Generade

Applied Genomics for Life

DNA variants - the big databases

Variant effect prediction course

Avans Hogeschool Breda

Peter Taschner, Lector Genome-based Health

Hogeschool Leiden

taschner@generade.nl

Lay-out

- Warnings and Tips
- Why use genetics databases?
- Population databases
- General databases
- Cancer databases
- Storing search information

Warning

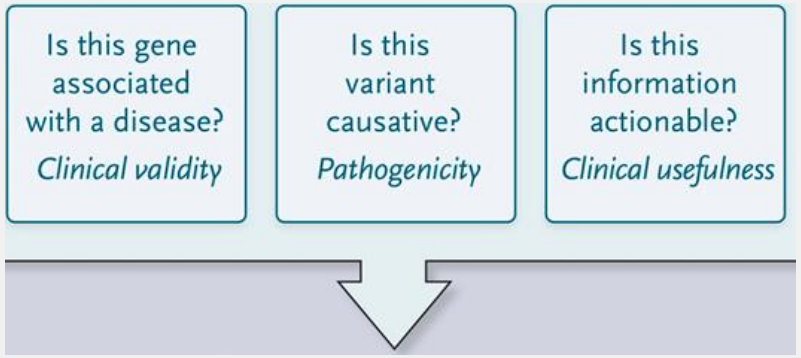
- **Use common sense when searching databases:**
  - Databases do not replace lack of background knowledge
  - Most databases have not been developed and certified for diagnostic use
  - Information can be incorrect due to automatic updates or errors on data-entry
  - Search options may always return results
  - Clear manuals without errors are rare
- **nonsense in = nonsense out**

Tips

- **Use common sense:**
  - Are search results consistent with expectations?
    - Same reference sequences, genome builds?
  - Consider the purpose of the database and its options
    - YouTube movies
  - Website functionality may work in specific browsers only
  - Trouble-shooting: See Help or FAQ (Frequently Asked Questions) files

Why use genetic databases?

NGS variant data interpretation



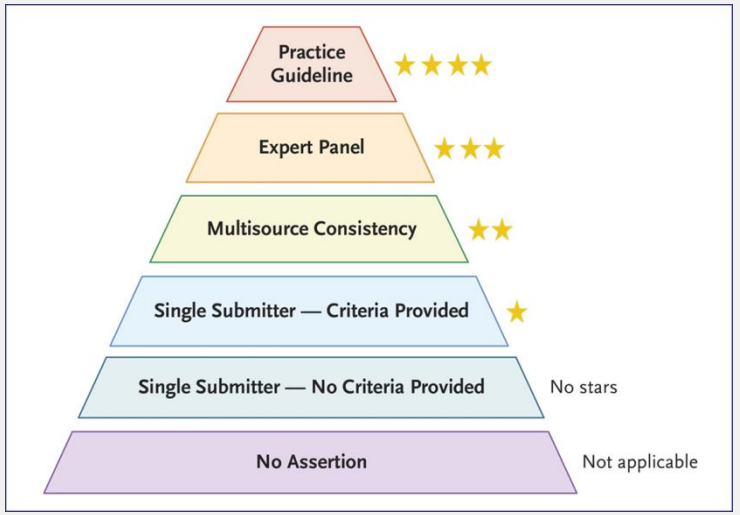
Clinical report

Relevant information from databases:

1) Gene associated with patient phenotype: Y/N
2) Variant causes disease phenotype: Y/N
3) Risk assessment
4) Which treatment options exist: None/A/B

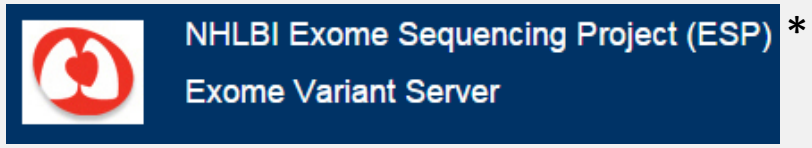
Genetic databases

- Many different variant databases
  - Different types of information
  - Varying quality
- Variant annotation
  - Functional effect prediction: difficult
- Strategy for identification of variants
  - Be careful when excluding variants in dbSNP and frequency databases
  - dbSNP: now contains somatic and disease-causing variants
  - Frequency databases: non-penetrant disease-causing variants
- Discriminate between somatic and germline variants
  - Some combinations of germline variants are lethal, but may occur in specific (cancer) tissues

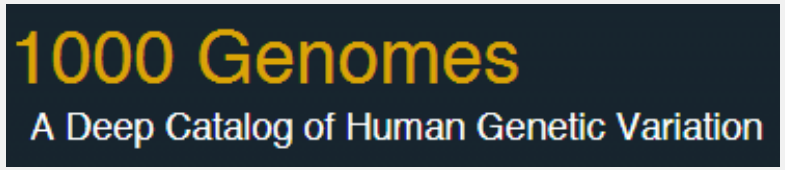


Rehm HL et al. N Engl J Med 2015;372:2235-2242

Genetic variant databases



*



*

*

*

Germline

Somatic

Clinical information:
Little

Little

Little

Depends on curator

Little

Sometimes

* integrated in Ensembl

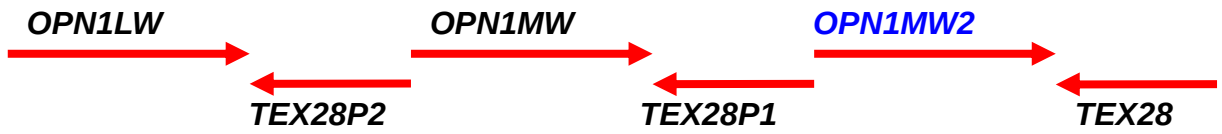
Which reference sequence is used?

Differences between genome builds – hg19

- *OPN1MW* Opsin 1, medium-wave sensitive
  - Color blindness
  - Chromosome X CNV

region 153,409,725-153,523,438 (GRCh37)

Different genomic coordinates
Different (copy) number of genes

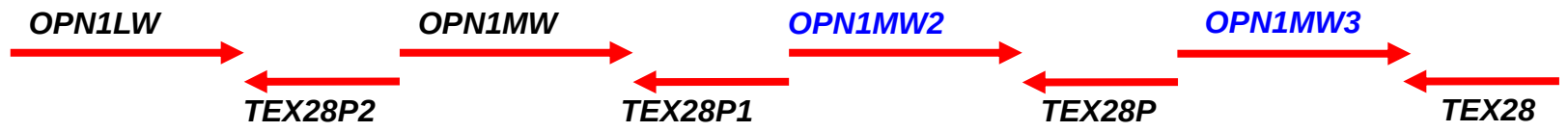


			Annotation	Build
NCBI map viewer	***OPN1MW*** **NM_000513.2**	***OPN1MW2*** **NM_001048181.2**	**Automatic**	**GRCh37.p13**
UCSC	***OPN1MW*** **NM_000513.2** ***OPN1MW2*** **NM_001048181.2**	***OPN1MW*** **NM_00513.2** ***OPN1MW2*** **NM_001048181.2**	**Automatic (Track dependent)**	**GRCh37**
Ensembl 73	***OPN1MW*** **3 transcripts**	***OPN1MW2*** **3 transcripts**	**Automatic Manual (Vega)**	**GRCh37.p12**

Differences between genome builds – hg38

- *OPN1MW* Opsin 1, medium-wave sensitive
  - Color blindness
  - Chromosome X CNV

region 154,144,224-154,295,356 (GRCh38)



				Annotation	Build
NCBI map viewer	*OPN1MW* NM_000513.2	*OPN1MW2* NM_001048181.2	*OPN1MW3* NM_001330067.2	Automatic	GRCh38.p7
UCSC	*OPN1MW* NM_001048181.2	*OPN1MW2* NM_001048181.2	*OPN1MW3* NM_00513.2	Automatic (Track dependent)	GRCh38
Ensembl 95	*OPN1MW* 3 transcripts	*OPN1MW2* 3 transcripts	*OPN1MW3* 3 transcripts	Automatic Manual (Havanna)	GRCh38.p12

Reference genome: not all genes

Generade
Applied Genomics for Life

TAS2R45 taste 2 receptor member 45 [*Homo sapiens* (human)]

Gene ID: 259291, updated on 29-Mar-2018

Summary

Official Symbol	TAS2R45 provided by HGNC
Official Full Name	taste 2 receptor member 45 provided by HGNC
Primary source	HGNC:HGNC:18876
See related	MIM:613967
Gene type	protein coding
RefSeq status	VALIDATED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	GPR59; T2R45; ZG24P
Annotation information	Annotation category: only annotated on alternate loci in reference assembly

- Sequence missing from reference genome:
- *TAS2R45* gene (Craig Venter)
- Now: CNV (Alt_Ref_Locus)

Genomic context

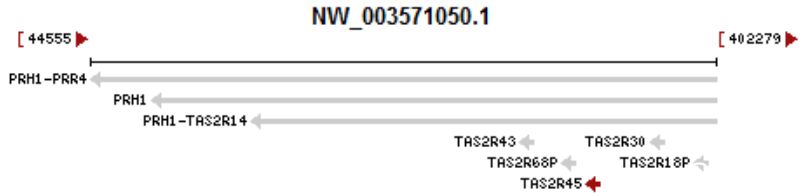
Location: chromosome: 12

See TAS2R45 in Genome Data Viewer Map Viewer

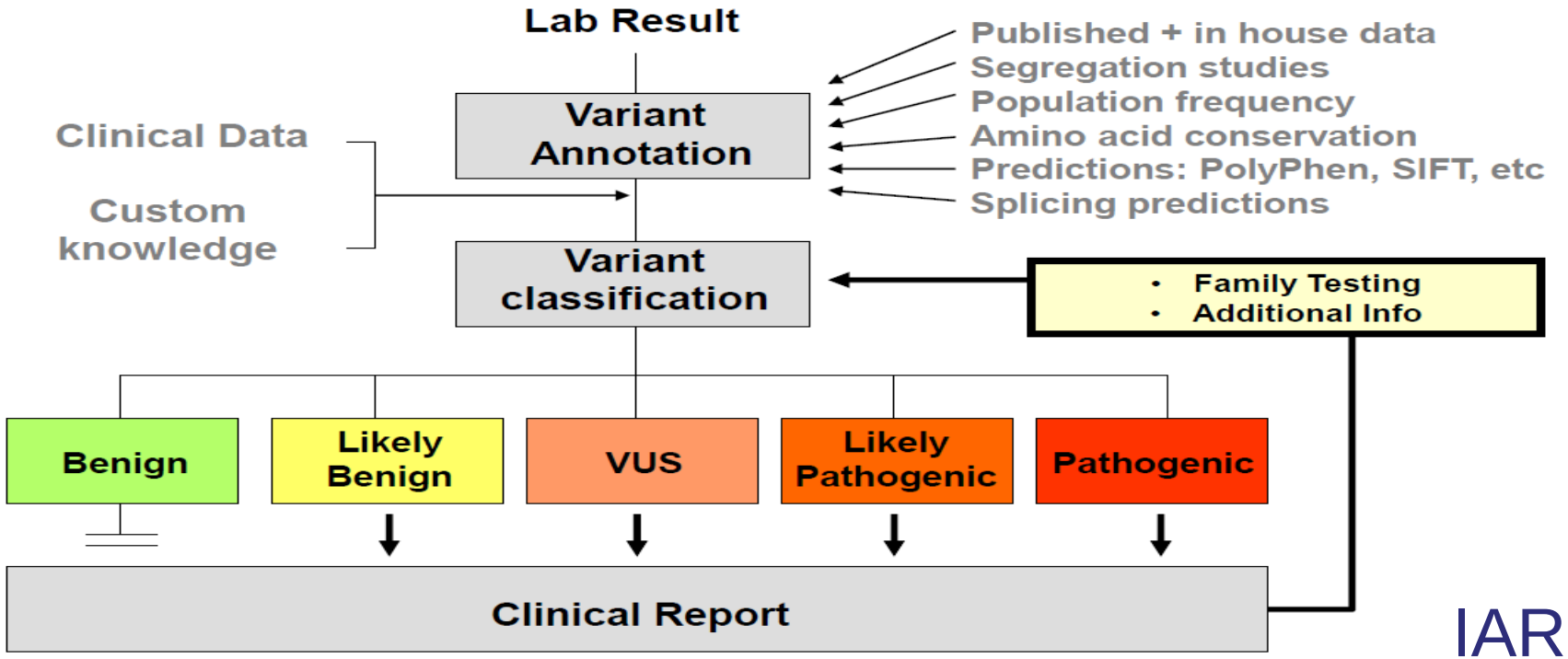
Exon count: 1

Annotation release	Status	Assembly	Chr	Location
109	current	GRCh38.p12 (GCF_000001405.38)	12 (ALT_REF_LOCI_1)	NW_003571050.1 (327525..328424, complement)
105	previous assembly	GRCh37.p13 (GCF_000001405.25)	12 Unlocalized Scaffold (PATCHES)	NW_003571050.1 (327525..328424, complement)

