DIU Génétique et Reproduction

Introduction à l'analyse des Variants

18 Novembre 2022

Philippe Dessen
Institut Gustave Roussy, Villejuif

dessen@igr.fr

http://pdessen.free.fr/KB

Abdelkader Heddar
Hôpital Bicêtre

abdelkader.heddar@aphp.fr

Recherche de la nature pathogène des variants

exemple : variants de HARS2, dans le syndrome de Perrault

Recherche de HARS2 sur Uniprot et UCSC et autres bases

Bases de polymorphismes ..

Catégorisation du phénotype

La logique d'analyse voudrait que l'on parte de variants identifiés expérimentalement pour rechercher dans les différentes bases de mutations ou polymorphisme, la nature du phénotype et une relation avec la maladie

Pour introduire par un exemple on prendra les variants Leu200Val et Val368Leu dans le gène de HARS2

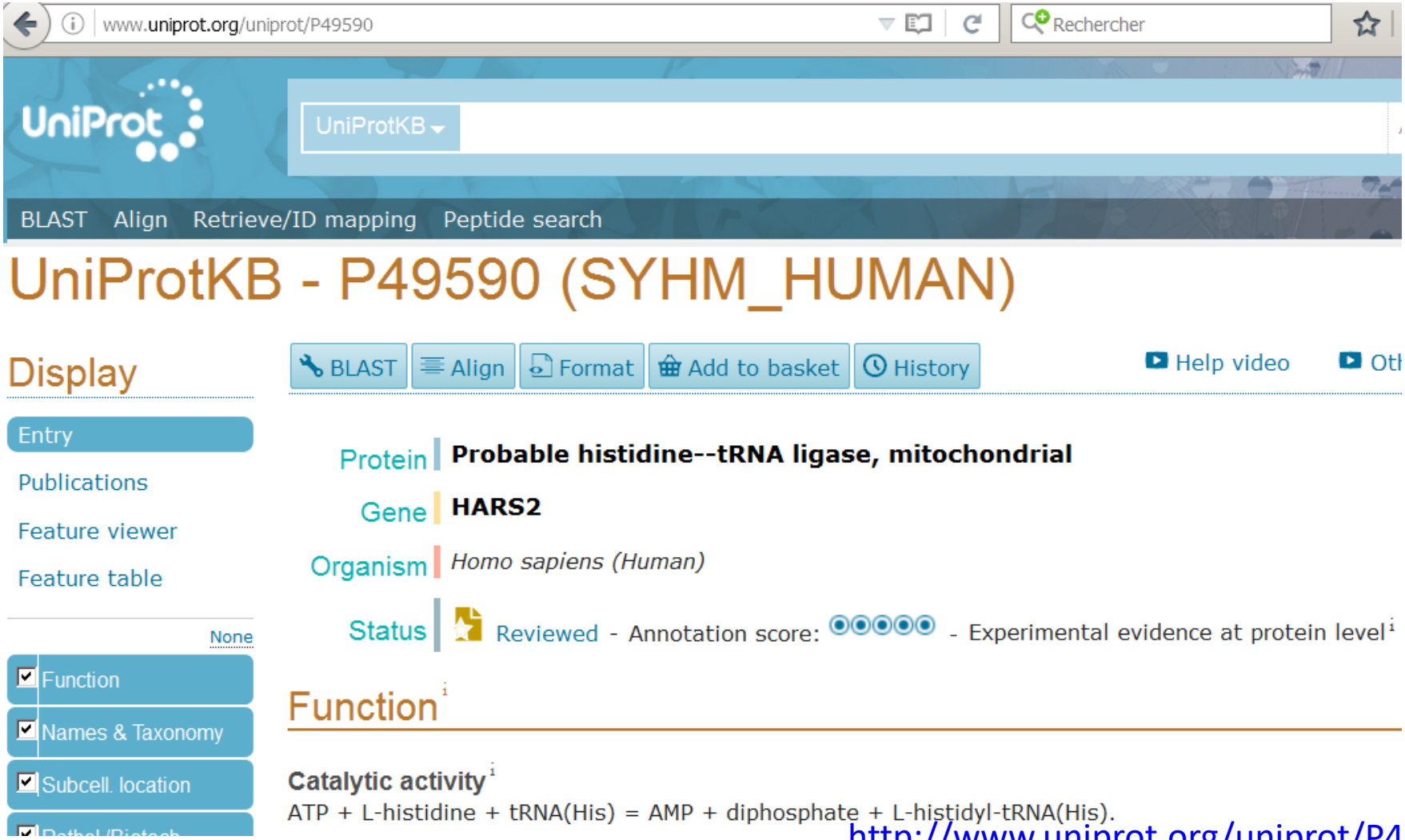
Le syndrome de Perrault (SP) est caractérisé par l'association d'une dysgénésie ovarienne chez une femme ayant une surdité neurosensorielle. Dans de récentes études, quelques auteurs ont décrit des anomalies neurologiques, en particulier une ataxie cérébelleuse progressive et un déficit intellectuel.

https://www.orpha.net/consor/cgi-bin/OC_Exp.php?lng=FR&Expert=2855

UniProt [1]

UniProt

Uniprot : base internationale sur les protéines : séquences, organisation , fonctions, évolution
Contient une section **Variants/Diseases**



http://www.uniprot.org/uniprot/P49590

UniProt [2]

Tools ▾ Download Add Highlight ▾ Copy sequence

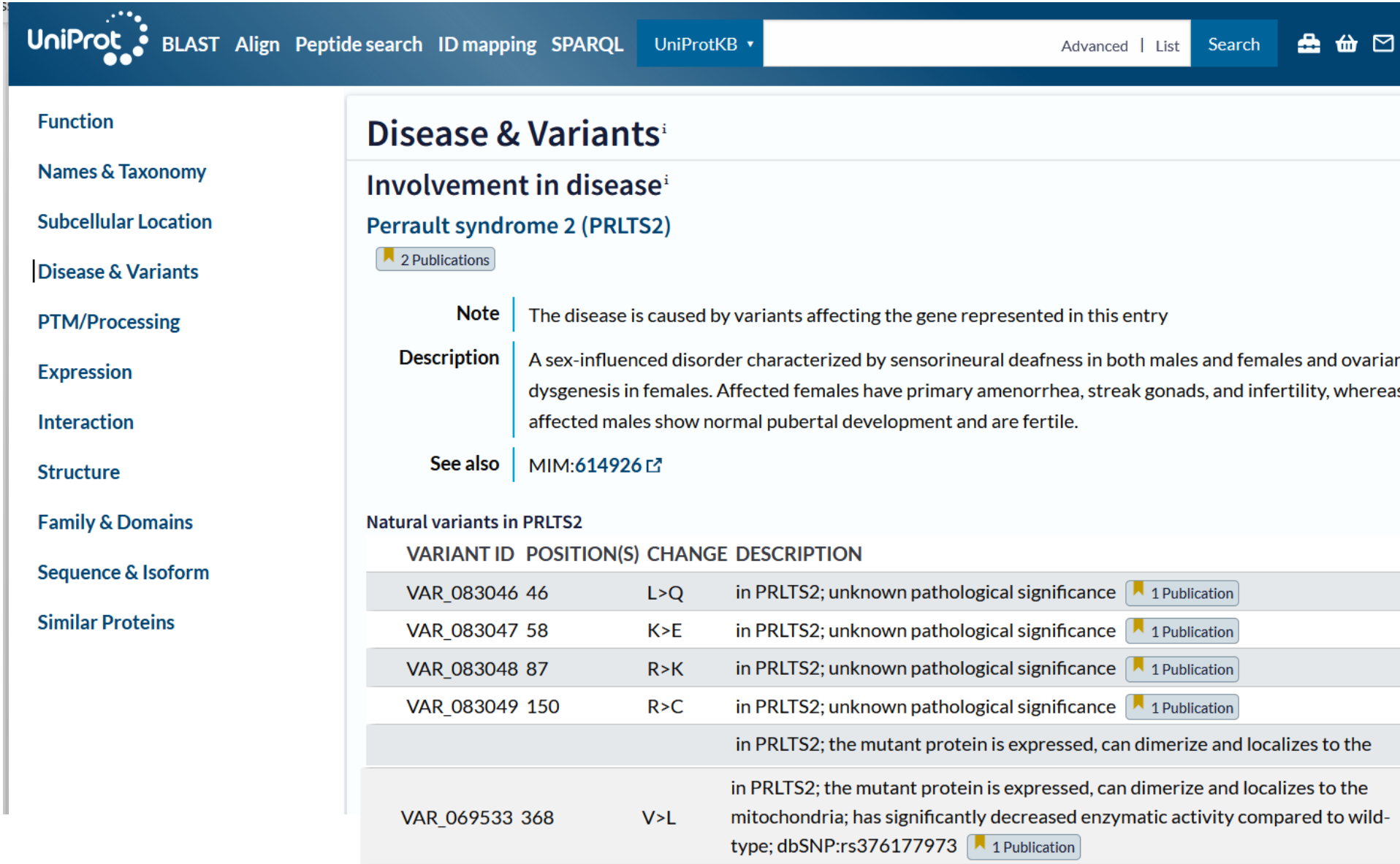
Length 506
Mass (Da) 56,888

Last updated 1996-02-01 v1
Checksum[i] E1CE879837AE26E7

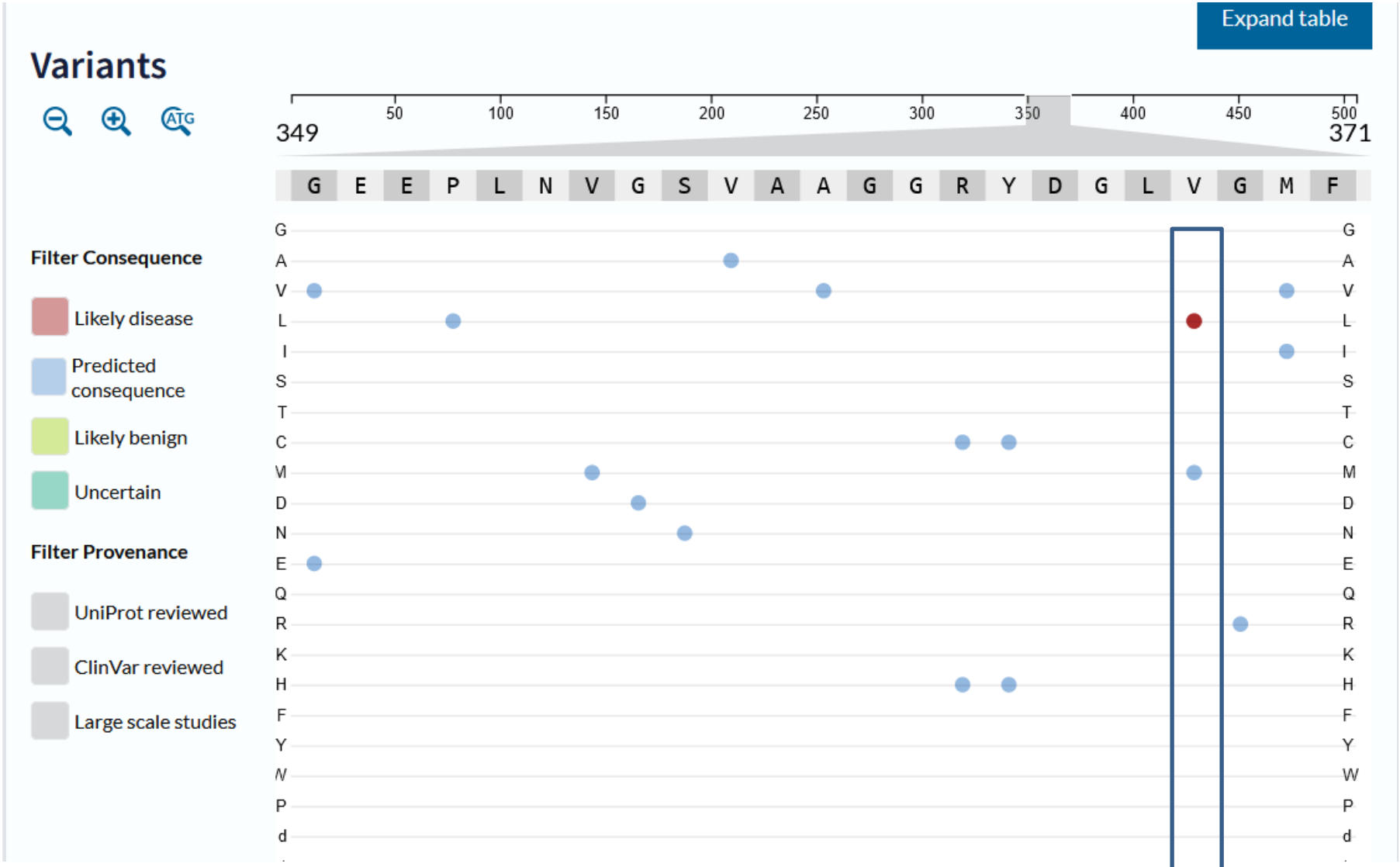
```
        10         20         30         40         50         60         70         80
MPLLGLLPRR AWASLLSQLL RPPCASCTGA VRCQSQVAEA VLTSQLKAHQ EKPNFIIKTP KGTRDLSPQH MVVREKILDL

        90        100        110        120        130        140        150        160
VISCFKRHGA KGMDTPAFEL KETLTEKYGE DSGLMYDLKD QGGELLSLRY DLTVPFARYL AMNKVKKMKR YHVGKVWRRE

       170        180        190        200        210        220        230        240
SPTIVQGRYR EFCQCDFDIA GQFDPMIPDA ECLKIMCEIL SGLQLGDFLI KVNDRRIVDG MFAVCGVPES KFRAICSSID

       250        260        270        280        290        300        310        320
KLDKMAWKDV RHEMVVKKGL APEVADRIGD YVQCHGGVSL VEQMFQDPRL SQNKQALEGL GDLKLLFEYL TLFGIADKIS

       330        340        350        360        370        380        390        400
FDLSLARGLD YYTGVIYEAV LLQTPTQAGE EPLNVGSVAA GGRYDGLVGM FDPKGHKVPC VGLSIGVERI FYIVEQRMKT

       410        420        430        440        450        460        470        480
KGEKVRTTET QVFVATPQKN FLQERLKLIA ELWDSGIKAE MLYKNNPKLL TQLHYCESTG IPLVVIIGEQ ELKEGVIKIR

       490        500
SVASREEVAI KRENFVAEIQ KRLSES
```

Val368Leu ou Leu200Val

UniProt [3]


UniProt
BLAST
Align
Peptide search
ID mapping
SPARQL
UniProtKB
Advanced | List
Search
Function
Names & Taxonomy
Subcellular Location
Disease & Variants
PTM/Processing
Expression
Interaction
Structure
Family & Domains
Sequence & Isoform
Similar Proteins
Disease & Variants
Involvement in disease
Perrault syndrome 2 (PRLTS2)
2 Publications
Note
The disease is caused by variants affecting the gene represented in this entry
Description
A sex-influenced disorder characterized by sensorineural deafness in both males and females and ovarian dysgenesis in females. Affected females have primary amenorrhea, streak gonads, and infertility, whereas affected males show normal pubertal development and are fertile.
See also
MIM:614926
Natural variants in PRLTS2
VARIANT ID POSITION(S) CHANGE DESCRIPTION
VAR_083046 46 L>Q in PRLTS2; unknown pathological significance 1 Publication
VAR_083047 58 K>E in PRLTS2; unknown pathological significance 1 Publication
VAR_083048 87 R>K in PRLTS2; unknown pathological significance 1 Publication
VAR_083049 150 R>C in PRLTS2; unknown pathological significance 1 Publication
in PRLTS2; the mutant protein is expressed, can dimerize and localizes to the
VAR_069533 368 V>L in PRLTS2; the mutant protein is expressed, can dimerize and localizes to the mitochondria; has significantly decreased enzymatic activity compared to wild-type; dbSNP:rs376177973 1 Publication

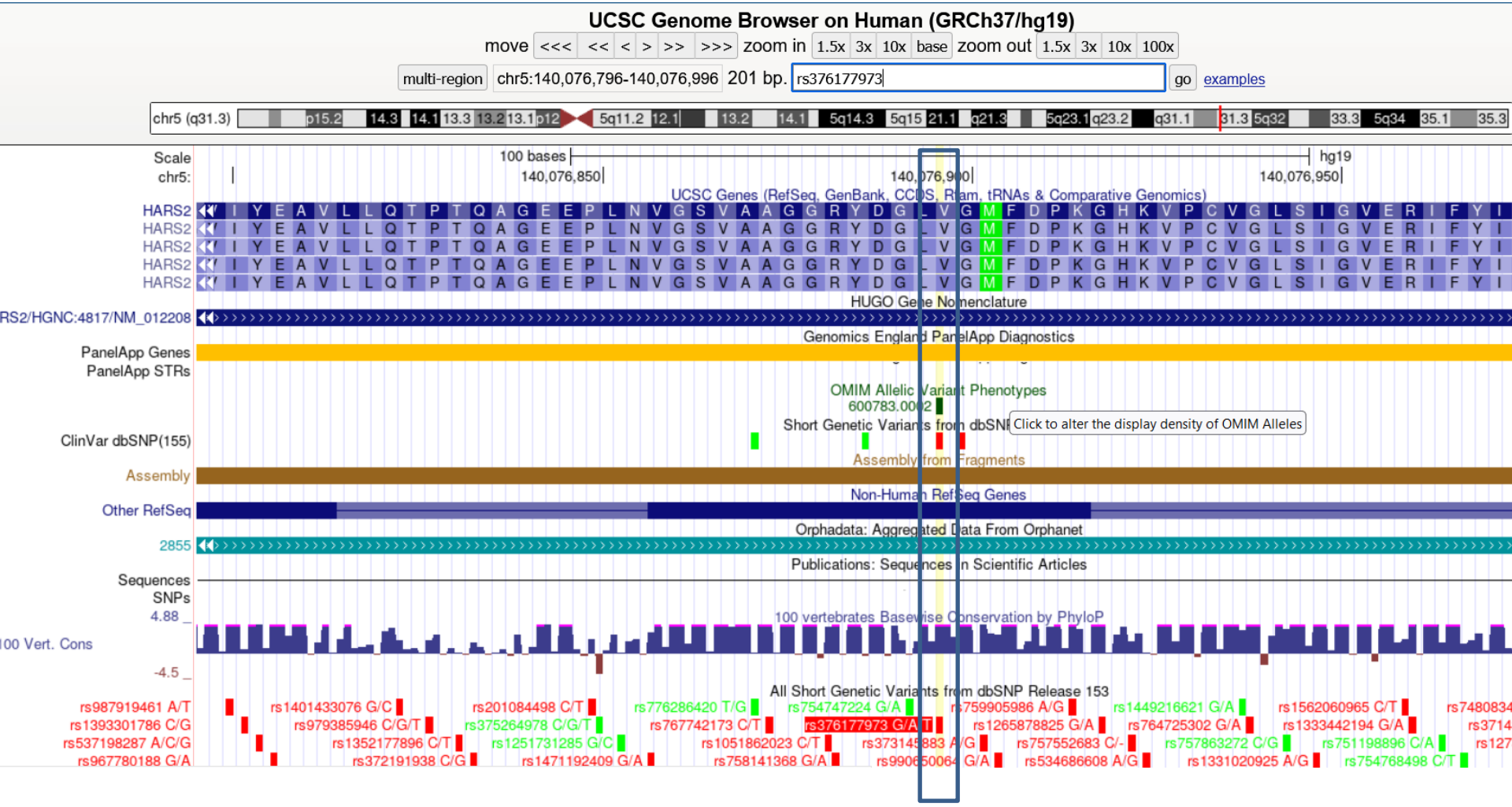
UniProt [4]



Leu200Val	rs397515410	G>A
Val368Leu	rs276177973	G>T

UCSC [1]

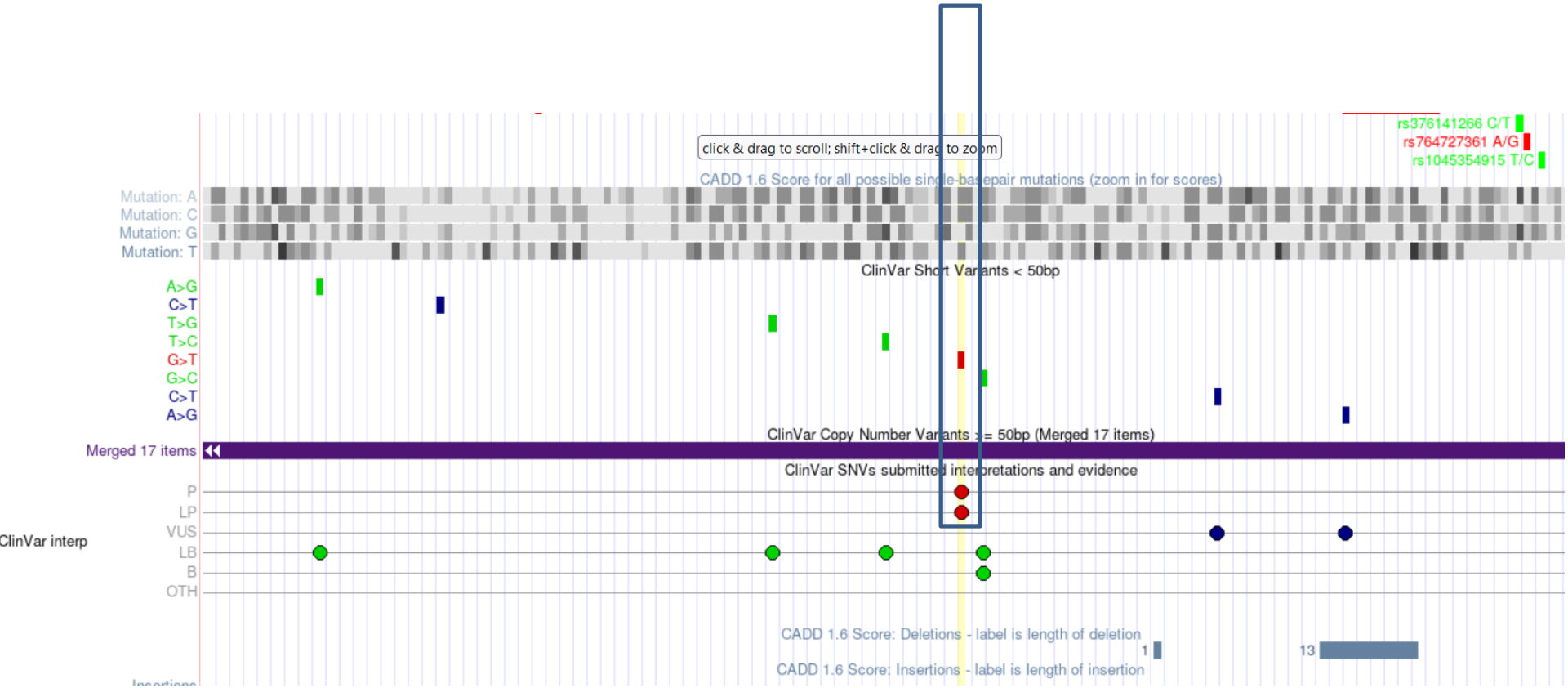
Recherche sur le GoldenPath (hg19) du SNP rs376177973



