Mutation taster

http://www.mutationtaster.org/

Принцип работы

- MutationTaster использует классификатор Байеса, чтобы в конечном итоге предсказать потенциальный риск изменения. Классификатор Байеса получает результат всех тестов и особенностей изменений и вычисляет вероятность патогенности.

Используются 3 модели, нацеленные на различные типы изменений

- 1 – (without_aae (amino acid модели) – не синонимичные и интронные изменения, которые не приводят к изменению аминокислоты;
- 2 – (simple_aae модели) – те, которые ведут к замене / вставке / удалению одной аминокислоты;
- 3 – (complex_aae модель) – более сложных изменений аминокислотной последовательности например, мутации внесения преждевременного стоп-кодона и сдвиги рамки считывания.

- Для обучения классификатора и для того, чтобы программа могла оценить замену были проанализированы данные более чем 390 000 известных мутаций из базы HGMD Professional и более 6 800 000 безвредных SNP и полиморфизмов Indel из проекта 1000 геномов.

mutation t@sting

- documentation | FAQs
- single query
- query chromosomal positions
- QueryEngine
- MutationDistiller (public beta)
- RegulationSpotter (public beta)
- other applications | team
- slides ESHG2017 Copenhagen

Gene npc1 HGNC gene symbol, NCBI Gene ID, Ensembl gene ID show available transcripts clear input

Transcript ENST00000269228 Ensembl transcript ID

Choose the transcript:

- (•) ENST00000269228 (*protein_coding*, 5157 bases) NM_000271
- ( ) ENST00000412552 (*protein_coding*, 3171 bases)
- ( ) ENST00000540608 (*processed_transcript*, 2790 bases)
- ( ) ENST00000588867 (*retained_intron*, 2000 bases)
- ( ) ENST00000591075 (*retained_intron*, 809 bases)
- ( ) ENST00000587223 (*processed_transcript*, 590 bases)
- ( ) ENST00000586718 (*retained_intron*, 574 bases)
- ( ) ENST00000591955 (*retained_intron*, 561 bases)
- ( ) ENST00000587163 (*retained_intron*, 431 bases)
- ( ) ENST00000590301 (*retained_intron*, 410 bases)

Position / snippet refers to ( ) coding sequence (ORF) ( ) transcript (cDNA sequence) (•) gene (genomic sequence)

Alteration

all types by sequence

TATGGAG[a/t]GTGT

enter a few bases around your alteration

options

☐ show nucleotide alignment

Format:

ACTGTC[A/*T*] GTGTF	A substituted by *T*
ACTGTC[AG/*T*] GTGTF	AG substituted by *T*
ACTGTC[ACGT/-] GTGTF	ACGT deleted
ACTGTC[-/*AA*] GTGTF	*AA* inserted

single base exchange by position

enter **position**

and **new base**

insertion or deletion by position

enter **positions** of

...**last wild type base** before alteration

...**first wild type base** after alteration

and the **inserted bases** (if applicable)

Name of alteration if you would like to have a name for this alteration in the output later on, please type in here

continue

2 варианта прогноза.

Polymorphism – полиморфизм

Disease causing – вызывает заболевание

Prediction

disease causing

Model: ***simple_aae***, prob: 0.999999997312517 (explain)

Summary

hyperlink

- **amino acid sequence changed**
- **protein features (might be) affected**
- **splice site changes**

analysed issue	analysis result
name of alteration	no title
alteration (phys. location)	chr18:21153507T>A show variant in all transcripts IGV
HGNC symbol	NPC1
Ensembl transcript ID	ENST00000269228
Genbank transcript ID	NM_000271
UniProt peptide	O15118
alteration type	single base exchange
alteration region	CDS
DNA changes	c.89A>T cDNA.644A>T g.13356A>T
AA changes	E30V Score: 121 explain score(s)
position(s) of altered AA if AA alteration in CDS	30
frameshift	no

known variant

database	homozygous (A/A)	heterozygous	allele carriers
1000G	-	-	-
ExAC	0	1	1

regulatory features

H3K36me3, Histone, Histone 3 Lysine 36 Tri-Methylation
H3K27ac, Histone, Histone 3 Lysine 27 Acetylation
H4K20me1, Histone, Histone 4 Lysine 20 mono-methylation

phyloP / phastCons

	PhyloP	PhastCons
(flanking)	0.627	1
	3.954	1
(flanking)	4.769	1

explain score(s) and/or inspect your position(s) in in UCSC Genome Browser

splice sites

effect	gDNA position	score	wt detection sequence	exon-intron border
Acc increased	13346	wt: 0.55 / mu: 0.67	wt: TGTTTTCACAGTCCTGTGTTTGGTATGGAGAGTGTGGAATT mu: TGTTTTCACAGTCCTGTGTTTGGTATGGAGTGTGTGGAATT	gttt|GGTA
Acc increased	13347	wt: 0.70 / mu: 0.78	wt: GTTTTCACAGTCCTGTGTTTGGTATGGAGAGTGTGGAATTG mu: GTTTTCACAGTCCTGTGTTTGGTATGGAGTGTGTGGAATTG	tttg|GTAT
Donor marginally increased	13358	wt: 0.2785 / mu: 0.3029 (marginal change - not scored)	wt: GAGAGTGTGGAATTG mu: GAGTGTGTGGAATTG	GAGT|gtgg
Donor gained	13349	0.73	mu: TTTGGTATGGAGTGT	TGGT|atgg

distance from splice site	32
Kozak consensus sequence altered?	N/A

Prediction

polymorphism

Model: *simple_aae*, prob: 0.970101697530678 (explain)

Summary

- **amino acid sequence changed**
- **protein features (might be) affected**

analysed issue	analysis result
name of alteration	no title
alteration (phys. location)	chr18:21134845G>A
HGNC symbol	NPC1
Ensembl transcript ID	ENST00000269228
UniProt peptide	O15118
alteration type	single base exchange
alteration region	CDS
DNA changes	c.1430C>T cDNA.1985C>T g.32018C>T
AA changes	T477M Score: 81 explain score(s)
position(s) of altered AA if AA alteration in CDS	477
frameshift	no
dbSNP / TGP / HGMD(public) / ClinVar	no SNPs in altered region found
regulatory features	H3K36me3, Histone, Histone 3 Lysine 36 Tri-Methylation
phyloP / phastCons	PhyloP PhastCons (flanking) 4.153 1 1.655 1 (flanking) 0.337 0.999 explain score(s) and/or inspect your position(s) in in UCSC Genome Browser

splice sites	effect	gDNA position	score	wt detection sequence	exon-intron border
	Acc marginally increased	32008	wt: 0.6887 / mu: 0.7528 (marginal change - not scored)	wt: TCTGCTTGGCCCCTCTTTCACCGTATAACACGAACTGCACC mu: TCTGCTTGGCCCCTCTTTCACCGTATAACATGAACTGCACC	tcac|CGTA
	Donor marginally	32016	wt: 0.5908 / mu: 0.6304 (marginal change -	wt: TATAACACGAACTGC	TAAC|acga

Prediction disease causing

Model: *complex_aae*, prob: 1 (classification due to NMD, real probability is shown anyway) (explain)

Summary hyperlink

- **NMD**
- **amino acid sequence changed**
- **frameshift**
- **protein features (might be) affected**
- **splice site changes**

analysed issue	analysis result
name of alteration	no title
alteration (phys. location)	chr18:21153507_21153507delT
HGNC symbol	NPC1
Ensembl transcript ID	ENST00000269228
Genbank transcript ID	NM_000271
UniProt peptide	O15118
alteration type	deletion
alteration region	CDS
DNA changes	c.89_89delA cDNA.644_644delA g.13356_13356delA
AA changes ★	E30Gfs*29
position(s) of altered AA if AA alteration in CDS	30 (frameshift or PTC - further changes downstream)
frameshift	yes
known variant	Variant was neither found in ExAC nor 1000G. Search ExAC.
regulatory features	H3K36me3, Histone, Histone 3 Lysine 36 Tri-Methylation H3K27ac, Histone, Histone 3 Lysine 27 Acetylation H4K20me1, Histone, Histone 4 Lysine 20 mono-methylation
phyloP / phastCons	PhyloP PhastCons (flanking) 0.627 1 3.954 1 (flanking) 4.769 1 explain score(s) and/or inspect your position(s) in in UCSC Genome Browser

splice sites

effect	gDNA position	score	wt detection sequence	exon-intron border
Acc increased	13347	wt: 0.70 / mu: 0.86	wt: GTTTTCACAGTCCTGTGTTTGGTATGGAGAGTGTGGAATTG	tttg|GTAT
			mu: GTTTTCACAGTCCTGTGTTTGGTATGGAGGTGTGGAATTGC	
Acc gained	13353	0.49	mu: ACAGTCCTGTGTTTGGTATGGAGGTGTGGAATTGCATATGG	atgg|AGGT
Donor gained	13352	0.66	mu: GGTATGGAGGTGTGG	TATG|gagg

distance from splice site	32

Prediction

disease causing Model: *simple_aae*, prob: 0.999999999994966 (classification due to ClinVar, real probability is shown anyway) (explain)

Summary

- **amino acid sequence changed**
- **known disease mutation at this position (HGMD CM992942)** ★
- **known disease mutation: rs80358257 (pathogenic)**
- **protein features (might be) affected**
- **splice site changes**

analysed issue	analysis result
name of alteration	no title
alteration (phys. location)	chr18:21118528G>C show variant in all transcripts
HGNC symbol	NPC1
Ensembl transcript ID	ENST00000269228
Genbank transcript ID	NM_000271
UniProt peptide	O15118
alteration type	single base exchange
alteration region	CDS
DNA changes	c.3019C>G cDNA.3574C>G g.48335C>G
AA changes	P1007A Score: 27 explain score(s)
position(s) of altered AA if AA alteration in CDS	1007
frameshift	no
known variant	Reference ID: rs80358257

database	homozygous (C/C)	heterozygous	allele carriers
1000G	0	1	1
ExAC	0	14	14

★ **known disease mutation: rs80358257 (pathogenic for *Niemann-Pick disease type C1*)** dbSNP NCBI variation viewer
known disease mutation at this position, please check HGMD for details (HGMD ID CM992942)

mutation t@sting

Gene npc1 HGNC gene symbol, NCBI Gene ID, Ensembl gene ID show available transcripts clear input

Transcript ENST00000269228 Ensembl transcript ID

Position / snippet refers to ◉ coding sequence (ORF) ○ transcript (cDNA sequence) ○ gene (genomic sequence)

Alteration

all types by sequence

enter a few bases around your alteration

Format:

