Study ([LHS](#))
- COPD Genetic Epidemiology ([COPDGene](#))
- Severe Asthma Research Project ([SARP](#))
- Pulmonary Arterial Hypertension population ([PAH](#))
- Acute Lung Injury cohort (ALI)
- Cystic Fibrosis cohort ([CF](#))
- PennCATH
- Cleveland Clinic Genebank
- Massachusetts General Hospital Premature Coronary Artery Disease Study ([MGH PCAD](#))
- Heart Attack Risk in Puget Sound ([HARPS](#))
- Translational Research Underlying Disparities in Myocardial Infarction Patients' Health Status ([TRIUMPH](#))

[Privacy](#)

[Terms](#)

Resources for Human Genetics

Locus Specific Databases

Diverse
coordinate
systems

Genome wide datasets (1000 genomes; UK10K; GWAS)

Causal
mutations
unknown

Integrated Resources for Human Genetics Research

Human data

Number of	2012-03	2011-03
dbSNP Short Genetic Variations	52,266,522	More than 30 million
HGMD variants 	65,591	57,930
Somatic mutations 	63,849	38,487
LRG submitted variants 	408	280
 genotypes	1.4 billion	20 million

Structural variants

Species/ number of structural variant	2012-03	2011-03
Human	4,597,857	72,276
Mouse	1,753,767	10,909
Dog	4,883	730
Pig	280	37

Variation annotation – Phenotypes

Source	No. of Human variants with phenotypes
HGMD	65,591
COSMIC	63,767
Open access db	52,489
SNPedia via Das	28,401
OMIM	14,352
GWAS catalog	5,536
EGA	384

Ensembl: phenotype view

Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Human (GRCh37) ▾ Phenotype: Glaucoma

Phenotype-based displays
Locations on genome

Configure this page

Manage your data

Export data

Bookmark this page

Download view as CSV

Locations of variants associated with Glaucoma

Click on the image above to jump to a chromosome, or click and drag to select a region

Key

Feature type	Colour
Variation	1.0 3.0 4.0 5.0 6.0 7.0 8.0 >10

Less significant -log(p-values) → More significant -log(p-values)

Variants associated with phenotype Glaucoma

Show	All	entries	Filter			
Show/hide columns						
Genomic location (strand)	Ensembl ID	Variant source	Reported gene(s)	Phenotype(s) associated with this variant	Phenotype(s) source(s)	P value (negative log)
9:22063652-	rs4977756	dbSNP	CDKN2B-AS1	GLIOMA SUSCEPTIBILITY 5; Glaucoma; Glioma	NHGRI GWAS	14.0

Ensembl: adding disease associations

- Incorporate GWAS association (NHGRI catalog)

Human (GRCh37) Location: 2:193,050,820-193,051,820 Variation: rs10497726

Variation displays

- Summary
- Gene/Transcript (3)
- Population genetics (95)
- Individual genotypes (3150)
- Context
- Linked variations
- Phenotype Data (1)**
- Phylogenetic Context (4)
- External Data

Configure this page

Manage your data

Export data

Bookmark this page

Variation: rs10497726

Variation class: SNP (source [dbSNP 131](#) - Variants (including SNPs and indels) imported from dbSNP)

Synonyms: Affy GeneChip 100K Array SNP_A-1671718
Illumina_Human660W-quad rs10497726
Illumina_Human1M-duoV3 rs10497726
[dbSNP rs57991142](#)

Present in: 1000 genomes (1000 genomes - High coverage - CEU trio, 1000 genomes - Low coverage), Phenotype-associated variations (GWAS_BMC_PAPER), HapMap (HapMap Phase II, HapMap Phase III), Clinical/LSDB variations from dbSNP, ENSEMBL:Venter

Alleles: C/A (Ambiguity code: M)
Ancestral allele: C

Location: This feature maps to 1 genomic location.
2:193051320 (forward strand) [Jump to region in detail](#)

« Linked variations **Phenotype Data** Phylogenetic Context »

Disease/Trait	Source	Study	Associated Gene(s)	Associated variant	Strongest risk allele	P value
CVD outcomes-Coronary Heart Disease including Myocardial Infarction, Coronary Insufficiency, CHD death [View on Karyotype]	[Open Access GWAS Database]	pubmed/17903304 TMEFF2	rs10497726			1.98E-05

Ensembl: integrating variation ancestry

e!Ensembl Account | Logout | BLAST/BLAT | BioMart | Tools | Downloads | Help | Documentation | Mirrors  

Human (GRCh37) Location: 2:193,050,820-193,051,820 Variation: rs10497726

Variation displays	Variation: rs10497726
<ul style="list-style-type: none">- Summary- Gene/Transcript (3)- Population genetics (95)- Individual genotypes (3150)- Context- Linked variations- Phenotype Data (1)Phylogenetic Context (4)External Data	<p>Variation class SNP (source dbSNP 131 - Variants (including SNPs and indels) imported from dbSNP)</p> <p>Synonyms Affy GeneChip 100K Array SNP_A-1671718 Illumina_Human660W-quad rs10497726 Illumina_Human1M-duoV3 rs10497726 dbSNP rs57991142</p> <p>Present in 1000 genomes (1000 genomes - High coverage - CEU trio, 1000 genomes - Low coverage), Phenotype-associated variations (GWAS_BMC_PAPER), HapMap (HapMap Phase II, HapMap Phase III), Clinical/LSDB variations from dbSNP, ENSEMBL:Venter</p> <p>Alleles C/A (Ambiguity code: M) <i>Ancestral allele:</i> C</p> <p>Location This feature maps to 1 genomic location. 2:193051320 (forward strand) Jump to region in detail</p> <p>Configure this page</p> <p>Manage your data</p> <p>Export data</p> <p>Bookmark this page</p>

