Raphaël Helaers
raphael.helaers@uclouvain.be

Table of content

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

1 Introduction

1.1 Background

The field of human genetics is being revolutionized by exome and genome sequencing. A massive amount of data is being produced at ever-increasing rates. Targeted exome sequencing can be completed in a few days using NGS, allowing for new variant discovery in a matter of weeks. The technology generates considerable numbers of false positives, and the differentiation of sequencing errors from true mutations is not a straightforward task. Moreover, the identification of changes-of-interest from amongst tens of thousands of variants requires annotation drawn from various sources, as well as advanced filtering capabilities.

We have developed Highlander, a Java software coupled to a local database, in order to centralize all variant data and annotations from the lab, and to provide powerful filtering tools that are easily accessible to the biologist. Data can be generated by any NGS machine, (such as Illumina's HiSeq or MiSeq, or Life Technologies' Solid or Ion Torrent) and most variant callers (such as Broad Institute's GATK). Variant calls are annotated using DBNSFP (providing predictions from 6 different programs, splicing predictions, prioritization scores from CADD and VEST, and MAF from 1000G and ESP), ExAC, GoNL and SnpEff, subsequently imported into the database. The database is used to compute global statistics, allowing for the discrimination of variants based on their representation in the database. The Highlander GUI easily allows for complex queries to this database, using shortcuts for certain standard criteria, such as "sample-specific variants", "variants common to specific samples" or "combined-heterozygous genes". Users can browse through query results using sorting, masking and highlighting of information. Highlander also gives access to useful additional tools, including visualization of the alignment, an algorithm that checks all available alignments for allele-calls at specific positions, and a module to explore the 'variant burden' gene by gene.

1.2 Goal and scope

Highlander is a Java software coupled to a local MySQL database that aims to centralize all variant data coming from (targeted) exome- and whole genome sequencing experiments. It provides annotations, visualizations and powerful filtering tools that allow to detect changes-of-interest amongst the complete list of variants detected in a sample. It was developed by Raphaël Helaers, researcher at the De Duve Institute - Université Catholique De Louvain (UCL, Belgium) (http://sites.uclouvain.be/highlander/).

The analysis in Highlander is applicable to all variant data (SNVs and short indels) that originate from (targeted) exome- and whole genome sequencing experiments (e.g. MENDELIOME panel, WES).

1.3 Citing Highlander

Highlander is an Open Source software, under the GPLv3 licence. If you publish results using Highlander, please cite it using:

*Helaers R. and Vikkula M. – **Highlander: variant filtering made easy** (submitted)*

1.4 Definitions

- BAM: Binary Alignment/Map format
- IGV: Integrative Genomics Viewer
- Indel: insertion/deletion
- MNP: multiple nucleotide polymorphism
- MySQL: My Standard Query Language
- RDP: remote desktop
- SNV: single nucleotide variant
- VCF: Variant Call Format
- WES: whole exome sequencing

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

2 Installation

The Highlander system contains a server and a client component.

The Highlander software is composed on the server side of:

- A MySQL (v5.5+) Highlander Database
- An apache webserver for the distant consultation of BAM and VCF files.

The Highlander client can be downloaded from http://sites.uclouvain.be/highlander/download.html and can be installed locally on any number of computers. It only requires a computer running Windows, Mac OS X or Unix with Java 7 or above.

The "client" is downloadable with an embedded demonstration database, that doesn't need the server component and works all alone.

3 Pre-analysis steps

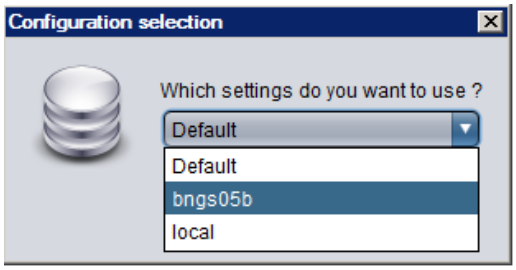
Before performing an analysis in Highlander, the samples need to be uploaded to the database. This is a two-step process that is executed by the Highlander administrators.

For VCF import, a Java companion tool (called "dbBuilder") is provided with Highlander. dbBuilder extracts all variants from the VCF, annotates them automatically using various biological databases, and imports them into the Highlander database.

4 Use of the software

4.1 Connection to the Highlander client

When launching Highlander, a screen to select the desired configuration appears.

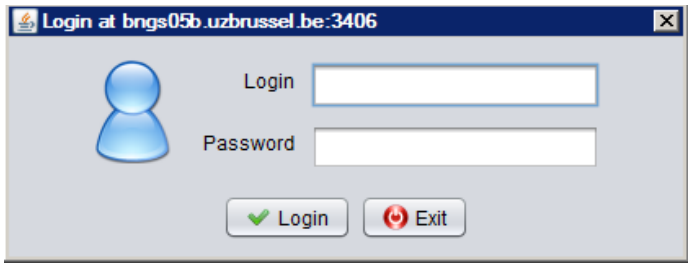


screen to select the highlander configuration.

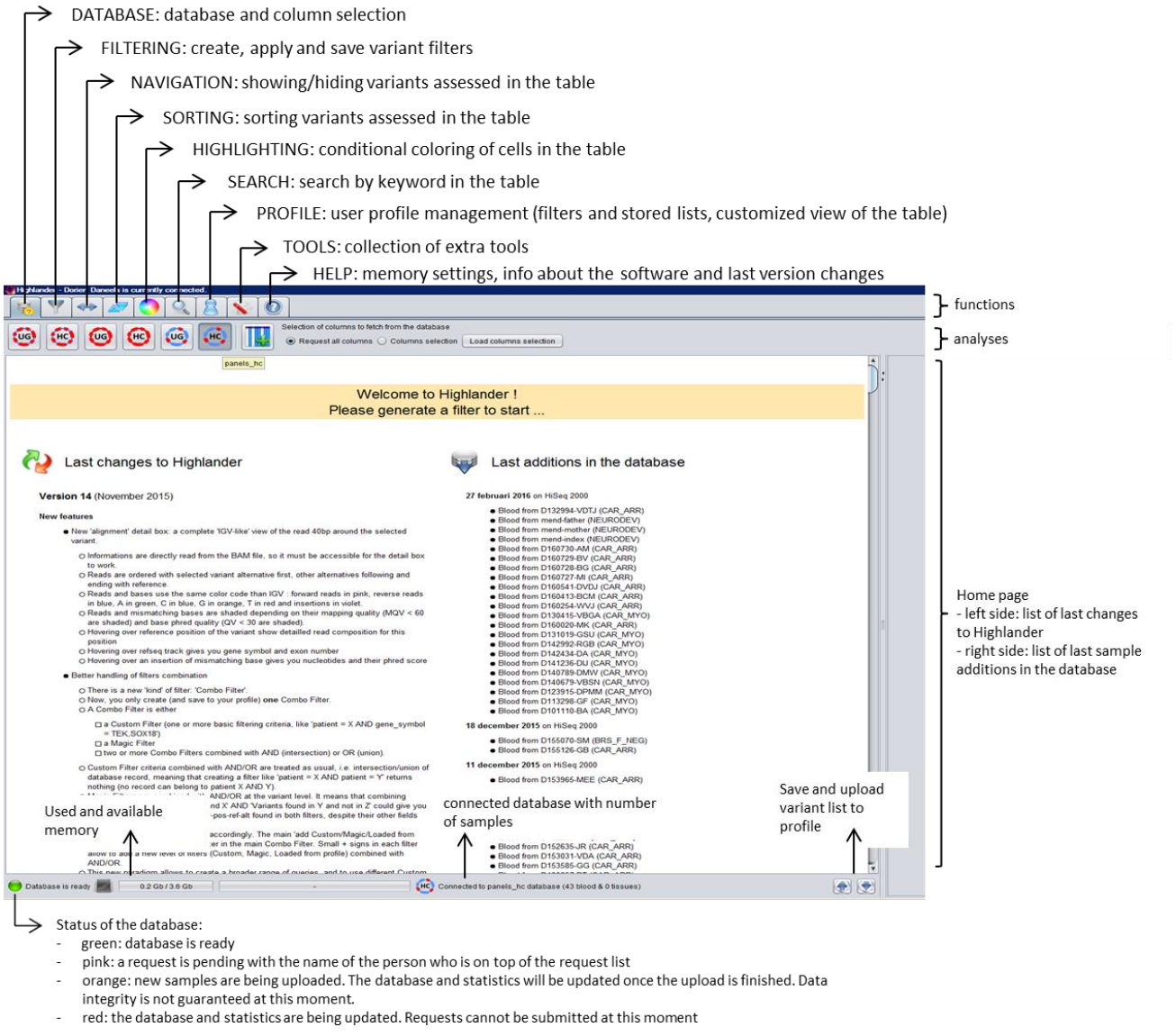
In a next step the user is asked to specify a login and password. This login is user-specific and was created by and provided to the user by the Highlander administrators. When the user logs in for the first time, the password equals the login. The user has the option to change his password (see below). The highlander client will now open.

The Highlander client comes with a standalone embedded database with public data from 1000 Genomes (chr22 of 80 samples). To test Highlander using this dataset, select the "demo" settings (instead of "default") in the dropdown list that appears when launching Highlander. Enter 'demo' as login and leave password field empty.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels



Login screen to access highlander.

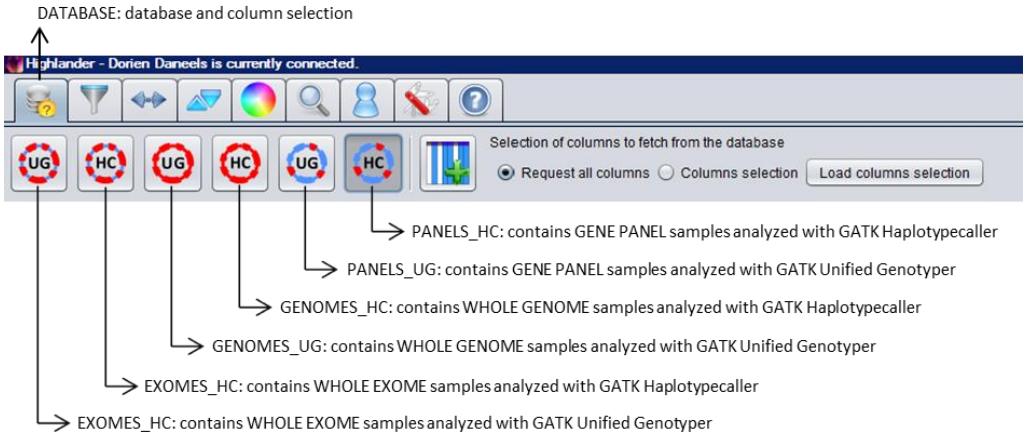


First screen that appears when opening the Highlander client with an explanation of all functions.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.2 Selection of a Highlander database

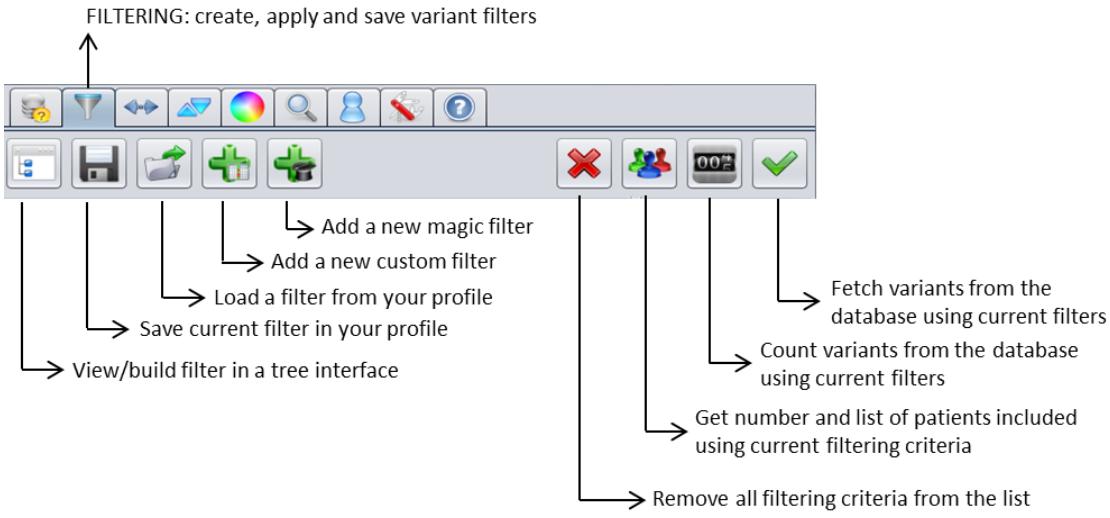
Before starting an analysis, the user should select the analysis (sub-database) in which the sample-of-interest is located. This is done by clicking on the database icon (that is selected by default) and clicking on one of several icons, depending on the type of the sample-of-interest. Highlander database can be subdivided in any number of sub-databases, *e.g.*:



DATABASE option with an explanation of the six databases present in the Highlander system.