ACTGTC[A/*T*] GTGTF	A substituted by *T*
ACTGTC[AG/*T*] GTGTF	AG substituted by *T*
ACTGTC[ACGT/-] GTGTF	ACGT deleted
ACTGTC[-/*AA*] GTGTF	*AA* inserted

options
☐ show nucleotide alignment

single base exchange by position

enter **position** 3019
and **new base** G

c.3019C>G

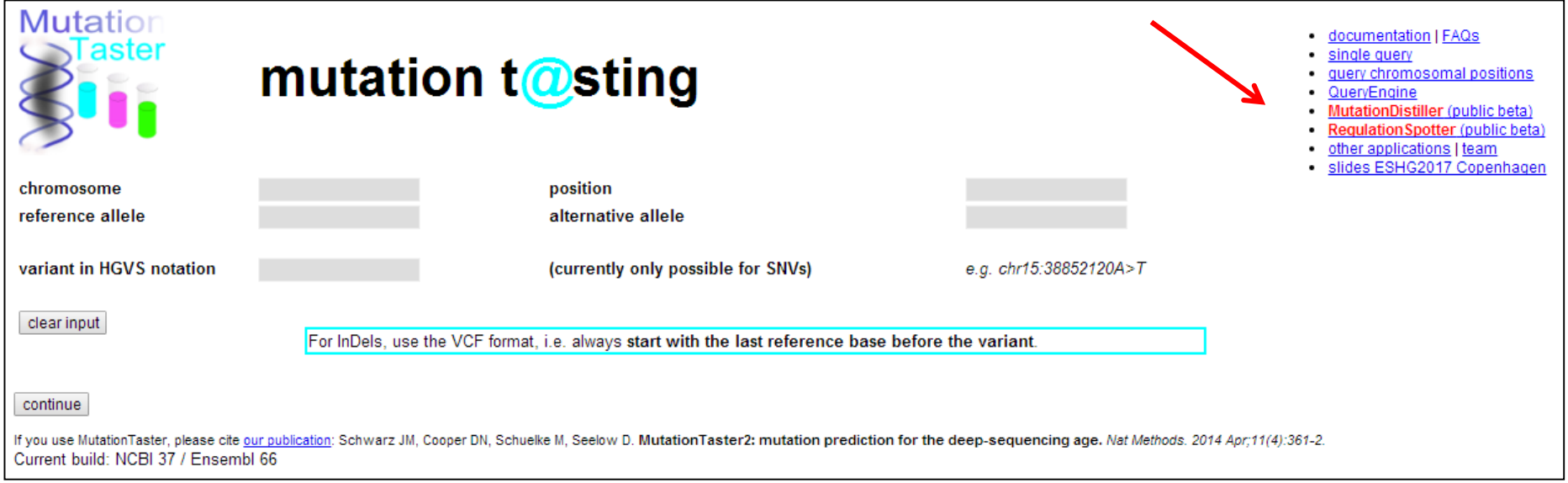
insertion or deletion by position

enter **positions** of
...**last wild type base** before alteration
...**first wild type base** after alteration
and the **inserted bases** (if applicable)

Name of alteration if you would like to have a name for this alteration in the output later on, please type in here

continue

Current build: NCBI 37 / Ensembl 69


Mutation Taster
mutation t@sting
documentation | FAQs
single query
query chromosomal positions
QueryEngine
MutationDistiller (public beta)
RegulationSpotter (public beta)
other applications | team
slides ESHG2017 Copenhagen
chromosome
position
reference allele
alternative allele
variant in HGVS notation
(currently only possible for SNVs)
e.g. chr15:38852120A>T
clear input
For InDels, use the VCF format, i.e. always start with the last reference base before the variant.
continue
If you use MutationTaster, please cite our publication: Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014 Apr;11(4):361-2.
Current build: NCBI 37 / Ensembl 66

mutation t@sting

QueryEngine

We offer automated MutationTaster analysis of variants from *Next Generation Sequencing* projects. Variants must be in VCF format and refer to GRCh37 / hg19. After your VCF file has been analysed, the link to download the results (archived as .zip) will be send via E-mail to you. For this reason, you have to provide a valid E-mail address. Look up more details in the documentation.

VCF file Выберите файл Файл не выбран sample file clear input

Please zip or gzip large files!

Format:

```
#CHROM   POS    ID  REF  ALT  QUAL  FILTER  INFO                                                                  FORMAT       SAMPLE
chr1     10199  .   A    C    4.77  .       DP=2;AF1=0.5003;CI95=0.25,0.75;DP4=1,0,0,1;MQ=60;FQ=-3.1;PV4=1,1,1,1  GT:PL:DP:GQ  0/1:33,0,28:2:30
```

(tab delimited) The coordinates **must** refer to GRCh37 (also called hg19).

Project name

Unfortunately, E-Mail notification does currently not work. We are sorry for the inconvenience and try to solve the problem as soon as possible

E-mail address

generate HTML files (slower)

Analysis settings

search for homozygous variants yes

combine neighbouring variants yes

filter against TGP homozygous in 4 or more TGP samples

heterozygous in 20 or more TGP samples

minimum coverage 4

Submit

queue status 2018-06-04 00:24 CEST (refresh)

low load - our bored server is looking forward to analysing your data

0 jobs running, 0 queued, 1566.4 millions alterations analysed.

DB queries	free slots (0 running, 16 of 16 available - at 0% capacity)
large jobs	free slots (0 running, 20 of 20 available - at 0% capacity)

- analyse complete VCF
  - ...but only exons with 10 bases intron flanking
- analyse variants on chr
  - ...but only exons with 10 bases intron flanking
- analyse custom regions (select to enter)
- exclude custom regions (select to enter)

PolyPhen-2

http://genetics.bwh.harvard.edu/pph2/

PolyPhen-2

PolyPhen-2 (от слов Polymorphism «полиморфизм» и Phenotyping «фенотипирование») - программа, которая прогнозирует возможное влияние аминокислотной замены на структуру и функцию белка. Имеет высокой качество выравнивания в эволюционном ряду.

Две пары наборов данных были использованы для тестирования и прогнозирования в PolyPhen-2.

-Первая пара **HumDiv**, была составлена из всех повреждающих аллелей с известными воздействиями на функцию белка, включая заболевания, присутствующие в базе данных UniProtKB, вместе с различиями между белками человека и его близкородственными гомологами млекопитающими.

HumDiv модель следует использовать для оценки редких аллелей в локусах потенциально участвующих в сложных фенотипах, при плотной картографированности регионов, а также при анализе естественного отбора, где даже мягко вредные аллели должны быть рассмотрены как патогенные.

-Вторая пара, **HumVar**, состоял из всех патогенных мутаций человека из базы UniProtKB, вместе с общечеловеческими SNP (с частотой > 1%).

HumVar модель следует использовать в диагностике Менделевских заболеваний, поскольку требует сравнение искомых мутаций со всеми остальных человеческими изменениями патогенными или полиморфными. Именно на графу с данными HumVar необходимо обращать тщательное внимание.

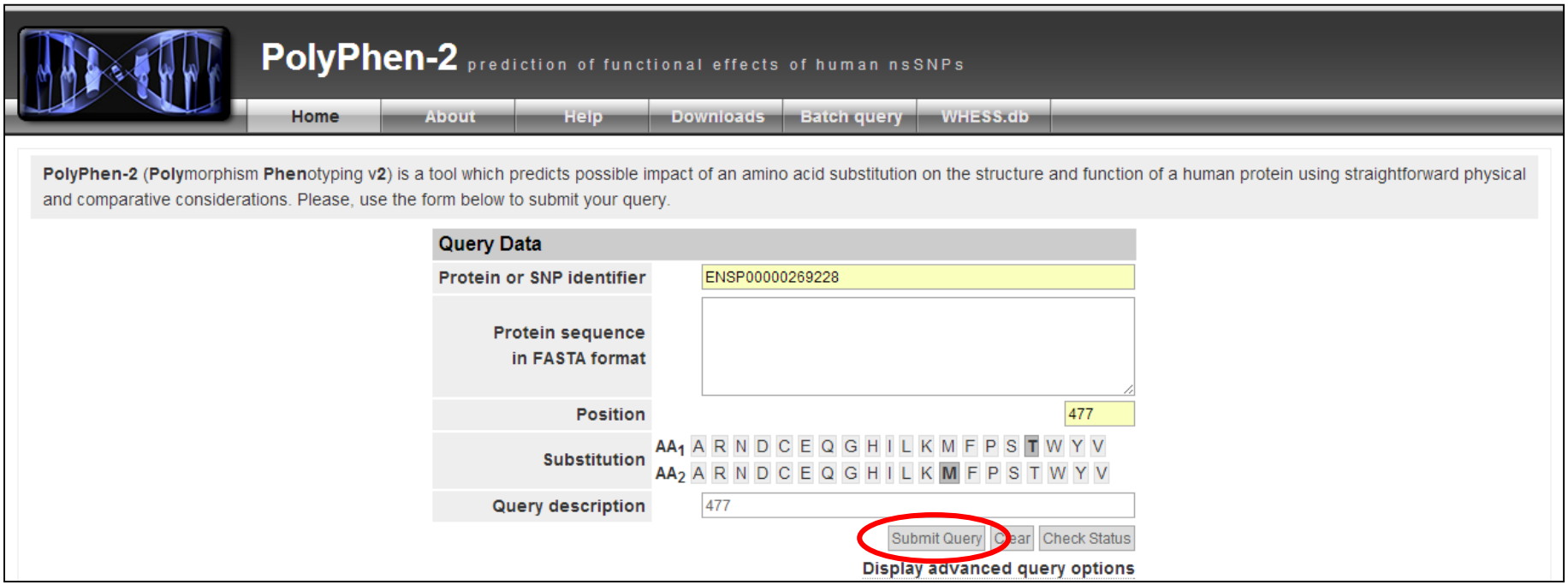
PolyPhen-2

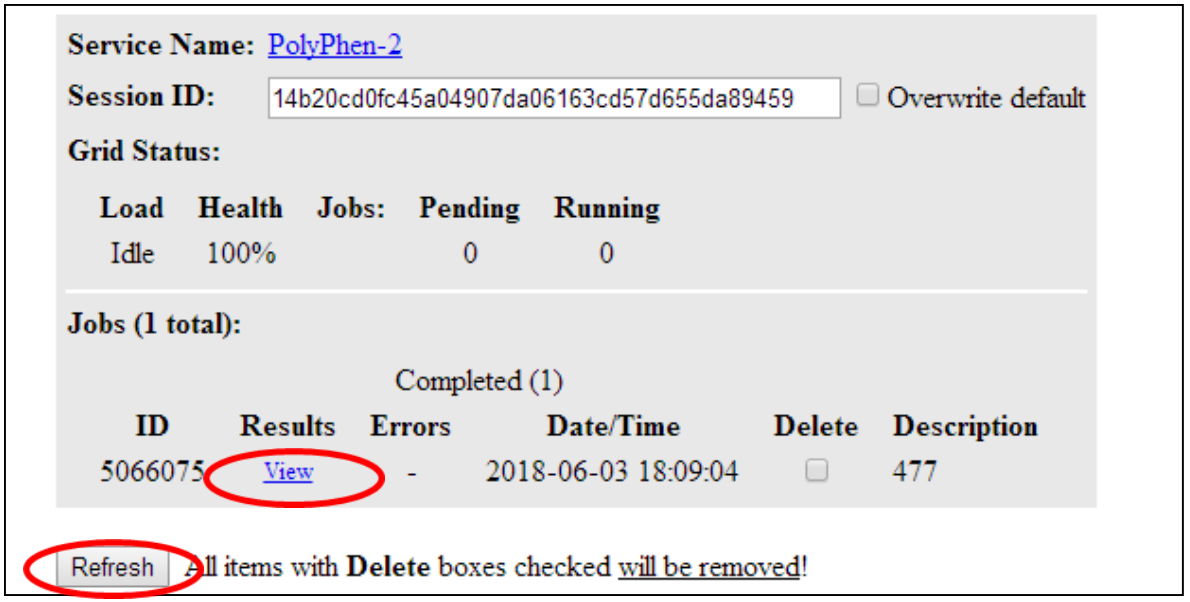
Benign – доброкачественная замена

Possibly damaging – возможно повреждающая (меньшая уверенность)

Probably damaging – вероятно повреждающая (большая уверенность)

Unknown – неизвестно


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.
Query Data
Protein or SNP identifier
ENSP00000269228
Protein sequence in FASTA format
Position
477
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
477
Submit Query
Clear
Check Status
Display advanced query options


Service Name: PolyPhen-2
Session ID: 14b20cd0fc45a04907da06163cd57d655da89459
Overwrite default
Grid Status:
Load Health Jobs: Pending Running
Idle 100% 0 0
Jobs (1 total):
Completed (1)
ID Results Errors Date/Time Delete Description
5066075 View - 2018-06-03 18:09:04 477
Refresh
All items with Delete boxes checked will be removed!