Separate contig



Variant annotation and classification



Courtesy: Brigit Funke

Share genotype and phenotype information with gene variant databases

IARC classification

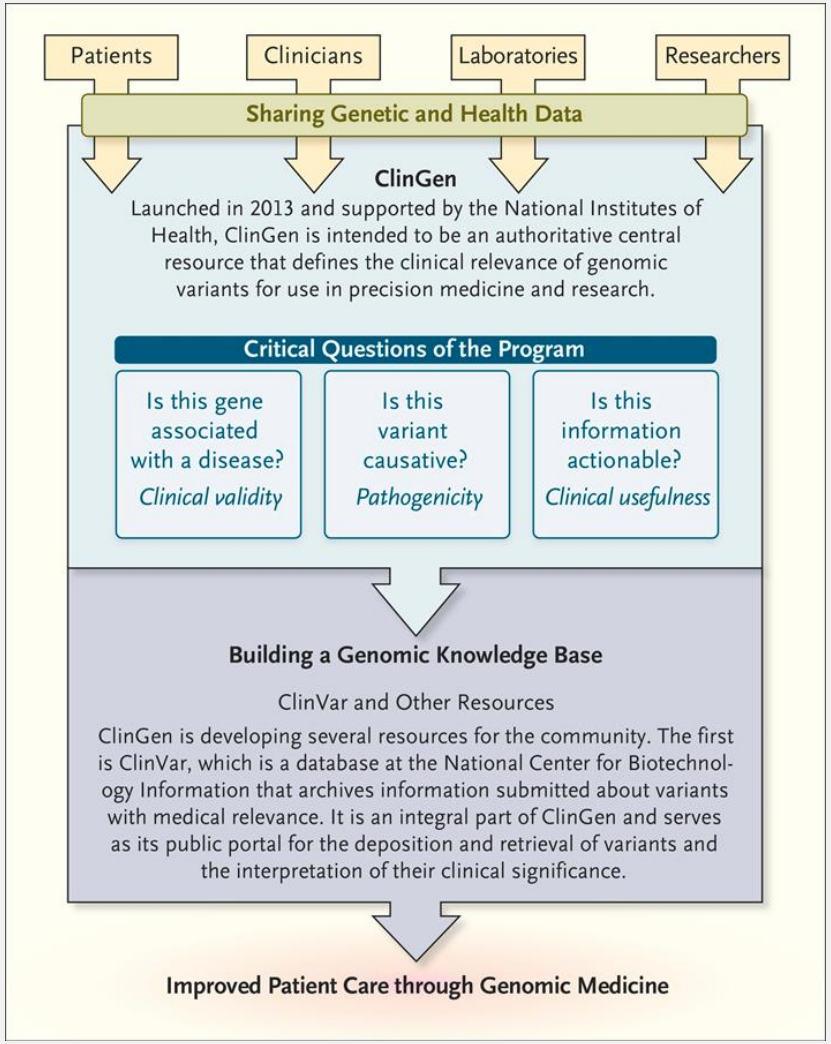
Table 3

Proposed Classification System for Sequence Variants Identified by Genetic Testing

Class	Description	Probability of being Pathogenic
5	Definitely Pathogenic	>0.99
4	Likely Pathogenic	0.95–0.99
3	Uncertain	0.05–0.949
2	Likely Not Pathogenic or of Little Clinical Significance	0.001–0.049
1	Not Pathogenic or of No Clinical Significance	<0.001

Plon et al 2008. Hum Mutat. 29: 1282–1291.

Sharing genetic data between databases



Rehm HL *et al*. N Engl J Med 372:2235-2242 (2015)

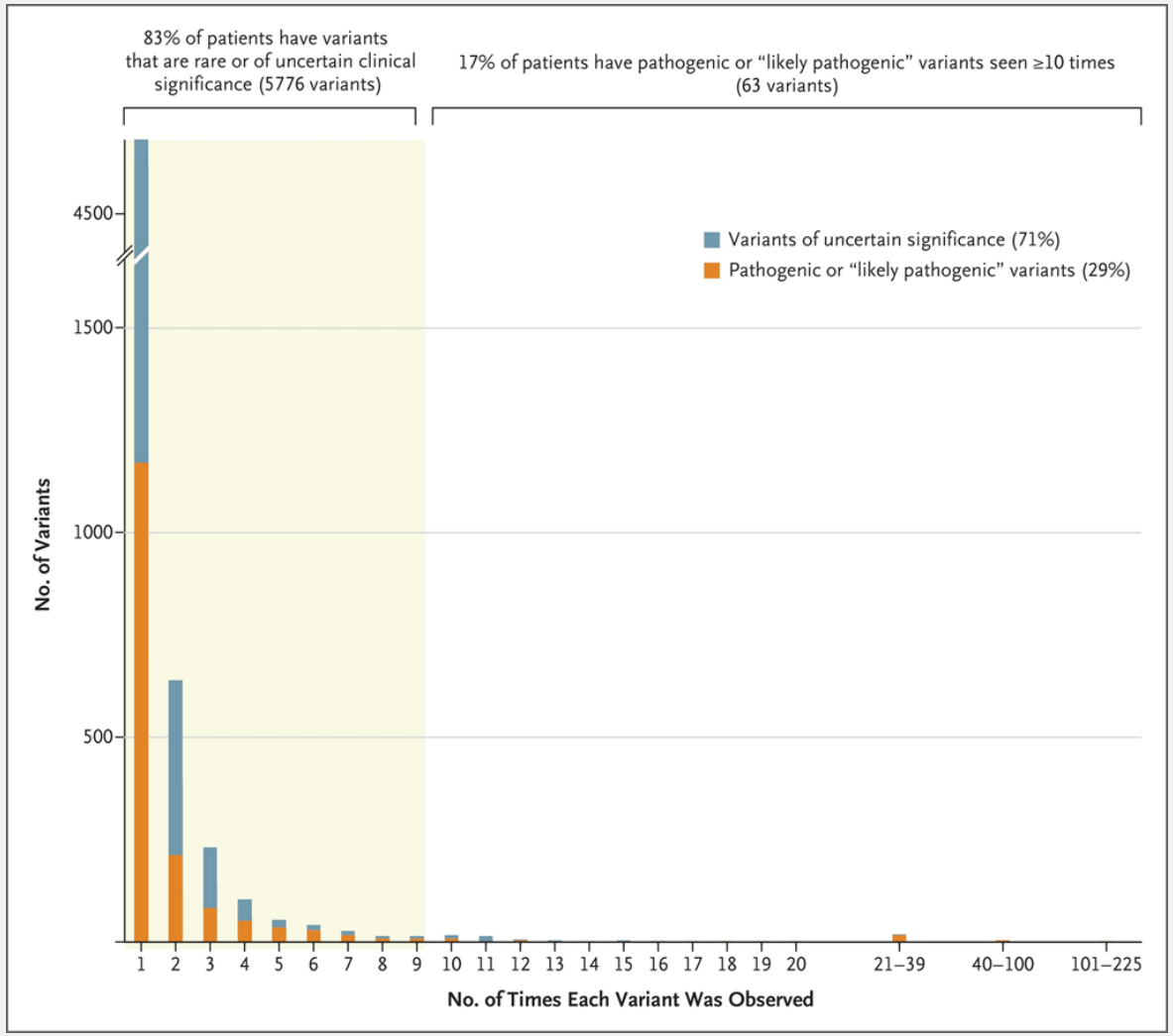
Data for evaluation of genetic variants

Generade
Applied Genomics for Life

- Variant information at DNA-, RNA- and protein level per gene
- Variant allele (homozygote, maternal, paternal)
- Co-segregation with disease in family
- Detection techniques and material tested
- Results functional tests
- Variants co-occurring or observed together:
  - Which combination of variants causes disease?
- Case-control:
  - Numbers of affected and healthy carriers, healthy controls
- Frequency of variants in the population (Minor allele frequency: MAF)
- Variant classification: clinical relevance?

Preferred source:
Reliable databases

ClinGen: number of observations per variant



Rehm HL et al. N Engl J Med 2015;372:2235-2242

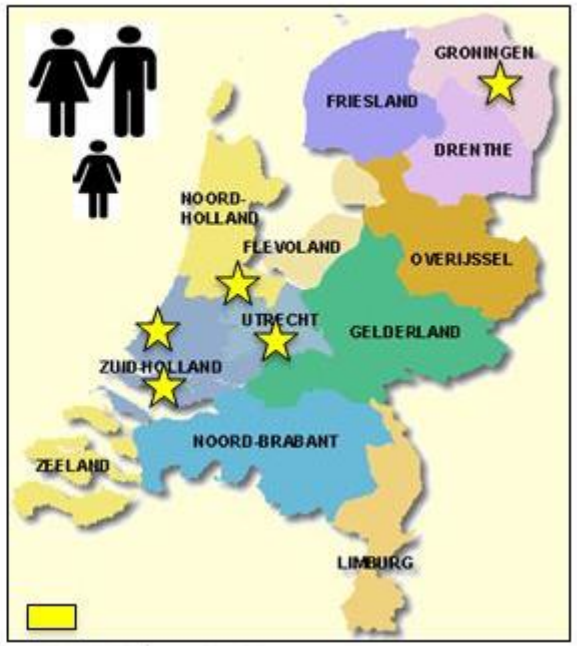
Class 3 – 5 variants in 15.000 patients with hereditary disorder

Most disease-causing variants are rarely observed

GoNL: variants in the Dutch population

Generade
Applied Genomics for Life

- Unique family-based design: 250 trios
  - 230 x 2 parents - 1 offspring
  - 10 x 2 parents - 2 offspring
  - 10 x 2 parents - 1 MZ twin offspring



- Specifications:
  - Families equally distributed over the Dutch provinces
  - Genomic DNA, paired-end sequencing on HiSeq2000, 12x coverage
  - Trios allow phase information; accurate haplotypes
  - Other results: Structural variation, detection *de novo* variants

Purpose filtering:
Removal/ annotation of frequent variants in Dutch patients

No phenotype information

http://www.nlgenome.com/
http://www.bbmriwiki.nl

BBMRI·NL
GoNL

genome Aggregation Database (gnomAD)– *CREBBP*

Generade
Applied Genomics for Life

CREBBP CREB binding protein

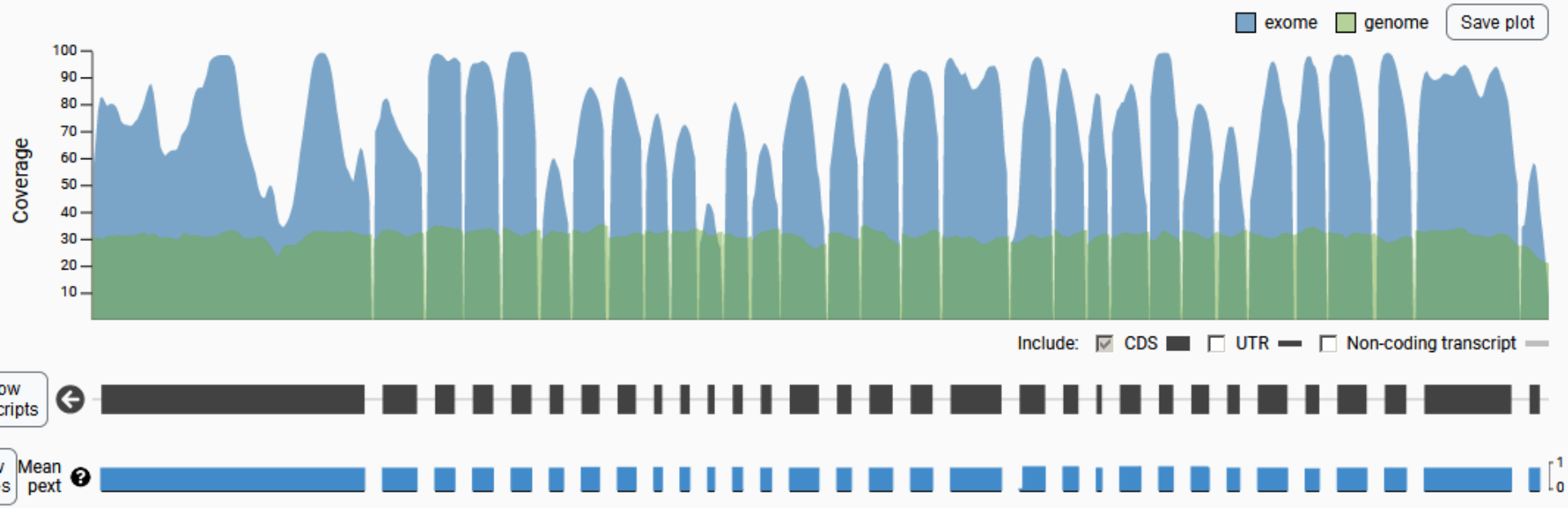
Dataset gnomAD v2.1.1 gnomAD SVs

http://gnomad.broadinstitute.org

Ensembl gene ID	ENSG00000005339
Ensembl transcript ID	ENST00000262367 (canonical)
UCSC Browser	16:3775056-3930728
GeneCards	CREBBP
OMIM	600140

Gene Constraint

Category	Exp. no. SNVs	Obs. no. SNVs	Constraint metrics	
Synonymous	567.3	714	Z = -4.84	o/e = 1.26 (1.18 - 1.34)
Missense	1418.2	1005	Z = 3.9	o/e = 0.71 (0.67 - 0.75)
LoF	118.3	3	pLI = 1	o/e = 0.03 (0.01 - 0.07)



genome Aggregation Database (gnomAD)

Generade
Applied Genomics for Life

Population	gnomAD	
	exomes	genomes
African/African American	8,128	4,359
Latino	17,296	424
Ashkenazi Jewish	5,040	145
East Asian	9,197	780
Finnish	10,824	1,738
Non-Finnish European	56,885	7,718
South Asian	15,308	*
Other	3,070	544
Female	57,787	6,967
Male	67,961	8,741
Total	**125,748**	**15,708**

http://gnomad.broadinstitute.org

Ensembl Custom Data VEP

Generade Applied Genomics for Life

http://www.ensembl.org/Tools/VEP VEP-Video https://www.youtube.com/watch?v=rSIG_OVzyLU

Variant Effect Predictor

New job Clear form

Species: Human (Homo sapiens)

Assembly: GRCh38.p12 *(If you are looking for VEP for Human GRCh37, please go to GRCh37 website.)*

Name for this job (optional):

Input data:

Either paste data:

```
AGT:c.803T>C
9:g.22125504G>C
ENST00000003084:c.1431_1433delTTC
19:g.110747_110748insT
LRG_101t1:c.1019T>C
```

Examples: Ensembl default, VCF, Variant identifiers, HGVS notations

Or upload file: Browse... No file selected.