4.3 Variant filtering in Highlander

After selecting the appropriate database, a filter should be created or loaded to start the variant analysis. This is done by selecting the filtering function. There are two types of filters: custom and magic filters.

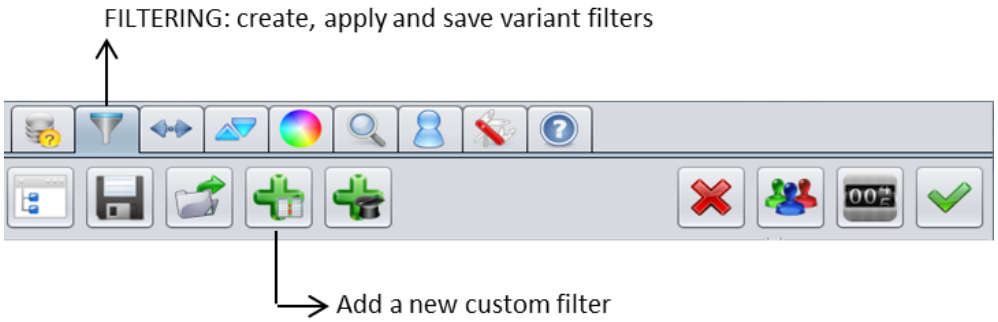


Overview of the FILTERING function

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.3.1 Custom filters

Creating a custom filter starts by selecting the filter function and clicking on the 'custom filter' icon.



FILTERING function with the selection of a custom filter.

Filters can be created for any column of the database (i.e. database fields). A column represents a type of information related to the sample or variant. Every column belongs to a certain category (i.e. database fields categories). An overview of these categories can be found in the table below. A description of all database fields is available within highlander, using tooltips (just leave your mouse cursor on a field name).

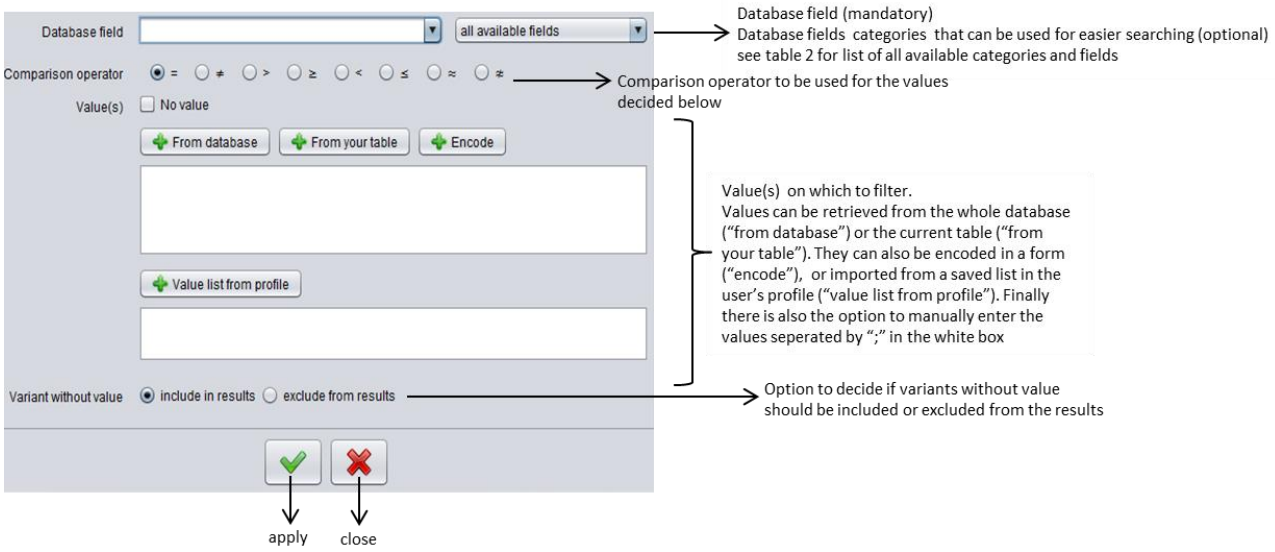
Overview and short description of the available database field categories in the FILTER function.

Database fields category	Description
Allele frequencies in population	Variant allele counts and allele frequencies for multiple public variant databases (e.g. dbSNP, 1000genomes, ExAc, ...)
Calling in other analysis	Flag set to "true" if the same variant has been found in other databases of the highlander system
Change details	Variant details (e.g. position and change at genomic, cDNA and protein level, gene symbol, ...)
Change statistics	Number of blood/tissue samples in the different Highlander databases having the same variant at the same position as the one selected.
Change statistics by group	Number of blood/tissue samples in the different Highlander databases, subdivided by the different pathologies, having the same variant at the same position as the one selected.
Confidence	Variant confidence scores (e.g. mapping quality, strand bias, ...)
Conservation scores	Variant conservation scores (e.g. gerp, phyloP, phastCons, ...)
Coverage	Variant read- and allelic depth details
Effect prediction	Variant pathogenicity predictions (e.g. SIFT, PolyPhen2, CADD, ...)
Error detection	Flag set to "true" if the same variant has been detected as an error by an internal algorithm
Gene statistics	Number of blood/tissue samples in the different Highlander databases carrying a variant in the same gene as the selected variant
Gene statistics by group	Number of blood/tissue samples in the different Highlander databases, subdivided by the different pathologies, carrying a variant in the same gene as the selected variant
Genotype	Variant zygosity and genotype quality
Miscellaneous	Time of insertion of this variant in the database
Other scores	Other variant pathogenicity predictions (e.g. aGVGD, effect on splicing, ...)
Related identifiers	Variant identifiers in different databases (e.g. clinvar, cosmic, dbsnp, ensembl, hgmd, ...)

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

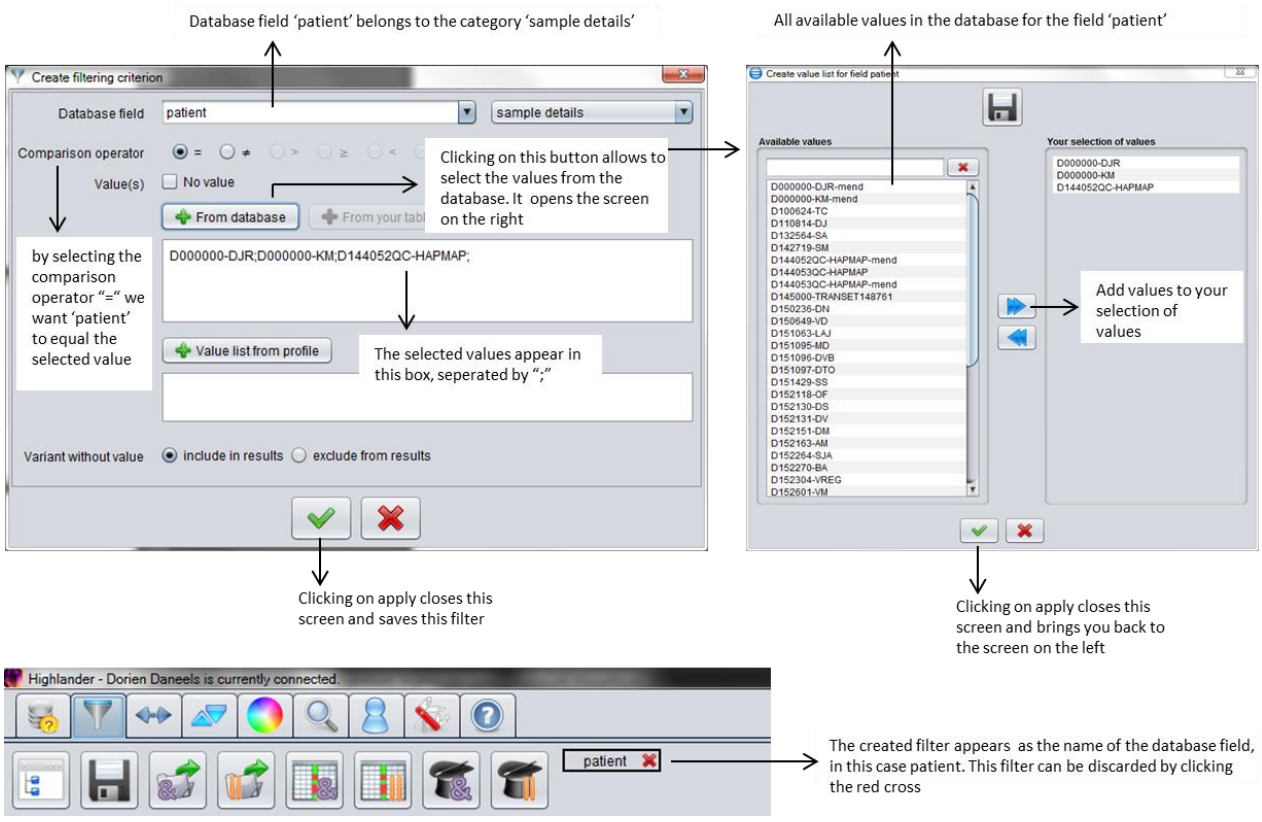
Sample details	Sample related data that was specified during sample upload
User annotations	Variant annotations and comments introduced by the user (e.g. in silico evaluation, variant pathogenicity, ...)

Next figure shows the screen to create a custom filter. The user should select a database field and define on which values the filter should be applied. At the bottom the user has the option to decide if variants without value should be included or excluded from the results. When finished, the user should click on the green check mark to save the filter or on the red cross to discard the filter.



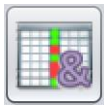
Screen that opens when selecting the option 'create custom filter'.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels



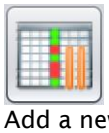
Example of creating a custom filter. The selected database field is 'patient' that belongs to the 'sample details' category. The filter will select the samples for which database field 'patient' equals 'D000000-DJR' or 'D000000-KM' or 'D144052-QC'.

Filters can be added to an existing custom filter using the logical operator "AND" (i.e. results will be the INTERSECTION of the different filters). This can be done by clicking on the icon shown below.



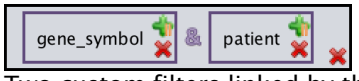
Add a new custom filter to an existing one using the logical operator "AND".

Filters can be added to an existing custom filter using the logical operator "OR" (i.e. results will be the UNION of the different filters). This can be done by clicking on the icon shown below.



Add a new custom filter to an existing one using the logical operator "OR".

If two filters are added using the logical operator "AND", the symbol for "OR" appears in the upper right corner of the filter box. This can be used to add a new sub-criterion to the filter, using the logical operator "OR" (i.e. Filter will be the union of sub-criteria).



Two custom filters linked by the 'AND' operator.

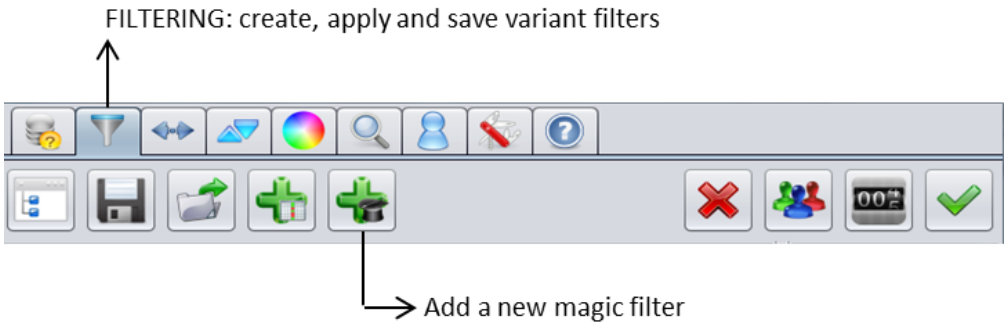
CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

Conversely, if two filters are added using the logical operator "OR", the symbol "AND" appears in the upper right corner of the filter box allowing to add a new sub-criterion to the filter using the logical operator "AND".

