



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

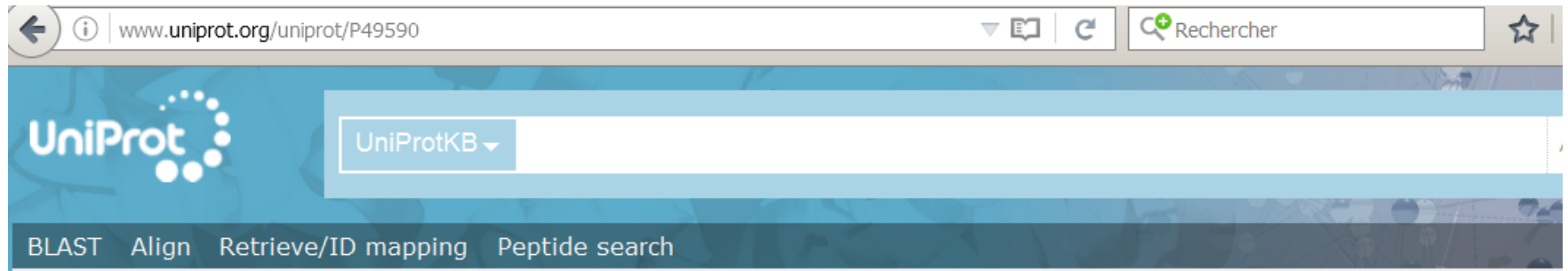
Le syndrome de Perrault (SP) est caractérisé par l'association d'une dysgénésie ovarienne chez une femme ayant une surdité neurosensorielle. Dans de récentes études, quelques auteurs ont décrit des anomalies neurologiques, en particulier une ataxie cérébelleuse progressive et un déficit intellectuel.

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**



UniProtKB - P49590 (SYHM_HUMAN)

Display



[Help video](#)

[Otl](#)

Entry

Publications

Feature viewer

Feature table

None

☒ Function

☒ Names & Taxonomy

☒ Subcell. location

☒ Pathol./Diseas

Protein | Probable histidine--tRNA ligase, mitochondrial

Gene | HARS2

Organism | *Homo sapiens (Human)*

Status | Reviewed - Annotation score: - Experimental evidence at protein levelⁱ

Functionⁱ

Catalytic activityⁱ

ATP + L-histidine + tRNA(His) = AMP + diphosphate + L-histidyl-tRNA(His).

<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

Checksumⁱ E1CE879837AE26E7

¹⁰ MPLLGLLPRR	²⁰ AWASLLSQLL	³⁰ RPPCASCTGA	⁴⁰ VRCQSQVAEA	⁵⁰ VLTSQLKAHQ	⁶⁰ EKPNFIIKTP	⁷⁰ KGTRDLSPQH	⁸⁰ MVVREKILDL
⁹⁰ VISCFKRHGA	¹⁰⁰ KGMDTPAFEL	¹¹⁰ KETLTEKYGE	¹²⁰ DSGLMYDLKD	¹³⁰ QGGELLSLRY	¹⁴⁰ DLTVPFARYL	¹⁵⁰ AMNKVKMKR	¹⁶⁰ YHVGKVVRR
¹⁷⁰ SPTIVQGRYR	¹⁸⁰ EFCQCDFDIA	¹⁹⁰ GQFDPMPIDA	²⁰⁰ ECLKIMCEIL	²¹⁰ SGLQLGDFLI	²²⁰ KVNDRRIVDG	²³⁰ MFAVCGVPES	²⁴⁰ KFRAICSSID
²⁵⁰ KLDKMAWKDV	²⁶⁰ RHEMVVKKGL	²⁷⁰ APEVADRIGD	²⁸⁰ YVQCHGGVSL	²⁹⁰ VEQMFQDPRL	³⁰⁰ SQNKQALEGL	³¹⁰ GDLKLLFEYL	³²⁰ TLFGIADKIS
³³⁰ FDLSLARGLD	³⁴⁰ YYTGVIEYAV	³⁵⁰ LLQTPTQAGE	³⁶⁰ EPLNVGSVAA	³⁷⁰ GGRYDGLVGM	³⁸⁰ FDPKGHKVPC	³⁹⁰ VGLSIGVERI	⁴⁰⁰ FYIVEQRMKT
⁴¹⁰ KGEKVRTTET	⁴²⁰ QVFVATPQKN	⁴³⁰ FLQERLKLIA	⁴⁴⁰ ELWDSGIKAE	⁴⁵⁰ MLYKNNPKLL	⁴⁶⁰ TQLHYCESTG	⁴⁷⁰ IPLVVIIEGQ	⁴⁸⁰ ELKEGVIKIR
⁴⁹⁰ SVASREEVAI	⁵⁰⁰ KRENFVAEIQ	KRLSES					

Val368Leu ou Leu200Val



- Function
- Names & Taxonomy
- Subcellular Location
- Disease & Variants**
- PTM/Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence & Isoform
- Similar Proteins

Disease & Variantsⁱ

Involvement in diseaseⁱ

Perrault syndrome 2 (PRLTS2)

2 Publications

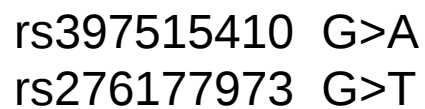
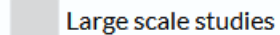
Note | The disease is caused by variants affecting the gene represented in this entry

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.

