

Welcome to SNP Analysis!

SNP Analysis is a free-to-use and free-to-modify executable Java program, which can automatically apply single nucleotide polymorphisms to the genetic sequence of a protein.

Getting started:

1. Ensure that “Java” and “Java SE Development Kit” are installed on your computer (<https://www.oracle.com/java/technologies/javase-downloads.html>)
2. Download SNP_Analysis_v1.3
3. Go to UniProt (<https://www.uniprot.org/>) and search for target protein
 - a. Identify the canonical sequence
 - b. Look for the nucleotide sequence (FASTA) by using the CCDS Link (red box) and save it for later
 - c. Identify the right RefSeq (HGVS) for your protein (green box)
 - d. If you want to perform additional analysis using Provean, look for the ENSP number (purple box)

3. [Cross-references](#)

Sequence databases

Select the link	M57230 mRNA Translation: AAA59155.1
destinations:	AB015706 mRNA Translation: BAA78112.1
<input checked="" type="radio"/> EMBL ⁱ	AB102802 mRNA Translation: BAD89393.1
<input type="radio"/> GenBank ⁱ	EF064722 Genomic DNA Translation: ABK41905.1
<input type="radio"/> DDBJ ⁱ	AC008914 Genomic DNA No translation available.
	AC016596 Genomic DNA No translation available.
	CH471123 Genomic DNA Translation: EAW54936.1
CCDS ⁱ	CCDS3971.1 [P40189-1]
	CCDS47209.1 [P40189-2]
	CCDS54856.1 [P40189-3]
PIR ⁱ	A36337
RefSeq ⁱ	NP_001177910.1, NM_001190981.1 [P40189-3]
	NP_002175.2, NM_002184.3 [P40189-1]
	NP_786943.1, NM_175767.2 [P40189-2]

Genome annotation databases

Ensembl ⁱ	ENST00000336909; ENSP00000338799; ENSG00000134352 [P40189-1] ENST00000381287; ENSP00000370687; ENSG00000134352 [P40189-2] ENST00000381294; ENSP00000370694; ENSG00000134352 [P40189-3] ENST00000381298; ENSP00000370698; ENSG00000134352 [P40189-1] ENST00000502326; ENSP00000462158; ENSG00000134352 [P40189-1] ENST00000522633; ENSP00000435399; ENSG00000134352 [P40189-2]
GeneID ⁱ	3572
KEGG ⁱ	hsa:3572
UCSC ⁱ	uc003jqj.4, human [P40189-1]

4. Go to NCBI Website (<https://www.ncbi.nlm.nih.gov>), select SNP and search for target protein
 - a. Select “missense”
 - b. Scroll all the way down to the bottom of the page to download SNP list
 - c. Select “File” and „Summary“ as shown

rs138205220 [Homo sapiens]
20.

Variant type: SNV
Alleles: A>C,G [Show Flanks]
Chromosome: 5:55941607 (GRCh38)
5:55237435 (GRCh37)

Canonical SPDI: NC_000005.10:55941608:A:C,NC_000005.10:55941608:A:G
Gene: IL6ST (Varview)
Functional Consequence: missense_variant,non_coding_transcript_variant,synonymous_variant,3_prime_UTR_variant

Validated: by frequency,by alfa,by cluster
MAF: G=0.000358/4 (ALFA)
G=0.000016/2 (TOPMED)
G=0.000077/1 (GoESP)
...more

HGVS: NC_000005.10:g.55941607A>C, NC_000005.10:g.55941608A>G, NC_000005.9:g.55237435A>C, NC_000005.9:g.55237435A>C, NM_002184.4:c.2232T>G, NM_002184.4:c.2232T>C, NM_002184.4:c.2232T>G, NM_175767.2:c.1145G>C, NM_175767.2:c.1145G>C

<< First < Prev P

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

Send to: >

Choose Destination

File Clipboard

Collections

Download 577 items.