Two custom filters linked by the 'OR" operator.

4.3.2 Magic filters

Creating a magic filter starts by selecting the filter function and clicking on the 'magic filter' icon.

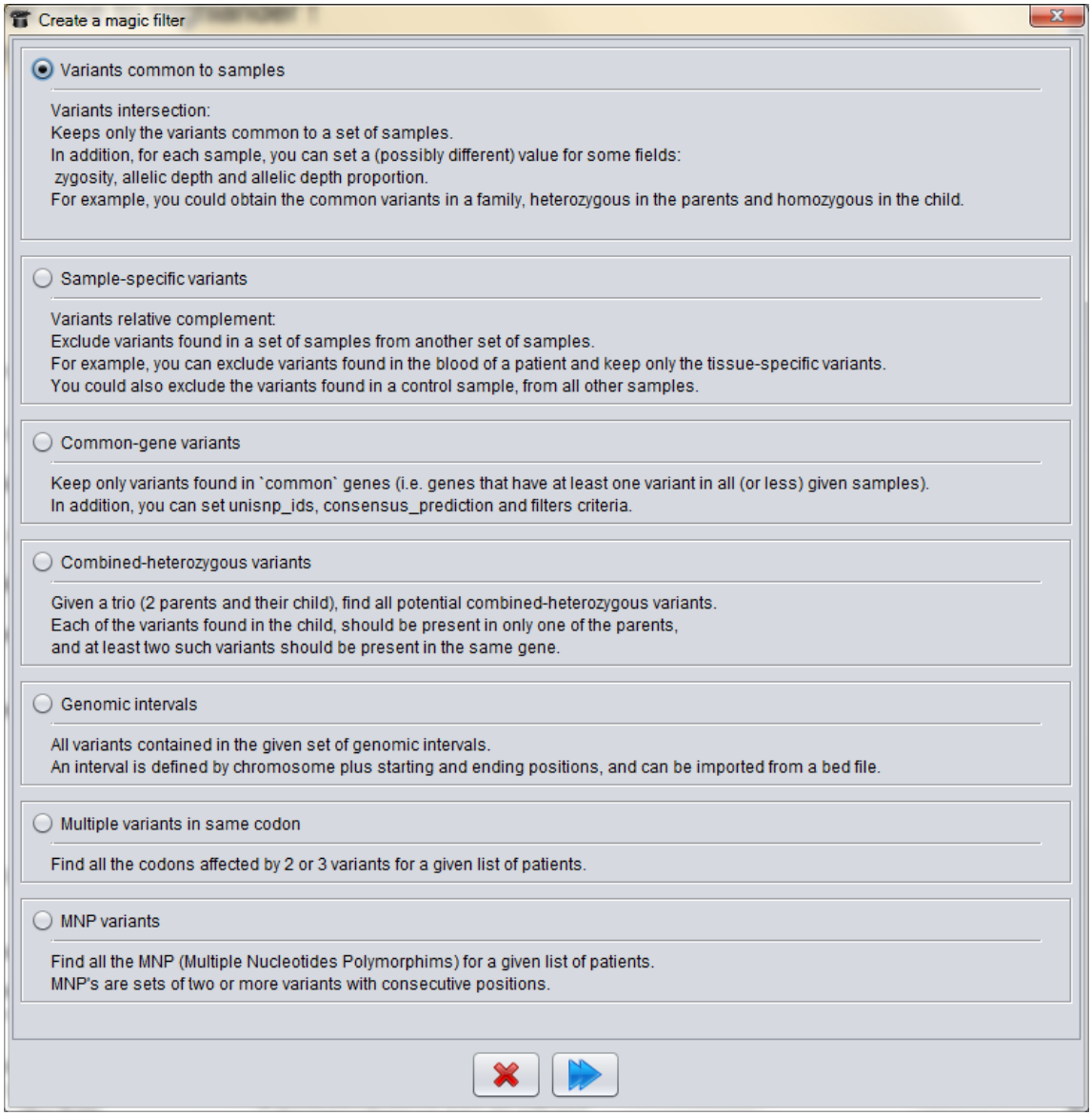


FILTERING function with the selection of a magic filter.

Seven different magic filters exist and are described below: variants common to samples, sample-specific variants, common-gene variants, combined-heterozygous variants, genomic intervals, multiple variants in same codon, MNP variants.

Each magic filter allows a pre-filter: only variants resulting from this pre-filter will be taken into account in the magic filter.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

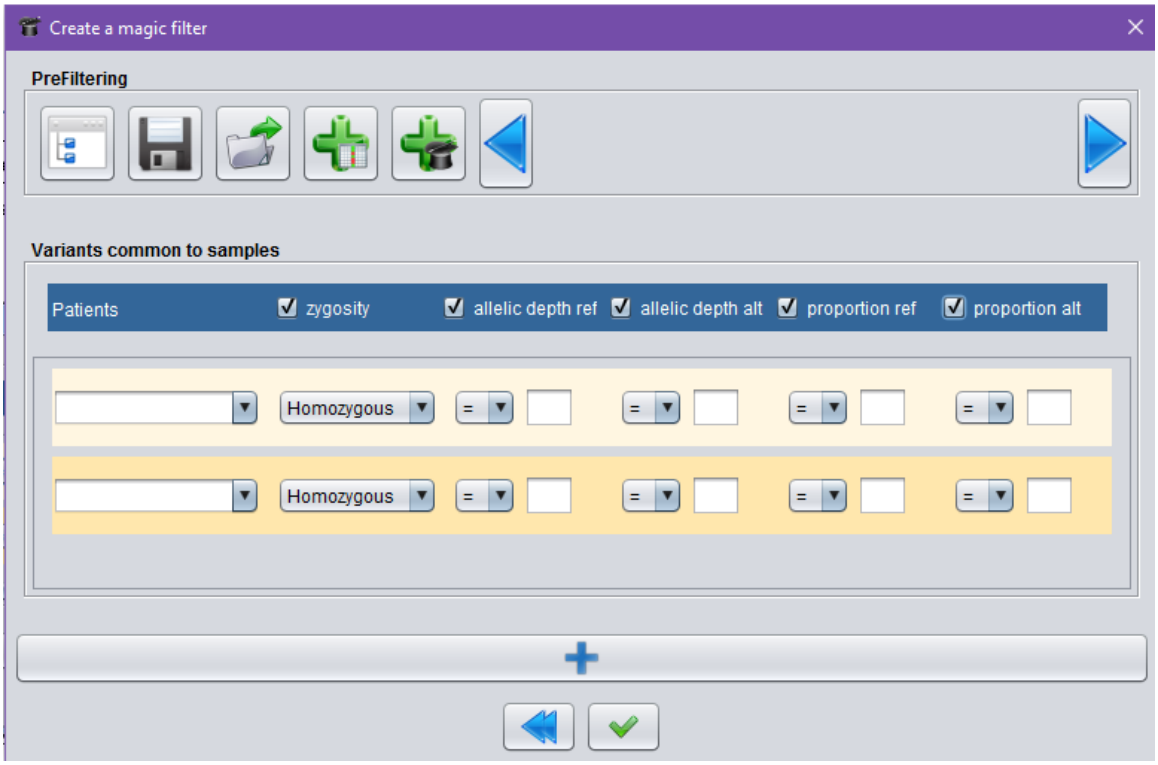


Overview of the seven different magic filters.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.3.2.1 Magic filter - Variants common to samples

This magic filter is designed to select the variants that are common to a set of samples. For each sample, the user can set a (possibly different) value for the zygosity, allelic depth and allelic depth proportion fields. This filter can for example be used to obtain common variants in multiple affected family members or to obtain variants that are heterozygous in the parents and homozygous in the child.

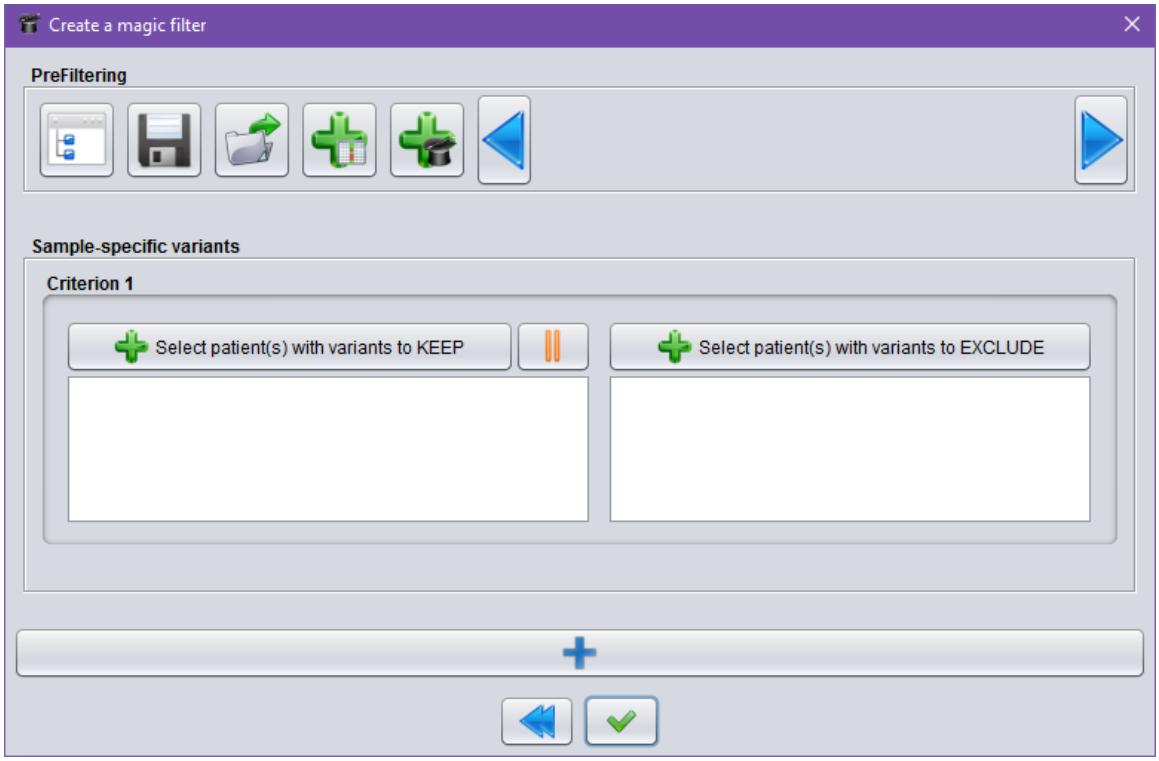


Box with which to create the magic filter 'variants common to samples'.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.3.2.2 Magic filter - Sample-specific variants

This magic filter is designed to exclude variants from a set of samples that are also present in another set of samples. This filter can for example be used to exclude the variants found in controls from a set of affected samples, or exclude variants in a patient's blood sample and keep only the tissue specific variants. This filter is also useable for a classic trio analysis, where one wants to exclude the variants found in the parents and only keep the 'de-novo' variants in the child.

