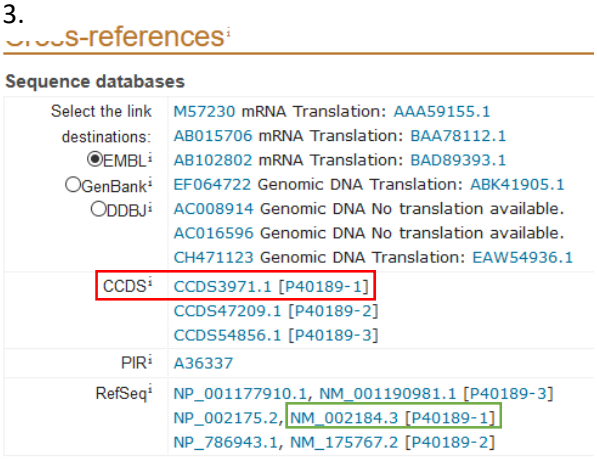
Welcome to SNP_Analysis!

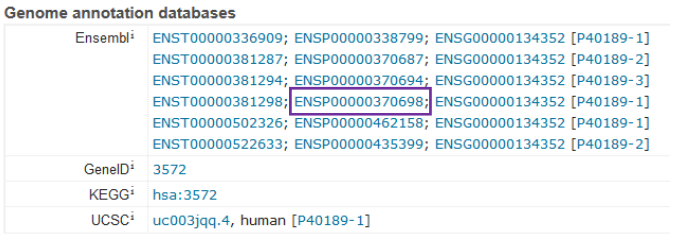
SNP Analysis is a free-to-use and free-to-modify executable Java program, which can automatically apply single nucleotide polymorphisms to the genetic sequence of a protein.

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.





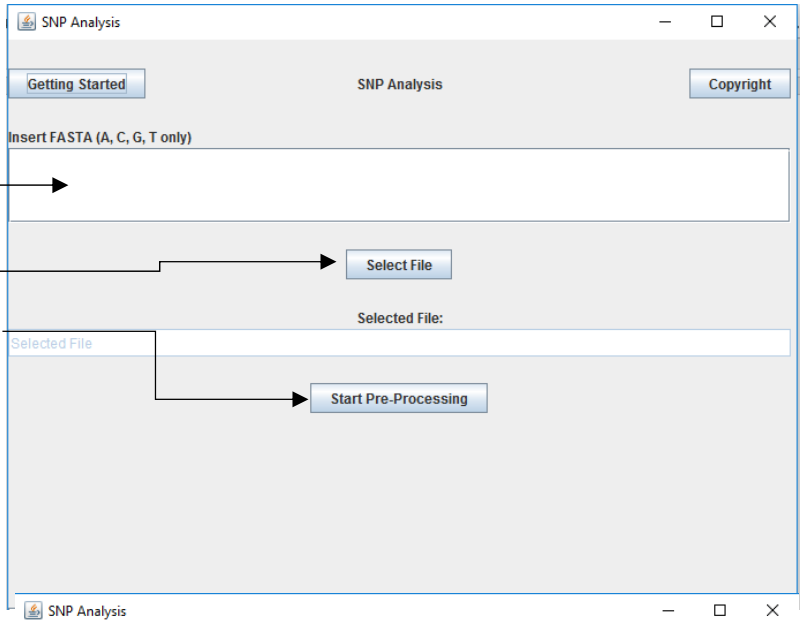
4. Go to NCBI Website (https://www.ncbi.nlm.nih.gov), select SNP and search for target protein
    a. Select "missense"
    b. Scroll all the way down to the bottom of the page to download SNP list
    c. Select "File" and „Summary" as shown

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

5. Let’s get started!

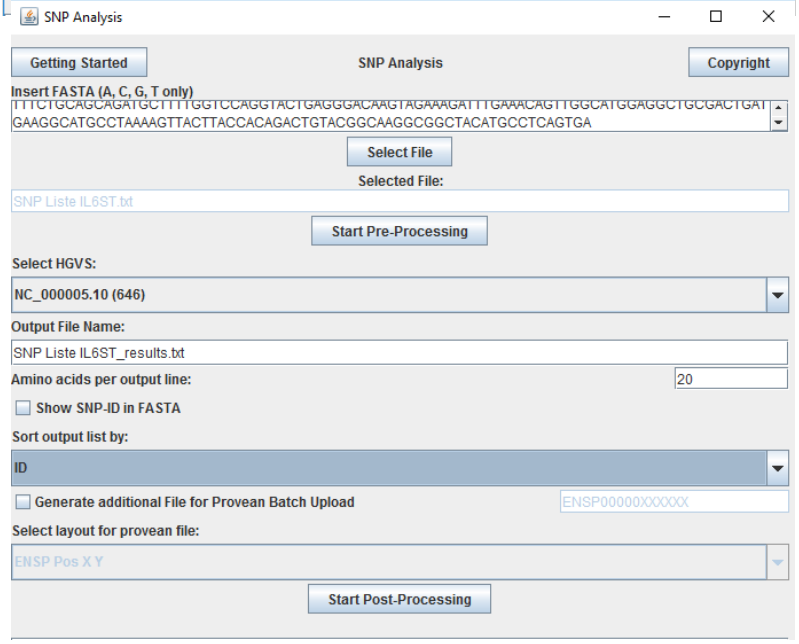
Pre-Processing:

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“

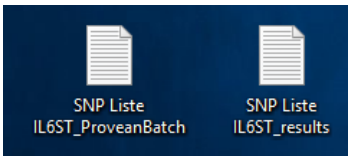


Post-Processing:

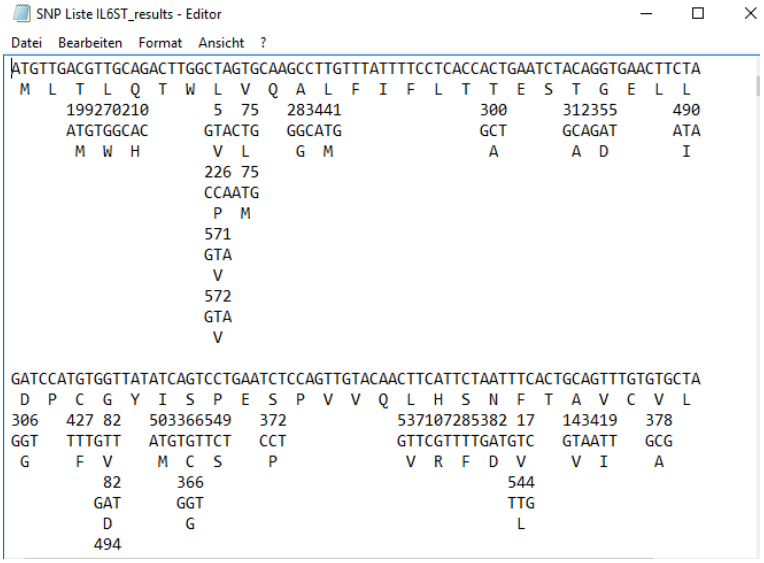
1. Select HGVS use the RefSeq you found in UniProt (Getting started, 3c), the program will show you how many mathcing SNPs it found with each HGVS
2. Name your Output File
3. Choose how many amino acids you want to have displayed per output line
4. Select a sorting option
    a. Sort by SNP ID as sorted in NCBI
    b. Sort by name (rs code)
    c. Sort by clinical sgnificance (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:



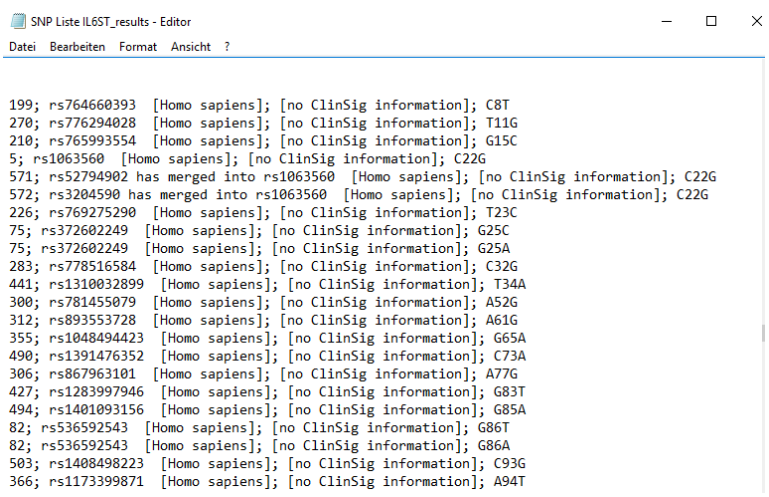
1. You should now find two txt files on your Desktop (if you chose to create a file for Provean Upload)
2. The result file contains the nucleotide sequence, the corresponding amino acid sequence, the aligned SNPs and their ID numbers in the upper part.
   The lower part contains a list with all of the 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_results - Editor

Datei Bearbeiten Format Ansicht ?

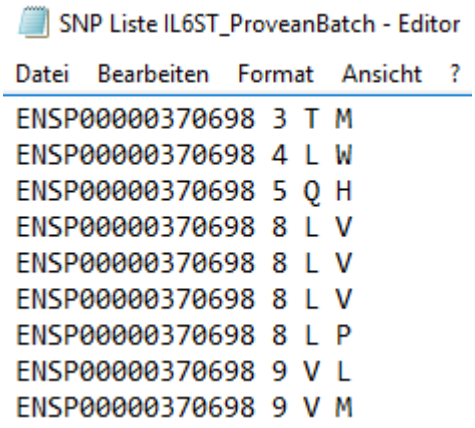
```
ATGTTGACGTTGCAGACTTGGCTAGTGCAAGCCTTGTTTATTTTCCTCACCACTGAATCTACAGGTGAACTTCTA
 M  L  T  L  Q  T  W  L  V  Q  A  L  F  I  F  L  T  T  E  S  T  G  E  L  L
      199270210       5  75   283441              300      312355      490
      ATGTGGCAC      GTACTG   GGCATG               GCT      GCAGAT      ATA
       M  W  H        V  L     G  M                A        A  D        I
                     226 75
                     CCAATG
                      P  M
                     571
                     GTA
                      V
                     572
                     GTA
                      V

GATCCATGTGGTTATATCAGTCCTGAATCTCCAGTTGTACAACTTCATTCTAATTTCACTGCAGTTTGTGTGCTA
 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           GTTCGTTTTGATGTC   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
```


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 V
ENSP00000370698 8 L P
ENSP00000370698 9 V L
ENSP00000370698 9 V M
```