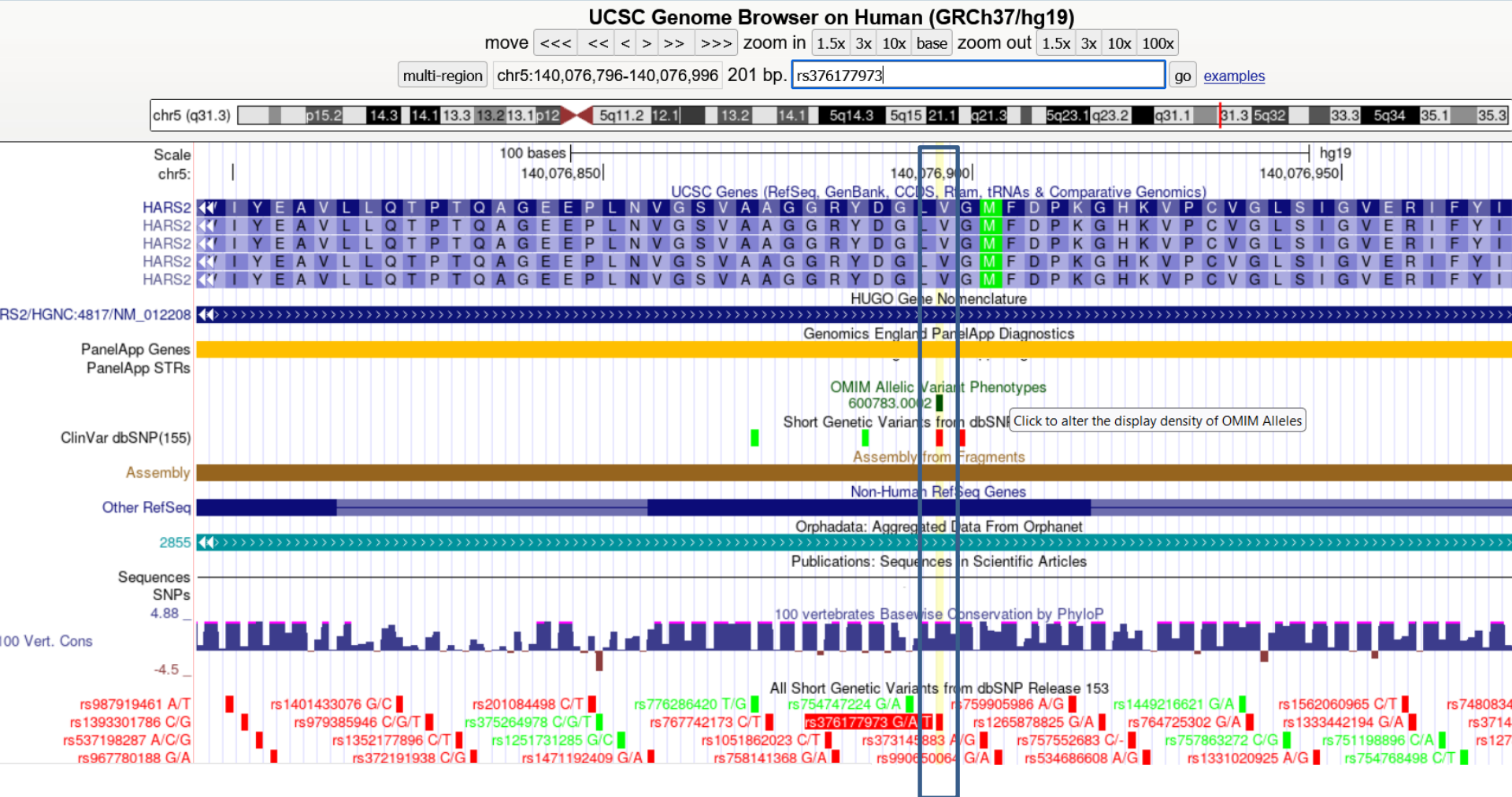
See also | [MIM:614926](#) ↗

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



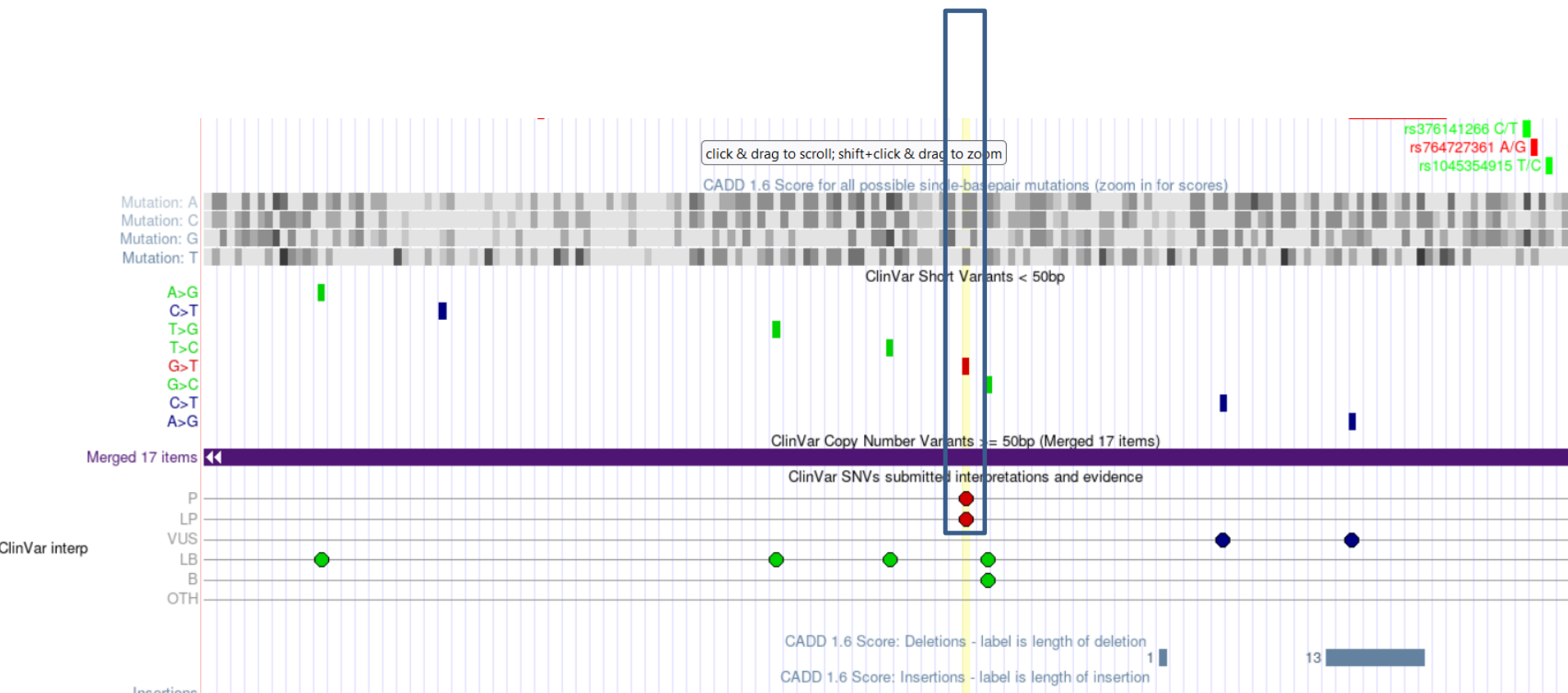
Recherche sur le GoldenPath (hg19) du SNP rs376177973



Val368Leu SNP rs376177973

hg19 : [chr5:140076896-140076896](https://genome.ucsc.edu/chr5:140076896-140076896)<https://genome.ucsc.edu>

UCSC [2]



Un clic sur le variant rouge -> documentation

ClinVar SNVs submitted interpretations and evidence

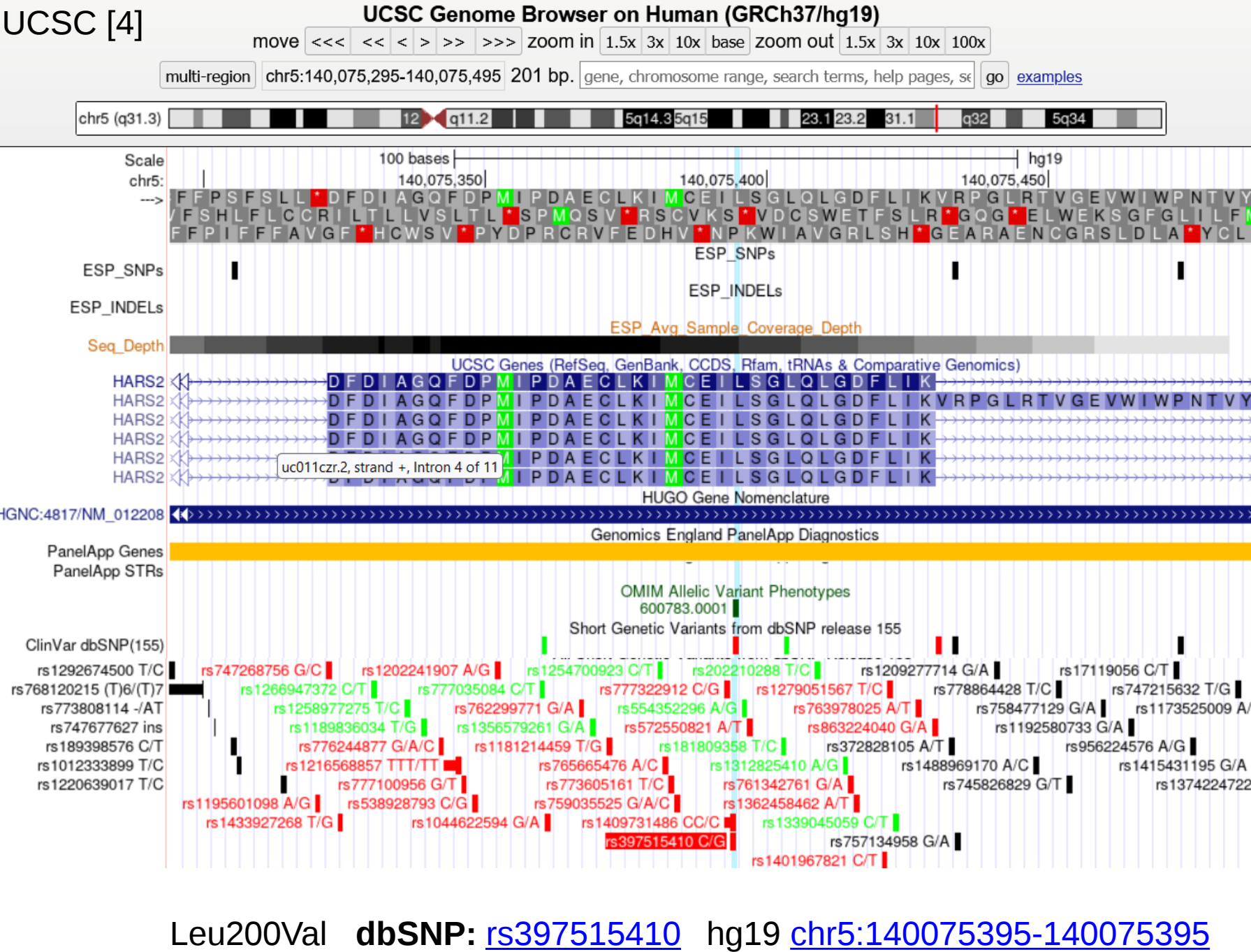
Position: [chr5:140076896-140076896](#)

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	-



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 of 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 dépôt de polymorphisme



National Library of Medicine

National Center for Biotechnology Information

dbSNP

SNP

HARS2

Create alert Advanced

Clinical Significance

- likely benign
- ✓ likely pathogenic
- ✓ pathogenic
- ✓ pathogenic likely pathogenic

Validation Status

- by-ALFA
- by-cluster
- by-frequency

Publication

- LitVar Annotated
- PubMed Linked

Function Class

- inframe insertion
- intron
- missense

Variation Class

- delins

Annotation

- somatic

clear

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

Send to: ▼

Search results

Items: 16

[Reference SNP \(rs\)](#)

<https://www.ncbi.nlm.nih.gov/books/NBK21088/>

i Filters activated: likely pathogenic, pathogenic likely pathogenic, pathogenic. [Clear all](#) to show 3907 items.

☐ rs140540222 [*Homo sapiens*]

1.

Variant type: SNV

Alleles: C>T [\[Show Flanks\]](#)

Chromosome: 5:140695556 (GRCh38)

5:140075141 (GRCh37)

Canonical SPDI: NC_000005.10:140695555:C:T

Gene: HARS2 [\(Varview\)](#)

Functional Consequence: coding_sequence_variant,intron_variant,missense_variant

Clinical significance: likely-pathogenic,uncertain-significance

Validated: by frequency,by alfa,by cluster

MAF: T=0.000045/2 [\(ALFA\)](#)

T=0.000049/6 (ExAC)

T=0.000056/14 (GnomAD_exomes)

[...more](#)

HGVSc: NC_000005.10:g.140695556C>T NC_000005.9:g.140075141C>T NG_032158.1:g.831G>A

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_001350405.1:p.Leu130Val, NP_001205000.1:p.Leu175Val, NP_001205001.1:p.Leu50Val

...le:

Leu200Val rs397515410
Coordonnées relatives aux transcrits

[Clinical Significance](#)
 ✓ 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](#)

clear [Display Settings:](#) Summary [Send to:](#)

Filters activated: likely benign, likely pathogenic, pathogenic likely pathogenic, pathogenic. [Clear all to show 1 items.](#)

☐ **rs376177973** [*Homo sapiens*]

1.

Variant type:	SNV
Alleles:	G>A,T [Hide Flanks]
	AGTGATCTATGAAGCAGTGCTGCTGCAGACCCCAACTCAGGCTGGGGAGG AGCCCCCTGAATGTGGGCAGTGTGGCTGCTGGTGGGCCTATGATGGGCTG [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:n.Val374Met, NP_001350464.1:n.Val374Leu

Attention :

La nomenclature HGVS est
nécessaire pour une bonne
identification

ClinVar [1]

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

**National Library of Medicine**
National Center for Biotechnology Information

Log in

ClinVar

ClinVar

HARS2[gene]

Search

Create alert

Advanced

Home

About

Access

Help

Submit

Statistics

FTP

Clinical significance clear
Conflicting interpretations (0)
Benign (0)
Likely benign (0)
Uncertain significance (0)
✓ **Likely pathogenic** (6)
✓ **Pathogenic** (4)
Molecular consequence clear
Frameshift (0)
✓ **Missense** (8)
Nonsense (0)
Splice site (0)
ncRNA (0)
Near gene (0)

Search results
Display options **Sort by Location** **Download** Items: 8
Filters activated: Likely pathogenic, Pathogenic, Missense. [Clear all](#) to show 139 items.
The following terms were not found in ClinVar: clinsig pathogenic low penetrance[Properties], clinsig established risk allele[Properties].