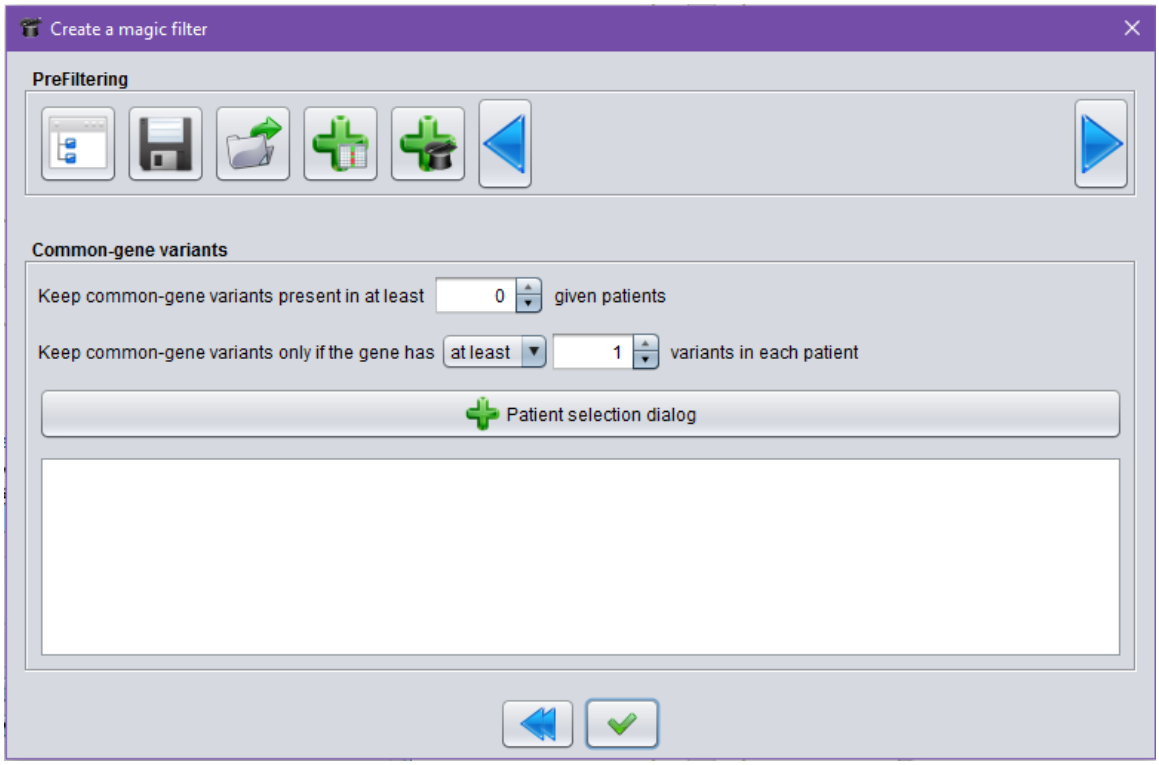


Box with which to create the magic filter 'sample-specific variants'.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.3.2.3 Magic filter - Common-gene variants

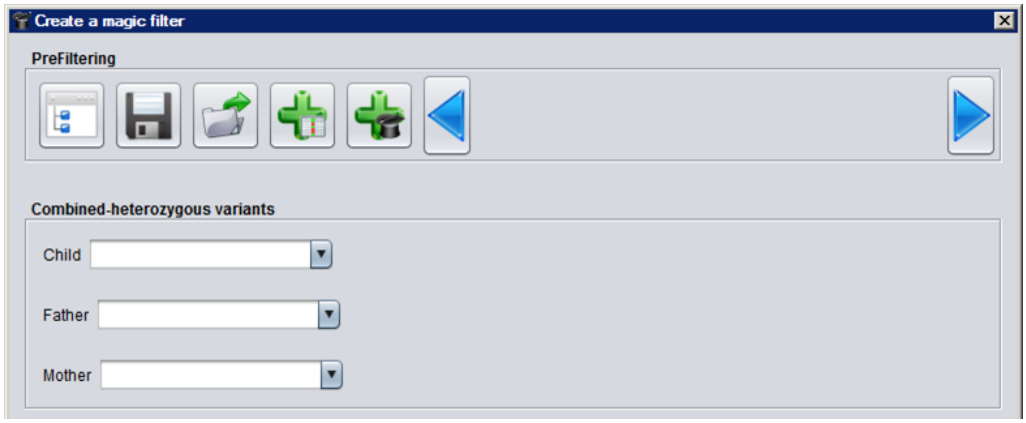
This magic filter is designed to select variants found in "common" genes (i.e. genes that have at least one variant in the selected samples).



Box with which to create the magic filter 'common-gene variants'.

4.3.2.4 Magic filter - Combined-heterozygous variants

This magic filter is designed to select all potential combined-heterozygous (i.e. compound heterozygous) variants from a trio (two parents and one affected child). Each of the variants found in the affected child, should be present in only one of the parents and at least two such variants should be present in the same gene.

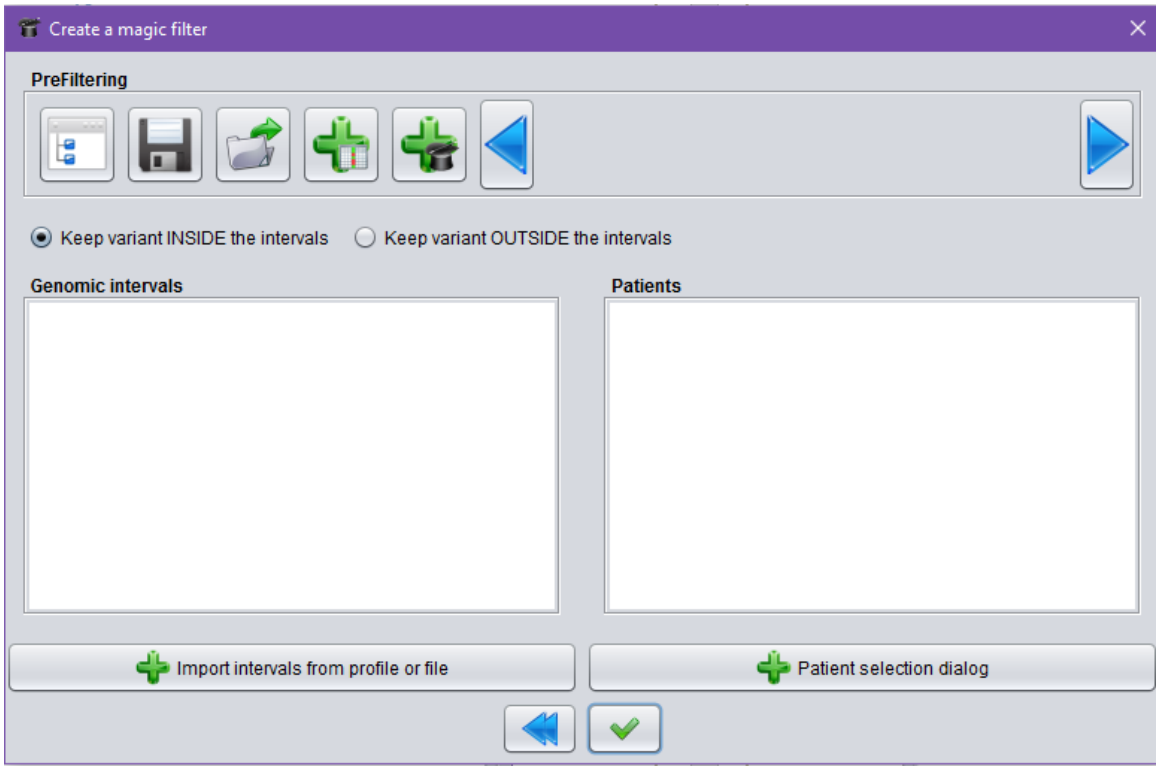


Box with which to create the magic filter 'combined-heterozygous variants'.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.3.2.5 Magic filter - Genomic intervals

This magic filter is designed to in- or exclude all variants contained in a given set of genomic intervals for the selected samples. An interval is defined by the chromosome and the start- and end position (i.e. 6:13000000-14000000) and can be either entered manually or imported from a region file saved in the user's profile.

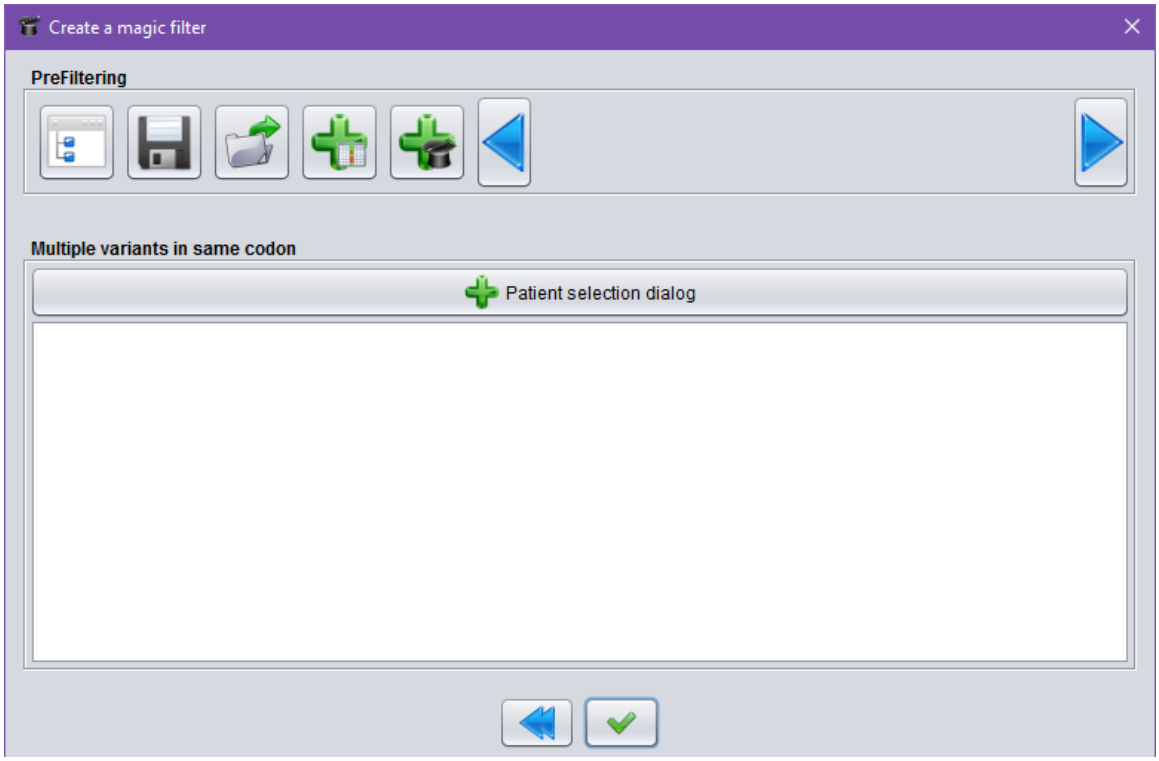


Box with which to create the magic filter 'genomic intervals'.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.3.2.6 Magic filter - Multiple variants in same codon

This magic filter is designed to find all the codons affected by two or three variants for a given list of samples.



Box with which to create the magic filter 'multiple variants in same codon'.

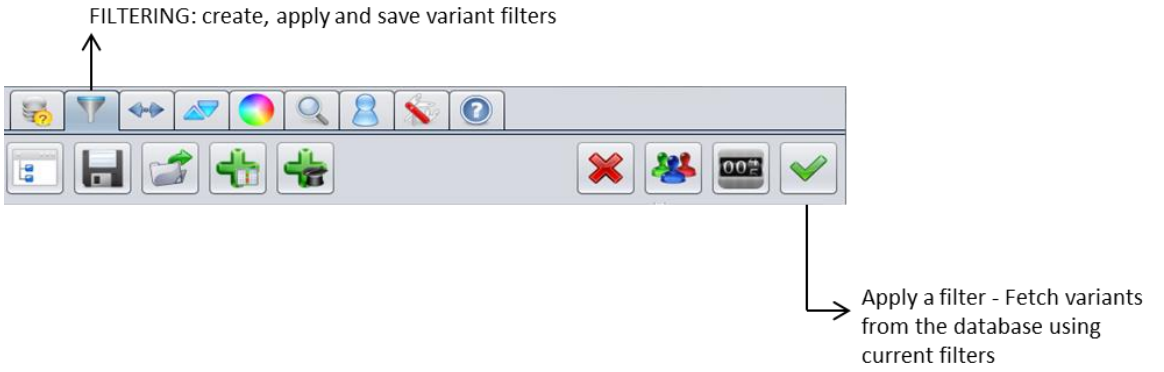
4.3.2.7 Magic filter - MNP variants

This filter is designed to select all multiple nucleotide polymorhism (MNP) variants for a given list of samples. MNPs are sets of two or more variants with consecutive genomic positions. This filter allows to detected potential errors of nomenclature in case there are two or more consecutive substutions on the same strand. By default, Highlander considers two consecutive substitutions as being in *trans*. Therefore, if they are in *cis*, the protein nomenclature is not correct.