Val368Leu SNP rs376177973 hg19 : [chr5:140076896-140076896](https://genome.ucsc.edu)

https://genome.ucsc.edu

UCSC [2]



Un clic sur le variant rouge -> documentation

UCSC [3]

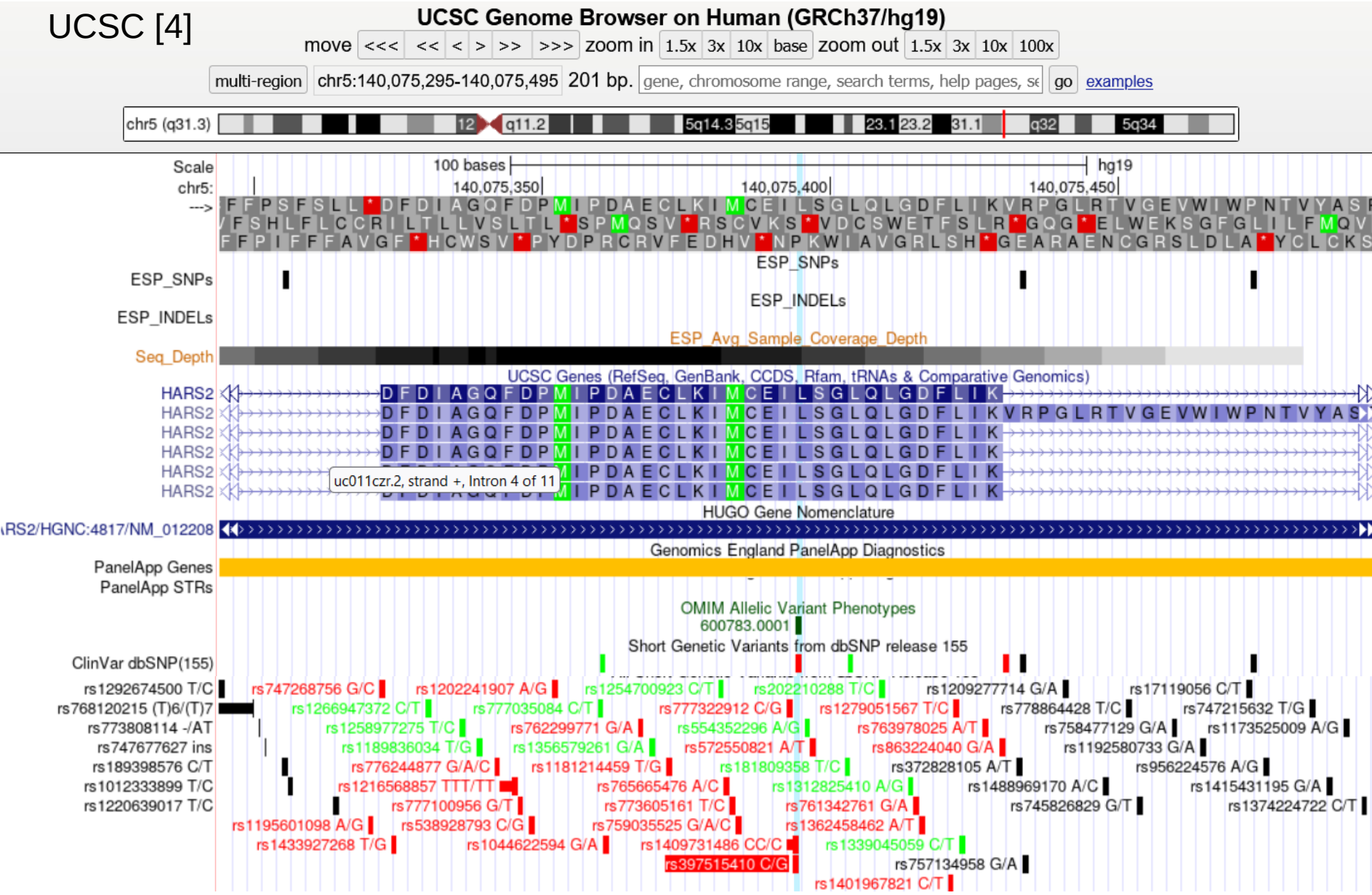
ClinVar SNVs submitted interpretations and evidence

Position: chr5:140076896-140076896
Band: 5q31.3
Genomic Size: 1

There are 1 submissions at this position with clinical significance 'Pathogenic'.

Link to ClinVar with Variant ID	NM_012208.4(HARS2):c.1102G>T (p.Val368Leu)
SubmittedGeneSymbol	HARS2
Molecular Consequence	missense variant
dbSNP ID	376177973
Variation ID	208286
Clinical Significance	Pathogenic
Variant Review Status	☆☆☆☆ based on: no assertion criteria provided
Date Last Evaluated	Apr 19, 2011
Description	-
Submitted Phenotype Info	PERRAULT SYNDROME 2
Reported Phenotype Info	C3554105:Perrault syndrome 2
Submission Review Status	no assertion criteria provided
Collection Method	literature only
Origin Counts	germline:na
Submitter	OMIM
SCV	SCV000056589.4
Explanation of Interpretation	-

UCSC [4]



Leu200Val **dbSNP:** rs397515410 hg19 chr5:140075395-140075395

UCSC [5]

All Short Genetic Variants from dbSNP Release 153 (rs397515410)

dbSNP: rs397515410
Position: chr5:140075395-140075395
Band: 5q31.3
Genomic Size: 1
View DNA for this feature (hg19/Human)

Reference allele: C
Alternate allele: G
Allele frequency counts:

Allele	TOPMED	PAGE_STUDY	GnomAD
C	**125567/125568 (0.999992)**	**78698/78698 (1.000000)**	**31397/31398 (0.999968)**
G	1/125568 (0.000008)	0/78698 (0.000000)	1/31398 (0.000032)

Functional effects: missense_variant, coding_sequence_variant, 5_prime_UTR_variant
ClinVar: RCV000032820.7 (pathogenic-likely-pathogenic)
Submitted by: GNOMAD, ILLUMINA, OMIM-CURATED-RECORDS, PAGE_CC, TOPMED
Publications in PubMed: PMID517579 , PMID21464306
Variation class/type: snv

Interesting or anomalous conditions noted by UCSC:

- Variant is in ClinVar.
- Variant is in ClinVar with clinical significance of pathogenic and/or likely pathogenic.
- Variant is "rare", i.e. has a Minor Allele Frequency of less than 1% in all projects reporting frequencies
- This variant overlaps another variant with a different type/class.

Bases de variants

Attention : la localisation dépend du système de coordonnées des variants sur le génome :

choisir le « build » adéquat :

- hg19 (2009) majoritaire dans les bases de polymorphismes
- hg38 (2013)

Normalement , les sorties tabulées donnent les coordonnées et la transition en bases du variant (REF > ALT) pour un build donné (hg19 ou hg38)

On peut donc accéder aux informations des BD avec soit

- la localisation génomique,
- un code répertorié (ex SNP : rsxxx)
- le symbole du gène ….

dbSNP Single Nucleotide Polymorphism (NCBI, Bethesda, Us)

https://www.ncbi.nlm.nih.gov/SNP/overview.html

A Database of Single Nucleotide Polymorphisms : A key aspect of research in genetics is associating sequence variations with heritable phenotypes.

Exome Variant server (EVS) (Washington, Us)

https://evs.gs.washington.edu/EVS/

gnomAD (Broad Institute, Boston, Us)

https://gnomad.broadinstitute.org/

The Genome Aggregation Database (gnomAD) is a resource developed by an international coalition of investigators, with the goal of aggregating and harmonizing both exome and genome sequencing data from a wide variety of large-scale sequencing projects, and making summary data available for the wider scientific community.

Varsome (US)

https://varsome.com/

VarSome is a search engine, aggregator and impact analysis tool for human genetic variation and a community-driven project aiming at sharing global expertise on human variants.

M-CAP (US)

http://bejerano.stanford.edu/mcap/

Mendelian Clinically Applicable Pathogenicity (M-CAP) Score M-CAP is the first pathogenicity classifier for rare missense variants in the human genome that is tuned to the high sensitivity required in the clinic (see Table). By combining previous pathogenicity scores (including SIFT, Polyphen-2 and CADD) with novel features and a powerful model, we attain the best classifier at all thresholds, reducing a typical exome/genome rare (<1%) missense variant (VUS) list from 300 to 120, while never mistaking 95% of known pathogenic variants as benign.

Varity (US)

http://varity.varianteffect.org/

VARITY (Improved pathogenicity prediction for rare human missense variants) Web Application User Guide. This web application provides: 1) Search and visualize VARITY predictions, features and feature contributions for all possible single nucleotide change missense variants for each o 18,239 human proteins

dbVar (Bethesda, NCBI, Us)

https://www.ncbi.nlm.nih.gov/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

http://pdessen.free.fr/atlas_links.html

OMIM Online Mendelian Inheritance in Man" (John Hopkins, Baltimore, Us)

https://www.omim.org/

OMIM is a catalog of human genes and genetic disorders authored and edited by Dr. Victor A. McKusick and his colleagues at Johns Hopkins and elsewhere, and developed for the World Wide Web by NCBI, the National Center for Biotechnology Information

ORPHANET : Database of rare diseases and orphan drugs (INSERM, Paris, Fr)

https://www.orpha.net/consor/cgi-bin/index.php

This project is the result of a commonly observed fact: rare diseases are difficult to deal with for medical practitioners. This is due to their restricted knowledge of the diseases' natural history, the patient care required, treatment, and sometimes even of its existence.

MedGen (NCBI, Bethesda, Us)

https://www.ncbi.nlm.nih.gov/medgen/

MedGen is NCBI portal to information about human disorders and other phenotypes having a genetic component. MedGen is structured to serve health care professionals, the medical genetics community, and other interested parties by providing centralized access to diverse types of content.

dbGap (NCBI, Bethesda, Us)

https://www.ncbi.nlm.nih.gov/gap/

The database of Genotypes and Phenotypes (dbGaP) was developed to archive and distribute the data and results from studies that have investigated the interaction of genotype and phenotype in Humans.

ClinVar (NCBI, Bethesda, Us)

https://www.ncbi.nlm.nih.gov/clinvar/

ClinVar is designed to provide a freely accessible, public archive of reports of the relationships among human variations and phenotypes, with supporting evidence.

ClinGen : Clinical Genome resource (NIH, Bethesda, Us)

https://www.clinicalgenome.org/

ClinGen is a National Institutes of Health (NIH)-funded resource dedicated to building an authoritative central resource that defines the clinical relevance of genes and variants for use in precision medicine and research.

HuGE Navigator (Us)

https://phgkb.cdc.gov/PHGKB/phgHome.action?action=home

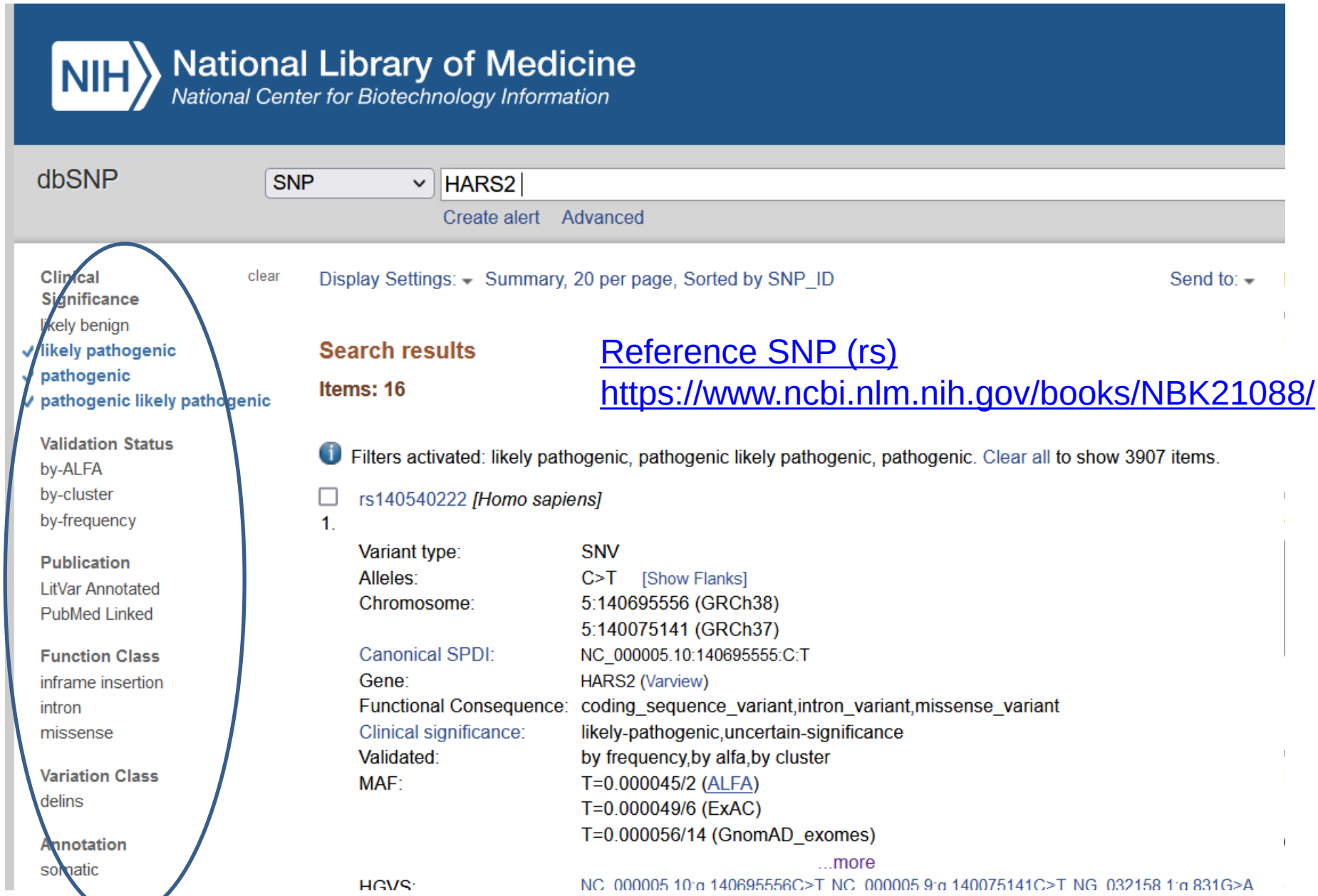
HuGE Navigator provides access to a continuously updated knowledge base in human genome epidemiology, including information on population prevalence of genetic variants, gene-disease associations, gene-gene and gene- environment interactions, and evaluation of genetic tests

DisGeNET (Es)

https://www.disgenet.org/home/

DisGeNET is a discovery platform containing one of the largest publicly available collections of genes and variants associated to human diseases (Piñero et al., 2016; Piñero et al., 2015).

dbSNP [1] Base originale de dépot de polymorphisme



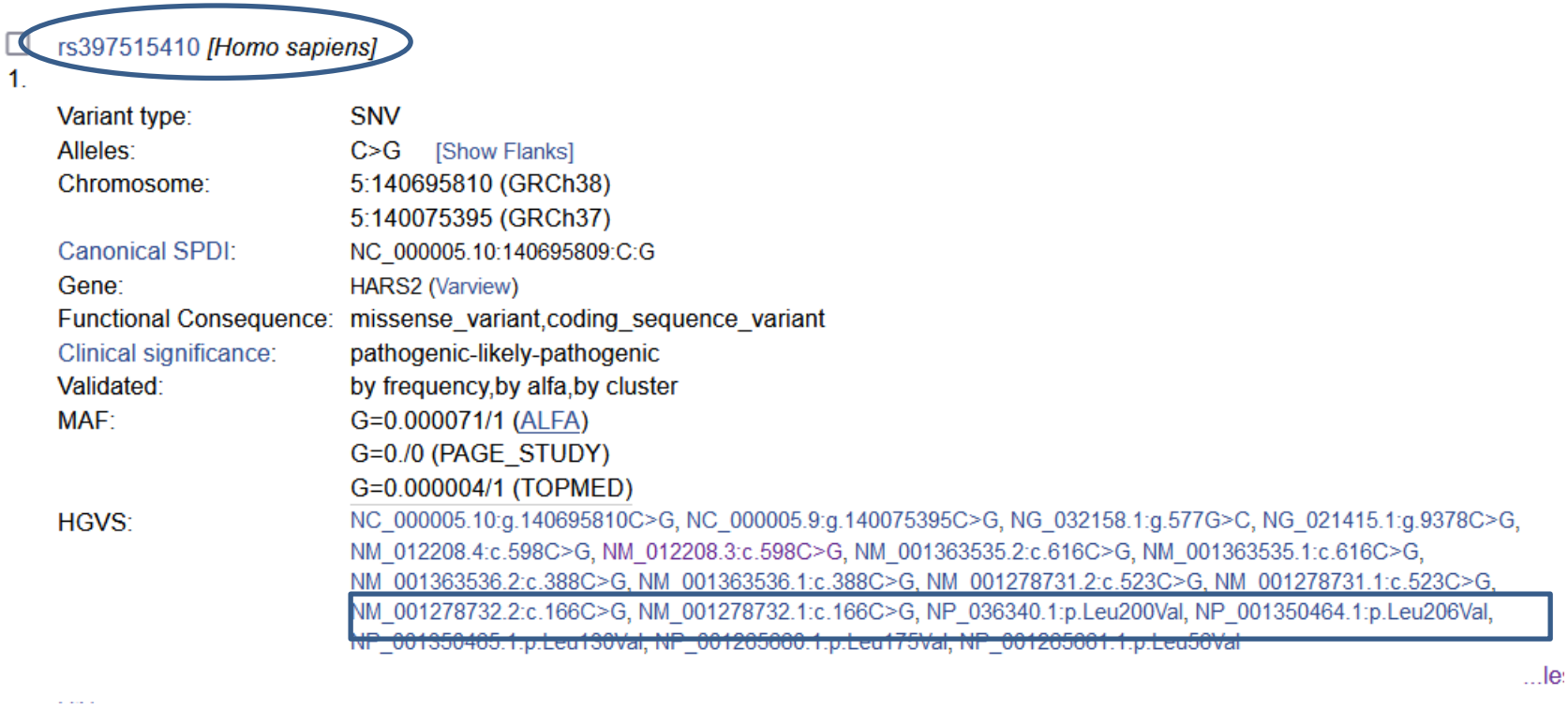
[Reference SNP (rs)](https://www.ncbi.nlm.nih.gov/books/NBK21088/)
https://www.ncbi.nlm.nih.gov/books/NBK21088/

Filtres

https://www.ncbi.nlm.nih.gov/SNP

dbSNP [2] Suite du listing

Ici, on liste les variants dans l'ordre du nom (rs - pour reference snp)



rs397515410 *[Homo sapiens]*

1.

Variant type:	SNV
Alleles:	C>G [Show Flanks]
Chromosome:	5:140695810 (GRCh38) 5:140075395 (GRCh37)
Canonical SPDI:	NC_000005.10:140695809:C:G
Gene:	HARS2 (Varview)
Functional Consequence:	missense_variant,coding_sequence_variant
Clinical significance:	pathogenic-likely-pathogenic
Validated:	by frequency,by alfa,by cluster
MAF:	G=0.000071/1 (ALFA) G=0./0 (PAGE_STUDY) G=0.000004/1 (TOPMED)
HGVS:	NC_000005.10:g.140695810C>G, NC_000005.9:g.140075395C>G, NG_032158.1:g.577G>C, NG_021415.1:g.9378C>G, NM_012208.4:c.598C>G, NM_012208.3:c.598C>G, NM_001363535.2:c.616C>G, NM_001363535.1:c.616C>G, NM_001363536.2:c.388C>G, NM_001363536.1:c.388C>G, NM_001278731.2:c.523C>G, NM_001278731.1:c.523C>G, NM_001278732.2:c.166C>G, NM_001278732.1:c.166C>G, NP_036340.1:p.Leu200Val, NP_001350464.1:p.Leu206Val, NP_001350465.1:p.Leu130Val, NP_001265060.1:p.Leu175Val, NP_001265061.1:p.Leu58Val

...le

Leu200Val rs397515410

Coordonnées relatives aux transcrits

dbSNP [3] Suite du listing

Clinical Significance clear
- ✓ likely benign
- ✓ likely pathogenic
- ✓ pathogenic
- ✓ pathogenic likely pathogenic

Validation Status
by-ALFA
by-cluster
by-frequency

Publication
LitVar Annotated
PubMed Linked

Function Class
missense

Global MAF
Custom range...

Clear all

Show additional filters

Display Settings: Summary Send to:

Filters activated: likely benign, likely pathogenic, pathogenic likely pathogenic, pathogenic. Clear all to show 1 items.