Or provide file URL:

Transcript database to use:

- Ensembl/GENCODE transcripts
- Ensembl/GENCODE basic transcripts
- RefSeq transcripts
- Ensembl/GENCODE and RefSeq transcripts

Variants and frequency data | Co-located variants and frequency data

Variants and frequency data

Find co-located known variants: Yes

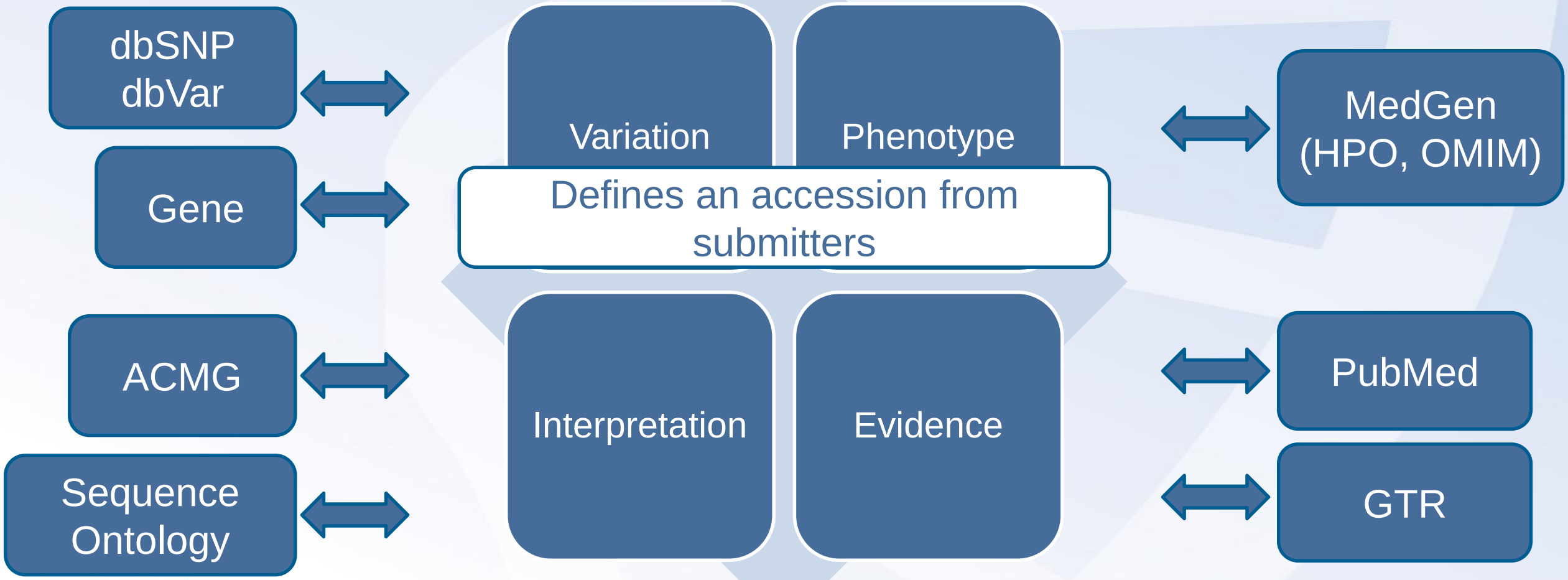
Frequency data for co-located variants:

- [x] 1000 Genomes global minor allele frequency
- [ ] 1000 Genomes continental allele frequencies
- [ ] ESP allele frequencies
- [x] gnomAD (exomes) allele frequencies

Report allele frequencies from the genome Aggregation Database (exomes)

PubMed IDs for citations of co-located variants:

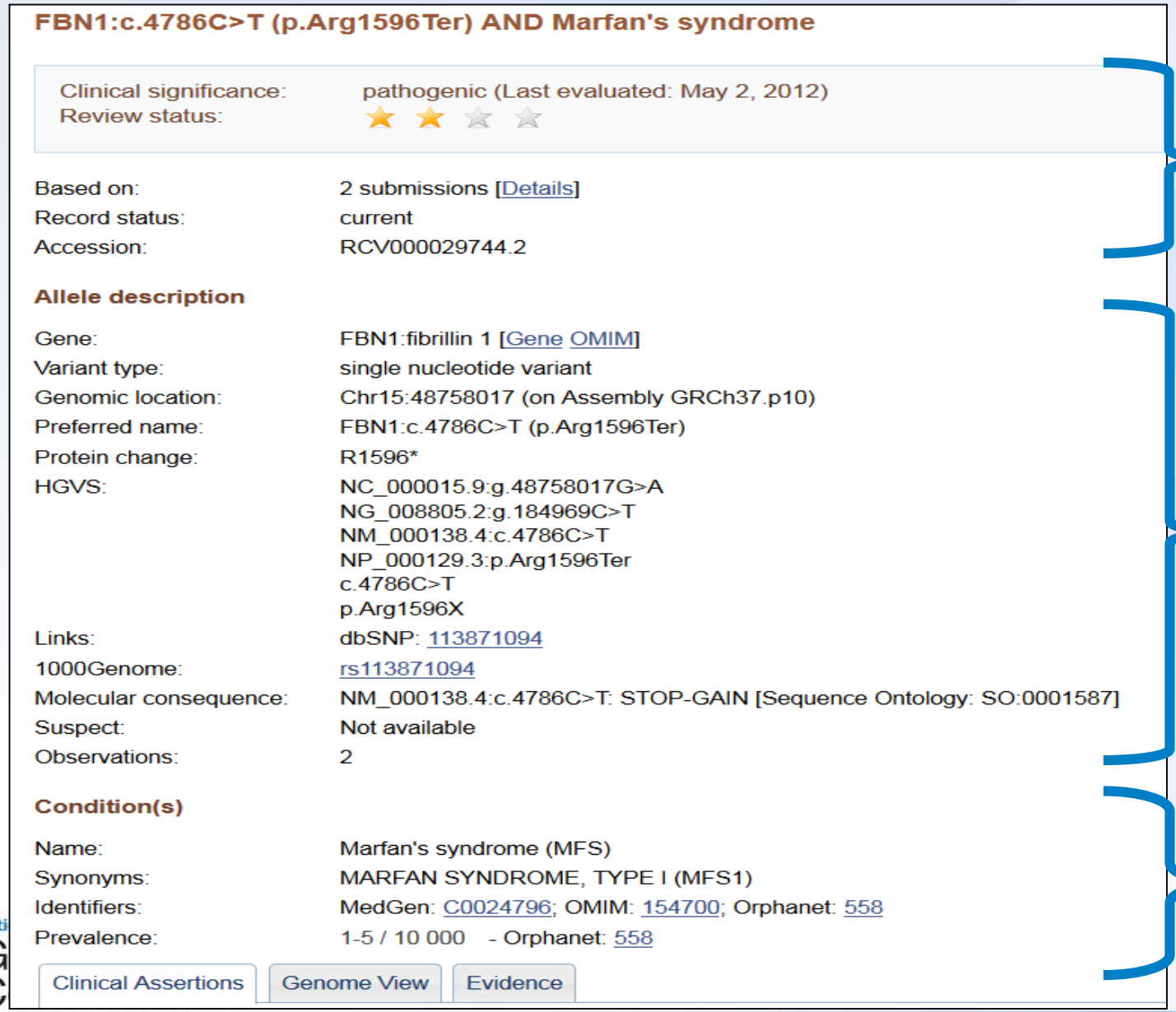
ClinVar integrates four domains of information



National Society of Genetic Counselors

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

ClinVar detailed display



FBN1:c.4786C>T (p.Arg1596Ter) AND Marfan's syndrome

Clinical significance:	pathogenic (Last evaluated: May 2, 2012)
Review status:	★★☆☆
Based on:	2 submissions [Details]
Record status:	current
Accession:	RCV000029744.2

Allele description

Gene:	FBN1:fibrillin 1 [Gene OMIM]
Variant type:	single nucleotide variant
Genomic location:	Chr15:48758017 (on Assembly GRCh37.p10)
Preferred name:	FBN1:c.4786C>T (p.Arg1596Ter)
Protein change:	R1596*
HGVS:	NC_000015.9:g.48758017G>A NG_008805.2:g.184969C>T NM_000138.4:c.4786C>T NP_000129.3:p.Arg1596Ter c.4786C>T p.Arg1596X
Links:	dbSNP: 113871094
1000Genome:	rs113871094
Molecular consequence:	NM_000138.4:c.4786C>T: STOP-GAIN [Sequence Ontology: SO:0001587]
Suspect:	Not available
Observations:	2

Condition(s)

Name:	Marfan's syndrome (MFS)
Synonyms:	MARFAN SYNDROME, TYPE I (MFS1)
Identifiers:	MedGen: C0024796; OMIM: 154700; Orphanet: 558
Prevalence:	1-5 / 10 000 - Orphanet: 558

Clinical Assertions | Genome View | Evidence

Interpretation
- Significance
- Review status *
- Accession.version *

Allele summary
- Gene
- Variant type
- Genomic location
- HGVS expressions*
- Molecular consequence*
- Links*
- Frequency*

Phenotype summary
- Names
- Links*
- Age of onset *
- Prevalence *

* May be provided by NCBI

ClinVar detailed display

Clinical Assertions | Genome View | Evidence

Help

Submission Accession	Submitter	Review Status	Clinical Significance (Last evaluated)	Origin	Method	Consequence	Citations
SCV000052397	LabCorp	classified by single submitter	pathogenic (Aug 18, 2011)	germline	curation, clinical testing		PubMed (6)
SCV000058856	Partners Healthcare/ Harvard Medical School	classified by single submitter	pathogenic (May 2, 2012)	germline	clinical testing		PubMed (3)

National Society of Genetic Counselors

ClinVar detailed display

Clinical Assertions | Genome View | Evidence

Summary from all submissions Help

Ethnicity	Origin	Affected	Alleles observed	Families	Chromosomes tested	Number Tested	Family history	Method
not provided	germline	yes	6	not provided	not provided	6	not provided	curation
not provided	germline	unknown	2	2	not provided	not provided	not provided	clinical testing

Citations

PubMed

Evaluation and application of denaturing HPLC for mutation detection in Marfan syndrome: Identification of 20 novel mutations and two novel polymorphisms in the FBN1 gene.
Mátyás G, De Paepe A, Halliday D, Boileau C, Pals G, Steinmann B.
Hum Mutat. 2002 Apr;19(4):443-56.
PubMed [citation] PMID: 11933199

Preimplantation genetic diagnosis for Marfan syndrome.
Spits C, De Rycke M, Verpoest W, Lissens W, Van Steirteghem A, Liebaers I, Sermon K.
Fertil Steril. 2006 Aug;86(2):310-20. Epub 2006 Jun 6.
PubMed [citation] PMID: 16756980

See all PubMed Citations (6)

Details of each submission

- From LabCorp, SCV000052397.1
- From Partners Healthcare/ Harvard Medical School, SCV000058856.1

National Society of Genetic Counselors

ClinVar detailed display

Clinical Assertions | Genome View | Evidence

Summary from all submissions Help

Ethnicity	Origin	Affected	Alleles observed	Families	Chromosomes tested	Number Tested	Family history	Method

▾ **From Partners Healthcare/ Harvard Medical School, SCV000058856.1**

Ethnicity	Alleles Observed	Chromosomes Tested	Family History	Method	Citations
not provided	not provided	not provided	not provided	clinical testing	PubMed (3)

Description

The Arg1596X variant has been reported in 4 individuals with clinical features of Marfan syndrome and was shown to segregate with clinical feature in one reported family (Loeys 2001, De Backer 2007, Magyar 2009,). In addition, this variant has been identified in one individual with clinical features of Marfan syndrome by our laboratory. This nonsense variant leads to a premature termination codon at position 1596, which is predicted to lead to a truncated or absent protein. Heterozygous loss of function of the FBN1 gene is an established disease mechanism in Marfan syndrome. In summary, this variant meets our criteria to be classified as pathogenic (http://pcpgm.partners.org/LMM).

#	Sample				Method		Observation		
	Origin	Affected	Number tested	Tissue	Purpose	Method	Variant Alleles	Families	Co-occurrences
1	germline	unknown	not provided	not provided	assert pathogenicity	clinical testing	not provided	2	not provided

▸ From Partners Healthcare/ Harvard Medical School, SCV000058856.1

Genetic Counselors

Ensembl Variation: *TSC2* Frameshift Variants

Generade
Applied Genomics for Life

Variant table

Too many data to display

There are 23,978 variants for this Gene, which is too many to display in this page, so **only exonic variants** are displayed. Please use BioMart to extract all data.

Filter: Global MAF: All | SIFT: All | PolyPhen: All | Consequences: frameshift variant | Filter Other Columns

Show/hide columns | Search...

Variant ID	Chr: bp	Alleles	Class	Source	Evidence	Clin. Sig.	Conseq. Type	AA	AA co-ord	Transcript
rs137854295	16:2048619	G/-	deletion	dbSNP		?	frameshift variant	A/X	2	ENST00000219476.8
rs137854019	16:2048636	A/-	deletion	dbSNP		?	frameshift variant	K/X	7	ENST00000219476.8
rs796053515	16:2048647	T/-	deletion	dbSNP	-	⚠	frameshift variant	L/X	11	ENST00000219476.8
rs397515020	16:2048660	T/-	deletion	dbSNP		?	frameshift variant	F/X	15	ENST00000219476.8
rs397514958	16:2048660-2048666	TAAGATT/-	deletion	dbSNP		?	frameshift variant	FKI/X	15-17	ENST00000219476.8
rs863225040	16: between 2048660 & 2048661	-/T	insertion	dbSNP		⚠	frameshift variant	-/X	16-15	ENST00000219476.8
rs137854178	16: between 2048724 & 2048725	-/G	insertion	dbSNP		?	frameshift variant	E/GX	37	ENST00000219476.8

Human Gene Mutation Database (www.hgmd.org)

Generade
Applied Genomics for Life

The Human Gene Mutation Database

at the Institute of Medical Genetics in Cardiff

HGMD®

QIAGEN

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

Gene symbol Go! Symbol: Missense/nonsense Go!