PolyPhen-2 report for O15118 T477M

Query

Protein Acc	Position	AA_1	AA_2	Description
O15118	477	T	M	Canonical; RecName: Full=Niemann-Pick C1 protein; Flags: Precursor; Length: 1278

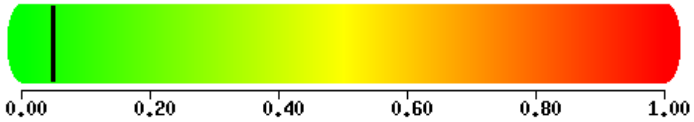
Results

Prediction/Confidence

PolyPhen-2 v2.2.2r398

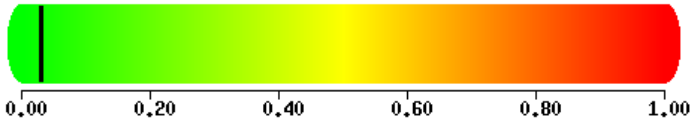
HumDiv

This mutation is predicted to be **BENIGN** with a score of **0.049** (sensitivity: **0.94**; specificity: **0.83**)



HumVar

This mutation is predicted to be **BENIGN** with a score of **0.030** (sensitivity: **0.94**; specificity: **0.59**)



Details

Multiple sequence alignment

UniProtKB/UniRef100 Release 2011_12 (14-Dec-2011)

```
QUERY                D-----NE-------TVTLQDICLAPLSPY------NTNCTIL-SVLNYFQNSHS-VLDHK--KGDD------FF-
sp|F7INK8#1          N-----NE-------TVTLQDICLAPLSPY------NTNCTIM-SVLNYFQNSHS-VLDHK--IGDD------FF-
sp|Q8MKD8#1          N-----NE-------TVTLQDICVAPLSPY------NKNCTIL-SVLNYFQNSHS-MLDHE--IGDD------FF-
sp|Q9N0Q0#1          N-----NE-------TVTLQDICVAPLSPY------NKNCTIL-SVLNYFQNSHS-MLDHE--IGDD------FF-
sp|Q8MI49#1          N-----NE-------TVTLQDICVAPLSPY------NKNCTIL-SVLNYFQNSHS-MLDHE--IGDD------FF-
sp|G1M5K1#1          N-----NE-------TVTLQDICVAPLSPY------NKNCTIL-SVLNYFQNSHS-LLDHK--IGDD------FF-
sp|F6XL99#1          N-----NE-------TVTLQDICLAPLSPY------NKNCTIM-SVLNYFQNSHS-MLDHK--VEDD------FF-
sp|UPI00017966B6#1   N-----NE-------TVTLQDICLAPLSPY------NKNCTIM-SVLNYFQNSHS-MLDHK--VEDD------FF-
sp|G3TMS9#1          N-----NQ-------TVMLRDICLAPLSPY------NKNCTIL-SVLNYFQNSHS-VLDHK--VGDE------FY-
sp|F1SBB5#1          N-----NE-------TVTLQDICLAPLSPY------NKNCTIL-SVLNYFQNSHS-VLDHQ--VGDF------FF-
sp|G1SY19#1          N-----NE-------TVTLQDICLAPLSPY------NKNCTIL-SVLNYFQNSHS-MLDHE--QGDD------FF-
sp|P56941#1          N-----NE-------TVTLQDICLAPLSPY------NKNCTIL-SVLNYFQNSHS-VLDHQ--VGDF------FF-
sp|Q9GK52#1          N-----NE-------TVTLQDICVAPLSPY------NKNCTIM-SVLNYFQNSHS-MLDHK--IGDD------FY-
sp|UPI0002233BEE#1   N-----NQ-------TVMLRDICLAPLSPY------NKNCTIL-SVLNYFQNSHS-VLDHK--VGDE------FY-
sp|B0JYK2#1          N-----NE-------TVTLRDICVAPLSPY------NQNCTIL-SVLNYFQNSHS-VLDHQ--VGDD------FF-
sp|F1PB50#1          N-----NE-------TVTLQDICVAPLSPY------NKNCTIM-SVLNYFQNSHS-MLDHK--IGDD------FY-
sp|F1PSW3#1          N-----NE-------TVTLQDICVAPLSPY------NKNCTIM-SVLNYFQNSHS-MLDHK--IGDD------FY-
sp|Q9GLC9#1          N-----NE-------TVTLRDICVAPLSPY------NQNCTIL-SVLNYFQNSHS-VLDHQ--VGDD------FF-
```

Shown are 75 amino acids surrounding the mutation position (marked with a black box). An interactive version of the complete alignment is also available.

PolyPhen-2 report for O15118 A558S

Query

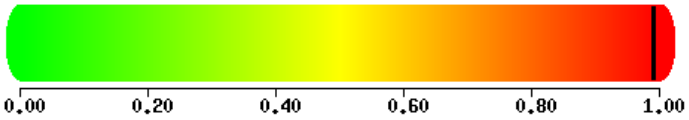
Protein Acc	Position	AA_1	AA_2	Description
O15118	558	A	S	Canonical; RecName: Full=Niemann-Pick C1 protein; Flags: Precursor; Length: 1278

Results

Prediction/Confidence *PolyPhen-2 v2.2.2r398*

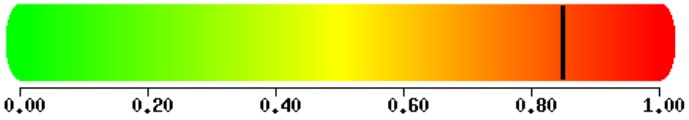
HumDiv

This mutation is predicted to be **PROBABLY DAMAGING** with a score of **0.990** (sensitivity: **0.72**; specificity: **0.97**)



HumVar

This mutation is predicted to be **POSSIBLY DAMAGING** with a score of **0.848** (sensitivity: **0.73**; specificity: **0.88**)



Details

Multiple sequence alignment *UniProtKB/UniRef100 Release 2011_12 (14-Dec-2011)*

Shown are 75 amino acids surrounding the mutation position (marked with a black box). An interactive version of the complete alignment is also available.

SIFT+PROVEAN

http://provean.jcvi.org/index.php

SIFT

SIFT прогнозирует, влияет ли замена (substitution) аминокислоты на функцию (function) белка. Предсказание основано на степени сохранения аминокислотных остатков в эволюционных последовательностях. SIFT основан на предпосылке, что эволюция белка коррелирует с функцией белка. Программа получает запрашиваемую последовательность и использует множественную информацию выравнивания.

SIFT проверка является многоступенчатой процедурой, при которой (1) происходит поиск подобных последовательностей генов, (2) выбирается тесно связанные последовательности, которые могут имеют схожую белковую функцию, (3) получают выравнивание этих выбранных последовательностей, и (4) вычисляют вероятности, используя специальную формулу.

TOLERATED – толерантная, допустимая, переносимая, терпимая
DAMAGING – повреждающая замена
N/A – нет данных/не известно

Позиции с вероятностью менее 0,05 – **DAMAGING**
Позиции с вероятностью больше или равны 0,05 – **TOLERATED**

J. Craig Venter INSTITUTE

SIFT

JCVI Home | SIFT Home | Help | Team | Contact us

- → SIFT Home
- → Help
- → Contact us

Code release
License
Source Code JCVI-SIFT v. 1.03
Code & exe (Sun, Linux)

FTP download
SIFT Human DB (release 63)
SIFT dbSNP DB (build 132)

Related links
Human genome assembly GRCh37
Ensembl annotation release 63
NCBI dbSNP Build 132
NCBI BLink

Updates
Aug 2011: SIFT Human DB updated to support GRCh37 Ensembl release 63

Apr 2011: SIFT dbSNP DB updated to support NCBI dbSNP build 132

SIFT predicts whether an **amino acid substitution affects protein function**. SIFT prediction is based on the degree of conservation of amino acid residues in sequence alignments derived from closely related sequences, collected through PSI-BLAST. SIFT can be applied to naturally occurring **nonsynonymous polymorphisms or laboratory-induced missense mutations.**

** **PROVEAN project** New **

Visit our new PROVEAN project to get functional predictions from multiple tools. We welcome your feedback or questions.

New features in PROVEAN Human Genome Variants DB:

- Single submission returns functional predictions from **SIFT** and **PROVEAN**. PROVEAN is a new prediction tool which works for **both SNPs and indels**. (Choi et al., 2012, PLOS ONE)
- Updated versions of Ensembl gene annotation (GRCh37 Ensembl 66) and NCBI dbSNP database (Build 137).
- New database structure to support fast retrieval for genome-wide analysis.

Human Genome DB	Tool Description
SIFT/PROVEAN Human SNPs	Get **SIFT** and **PROVEAN** predictions for SNPs and indels (Ensembl 66) (Sample format)
SIFT Human SNPs	Get SIFT predictions for nonsynonymous SNPs (Ensembl 63) (Sample format)
	Other human genome tools: • Restrict to Coding Variants (Sample format) • Classify Human indels (Sample format)
SIFT Human Protein DB	**Tool Description (Ensembl 63)**
SIFT Human Protein	Get SIFT predictions for nonsynonymous AA substitutions (Ensembl ENSP ID)
SIFT dbSNP DB	**Tool Description (dbSNP Build 132)**
SIFT dbSNP rs IDs	Get SIFT predictions for dbSNP SNPs including non-human species (NCBI rs ID)
SIFT dbSNP Protein	Get SIFT predictions for dbSNP proteins including non-human species (RefSeq ID or GI number)
SIFT Single Protein Tools	**Tool Description**
SIFT BLink	Run SIFT analysis on single protein using precomputed BLAST from NCBI BLink (RefSeq ID or GI number)

Referencing SIFT

Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* *2009*;4(7):1073-81. PubMed PDF

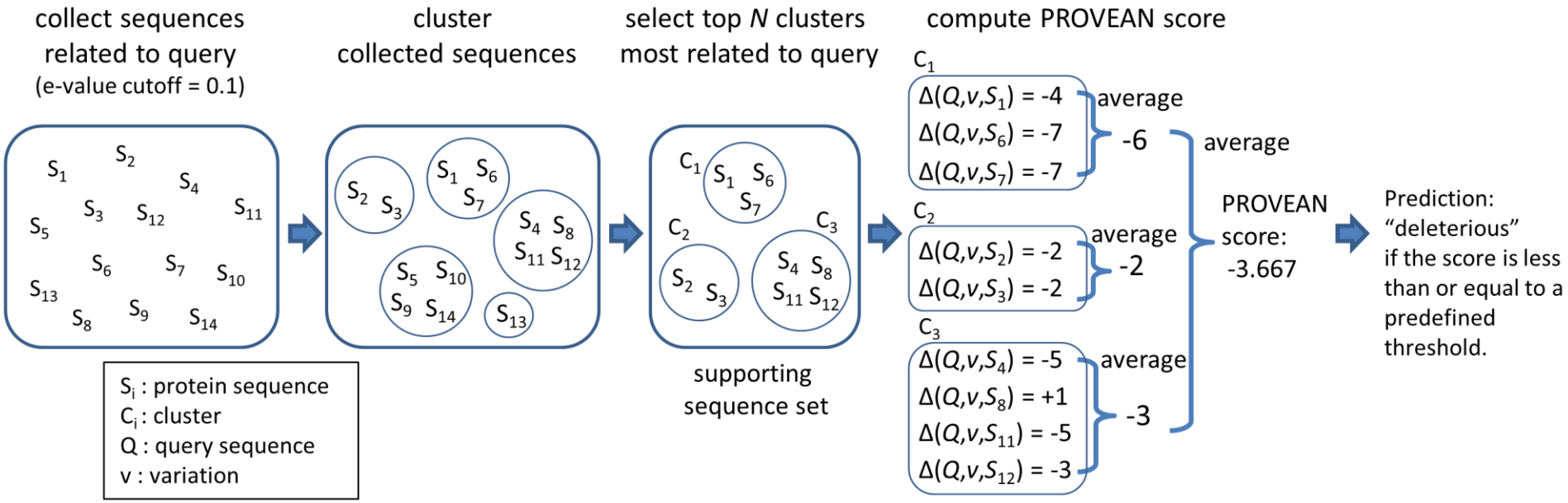
Ng PC, Henikoff S. Predicting the Effects of Amino Acid Substitutions on Protein Function *Annu Rev Genomics Hum Genet.* 2006;7:61-80. PubMed PDF Supp

Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003 Jul 1;31(13):3812-4. PubMed

Ng PC, Henikoff S. Accounting for Human Polymorphisms Predicted to Affect Protein Function. *Genome Res.* 2002 Mar;12(3):436-46. PubMed Data

Ng PC, Henikoff S. Predicting Deleterious Amino Acid Substitutions. *Genome Res.* 2001 May;11(5):863-74. PubMed

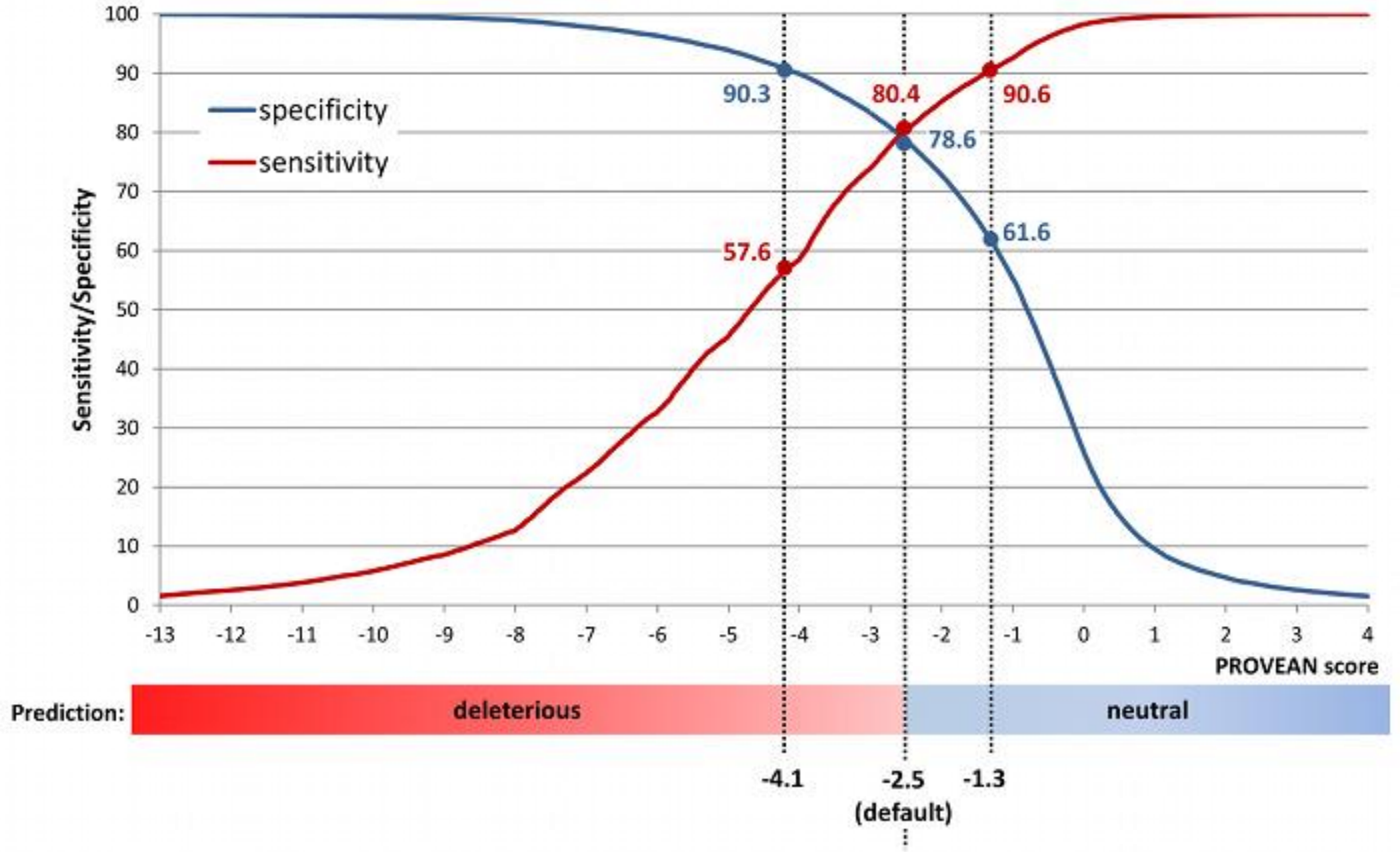
PROVEAN (Protein Variation Effect Analyzer)



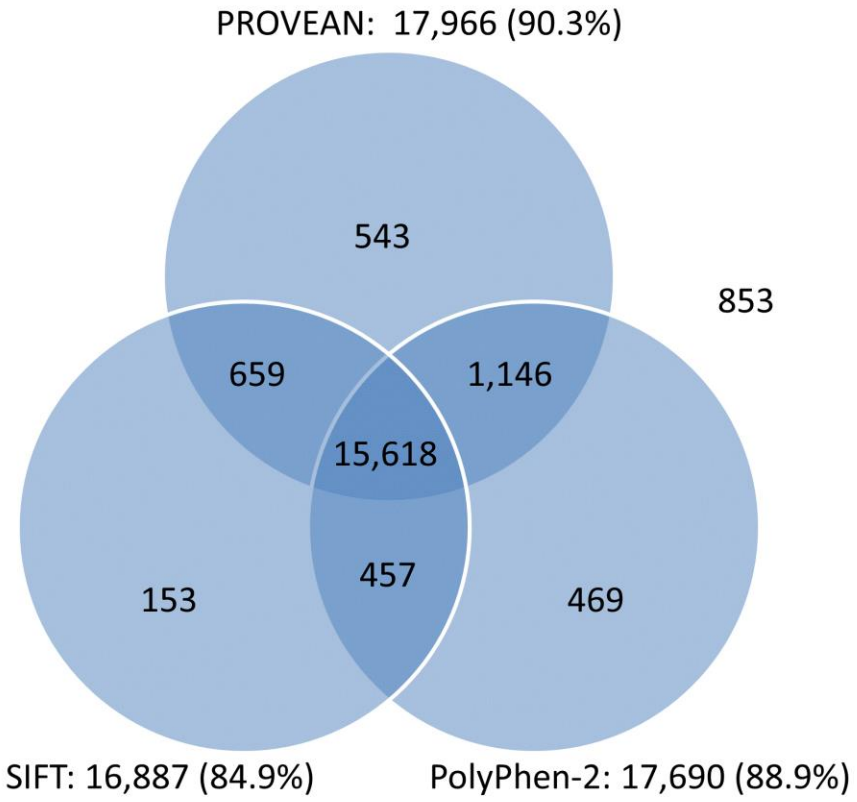
Кратко. 30 групп близкородственных последовательностей образуют опорную последовательность, которая будет использоваться для генерации прогноза.
PROVEAN считает баллы внутри каждого кластера по количеству схожих аминокислот. Оценки усредняются внутри и между кластерами для генерации окончательного счета.

Счет равен или ниже -2,5 - **DELETERIOUS** – **вредный** эффект
выше -2,5 – **NEUTRAL** – **нейтральный** эффект

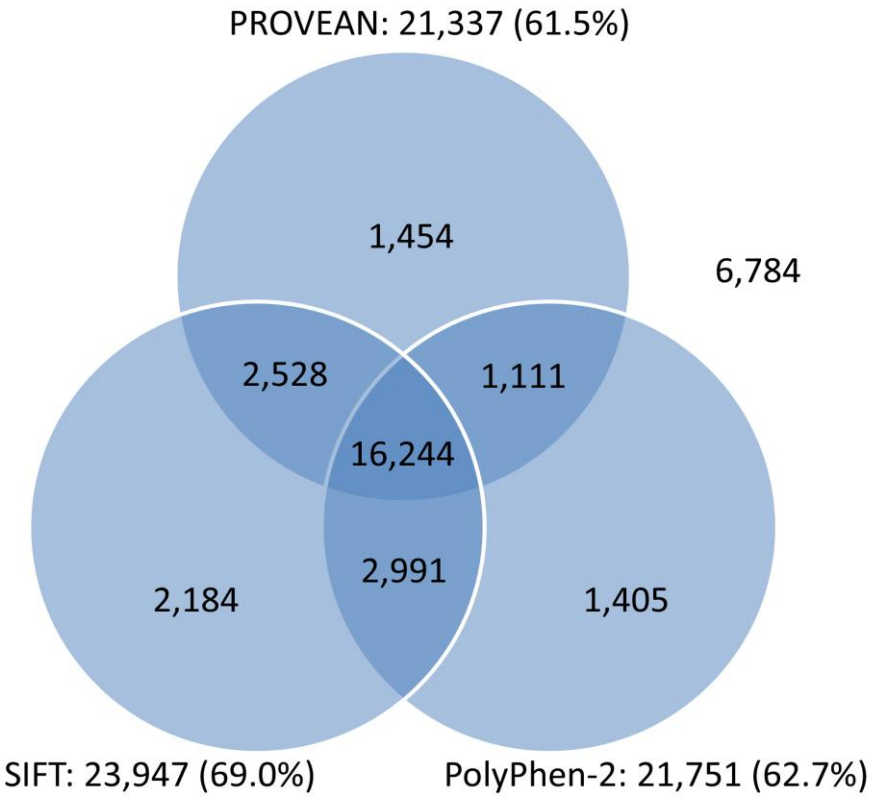
PROVEAN


specificity
sensitivity
Sensitivity/Specificity
100
90
80
70
60
50
40
30
20
10
0
90.3
80.4
90.6
78.6
61.6
57.6
-13
-12
-11
-10
-9
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
PROVEAN score
Prediction:
deleterious
neutral
-4.1
-2.5
(default)
-1.3

PROVEAN



19,898 disease variants



34,701 common polymorphisms

- PROVEAN Tools
  - PROVEAN Protein
  - PROVEAN Protein Batch
    - Human
    - Mouse
  - PROVEAN Genome Variants
    - Human
    - Mouse
- About
- FAQ
- News
- Download
- Help
- Contact Us
- Related Links

PROVEAN (**Pro**tein **V**ariation **E**ffect **An**alyzer) is a software tool which predicts whether an amino acid substitution or indel has an impact on the biological function of a protein.

PROVEAN is useful for filtering sequence variants to identify nonsynonymous or indel variants that are predicted to be functionally important.

