SNP Toolbox
Quick Start Guide

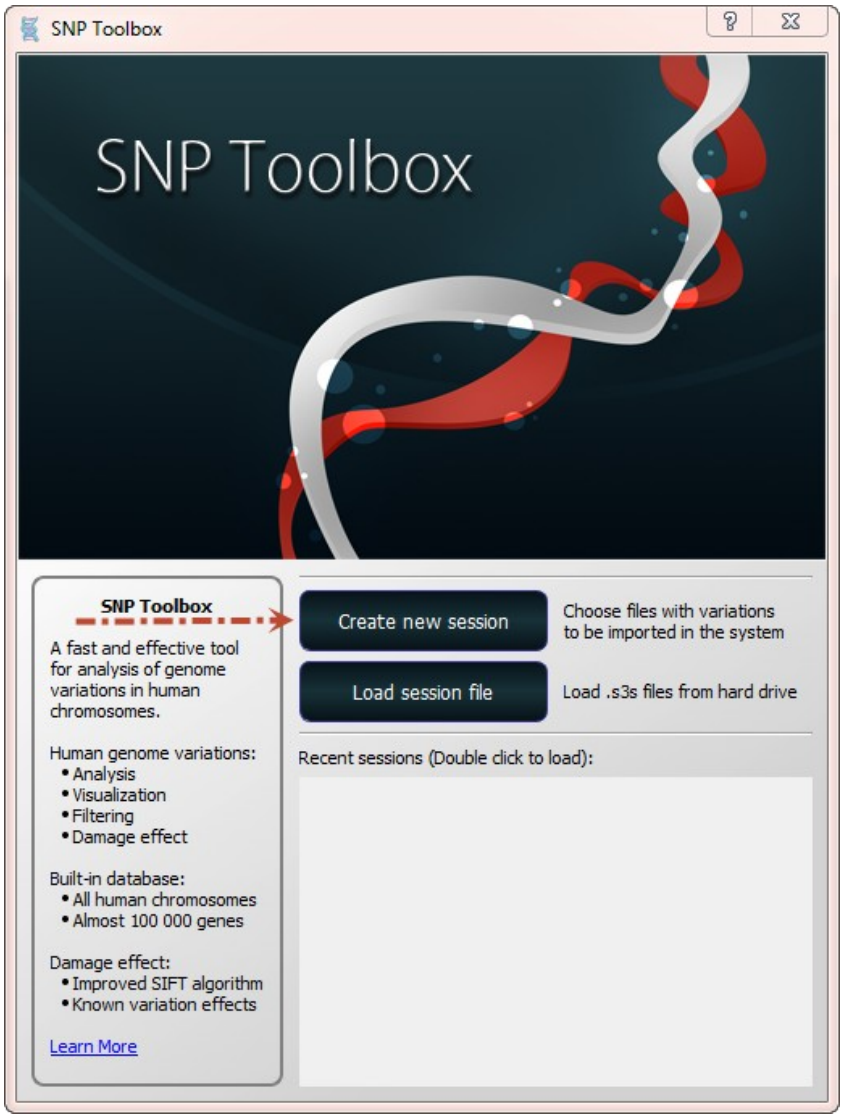
This document will give you a brief introduction to SNP Toolbox. This is a software suite for visualization, filtration and annotation of human genome variations. You can find more information in the full documentation.

To start working with variations you will need:

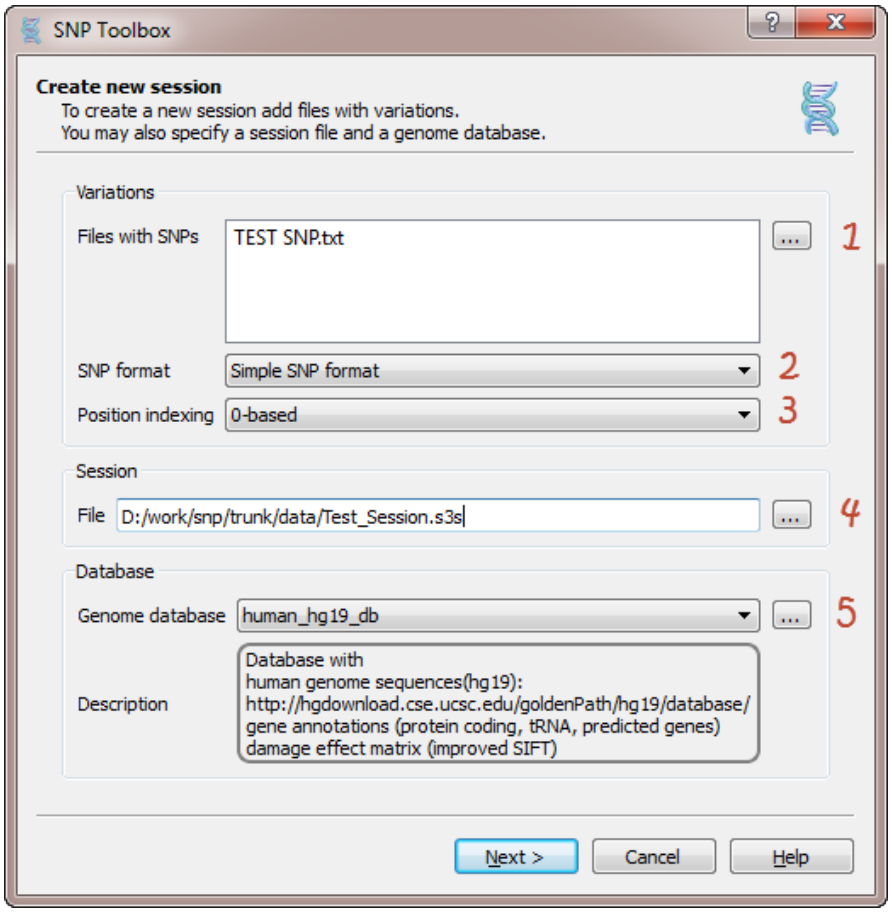
- Special database *.s3db* file (with sequences, annotations and damage effect matrix)
- File(s) with variations

Having these files you can create a *session*.

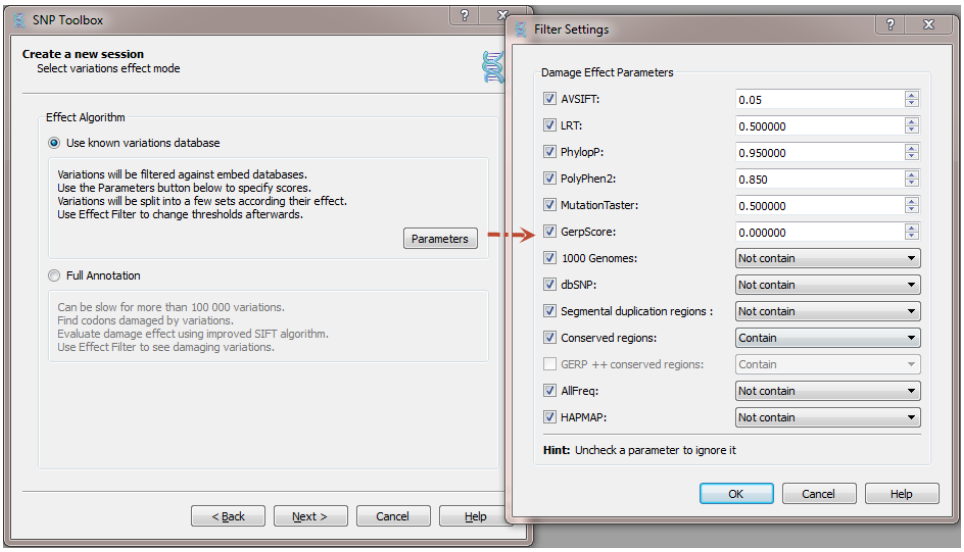
Create a new session



Click “Create new session” in Welcome Page to start a session.