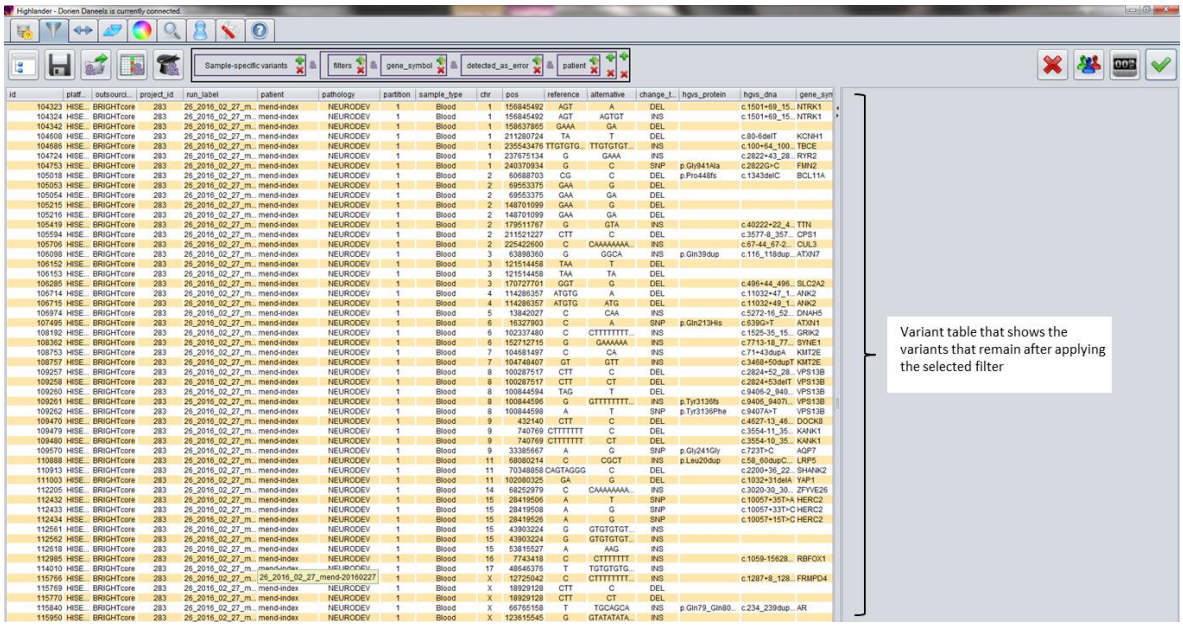
CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.3.3 Applying a filter

A filter can be applied by clicking on the green check-mark. When this is done, the variants selected based on the current filtering criteria are fetched from the database and shown in the variant table.



Applying a filter.



Variant table that shows the variants that remain after applying the selected filter. Each line represents one variant.

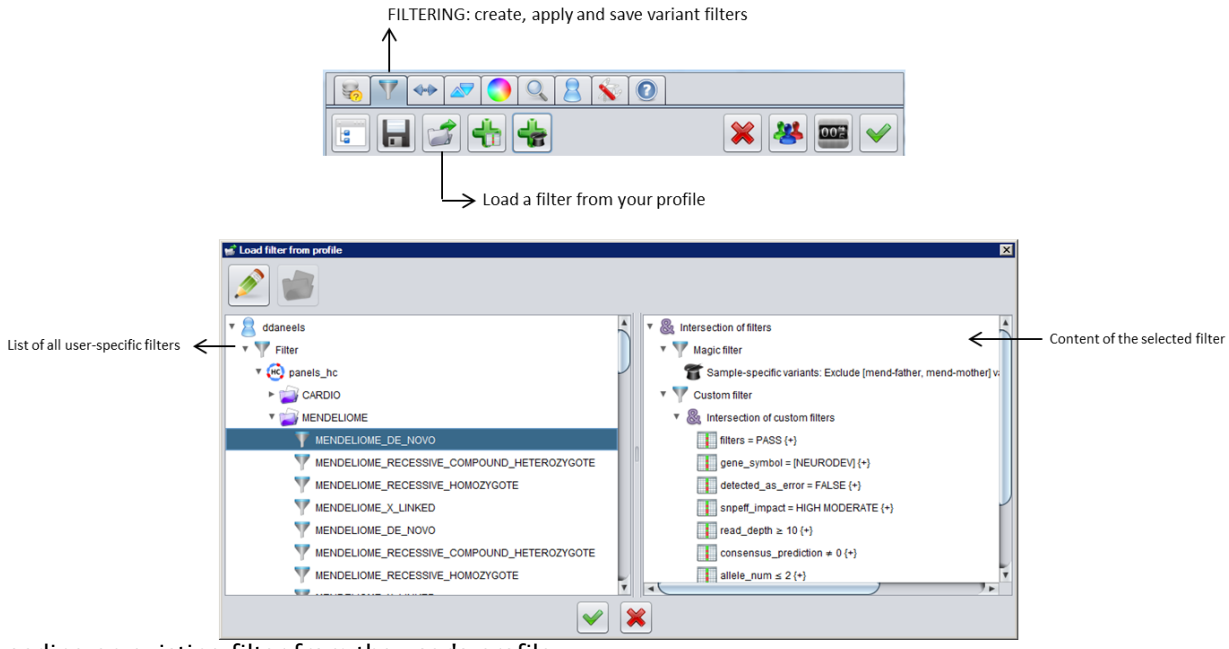
CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.3.4 Loading and modifying an existing filter

Besides creating a new filter, there is also the possibility to load and modify an existing filter from the user's profile.

4.3.4.1 Loading an existing filter from the user's profile

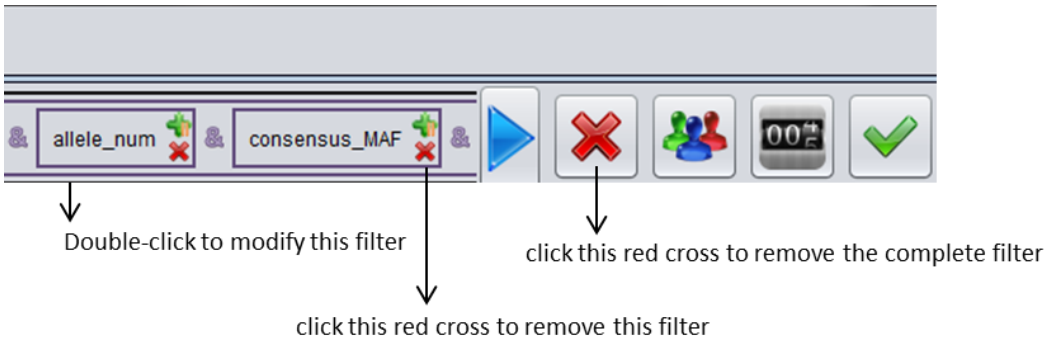
Loading an existing filter starts by selecting the FILTERING function and clicking on the 'load filter' icon. A screen opens that shows all the filters saved in the user's profile.



Loading an existing filter from the user's profile.

4.3.4.2 Modifying an existing filter

When a filter is selected, each criterion can be changed by double-clicking on the filter or removed by clicking on the red cross.

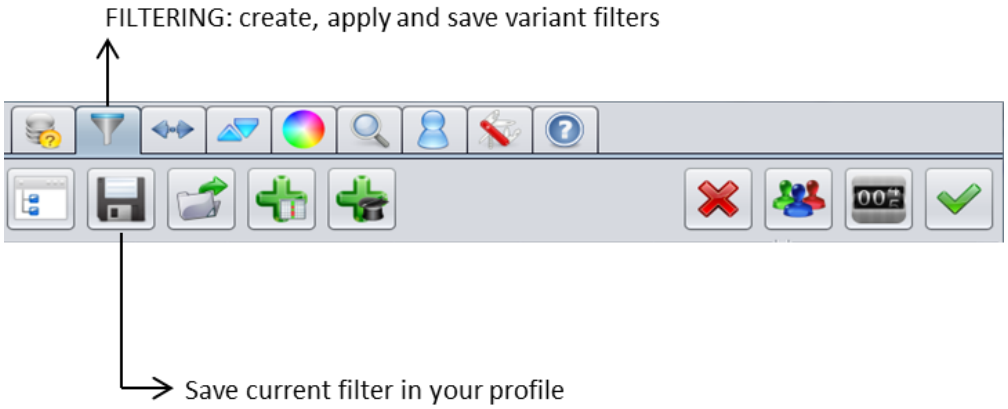


Modifying and deleting a filter.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.3.5 Saving a filter

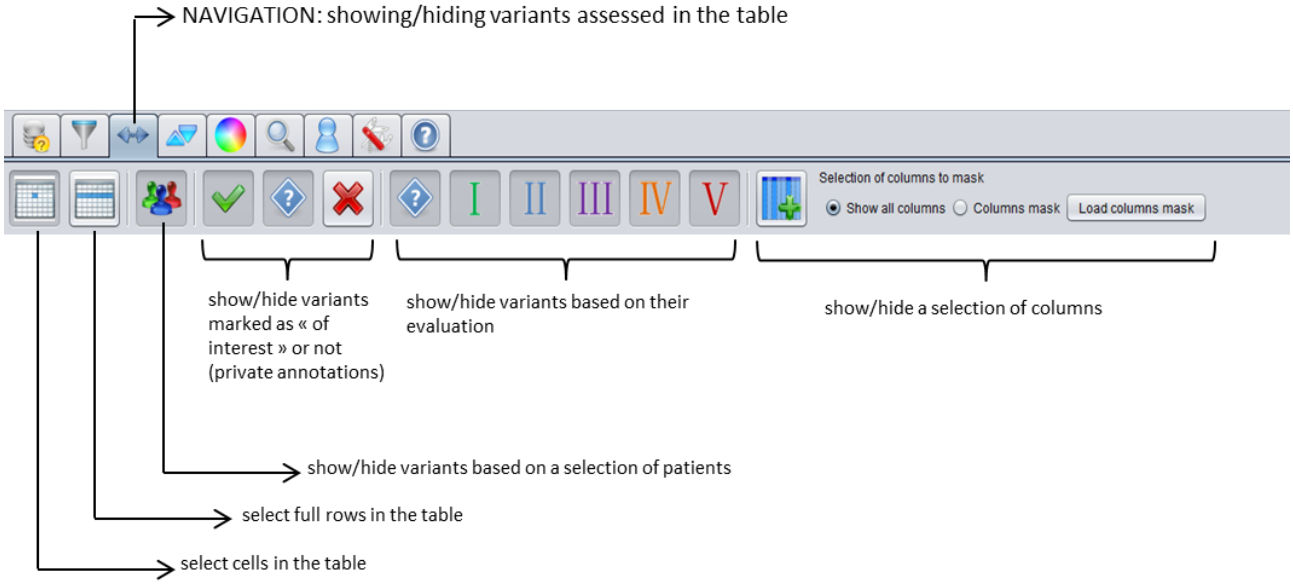
A filter can be saved to the user's profile by clicking on the 'save' icon in the FILTERING function.



Saving a filter to the user's profile.

4.4 Navigation function

The NAVIGATION function is designed to show and hide variants assessed in the variant table. Next figure gives an overview of the available options in the navigation function.

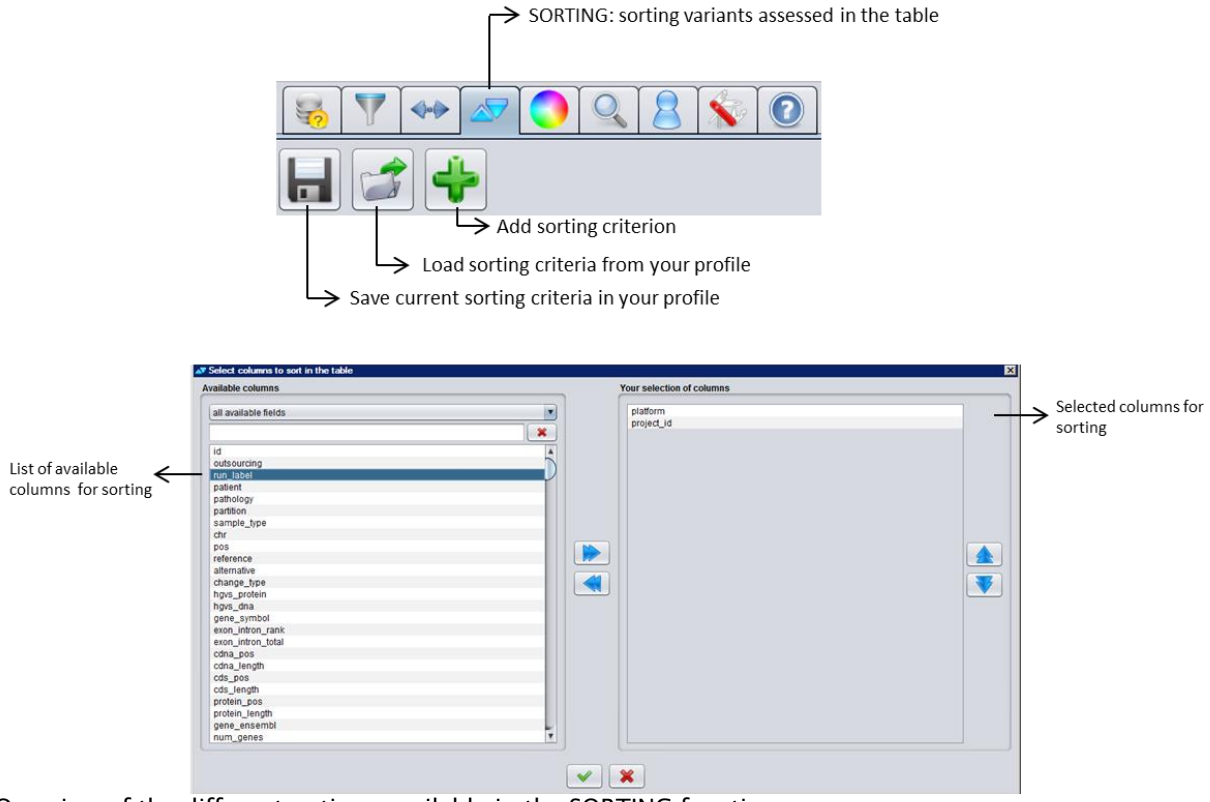


Overview of the different options available in the NAVIGATION function.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.5 Sorting function