The performance of PROVEAN is comparable to popular tools such as SIFT or PolyPhen-2 [1]. Read more.

A fast computation approach to obtain pairwise sequence alignment scores enabled the generation of precomputed PROVEAN predictions for 20 single AA substitutions and a single AA deletion at every amino acid position of all protein sequences in human and mouse [2].

This work is funded by the National Institutes of Health [grant number 5R01HG004701-04].

References:

1. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS ONE 7(10): e46688.
2. Choi Y (2012) A Fast Computation of Pairwise Sequence Alignment Scores Between a Protein and a Set of Single-Locus Variants of Another Protein. *In Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine* (BCB '12). ACM, New York, NY, USA, 414-417.
(* This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in ACM BCB '12. http://doi.acm.org/10.1145/2382936.2382989)

PROVEAN web server functions are currently using PROVEAN v1.1.3.

PROVEAN Tool	Species	Description
PROVEAN Protein	Any species	This tool provides PROVEAN prediction for a protein sequence from any organisms. [details] • Input: **A protein sequence from any organism** and amino acid variants of interest. See example. • Output: PROVEAN scores and predictions. See example.
PROVEAN Protein Batch	- Human - Mouse	This tool provides PROVEAN and SIFT predictions for a list of protein variants. [details] • Input: A list of **protein variants**. See example. • Output: Scores and predictions from PROVEAN and SIFT. See example.
PROVEAN Genome Variants	- Human - Mouse	This tool provides PROVEAN and SIFT predictions for a list of genome variants. It is based on the assembly of the species and the Ensembl genome annotation. [details] • Input: A list of **genomic variants**. See example. • Output: Changes at protein level, their scores and predictions from PROVEAN and SIFT, and accessory information (dbSNP rs IDs, gene description, PFAM domain, GO terms, etc.). See example.

SIFT+PROVEAN

- PROVEAN Tools
  - PROVEAN Protein
  - PROVEAN Protein Batch
    - Human
    - Mouse
  - PROVEAN Genome Variants
    - Human
    - Mouse
- About
- FAQ
- News
- Download
- Help
- Contact Us
- Related Links

PROVEAN HUMAN PROTEIN BATCH

This tool provides PROVEAN prediction for all human proteins and variants. It also shows SIFT predictions when precomputed scores are available.

- Input: A list of **human protein variants**. See example.
- Output: PROVEAN scores and predictions along with available SIFT predictions. See example.

Enter a list of human protein variants

<protein ID>,<position>,<reference amino acids>,<variant amino acids>,<comment>

Paste in your protein variants: [format]

```
ENSP00000269228 477 T M
ENSP00000269228 558 A S
ENSP00000269228 181 A T
```

Example (upload example)

```
ENSP00000224605 63 A S
ENSP00000224605 55 D G
ENSP00000443112 651 T .
ENSP00000359240 59 Q QA
NP_000483.3 508 F .
P13569 508 F .
```

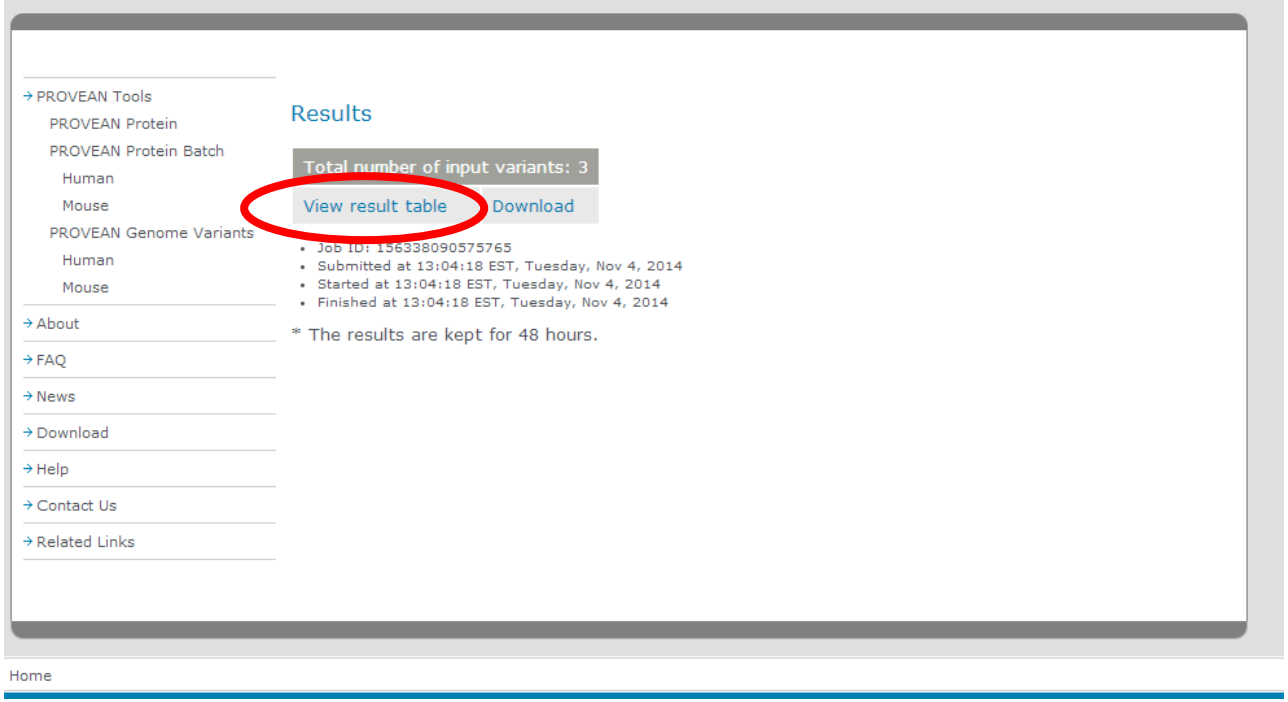
Or upload a file containing variants (1MB limit):

Выберите файл Файл не выбран

Email (optional)

If provided, results will also be sent via email.

Отправить Сбросить



EAN Protein Batch Result (Download)

TION		PROTEIN SEQUENCE CHANGE				PROVEAN PREDICTION				SIFT PREDICTION			
NO.	INPUT	PROTEIN_ID	POSITION	RESIDUE_REF	RESIDUE_ALT	SCORE	PREDICTION (cutoff=-2.5)	#SEQ	#CLUSTER	SCORE	PREDICTION (cutoff=0.05)	MEDIAN_INFO	#SEQ
1	ENSP00000269228 477 T M	ENSP00000269228	477	T	M	-2.37	Neutral	95	30	0.182	Tolerated	2.82	79
2	ENSP00000269228 558 A S	ENSP00000269228	558	A	S	-2.40	Neutral	95	30	0.034	Damaging	2.82	80
3	ENSP00000269228 181 A T	ENSP00000269228	181	A	T	-2.67	Deleterious	95	30	0.042	Damaging	2.82	77

Доп.

#CLUSTER - Количество кластеров используемое для предсказания (по умолчанию 30)

#SEQ - Количество последовательностей используемое для предсказания, у которых в положении имеется искомая аминокислота (чем больше, тем лучше)

MEDIAN_INFO. Диапазон от 0 до 4,32, в идеале число будет между 2,75 и 3,5. Применяется для измерения разнообразия последовательностей, используемых для прогнозирования.

Предупреждение, если больше, чем 3,25, это указывает, что прогноз был основан на тесно связанных последовательностях без необходимого разнообразия.

SIFT+PROVEAN

- PROVEAN Tools
  - PROVEAN Protein
  - PROVEAN Protein Batch
    - Human
    - Mouse
  - PROVEAN Genome Variants
    - Human
    - Mouse
- About
- FAQ
- News
- Download
- Help
- Contact Us
- Related Links

PROVEAN HUMAN GENOME VARIANTS

This tool provides PROVEAN and SIFT predictions for a list of human genome variants.

- Input: A list of **human genomic variants**. See example.
- Output: PROVEAN scores and predictions along with available SIFT predictions. See example.

Step 1. Enter a list of genomic coordinates and variants

<chromosome>,<position>,<reference allele>,<variant allele>

Paste in your coordinates and variants: [format]

Example (upload example)

```
1,100382265,C,G,user comment 1
1,100380997,A,G,user comment 2
22,30163533,A,C
X,12905093,A,T
2,230633386,G,C
1,100382265,C,A
7,117199641,ATCA,.
7,117199647,TTT,.
10,50184923,TGG,.
12,121438957,ACC,.
1,43217995,G,GCCA
10,102762472,G,GGCG
9,117856130,T,G
9,117856135,C,G
```

Or upload a file containing variants (5MB limit):

Выберите файл Файл не выбран

Step 2. Select gene annotation (optional)

- ☐ Ensembl Gene ID
- ☐ Associated Gene Name
- ☐ Ensembl Transcript ID
- ☐ Transcript Status
- ☐ Gene Description
- ☐ % GC Content
- ☐ Chromosome band
- ☐ Ensembl Protein Family ID
- ☐ Ensembl Family Description
- ☐ UniProt/SwissProt ID
- ☐ RefSeq Protein ID
- ☐ MIM Disease Accession
- ☐ PFAM ID
- ☐ TIGRFam ID
- ☐ Interpro ID
- ☐ GO Term Accession
- ☐ GO Slim GOA Accession

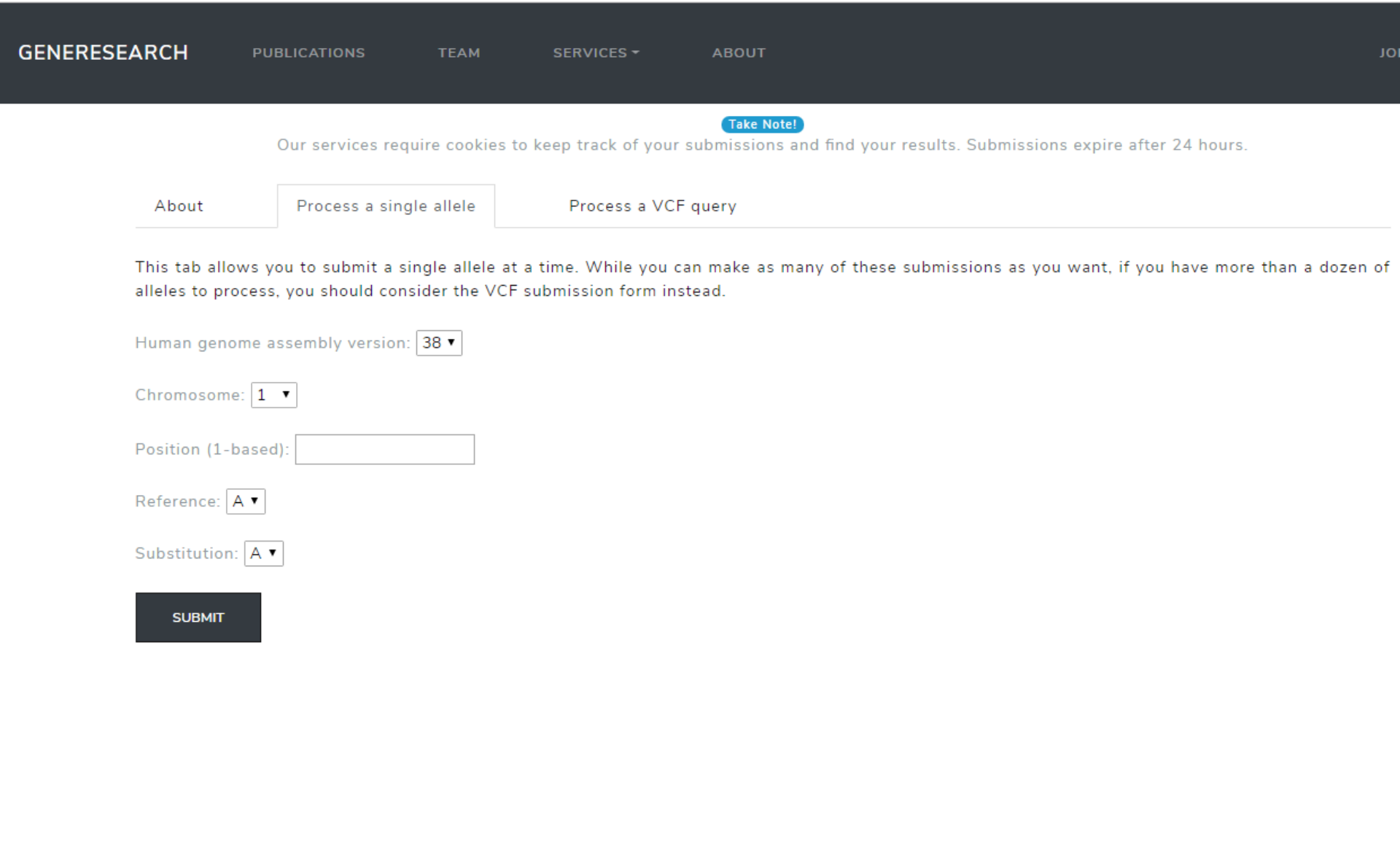
Parameters

Assembly/Annotation: Human GRCh37 Ensembl 66 ▼

Email (optional)