Variation Location	Gene(s)	Protein change	Condition(s)	Clinical significance (Last reviewed)	Review status
☐ NM_012208.4(HARS2):c.137T>A 1. (p.Leu46Gln) GRCh37: Chr5:140073204 GRCh38: Chr5:140693619	HARS2	L46Q, L52Q	Perrault syndrome 2	Likely pathogenic (May 20, 2019)	no assertion criteria provided
☐ NM_012208.4(HARS2):c.598C>G 4. (p.Leu200Val) GRCh37: Chr5:140075395 GRCh38: Chr5:140695810	HARS2	L200V, L130V, L56V, L206V, L175V	Perrault syndrome 2	Pathogenic/Likely pathogenic (Jun 27, 2013)	no assertion criteria provided
☐ NM_012208.4(HARS2):c.1102G>T 7. (p.Val368Leu) GRCh37: Chr5:140076896 GRCh38: Chr5:140697311	HARS2	V368L, V374L, V298L, V343L, V224L	Perrault syndrome 2	Pathogenic/Likely pathogenic (Jun 27, 2013)	no assertion criteria provided

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

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]

← → ↺ <https://www.ncbi.nlm.nih.gov/variation/view> ☆

NIH National Library of Medicine
National Center for Biotechnology Information

Variation Viewer

Share this page Reset All More Info ▾

New to Variation Viewer? [Read our quick overview!](#) X

Homo sapiens
(human)

Assembly: GRCh38.p13 (GCF_000001405.39) • Chr 5 (NC_000005.10)

Search assembly
HARS2 Examples ▸

◀ Pick Assembly
◀ Tracks and User Data
◀ History
◀ Assembly Region Details

NC_000005.10: 140,690,636 - 140,700,106

Region ▾ Gene HARS2 Transcript NM_012208.4 Exons: click an exon to zoom in, mouse over to see details

NC_000005.10 ◀ ▶ 🔍 ⚙️ 📄

140,691 K 140,692 K 140,693 K 140,694 K 140,695 K 140,696 K 140,697 K 140,698 K 140,699 K

Genes, NCBI Homo sapiens Annotation Release 110, 2022-04-08

HARS1 [+14]

HARS2 [+10]

Clinical, dbSNP b155 v2

Live RefSNPs, dbSNP b155 v2

dbVar Clinical Structural Variants (nstd102)

dbVar Non-Pathogenic Clinical Structural Variants (subset of nstd102)

dbVar Pathogenic Clinical Structural Variants (subset of nstd102)

ClinVar variants with precise endpoints

Attention :
Ici hg38 (GrCh38)

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

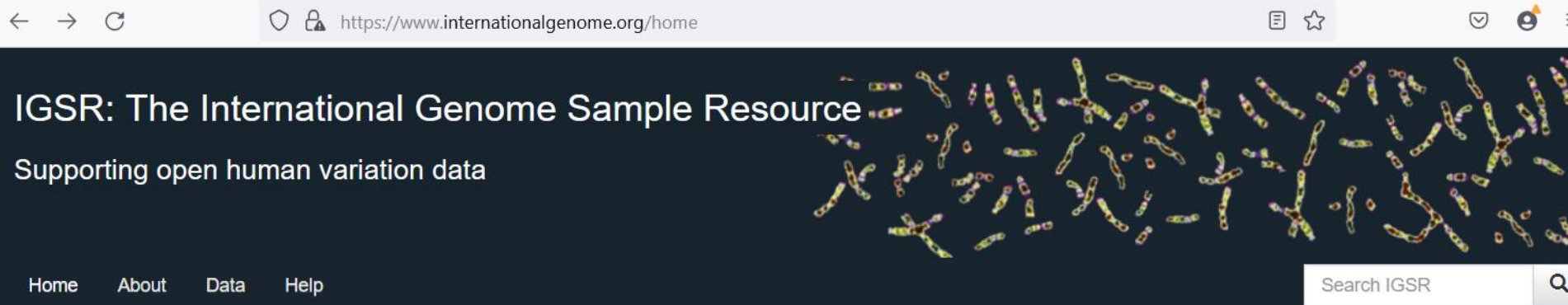
Variation Viewer[2]

Filter by	Download edit columns										Items 1 - 9 of 9	<< First	< Prev	Page 1 of 1	Next >	Last >>
Source database	Variant ID	Location	Variant type	Gene	Molecular consequences	Most severe clinical significance	1000G MAF	GO-ESP MAF	ExAC MAF	Publication						
☒ dbSNP (9)																
☐ dbVar (0)																
In ClinVar																
☒ Yes (9)																
☐ No (0)																
Most severe clinical significance																
☒ pathogenic (1)																
☒ pathogenic-likely-pathogenic (1)																
☒ likely-pathogenic (3)																
☐ drug-response (0)																
☐ confers sensitivity (0)																
☐ other (0)																
☐ conflicting-data-from-submitters (0)																
☒ conflicting-interpretations-of-pathogenicity (1)																
☐ risk-factor (0)																
☐ association (0)																
☐ protective (0)																
☒ uncertain-significance (3)																
☐ not-provided (0)																
☒ likely-benign (0)																
☐ benign-likely-benign (0)																
☐ benign (0)																
☐ affects (0)																
More...																
Variant type																
☒ single nucleotide variant (9)																
☐ copy number variation (0)																

Complet : sélection multiple et filtres

pathogenic
benign
likely benign
uncertain significance

...



IGSR: The International Genome Sample Resource

Supporting open human variation data

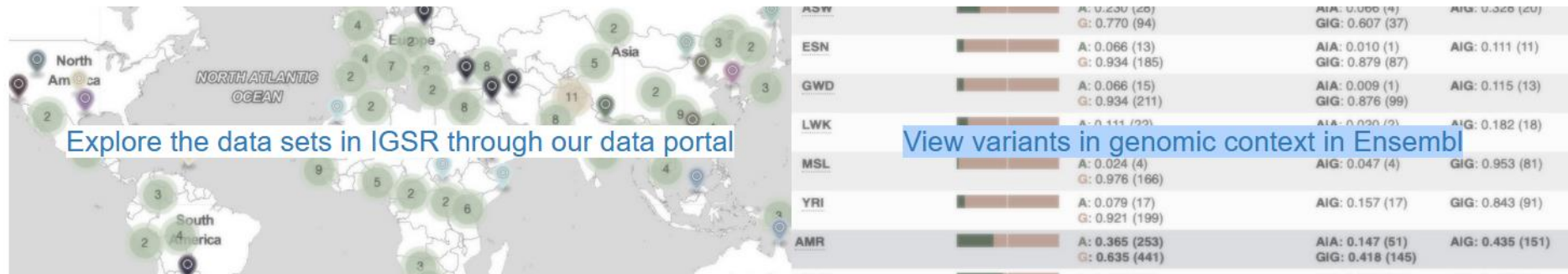
Home About Data Help

Search IGSR

The International Genome Sample Resource

The 1000 Genomes Project created a catalogue of common human genetic variation, using openly consented samples from people who declared themselves to be healthy. The reference data resources generated by the project remain heavily used by the biomedical science community.

The International Genome Sample Resource (IGSR) maintains and shares the human genetic variation resources built by the 1000 Genomes Project. We also update the resources to the current reference assembly, add new data sets generated from the 1000 Genomes Project samples and add data from projects working with other openly consented samples.