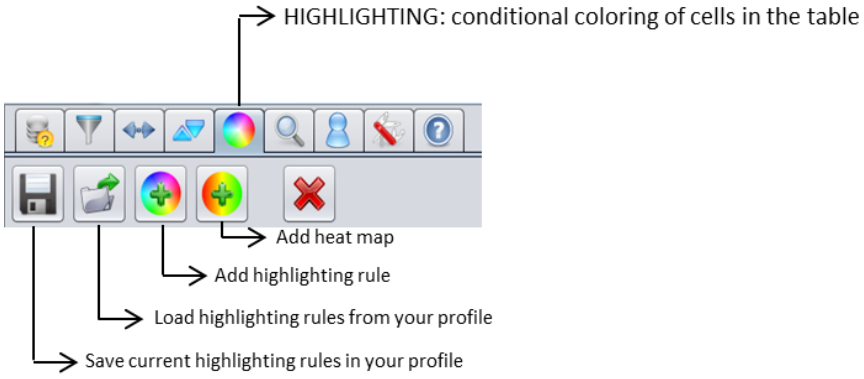
The SORTING function is designed to sort the variants assessed in the variant table. Next figure gives an overview of the available options in the sorting function. Sorting criteria can be created, but also loaded from and saved in the user's profile.



Overview of the different options available in the SORTING function.

4.6 Highlighting function

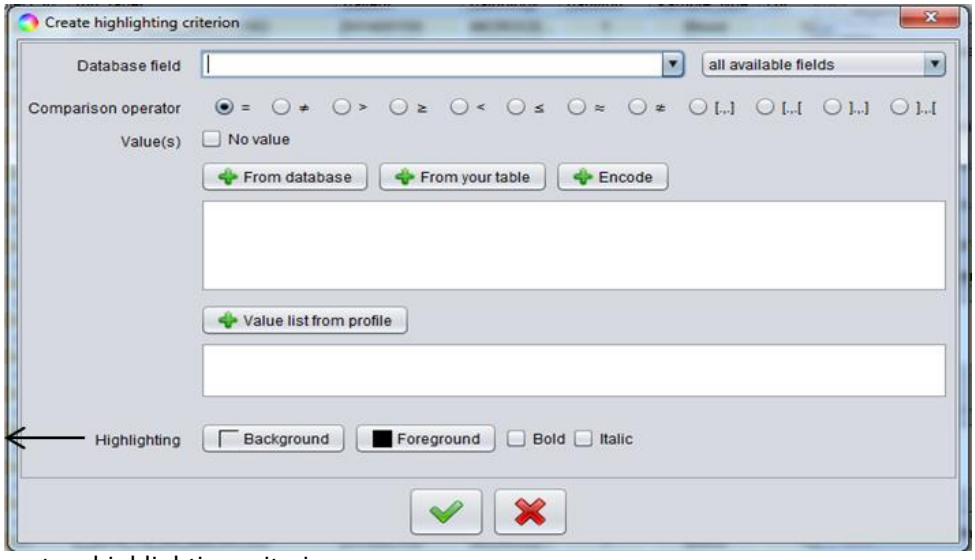
The HIGHLIGHTING function is designed to put a selection of cells in the variant table under attention, by means of colors or fonts. Next figure gives an overview of the available options in the highlighting function. Highlighting criteria can be created, but also loaded from and saved in the user's profile.



Overview of the different options available in the HIGHLIGHTING function.

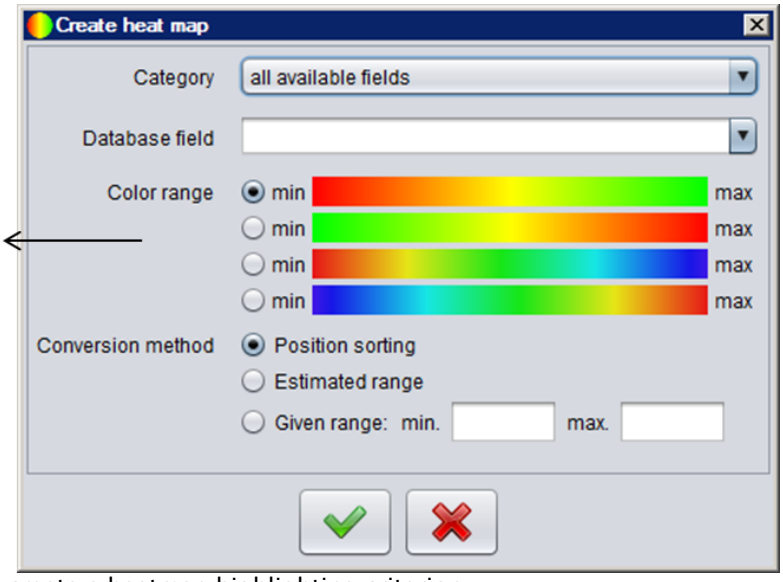
CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

Clicking on the button 'add highlighting rule' opens a window similar as the one for defining a custom filter, where the colors and font styles for highlighting can be set at the bottom.



Box with which to create a highlighting criterion.

Clicking on the button 'add heatmap' opens a window where the color ranges for the selected database field can be set.

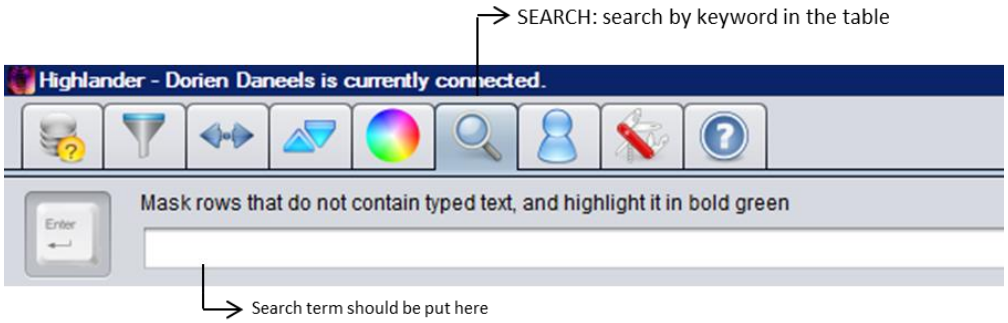


Box with which to create a heatmap highlighting criterion.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.7 Search function

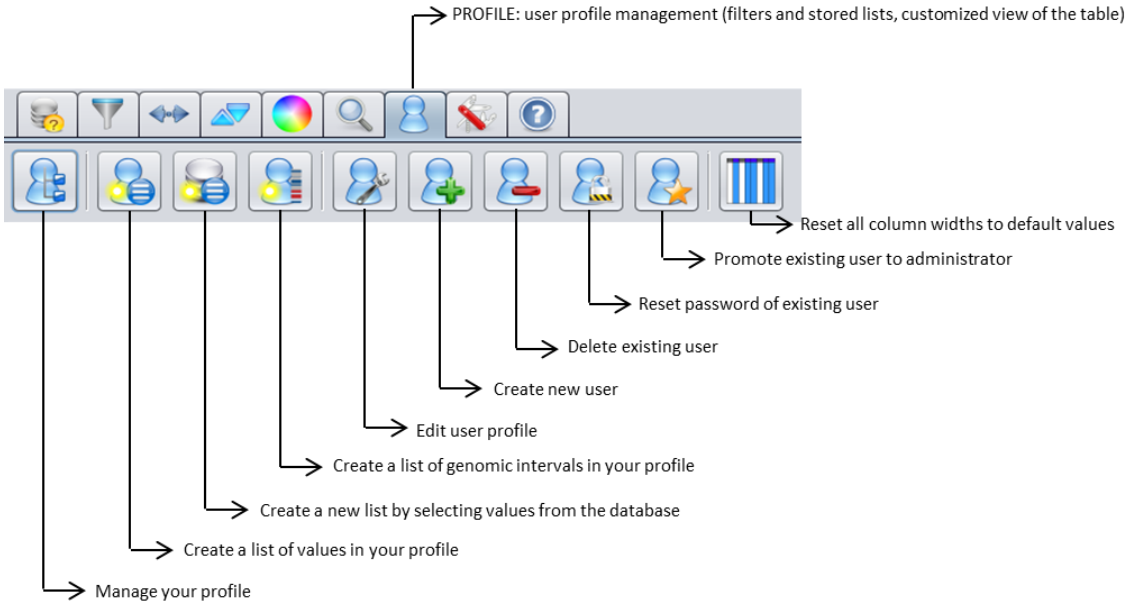
The search function is designed to search for any type of text or values in the variant table. Rows containing the searched item are highlighted in bold green. Next figure gives an overview of the available options in the search function.



Overview of the different options available in the SEARCH function.

4.8 Profile function

The profile function is designed to manage the user's profile. Next figure gives an overview of the available options in the profile function.

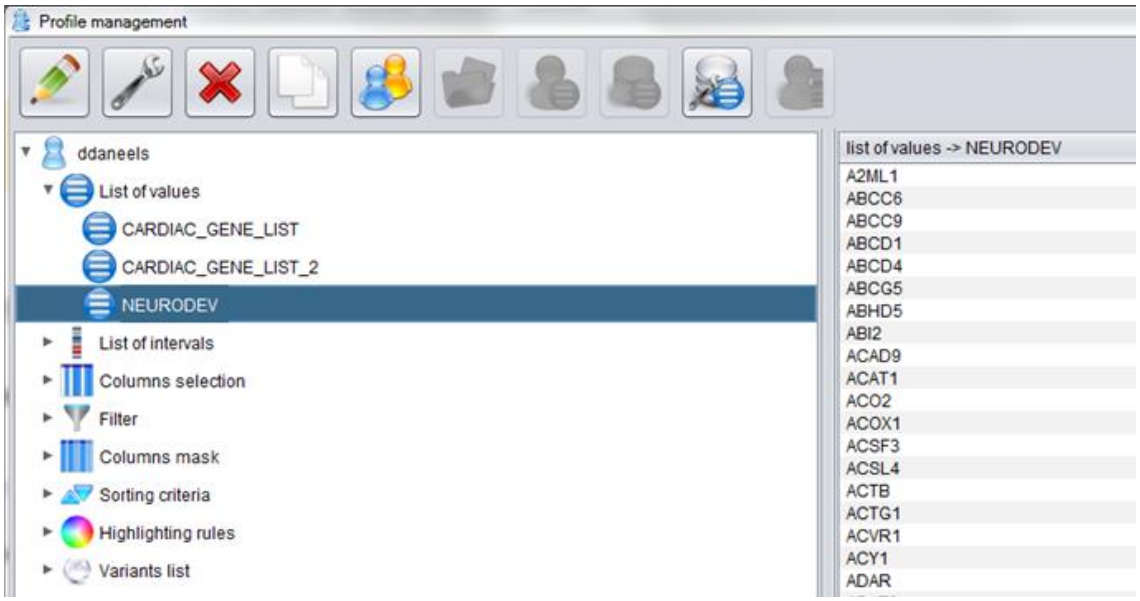


Overview of the different options available in the PROFILE function.