GENERESEARCH

http://score.generesearch.ru/services/badmut/



Анализ сайта сплайсинга

Human Splicing Finder

http://www.umd.be/HSF3/

Start an Analysis

Select an analysis type

Analyze mutation(s) ▾

---------[Select]---------
Analyze a sequence
Analyze mutation(s)
Branch point sequence analysis
Splice site analysis
Multiple transcript analysis

Ensembl database release: **75** (February 2014)

the longest transcript: Yes No

e analyzed sequence: Yes No

More database options

Gene Name (e.g. DMD) ▾ NPC2

Paste your mutant list (e.g. c.4250T>A)

```
c.441+1G>A
```

For more information on how to fill in this form, please refer to the help located at the bottom of this page.

Show advanced parameters

Select all or a subset of matrices and set their thresholds to display Splicing Elements.

Proceed to analysis! **Clear and Abort**

Detailed help

NPC2 c.441+1G>A

Sequences

Reference sequence

NPC2 Gene > ENST00000555619 Transcript > Exon number: **4** (78 bp) + **100** intronic nucleotides at exon ends

```
  1 gaagtttgct actgtacatc taggattcat agttaaacca attatggata gaagtcaggt cctattttca ttttctcaa ttatttttct ctttctccag
101 ATAAAACTGG TGGTGGAGTG GCAACTTCAG GATGACAAAA ACCAAAGTCT CTTCTGCTGG GAAATCCCAG TACAGATCgt aagtctatct gggggtgaga
201 gggcatgggt ggagggaaga aagtggagga gaaatcagac tgaaactaaa tcagtgccat aagataaaag gaatttca
```

Total sequence length: 278 nucleotides

Mutant sequence

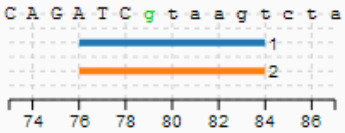
```
  1 gaagtttgct actgtacatc taggattcat agttaaacca attatggata gaagtcaggt cctattttca ttttctcaa ttatttttct ctttctccag
101 ATAAAACTGG TGGTGGAGTG GCAACTTCAG GATGACAAAA ACCAAAGTCT CTTCTGCTGG GAAATCCCAG TACAGATCat aagtctatct gggggtgaga
201 gggcatgggt ggagggaaga aagtggagga gaaatcagac tgaaactaaa tcagtgccat aagataaaag gaatttca
```

Total sequence length: 278 nucleotides

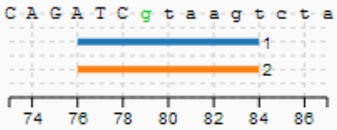
The underlined sequences are analyzed by HSF.

Interpreted Data

This table shows only relevant results related to the mutation position and context.

Predicted signal	Prediction algorithm	cDNA Position	Interpretation
Broken WT Donor Site	1 - HSF Matrices		Alteration of the WT donor site, most probably affecting splicing.
	2 - MaxEnt		

Raw Data Tables

Predicted signal	Prediction algorithm	cDNA Position	Interpretation
Broken WT Donor Site	1 - HSF Matrices 2 - MaxEnt		Alteration of the WT donor site, most probably affecting splicing.

Raw Data Tables

In the tables below, positions in sequence for the 5' intron are labeled as negative and as positive for the 3' intron.
Variations in the tables below are noted in colored boxes, according to the following scale:

Site broken	0% - 25% variation	26% - 50% variation	51% - 75% variation	76% - 100% variation	New site<

Potential Splice Sites | Potential Branch Points | Enhancer motifs | Silencer motifs | Other splicing motifs

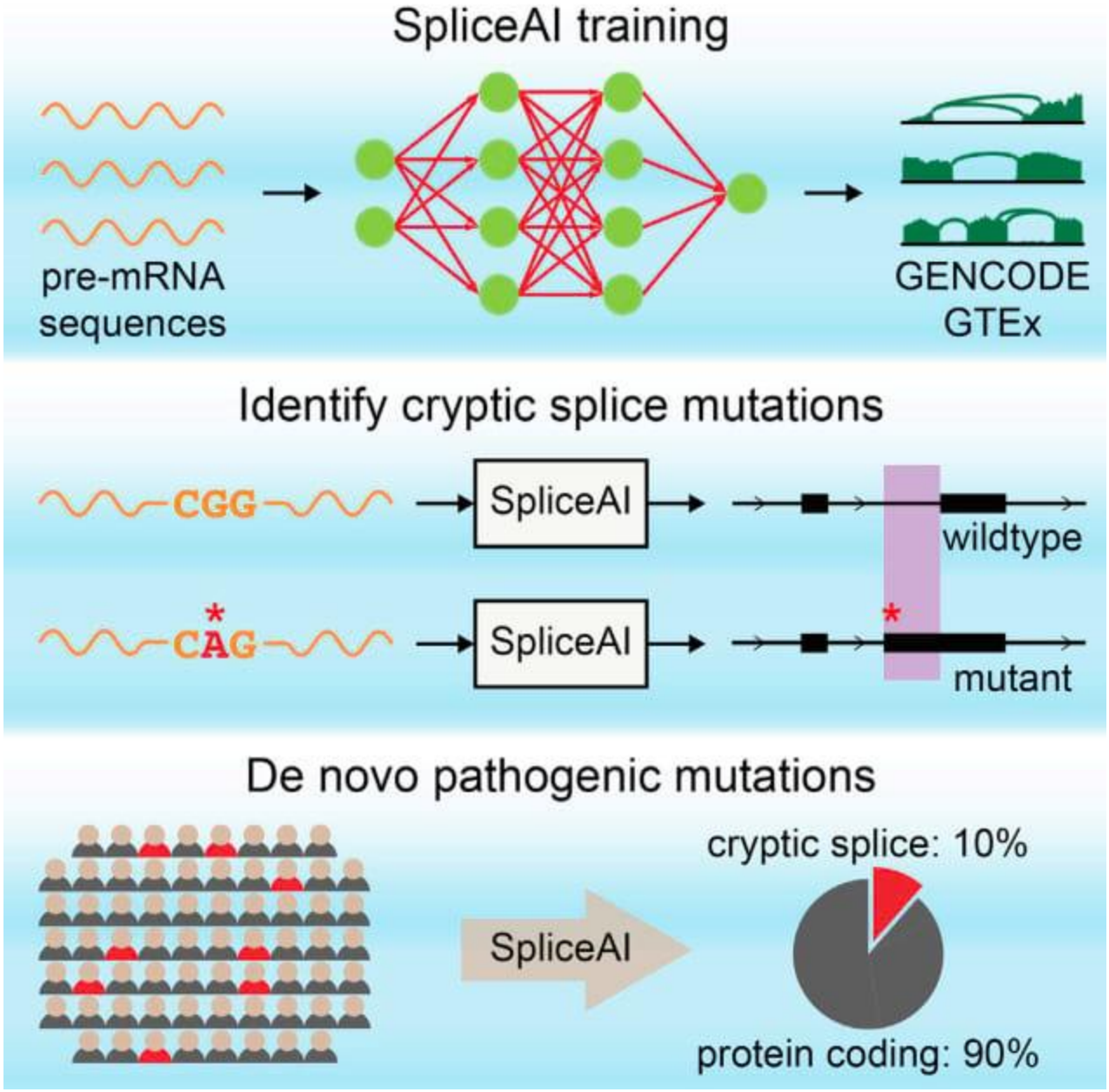
HSF Matrices

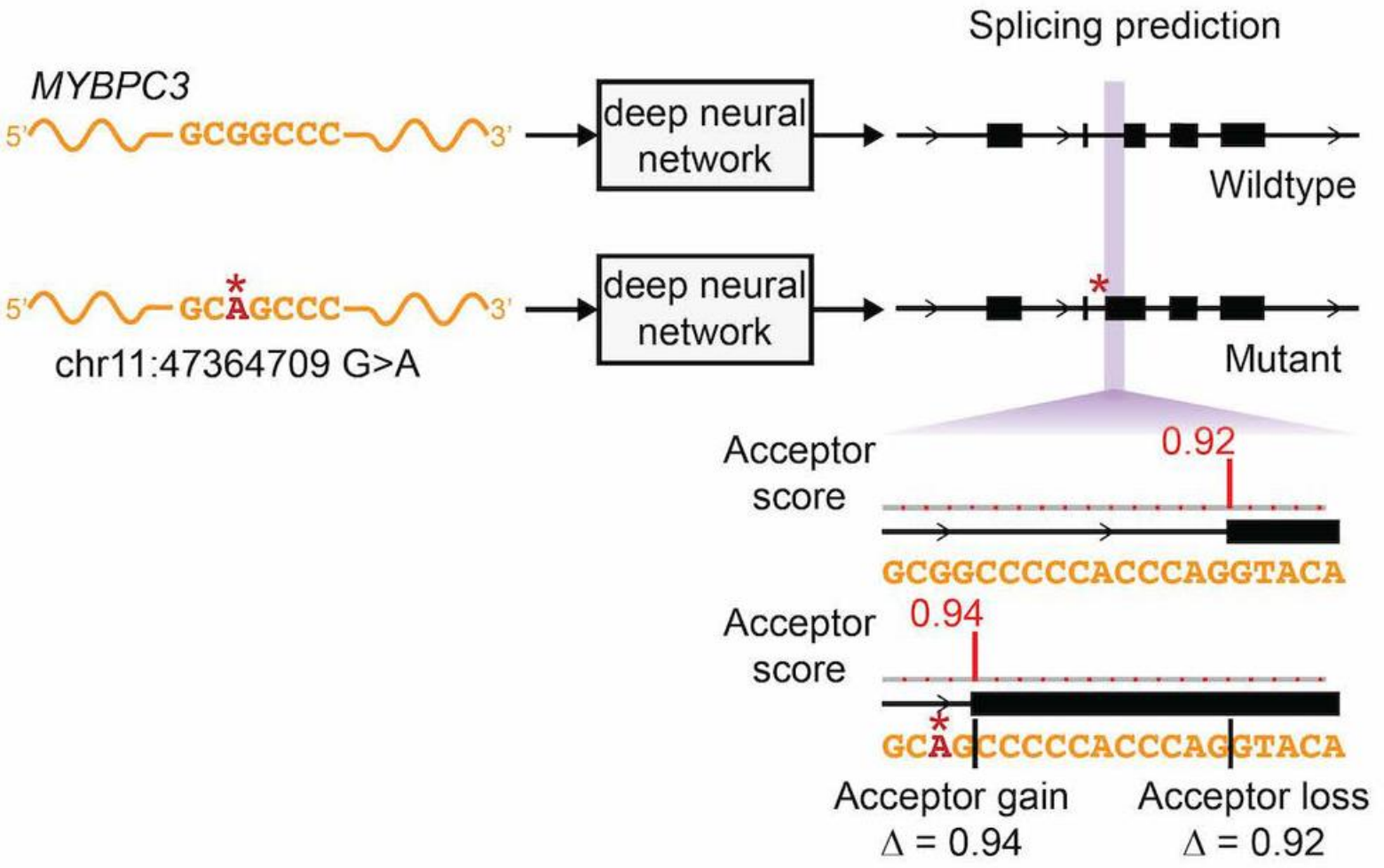
Sequence Position	cDNA Position	Splice site type	Motif	New splice site	Wild Type	Mutant	If cryptic site use, exon length variation	Variation (%)
172	72	Acceptor	ACAGATCgtaagtc	acagatcataagTC	66.67	66.74	NA	+0.1
176	76	Donor	ATCgtaagt	ATCataagt	83.28	56.45	87	WT site broken -32.22