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

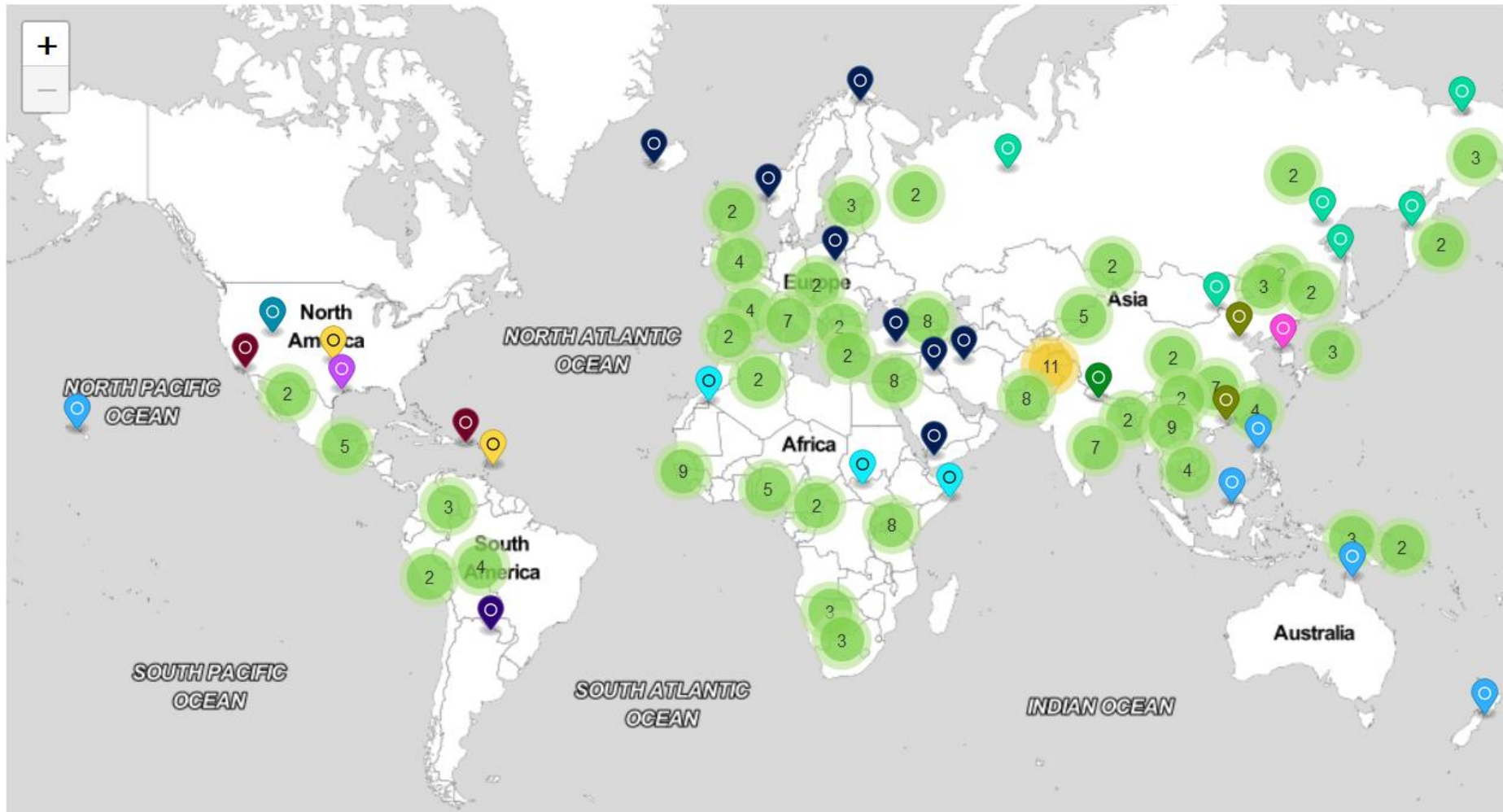
Technology view

Filter by technology ▾

Filter by data collection ▾

Download the list

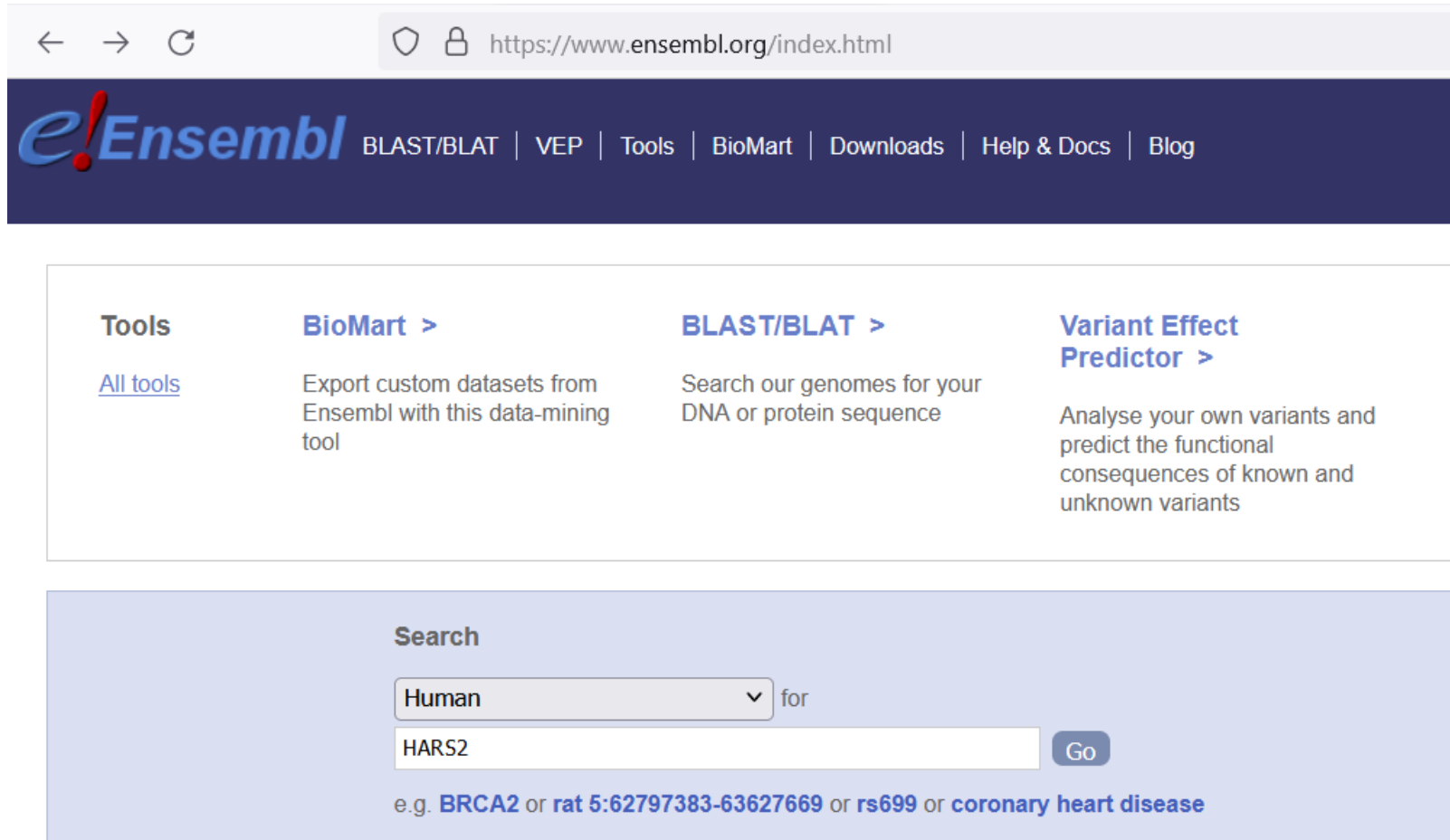
Map colour key



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

enSembl[1]

link depuis IGSR



The screenshot shows the Ensembl genome browser homepage. At the top, there is a navigation bar with the Ensembl logo and links to BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. Below this, there is a section with four main tool categories: Tools (with a link to 'All tools'), BioMart (with a description of exporting custom datasets), BLAST/BLAT (with a description of searching genomes), and Variant Effect Predictor (with a description of analyzing variants). At the bottom, there is a search bar with a dropdown menu set to 'Human' and a text input field containing 'HARS2'. A 'Go' button is next to the input field. Below the search bar, there is a hint text: 'e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease'.

← → ↻ <https://www.ensembl.org/index.html>

e!Ensembl [BLAST/BLAT](#) | [VEP](#) | [Tools](#) | [BioMart](#) | [Downloads](#) | [Help & Docs](#) | [Blog](#)

Tools
[All tools](#)

BioMart >
Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >
Search our genomes for your DNA or protein sequence

Variant Effect Predictor >
Analyse your own variants and predict the functional consequences of known and unknown variants

Search

Human ▼ for
HARS2 [Go](#)

e.g. [BRCA2](#) or [rat 5:62797383-63627669](#) or [rs699](#) or [coronary heart disease](#)

Attention : ici par défaut avec enSembl : build hg38

<https://www.ensembl.org>



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 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

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

X

Consequences: missense variant, pr...(2/30)

Filter Other Columns

Variant ID	Chr: bp	Alleles	Global MAF	Class	Source	Evidence	Clin. Sig.	Conseq. Type	AA	AA coord	SIFT	Poly-Phen	CADD	REVEL	MetaLR	Mutation Assessor	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		-	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		-	missense variant	R/C	363	0	1	32	0.793	0.712	0.996	ENST00000230771.9
rs758141368	5:140697297	G/A	-	SNP	dbSNP		-	missense variant	R/H	363	0	0.999	29	0.877	0.729	0.996	ENST00000230771.9
rs746757469	5:140697300	A/G	-	SNP	dbSNP		-	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			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			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			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		-	missense variant	M/V	370	0.01	0.005	20	0.117	0.11	0.556	ENST00000230771.9



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

Alleles

Change tolerance

Location

Co-located variant

Evidence status ⓘ

Clinical significance ⓘ

HGVS names

Synonyms

Original source

About this variant

Explore this variant ⓘ

missense variant | See all predicted consequences

G/A/T | Ancestral: G | Highest population MAF: < 0.01

CADD: A:26.3, T:25.6 | GERP: 2.82

Chromosome 5:140697311 (forward strand) | VCF: 5 140697311 rs376177973 G A,T

HGMD-PUBLIC CM112637

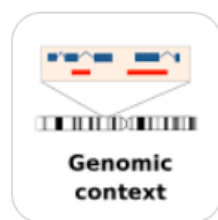


This variant has 84 HGVS names - Show ⓘ

This variant has 4 synonyms - Show ⓘ

Variants (including SNPs and indels) imported from dbSNP (release 154) | View in dbSNP ⓘ

This variant overlaps 14 transcripts, is associated with 1 phenotype and is mentioned in 1 citation.



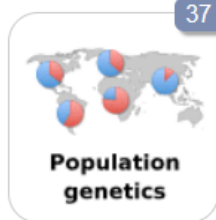
Genomic context



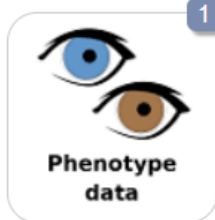
Genes and regulation



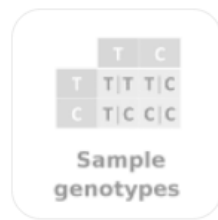
Flanking sequence



Population genetics



Phenotype data



Sample genotypes



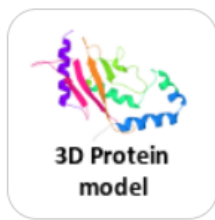
Linkage disequilibrium



Phylogenetic context



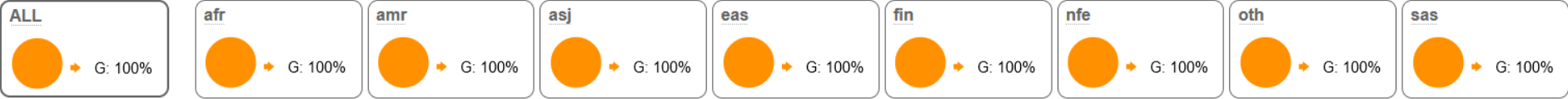
Citations



3D Protein model

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					

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 variants are shown in **green** text. To change the display of the flanking sequence (e.g. hide the other variants, change the length of the flanking sequence), use the "Configure" button.