Table:	Description:	Public entries: This site. Academic/non-profit users only	Total entries: HGMD Professional 2018.4
	Mutation totals (as of 2019-04-01)	**171400**	**248700**
Gene symbol	The gene description, gene symbol (as recommended by the HUGO Nomenclature Committee) and chromosomal location is recorded for each gene. In cases where a gene symbol has not yet been made official, a provisional symbol has been adopted which is denoted by lower-case letters.	7038	10389
cDNA sequence	cDNA reference sequences are provided, numbered by codon.	7088	10554
Genomic coordinates	Genomic (chromosomal) coordinates have been calculated for missense/nonsense, splicing, regulatory, small deletions, small insertions and small indels.	0	222169
HGVS nomenclature	Standard HGVS nomenclature has been obtained for missense/nonsense, splicing, regulatory, small deletions, small insertions and small indels.	0	222557
Missense/nonsense	Single base-pair substitutions in coding regions are presented in terms of a triplet change with an additional flanking base included if the mutated base lies in either the first or third position in the triplet.	95591	142868
Splicing	Mutations with consequences for mRNA splicing are presented in brief with informat ecifying the relative position of the lesion with respect to a numbered intron donor or acceptor splice site. Positions given as positive integers ref 3' (downstream) location, negative integers refer to a 5' (upstream) location.	15576	21773
Regulatory	Substitutions causing regulatory abnormalities are logged in with thirty nucleotides flanking the site of the mutation on both sides. The location of the mutation relative to the transcriptional initiation site, initiaton codon, polyadenylation site or termination codon is given.	3248	4253
Small deletions	Micro-deletions (20 bp or less) are presented in terms of the deleted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character (^).	25641	36227
Small insertions	Micro-insertions (20 bp or less) are presented in terms of the inserted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character (^).	10686	15267
Small indels	Micro-indels (20 bp or less) are presented in terms of the deleted/inserted bases in lower case plus, in upper case, 10 bp DNA sequence flanking both sides of the lesion. The numbered codon is preceded in the given sequence by the caret character (^).	2451	3327
Gross deletions	Information regarding the nature and location of each lesion is logged in narrative form because of the extremely variable quality of the original data reported.	12969	17947
Gross insertions	Information regarding the nature and location of each lesion is logged in narrative form because of the extremely variable quality of the original data reported.	3108	4451
Complex rearrangements	Information regarding the nature and location of each lesion is logged in narrative form because of the extremely variable quality of the original data reported.	1652	2062
Repeat variations	Information regarding the nature and location of each lesion is logged in narrative form because of the extremely variable quality of the original data reported.	478	525

Human Gene Mutation Database (www.hgmd.org)

Generade
Applied Genomics for Life

Gene Symbol	Chromosomal location	Gene name	cDNA sequence	Extended cDNA	Mutation viewer
CREBBP (Aliases: available to subscribers)	16p13.3	CREB-binding protein (Aliases: available to subscribers)	NM_004380.2	Not available	Available to subscribers

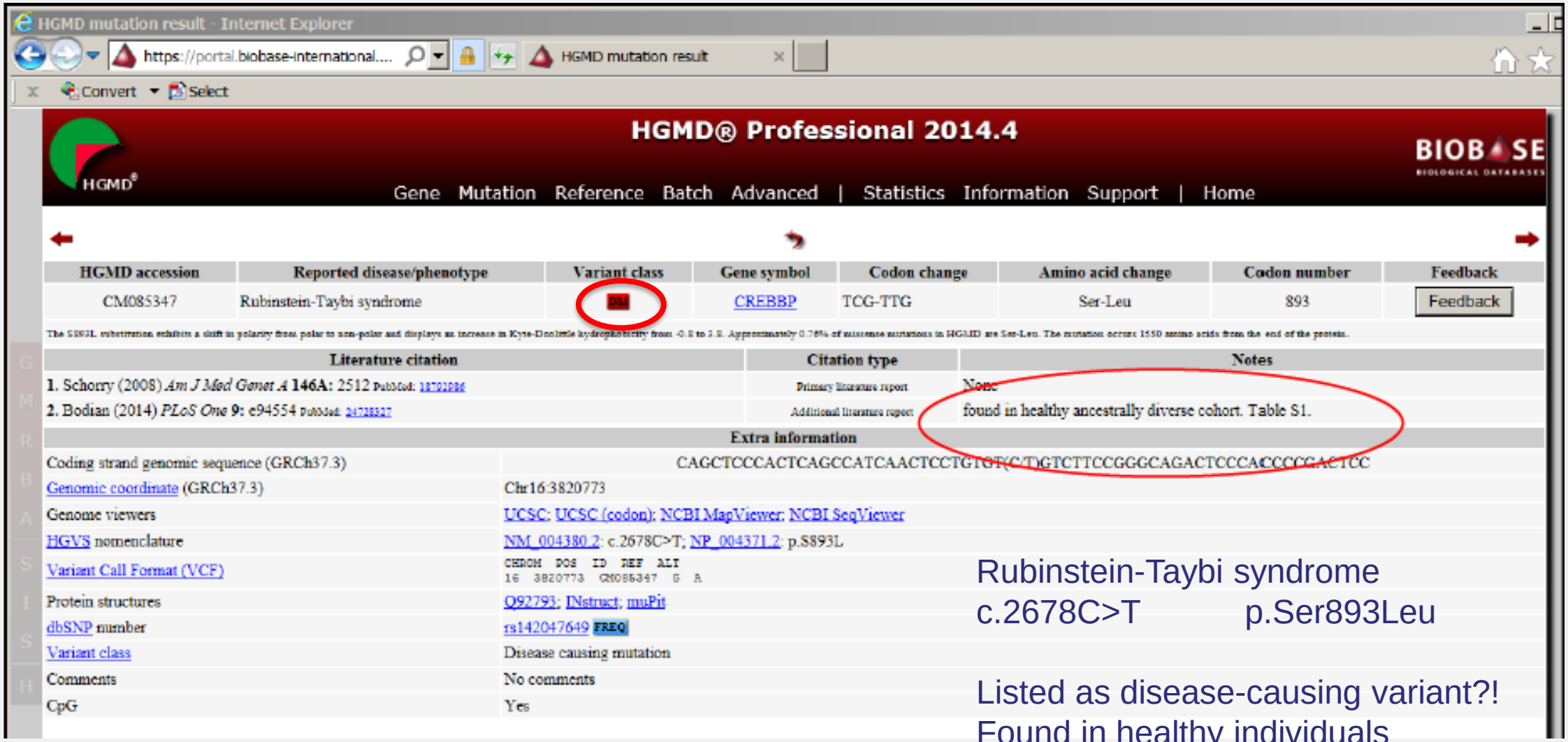
Mutation type	Number of mutations	Mutation data by type (register or log in)
Missense/nonsense	78	Get mutations
Splicing	20	Get mutations
Regulatory	0	No mutations
Small deletions	41	Get mutations
Small insertions	23	Get mutations
Small indels	3	Get mutations
Gross deletions	44	Get mutations
Gross insertions/duplications	26	Get mutations
Complex rearrangements	3	Get mutations
Repeat variations	0	No mutations
Get all mutations by type	Available to subscribers	

CM085337	CAA-TAA	Gln-Term	676	Available to subscribers	Rubinstein-Taybi syndrome	Schorry (2008) Am J Med Genet A 146A, 2512 Additional report available to subscribers
CM085342	CGA-TGA	Arg-Term	768	Available to subscribers	Rubinstein-Taybi syndrome	Schorry (2008) Am J Med Genet A 146A, 2512 Additional report available to subscribers
CM085347	TCG-TTG	Ser-Leu	893	Available to subscribers	Rubinstein-Taybi syndrome	Schorry (2008) Am J Med Genet A 146A, 2512 Additional report available to subscribers
CM098474	CAG-TAG	Gln-Term	948	Available to subscribers	Rubinstein-Taybi syndrome	Wieczorek (2009) Am J Med Genet A 149A, 2849 Additional report available to subscribers

Information extracted from literature
Limited information accessible after free registration for academia

HGMD Professional - *CREBBP*

Generade
Applied Genomics for Life

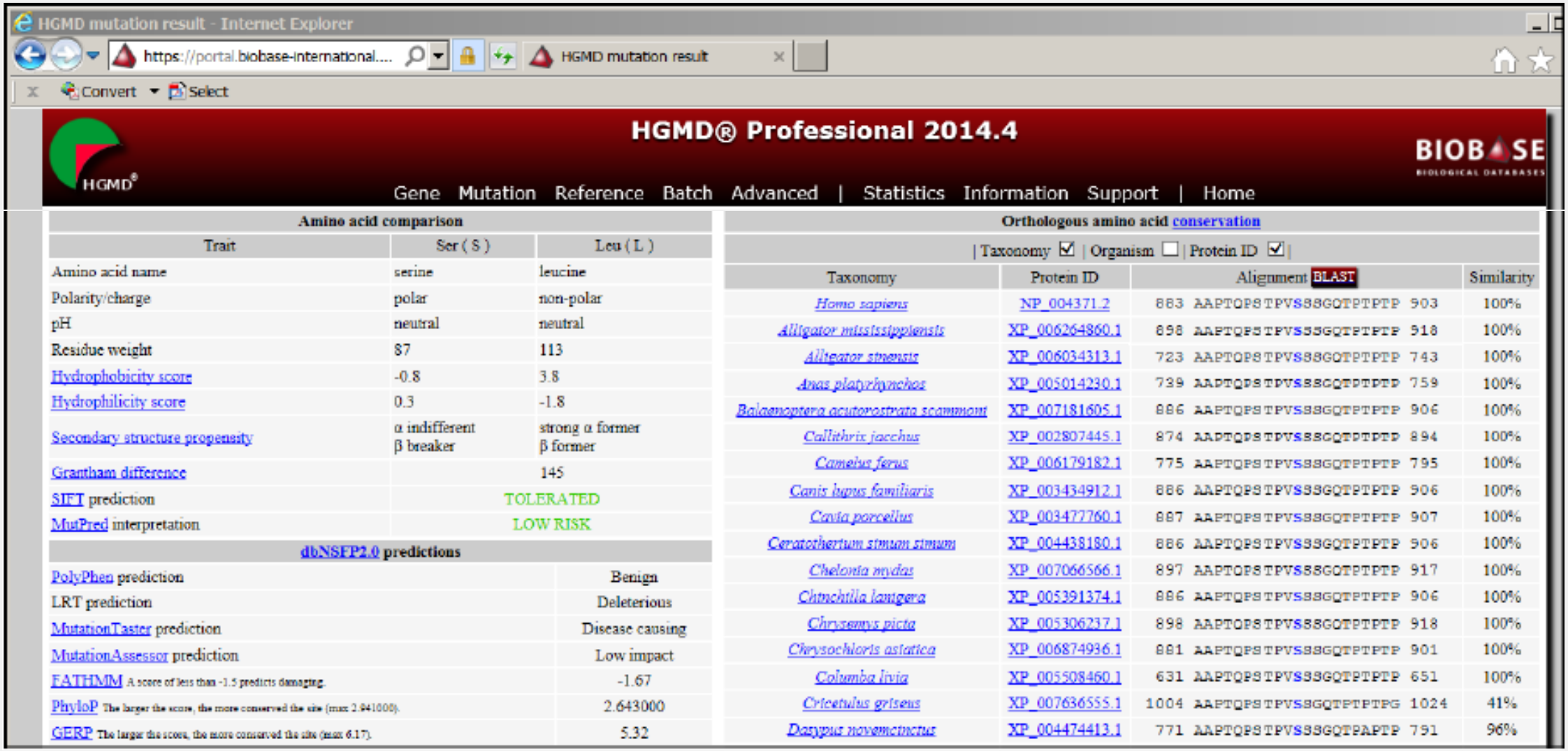


Rubinstein-Taybi syndrome
c.2678C>T p.Ser893Leu

Listed as disease-causing variant?!
Found in healthy individuals

HGMD Professional - *CREBBP*

Generade
Applied Genomics for Life



Amino acid comparison

Trait	Ser (S)	Leu (L)
Amino acid name	serine	leucine
Polarity/charge	polar	non-polar
pH	neutral	neutral
Residue weight	87	113
Hydrophobicity score	-0.8	3.8
Hydrophilicity score	0.3	-1.8
Secondary structure propensity	α indifferent β breaker	strong α former β former
Grantham difference		145
SIFT prediction		TOLERATED
MutPred interpretation		LOW RISK

dbNSFP2.0 predictions

PolyPhen prediction	Benign
LRT prediction	Deleterious
MutationTaster prediction	Disease causing
MutationAssessor prediction	Low impact
FATHMM A score of less than -1.5 predicts damaging.	-1.67
PhyloP The larger the score, the more conserved the site (max 2.841000).	2.643000
GERP The larger the score, the more conserved the site (max 6.17).	5.32