1. rs376177973 *[Homo sapiens]*

Variant type: SNV
Alleles: G>A,T [Hide Flanks]

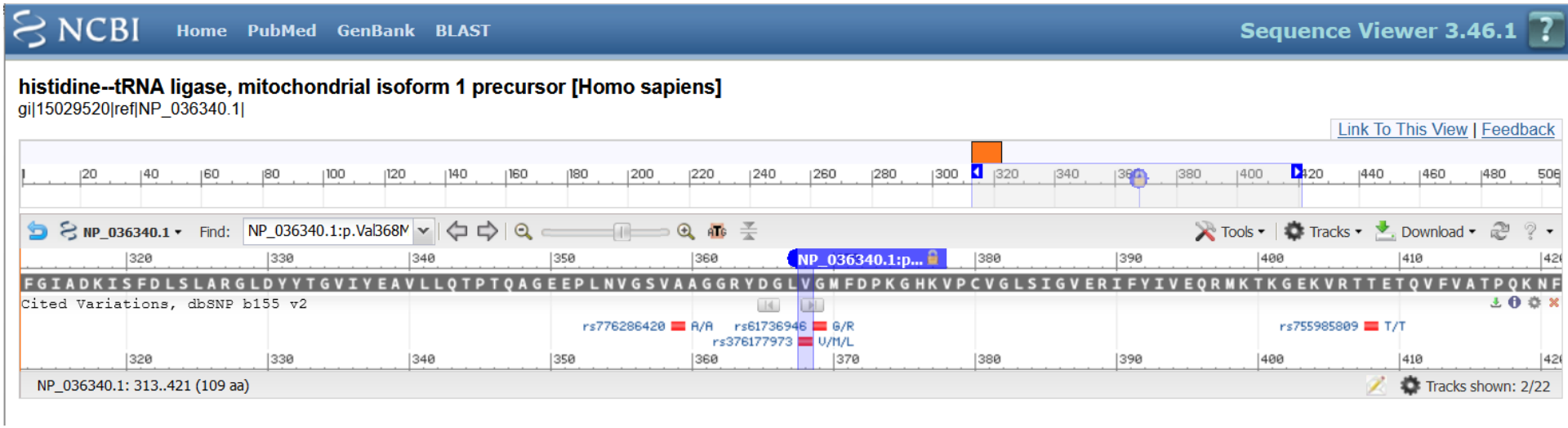
```
AGTGATCTATGAAGCAGTGCTGCTGCAGACCCCAACTCAGGCTGGGGAGG
AGCCCCTGAATGTGGGCAGTGTGGCTGCTGGTGGGCGCTATGATGGGCTG
[G/A/T]
TGGGCATGTTTGACCCCAAGGGCCACAAGGTGCCATGTGTGGGACTCAGC
ATTGGGGTTGAGCGAATCTTCTACATTGTGGAGCAGAGGATGAAGGTAGG
```

Chromosome: 5:140697311 (GRCh38)
5:140076896 (GRCh37)
Canonical SPDI: NC_000005.10:140697310:G:A,NC_000005.10:140697310:G:T
Gene: HARS2 (Varview)
Functional Consequence: missense_variant,coding_sequence_variant
Clinical significance: likely-pathogenic
Validated: by frequency,by alfa,by cluster
MAF: T=0./0 (ALFA)
A=0./0 (TWINSUK)
A=0.000004/1 (TOPMED)
...more

HGVS: NC_000005.10:g.140697311G>A, NC_000005.10:g.140697311G>T, NC_000005.9:g.140076896G>A, NC_000005.9:g.140076896G>T, NG_021415.1:g.10879G>A, NG_021415.1:g.10879G>T, NM_012208.4:c.1102G>A, NM_012208.4:c.1102G>T, NM_012208.3:c.1102G>A, NM_012208.3:c.1102G>T, NM_001363535.2:c.1120G>A, NM_001363535.2:c.1120G>T, NM_001363535.1:c.1120G>A, NM_001363535.1:c.1120G>T, NM_001363536.2:c.892G>A, NM_001363536.2:c.892G>T, NM_001363536.1:c.892G>A, NM_001363536.1:c.892G>T, NM_001278731.2:c.1027G>A, NM_001278731.2:c.1027G>T, NM_001278731.1:c.1027G>A, NM_001278731.1:c.1027G>T, NM_001278732.2:c.670G>A, NM_001278732.2:c.670G>T, NM_001278732.1:c.670G>A, NM_001278732.1:c.670G>T, NP_036340.1:p.Val368Met, NP_036340.1:p.Val368Leu, NP_001350464.1:p.Val374Met, NP_001350464.1:p.Val374Leu

Attention :
La nomenclature HGVS est nécessaire pour une bonne identification

dbSNP [4]



Echelle en AA (368)

Chromosome: 5:140697311 (GRCh38)
5:140076896 (GRCh37)

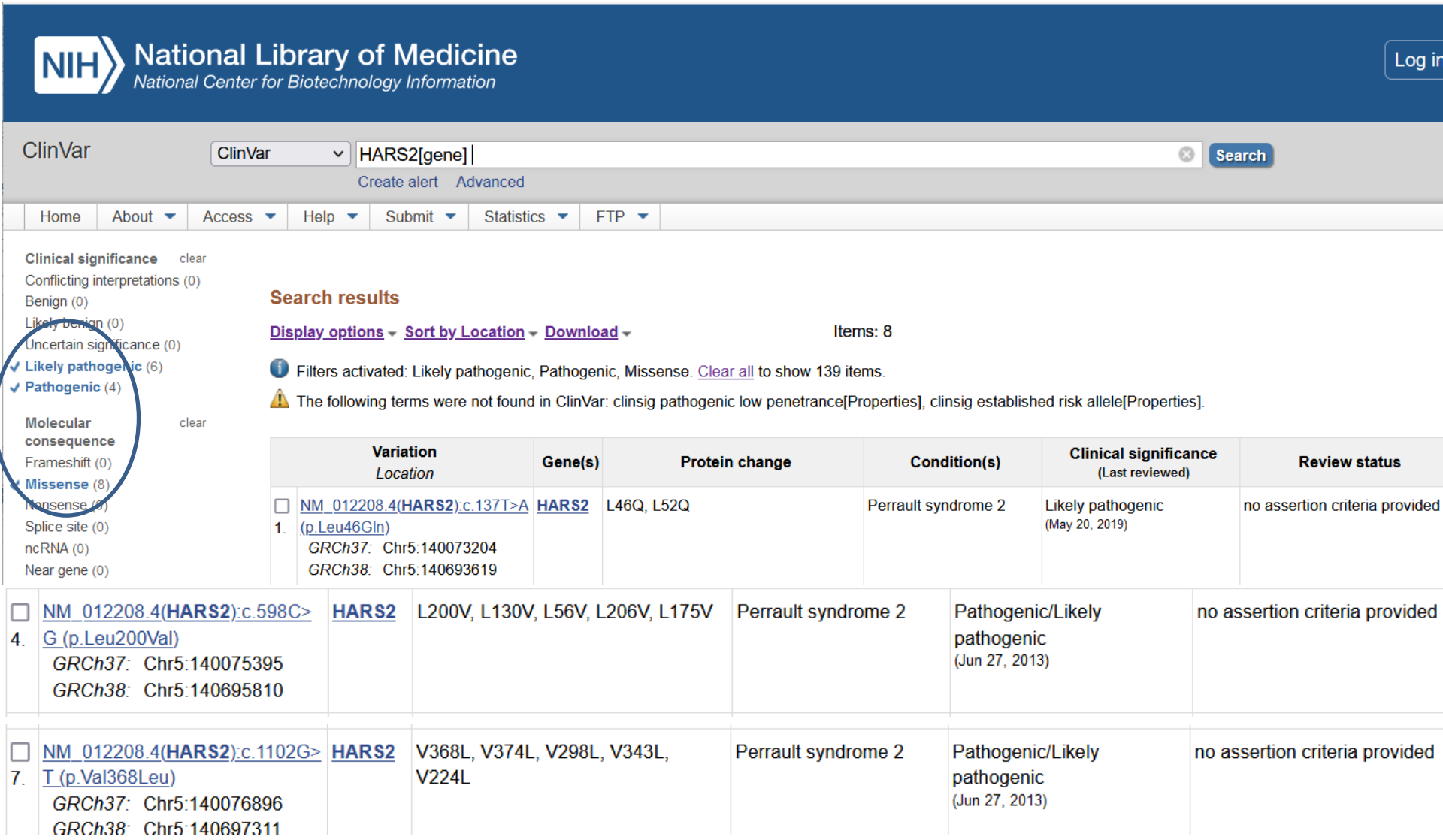
Val368Leu rs376177973
hg19 5:140076896
hg38 5;140697311

Chromosome: 5:140695810 (GRCh38)
5:140075395 (GRCh37)

Leu200Val rs397515410 G>A
hg19 5:140075395
hg38 5:140695810

ClinVar [1]

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



https://www.ncbi.nlm.nih.gov/clinvar/?gr=0&term=HARS2[gene]

ClinVar [2]

NM_012208.4(HARS2):c.1102G>T (p.Val368Leu)

Interpretation: Pathogenic/Likely pathogenic

Review status: ☆☆☆☆ no assertion criteria provided
Submissions: 2
First in ClinVar: Feb 2, 2016
Most recent Submission: Oct 30, 2020
Last evaluated: Jun 27, 2013

Variant details

Conditions

Gene(s)

NM_012208.4(HARS2):c.1102G>T (p.Val368Leu)

Allele ID: 204531
Variant type: single nucleotide variant
Variant length: 1 bp
Cytogenetic location: 5q31.3
Genomic location: 5: 140697311 (GRCh38) GRCh38 UCSC
5: 140076896 (GRCh37) GRCh37 UCSC

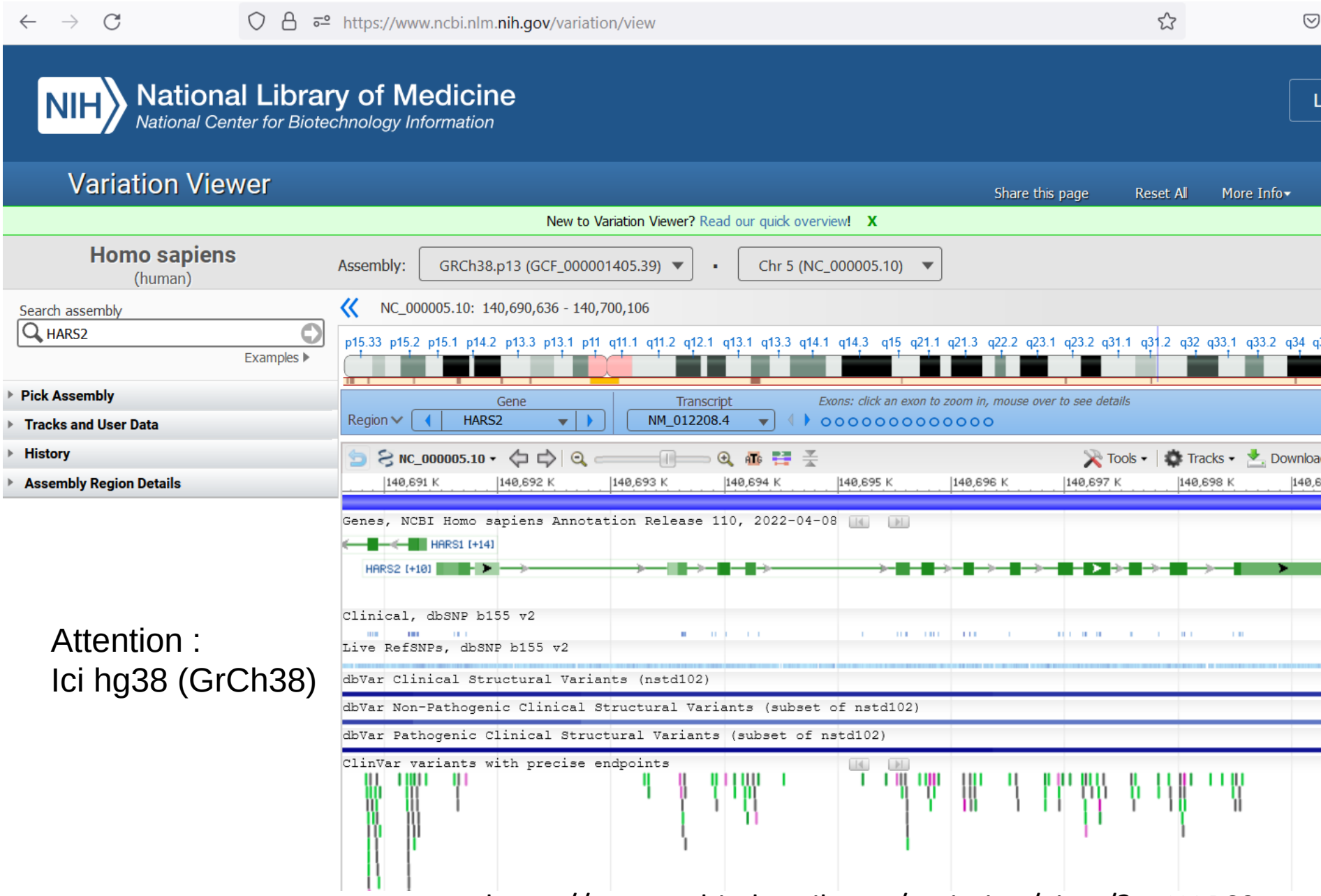
HGVS:

Nucleotide	Protein	Molecular consequence
NM_012208.4:c.1102G>T MANE SELECT	NP_036340.1:p.Val368Leu	missense
NM_001278731.2:c.1027G>T	NP_001265660.1:p.Val343Leu	missense
NM_001278732.2:c.670G>T	NP_001265661.1:p.Val224Leu	missense

... more HGVS

Protein change: V368L, V374L, V298L, V343L, V224L
Other names: -
Canonical SPDI: NC_000005.10:140697310:G:T
Functional consequence: -
Global minor allele frequency (GMAF): -
Allele frequency: NHLBI Exome Sequencing Project (ESP) Exome Variant Server 0.00008
Links: ClinGen: CA350908
UniProtKB: P49590#VAR_069533
OMIM: 600783.0002
dbSNP: rs376177973
VarSome

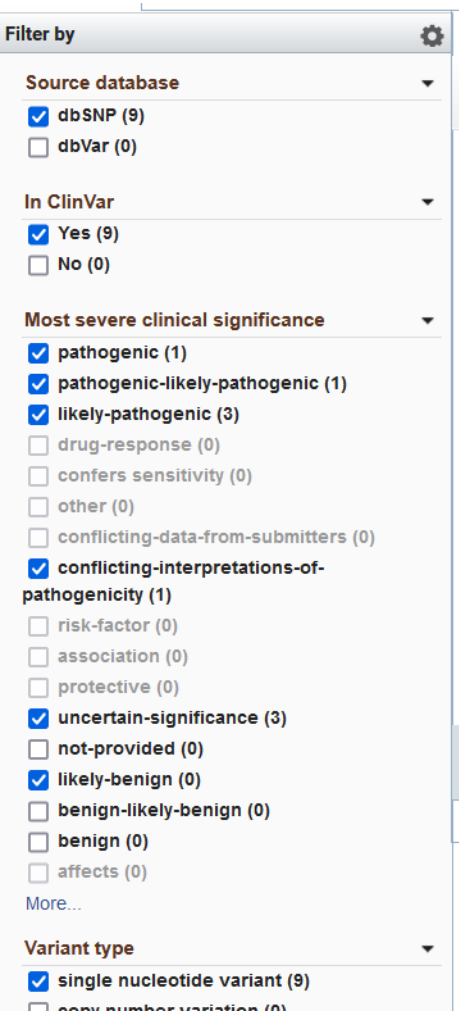
Variation Viewer[1]

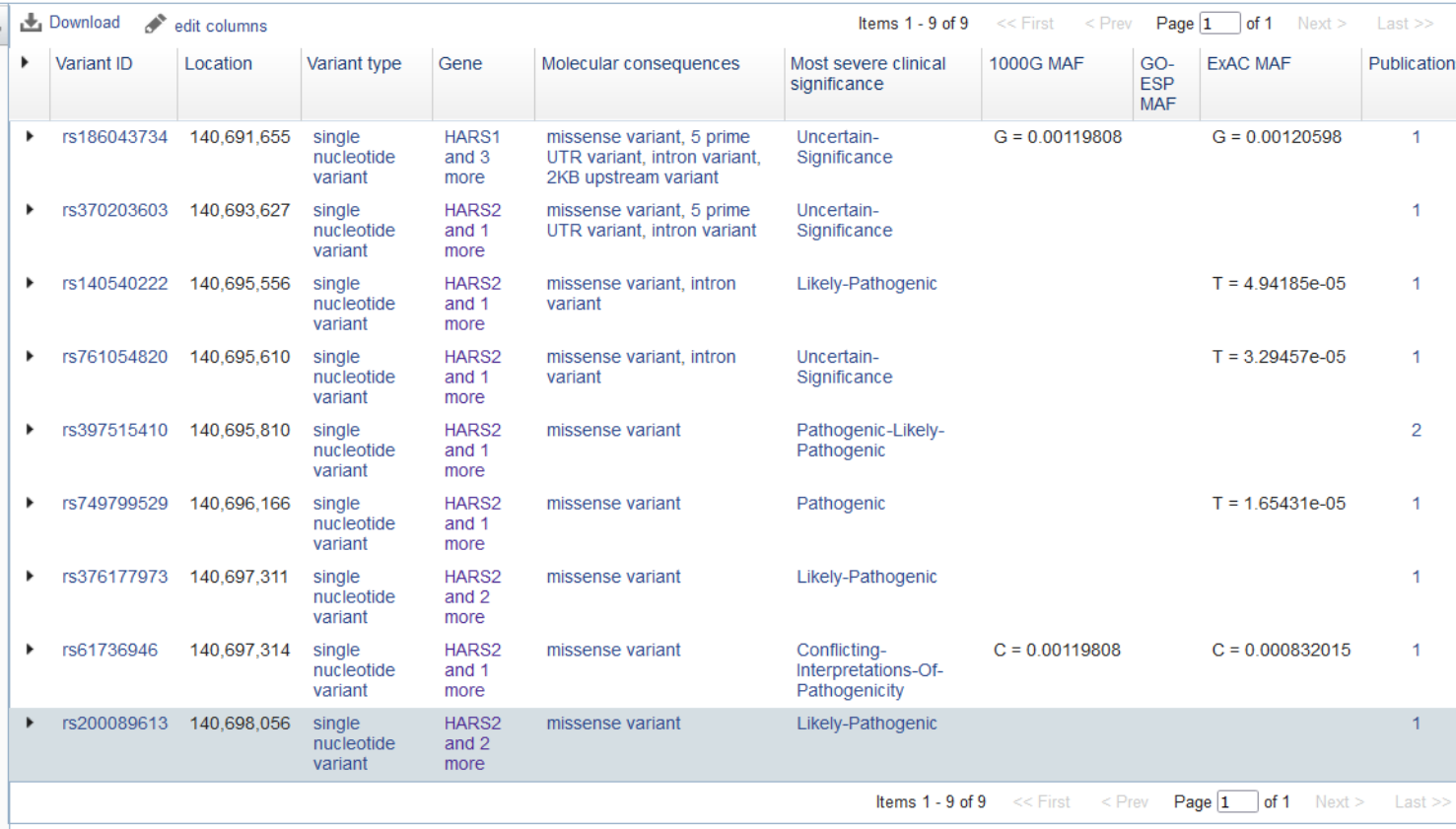


Attention :
Ici hg38 (GrCh38)

https://www.ncbi.nlm.nih.gov/variation/view/?q=HARS2

Variation Viewer[2]





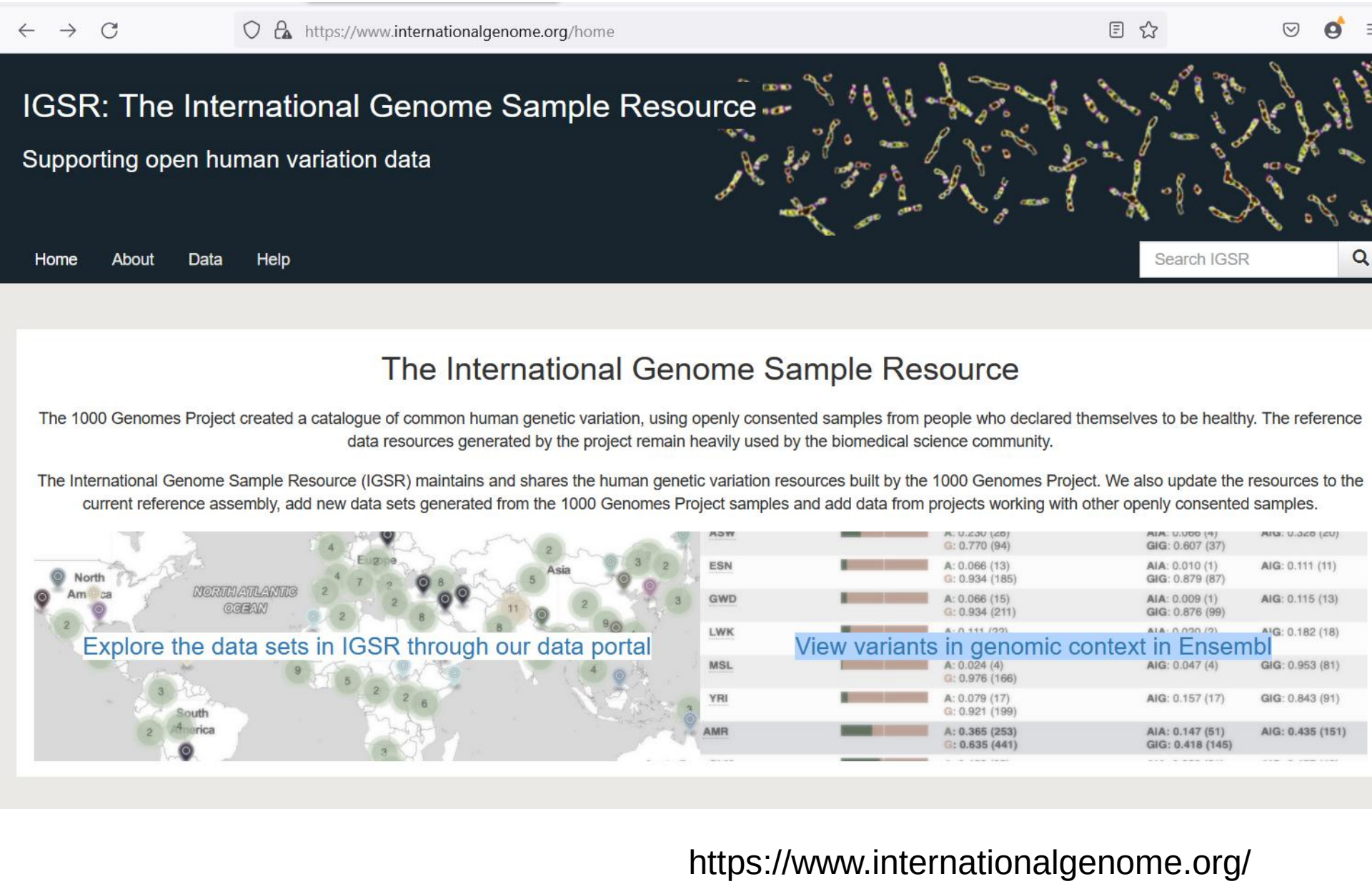
Complet : sélection multiple et filtres

pathogenic
benign
likely benign
uncertain significance
…

IGSR[1]

IGSR

suite du programme 1000 Genomes : n'est pas interrogeable



https://www.internationalgenome.org/

IGSR[2]

Overview

The International Genome Sample Resource (IGSR) was established at EMBL-EBI in January 2015. The resource was established with three main aims, to:

1. Ensure maximal usefulness and relevance of the existing 1000 Genomes data resources
2. Extend the resource for the existing populations
3. Expand the resource to new populations

The first aim will start with the remapping of the existing low coverage and exome data to the new version of the human reference assembly, GRCh38. The second aim will bring together other functional and sequence data that has been generated on the 1000 Genomes cell lines, such as the Geuvadis RNA-Seq data and the high coverage and long read data that the 1000 Genomes Structural Variant group is continuing to generate, in order to present a uniform analysis set. The final aim is to expand the resource to new populations; the IGSR has been funded to support the addition of new populations to the 1000 Genomes dataset and this document describes the principles of that process.