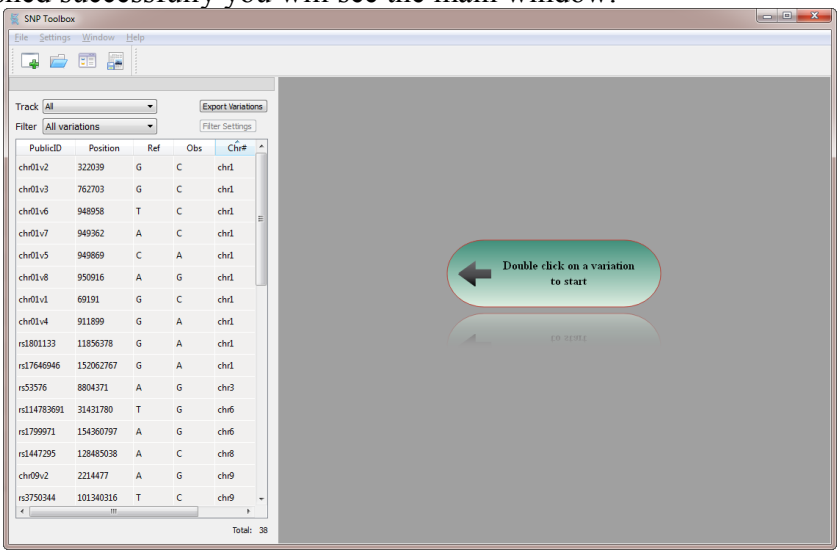


You just need to specify SNP file(s) clicking on button “1” and format of the file(s) (Simple SNP format or VCF4) in combo box “2”. You may also specify SNP position indexing (combo box “3”). SNP Toolbox will create a session file and select a database file, but you can change that if you want.
Select “data/TEST_SNP.txt” file to load sample variations.
Click the “Next” button to specify an annotation algorithm to be used.



Effect Algorithm will retrieve annotations from embedded database. Scores and IDs will be used for filtration of variations. You can specify thresholds clicking the "Parameters" button. Variations will be split into three sets: "Perspective" for *de novo* damaging annotations, "Discarded" for variations discarded by at least one filter and "Unknown" for variations which effect is unknown.
Full Annotation algorithm will retrieve tolerance scores from damage effect matrices and all the scores that uses the previous algorithm. No filtration is applied. This algorithm is slow for more than 100 000 variations since it requires a lot of computations and queries to the databases.

When you click the "Next" button, SNP Toolbox will load your variation files and store them into the embedded database (session file actually). This can take a while depending on a number of variations in file(s).
If import is finished successfully you will see the main window:

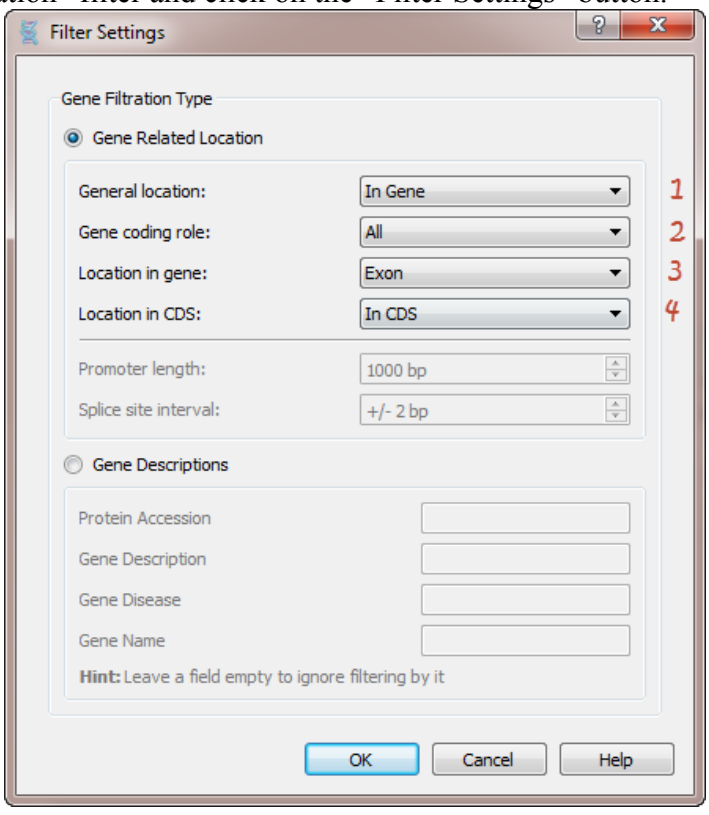


Variation Navigator

The left part of the window is Variation Navigator. All variations from your files are available in Variation Navigator.

Try switching tracks if you selected the "Effect Algorithm".

You can sort variations by any column or filter them (there are 4 different filters). It's important that a filter is applied to all variations of the **selected track**, but not to a current set. Also a variation is not discarded by a filter if filtration conditions are satisfied at least in one gene intersected by variation. Click on "Filter Settings" to edit filtration conditions.
Available filters:
 All variations – shows all variations of current session.
 Effect – filter by damage effect of variations
 Gene location – filter by gene related location (e.g. inside a gene, inside CDS, inside Exon, etc.)
 Variation characteristic – filter by characteristics of variations (e.g. region, chromosome, nucleotide)
Let's have a look at a filtration example:
Select the "Gene location" filter and click on the "Filter Settings" button.