« [Phenotype Data](#) [Phylogenetic Context](#) [help](#) » [External Data](#)

Alignment:

		M
Homo sapiens	ATAGACTTGC	C
Ancestral sequences	ATAGACTTGCC	GATCCTTTCA
Pan troglodytes	ATAGACTTGCCG	GATCCTTTCA
Ancestral sequences	ATAGACTTGCT	GATCCTTTCA
Gorilla gorilla	ATAGACTTGCT	GATCCTTTCA
Ancestral sequences	ATAGACTTGCT	GATCCTTTCA
Pongo pygmaeus	ATAGACTTGCT	GATCCTTTCA
Ancestral sequences	ATAGACTTGCT	GATCCTTTCA
Macaca mulatta	CTAGACTTGCT	GATCCTTTCA
Ancestral sequences	ATAGACTTGCT	GATCCTTTCA
Callithrix jacchus	ATAGACTTGCT	GATCCTTTCA

Variation annotation- Variant Effect Predictor

Input file

Species: Human (Homo sapiens): GRCh37

Name for this upload (optional):

Paste file:

Upload file: Choose File No file chosen

or provide file URL:

Input file format: Ensembl default

Options

Get regulatory region consequences (human and mouse only):

Type of consequences to display: Ensembl terms

Check for existing co-located variants: Yes

Return results for variants in coding regions only:

Show HGNC identifier for genes where available:

Show Ensembl protein identifiers where available:

Show HGVS identifiers for variants where available: No

Non-synonymous SNP predictions (human only)

SIFT predictions: No

PolyPhen predictions: No

Condel consensus (SIFT/PolyPhen) predictions: No

Frequency filtering of existing variants (human only)

Filter variants by frequency:

NB: Enabling frequency filtering may be very slow for large datasets

Filter: Exclude variants with MAF greater than 0.1 in any 1KG low coverage population

Next >

What can Ensembl tell me about my allele change?

All species

- Ensembl: 50+ species
- Ensembl genomes: 300+

Support for multiple file formats:

- VCF, Pileup, HGVS, rsIDs

Filter input by frequency data



Variation annotation – VEP output

Variant Effect Predictor Results:

[Download text version](#)

Show	All	entries	Show/hide columns			Filter	CSV	
Uploaded Variation	Location	Allele	Gene	Feature	Feature type	Consequence	Position in cDNA	Position in CDS
12_1017956_T/A	12:1017956	A	ENSG00000060237	ENST00000537501	Transcript	STOP_LOST	4612	973
1_881907_-C	1:881906-881907	C	ENSG00000188976	ENST00000327044	Transcript	FRAMESHIFT_CODING	1728-1729	1678-167
21_34029195_A/G	21:34029195	G	ENSG00000159082	ENST00000382491	Transcript	NON_SYNONYMOUS_CODING	2707	2582
21_34029195_A/G	21:34029195	G	ENSG00000159082	ENST00000322229	Transcript	NON_SYNONYMOUS_CODING	2597	2597
21_34029195_A/G	21:34029195	G	ENSG00000159082	ENST00000357345	Transcript	NON_SYNONYMOUS_CODING	2722	2597

Extendable tools, Cloud ready



BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors [Account · Logout](#)

We provide a number of ready-made tools for processing your data. At the moment, small datasets can be uploaded to our servers and processed online; for larger datasets, we provide an API script that can be downloaded (you will also need to [install our Perl API](#) to use these).

In the near future we aim to offer an intermediate service, whereby medium-to-large data sets can be submitted to a queue, similar to BLAST.

Currently available:

Tool	Description		
Assembly converter	Map your data to the current assembly. Accepted file formats: GFF , GTF , BED , PSL N.B. Export is currently in GFF only	 Upload your data	 Download Perl script
ID History converter	Convert a set of Ensembl IDs from a previous release into their current equivalents.	 Upload your data (max 30 ids)	 Download Perl script
Region Report	Export standard data sets (genes, sequence, variations, etc) from one or more regions, in either GFF or simple text format	 Upload your data (max 5MB total, or 1MB for variation/regulation data)	 Download Perl script
Variant Effect Predictor	(Formerly SNP Effect Predictor). Upload a set of variants in our standard format or VCF format and export a file containing consequence types. Uploaded tracks can also be viewed on Location pages. See also full documentation	 Upload your data (max 750 variants)	 Download Perl script

Ensembl release 66 - Feb 2012 © [WTSI](#) / [EBI](#)

[About Ensembl](#) | [Contact Us](#) | [Help](#)

[Permanent link](#)



VEP Plugins

Plugin name	Description
Downstream	provides the predicted downstream protein sequence, relative change in length relative to the reference protein
Draw GD	Draws transcript models using GD
LD.pm	finds variants in LD with overlapping existing variants
Non-synonymous	limits output to non-synonymous
CCDS filter	limits output to variants in CCDS transcripts
TSS distance	calculates the distance from TSS for upstr variant
Conservation	retrieves a conservation score from Compara
Blosum 62	adds BLOSUM 62 substitution matrix score for variant