Variants

Coding sequence

Focus variant

Frameshift

Inframe deletion

Intronic

Missense

Splice region

Stop gained

Synonymous

Markup

loaded

```

CARTBAASWGYRGAKAMTTTATTTCTSTCCAGGTGKGRYRTCCSTACTRRRRCAVATRTT
TCAVGVWYCCAGAYTRTSCSMKAACAARHARRYMCTRGAGGGCCTSRRAGAMYAAAGCT
GYTATTTTGAATACSTGAHTTTATTTGCMWYTGCTGRTWAGGTAARMKGAVTGSPRRGTGG
ACCTCYTGCTGRSYTCTAGGGYITCARRGYCCBRAGTYTARWYTDGGACTGACTGTAAT
STKKYSYCCACAGAYMTCTTTGACYTCAGYYTSGCTBVRRCCTAGACTRCTAYACAGG
AGTGWISTVTRAAGCRGTGCYGYRCASACCECAAYTSASXWGDGGAGGAGCYBYTSAA
TRTGGRCARTGYGGCKGYTGCGGGYRSTRYGATGGRCCTGDTSGCRTTRTTGRCBYCRA
GGRCCRCAARGKTRCYRTGTGTRRGAYTSAGCRTTGGSDYAGAGYRMATMTYYAYRYGT
RGARCRGARGRRTGAAGGTAGGTCCTWGAAYAGGTGTGASRTSRRCRTRGABACYTAARAAC
SCTYYTGTYTCTGGAGRTGTAGTYGSAGTGGTTTTCTGATKATTYTTTTTSTSTTKATWT
TTGGAATTGSBGRTGRAARRGAGGTKTTTATTAGKTKTACYTTYTTCTCTCTYWYTAGRC
CRRAGGTGAGAAGGWRYRGACTRCRGAVACTCAARTGYTYGTGGCYACASCRCA SAAGAA
CTTTSTCMAAGAAAYDGYTRAARBTTAYTGSMGAGCTTTGGGWTWCTGCAATCMARGTRTR
RTGGARYYGATAYCYGAGCCA

```

D : non C = A,T,G

gnomAD[1]



Genome Aggregation Database

gnomAD v2.1.1



Search by gene, region, or variant

Or

- [Find co-occurrence of two variants](#)
- [Download gnomAD data](#)
- [Read gnomAD publications](#)

Please note that gnomAD v2.1.1 and v3.1.2 have substantially different but overlapping sample compositions and are on different genome builds. For more information, see ["Should I switch to the latest version of gnomAD?"](#)

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/>

HARS2 histidyl-tRNA synthetase 2, mitochondrial

Dataset gnomAD v2.1.1 gnomAD SVs v2.1

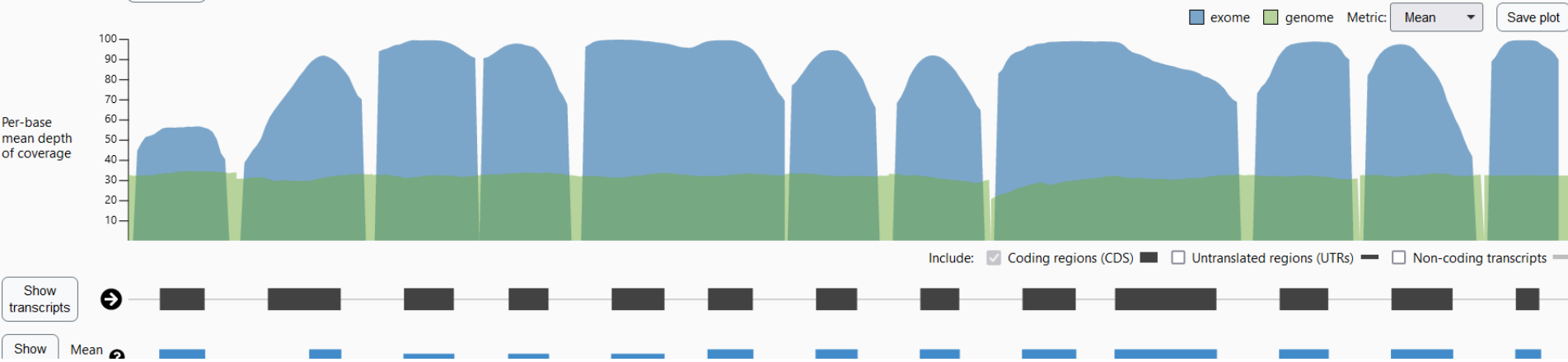
Genome build GRCh37 / hg19
Ensembl gene ID ENSG00000112855.10
Ensembl canonical transcript ENST00000230771.3
Other transcripts ENST00000513688.1, ENST00000510104.1, and 12 more
Region 5:140071011-140078889
External resources Ensembl, UCSC Browser, and more

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).

Viewing full gene. Zoom in



ClinVar variants

☒ Pathogenic / likely pathogenic

☒ Uncertain significance / conflicting

☐ Benign / likely benign

☐ Other

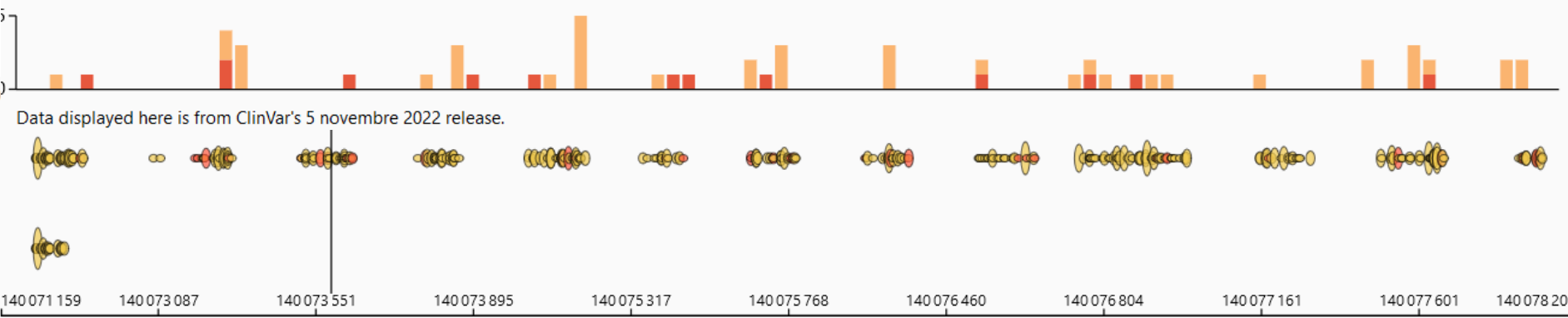
☒ pLoF

☒ Missense / Inframe indel

☐ Synonymous

☐ Other

☐ Only show ClinVar variants that are in gnomAD



☒ pLoF

☒ Missense / Inframe indel

☐ Synonymous

☐ Other

☒ Exomes ☒ SNVs ☐ Filtered variants

☒ Genomes ☒ Indels

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		p.Leu200Val	missense		Pathogenic/Likely p...		1
5-140076896-G-T		p.Val368Leu	missense		Pathogenic/Likely p...		2

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

External Resources

- [dbSNP \(rs767814719\)](#)
- [UCSC](#)
- [ClinGen Allele Registry \(CA3444621\)](#)

Feedback

[Report an issue with this variant](#)

Population Frequencies ⓘ

Population	<u>Allele Count</u>	<u>Allele Number</u>	<u>Number of Homozygotes</u>	<u>Allele Frequency</u>
▸ 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



NHLBI Exome Sequencing Project (ESP)

Exome Variant Server

Variant Results**Coverage Results**

Gene Name: [HARS2](#) (+)

Gene ID: [23438](#) (+)

[Chromosome 5: 140071011 - 140078903](#)

Genes in this region: [HARS\(-\)](#) [HARS2\(+\)](#) [ZMAT2\(+\)](#)

Select Data Set(s)

Check at least one data set below.

Select	Number Variations	Population
☒	51	EuropeanAmerican
☒	45	AfricanAmerican

Display Results

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

EVS[2]

←

evs.gs.washington.edu/EVS/ServletManager?variantType=snp&popID=EuropeanAmerican&popID=Afr

Rechercher

☆

📁

↓

🏠

📧

☰

Gene Name: [HARS2](#) (Gene ID: 23438) (+)

Chromosome 5: 140071011 - 140078903

Genes in this region: [HARS\(-\)](#) [HARS2\(+\)](#) [ZMAT2\(+\)](#)

Population: EuropeanAmerican, AfricanAmerican

GWAS Catalog: [HARS](#) [HARS2](#) [ZMAT2](#)

KEGG Pathway: [HARS](#) [HARS2](#) [ZMAT2](#)

Sanger COSMIC: [HARS](#) [HARS2](#) [ZMAT2](#)

PPI STRING 9.0: [HARS](#) [HARS2](#) [ZMAT2](#)

OMIM: [HARS](#) [HARS2](#) [ZMAT2](#)

Variation Color Code:

splice or nonsense or frameshift

missense

coding-synonymous

coding

utr

codingComplex

Download Options

File Format

Zip Format

download

Add or Remove Columns (Description of Columns)

☒ dbSNP rs ID

☒ Genotype Count

☒ cDNA Size

☐ NCBI 37 Allele

☒ Alleles

☐ MAF (%)

☒ Protein Change

☐ Chimp Allele

☒ EA Allele Count

☒ Sample Read Depth

☒ Conservation (GERP)

☐ Illumina HumanExome Chip

☒ AA Allele Count

☒ Genes

☐ Conservation (phastCons)

☐ GWAS Hits

☒ Allele Count

☒ Gene Accession #

☒ Grantham Score

☐ EA Est. Age (kyrs)

☒ EA Genotype Count

☒ GVS Function

☒ PolyPhen Prediction

☐ AA Est. Age (kyrs)

☒ AA Genotype Count

☒ cDNA Change

☐ Clinical Link

☐ GRCh38 Position

Sort Variants by

Variant Pos

Select Population

All

Select Transcript

Union of Transcripts

If "Select Transcript" above is set to "Union of Transcripts", and if multiple transcripts of a gene are involved in a variant and the function annotations for the variant are the same, only one representative transcript is listed in the table below for the reasons of speed and space. But annotations for each individual transcript are fully listed in the downloaded file if one chooses to download the data.