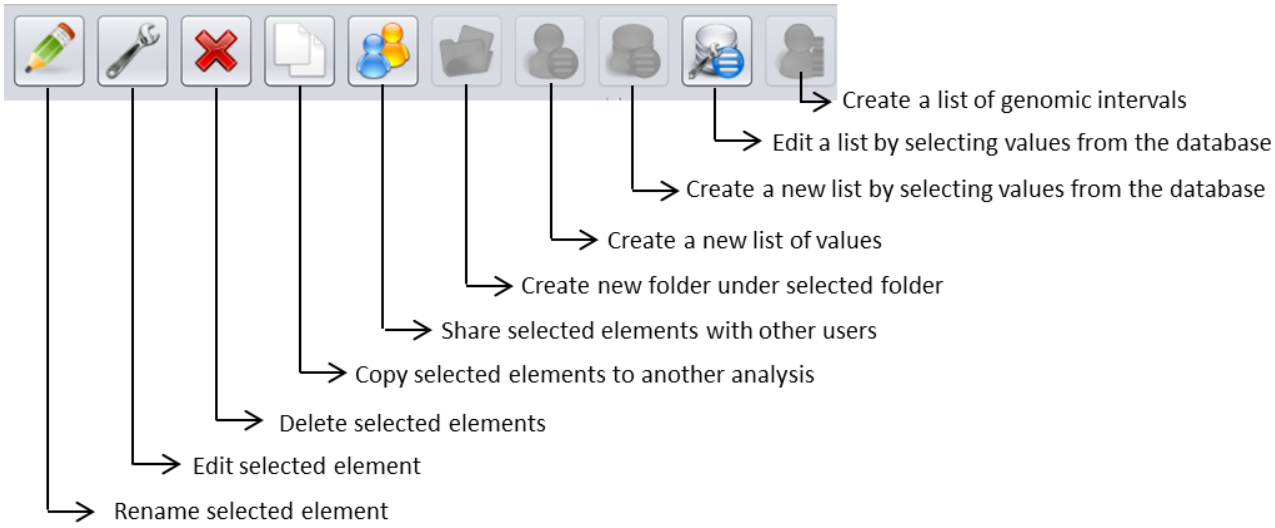
CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.8.1 Manage user's profile

Clicking on the button 'manage your profile' opens a windows that shows an overview of all the stored items in the user's profile. The different options to handle these items are described in next figure: one can rename, edit, delete, copy and share a selected element.



Overview of the profile management option in the PROFILE function.

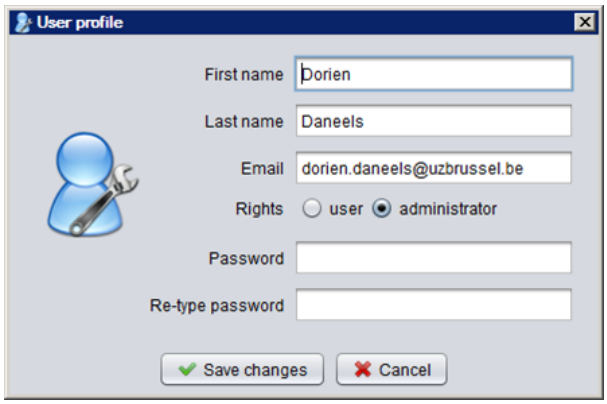


Overview of the options in the profile management options of the PROFILE function.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.8.2 Edit user's profile

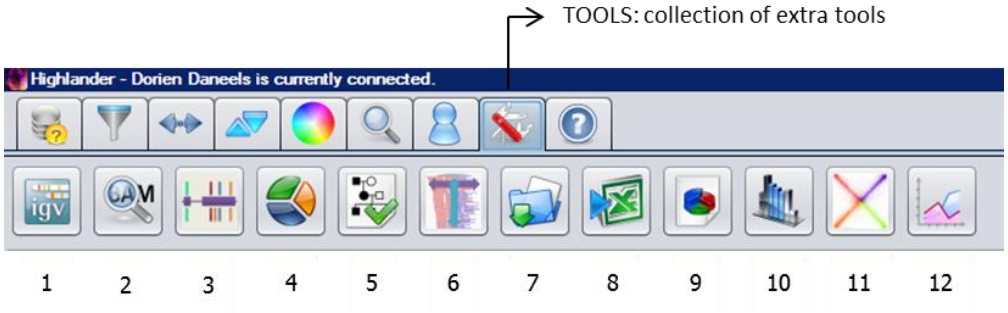
Clicking on the button 'edit your profile' in the PROFILE function opens a window in which you can change your name, email-address and password.



Box with which to edit the user's profile.

4.9 Tools function

Besides different filtering options, highlander also contains a collection of tools to aid with the variant interpretation.



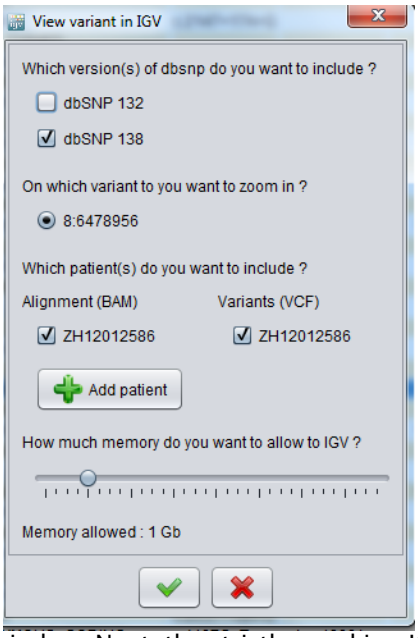
1. IGV: visualization of the aligned reads covering a selected variant for a selection of samples.
2. BAM viewer: allows to determine the proportion of samples where the selected variant is present.
3. Burden test using current custom filters
4. View selected variant statistics from another analysis
5. Check familial relationships between patients
6. Get average coverage for a given set of genes or regions
7. Download, BAM and VCF files for current analysis
8. Export current table content to MS Excel
9. View all detailed run statistics
10. View all detailed run statistics charts
11. View run report associated to the selected variant (currently not in use)
12. View run summary associated to the selected variant (currently not in use)

Overview of the different tools available in the TOOLS function.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.9.1 IGV – aligned reads visualization

Reads under each variant can be viewed through the IGV software that is automatically installed with Highlander.



Selecting a dbSNP version allow to view, together with variants, SNPs listed in dbSNP. The dbSNP versions available depend on what has been installed initially.

If a variant has been selected, the software opens at its genomic position. If no variant is selected, IGV opens to a blank page.

Both BAM (the reads) and associated VCF (list of variants) of the selected sample can be include.

Reads and variants of several patients can be viewed simultaneously (use Add patient to add patient(s) from the database).

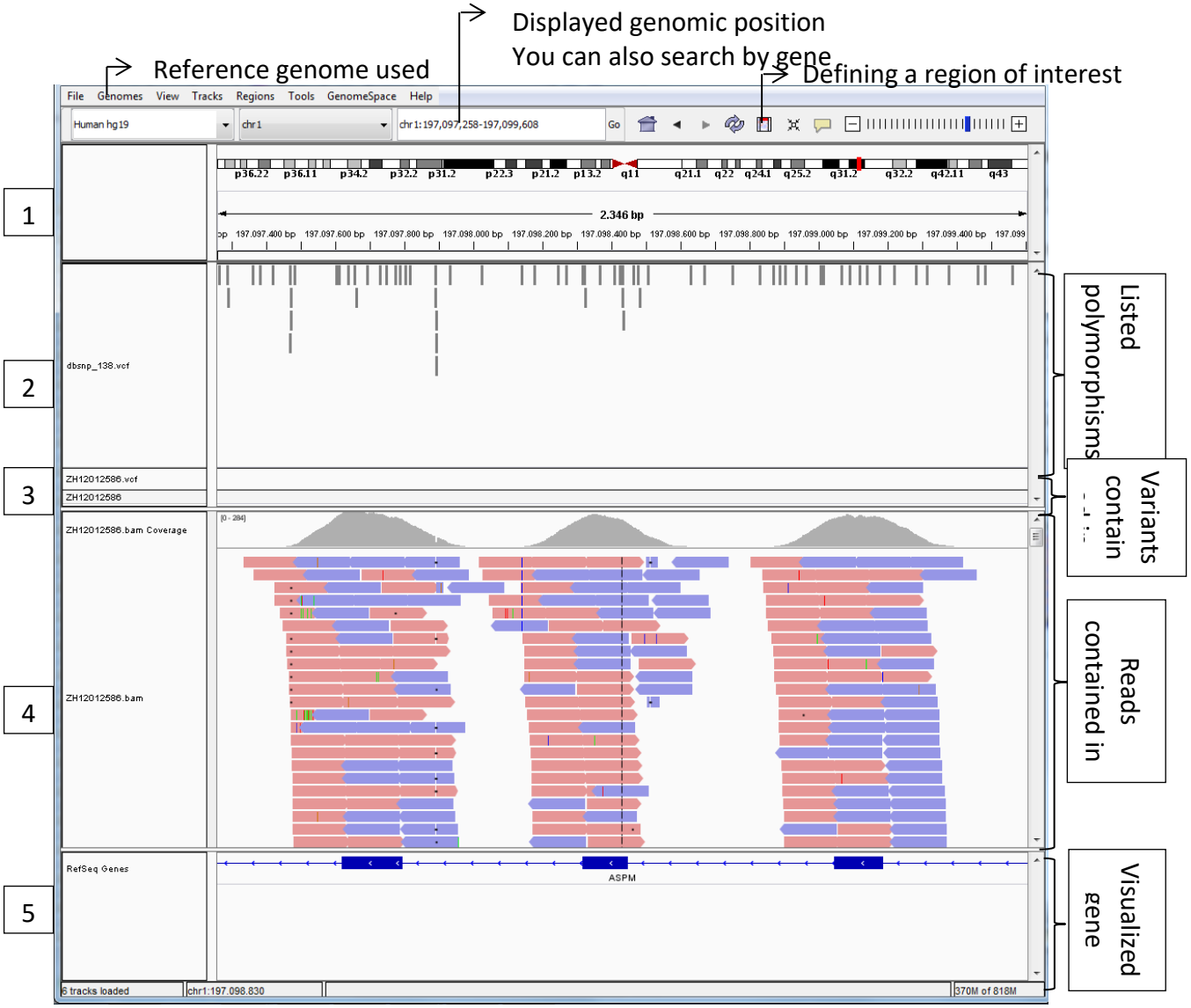
The amount of memory allocated to the software depends on the memory of the computer. The more allocated memory, the more IGV will work quickly.

When you click on the  a window titled "Console" opens and lets you know the status of the loading of the IGV window. Next, the strictly speaking IGV window opens.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

The window is divided into 5 tracks :

1. Display the location of the region analyzed on the chromosome
2. dbSNP: displays polymorphisms listed in the dbSNP selected version. One track by dbSNP version selected.
3. VCF: to view the variants listed in the VCF(s) selected. One track by VCF / sample selected.
4. BAM: used to display reads listed in BAM(s) selected. One track by BAM / sample selected.
5. Sequence and RefSeq genes: displays the reference sequence (selected in the top left) and the RefSeq gene (annotated by UCSC database) in the required position.



CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

By clicking the right button in the different panels, a series of options is available to customize the window. For more detailed information, see the "help" function of the software.

Useful features:

- **Change the size of the track**:
  - Right-click in the track
  - Positioning on a separation between 2 panels, left click and move up or down depending on the desired height.
- **Color/order/group** reads depending on the orientation of the sequence, the starting position: right-click on the BAM track.
- **Change position on the genome:** click anywhere in the window and drag left or right with mouse
- **Zoom**:
  - In the "chromosome" window, select the area to zoom
  - Use the + and - buttons at the top right of the window
- **Detailed information on polymorphism**:
  - Left-click on the polymorphism in the "dbSNP" track
  - Display of the genomic position, allele, frequency … if available
- **Detailed information on the variant:**
  - Left click on a variant in the "bam" track
  - Visualization of the genomic position, alleles and quality scores (Q score, mapping quality ...).
- **Detailed exon information:** left-click on a gene exon
- **Copy the reference sequence of a region:**
  - Select the region of interest: click on the "Define a region of interest", click in the window to the desired position to determine the lower and upper limits of the region of interest. A red banner appears.
  - Right-click on the red banner
- **Copy the consensus sequence of a region of the sample analyzed:**
  - Right-click the "bam" panel or a read
  - Select "copy consensus sequence"
- **Copy the sequence of a read:**
  - Right-click on the read
  - Select « copy read sequence »
- **Copy the sequence of an exon:**
  - Right-click on the exon
  - Select « Copy sequence »
- **See read pairs:**
  - Right-click on the BAM track
  - Select « View as pairs »
- **View the different gene transcripts analyzed:**
  - Click on the track « RefSeq Genes »
  - Select « Expanded » or « Squished », track will take less space.