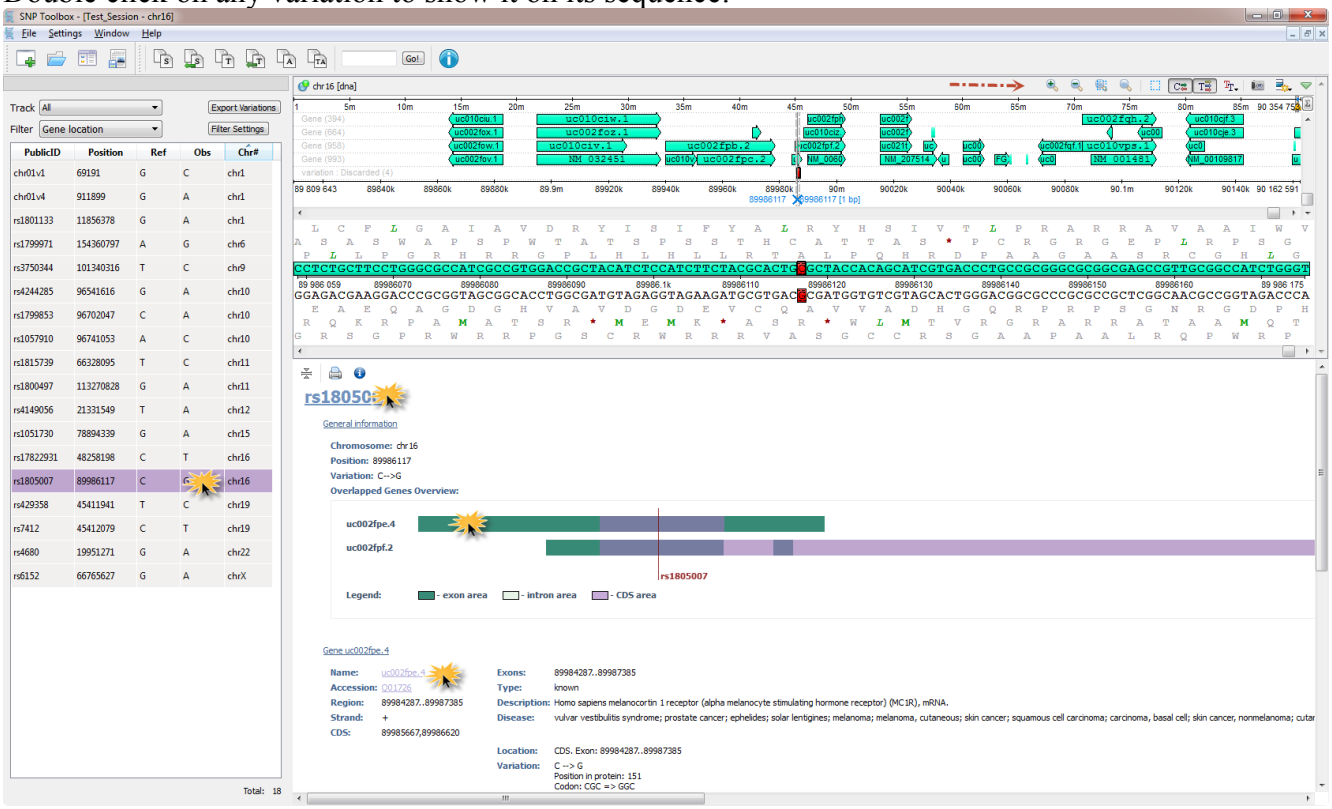


Then select:
1. If a mutation in a gene (or "Out of gene", "Promoter")
2. All genes: Protein and non-protein coding.
3. If a mutation in an exon (or "intron", "splice site")
4. If a mutation in CDS (or "Out of CDS". 5' or 3'-end)
With the constraints above all variations that are in coding exons will be shown.

You can setup any constraint by the filter, for instance:
 Mutations that are in CDS of coding genes («In Gene», «Protein Coding», «Whole Gene», «In CDS»))
 5'-end exons («In Gene», «All», «Exon», «Out of CDS. 5"-end»))
 In splice sites («In Gene», «All», «Splice site», «All»)), and select length of splice site in "Splice site interval"
 In promoter regions ("Gene promoter") and select length of promoter regions in "Promoter length"
 …

Visualization

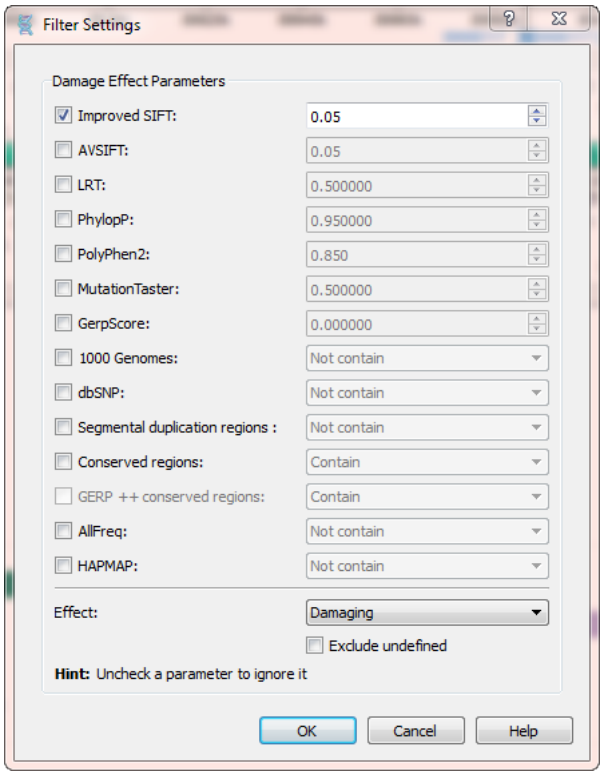
The right part is for visualization of sequences and annotations on them.
Double click on any variation to show it on its sequence:



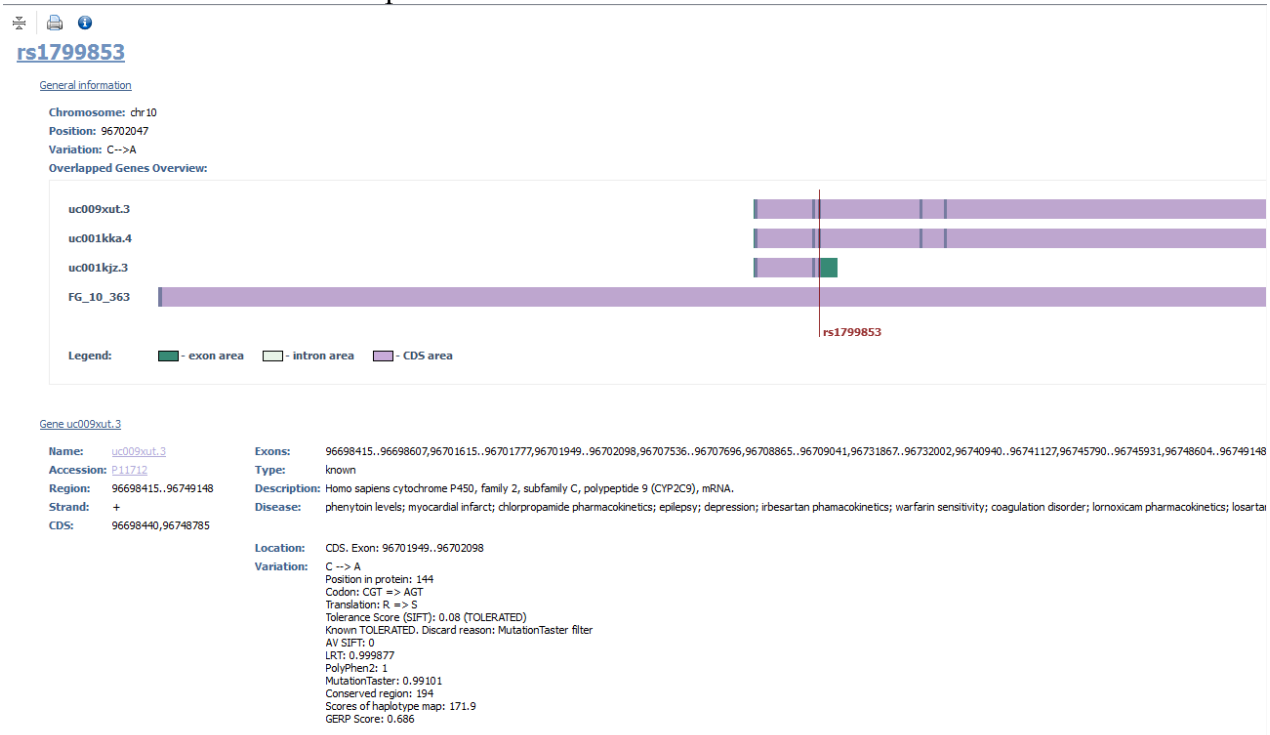
1. Gene overview. You can zoom it using button on the toolbar
2. Nucleotides and translations
3. Intersected genes overview. You can click on a gene to select it on the overview
4. Intersected genes report. You can click on a gene name or on its accession. Variations effect is shown as a text report.

Effect Filtration

To see all damaging variations select "Effect" filter.
Setup filtration conditions to select only damaging variations with a tolerance score threshold less than 0.05:



Variation Navigator now shows all damaging variations (having a score less than 0.05). Double click on rs1799853 to see a report.



As we see here, the variations intersect 4 genes. The variation damages one of the genes (because a score less than 0.05), tolerated in two genes and one of the genes is predicted so now score is available.
Click on “Variations Report” on the main toolbar to output a report for all variations into a text file.
You can change global tolerance score threshold in the “Settings” menu. The system takes this score to decide what to write in the report: DAMAGING or TOLERATED. Do not mix up this score with the score in the “Effect” filter which is needed for filtration only.