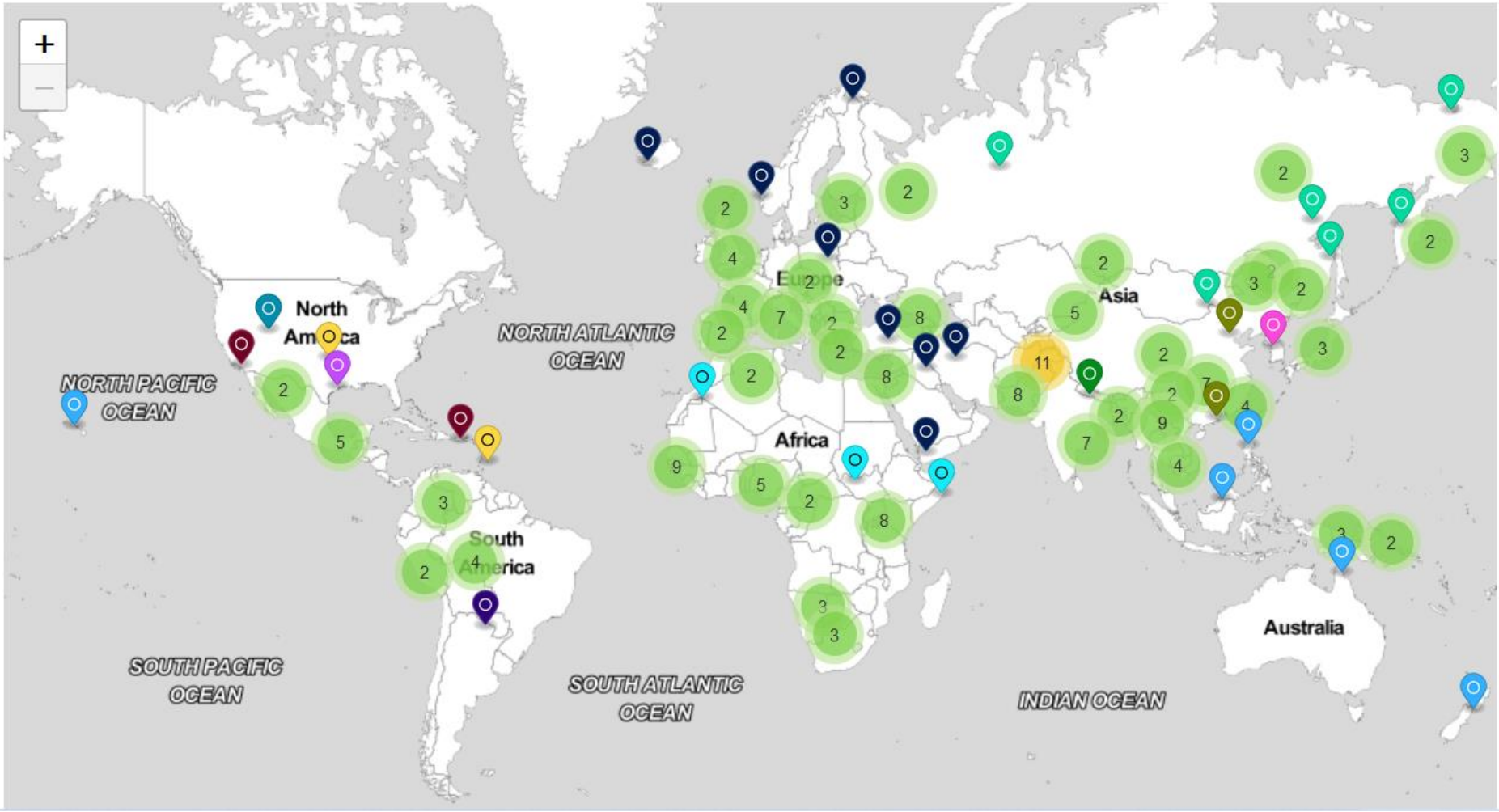
The IGSR recognises that the current 1000 Genomes Project samples do not reflect all populations. An important aim for the IGSR is to expand the populations represented in the collection and to ensure that the public data represents maximum possible population diversity. This will ensure that the 1000 Genomes dataset remains a valuable open resource for the community over the next five years. The IGSR will work with the groups who were unable to contribute samples to the 1000 Genomes Project prior to the completion of sample collection, and will investigate collaborations with other groups to ensure that population diversity gaps are filled.

IGSR[3]

Populations

Map view | Data collection view | Technology view

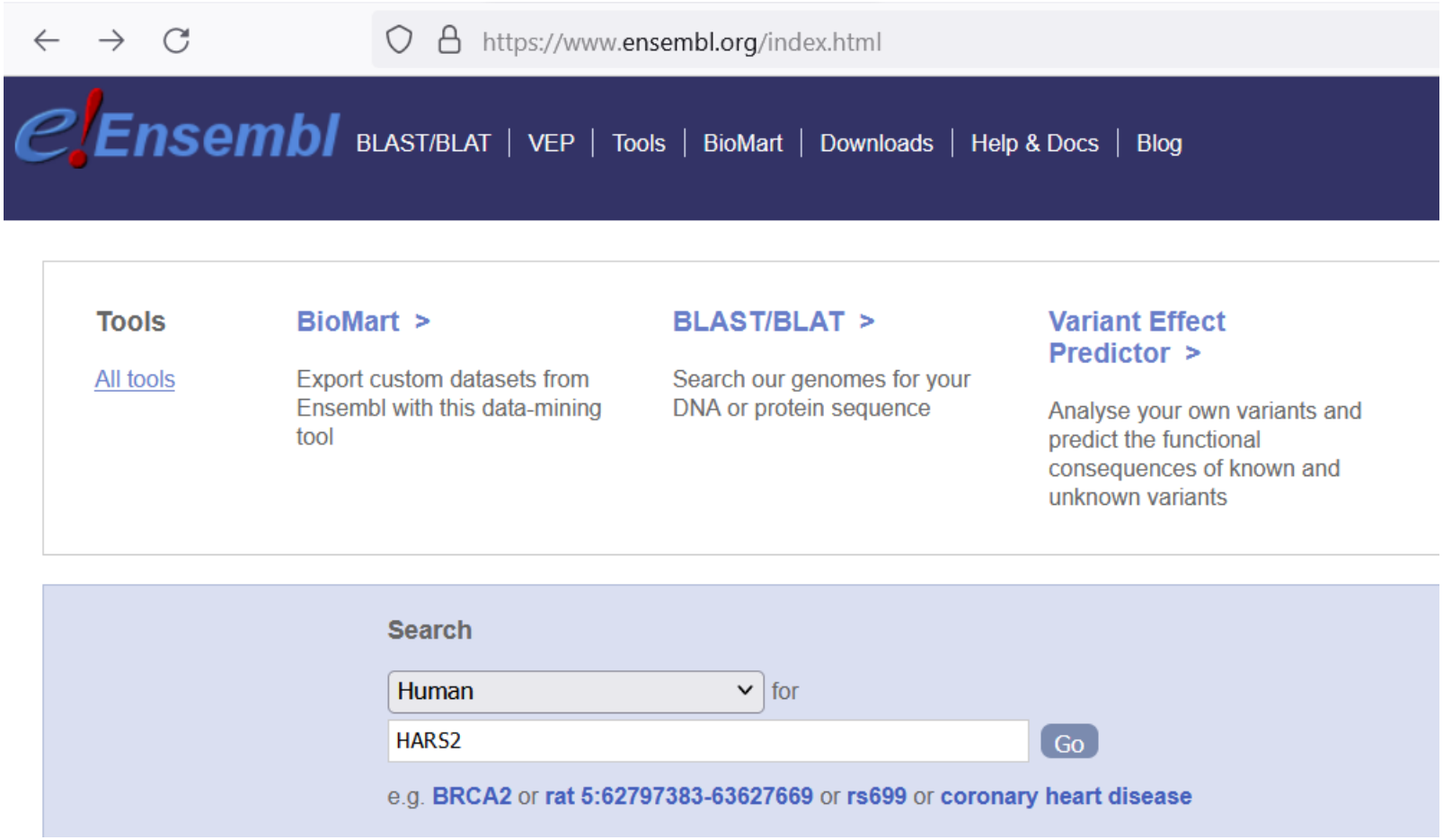
Filter by technology | Filter by data collection | Download the list | Map colour key



https://www.internationalgenome.org/data-portal/population

enSembl[1]

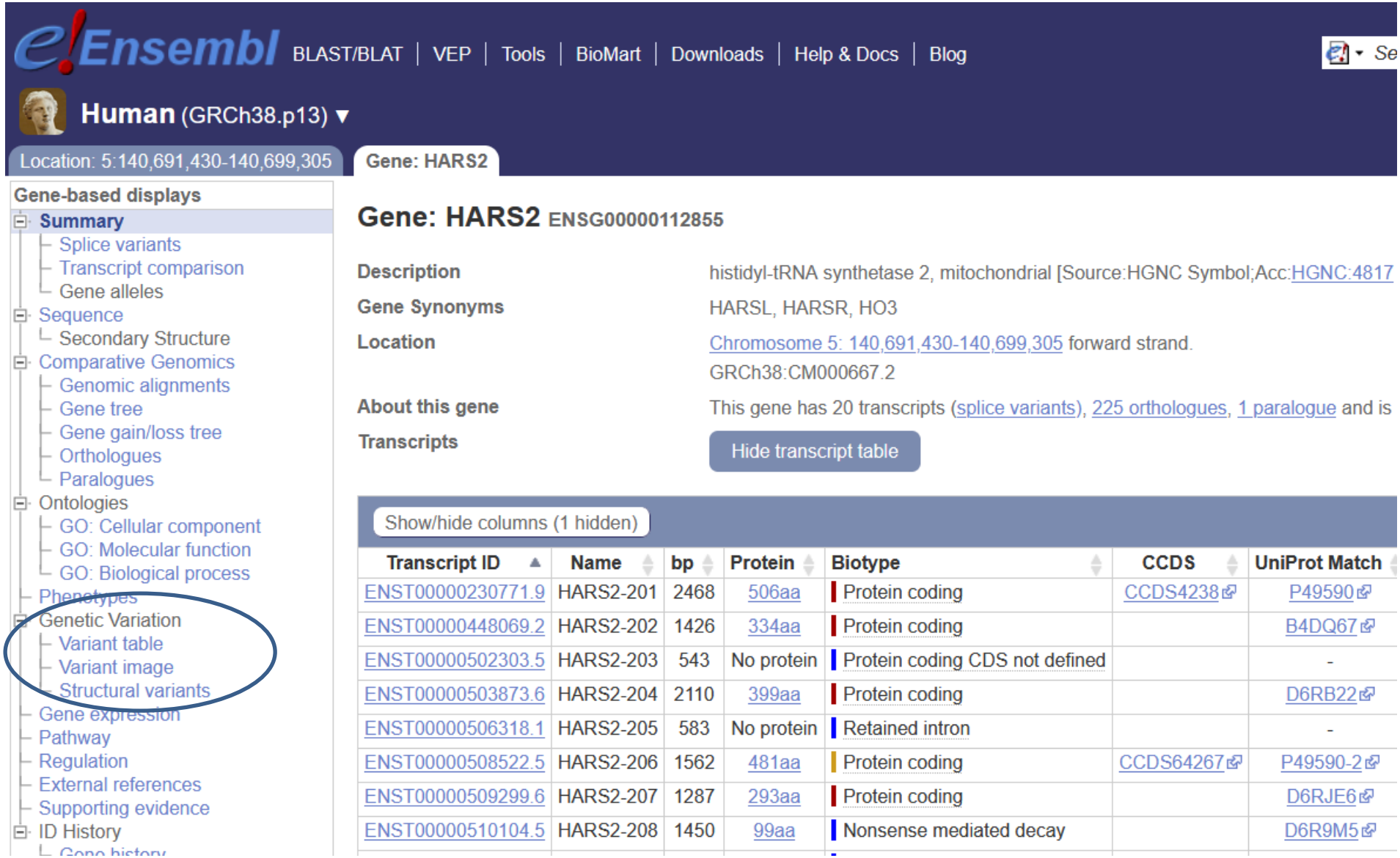
link depuis IGSR



Attention : ici par défaut avec enSembl : build hg38

https://www.ensembl.org

enSembl[2]


e!Ensembl
BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog
Human (GRCh38.p13) ▾
Location: 5:140,691,430-140,699,305
Gene: HARS2
Gene-based displays
Summary
Splice variants
Transcript comparison
Gene alleles
Sequence
Secondary Structure
Comparative Genomics
Genomic alignments
Gene tree
Gene gain/loss tree
Orthologues
Paralogues
Ontologies
GO: Cellular component
GO: Molecular function
GO: Biological process
Phenotypes
Genetic Variation
Variant table
Variant image
Structural variants
Gene expression
Pathway
Regulation
External references
Supporting evidence
ID History
Gene: HARS2 ENSG00000112855
Description
histidyl-tRNA synthetase 2, mitochondrial [Source:HGNC Symbol;Acc:HGNC:4817
Gene Synonyms
HARSL, HARSR, HO3
Location
Chromosome 5: 140,691,430-140,699,305 forward strand.
GRCh38:CM000667.2
About this gene
This gene has 20 transcripts (splice variants), 225 orthologues, 1 paralogue and is
Transcripts
Hide transcript table
Show/hide columns (1 hidden)
Transcript ID | Name | bp | Protein | Biotype | CCDS | UniProt Match
ENST00000230771.9 | HARS2-201 | 2468 | 506aa | Protein coding | CCDS4238 | P49590
ENST00000448069.2 | HARS2-202 | 1426 | 334aa | Protein coding | | B4DQ67
ENST00000502303.5 | HARS2-203 | 543 | No protein | Protein coding CDS not defined | | -
ENST00000503873.6 | HARS2-204 | 2110 | 399aa | Protein coding | | D6RB22
ENST00000506318.1 | HARS2-205 | 583 | No protein | Retained intron | | -
ENST00000508522.5 | HARS2-206 | 1562 | 481aa | Protein coding | CCDS64267 | P49590-2
ENST00000509299.6 | HARS2-207 | 1287 | 293aa | Protein coding | | D6RJE6
ENST00000510104.5 | HARS2-208 | 1450 | 99aa | Nonsense mediated decay | | D6R9M5

enSembl[3]

Variant table

Variant table

This table shows known variants for this gene. Use the 'Consequence Type' filter to view a subset of these.

Filter | Global MAF: All | SIFT: All | PolyPhen: All | Consequences: missense variant, pr...(2/30) | Filter Other Columns

Variant ID	Chr: bp	Alle-les	Glo-bal MAF	Class	Source	Evidence	Clin. Sig.	Conseq. Type	AA	AA coord	SIFT	Poly-Phen	CADD	REVE L	MetaL R	Mutat ion Asse ssor	Transcript
rs1562060357	5:140697279	G/A	-	SNP	dbSNP	-	-	missense variant	S/N	357	0	0.178	23	0.477	0.381	0.985	ENST00000230771.9
rs778446302	5:140697282	T/C	-	SNP	dbSNP	Ex AC gnom AD	-	missense variant	V/A	358	0	0.506	25	0.723	0.646	0.995	ENST00000230771.9
rs767742173	5:140697288	C/T	-	SNP	dbSNP	-	-	missense variant	A/V	360	0	0.52	25	0.465	0.308	0.966	ENST00000230771.9
rs1051862023	5:140697296	C/T	-	SNP	dbSNP	gnom AD	-	missense variant	R/C	363	0	1	32	0.793	0.712	0.996	ENST00000230771.9
rs758141368	5:140697297	G/A	-	SNP	dbSNP	Ex AC gnom AD	-	missense variant	R/H	363	0	0.999	29	0.877	0.729	0.996	ENST00000230771.9
rs746757469	5:140697300	A/G	-	SNP	dbSNP	Ex AC gnom AD	-	missense variant	Y/C	364	0	0.999	27	0.897	0.802	0.999	ENST00000230771.9
rs376177973	5:140697311	G/**A**/T	-	SNP	dbSNP	Ex AC gnom AD	⚠	missense variant	V/M	368	0	0.927	26	0.548	0.563	0.915	ENST00000230771.9
rs376177973	5:140697311	G/A/**T**	-	SNP	dbSNP	Ex AC gnom AD	⚠	missense variant	V/L	368	0	0.893	25	0.559	0.46	0.764	ENST00000230771.9
rs61736946	5:140697314	G/C	0.001 (C)	SNP	dbSNP	Ex AC gnom AD	✚	missense variant	G/R	369	0.01	0.935	24	0.464	0.449	0.9	ENST00000230771.9
rs373145883	5:140697317	A/G	-	SNP	dbSNP	Ex AC	-	missense variant	M/V	370	0.01	0.005	20	0.117	0.11	0.556	ENST00000230771.9

enSembl[4]

Human (GRCh38.p13) ▾

Location: 5:140,691,430-140,699,305 | Gene: HARS2 | Variant: rs376177973

Variant displays

- Explore this variant
- Genomic context
  - Genes and regulation
  - Flanking sequence
- Population genetics
- Phenotype data
- Sample genotypes
- Linkage disequilibrium
- Phylogenetic context
- Citations
- 3D Protein model

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

rs376177973 SNP

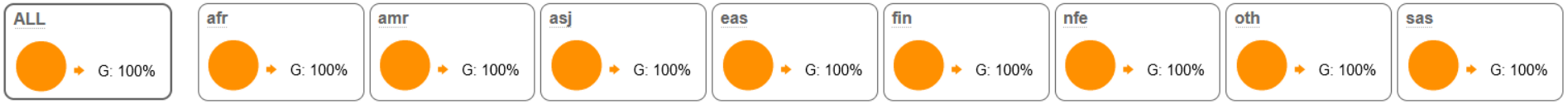
Most severe consequence	**missense variant** | See all predicted consequences
Alleles	G/A/T | Ancestral: **G** | Highest population MAF: **< 0.01**
Change tolerance	CADD: A:26.3, T:25.6 | GERP: 2.82
Location	Chromosome **5:140697311** (forward strand) | VCF: `5 140697311 rs376177973 G A,T`
Co-located variant	**HGMD-PUBLIC** CM112637
Evidence status	
Clinical significance	
HGVS names	This variant has **84** HGVS names - Show
Synonyms	This variant has **4** synonyms - Show
Original source	Variants (including SNPs and indels) imported from dbSNP (release 154) | View in dbSNP
About this variant	This variant overlaps 14 transcripts, is associated with 1 phenotype and is mentioned in 1 citation.

Explore this variant

enSembl[5]

Population genetics

gnomAD exomes r2.1.1 allele frequencies



Jump to: gnomAD exomes r2.1.1 (9) | gnomAD genomes v3.1.2 (11) | NCBI ALFA (12) | TOPMed (1) | UK10K (2) | NHLBI Exome Sequencing Project (2)

gnomAD exomes r2.1.1 (9)

Show/hide columns

Filter

Population	Allele: frequency (count)		
gnomADe:**ALL**	G: 1.000 (251454)	A: 2.386e-05 (6)	T: 7.953e-06 (2)
gnomADe:**afr**	G: 1.000 (16256)	A: 0.000	T: 0.000
gnomADe:**amr**	G: 1.000 (34590)	A: 0.000	T: 0.000
gnomADe:**asj**	G: 1.000 (10080)	A: 0.000	T: 0.000
gnomADe:**eas**	G: 1.000 (18394)	A: 0.000	T: 0.000
gnomADe:**fin**	G: 1.000 (21633)	A: 4.622e-05 (1)	T: 0.000
gnomADe:**nfe**	G: 0.9999 (113747)	A: 4.395e-05 (5)	T: 1.758e-05 (2)
gnomADe:**oth**	G: 1.000 (6138)	A: 0.000	T: 0.000
gnomADe:**sas**	G: 1.000 (30616)	A: 0.000	T: 0.000

enSembl[6]

