

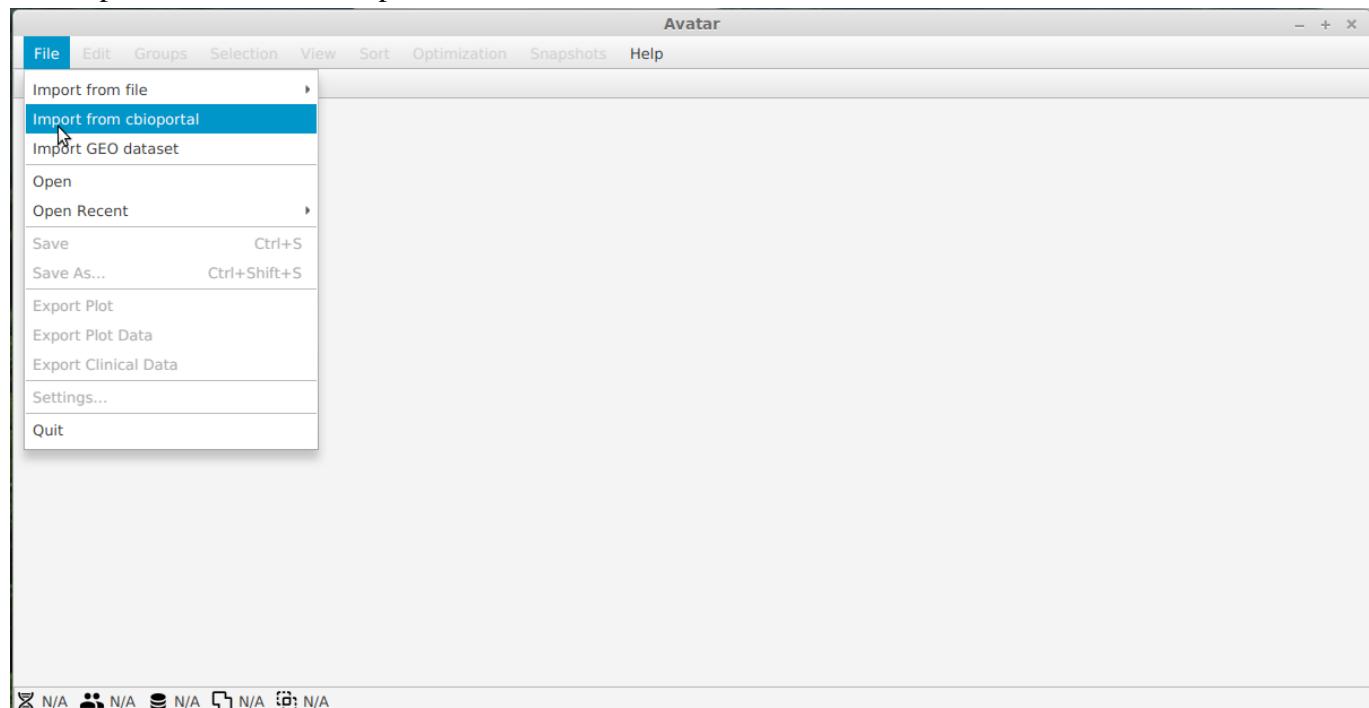
Manual & Walkthrough

This document contains screenshot and notes that show how Avatar is applied to analyze the Head and Neck Squamous Cell Carcinoma (HNSCC) data. The description is focused on the HNSCC usecase. Thus, not all features of Avatar are described here. The mouse pointer in the screenshots usually indicates the menu item or button to click next. The software is available for download on Github: <https://github.com/sysbio-bioinf/avatar>. Use the evaluation license provided.

1 Import Data

1.1 HNSCC TCGA Nature 2015

First, open the cBioPortal import from the menu.



Type *head* into the search field to search for the HNSCC studies. Select the *TCGA Nature 2015* study and click import.

The progress of the download is shown including the estimated duration to finish the download.

Select *expression data* as alteration data type to import.

The screenshot shows a search results table for studies containing the word 'head'. A modal dialog titled 'Select alteration data' is open over the table. The dialog lists four data types: mutation data, fusion data, expression data, and methylation data. The 'expression data' checkbox is checked. Below the checkboxes are 'Select' and 'Cancel' buttons. The background table shows study details like 'Head and Neck Squamous Cell Carcinoma (Broad, Science 2011)' and 'Head and Neck Squamous Cell Carcinoma (Johns Hopkins, Scienc...)'.