Orthologous amino acid conservation

| Taxonomy ☑ | Organism ☐ | Protein ID ☑ |

Taxonomy	Protein ID	Alignment BLAST	Similarity
Homo sapiens	NP_004371.2	883 AAPTQPSTPVSSSGQTPTPTP 903	100%
Alligator mississippiensis	XP_006264860.1	898 AAPTQPSTPVSSSGQTPTPTP 918	100%
Alligator sinensis	XP_006034313.1	723 AAPTQPSTPVSSSGQTPTPTP 743	100%
Anas platyrhynchos	XP_005014230.1	739 AAPTQPSTPVSSSGQTPTPTP 759	100%
Balaenoptera acutorostrata scammoni	XP_007181605.1	886 AAPTQPSTPVSSSGQTPTPTP 906	100%
Callithrix jacchus	XP_002807445.1	874 AAPTQPSTPVSSSGQTPTPTP 894	100%
Camelus ferus	XP_006179182.1	775 AAPTQPSTPVSSSGQTPTPTP 795	100%
Canis lupus familiaris	XP_003434912.1	886 AAPTQPSTPVSSSGQTPTPTP 906	100%
Cavia porcellus	XP_003477760.1	887 AAPTQPSTPVSSSGQTPTPTP 907	100%
Ceratotherium simum simum	XP_004438180.1	886 AAPTQPSTPVSSSGQTPTPTP 906	100%
Chelonia mydas	XP_007066566.1	897 AAPTQPSTPVSSSGQTPTPTP 917	100%
Chinchilla lanigera	XP_005391374.1	886 AAPTQPSTPVSSSGQTPTPTP 906	100%
Chrysemys picta	XP_005306237.1	898 AAPTQPSTPVSSSGQTPTPTP 918	100%
Chrysochloris asiatica	XP_006874936.1	881 AAPTQPSTPVSSSGQTPTPTP 901	100%
Columba livia	XP_005508460.1	631 AAPTQPSTPVSSSGQTPTPTP 651	100%
Cricetulus griseus	XP_007636555.1	1004 AAPTQPSTPVSSGQTPTPTPG 1024	41%
Dasypus novemcinctus	XP_004474413.1	771 AAPTQPSTPVSSSGQTPAPTP 791	96%

genome Aggregation Database (gnomAD)– *CREBBP*

Generade
Applied Genomics for Life

Variant ID	Source	Consequence	Annotation	Flags	Allele Count	Allele Number	Allele Frequency
16-3820749-G-A	E G	p.Pro901Leu	missense		3	282708	1.06e-5
16-3820750-G-A	E	p.Pro901Ser	missense		1	251364	3.98e-6
16-3820756-G-A	E	p.Pro899Ser	missense		1	251370	3.98e-6
16-3820765-C-G	E	p.Gly896Arg	missense		1	251360	3.98e-6
16-3820773-G-A	E G	p.Ser893Leu	missense		280	282748	9.9e-4
16-3820776-A-C	E	p.Val892Gly	missense		2	251400	7.96e-6

p.Ser893Leu:
280 out of 282748 alleles, frequency: 0.00099

http://gnomad.broadinstitute.org

Human gene information database

Generade
Applied Genomics for Life

OMIM® http://omim.org/

Online Mendelian Inheritance in Man®

An Online Catalog of Human Genes and Genetic Disorders

Updated March 30, 2018

POMGNT1

Advanced Search : OMIM, Clinical Synopses, Gene Map

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

Mirror site : mirror.omim.org

OMIM is supported by a grant from NHGRI, licensing fees, and generous contributions from people like you.

Search: 'POMGNT1'

Results: 16 entries. Show 100 | Download As | « First | ‹ Previous | Next › | Last »

1: * 606822. PROTEIN O-MANNOSE BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE; POMGNT1

Cytogenetic location: 1p34.1, Genomic coordinates (GRCh38): 1:46,188,680-46,220,304

Matching terms: pomgnt1

▶ Gene-Phenotype Relationships ▶ Links

OMIM – *POMGNT1*

Generade
Applied Genomics for Life

* 606822

PROTEIN O-MANNOSE BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE; POMGNT1

HGNC Approved Gene Symbol: POMGNT1

Cytogenetic location: 1p34.1 *Genomic coordinates (GRCh38): 1:46,188,680-46,220,304* (from NCBI)

Gene-Phenotype Relationships

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key
1p34.1	Muscular dystrophy-dystroglycanopathy (congenital with brain and eye anomalies), type A, 3	253280	AR	3
	Muscular dystrophy-dystroglycanopathy (congenital with mental retardation), type B, 3	613151	AR	3
	Muscular dystrophy-dystroglycanopathy (limb-girdle), type C, 3	613157	AR	3
	Retinitis pigmentosa 76	617123	AR	3

External Links
- Genome
- DNA
- Protein
- Gene Info
- Clinical Resources
- Variation
  - 1000 Genome
  - ClinVar
  - ExAC
  - gnomAD
  - GWAS Central
  - HGMD
  - HGVS
  - Locus Specific DBs
  - NHLBI EVS
  - PharmGKB
- Animal Models
- Cellular Pathways

OMIM – *POMGNT1* variants

Generade
Applied Genomics for Life

▼ **ALLELIC VARIANTS (23 Selected Examples):**

Table View | ClinVar

.0001 MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (CONGENITAL WITH BRAIN AND EYE ANOMALIES), TYPE A, 3

POMGNT1, IVS17DS, G-T, +1 dbSNP:rs587777821 RCV000004193

No HGVS descriptions

In 2 unrelated patients with muscle-eye-brain disease (MDDGA3; 253280), each the offspring of consanguineous parents, Yoshida et al. (2001) identified a splice site mutation (IVS17DS+1G-T) in the POMGNT1 gene, leading to deletion of amino acids leu472 to his513. Each patient was homozygous for the mutation.

606822 Table View

Download As

PROTEIN O-MANNOSE BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE; POMGNT1

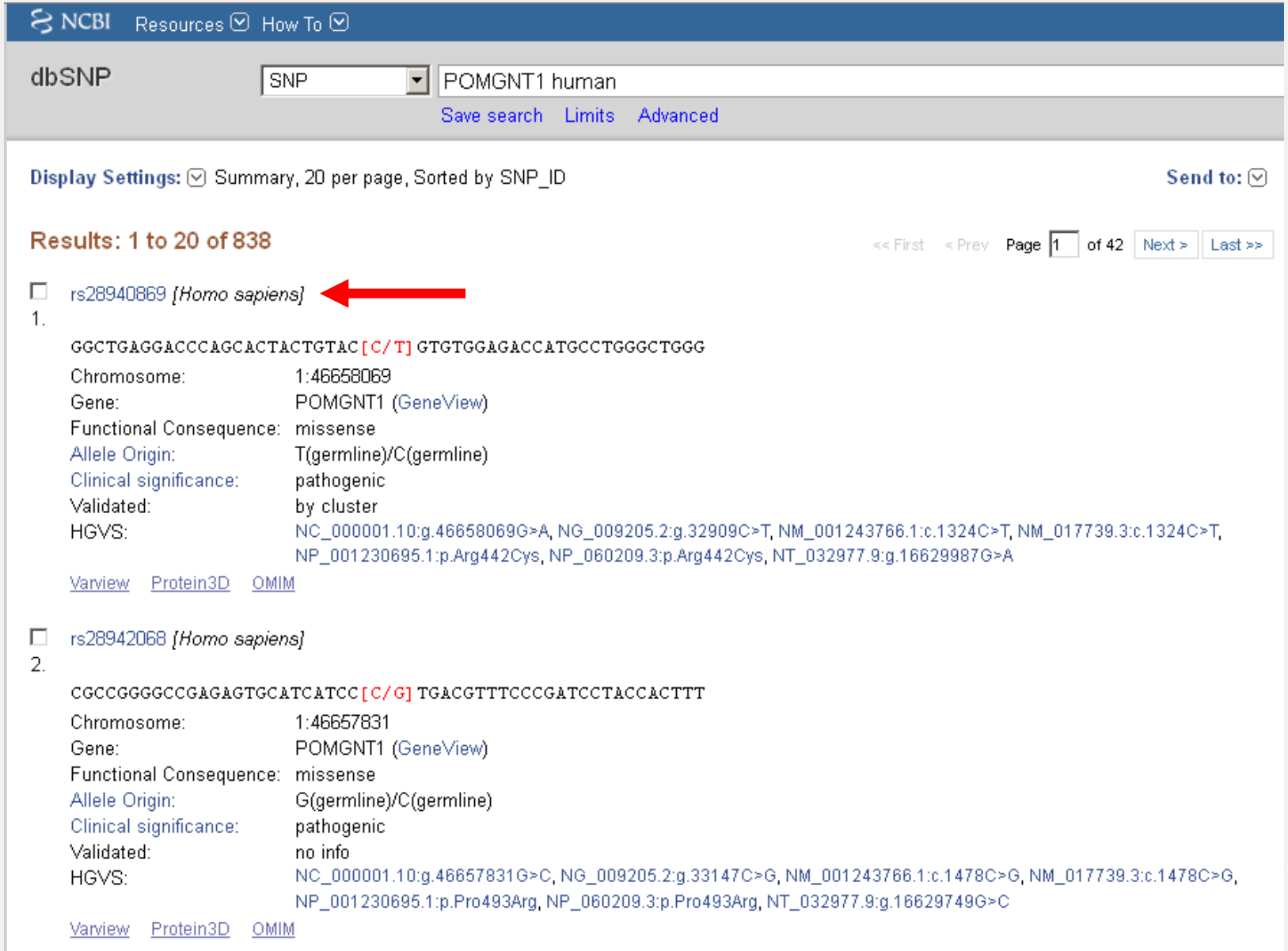
Allelic Variants (23 Selected Examples) : All ClinVar Variants

Number	Phenotype	Mutation	dbSNP	ExAC	ClinVar
.0001	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (CONGENITAL WITH BRAIN AND EYE ANOMALIES), TYPE A, 3	POMGNT1, IVS17DS, G-T, +1	[rs587777821]	-	[RCV000004193]
.0002	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (CONGENITAL WITH BRAIN AND EYE ANOMALIES), TYPE A, 3	POMGNT1, IVS17DS, G-A, +1	-	-	[RCV000004194]

dbSNP

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

Generade
Applied Genomics for Life



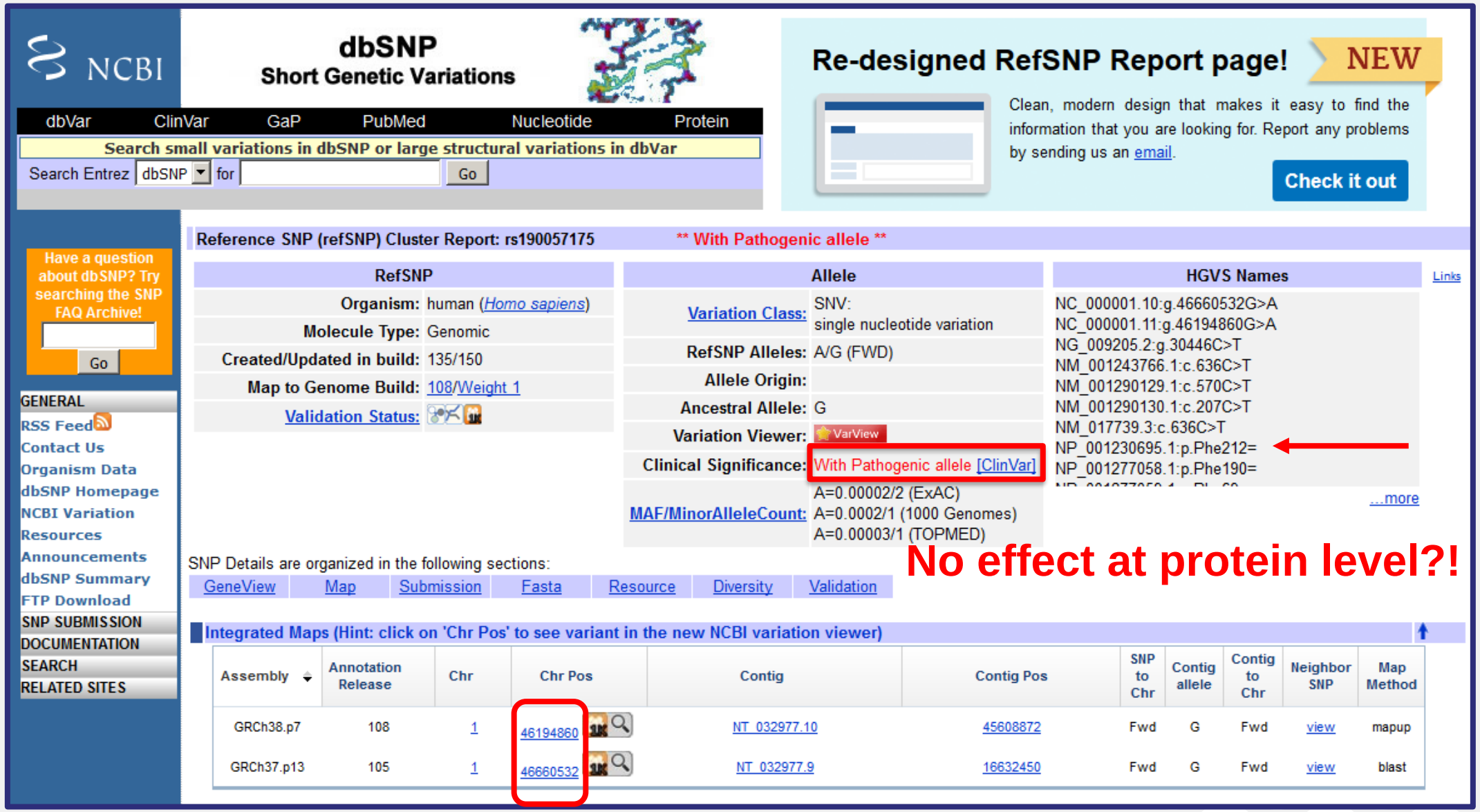
Before build 134:
Only frequent variants

Build 135 and later:
also disease-causing and somatic variants!