Flanking sequence

Download sequence | BLAST this sequence

Flanking sequence

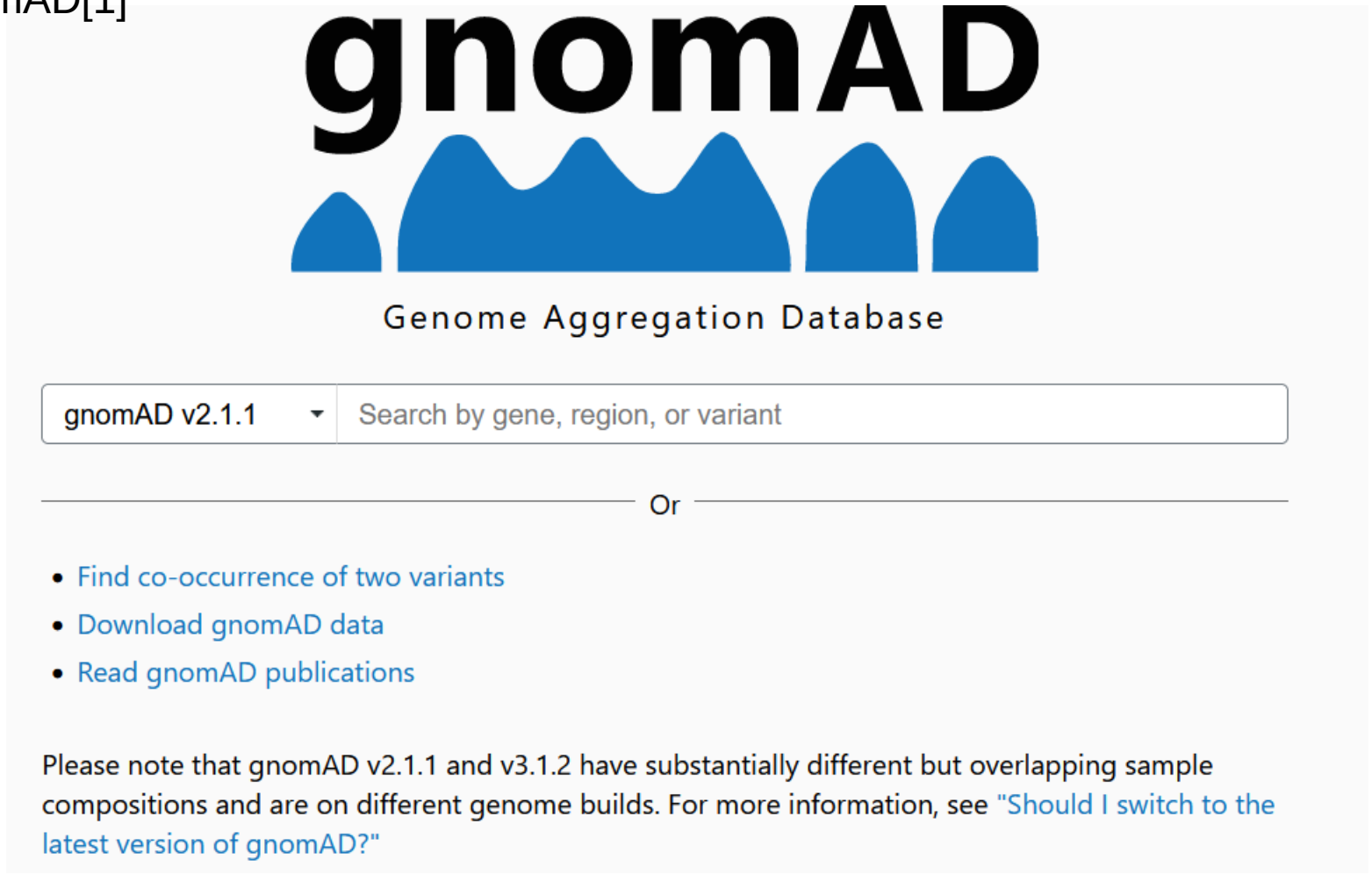
The sequence below is from the **reference genome** flanking the variant location. The variant is shown in **red** text. Neighbouring variar
To change the display of the flanking sequence (e.g. hide the other variants, change the length of the flanking sequence), use the "Cor

Variants: Coding sequence | Focus variant | Frameshift | Inframe deletion | Intronic | Missense | Splice region | Stop gained | Synonymous

Markup: loaded

D : non C = A,T,G

gnomAD[1]



Remplace la base ExAC
Même structure que ExAC mais avec une plus forte intégration de données
Une des meilleures références actuelles
Coordonnées en hg19

http://gnomad.broadinstitute.org/

gnomAD[2]

HARS2 histidyl-tRNA synthetase 2, mitochondrial

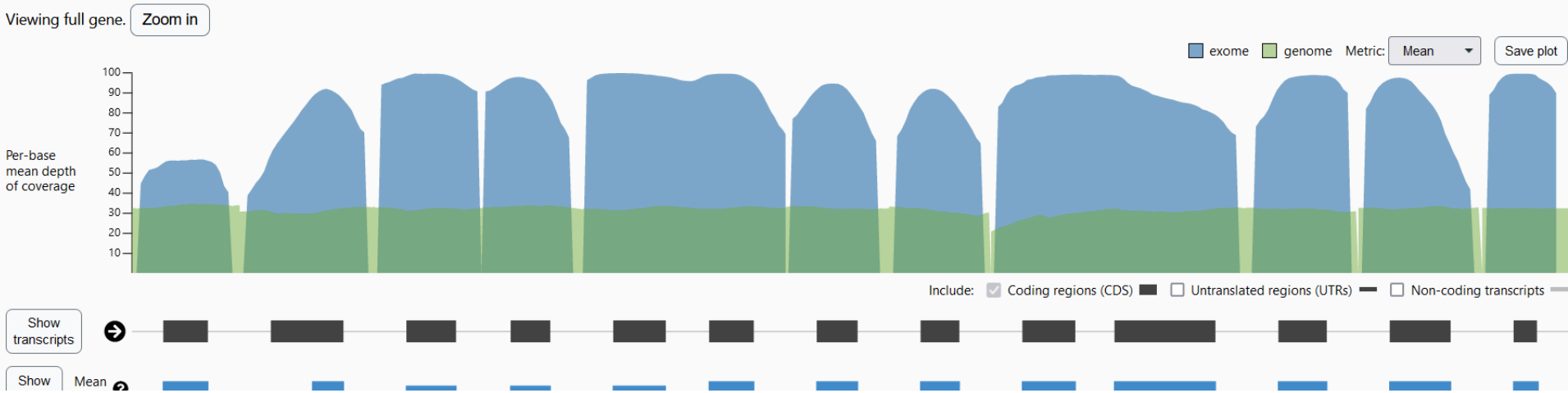
Dataset gnomAD v2.1.1 | gnomAD SVs v2.1

Genome build GRCh37 / hg19
Ensembl gene ID ENSG00000112855.10
Ensembl canonical transcript ENST00000230771.3
Other transcripts ENST00000513688.1, ENST00000510104.1, and 12 more
Region 5:140071011-140078889
External resources Ensembl, UCSC Browser, and more

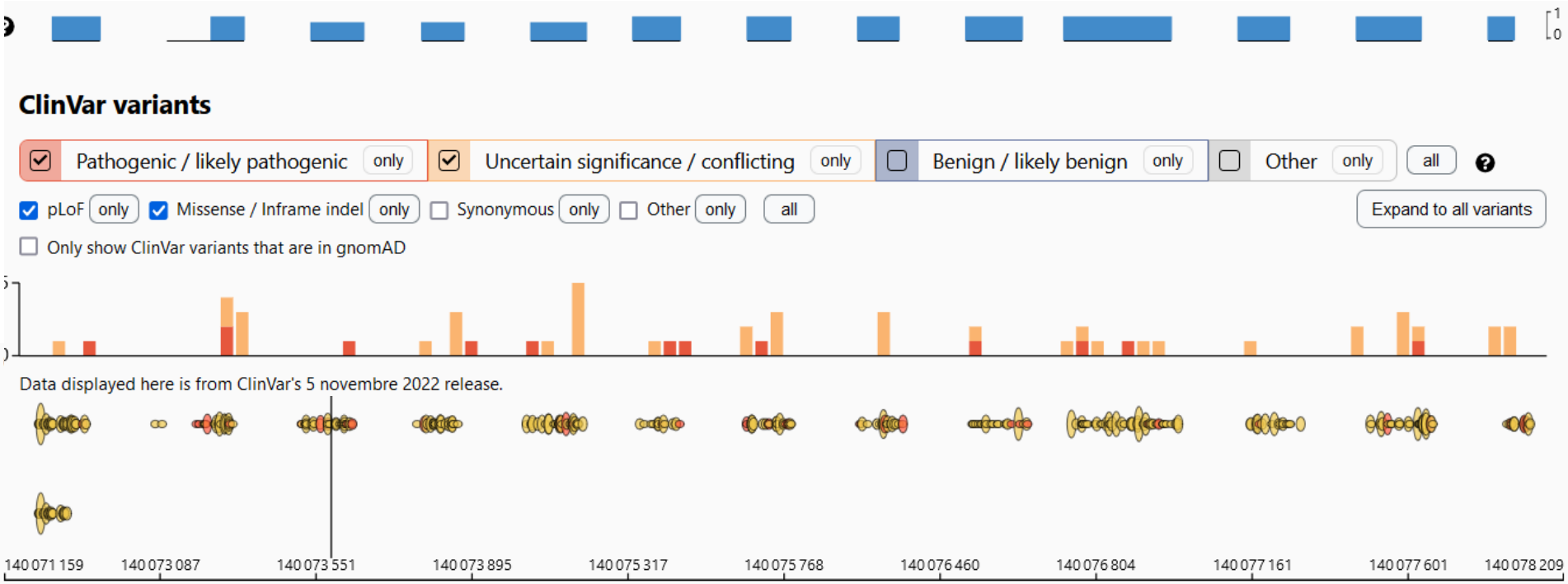
Constraint

Category	Expected SNVs	Observed SNVs	Constraint metrics
Synonymous	98.1	103	Z = -0.39 o/e = 1.05 (0.89 - 1.24)
Missense	267	265	Z = 0.04 o/e = 0.99 (0.9 - 1.1)
pLoF	26.4	12	pLI = 0 o/e = 0.45 (0.29 - 0.74)

Constraint metrics based on Ensembl canonical transcript (ENST00000230771.3).



gnomAD[3]



ClinVar variants

Pathogenic / likely pathogenic only | Uncertain significance / conflicting only | Benign / likely benign only | Other only | all

pLoF only | Missense / Inframe indel only | Synonymous only | Other only | all

Expand to all variants

Only show ClinVar variants that are in gnomAD

Data displayed here is from ClinVar's 5 novembre 2022 release.

140 071 159 | 140 073 087 | 140 073 551 | 140 073 895 | 140 075 317 | 140 075 768 | 140 076 460 | 140 076 804 | 140 077 161 | 140 077 601 | 140 078 209

pLoF only | Missense / Inframe indel only | Synonymous only | Other only | all

Exomes | Genomes | SNVs | Indels | Filtered variants

Search variant table

Export variants to CSV | Configure table

Note Only variants located in or within 75 base pairs of a coding exon are shown here. To see variants in UTRs or introns, use the region view.

The table below shows the HGVS consequence and VEP annotation for each variant's most severe consequence across all transcripts in this gene. Cases where the most severe consequence occurs in a non-canonical transcript are denoted with †. To see consequences in a specific transcript, use the transcript view.

Variant ID	Source	HGVS Consequence	VEP Annotation	LoF Curation	Clinical Significance	Flags	Allele Count
5-140075395-C-G	G	p.Leu200Val	missense		Pathogenic/Likely p...		1
5-140076896-G-T	E	p.Val368Leu	missense		Pathogenic/Likely p...		2

gnomAD[4]

HARS2 | gnomAD v2.1.1 | gnomAD × | 5-140076926-G-T | gnomAD v2.1.1 × +

Details et frequences en populations

https://gnomad.broadinstitute.org/variant/5-140076926-G-T?dataset=gnomad_r2_1

	Exomes	Genomes	Total
Filters	Pass	No variant	
Allele Count	1		1
Allele Number	251434		251434
Allele Frequency	0.000003977		0.000003977
Popmax Filtering AF (95% confidence)	—		
Number of homozygotes	0		0
Mean depth of coverage	85.0	31.2	

External Resources

- dbSNP (rs767814719)
- UCSC
- ClinGen Allele Registry (CA3444621)

Feedback

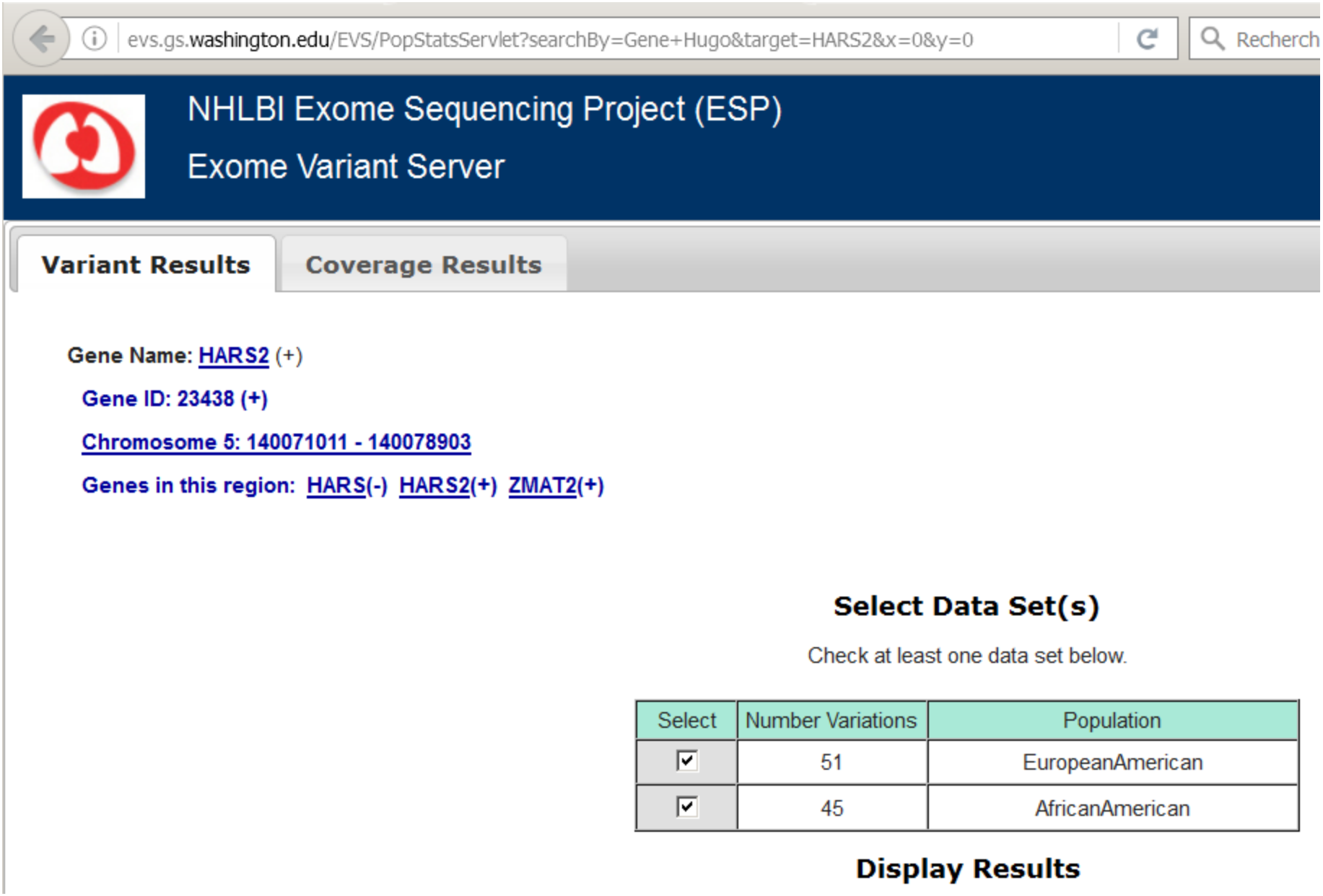
Report an issue with this variant

Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
East Asian	1	18394	0	0.00005437
African/African American	0	16254	0	0.000
Latino/Admixed American	0	34588	0	0.000
Ashkenazi Jewish	0	10078	0	0.000
European (Finnish)	0	21612	0	0.000
European (non-Finnish)	0	113754	0	0.000
Other	0	6138	0	0.000
South Asian	0	30616	0	0.000
XX	0	115538	0	0.000
XY	1	135896	0	0.000007359

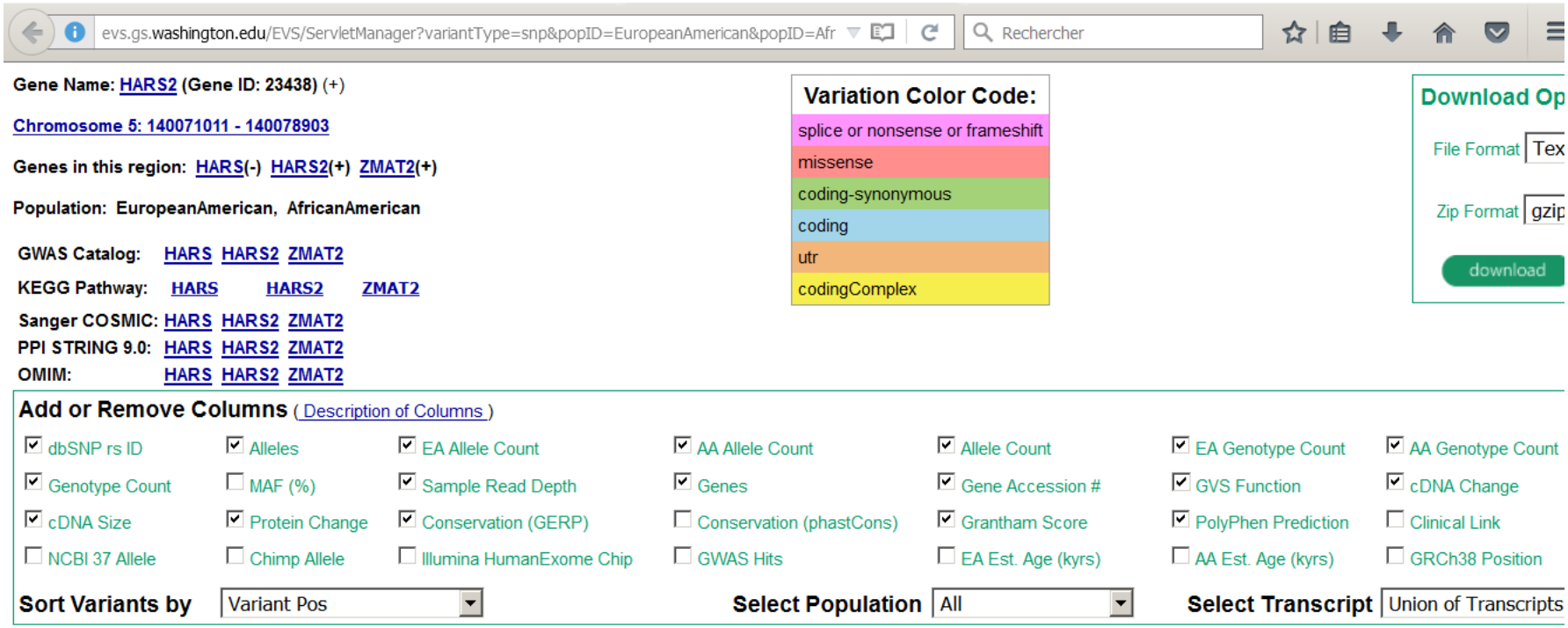
EVS[1]

http://evs.washington.edu/EVS



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

EVS[2]



Data en hg19 (GrCh37) Leu200Val variant : rs397515410 (chr5:140075395)
Data en hg19 (GrCh37) Val368Leu variant : rs376177693 (chr5:140076928)

Possibilité de sélectionner les colonnes :Voir le help
http://evs.gs.washington.edu/EVS/HelpDescriptions.jsp#EAGenotypeCount

EWS[3]

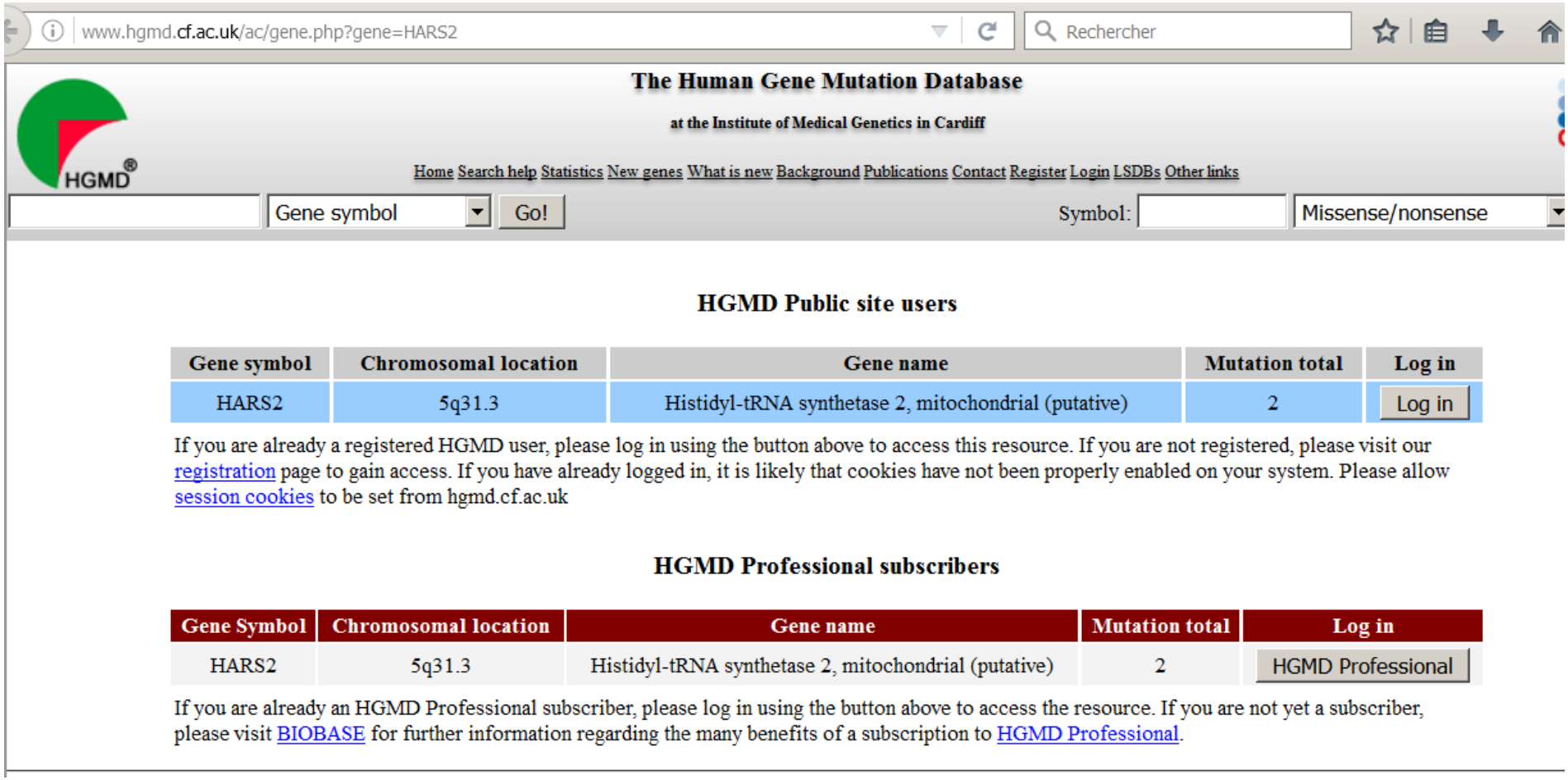
Variant GRCh37 Pos	rs ID	Alleles	EA Allele #	AA Allele #	All Allele #	EA Genotype #	AA Genotype #	All Genotype #	Avg. Sample Read Depth	Genes	mRNA Accession #	GVS Function	cDNA Change
5:140076850	rs375264978	C>G	G=1/C=8599	G=0/C=4406	G=1/C=13005	GG=0/GC=1 /CC=4299	GG=0/GC=0 /CC=2203	GG=0/GC=1 /CC=6502	65	HARS2	NM_012208.2	coding-synonymous	c.1056C>G
5:140076861	rs369536729	G>A	A=1/G=8599	A=0/G=4406	A=1/G=13005	AA=0/AG=1 /GG=4299	AA=0/AG=0 /GG=2203	AA=0/AG=1 /GG=6502	64	HARS2	NM_012208.2	missense	c.1067G>A
5:140076896	rs376177973	G>T	T=1/G=8599	T=0/G=4406	T=1/G=13005	TT=0/TG=1 /GG=4299	TT=0/TG=0 /GG=2203	TT=0/TG=1 /GG=6502	70	HARS2	NM_012208.2	missense	c.1102G>T
5:140076899	rs61736946	G>C	C=0/G=8600	C=41/G=4365	C=41/G=12965	CC=0/CG=0 /GG=4300	CC=0/CG=41 /GG=2162	CC=0/CG=41 /GG=6462	72	HARS2	NM_012208.2	missense	c.1105G>C
5:140076902	rs373145883	A>G	G=0/A=8600	G=1/A=4405	G=1/A=13005	GG=0/GA=0 /AA=4300	GG=0/GA=1 /AA=2202	GG=0/GA=1 /AA=6502	72	HARS2	NM_012208.2	missense	c.1108A>G
5:140076921	unknown	R>A1	A1=0/R=8254	A1=1/R=4263	A1=1/R=12517	A1A1=0/A1R=0 /RR=4127	A1A1=0/A1R=1 /RR=2131	A1A1=0/A1R=1 /RR=6258	72	HARS2	NM_012208.2	frameshift	c.1128del1
5:140076970	rs376141266	C>T	T=1/C=8599	T=0/C=4406	T=1/C=13005	TT=0/TC=1 /CC=4299	TT=0/TC=0 /CC=2203	TT=0/TC=1 /CC=6502	58	HARS2	NM_012208.2	coding-synonymous	c.1176C>T

cDNA Change	cDNA Size	Protein Change	Conservation (GERP)	Grantham Score	PolyPhen2 (Class:Score)
c.1056C>G	1521	p.(P352=)	-2.72	NA	unknown
c.1067G>A	1521	p.(G356D)	5.67	94	probably-damaging:1.0
c.1102G>T	1521	p.(V368L)	5.67	32	probably-damaging:0.999
c.1105G>C	1521	p.(G369R)	5.67	125	probably-damaging:1.0
c.1108A>G	1521	p.(M370V)	1.8	21	benign:0.08
c.1128del1	1521	p.(H376Qfs*23)	2.97	unknown	unknown
c.1176C>T	1521	p.(Y392=)	1.91	NA	unknown