Format

Summary

Sort by

SNP_ID

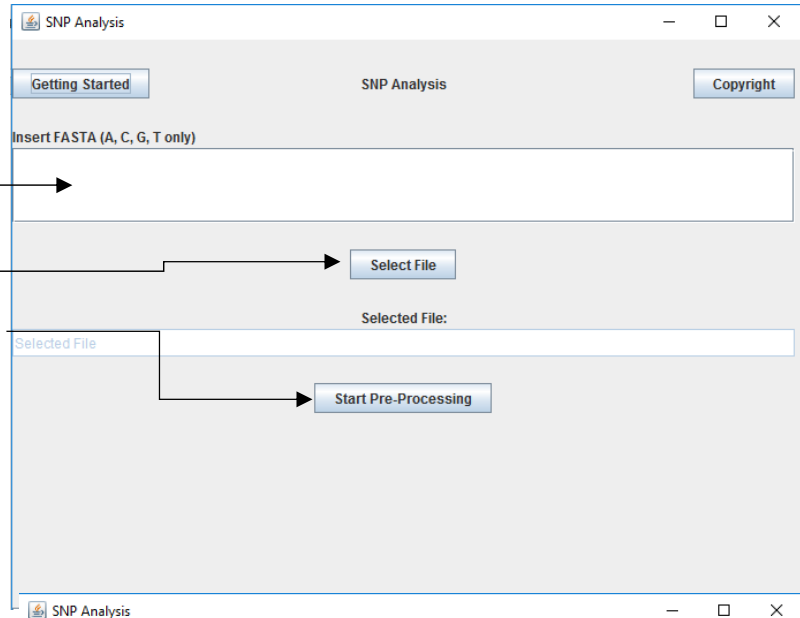
Create File

d. Click “Create File”. A file named “snp_result” should now have appeared in your Download folder

5. Let’s get started!

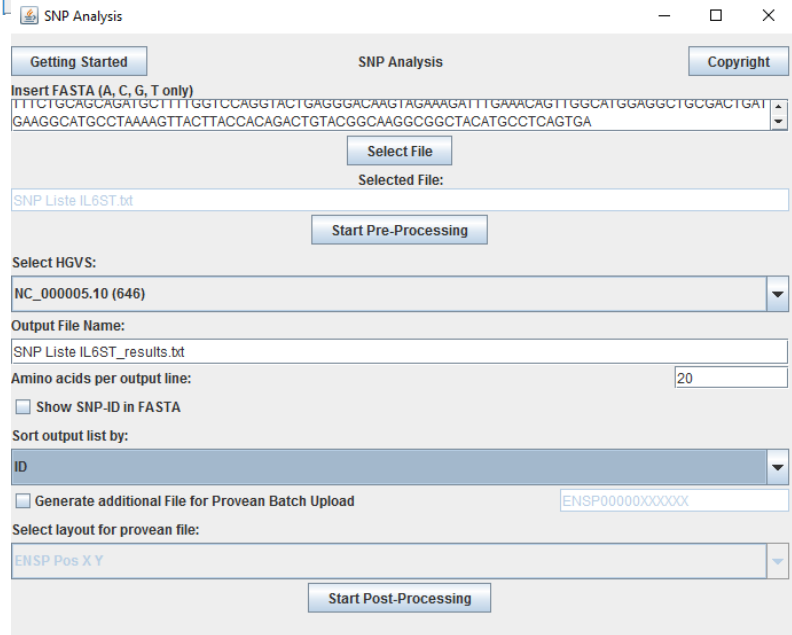
Pre-Processing:

1. Open your SNP Analysis tool
2. Use the “Getting started“-button to display a short Tutorial
3. Insert your protein’s FASTA
4. Click “Select File” and pick the previously downloaded SNP_result file from NCBI
5. Proceed by clicking “Start Pre-Processing”



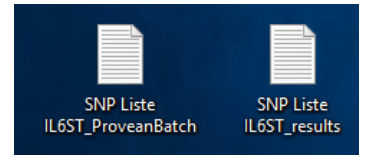
Post-Processing:

1. Select HGVS use the RefSeq you found in UniProt (Getting started, 3c), the program will show you how many matching SNPs it found with each HGVS
2. Name your Output File
3. Choose how many amino acids you want to have displayed per output line
4. Select a sorting option
 - a. Sort by SNP ID as sorted in NCBI
 - b. Sort by name (rs code)
 - c. Sort by clinical significance (Annotations in ClinVar)
 - d. Sort by Position (in the nucleotide Sequence)
5. If you wish to run a Provean analysis:
 - a. Tick the box
 - b. Insert the protein’s ENSP number
 - c. Choose a Layout for your Output File (“ENSP Pos X Y” is the format to use for the “Provean Protein Batch” [http://provean.jcvi.org/protein_batch_submit.php?species=human])
6. “Start Post-Processing”