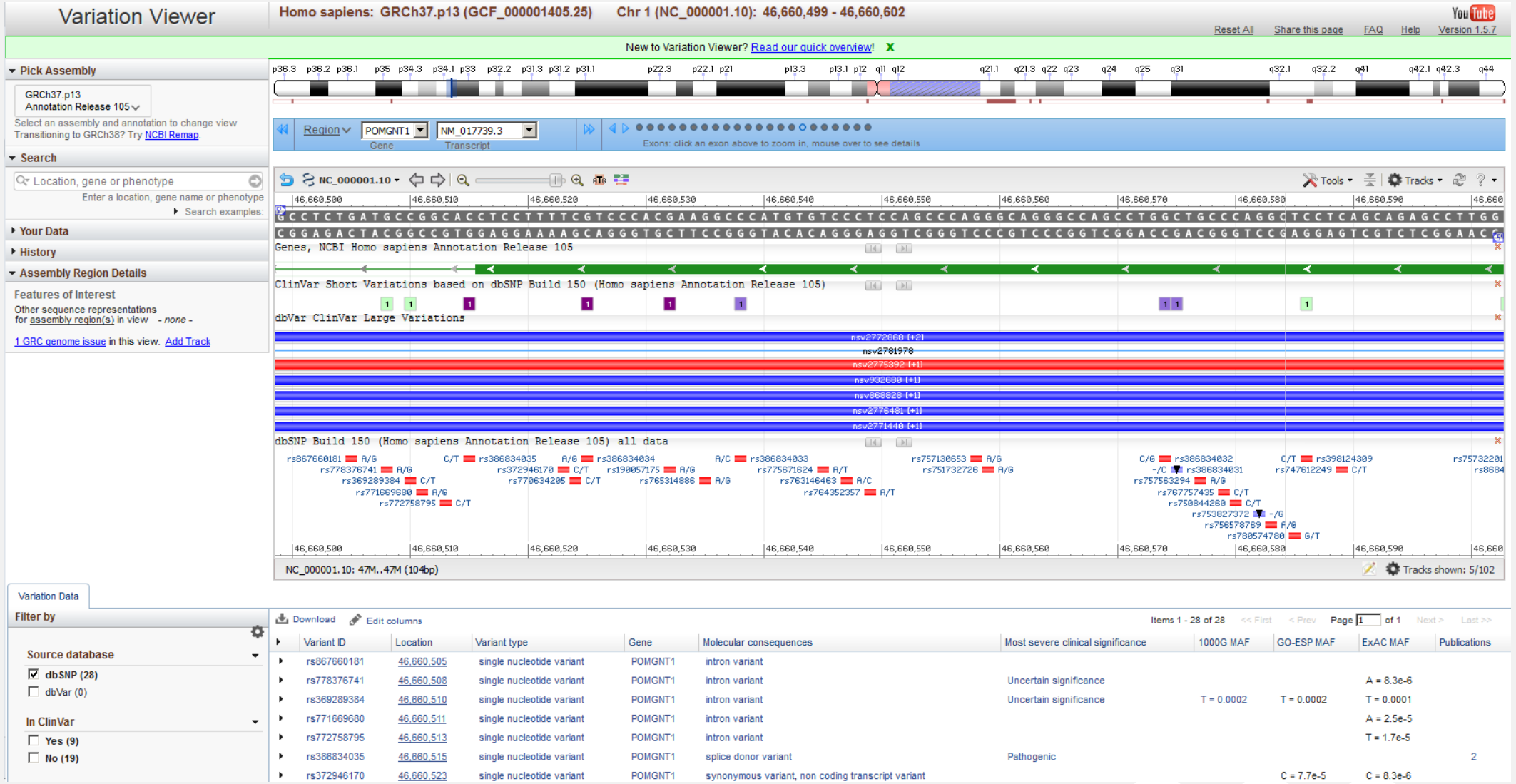
dbSNP *POMGNT1* variant

Generade
Applied Genomics for Life

"Classic View"



NCBI variation viewer – *POMGNT1*



http://www.ncbi.nlm.nih.gov/variation/view/

Relevant databases for cancer diagnostics

Table 1 Databases Relevant to Interpretation of Somatic Sequence Variants

Utility/function	Database	Location (web address)
Population databases to exclude polymorphisms	1000 Genomes Project[16]	http://browser.1000genomes.org
	Exome Variant Server	http://evs.gs.washington.edu/EVS
	dbSNP[17]	http://www.ncbi.nlm.nih.gov/snp
	dbVar[18]	http://www.ncbi.nlm.nih.gov/dbvar
	ExAC	http://exac.broadinstitute.org
Cancer-specific variant databases	Catalog of Somatic Mutations in Cancer[19]	http://cancer.sanger.ac.uk/cosmic
	My Cancer Genome	http://www.mycancergenome.org
	Personalized cancer therapy, MD Anderson Cancer Center	https://pct.mdanderson.org
	cBioPortal, Memorial Sloan Kettering Cancer Center[20]	http://www.cbioportal.org
	Intogen[21]	https://www.intogen.org/search
	ClinicalTrials.gov	https://clinicaltrials.gov
	IARC (WHO) TP53 mutation database[22]	http://p53.iarc.fr
	Pediatric Cancer Genome Project (St. Jude Children's Research Hospital—Washington University)	http://explorepcgp.org
	International Cancer Genome Consortium[23]	https://dcc.icgc.org
Sequence repositories and data hosts	NCBI Genome	http://www.ncbi.nlm.nih.gov/genome
	RefSeqGene[24]	http://www.ncbi.nlm.nih.gov/refseq/rsg
	Locus Reference Genomic[25]	http://www.lrg-sequence.org
	UCSC table browser[26]	https://genome.ucsc.edu/cgi-bin/hgTables
	Ensemble BioMart[27]	http://useast.ensembl.org/biomart/martview
Other disease/mutation databases useful in the context of variant interpretation for cancer genomics	ClinVar[28]	http://www.ncbi.nlm.nih.gov/clinvar
	Human Gene Mutation Database[29]	http://www.hgmd.org
	Leiden Open Variation Database[30]	http://www.lovd.nl
	dbNSFP (compiled database of precomputed *in silico* prediction scores for nonsynonymous SNVs)[31]	https://sites.google.com/site/jpopgen/dbNSFP
	Ensemble Variant Effect Predictor[15]	http://www.ensembl.org/info/docs/tools/vep/index.html

Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer

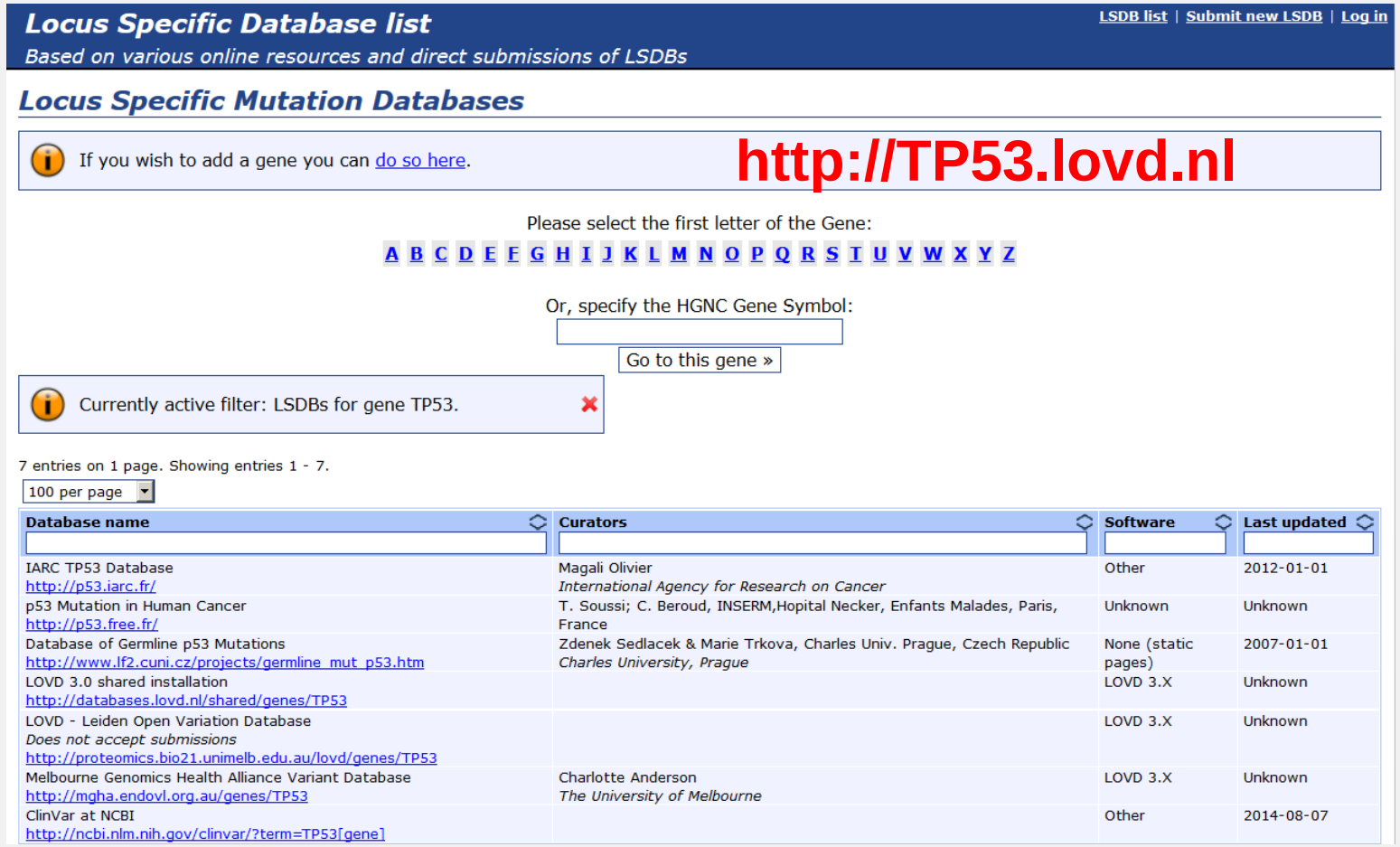
A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists

Marilyn M. Li,[*†] Michael Datto,[*†] Eric J. Duncavage,[*§] Shashikant Kulkarni,[*¶] Neal I. Lindeman,[*‖] Somak Roy,[*,**] Apostolia M. Tsimberidou,[*††] Cindy L. Vnencak-Jones,[*‡‡] Daynna J. Wolff,[*§§] Anas Younes,[*¶¶] and Marina N. Nikiforova[*,**]

Li MM et al. (2017) J Mol Diagn 19:5-23

Gene variant database lookup

- Search using http://Gene_symbol.LOVD.nl
  - http://TP53.lovd.nl redirects to list or TP53 database homepage
  - http://USH2A.lovd.nl redirects to list



Locus Specific Database list

Based on various online resources and direct submissions of LSDBs

LSDB list | Submit new LSDB | Log in

Locus Specific Mutation Databases

If you wish to add a gene you can do so here.

http://TP53.lovd.nl

Please select the first letter of the Gene:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Or, specify the HGNC Gene Symbol:

Go to this gene »

Currently active filter: LSDBs for gene TP53.

7 entries on 1 page. Showing entries 1 - 7.

100 per page

Database name	Curators	Software	Last updated
IARC TP53 Database http://p53.iarc.fr/	Magali Olivier *International Agency for Research on Cancer*	Other	2012-01-01
p53 Mutation in Human Cancer http://p53.free.fr/	T. Soussi; C. Beroud, INSERM,Hopital Necker, Enfants Malades, Paris, France	Unknown	Unknown
Database of Germline p53 Mutations http://www.lf2.cuni.cz/projects/germline_mut_p53.htm	Zdenek Sedlacek & Marie Trkova, Charles Univ. Prague, Czech Republic *Charles University, Prague*	None (static pages)	2007-01-01
LOVD 3.0 shared installation http://databases.lovd.nl/shared/genes/TP53		LOVD 3.X	Unknown
LOVD - Leiden Open Variation Database *Does not accept submissions* http://proteomics.bio21.unimelb.edu.au/lovd/genes/TP53		LOVD 3.X	Unknown
Melbourne Genomics Health Alliance Variant Database http://mgha.endovl.org.au/genes/TP53	Charlotte Anderson *The University of Melbourne*	LOVD 3.X	Unknown
ClinVar at NCBI http://ncbi.nlm.nih.gov/clinvar/?term=TP53[gene]		Other	2014-08-07