Classe de PolyPhen2

Unknown
Begnign
Probably damaging
damaging

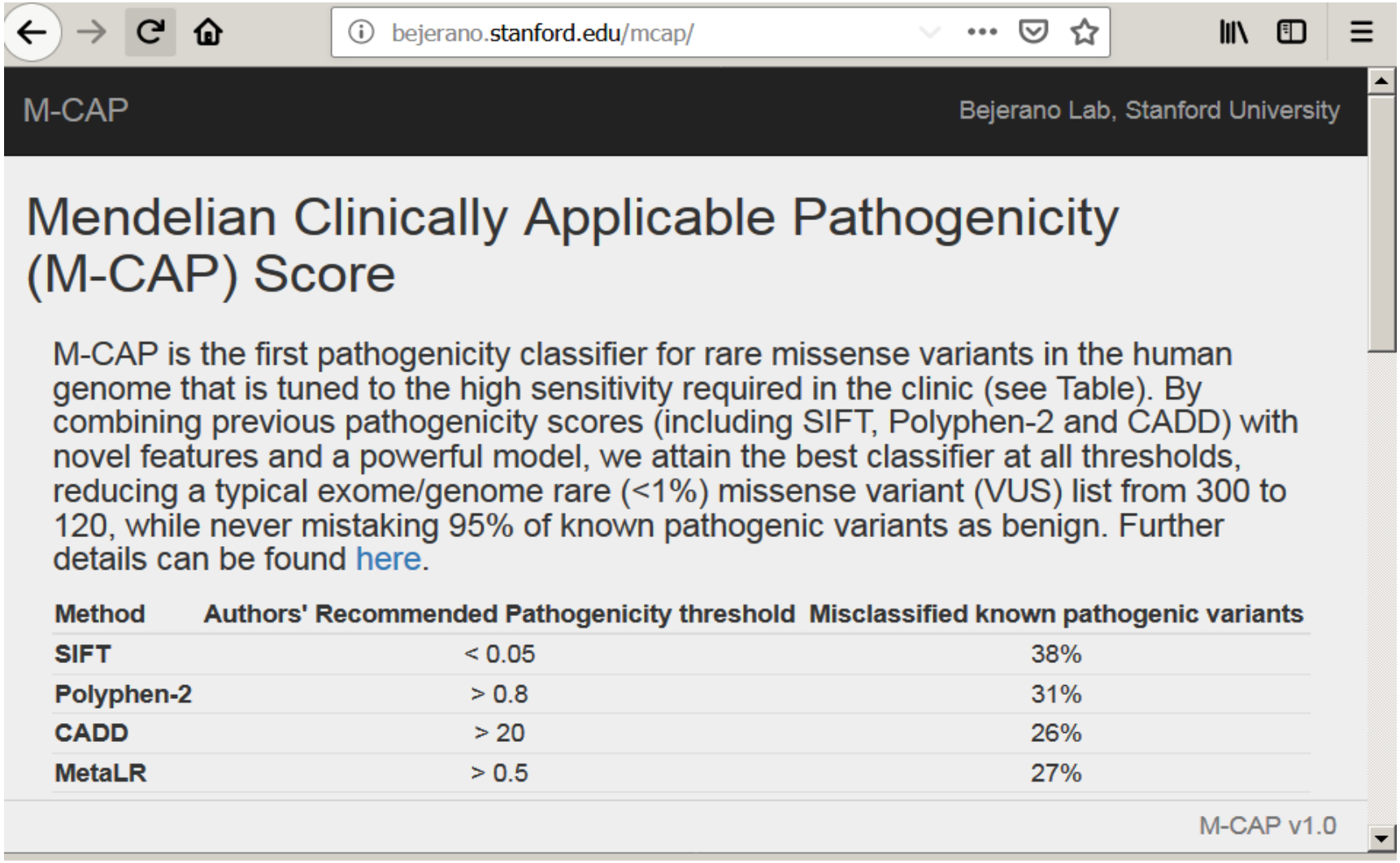
HGMD[1]



Commerciale

http://hgmd.cf.uk/

M-CAP [1]



http://bejerano.stanford.edu/mcap/

M-CAP [2]

Score a variant

Enter the GRCh37/hg19 coordinate for a missense variant to retrieve its M-CAP score.

GRCh37/hg19 | 5:140076896 | Go! | Demo!

GRCh37/hg19.5:140,076,896 Reference Allele G

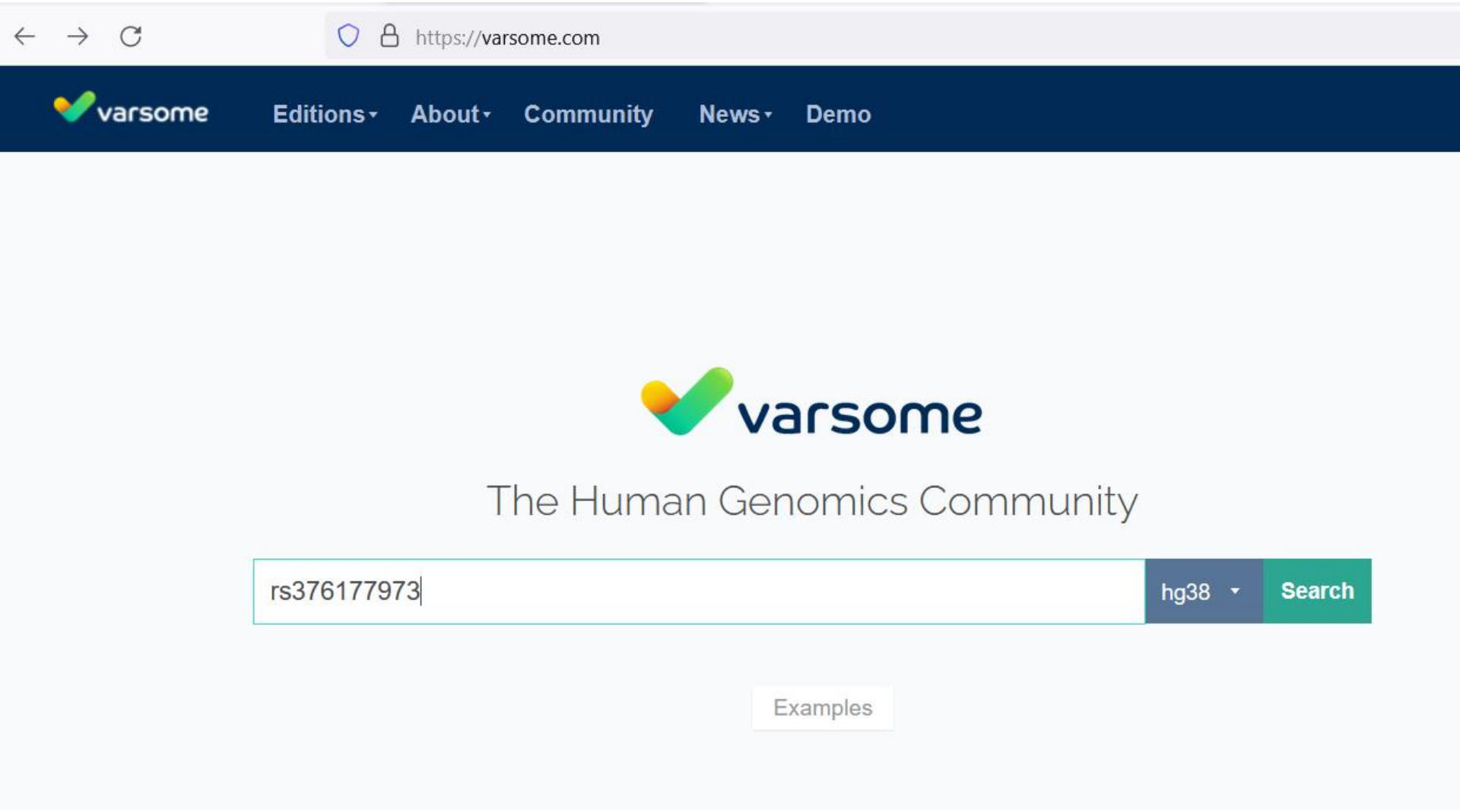
Alt Allele	M-CAP	95% sensitivity
A	0.069	Possibly Pathogenic
C	0.074	Possibly Pathogenic
T	0.075	Possibly Pathogenic

How to cite

Jagadeesh, K., Wenger, A., Berger, M., Guturu, H., Stenson, P., Cooper, D., Bernstein, J., and Bejerano, G. (2016). M-CAP eliminates a majority of variants with uncertain significance in clinical exomes at high sensitivity. Nature Genetics, 2016. 48 (12)

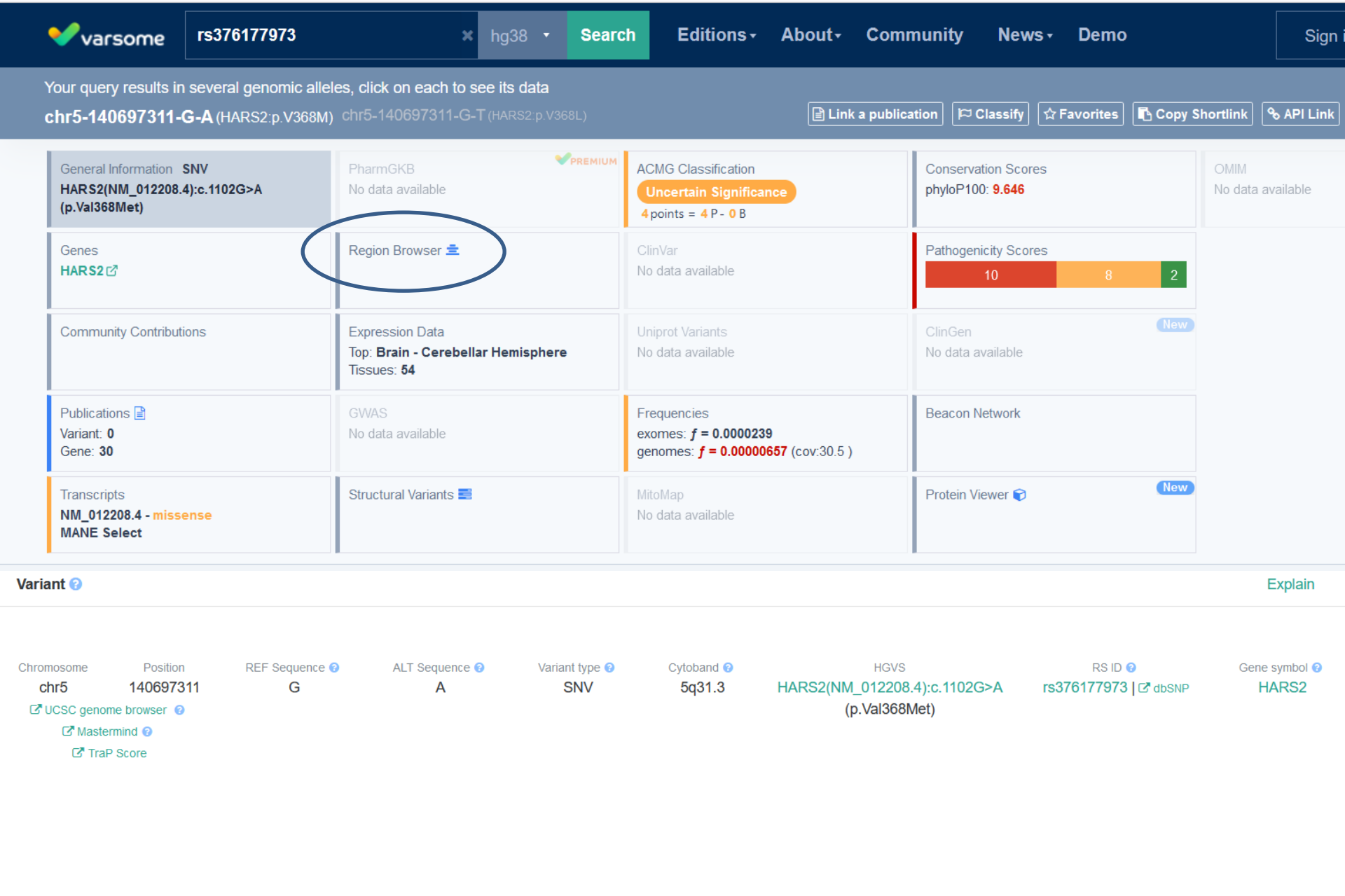
Val368Leu	hg19	5:140076896
Leu200Val	hg19	5:140075395

Varsome [1]



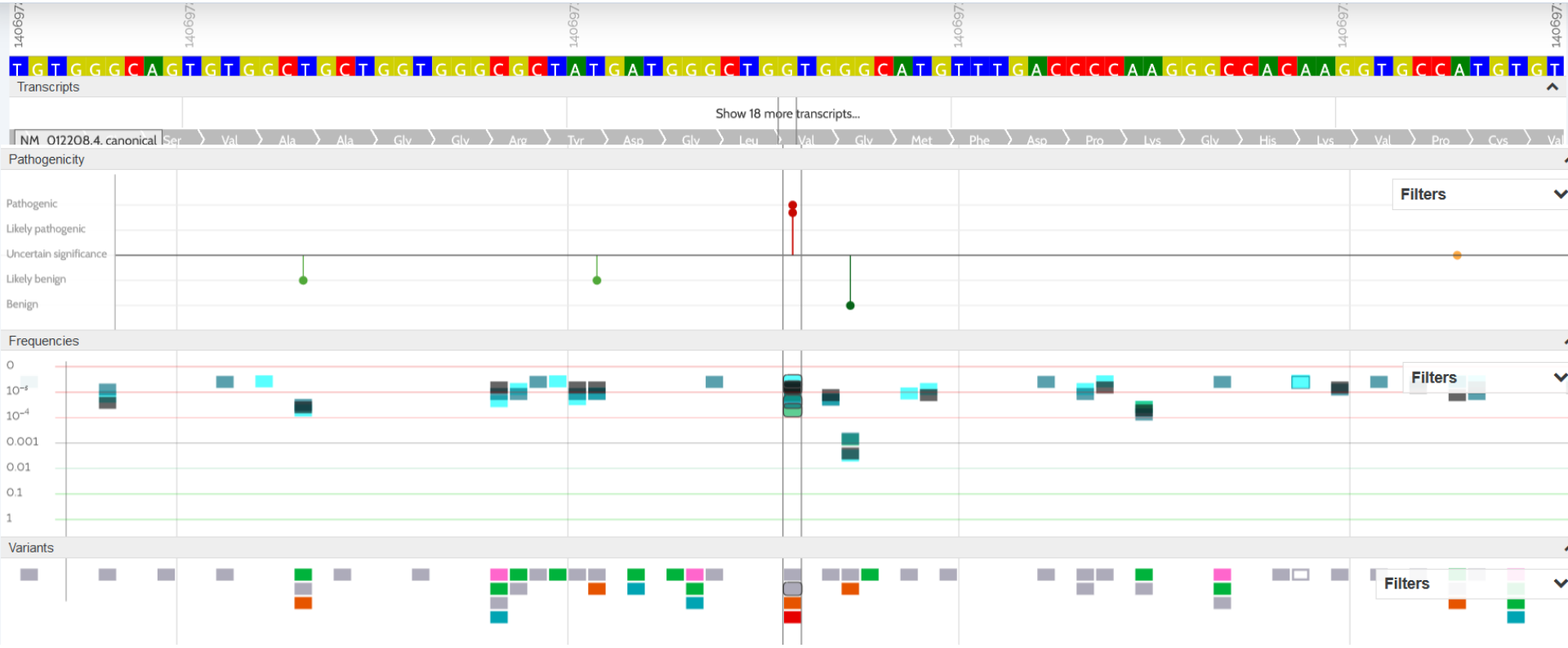
https://varsome.com/

Varsome [2]


varsome
rs376177973
hg38
Search
Editions
About
Community
News
Demo
Sign
Your query results in several genomic alleles, click on each to see its data
chr5-140697311-G-A (HARS2:p.V368M)
chr5-140697311-G-T (HARS2:p.V368L)
Link a publication
Classify
Favorites
Copy Shortlink
API Link
General Information SNV
HARS2(NM_012208.4):c.1102G>A
(p.Val368Met)
PharmGKB
No data available
PREMIUM
ACMG Classification
Uncertain Significance
4 points = 4 P - 0 B
Conservation Scores
phyloP100: 9.646
OMIM
No data available
Genes
HARS2
Region Browser
ClinVar
No data available
Pathogenicity Scores
10
8
2
Community Contributions
Expression Data
Top: Brain - Cerebellar Hemisphere
Tissues: 54
Uniprot Variants
No data available
New
ClinGen
No data available
Publications
Variant: 0
Gene: 30
GWAS
No data available
Frequencies
exomes: f = 0.0000239
genomes: f = 0.00000657 (cov:30.5)
Beacon Network
Transcripts
NM_012208.4 - missense
MANE Select
Structural Variants
MitoMap
No data available
New
Protein Viewer
Variant
Explain
Chromosome
chr5
UCSC genome browser
Mastermind
TraP Score
Position
140697311
REF Sequence
G
ALT Sequence
A
Variant type
SNV
Cytoband
5q31.3
HGVS
HARS2(NM_012208.4):c.1102G>A
(p.Val368Met)
RS ID
rs376177973 | dbSNP
Gene symbol
HARS2

Varsome [3]

Region browser



Liens fréquences et classes sur les bases équivalentes

Varsome [4] Population frequencies

Population frequencies

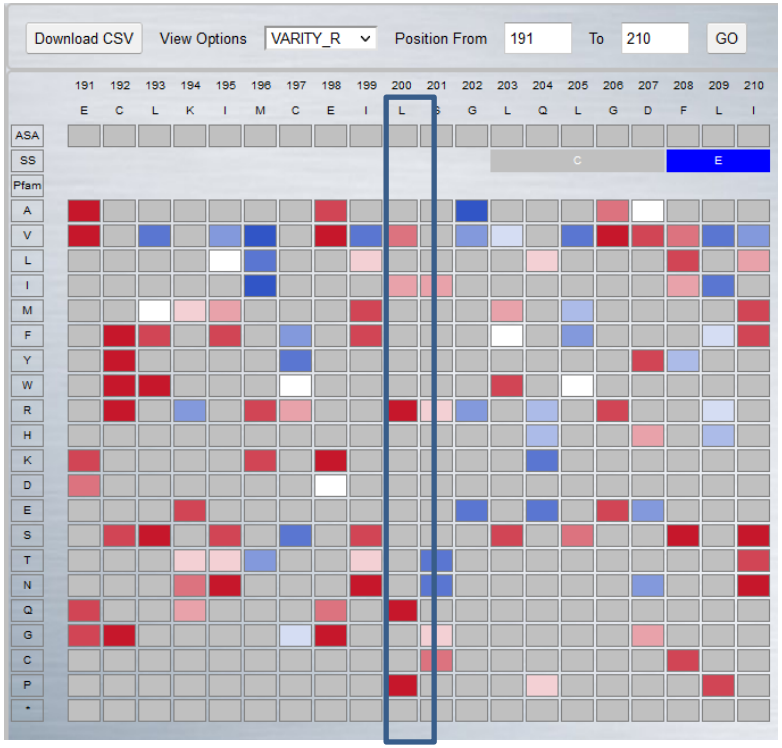
Population	Allele Count	Allele Number	Homozygotes	Allele Frequency
African	-	16 256	-	-
Ashkenazi Jewish	-	10 080	-	-
East Asian	-	18 394	-	-
European (Finnish)	1	21 634	-	0.0000462
European (Non-Finnish)	5	113 754	-	0.0000439
Latino	-	34 590	-	-
South Asian	-	30 616	-	-
Other	-	6 138	-	-
Total	**6**	**251 462**	-	0.0000239
Male	3	135 906	-	0.0000221
Female	3	115 556	-	0.0000259

Varity [1]