MaxEnt

Threshold values:
5' Motif: 3 3' Motif: 3

Sequence Position	cDNA Position	5' Motif					3' Motif				
		Ref Motif	Ref Score	Mut Motif	Mut Score	Variation (%)	Ref Motif	Ref Score	Mut Motif	Mut Score	Variation (%)
176	76	ATCgtaagt	10.44	ATCataagt	2.26	-78.35					


SpliceAI training
pre-mRNA sequences
GENCODE GTEx
Identify cryptic splice mutations
CGG
SpliceAI
wildtype
CAG
SpliceAI
mutant
De novo pathogenic mutations
SpliceAI
cryptic splice: 10%
protein coding: 90%


Splicing prediction
MYBPC3
5'
GCGGCCC
3'
deep neural network
Wildtype
5'
GCAGCCC
3'
chr11:47364709 G>A
deep neural network
Mutant
Acceptor score
0.92
GCGGCCCCCACCCAGGTACA
Acceptor score
0.94
GCAGCCCCCACCCAGGTACA
Acceptor gain
Δ = 0.94
Acceptor loss
Δ = 0.92

Программы для оценки патогенности

Использованные в презентации

- MutationTaster http://www.mutationtaster.org
- PolyPhen2 http://genetics.bwh.harvard.edu/pph2/
- PROVEAN+SIFT http://provean.jcvi.org/index.php
- Predictsnp http://loschmidt.chemi.muni.cz/predictsnp/
- GENERESEARCH http://score.generesearch.ru/services/badmut/
- Human Splicing Finder http://www.umd.be/HSF3/
- MITOMAP https://www.mitomap.org/MITOMAP
- Mitimpact http://mitimpact.css-mendel.it/
- mt-tRNA http://structure.bmc.lu.se/PON-mt-tRNA/
- HOPE http://www.cmbi.ru.nl/hope/input
- DUET http://biosig.unimelb.edu.au/duet/stability/

Анализ сайта сплайсинга

- GeneSplicer http://ccb.jhu.edu/software/genesplicer/
- NetGene2 http://www.cbs.dtu.dk/services/NetGene2/
- NNSplice http://www.fruitfly.org/seq_tools/splice.html
- FSPLICE http://www.softberry.com/berry.phtml?topic=fsplice&group=programs&subgroup=gfind

Дополнительно

- Genetic variation databases https://www.humgen.nl/SNP_databases.html
- Fathmm http://fathmm.biocompute.org.uk/
- MutationAssessor http://mutationassessor.org/r3/
- PANTHER http://www.pantherdb.org/
- PhDSNP http://snps.biofold.org/
- SNPs&GO https://snps-and-go.biocomp.unibo.it/snps-and-go/
- MutPred http://mutpred.mutdb.org/
- nsSNPAnalyzer http://snpanalyzer.uthsc.edu/

Анализ на белковом уровне

- MuPro http://mupro.proteomics.ics.uci.edu/
- PHYRE2 http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index
- 3Drefine http://sysbio.rnet.missouri.edu/3Drefine/
- SDM http://mordred.bioc.cam.ac.uk/~sdm/sdm.php

Сокращения в аннотированном файле

PolyPhen-2

Benign – **B** – доброкачественная замена
Possibly damaging – **P** – возможно повреждающая (меньшая уверенность)
Probably damaging – **D** – вероятно повреждающая (большая уверенность)
Unknown – **N/A** – **.** – нет данных/не известно

SIFT

TOLERATED – **T** – толерантная, допустимая, переносимая, терпимая
DAMAGING – **D** – повреждающая замена
N/A – **.** – нет данных/не известно

PROVEAN

DELETERIOUS – **D** – **вредный** эффект
NEUTRAL – **N** – **нейтральный** эффект
N/A – **.** – нет данных/не известно

MUTATION TASTER

disease causing – **D** – вызывающий заболевание (мутация)

disease causing automatic - **A** – с точно установленным патогенным влиянием в базах dbSNP / TGP / ClinVar / HGMD

polymorphism – **P** – полиморфизм

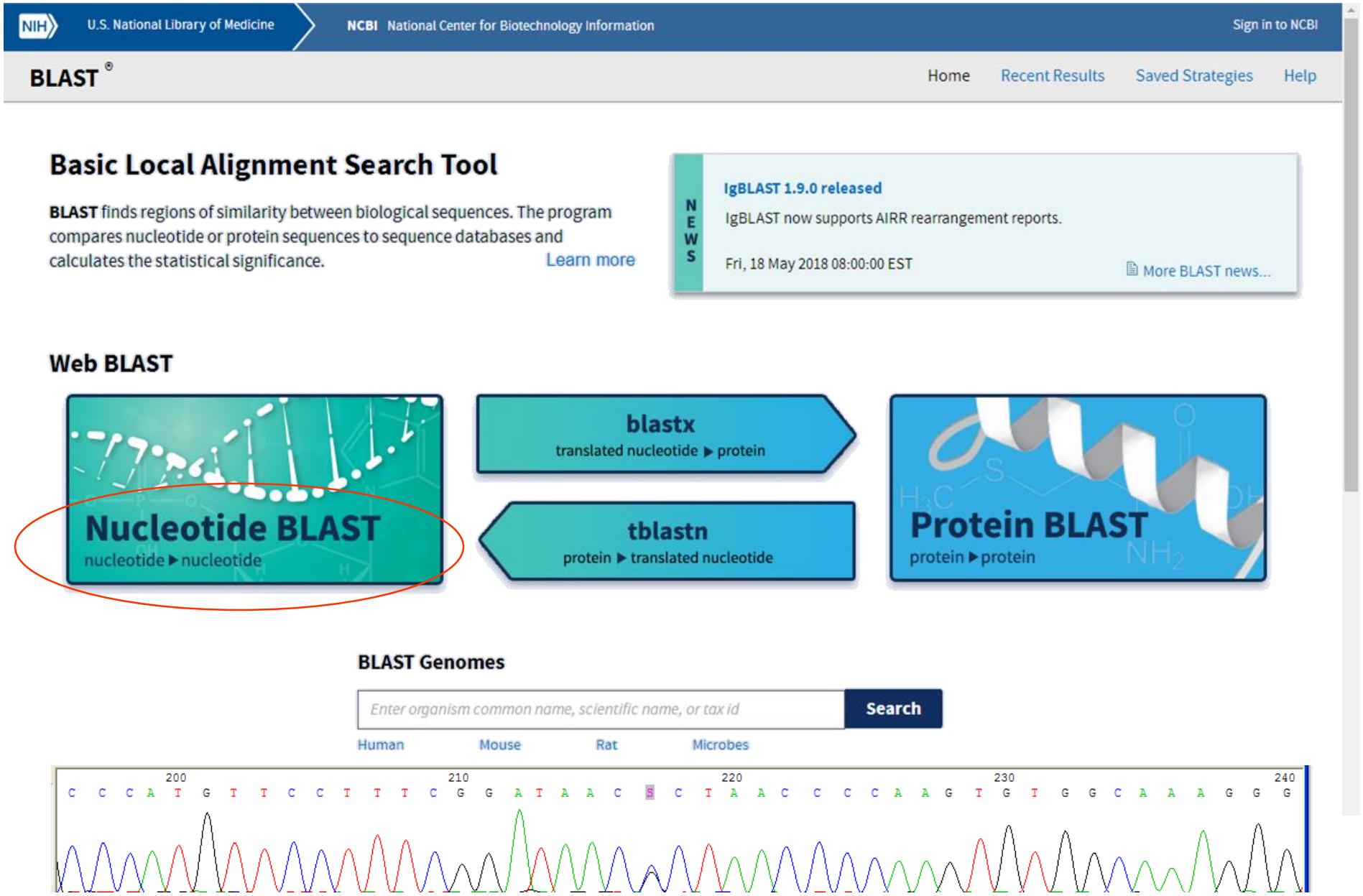
polymorphism automatic – **N???** – с точно установленным НЕ патогенным влиянием в базах dbSNP / TGP / ClinVar / HGMD