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

Benign

Probably damaging
damaging



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](#)

HGMD Public site users

Gene symbol	Chromosomal location	Gene name	Mutation total	Log in
HARS2	5q31.3	Histidyl-tRNA synthetase 2, mitochondrial (putative)	2	Log in

If you are already a registered HGMD user, please log in using the button above to access this resource. If you are not registered, please visit our [registration](#) page to gain access. If you have already logged in, it is likely that cookies have not been properly enabled on your system. Please allow [session cookies](#) to be set from hgmd.cf.ac.uk

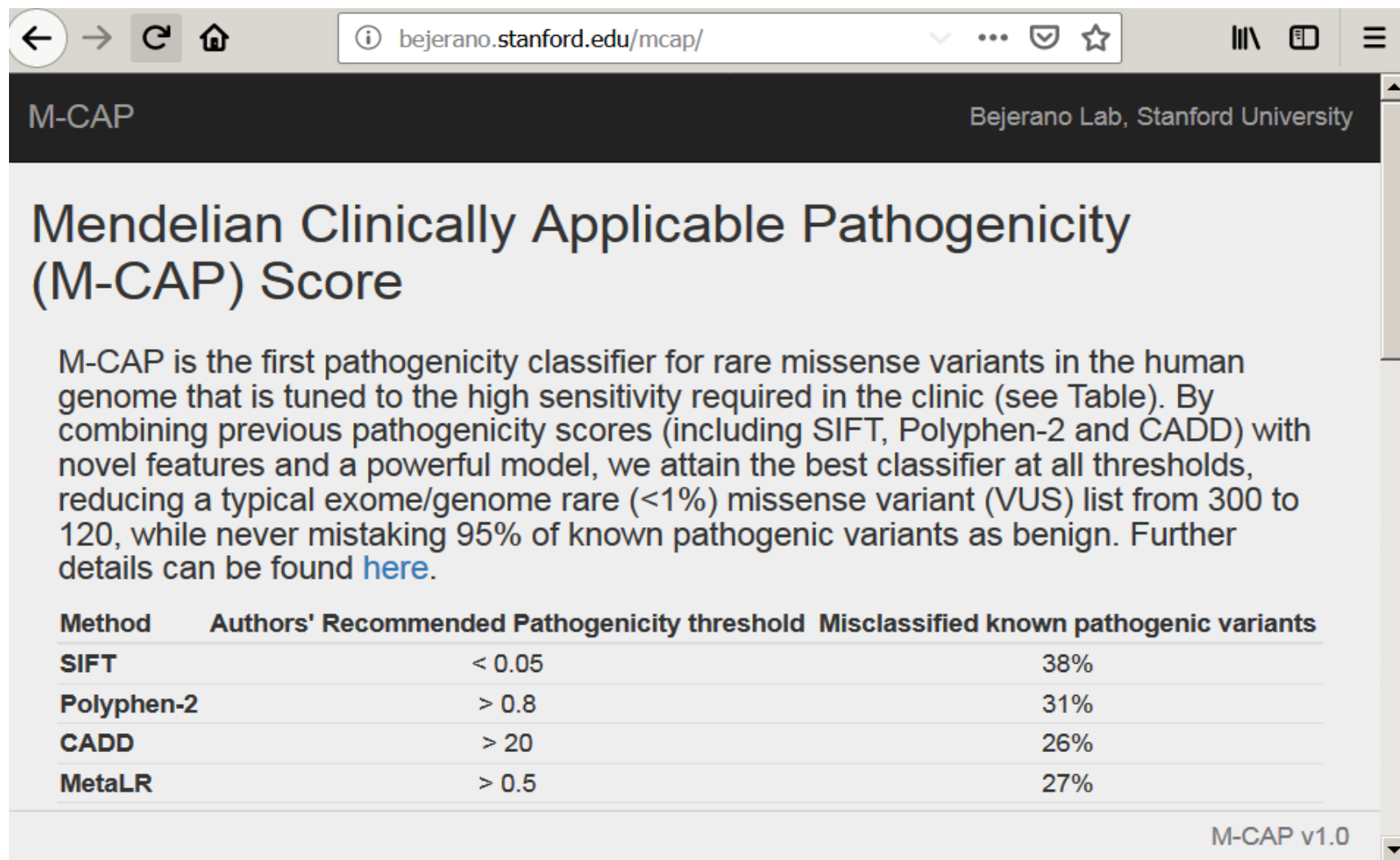
HGMD Professional subscribers

Gene Symbol	Chromosomal location	Gene name	Mutation total	Log in
HARS2	5q31.3	Histidyl-tRNA synthetase 2, mitochondrial (putative)	2	HGMD Professional

If you are already an HGMD Professional subscriber, please log in using the button above to access the resource. If you are not yet a subscriber, please visit [BIOBASE](#) for further information regarding the many benefits of a subscription to [HGMD Professional](#).

Commerciale

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



The screenshot shows a web browser window with the address bar displaying `bejerano.stanford.edu/mcap/`. The page has a dark header with "M-CAP" on the left and "Bejerano Lab, Stanford University" on the right. The main content area has a light gray background and features the title "Mendelian Clinically Applicable Pathogenicity (M-CAP) Score". Below the title is a paragraph explaining that M-CAP is a pathogenicity classifier for rare missense variants, combining previous scores (SIFT, Polyphen-2, CADD) with novel features to reduce the number of variants in a list while maintaining high sensitivity. A table compares M-CAP to other methods, showing its superior performance in reducing misclassified known pathogenic variants. The footer of the page indicates "M-CAP v1.0".

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. Further details can be found [here](#).

Method	Authors' Recommended Pathogenicity threshold	Misclassified known pathogenic variants
SIFT	< 0.05	38%
Polyphen-2	> 0.8	31%
CADD	> 20	26%
MetaLR	> 0.5	27%

M-CAP v1.0

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



The Human Genomics Community

hg38 ▾

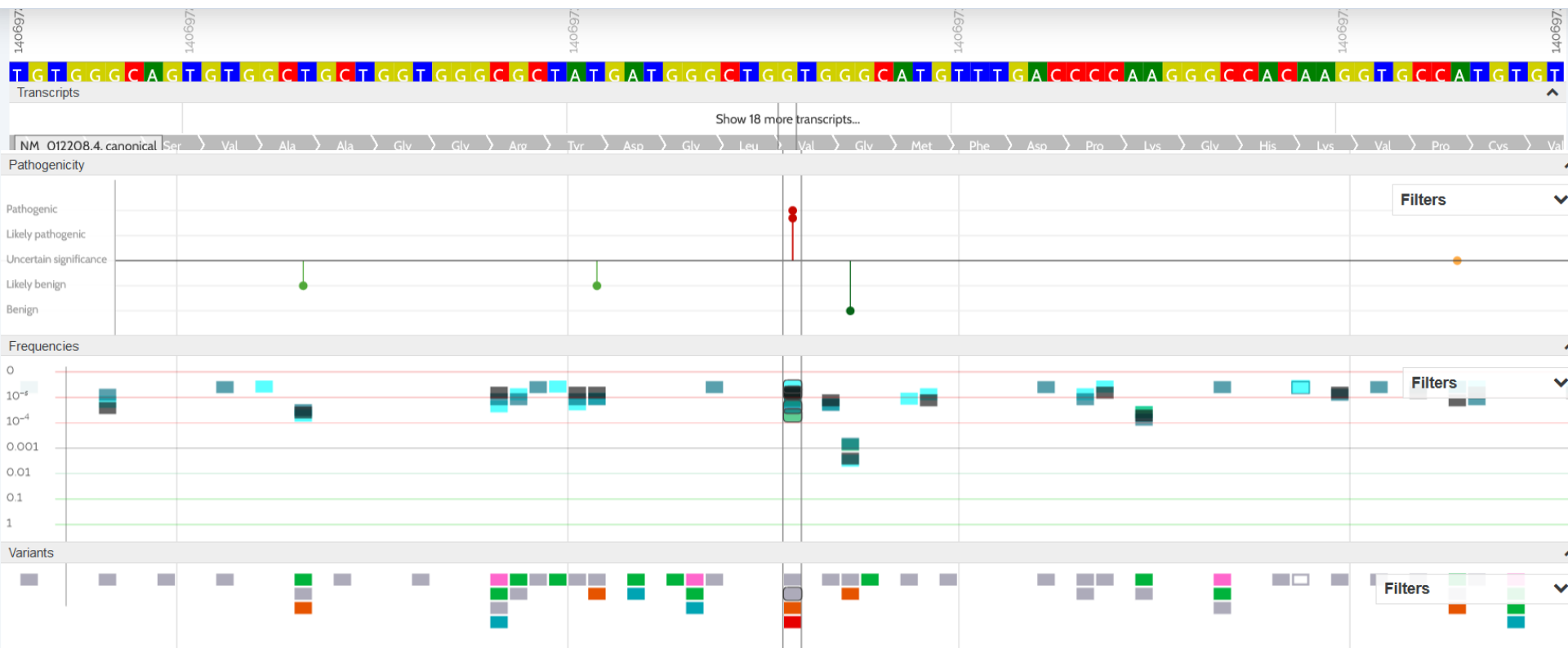
Search

Examples

<https://varsome.com/>

Chromosome	Position	REF Sequence	ALT Sequence	Variant type	Cytoband	HGVS	RS ID	Gene symbol
chr5	140697311	G	A	SNV	5q31.3	HARS2(NM_012208.4):c.1102G>A (p.Val368Met)	rs376177973 dbSNP	HARS2
UCSC genome browser								
Mastermind								
TraP Score								

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]