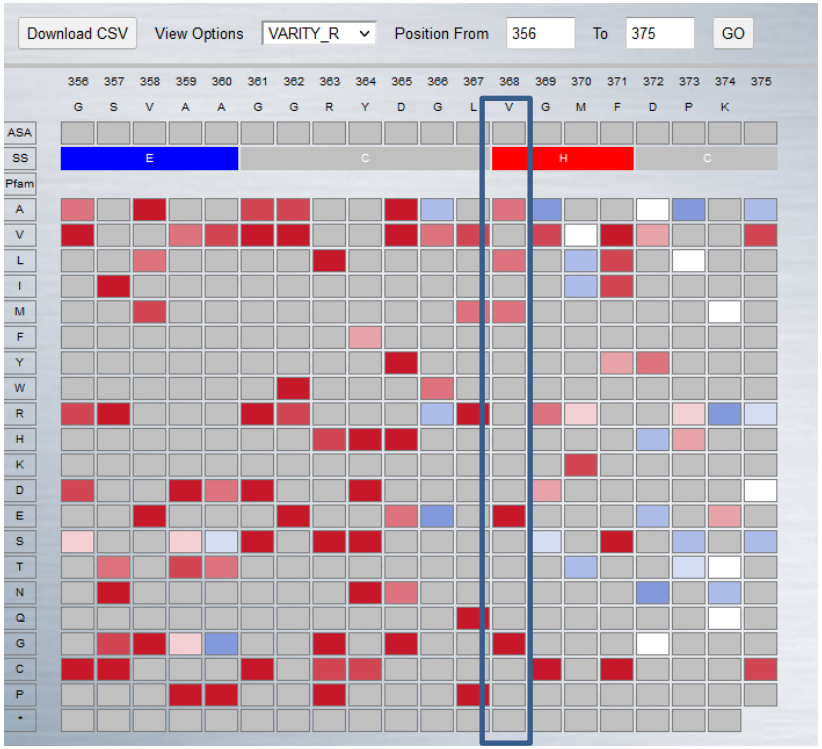
Varity

http://varity.varianteffect.org/

Varity [2]



Leu200Val



Val368Leu

Un clic sur une position donne un tableau complet

Varity [3]

Un clic sur une position donne un tableau complet

Position: 368 [V ⇨ L]

VARITY_R: 0.818 VARITY_R_LOO: 0.796 -log10(MAF): 5.099

VARITY_ER: 0.747 VARITY_ER_LOO: 0.780 ClinVar: [nan, 'Likely pathogenic']

Feature	Value	Contribution[R]	Feature	Value	Contribution[R]	Feature	Value	Contribution[R]	Feature	Value	Contribution[R]
PROVEAN	-2.750	5.71%	**β Sheet**	0.000	-0.27%	**Δ PKb**	0.020	4.65%	**Δ Hydrophobic**	0.000	-0.07%
SIFT	0.001	37.12%	**α Helix**	1.000	2.79%	**Δ PKb**	-0.020	-1.39%	**Δ Polar**	0.000	-0.00%
EVMutation	-5.738	51.26%	**Coiled coil**	0.000	3.92%	**Δ Isoelectronic Point**	0.400	2.79%	**Δ Ionizable**	0.000	-0.00%
LRT	0.000	10.36%	**Buried Surface Area**		-2.32%	**Δ percentage buried residues**	15.000	6.37%	**Δ Aromatic**	0.000	1.20%
GERP++	5.670	-4.05%	**# of H bond**		-1.20%	**Δ Average volumn of buried residues**	-26.000	1.73%	**Δ H-bond**	0.000	0.27%
PholyP	1.166	2.26%	**# of Salt Bridge**		0.00%	**Δ Van der Waals volume**	-19.000	-0.00%	**Δ Sulfur Containing**	0.000	0.46%
PhastCons	0.940	-3.98%	**# of Disulfide Bond**		0.00%	**Δ Accessible surface area of side chain**	-20.000	-1.99%	**Δ Essential to Human**	0.000	0.20%
SiPhy	19.785	10.89%	**# of Covelent Bond**		0.00%	**Δ Cyclic**	0.000	-2.46%	**Δ Size**	0.000	0.53%
BLOSUM100	0.036	3.78%	**Solvation Energy**		-4.52%	**Δ Charged**	0.000	-0.07%			
IN/OUT pFam Domain	1.000	3.72%	**Δ Molecular Weight**	-14.030	-0.00%	**Δ Postive Charged**	0.000	-0.00%			
Accessible Surface Area		-10.49%	**Δ PKa**	-0.040	0.60%	**Δ Negative Charged**	0.000	-0.00%			

VARITY_ER Feature Contribution VARITY_R Feature Contribution CLOSE

ACMG [1]

© American College of Medical Genetics and Genomics **ACMG STANDARDS AND GUIDELINES** | Genetics inMedicine

Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Sue Richards, PhD[1], Nazneen Aziz, PhD[2,16], Sherri Bale, PhD[3], David Bick, MD[4], Soma Das, PhD[5], Julie Gastier-Foster, PhD[6,7,8], Wayne W. Grody, MD, PhD[9,10,11], Madhuri Hegde, PhD[12], Elaine Lyon, PhD[13], Elaine Spector, PhD[14], Karl Voelkerding, MD[13] and Heidi L. Rehm, PhD[15]; on behalf of the ACMG Laboratory Quality Assurance Committee

Official Journal of the American College of Medical Genetics and Genomics **ORIGINAL RESEARCH ARTICLE** | Genetics inMedicine

Open

Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria

Keith Nykamp, PhD[1], Michael Anderson, PhD[1], Martin Powers, MD[1], John Garcia, PhD[1], Blanca Herrera, PhD[1], Yuan-Yuan Ho, PhD[1], Yuya Kobayashi, PhD[1], Nila Patil, PhD[1], Janita Thusberg, PhD[1], Marjorie Westbrook, PhD[1], The Invitae Clinical Genomics Group[2] and Scott Topper, PhD, FACMG[1]

ACMG [2]

Table 1 Population, disease-specific, and sequence databases

Population databases	
Exome Aggregation Consortium http://exac.broadinstitute.org/	Database of variants found during exome sequencing of 61,486 unrelated individuals sequenced as part of various disease-specific and population genetic studies. Pediatric disease subjects as well as related individuals were excluded.
Exome Variant Server http://evs.gs.washington.edu/EVS	Database of variants found during exome sequencing of several large cohorts of individuals of European and African American ancestry. Includes coverage data to inform the absence of variation.
1000 Genomes Project http://browser.1000genomes.org	Database of variants found during low-coverage and high-coverage genomic and targeted sequencing from 26 populations. Provides more diversity compared to the Exome Variant Server but also contains lower-quality data, and some cohorts contain related individuals.
dbSNP http://www.ncbi.nlm.nih.gov/snp	Database of short genetic variations (typically ≤50 bp) submitted from many sources. May lack details of the originating study and may contain pathogenic variants.
dbVar http://www.ncbi.nlm.nih.gov/dbvar	Database of structural variation (typically >50 bp) submitted from many sources.
Disease databases	
ClinVar http://www.ncbi.nlm.nih.gov/clinvar	Database of assertions about the clinical significance and phenotype relationship of human variations.
OMIM http://www.omim.org	Database of human genes and genetic conditions that also contains a representative sampling of disease-associated genetic variants.
Human Gene Mutation Database http://www.hgmd.org	Database of variant annotations published in the literature. Requires fee-based subscription to access much of the content.
Locus/disease/ethnic/other-specific databases	
Human Genome Variation Society http://www.hgvs.org/dblist/dblist.html Leiden Open Variation Database http://www.lovd.nl	The Human Genome Variation Society site developed a list of thousands of databases that provide variant annotations on specific subsets of human variation. A large percentage of databases are built in the Leiden Open Variation Database system.
DECIPHER http://decipher.sanger.ac.uk	A molecular cytogenetic database for clinicians and researchers linking genomic microarray data with phenotype using the Ensembl genome browser.
Sequence databases	
NCBI Genome http://www.ncbi.nlm.nih.gov/genome	Source of full human genome reference sequences.
RefSeqGene http://www.ncbi.nlm.nih.gov/refseq/rsg Locus Reference Genomic (LRG) http://www.lrg-sequence.org	Medically relevant gene reference sequence resource.
MitoMap http://www.mitomap.org/MITOMAP/HumanMitoSeq	Revised Cambridge reference sequence for human mitochondrial DNA.

ACMG [3]

Variant classification according to the ACMG Guidelines

The results from the variant classification in different Norwegian laboratories.

Mari Ann Kulseth
AMG - OUS

Variant classification

Class 5: pathogenic
Class 4: likely pathogenic
Class 3: VUS – variant of uncertain significance
Class 2: likely benign
Class 1: benign

http://www.nshg.no/wp-content/uploads/2017/11/2017-Workshop-varianttolkning.pdf

NGS-Diag [1]

NGS-Diag

Homogénéisation de l'interprétation de variants de séquence générés par les analyses en NGS		
Date de création : 20/12/2017	Date de révision :	Version : 1
Date de 1ère Application : 01/06/2018	N° document : NGSDIAG_001	
Approbation par le board du réseau le 08/02/2018		

1 Liste des arguments à prendre en compte pour l'interprétation des variants

L'interprétation des résultats repose sur un faisceau d'arguments. Ces arguments ont un poids plus ou moins important dans l'interprétation du variant mis en évidence :

PVS/PS/PM/PP : Argument très fort (Pathogenic Very Strong) / fort (Pathogenic Strong) / moyen (Pathogenic Moderate) / faible (Pathogenic Poor) en faveur de la pathogénicité du variant.

BA/BS/BP : Argument suffisant (Benign stand Alone)/fort (Benign Strong) /faible (Benign Poor) en faveur du caractère bénin du variant.

L'interprétation des résultats est sous la responsabilité exclusive du biologiste et consiste à combiner ces arguments pondérés afin d'assigner une des 5 classes suivantes au variant étudié :

- **Classe 1 : Variant bénin**
- **Classe 2 : Variant probablement bénin**
- **Classe 3 : Variant de signification inconnue**
- **Classe 4 : Variant probablement pathogène**
- **Classe 5 : Variant pathogène**

Document (français) très complet sur l interprétation de variants

http://anddi-rares.org/assets/files/recommandations-ngs-diag-2018.pdf

Annotation des variants et impact des variations ou mutations

ANNOVAR :
http://annovar.openbioinformatics.org/en/latest/user-guide/filter/

dbSNFP :
https://sites.google.com/site/jpopgen/dbNSFP

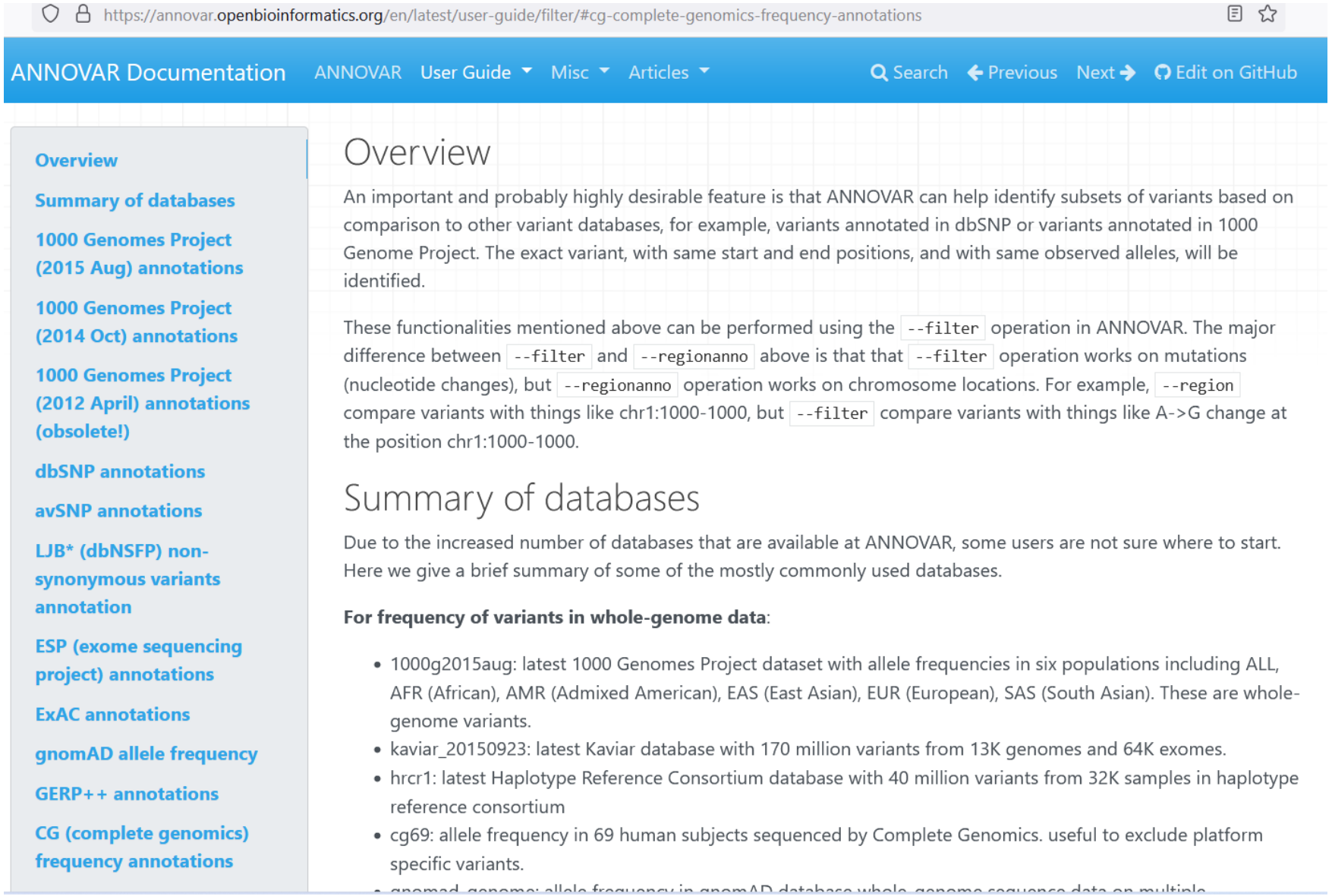
Methodes de prédiction

Table 2: Summary of deleteriousness prediction methods analyzed in our study

Name	Category	Score used for analysis	Deleterious threshold	Information used
SIFT	Function prediction	1 – Score	>0.95	Protein sequence conservation among homologs
PolyPhen-2	Function prediction	Score	>0.5	Eight protein sequence features, three protein structure features
LRT	Function prediction	Score *0.5 (if Omega ≥ 1) or 1 – Score *0.5 (if Omega <1)	P	DNA sequence evolutionary model
MutationTaster	Function prediction	Score (if A or D) or 1 – Score (if N or P)	>0.5	DNA sequence conservation, splice site prediction, mRNA stability prediction and protein feature annotations
Mutation Assessor	Function prediction	(Score-Min)/(Max – Min)	>0.65	Sequence homology of protein families and sub-families within and between species
FATHMM	Function prediction	1 – (Score-Min)/(Max – Min)	≥0.45	Sequence homology
GERP++ RS	Conservation score	Score	>4.4	DNA sequence conservation
PhyloP	Conservation score	Score	>1.6	DNA sequence conservation
SiPhy	Conservation score	Score	>12.17	Inferred nucleotide substitution pattern per site
PON-P	Ensemble score	Score	P	Random forest methodology-based pipeline integrating five predictors
PANTHER	Function prediction	Score	P	Phylogenetic trees based on protein sequences
PhD-SNP	Function prediction	Score	P	SVM-based method using protein sequence and profile information
SNAP	Function prediction	Score	P	Neural network-based method using DNA sequence information as well as functional and structural annotations
SNPs&GO	Function prediction	Score	P	SVM-based method using information from protein sequence, protein sequence profile and protein function
MutPred	Function prediction	Score	>0.5	Protein sequence-based model using SIFT and a gain/loss of 14 different structural and functional properties
KGGSeq	Ensemble score	Score	P	Filtration and prioritization framework using information from three levels: genetic level, variant-gene level and knowledge level
CONDEL	Ensemble score	Score	>0.49	Weighted average of the normalized scores of five methods
CADD	Ensemble score	Score	>15	63 distinct variant annotation retrieved from Ensembl Variant Effect Predictor (VEP), data from the ENCODE project and information from UCSC genome browser tracks

Dong et al, Human Molecular Genetics, 2015, Vol. 24, No. 8 | 2127

ANNOVAR [1] Documenation des annottations ANNOVAR dans les différentes bases



https://annovar.openbioinformatics.org/en/latest/user-guide/filter/

ANNOVAR [2]

Annotation ANNOVAR

- `*_annoTable.txt` from the annotator via ANNOVAR

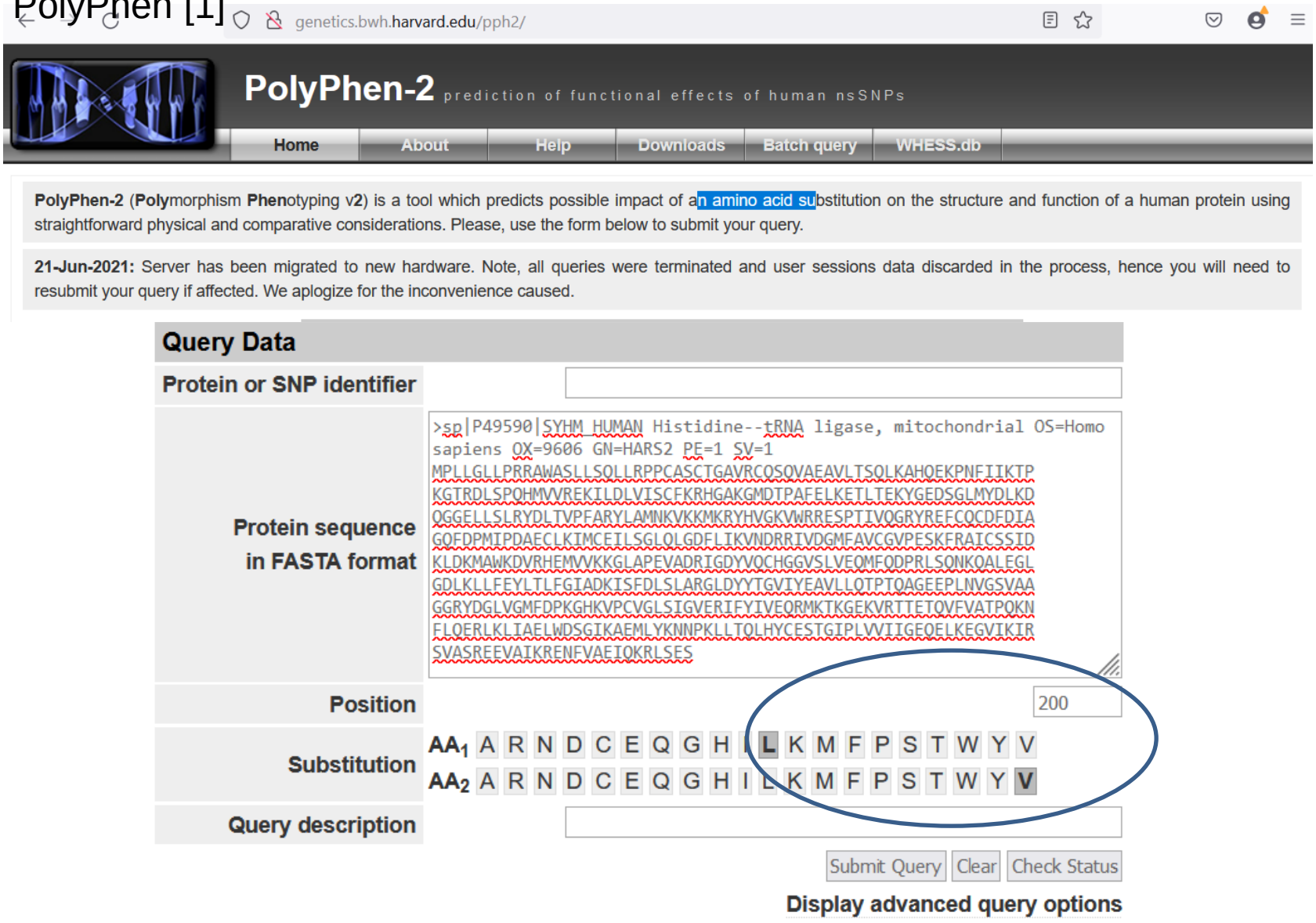
Column Names	Description
Chr	Chromosome number
Start	Start position
End	End position
Ref	Reference base(s)
Alt	Alternate non-reference alleles called on at least one of the samples
COSMIC ID	COSMIC ID
Func.refGene	Regions (e.g., exonic, intronic, non-coding RNA)) that one variant hits; please click here for details.
Gene.refGene	Gene name associated with one variant
ExonicFunc.refGene	Exonic variant function, e.g., nonsynonymous, synonymous, frameshift insertion.please click here for details.
AAChange.refGene	Amino acid change. For example, SAMD11:NM_152486:exon10:c.T1027C:p.W343R stands for gene name, Known RefSeq accession, region, cDNA level change, protein level change.

http://annovar.openbioinformatics.org/en/latest/user-guide/filter/

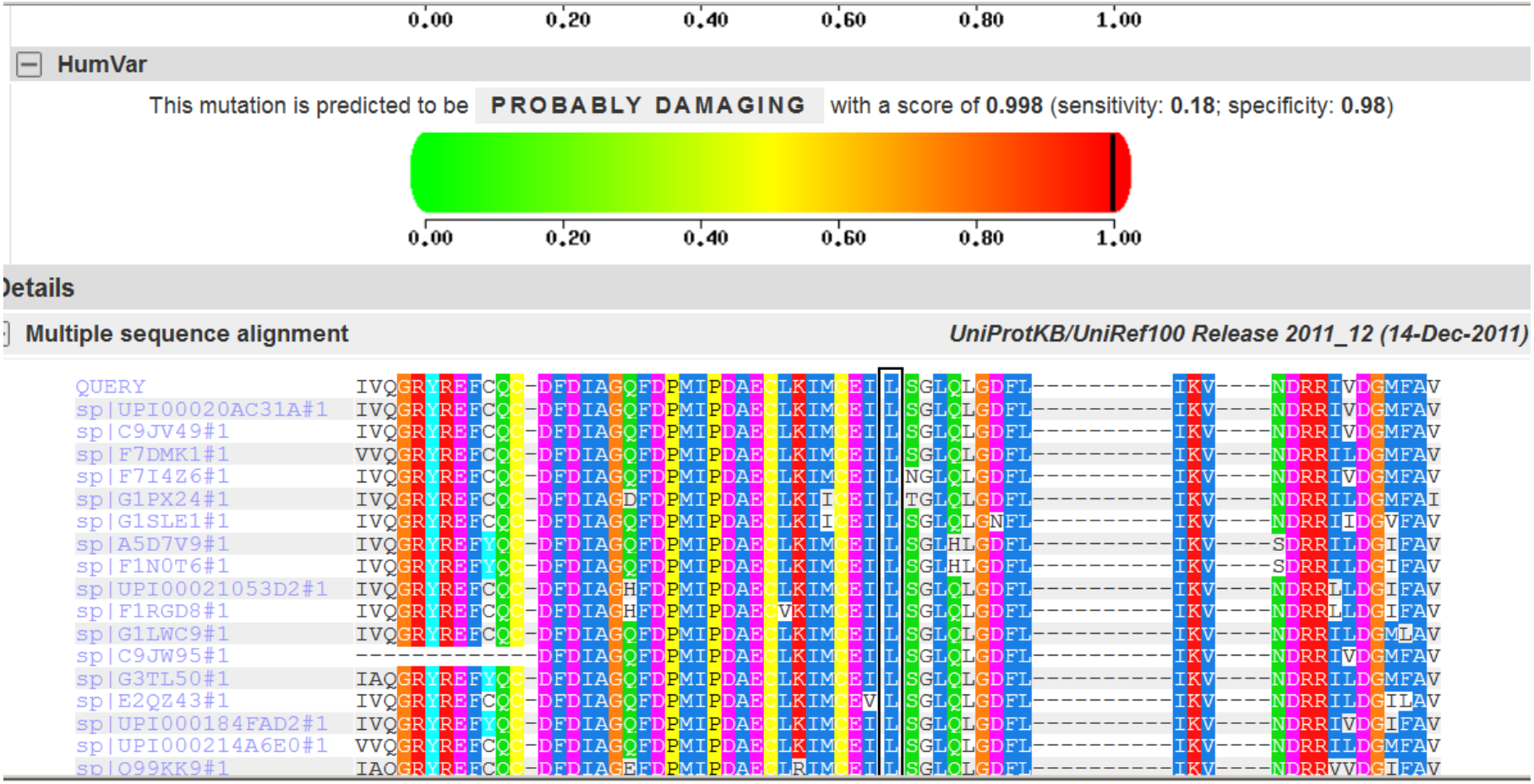
Annotation ANNOVAR (2)

SIFT_score	SIFT score. See the dbNSFP information table for details.
SIFT_pred	SIFT prediction. See the dbNSFP information table for details.
Polyphen2_HDIV_score	Pholyphen2 score based on HDIV. See the dbNSFP information table for details.
Polyphen2_HDIV_pred	Pholyphen2 prediction based on HDIV. See the dbNSFP information table for details.
Polyphen2_HVAR_score	Polyphen2 score based on HVAR. See the dbNSFP information table for details.
Polyphen2_HVAR_pred	Polyphen2 prediction based on HVAR. See the dbNSFP information table for details.
LRT_score	LRT score. See the dbNSFP information table for details.
LRT_pred	LRT prediction. See the dbNSFP information table for details.
MutationTaster_score	MutationTaster score. See the dbNSFP information table for details.
MutationTaster_pred	MutationTaster prediction. See the dbNSFP information table for details.
MutationAssessor_score	MutationTaster score. See the dbNSFP information table for details.