Output Files:

1. You should now find two txt files on your Desktop (if you chose to create a file for Provean Upload)
2. The result file contains the nucleotide sequence, the corresponding amino acid sequence, the aligned SNPs and their ID numbers in the upper part.
The lower part contains a list with all of the protein's SNPs
3. The "_ProveanBatch" file contains a list ready for upload to Provean (http://provean.jcvi.org/protein_batch_submit.php?species=human)



SNP Liste IL6ST_ProveanBatch - Editor

Datei Bearbeiten Format Ansicht ?

```

ENSP00000370698 3 T M
ENSP00000370698 4 L W
ENSP00000370698 5 Q H
ENSP00000370698 8 L V
ENSP00000370698 8 L V
ENSP00000370698 8 L P
ENSP00000370698 9 V L
ENSP00000370698 9 V M
    
```

SNP Liste IL6ST_results - Editor

Datei Bearbeiten Format Ansicht ?

```

ATGTTGACGTTGCAGACTTGGCTAGTGCAAGCCTTGTTATTTTCTCACCAGTGAATCTACAGGTGAACTTCTA
M L T L Q T W L V Q A L F I F L T T E S T G E L L
199270210 5 75 283441 300 312355 490
ATGTGGCAC GTACTG GGCATG GCT GCAGAT ATA
M W H V L G M A A D I
226 75
CCAATG
P M
571
GTA
V
572
GTA
V
GATCCATGTGGTTATATCAGTCTGAATCTCAGTTGTACAACCTCATTCTAATTTCACTGCAGTTTGTGTGCTA
D P C G Y I S P E S P V V Q L H S N F T A V C V L
306 427 82 503366549 372 537107285382 17 143419 378
GGT TTTGTT ATGTGTTCT CCT GTTCGTTTGTATGTC GTAATT GCG
G F V M C S P V R F D V V I A
82 366 544
GAT GGT TTG
D G L
494
    
```

SNP Liste IL6ST_results - Editor

Datei Bearbeiten Format Ansicht ?

```

199; rs764660393 [Homo sapiens]; [no ClinSig information]; C8T
270; rs776294028 [Homo sapiens]; [no ClinSig information]; T11G
210; rs765993554 [Homo sapiens]; [no ClinSig information]; G15C
5; rs1063560 [Homo sapiens]; [no ClinSig information]; C22G
571; rs52794902 has merged into rs1063560 [Homo sapiens]; [no ClinSig information]; C22G
572; rs3204590 has merged into rs1063560 [Homo sapiens]; [no ClinSig information]; C22G
226; rs769275290 [Homo sapiens]; [no ClinSig information]; T23C
75; rs372602249 [Homo sapiens]; [no ClinSig information]; G25C
75; rs372602249 [Homo sapiens]; [no ClinSig information]; G25A
283; rs778516584 [Homo sapiens]; [no ClinSig information]; C32G
441; rs1310032899 [Homo sapiens]; [no ClinSig information]; T34A
300; rs781455079 [Homo sapiens]; [no ClinSig information]; A52G
312; rs893553728 [Homo sapiens]; [no ClinSig information]; A61G
355; rs1048494423 [Homo sapiens]; [no ClinSig information]; G65A
490; rs1391476352 [Homo sapiens]; [no ClinSig information]; C73A
306; rs867963101 [Homo sapiens]; [no ClinSig information]; A77G
427; rs1283997946 [Homo sapiens]; [no ClinSig information]; G83T
494; rs1401093156 [Homo sapiens]; [no ClinSig information]; G85A
82; rs536592543 [Homo sapiens]; [no ClinSig information]; G86T
82; rs536592543 [Homo sapiens]; [no ClinSig information]; G86A
503; rs1408498223 [Homo sapiens]; [no ClinSig information]; C93G
366; rs1173399871 [Homo sapiens]; [no ClinSig information]; A94T
    
```