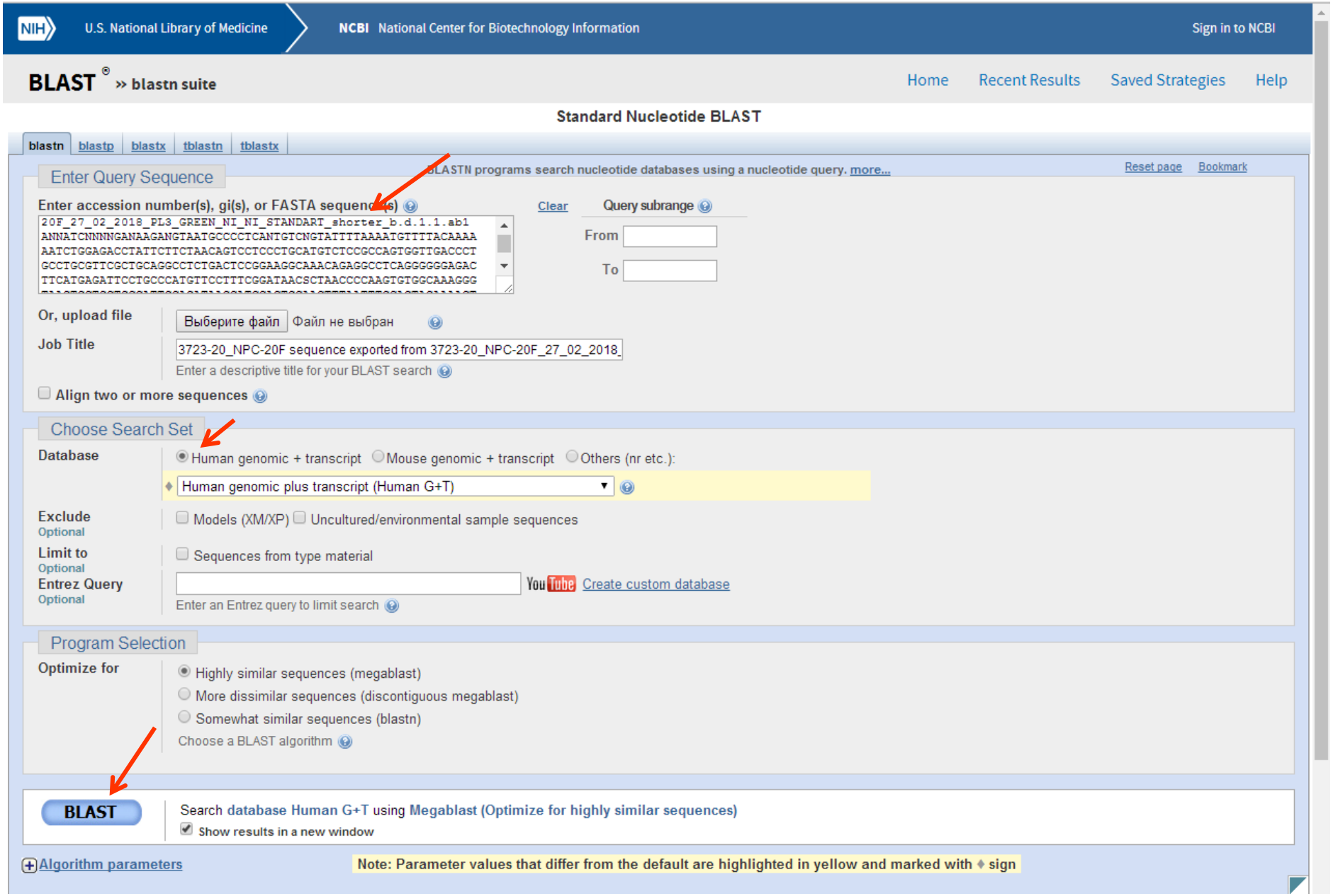
n/a – **.** – нет данных

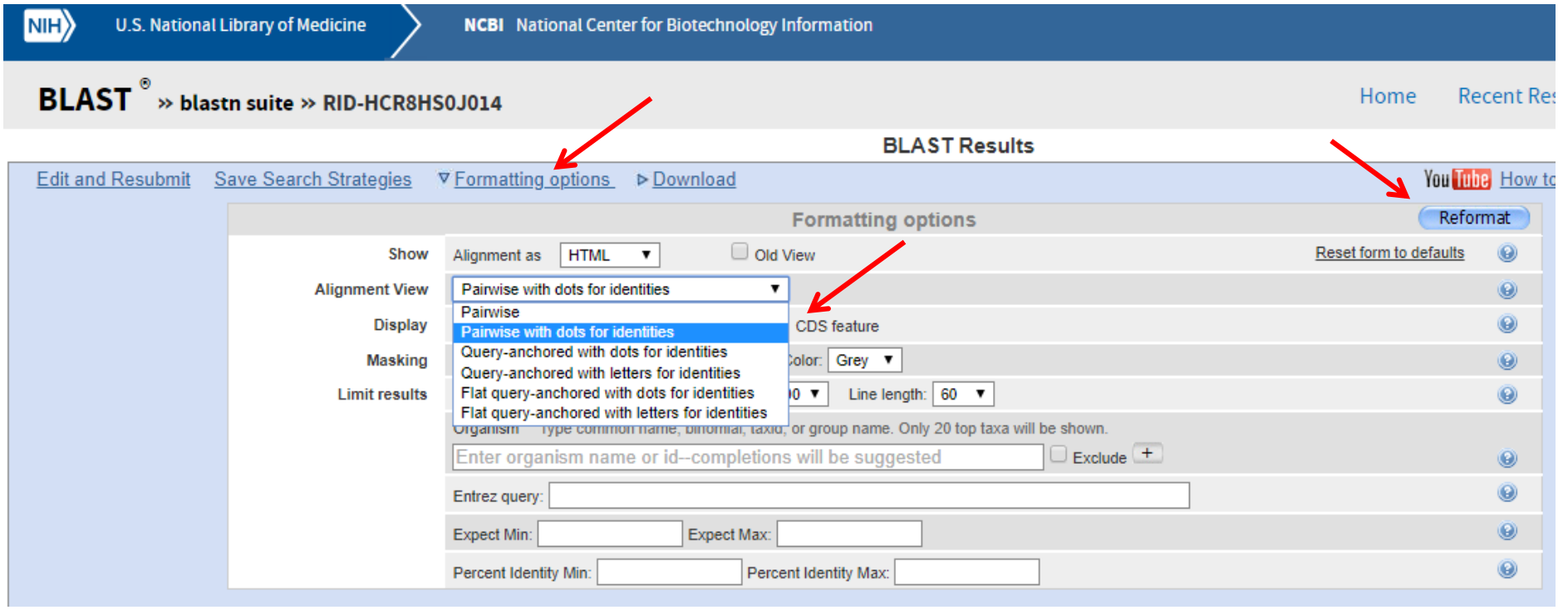
Заключение

1) Пользоваться рекомендуется несколькими программами, чтобы получить более точный результат.
2) Необходимо искать более усовершенствованные и простые в использовании базы данных и программы.
3) Надо иметь в виду, что данные регулярно обновляются, то есть предсказания в некоторых случаях могут меняться с течением времени.
4) Необходимо дополнительно проверять данные из баз dbSNP и HGMD. Обязательно проверять статьи для ранее выявленных замен.
5) Оценивать замены необходимо не только в кодирующей области, а во всем гене, включая промоторные области.
6) Если есть возможность, то для самого точного подтверждения патогенности варианта рекомендуется проводить популяционный и функциональный тест.
7) Сами разработчики указывают на то, что каждый результат необходимо рассматривать самому и анализировать числовой счет самостоятельно, особенно когда результат близок к пороговому значению.

https://blast.ncbi.nlm.nih.gov/Blast.cgi




U.S. National Library of Medicine
NCBI National Center for Biotechnology Information
Sign in to NCBI
BLAST » blastn suite
Home
Recent Results
Saved Strategies
Help
Standard Nucleotide BLAST
blastn
blastp
blastx
tblastn
tblastx
Enter Query Sequence
BLASTN programs search nucleotide databases using a nucleotide query. more...
Reset page
Bookmark
Enter accession number(s), gi(s), or FASTA sequence(s)
Clear
Query subrange
From
To
20F_27_02_2018_PL3_GREEN_NI_NI_STANDART_shorter_b.d.1.1.ab1
ANNATCNNNNGANAAGANGTAATGCCCCTCANTGTCNGTATTTTAAAATGTTTTACAAAA
AATCTGGAGACCTATTCTTCTAACAGTCCTCCCTGCATGTCTCCGCCAGTGGTTGACCCT
GCCTGCGTTCGCTGCAGGCCTCTGACTCCGGAAGGCAAACAGAGGCCTCAGGGGGGAGAC
TTCATGAGATTCCTGCCCATGTTCCTTTCGGATAACSCTAACCCCAAGTGTGGCAAAGGG
Or, upload file
Выберите файл
Файл не выбран
Job Title
3723-20_NPC-20F sequence exported from 3723-20_NPC-20F_27_02_2018_
Enter a descriptive title for your BLAST search
Align two or more sequences
Choose Search Set
Database
Human genomic + transcript
Mouse genomic + transcript
Others (nr etc.):
Human genomic plus transcript (Human G+T)
Exclude
Optional
Models (XM/XP)
Uncultured/environmental sample sequences
Limit to
Optional
Sequences from type material
Entrez Query
Optional
Create custom database
Enter an Entrez query to limit search
Program Selection
Optimize for
Highly similar sequences (megablast)
More dissimilar sequences (discontiguous megablast)
Somewhat similar sequences (blastn)
Choose a BLAST algorithm
BLAST
Search database Human G+T using Megablast (Optimize for highly similar sequences)
Show results in a new window
Algorithm parameters
Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign


U.S. National Library of Medicine
NCBI National Center for Biotechnology Information
BLAST » blastn suite » RID-HCR8HS0J014
Home
Recent Res
BLAST Results
Edit and Resubmit
Save Search Strategies
Formatting options
Download
How to
Formatting options
Reformat
Show
Alignment as
HTML
Old View
Reset form to defaults
Alignment View
Pairwise with dots for identities
Pairwise
Pairwise with dots for identities
Query-anchored with dots for identities
Query-anchored with letters for identities
Flat query-anchored with dots for identities
Flat query-anchored with letters for identities
Display
CDS feature
Masking
Color:
Grey
Limit results
Line length:
60
Organism
Type common name, binomial, taxid, or group name. Only 20 top taxa will be shown.
Enter organism name or id--completions will be suggested
Exclude
Entrez query:
Expect Min:
Expect Max:
Percent Identity Min:
Percent Identity Max:

Download GenBank Graphics ▼ Next ▲ Previous Descriptions

Homo sapiens NPC intracellular cholesterol transporter 1 (NPC1), mRNA

Sequence ID: NM_000271.4 Length: 4827 Number of Matches: 1

Range 1: 3183 to 3316 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
244 bits(132)	2e-62	133/134(99%)	0/134(0%)	Plus/Plus

```
Query  107   CAGTGGTTGACCCTGCCTGCGTTCGCTGCAGGCCTCTGACTCCGGAAGGCAAACAGAGGC  166
Sbjct  3183  ............................................................  3242

Query  167   CTCAGGGGGGAGACTTCATGAGATTCCTGCCCATGTTCCTTTCGGATAACSCTAACCSCA  226
Sbjct  3243  ..........................................C.........C.......  3302

Query  227   AGTGTGGCAAAGGG  240
Sbjct  3303  ..............  3316
```

Related Information

Gene - associated gene details
UniGene - clustered expressed sequence tags
GEO Profiles - microarray expression data

Download GenBank Graphics ▼ Next ▲ Previous Descriptions

Homo sapiens chromosome 18, GRCh38.p12 Primary Assembly

Sequence ID: NC_000018.10 Length: 80373285 Number of Matches: 1

Range 1: 23538458 to 23538762 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
553 bits(299)	3e-155	302/305(99%)	0/305(0%)	Plus/Minus

Features: NPC intracellular cholesterol transporter 1 isoform X3
NPC intracellular cholesterol transporter 1 isoform X1

```
Query  19        GTAATGCCCCTCANTGTCNGTATTTTAAAATGTTTTACAAAAAATCTGGAGACCTATTCT  78
Sbjct  23538762  ............C....A.............................................  23538703

Query  79        TCTAACAGTCCTCCCTGCATGTCTCCGCCAGTGGTTGACCCTGCCTGCGTTCGCTGCAGG  138
Sbjct  23538702  ............................................................  23538643

Query  139       CCTCTGACTCCGGAAGGCAAACAGAGGCCTCAGGGGGGAGACTTCATGAGATTCCTGCCC  198
Sbjct  23538642  ............................................................  23538583

Query  199       ATGTTCCTTTCGGATAACSCTAACCCCAAGTGTGGCAAAGGGTAAGTGCTGCTGCCATTG  258
Sbjct  23538582  ................C...........................................  23538523

Query  259       CAGATAAGCATCCACTGCAACTTTAATTTGCAGTAGAAAACTAGGAGAGGACTGGGCTAA  318
Sbjct  23538522  ............................................................  23538463

Query  319       GACAG  323
Sbjct  23538462  .....  23538458
```

Related Information

PubChem BioAssay - bioactivity screening
New Genome Data Viewer - aligned genomic context