International Agency for Research on Cancer

World Health Organization

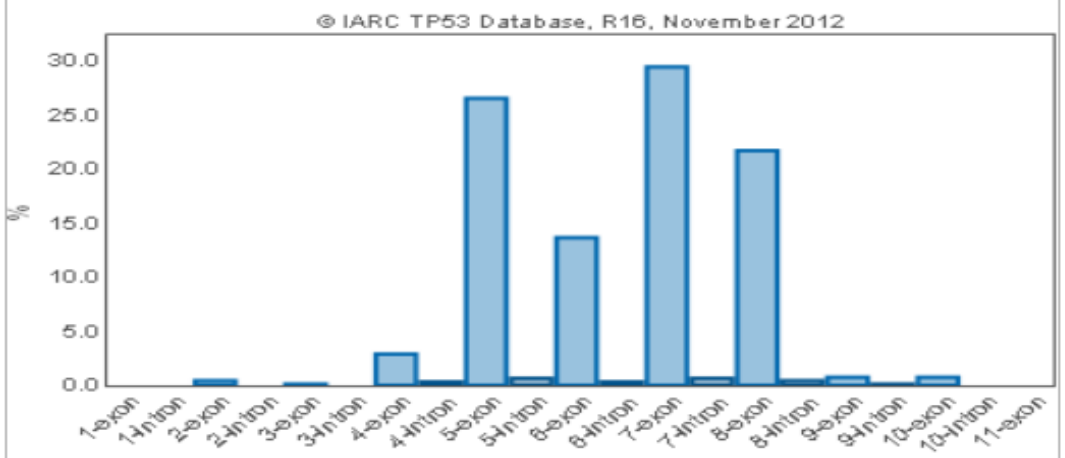
IARC TP53 Database

Generade
Applied Genomics for Life

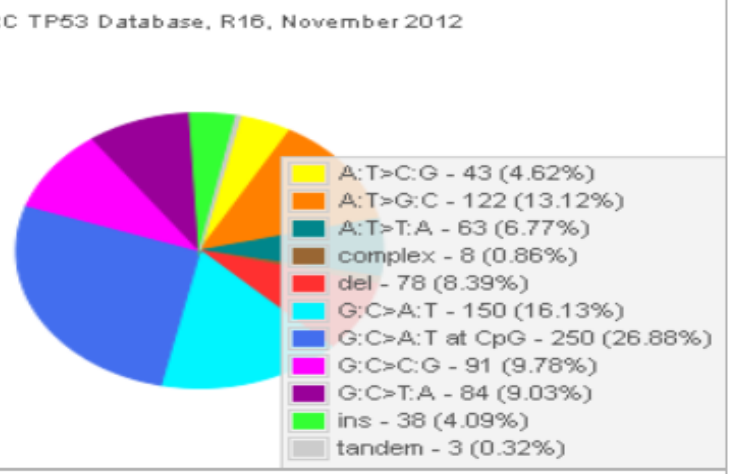
Mutation type | Mutation effect

Query: somatic variants in haematopoietic tissues

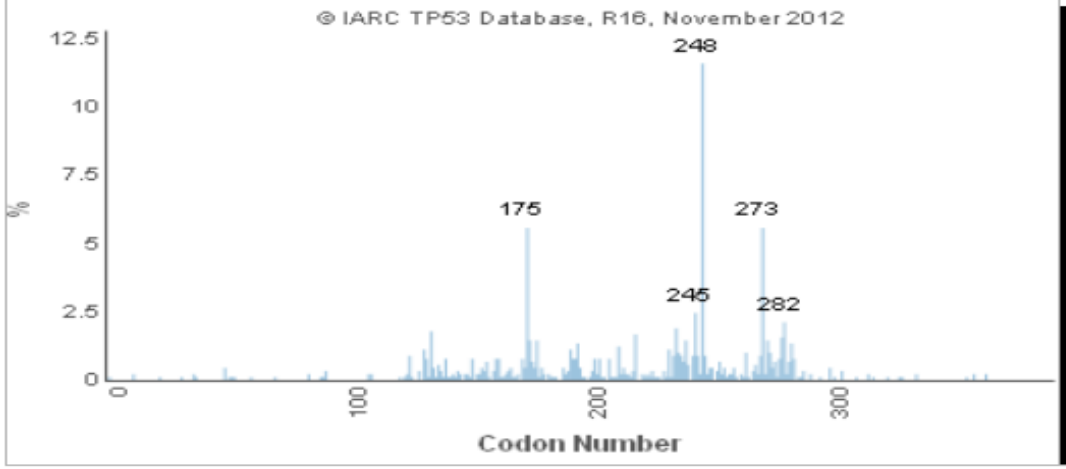
Exon/Intron Distribution (N = 930)



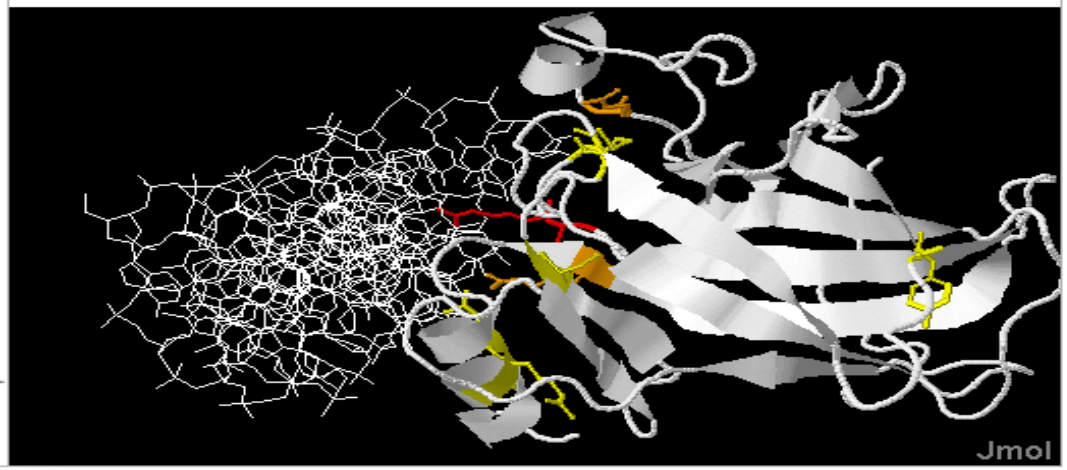
Mutation Type (N = 930)



Codon distribution (N = 896)



3D - Graph



COSMIC
Catalogue Of Somatic Mutations In Cancer

https://cancer.sanger.ac.uk/cosmic

Projects | Data | Tools | News | Help | About | Genome Version | Search COSMIC... SEARCH

Gene

TP53

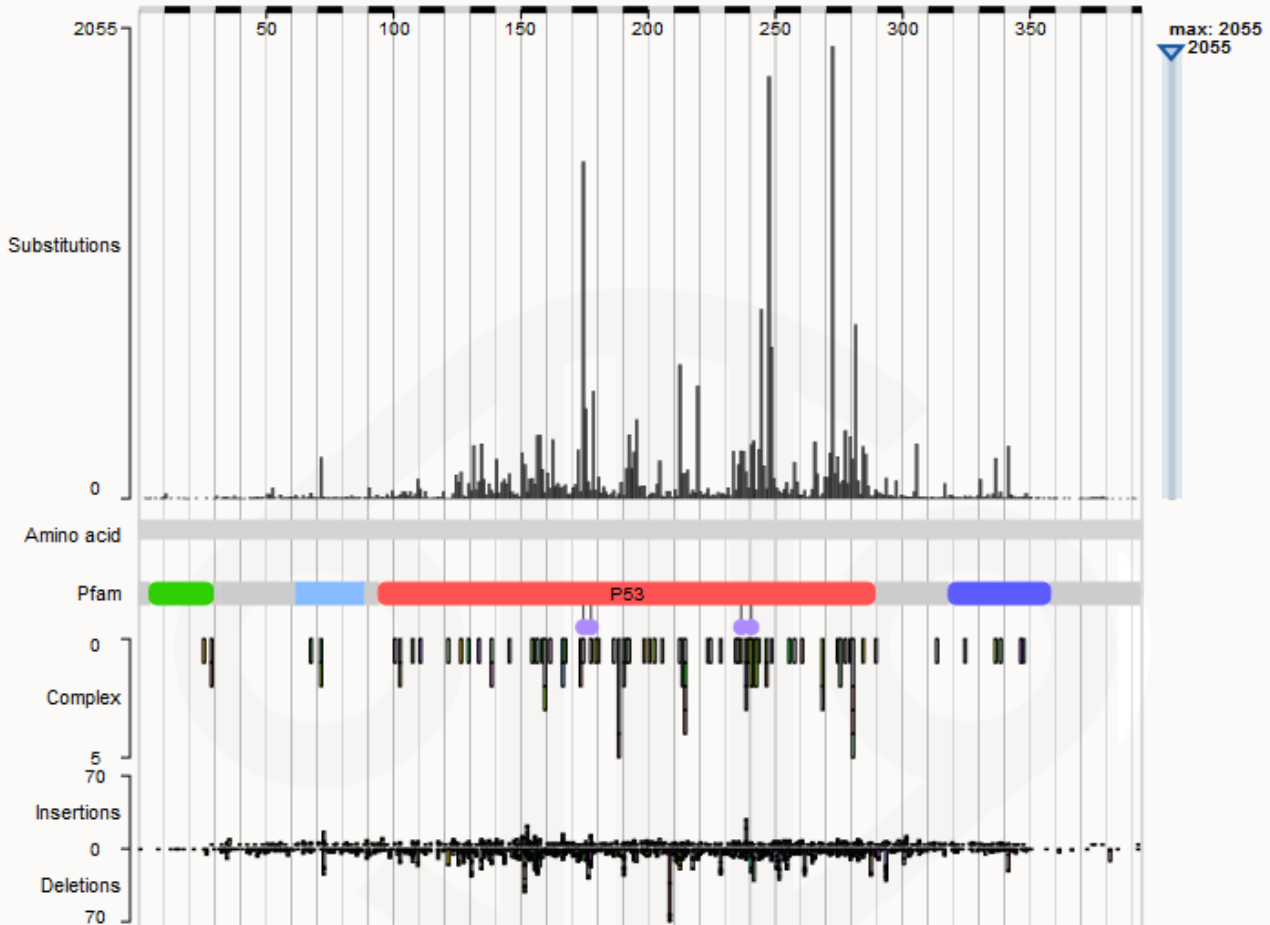
GRCh38

- Gene view
- Overview
- External links
- Drug resistance
- Tissue distribution
- Genome browser
- Mutation distribution
- Variants
- References

Reset page

Gene view

The gene view histogram is a graphical view of mutations across TP53. These mutations are displayed at the length of the gene by default. Restrict the view to a region of the gene by dragging across the histogram to or by using the sliders in the filters panel to the left. Show more



Filters

Show advanced filters

Range Show input fields

1 — 394

1 99 198 296 394

Coordinate system

- (•) Amino-acid
- ( ) cDNA

Apply filters | Reset filters

Filters

Hide advanced filters

Range Show input fields

1 — 394

1 99 198 296 394

Coordinate system

- (•) Amino-acid
- ( ) cDNA

Apply filters | Reset filters

Tissue

Choose site

Choose sub-site

Choose histology

Choose sub-histology

Screen type

- (•) Whole genome screen
- ( ) Targeted screen
- ( ) All

Somatic status

- [ ] Confirmed somatic
- [ ] Previously reported
- [ ] Variant of unknown origin

Sample type

- [ ] Cultured
- [ ] Tumour sample

Mutation impact

Distribution of *TP53* variants across tissues

Generade
Applied Genomics for Life

COSMIC
Catalogue Of Somatic Mutations In Cancer

Projects | Data | Tools | News | Help | About | Genome Version | Search COSMIC... SEARCH | Login

Gene
TP53

- Gene view
- Overview
- External links
- Drug resistance
- Tissue distribution
- Genome browser
- Mutation distribution
- Variants
- References

Reset page

Filters
Show advanced filters

Range Show input fields

1 — 394

1 99 198 296 394

Coordinate system

- Amino-acid
- cDNA

Apply filters | Reset filters

GRCh38 · COSMIC v84

Tissue distribution

The table shows the distribution of mutations across the primary tissue types that are curated by COSMIC. Histograms show the percentage of mutated samples for point mutations, CNV data and gene expression data. Moving your mouse over the histograms will show additional data.

You can see additional information about the data presented here in the help pages.

Show All entries

Search:

Tissue	Point Mutations		Copy Number Variation		Gene Expression		Methylation	
	% Mutated	Tested	Variant %	Tested	% Regulated	Tested	% Diff. Methylated	Tested
Adrenal gland		1183		268		79		-
Autonomic ganglia		994		-		-		-
Biliary tract		1200		-		-		-
Bone		1438		170		-		-
Breast		14227		1544		1104		707
Central nervous system		7881		1093		697		-
Cervix		1632		-		307		-
Endometrium		2066		598		602		-
Eye		295		-		-		-
Fallopian tube		5		-		-		-
Gastrointestinal tract (site indeterminate)		2		-		-		-

Generade
Applied Genomics for Life

Gene
BCR

GRCh38 · COSMIC v83

Gene view
Overview
External links
Drug resistance
Tissue distribution
Genome browser
Mutation distribution
Variants
References
Reset page

Filters

Show advanced filters

Range Show input fields

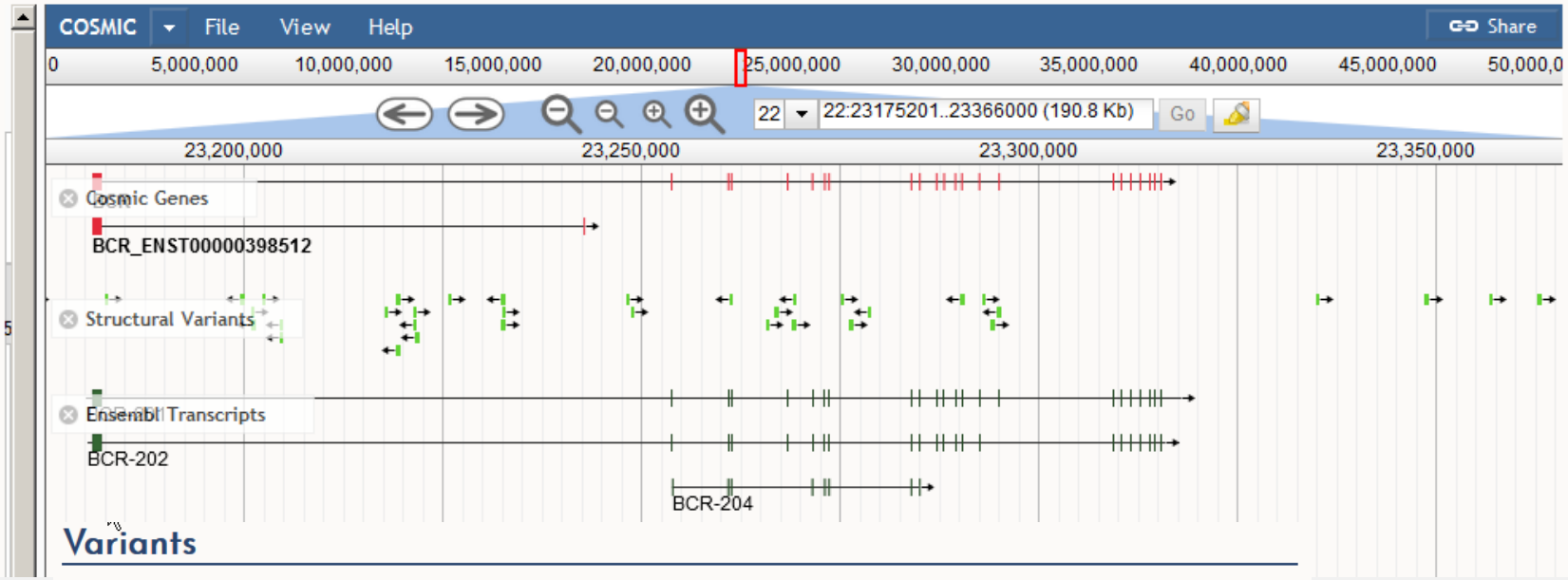
1 — 1272

1 319 637 954 1272

Genome browser

The genome browser shows COSMIC annotations for BCR in a genomic context. Show more

Note that you can also view the genome browser in a separate page.