PolyPhen [1]



PolyPhen-2 (**Poly**morphism **Phen**otyping v**2**) is a tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations. Please, use the form below to submit your query.


0.00
0.20
0.40
0.60
0.80
1.00
HumVar
This mutation is predicted to be PROBABLY DAMAGING with a score of 0.998 (sensitivity: 0.18; specificity: 0.98)
0.00
0.20
0.40
0.60
0.80
1.00
Details
Multiple sequence alignment
UniProtKB/UniRef100 Release 2011_12 (14-Dec-2011)
QUERY
sp|UPI00020AC31A#1
sp|C9JV49#1
sp|F7DMK1#1
sp|F7I4Z6#1
sp|G1PX24#1
sp|G1SLE1#1
sp|A5D7V9#1
sp|F1N0T6#1
sp|UPI00021053D2#1
sp|F1RGD8#1
sp|G1LWC9#1
sp|C9JW95#1
sp|G3TL50#1
sp|E2QZ43#1
sp|UPI000184FAD2#1
sp|UPI000214A6E0#1
sp|O99KK9#1

L200 V

SIFT [1]

Sorting Intolerant From Tolerant

Home Help Code Contact us

SIFT predicts whether an **amino acid substitution affects protein function** based on sequence homology and the physical properties of amino acids. SIFT can be a **missense mutations.**

UPDATE on 31 Mar 2022

- Move to new server

UPDATE on 6 Dec 2019

- Change SIFT4G Annotator to browse for genome locally instead of dynamically querying SIFT website due to change in A-STAR

Genome Tools

SNV / SNP prediction

SIFT For Genomes Predictions for human build 37, 38, and > 200 genomes

SIFT For Genomes (Online submission) (**Beta**) Predictions for some model organisms (e.g. human, mouse, worm, yeast).

SIFT nonsynonymous single nucleotide variants (genome-scale) (human build 37)

dbSNP rsIDs (SIFT4G predictions)

Single Protein Tools

SIFT Sequence

SIFT Related Sequence

SIFT Aligned Sequences

http://sift.bii.a-star.edu.sg/

SIFT Sequence

SIFT Sequence provides SIFT predictions for a given protein FASTA sequence. This will take 10-15 min because we must search your protein sequence against a database to pick the related sequences. You can also submit your protein sequence and related sequences or aligned sequences if you already have them.

Results are deleted after 24 hours, so please save them!
[Preventing connection failures]

Protein sequence

Name of file containing protein query sequence (fasta format).
Parcourir... Aucun fichier sélectionné.

-or-

Paste in your protein query sequence (Upload example) (fasta format).

```
sapiens OX=9606 GN=HARS2 PE=1 SV=1
MPLLGLLPRRAWASLLSQLLRPPCASCTGAVRCQSQVAEAVLTSQLKAHQEKPNFIIKTP
KGTRDLSPQHMVVREKILDLVISCFKRHGAKGMDTPAFELKETLTEKYGEDSGLMYDLKD
QGGELLSLRYDLTVPFARYLAMNKVKKMKRYHVGKVWRRESPTIVQGRYREFCQCDFDIA
GQFDPMIPDAECLKIMCEILSGLQLGDFLIKVNDRRIVDGMFAVCGVPESKFRAICSSID
KLDKMAWKDVRHEMVVKKGLAPEVADRIGDYVQCHGGVSLVEQMFQDPRLSQNKQALEGL
GDLKLLFEYLTLFGIADKISFDLSLARGLDYYTGVIYEAVLLQTPTQAGEEPLNVGSVAA
GGRYDGLVGMFDPKGHKVPCVGLSIGVERIFYIVEQRMKTKGEKVRTTETQVFVATPQKN
FLQERLKLIAELWDSGIKAEMLYKNNPKLLTQLHYCESTGIPLVVIIGEQELKEGVIKIR
SVASREEVAIKRENFVAEIQKRLSES
```

Enter the substitutions of interest [format]:

```
V368L
```

SIFT [3]

SIFT Results (Protein Sequences)

Your job id is 953a4b2e02 and is currently running.

If your browser times out before results are shown, please go to https://sift.bii.a-star.edu.sg//sift-bin/format.pl?953a4b2e02_sequences in **20 minutes**. Opening it up before 20 minutes will throw an error (in which case, just refresh).

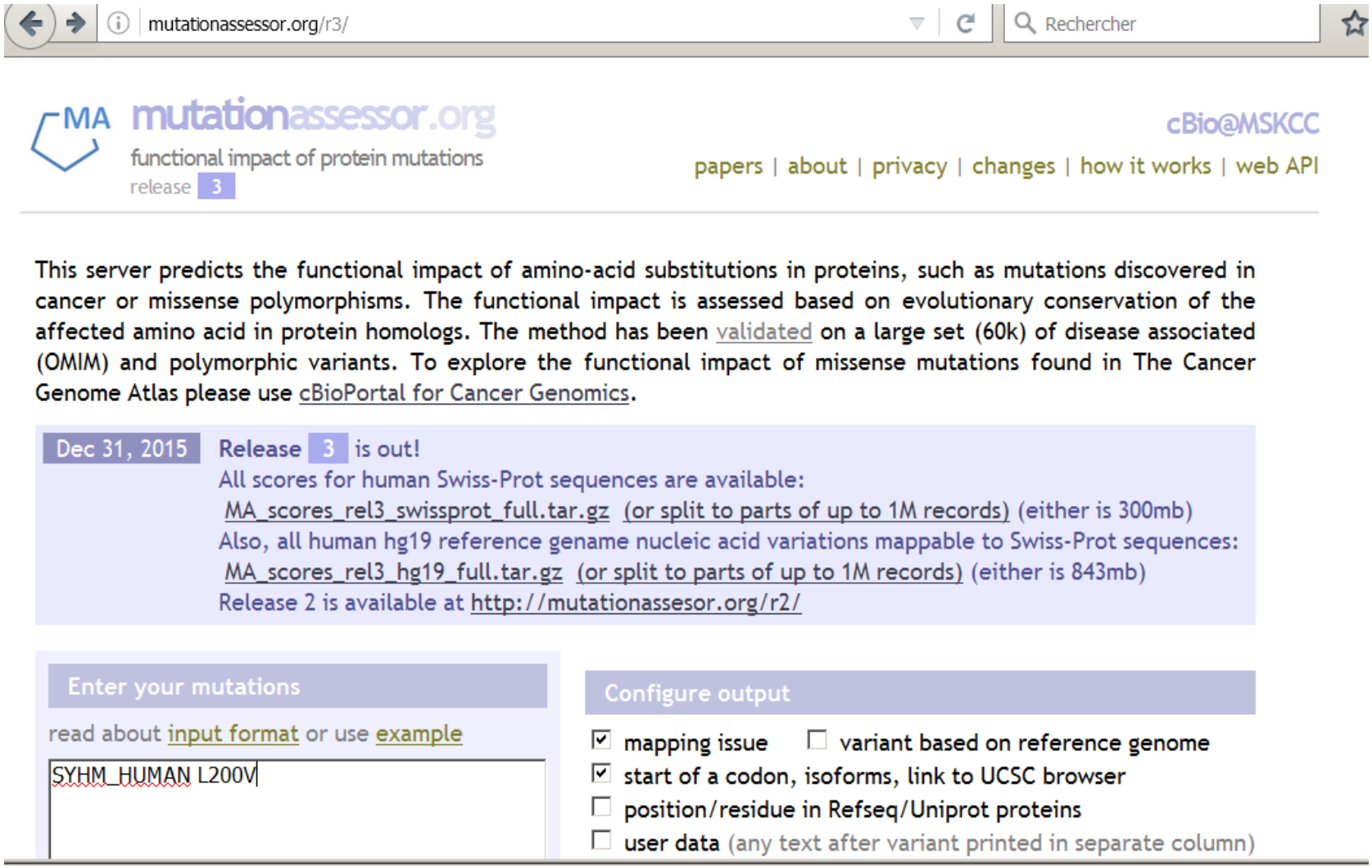
Problems? Contact us with your job id.
A notification will be sent to pdessen@free.fr once the results are ready for viewing.

Use web browser to view these files.
*Please note that tables will take some time to load.*Your results will be available here for the next 24 hours.
17 sequences were selected to be closely related to your query sequence.
PSIBLAST alignment of submitted sequences
Alignment in FASTA format (for modification)
The alignment taken from PSIBLAST is returned in msf format.
Note: Xes are placeholders at the beginning and end of sequences. While - means a gap in the alignment an X means a lack of information such as a partial alignment or incomplete sequence and do not contribute to the prediction.
Please check the sequences that have been chosen. If the sequences are too diverged from your query or the alignment is questionable, we suggest you modify the fasta-formatted file above and resubmit.

Predict Not Tolerated	Position	Seq Rep	Predict Tolerated
mwifvlcyrpqht	366G	1.00	keasdGN
ywvtsrqpnmkihgfedca	367L	1.00	L
whgdnyrqekspcfmal	368V	1.00	TIV
wcfmyihvlp	369G	1.00	rtqsdANKEG

pos	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
351E	0.81	0.14	0.00	0.36	1.00	0.00	0.02	0.01	0.01	0.06	0.01	0.01	0.03	0.03	0.21	0.02	0.03	0.03	0.01	0.00	0.01
352P	0.75	0.44	0.08	0.15	0.14	0.25	0.20	0.09	0.12	0.16	0.16	0.07	0.21	0.99	0.12	0.12	0.86	1.00	0.20	0.04	0.14
368V	1.00	0.02	0.01	0.00	0.00	0.01	0.00	0.00	0.18	0.00	0.03	0.01	0.00	0.01	0.00	0.00	0.00	0.12	1.00	0.00	0.00
369G	1.00	0.18	0.01	0.12	0.24	0.01	1.00	0.02	0.02	0.22	0.03	0.01	0.21	0.04	0.06	0.06	0.09	0.06	0.03	0.00	0.01

Mutation Assessor [1]



The functional impact is assessed based on evolutionary conservation of the affected amino acid in protein homologs

http://mutationassessor.org/r3/

Mutation Assessor [2]

Enter your mutations

read about input format or use example

SYHM_HUMAN L200V

download as .csv file | submit

Configure output

- [x] mapping issue
- [ ] variant based on reference genome
- [x] start of a codon, isoforms, link to UCSC browser
- [ ] position/residue in Refseq/Uniprot proteins
- [ ] user data (any text after variant printed in separate column)
- [ ] individual scores, misc. msa info
- [ ] number of COSMIC alterations and SNPs in RefSeq protein
- [ ] COSMIC ([ ] detailed) / SNPs in RefSeq position
- [ ] known functional regions of Uniprot protein
- [ ] nearby Pfam domains
- [ ] all Pfam domains in a protein
- [x] binding sites (directly computed from PDB structures)

[-] input/options. Mouseover column header to read short description. Read how it works document for more details.

	Mutation	AA variant	Gene	MSA	PDB	Func. Impact	FI score	Uniprot	Refseq	MSA height	Codon start position	Func. region	Protein bind.site	DNA/RNA bind.site	small.mol bind.site
1	SYHM_HUMAN L200V	L200V	HARS2	msa	pdb	high	3.585	SYHM_HUMAN	NP_036340	638	isoforms hg19:chr5:140075395				

*** all swissprot (release 2014_07) human gene positions have been pre-computed and are currently available.
*** submission of novel variations for scoring is temporarily unavailable.
server time spent : 5.537398 secs.
peak memory used : 3,150,400

Refseq position	Refseq residue	Func. region	Protein bind.site	DNA/RNA bind.site	small.mol bind.site	N.Cosmic	N.SNPs	Cosmic@position	regions@position	domain@position
368	V					37	16		VARIANT 368 368 V->L (in PRLTS2; the mutant protein is expressed, can dimerize and localizes to the mitochondria; has significantly decreased enzymatic activity compared to wild-type).	-133 aa ends: tRNA synthetase class II core domain (G, H, P, S and T) +43 aa starts: Anticodon binding domain

Exercices d analyse comparative des séquences des SYHM et de leur conservation au cours de l'évolution

Extraire les séquences de HARS2 humain (SYHM_HUMAN) de Uniprot

Faire un BLAST de cette séquence SYHM_HUMAN / UniProt
Filtrer sur SwissProt (17 séquences)

Récupérer les séquences en format FASTA sur le PC
(Download alignments (non compressés) par copier coller dans un fichier text

Utiliser le logiciel seaview4 pour aligner les séquences
(glisser le fichier fasta sur le logiciel)

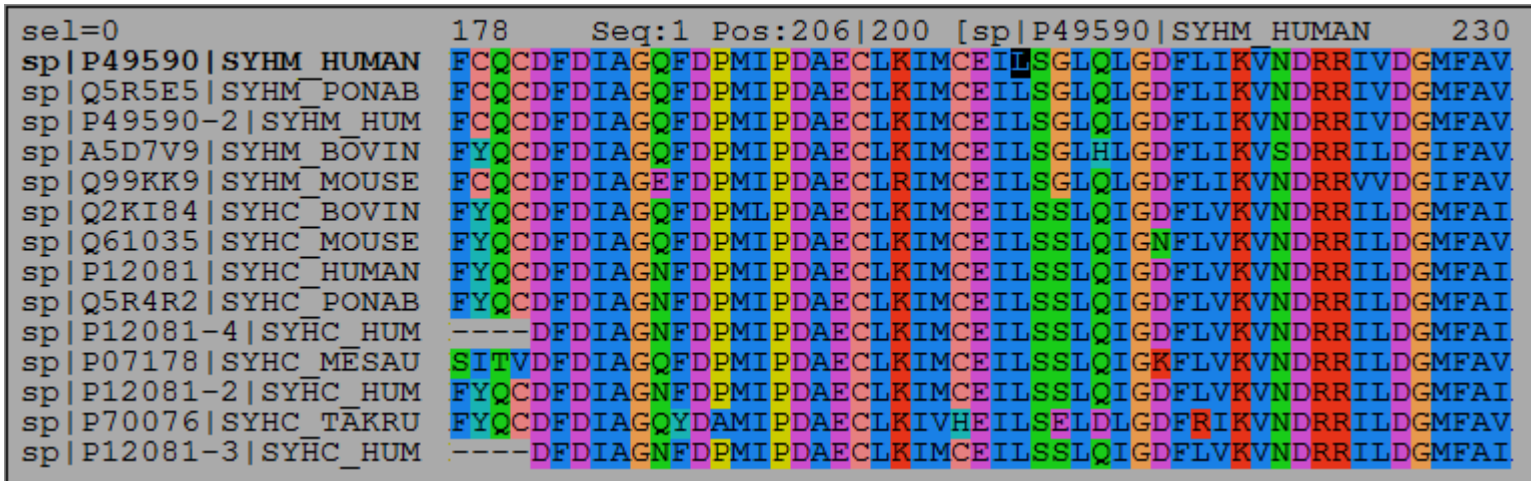
(récupérer le logiciel seaview4 (MS-Windows) sur
http://doua.prabi.fr/software/seaview

Faire glisser le fichier « fasta » sur la fenêtre de seaview4
Aligner les séquences (menu ALIGN)
Visualiser autour des positions 200 et 368

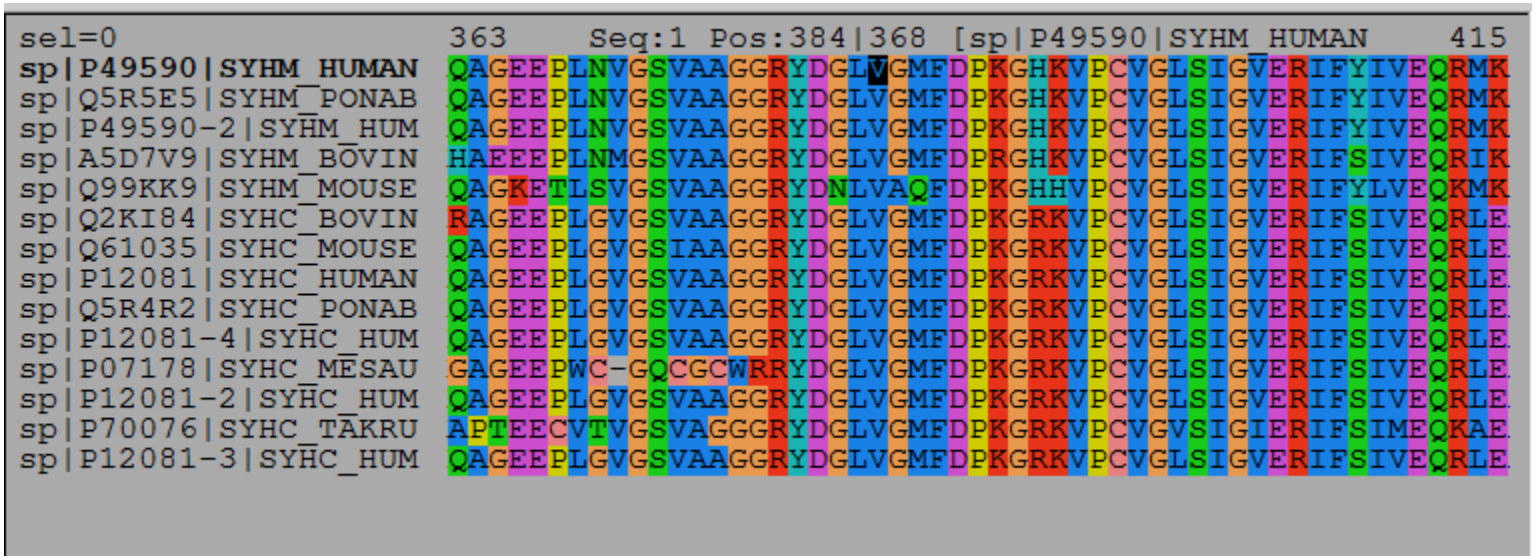
>sp|P49590|SYHM_HUMAN Histidine--tRNA ligase, mitochondrial OS=Homo sapiens OX=9606 GN=HARS2 PE=1 SV=1

MPLLGLLPRRAWASLLSQLLRPPCASCTGAVRCQSQVAEAVLTSQLKAHQEKPNFIIKTP
KGTRDLSPQHMVVREKILDLVISCFKRHGAKGMDTPAFELKETLTEKYGEDSGLMYDLKD
QGGELLSLRYDLTVPFARYLAMNKVKKMKRYHVGKVWRRESPTIVQGRYREFCQCDFDIA
GQFDPMIPDAECLKIMCEILSGLQLGDFLIKVNDRRIVDGMFAVCGVPESKFRAICSSID
KLDKMAWKDVRHEMVVKKGLAPEVADRIGDYVQCHGGVSLVEQMFQDPRLSQNKQALEGL
GDLKLLFEYLTLFGIADKISFDLSLARGLDYYTGVIYEAVLLQTPTQAGEEPLNVGSVAA
GGRYDGLVGMFDPKGHKVPCVGLSIGVERIFYIVEQRMKTKGEKVRTTETQVFVATPQKN
FLQERLKLIAELWDSGIKAEMLYKNNPKLLTQLHYCESTGIPLVVIIGEQELKEGVIKIR
SVASREEVAIKRENFVAEIQKRLSES

L 200 V



V 368 L



Une conservation est une indication de la pertinence de l'effet d'un variant

Comme exercice (en différé)

Toutes ces interrogations des bases de données peuvent être refaites avec d’autres variants !

Peut on noter des bases n’ayant pas de données pour certains variants
Les bases donnent t elle la même signification de pathogénicité ?

Comparer quand c est possible la fréquence des variants dans les différentes populations (cas de enSembl, et de gnomAD)