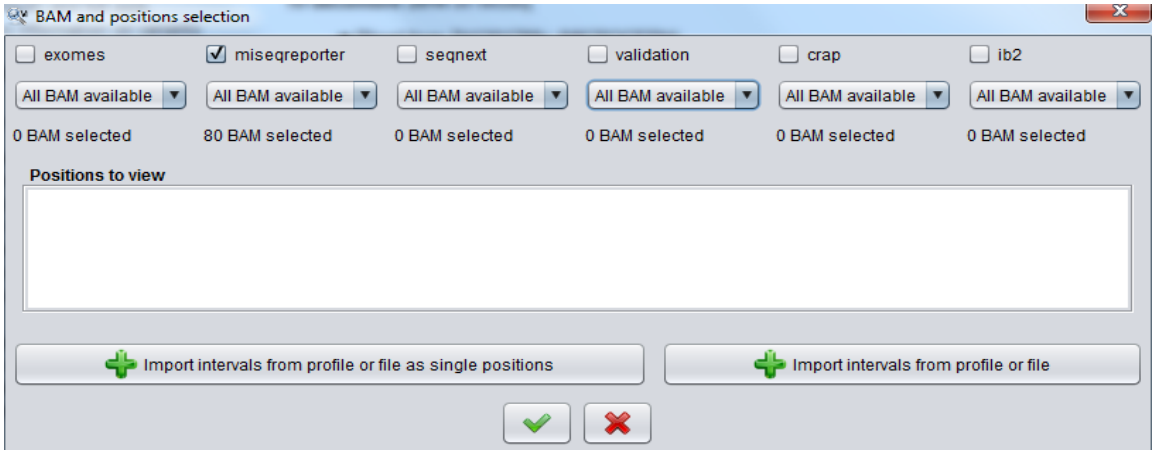
CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.9.2 Frequency of a variant in BAMs from the database

Function to look for the presence of variant, even at low level, for a certain position in different BAMs. This shows whether a variant is recurrent but not called by the pipeline, for example because it does not pass filters.

It is possible to:

- select BAMs to analyze: the entire database or individual BAMs in the various analyzes
- list different positions or select a list of pre-established genomic positions



An Excel file is then generated:

- A sheet by analyzed position
- Each sheet includes a column with the name of the BAM wherein a variant was found (one BAM per line), a column with the number of reads found at this position and then as many columns as allele found in this position in the various BAMs

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.9.3 Burden test using current cusotm filters

Function for comparing the frequency of variants of the database (filtered according to certain criteria) with respect to the frequency of ExAC or GoNL variants filtered with the same criteria

a. Toolbar

- Color scale corresponding to the frequency of variants
- Information on the number of different variants selected in the table. If a variant is identified in several samples in Highlander, it is counted once
- Information about the transcript used
- Zoom-function
- Chi-squared test showing if there are statistically more variants passing the filters applied in a gene in the Highlander database than in ExAC/GoNL

b. Gene representation

Variants are represented by vertical bars on both side the gene. The bars above represent variants in ExAC/GoNL, the bars underneath, variants in Highlander. The color of the bar varies with the frequency of the variant in the Highlander table or in ExAC/GoNL.

If several variants are found at the same position, they are place on top of each other.

When the mouse is stopped on a variant, an information window appears showing alleles, frequency and the number of times that this variant appears in the table. When the mouse is stopped on an exon, an information window appears taking the number of the exon, the genomic position of the beginning and the end of the exon.

c. Table of analyzed variants analyzed

The green highlighted variants are variants in ExAC/GoNL, while the blue are in Highlander. The columns displayed are the one selected in Highlander when the Burden test is launch.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.9.4 Statistics Associator

Function allowing to see if selected variants are present in another analysis

4.9.5 Pedigree checker

Function allowing to check familial relationships between patients

For the 'Common dbSNP' tabs, when comparing 2 patients P1 and P2, we first take all variants from P1 and P2 having a dbSNP id and a read depth > 10.
The total (number between parenthesis in the table) is the sum of snps unique to P1, unique to P2 and common between P1 and P2 (so the snps common are only counted ONCE).
The percentage reflects then the number of snps common to P1 and P2.

For the 'Adjusted dbSNP' tabs, when comparing 2 patients P1 and P2, we first take all variants from P1 and P2 having a dbSNP id and a read depth > 10.
The total (number between parenthesis in the table) is the sum of snps found in P1 and the snps found in P2 (so the snps common are counted TWICE).
The percentage reflects then TWICE the number of snps common to P1 and P2.
'Common dbSNP' gives you real numbers: percentage of snps common between 2 patients, from a real total number of snps.
The problem is that this percentage is generally low, due to a high number of sequencing errors.
'Adjusted dbSNP' gives you an abstract % and total (intersection counted twice), but it mitigates errors and gives better estimation of similarity between 2 samples.

Interpretation (for a complete exome):

Correspondence between the color of the case and the degree of parenthood

Color	**% common alleles**	**degree of parenthood**
red	< 60-62	none
orange	< 67-69%	Uncle / aunt
yellow	< 69-71%	Grandparents
green	> 71%	Relative/brother-sister

4.9.6 Get average coverage

Function to obtain, for a list of genes and for selected samples, the average coverage and the % of the target area covered more than 10x, 20x or 30x

- ➔ Display in a table either by gene for all the selected samples or by patient
- ➔ Color boxes varies depending on the % the target area well covered

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.9.7 Download BAM and VCF files

Function to download BAM and VCF files on a local computer

4.9.8 Export current table content to Excel or TSV


Function to export a table of variants in Excel or TSV format

- ➔ First sheet entitled "analysis + report date": table of variants as it was in Highlander when generating the report.
- ➔ First sheet entitled "Filters details":
  - The filter used to generate the table of variants
  - The software version of the Highlander
  - The user who generated the report
  - The date of the report

4.9.9 View ALL runs detailed statistics

Function for viewing all the statistical values of one or more run(s).

Table of possible statistics

Statistics	average_depth_of_target_coverage
Total_bead_deposition	coverage_wo_dup
whole_run_reads_produced	percent_of_target_covered_meq_1X_wo_dup
percent_bad_beads	percent_of_target_covered_meq_5X_wo_dup
whole_run_assigned_reads	percent_of_target_covered_meq_10X_wo_dup
percent_unassigned_reads	percent_of_target_covered_meq_20X_wo_dup
reads_produced	percent_of_target_covered_meq_30X_wo_dup
reads_mapped	percent_duplicates_picard
percent_total_mapped	percent_not_totally_on_target
low_mapqv_reads	Het_hom_ratio_ls
percent_low_mapqv	ti_tv_ratio_ls
reads_on	ti_tv_ratio_all_gatk
percent_on	ti_tv_ratio_known_gatk
reads_off	ti_tv_ratio_novel_gatk
percent_off	Sequence_duplication_prop
enrichment fold	coverage_ratio_chrom_xy
num_targets_not_covered	variants_all_called_analysis
target_bases_not_covered	variants_known_called_analysis
percent_of_target_covered_meq_1X	variants_novel_called_analysis
percent_of_target_covered_meq_5X	variants_all_called_analysis_pass
percent_of_target_covered_meq_10X	variants_known_called_analysis_pass
percent_of_target_covered_meq_20X	variants_novel_called_analysis_pass

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

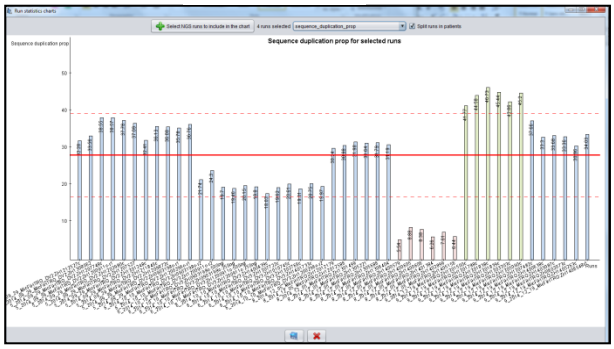
4.9.10 View ALL statistics charts

Function for comparing statistics as a histogram. It is possible for each graph to visualize statistics of individual samples.

Example of histogram comparison of several runs:

For all samples

Per sample



Red line: average
Dotted lines: 1 standard deviation
The bars below the mean - 1SD are red and the bars above the mean + 1SD are green, others are blue.

.

4.10 Help

Memory: allow to modify the PC memory allocated to Highlander.

About: provides access to the version of Highlander as well as the different libraries used by the software.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

4.11 Right panel: details of variants

When a variant is selected, a set of boxes appears in the right panel detailing information on the variant selected.

4.11.1 Customization

- The order of boxes can be changed by clicking and dragging the boxes up or down.
- To scroll the contents of a box, you must click on the plus sign on the box.
- Customization is automatically saved in the profile and remains at a future connection.

4.11.2 Type of boxes

1. Boxes corresponding to a column category:
    - allele frequency in population
    - calling in other analysis
    - change details
    - change statistics
    - change statistics by group
    - confidence
    - conservation scores
    - coverage
    - effect prediction
    - error detection
    - gene statistics
    - gene statistics by group
    - genotype
    - other scores
    - related identifiers
    - sample details
    - variant evaluation (see below)

2. Box with gene annotations in which the variant is: name of genes, IDs in databases, OMIM disease,…

3. Boxes collecting information on other transcripts : Polyphen and SnpEff predictions

4. Boxes collecting user annotations on the variant:
    - **Variant evaluation**: evaluation and comments visible by all users but editable only by authorized persons (defined when creating the run of the project using the Project Manager). This box contains different criteria for evaluating the variant. By default, when the test is not evaluated, there is a "?" in the field corresponding to the box or "zero" in the table of variants. When the test is evaluated, the table of variants is automatically up-to- date.
        - *evaluation*: allow to classify the variant according to 5 classes. At each class is assigned a color that allow to easily visualize in the table the different kind of evaluated variants (and in some “statistics” boxes)
            - I - Benign - Polymorphism.
            - II - Variant Likely Benign.
            - III - Variant of Unknown Significance.
            - IV - Variant Likely Pathogenic.
            - V - Pathogenic Mutation.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels

- *check_insilico*: this function is used to signal whether the variant seems true or not *in silico* (after viewing reads with IGV).
- *check_validated_change*: allowed to signal whether the variant was confirmed by another method.
- *check_somatic_change*: used to signal whether the variant is somatic (TRUE) or germinal (FALSE)
- *check_segregation*: used to signal whether the variant cosegregates in the family
  - SINGLE: no other sample in family
  - COSEG: variant cosegregates
  - CARRIERS: some unaffected carrier(s)
  - NO_COSEG: not in other affected(s)
  - NO_COSEG_OTHER: does not cosegregate in other families
  - NOT_CHECKED: not checked for segregation
- *evaluation_comments* : free field
- *history*: automatic field to trace any changes made in the box "variant evaluation" by recording the user login, the date of change and the type of change.

If several variants are selected, the value of the assessment applies to all these variants. This also works for private annotations but not for public annotations.

- **Public annotations**: annotations visible and editable by all users. Changes made in this field will appear in the table of variants when the request (or filter) is restarted.

- **Private annotations**: annotations visible only by the user (but these annotations can be shared with other users).
  - Allows to report whether the variant is interesting or not and record private comments on the variant or gene
  - If several variants are selected, the annotation applies to all variants.

- **Samples with same evaluated change**: shows the evaluation of the same variant in other patients as well as comments associated with it.

5. Boxes showing the statistics for the same change/gene: in blood/tissues, passing or not the GATK filters, with or without a damaging effect.

6. Other boxes :

- **External links:** contains a series of buttons that open the web browser directly with the variant/gene enrolled in different websites. The buttons appear only if Highlander found an entry in the website. Links to dbNSP, COSMIC, ExAC, Beacon of Beacons, UCSC position, Ensembl gene, Gene in OMIM, Gene in PubMed, Gene in NCBI, Gene in Entrez, Gene in LOVD, Gene in DIDA, Gene in Decipher, Gene in HGNC, Mutalyzer.

- **Alignment:** allows viewing 40bp on both sides of the variant without opening IGV.

CC-BY-SA – 2016 – Raphael Helaers, Dorien Daneels