Variants

Mutations | Fusions | CNV & Expression | Methylation

This tab displays a table of fusions for the selected gene. You can see more information in our help pages.

Show 10 entries Export: CSV TSV Search:

Mutation (Amino Acid)	Mutation ID (COSF)	Count	Mutation Type
BCR{ENST00000305877}:r.1_3458_JAK2{ENST00000381652}:r.1830_5285	COSF757	1	Fusion
BCR{ENST00000305877}:r.1_2030_JAK2{ENST00000381652}:r.2929_5285	COSF758	1	Fusion
BCR{ENST00000305877}:r.1_2030+455_JAK2{ENST00000381652}:r.2626_5285	COSF988	1	Fusion

Summary

An overview of the types of mutation observed.

Colour	Mutation type	Number of samples (%)
	Nonsense substitution	12 (3.74%)
	Missense substitution	196 (61.06%)
	Synonymous substitution	100 (31.15%)
	Inframe insertion	0 (0.00%)
	Frameshift insertion	16 (4.98%)
	Inframe deletion	0 (0.00%)
	Frameshift deletion	4 (1.25%)
	Complex mutation	0 (0.00%)
	Other	0 (0.00%)
	Total unique samples	321

External links

Links to bioinformatics resources that are related to BCR.

OMIM	151410
Transcript	ENST00000305877
Genome Browsers	Ensembl, UCSC
Copy Number	CONAN
NCBI Entrez Gene	n/a
CCDS	CCDS13806.1
UniProt	P11274
Pfam	P11274
TrEMBL	n/a
Atlas Genetic Oncology	BCR
HGNC	1014

COSMIC: *BCR-JAK2* fusion transcript

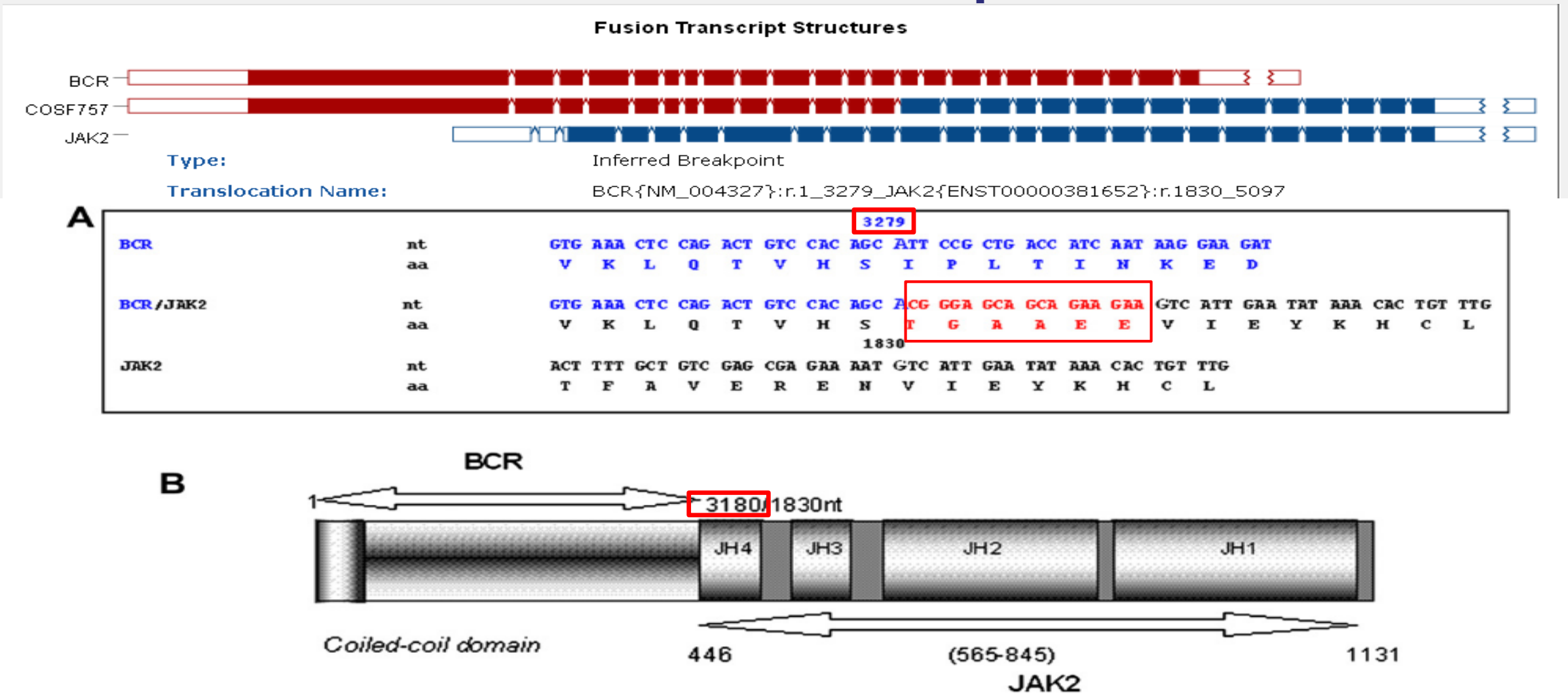
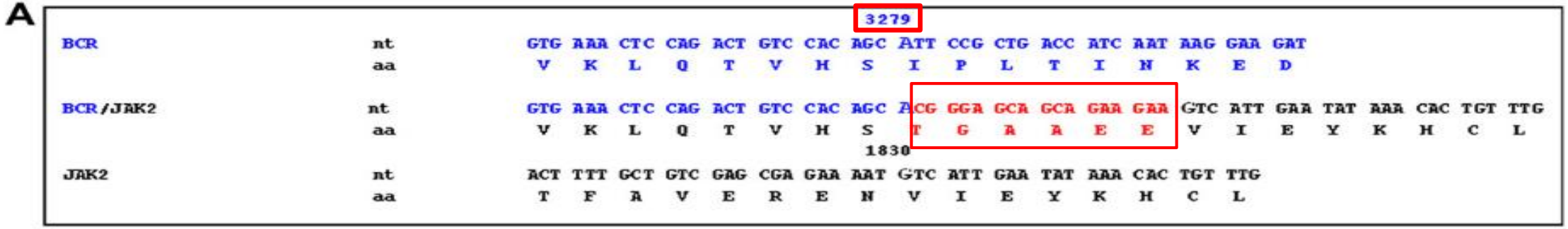


Fig. 4. (A) Sequences of the breakpoint region in the BCR—JAK2 chimeric transcript (nt, nucleotides; aa, amino acids) (17-bp insertion in red). Numbers indicate the position of the breakpoint in the corresponding RefSeq sequence: 3279 for the *BCR* gene (NM_004327) and 1830 for *JAK2* (NM_004972). (B) Schematic representation of the BCR—JAK2 fusion protein.

Cirmena et al. (2008) PMID:18503828

BCR-JAK2 fusion transcript description



Cirmena et al. (2008) PMID:18503828

COSMIC: BCR {NM_004327}:r.1_3279_JAK2{ENST00000381652}:r.1830_5097

BCR transcript NM_004327.3:

Exon number	Start (g.)	Stop (g.)	Start (c.)	Stop (c.)	*BCR* part
1	1	1875	-596	1279	r.-596_2683

JAK2 transcript NM_004972.3:

Exon number	Start (g.)	Stop (g.)	Start (c.)	Stop (c.)	*JAK2* part
1	1	386	-494	-109	
25	3786	5285	3292	*1392	r.1336_*1392

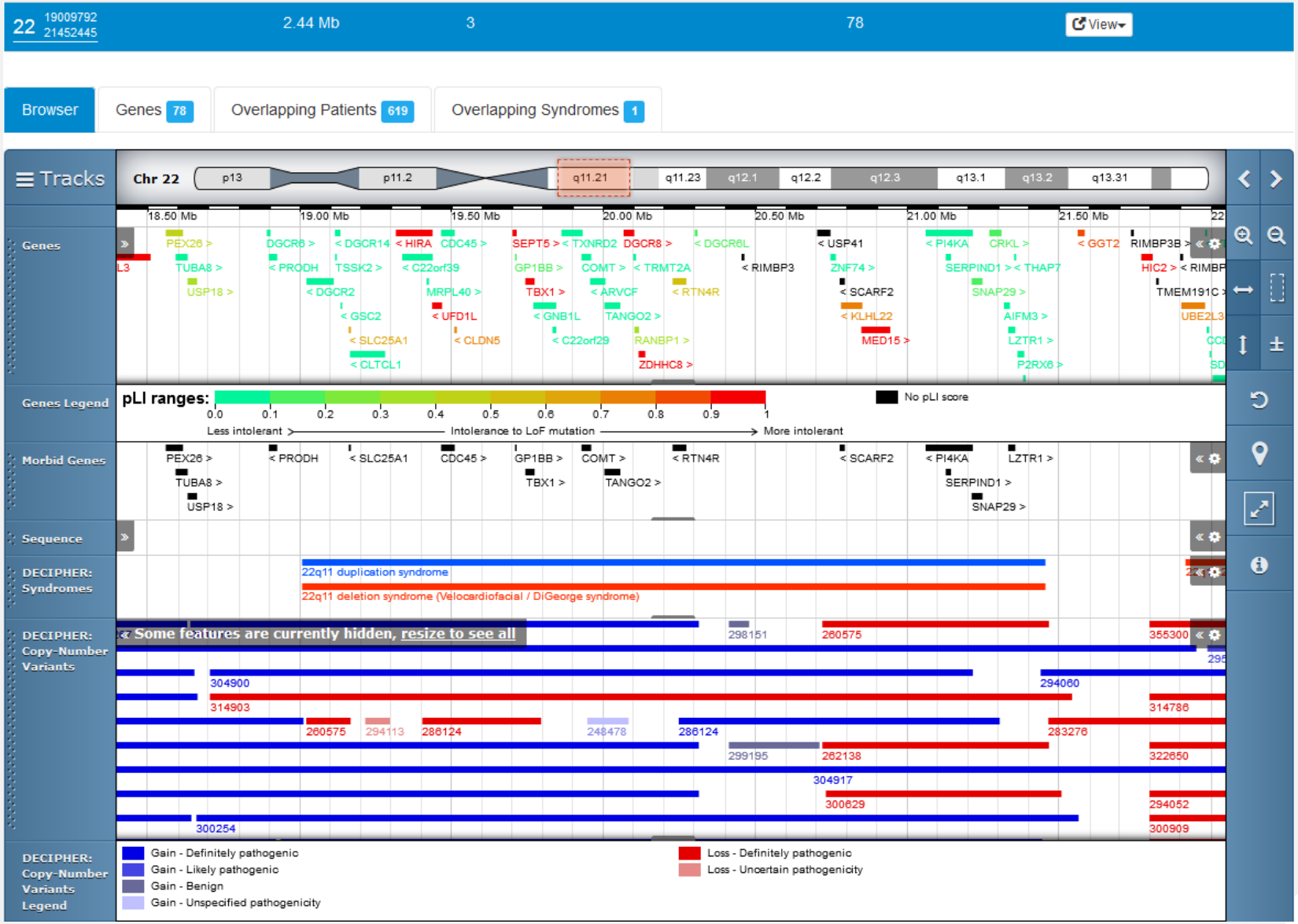
Correct description: NM_004327.3:r.-596_2683::cg gga gca gca gaa gaa::NM_004972.3:r.1336_*1392

22q11 deletion and duplication syndrome

Generade
Applied Genomics for Life

DECIPHER
GRCh37

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



Mostly rearrangements

Copy number variants

Now also other variants

Indication of disease-causing effect

Describing information from databases and tools

- Always:
  - URL (www.lovd.org) and date of visit (22-01-2019)
  - Database-IDs of records with information
  - Version numbers of database releases and software
  - Version number HGVS nomenclature (used by databases or tools)
  - Accession numbers and version numbers of reference sequences, genomes
  - Local references, legacy nomenclature systems
  - References of publications used(PMID, DOI)

Describing information used (or not)

- Use database-IDs of records with information and indicate:
  - Where no information has been found
  - Whether information is (un)reliable:
    - Accession numbers and version numbers absent
    - Use of deviating nomenclature systems
    - No or incomplete supporting evidence

Questions?

Thanks to:

Carli Tops, LUMC
Johan den Dunnen, LUMC
Brandi Kattman, ClinVar, NCBI