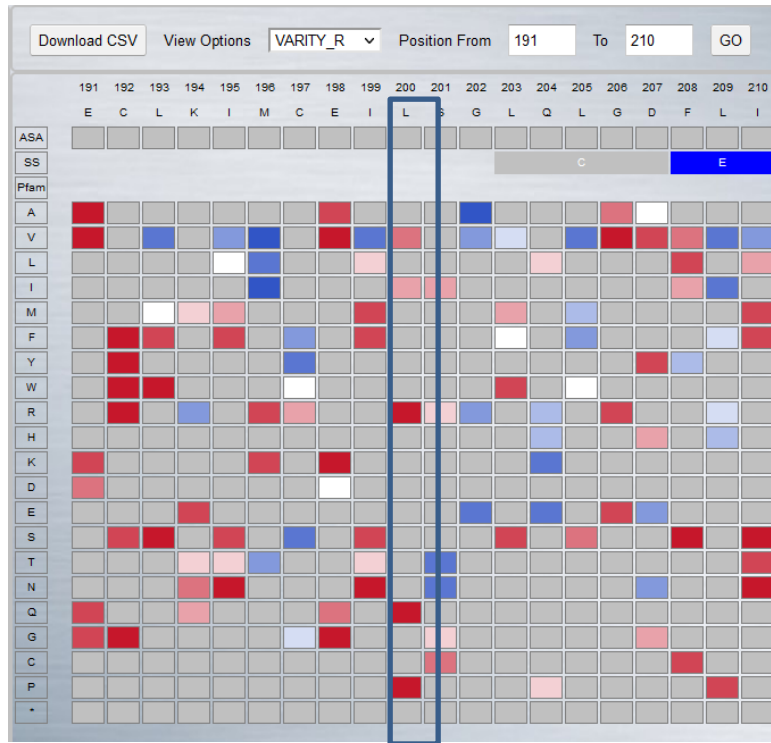
Varity



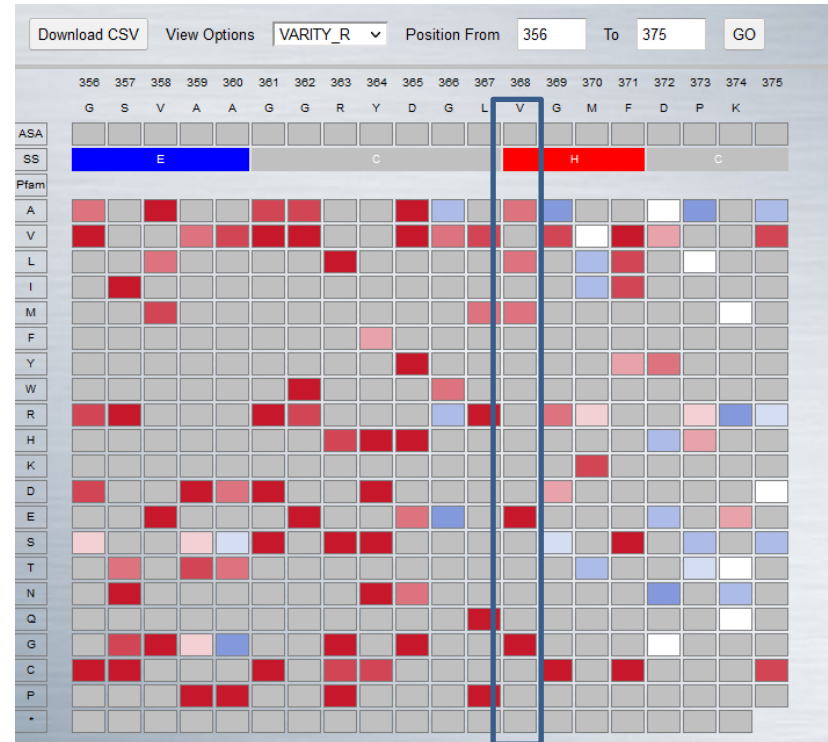
The screenshot shows a web browser window with the address bar displaying `varity.varianteffect.org`. The page title is "VARITY: Improved pathogenicity prediction for rare human missense variants". On the left sidebar, under "View VARITY Effect Maps", there are two search methods: (1) "View maps by Protein name" with a text input containing "HARS2" and a "View" button, and (2) "View maps by Uniprot ID" with an empty text input and a "View" button. The top right of the main content area contains a navigation bar with "Download CSV", "View Options" (with a dropdown menu showing "VARITY_R"), "Position From" (empty input), "To" (empty input), and a "GO" button. A small icon of a hand cursor is also present on the far right of this bar.

<http://varity.varianteffect.org/>

Varity [2]



Leu200Val



Val368Leu

Un clic sur une position donne un tableau complet

Variety [3]

Un clic sur une position donne un tableau complet

Position: 368 [V $\Rightarrow$ L]

VARIETY_R: 0.818
VARIETY_ER: 0.747

VARIETY_R_LOO: 0.796
VARIETY_ER_LOO: 0.780

-log10(MAF): 5.099
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 Isoelectric 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 volume 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 Covalent 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 Positive Charged	0.000	-0.00%			
Accessible Surface Area		-10.49%	Δ PKa	-0.040	0.60%	Δ Negative Charged	0.000	-0.00%			

VARIETY_ER Feature Contribution

VARIETY_R Feature Contribution

CLOSE

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, PhD¹, Nazneen Aziz, PhD^{2,16}, Sherri Bale, PhD³, David Bick, MD⁴, Soma Das, PhD⁵, Julie Gastier-Foster, PhD^{6,7,8}, Wayne W. Grody, MD, PhD^{9,10,11}, Madhuri Hegde, PhD¹², Elaine Lyon, PhD¹³, Elaine Spector, PhD¹⁴, Karl Voelkerding, MD¹³ and Heidi L. Rehm, PhD¹⁵; on behalf of the ACMG Laboratory Quality Assurance Committee

Official Journal of the American College of Medical Genetics and Genomics

ORIGINAL RESEARCH ARTICLE

Genetics
in Medicine

Open

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

Keith Nykamp, PhD¹, Michael Anderson, PhD¹, Martin Powers, MD¹, John Garcia, PhD¹, Blanca Herrera, PhD¹, Yuan-Yuan Ho, PhD¹, Yuya Kobayashi, PhD¹, Nila Patil, PhD¹, Janita Thusberg, PhD¹, Marjorie Westbrook, PhD¹, The Invitae Clinical Genomics Group² and Scott Topper, PhD, FACMG¹

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	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.
Leiden Open Variation Database http://www.lovd.nl	
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	Medically relevant gene reference sequence resource.
Locus Reference Genomic (LRG) http://www.lrg-sequence.org	
MitoMap http://www.mitomap.org/MITOMAP/	Revised Cambridge reference sequence for human mitochondrial DNA.
HumanMitoSeq	

Variant classification according to the ACMG Guidelines

The results from the variant
classification in different
Norwegian laboratories.

Variant classification

Mari Ann Kulseth
AMG - OUS

Class 5: pathogenic

Class 4: likely pathogenic

Class 3: VUS – variant of uncertain significance

Class 2: likely benign

Class 1: benign



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 1^{ère} 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-rare.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
KGSeq	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

ANNOVAR [1] Documentation des annotations ANNOVAR dans les différentes bases

Overview

Summary of databases

1000 Genomes Project (2015 Aug) annotations

1000 Genomes Project (2014 Oct) annotations

1000 Genomes Project (2012 April) annotations (obsolete!)

dbSNP annotations

avSNP annotations

LJB* (dbNSFP) non-synonymous variants annotation

ESP (exome sequencing project) annotations

ExAC annotations

gnomAD allele frequency

GERP++ annotations

CG (complete genomics) frequency annotations

https://annovar.openbioinformatics.org/en/latest/user-guide/filter/#cg-complete-genomics-frequency-annotations

ANNOVAR Documentation ANNOVAR User Guide ▾ Misc ▾ Articles ▾ Search ← Previous Next → Edit on GitHub

Overview

An important and probably highly desirable feature is that ANNOVAR can help identify subsets of variants based on comparison to other variant databases, for example, variants annotated in dbSNP or variants annotated in 1000 Genome Project. The exact variant, with same start and end positions, and with same observed alleles, will be identified.

These functionalities mentioned above can be performed using the `--filter` operation in ANNOVAR. The major difference between `--filter` and `--regionanno` above is that that `--filter` operation works on mutations (nucleotide changes), but `--regionanno` operation works on chromosome locations. For example, `--region` compare variants with things like chr1:1000-1000, but `--filter` compare variants with things like A->G change at the position chr1:1000-1000.

Summary of databases

Due to the increased number of databases that are available at ANNOVAR, some users are not sure where to start. Here we give a brief summary of some of the mostly commonly used databases.

For frequency of variants in whole-genome data:

- 1000g2015aug: latest 1000 Genomes Project dataset with allele frequencies in six populations including ALL, AFR (African), AMR (Admixed American), EAS (East Asian), EUR (European), SAS (South Asian). These are whole-genome variants.
- kaviar_20150923: latest Kaviar database with 170 million variants from 13K genomes and 64K exomes.
- hrcr1: latest Haplotype Reference Consortium database with 40 million variants from 32K samples in haplotype reference consortium
- cg69: allele frequency in 69 human subjects sequenced by Complete Genomics. useful to exclude platform specific variants.
- gnomad_genome: allele frequency in gnomAD database whole genome sequence data on multiple

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

- *_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/>

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-2

prediction of functional effects of human nsSNPs

[Home](#)[About](#)[Help](#)[Downloads](#)[Batch query](#)[WHESS.db](#)

PolyPhen-2 (Polymorphism Phenotyping v2) 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.

21-Jun-2021: Server has been migrated to new hardware. Note, all queries were terminated and user sessions data discarded in the process, hence you will need to resubmit your query if affected. We apologize for the inconvenience caused.

Query Data

Protein or SNP identifier

Protein sequence
in FASTA format

```
>sp|P49590|SYHM_HUMAN Histidine--tRNA ligase, mitochondrial OS=Homo  
sapiens OX=9606 GN=HARS2 PE=1 SV=1  
MPLLGILLPRRAWASLLSOLLRPPCASCTGAVRCOSOVAEAVLTSOLKAHOEKPNFIKTP  
KGTRDLSPOHMMVREKILDLVISCFKRHGAKGMDTPAFELKETLTEKYGEDSGLMYDLKD  
QGGELLSLRYDLTVPFARYLAMNKVKMKRYHVGKVVRRSPITVOGRYREFCQCFDIA  
GQFDPMTDAECLKTMCELSGLQLGDFLTKNDRRTVDGMFAVCGVPESKFRATCSST  
KLDKMAWKDVRHEMMVKKGLAPEVADRTGDYVOCHGGVSLVEOMFODPRLSONKOALEGL  
GDLKLLFEYLTLFGIADKISFDLSLARGLDYYTGVITYEAVLLOTPTOAGEEPLNVGSAVA  
GGRYDGLVGMFDPKGHKVPCVGLSIGVERIFYIVEORMKTKGEKVRTTETOVFVATPOKN  
FLOERLKLIAELWDSGIKAEMLYKNPKLLTOLHYCESTGIPLVVIIGEOELKEGVIKIR  
SVASREEVATKRENFVAETOKRSES
```

Position

200

Substitution

AA₁ A R N D C E Q G H I L K M F P S T W Y V
AA₂ A R N D C E Q G H I L K M F P S T W Y V

Query description

[Display advanced query options](#)

PolyPhen-2 (Polymorphism Phenotyping v2) 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.

HumVar

This mutation is predicted to be **PROBABLY DAMAGING** with a score of 0.998 (sensitivity: 0.18; specificity: 0.98)



A horizontal bar chart representing the HumVar score. The bar is colored with a gradient from green on the left to red on the right, with a yellow center. The x-axis is labeled from 0.00 to 1.00 in increments of 0.20. A vertical black line is positioned at the far right end of the bar, corresponding to a score of 0.998.



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 used to predict **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
MPLLGLLPRAWASLLSQLLRPPCASCTGAVRCQSQVAEAVLTSQLKAHQEKNFIKTP
KGTRDLSPQHMMVREKILDLVISCFKRHGAKGMDTPAFELKETLTEKYGEDSGLMYDLKD
QGGELLSRLYDLTVPFARYLAMNKVKMKRYHVGKVVRRRESPTIVQGRYREFCQCDFDIA
GQFDPMIPDAECLKIMCEILSGLQLGDFLIKVNDRRIVDGMFAVCGVPESKFRAICSSID
KLDKMAWKDVRHEMVVKKGLAPEVADRIGDYVQCHGGVSLVEQMFQDPRLSQNKQALEGL
GDLKLLFEYLTFLGIADKISFDLSLARGLDYYTGVIEAVLLQTPTQAGEEPLNVGSVAA
GGRYDGLVGMFDPKGHKVPCVGLSIGVERIFYIVEQRMKTGKEKVRTTETQVFVATPQKN
FLQERLKLIAELWDSGIKAEMLYKNNPKLLTQLHYCESTGIPLVVIIGEQLKEGVKIR
SVASREEVAIKRENFVAEIQRLSES
```

Enter the substitutions of interest [\[format\]](#):

V368L

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														
		mwifvl	366G	1.00		keas	dGN														
		ywvtr	367L	1.00		L															
		whgdn	368V	1.00		TIV															
		wcfmyihvlp	369G	1.00		rtqsd	ANKEG														
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]



cBio@MSKCC

[papers](#) | [about](#) | [privacy](#) | [changes](#) | [how it works](#) | [web API](#)

This server predicts the functional impact of amino-acid substitutions in proteins, such as mutations discovered in cancer or missense polymorphisms. The functional impact is assessed based on evolutionary conservation of the affected amino acid in protein homologs. The method has been validated on a large set (60k) of disease associated (OMIM) and polymorphic variants. To explore the functional impact of missense mutations found in The Cancer Genome Atlas please use [cBioPortal for Cancer Genomics](#).

Dec 31, 2015

Release **3** is out!

All scores for human Swiss-Prot sequences are available:

[MA_scores_rel3_swissprot_full.tar.gz](#) (or split to parts of up to 1M records) (either is 300mb)

Also, all human hg19 reference genome nucleic acid variations mappable to Swiss-Prot sequences:

[MA_scores_rel3_hg19_full.tar.gz](#) (or split to parts of up to 1M records) (either is 843mb)

Release 2 is available at <http://mutationassessor.org/r2/>

Enter your mutations

read about [input format](#) or use [example](#)

SYHM_HUMAN L200V

Configure output

- ☒ mapping issue ☐ variant based on reference genome
- ☒ 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)

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

- ☒ mapping issue ☐ variant based on reference genome
- ☒ 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
- ☒ 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
KGTRDLSPQHMOVREKILDLVISCFKRHGAKGMDTPAFELKETLTEKYGEDSGLMYDLKD
QGGELLSLRYDLTVPFARYLAMNKVKKMKRYHVGKVWRRESPTIVQGRYREFCQCDFDIA
GQFDPMIPDAECLKIMCEILSGLQLGDFLIKVNDRRIVDGMFAVCGVPESKFRAICSSID
KLDKMAWKDVRHEMVVKKGLAPEVADRIGDYVQCHGGVSLVEQMFQDPRLSQNKQALEGL
GDLKLLFEYLTFLFGIADKISFDLSLARGLDYYTGVIYEAVLLQTPTQAGEEPLNVGSVAA
GGRYDGLVGMFDPKGHKVPCVGLSIGVERIFYIVEQRMKTKEKVRTTETQVFVATPQKN
FLQERLKLIAELWDSGIKAEMLYKNNPKLLTQLHYCESTGIPLVVIIGEQLKEGVKIR
SVASREEVAIKRENFVAEIQKRLSES

L 200 V

```

sel=0
178      Seq:1 Pos:206|200 [sp|P49590|SYHM HUMAN      230
sp|P49590|SYHM HUMAN  FCQCDFDIAGQFDPMIPDAECLKIMCEILSGLQLGDFLIKVNDRRIVDGMFAV
sp|Q5R5E5|SYHM_PONAB  FCQCDFDIAGQFDPMIPDAECLKIMCEILSGLQLGDFLIKVNDRRIVDGMFAV
sp|P49590-2|SYHM_HUM  FCQCDFDIAGQFDPMIPDAECLKIMCEILSGLQLGDFLIKVNDRRIVDGMFAV
sp|A5D7V9|SYHM_BOVIN  FYQCDFDIAGQFDPMIPDAECLKIMCEILSGLHLGDFLIKVSDRRILDGIFAV
sp|Q99KK9|SYHM_MOUSE  FCQCDFDIAGEFDPMPDAECLRLMCEILSGLQLGDFLIKVNDRRVVDGIFAV
sp|Q2KI84|SYHC_BOVIN  FYQCDFDIAGQFDPMIPDAECLKIMCEILSSLQIGDFLVKVNDRRILDGMFAI
sp|Q61035|SYHC_MOUSE  FYQCDFDIAGQFDPMIPDAECLKIMCEILSSLQIGNFLVKVNDRRILDGMFAV
sp|P12081|SYHC_HUMAN  FYQCDFDIAGNFDPMIPDAECLKIMCEILSSLQIGDFLVKVNDRRILDGMFAI
sp|Q5R4R2|SYHC_PONAB  FYQCDFDIAGNFDPMIPDAECLKIMCEILSSLQIGDFLVKVNDRRILDGMFAI
sp|P12081-4|SYHC_HUM  ----DFDIAGNFDPMIPDAECLKIMCEILSSLQIGDFLVKVNDRRILDGMFAI
sp|P07178|SYHC_MESAU  SITVDFDIAGQFDPMIPDAECLKIMCEILSSLQIGKFLVKVNDRRILDGMFAV
sp|P12081-2|SYHC_HUM  FYQCDFDIAGNFDPMIPDAECLKIMCEILSSLQIGDFLVKVNDRRILDGMFAI
sp|P70076|SYHC_TAKRU  FYQCDFDIAGQYDAMI PDAECLKIVHEILSELDLGDFRIKVNDRRILDGMFAV
sp|P12081-3|SYHC_HUM  ----DFDIAGNFDPMIPDAECLKIMCEILSSLQIGDFLVKVNDRRILDGMFAI

```

V 368 L

```

sel=0
363      Seq:1 Pos:384|368 [sp|P49590|SYHM HUMAN      415
sp|P49590|SYHM HUMAN  QAGEEPLNVGSVAAGGRYDGLVGMFDPKGHKVPCVGLSIGVERIFYIVEQRMK
sp|Q5R5E5|SYHM_PONAB  QAGEEPLNVGSVAAGGRYDGLVGMFDPKGHKVPCVGLSIGVERIFYIVEQRMK
sp|P49590-2|SYHM_HUM  QAGEEPLNVGSVAAGGRYDGLVGMFDPKGHKVPCVGLSIGVERIFYIVEQRMK
sp|A5D7V9|SYHM_BOVIN  HAEFEPLNMGSAAGGRYDGLVGMFDPRGHKVPCVGLSIGVERIFSIVEQRIK
sp|Q99KK9|SYHM_MOUSE  QAGKETLSVGSVAAGGRYDNLVAQFDPKGHHVPCVGLSIGVERIFYLVEQKMK
sp|Q2KI84|SYHC_BOVIN  RAGEEPLGVGSVAAGGRYDGLVGMFDPKGRKVPCVGLSIGVERIFSIVEQRLE
sp|Q61035|SYHC_MOUSE  QAGEEPLGVGSIAAGGRYDGLVGMFDPKGRKVPCVGLSIGVERIFSIVEQRLE
sp|P12081|SYHC_HUMAN  QAGEEPLGVGSVAAGGRYDGLVGMFDPKGRKVPCVGLSIGVERIFSIVEQRLE
sp|Q5R4R2|SYHC_PONAB  QAGEEPLGVGSVAAGGRYDGLVGMFDPKGRKVPCVGLSIGVERIFSIVEQRLE
sp|P12081-4|SYHC_HUM  QAGEEPLGVGSVAAGGRYDGLVGMFDPKGRKVPCVGLSIGVERIFSIVEQRLE
sp|P07178|SYHC_MESAU  GAGEEPWC-GQCGCWRRYDGLVGMFDPKGRKVPCVGLSIGVERIFSIVEQRLE
sp|P12081-2|SYHC_HUM  QAGEEPLGVGSVAAGGRYDGLVGMFDPKGRKVPCVGLSIGVERIFSIVEQRLE
sp|P70076|SYHC_TAKRU  APTEECVTVGSVAGGRYDGLVGMFDPKGRKVPCVGVSIGIERIFSIMEQKAE
sp|P12081-3|SYHC_HUM  QAGEEPLGVGSVAAGGRYDGLVGMFDPKGRKVPCVGLSIGVERIFSIVEQRLE

```

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 enSeabl, et de gnomAD)