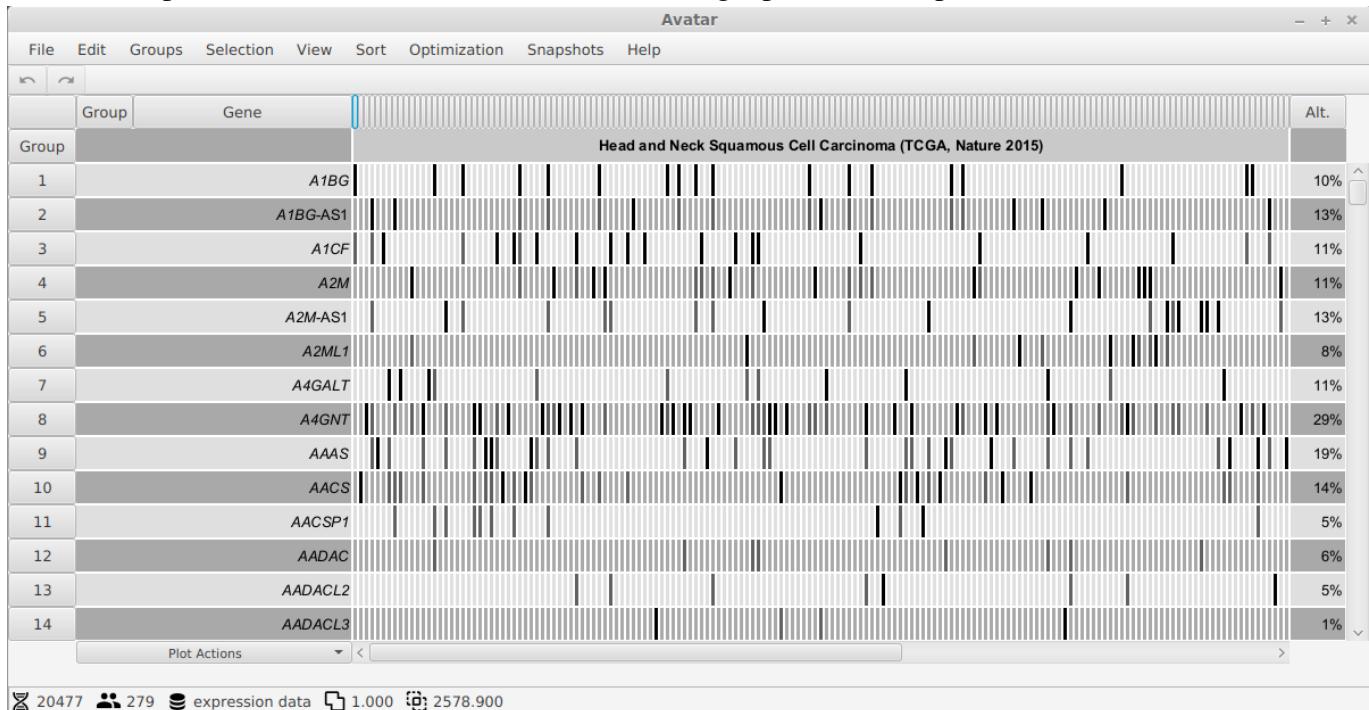
Study	Description
Head and Neck Squamous Cell Carcinoma (Broad, Science 2011)	Comprehensive profiling of 74 head and neck squamous cell carcinoma samples from the Broad Institute.
Head and Neck Squamous Cell Carcinoma (Johns Hopkins, Scienc...	Comprehensive profiling of 32 head and neck squamous cell carcinoma samples from Johns Hopkins.
Head and Neck Squamous Cell Carcinoma (TCGA, Nature 2011)	Comprehensive profiling of 279 head and neck squamous cell carcinoma samples from TCGA.
Head and Neck Squamous Cell Carcinoma (TCGA, PanCancer)	Comprehensive profiling of 279 head and neck squamous cell carcinoma samples from TCGA.
Head and Neck Squamous Cell Carcinoma (TCGA, Provisional)	Comprehensive profiling of 279 head and neck squamous cell carcinoma samples from TCGA.
MSS Mixed Solid Tumors (Van Allen, 2018)	Comprehensive profiling of 279 head and neck squamous cell carcinoma samples from TCGA.
Recurrent and Metastatic Head & Neck Cancer (MSKCC, JAMA)	Comprehensive profiling of 279 head and neck squamous cell carcinoma samples from TCGA.

Define an alteration as *overexpression* for a Z-score greater or equal to 1σ .

The screenshot shows a histogram titled 'Histogram: Number of Differential Expressions per Gene' with the x-axis labeled 'Number of affected Samples per Gene' and the y-axis labeled 'Number of Genes'. The distribution is right-skewed, peaking around 25-30 samples per gene. Below the histogram, there are two sections: 'Scenario' and 'Thresholds'. In the 'Scenario' section, 'overexpressed' is selected. In the 'Thresholds' section, the 'upper' threshold is set to 1.0σ . An 'Apply' button is at the bottom right of the dialog.

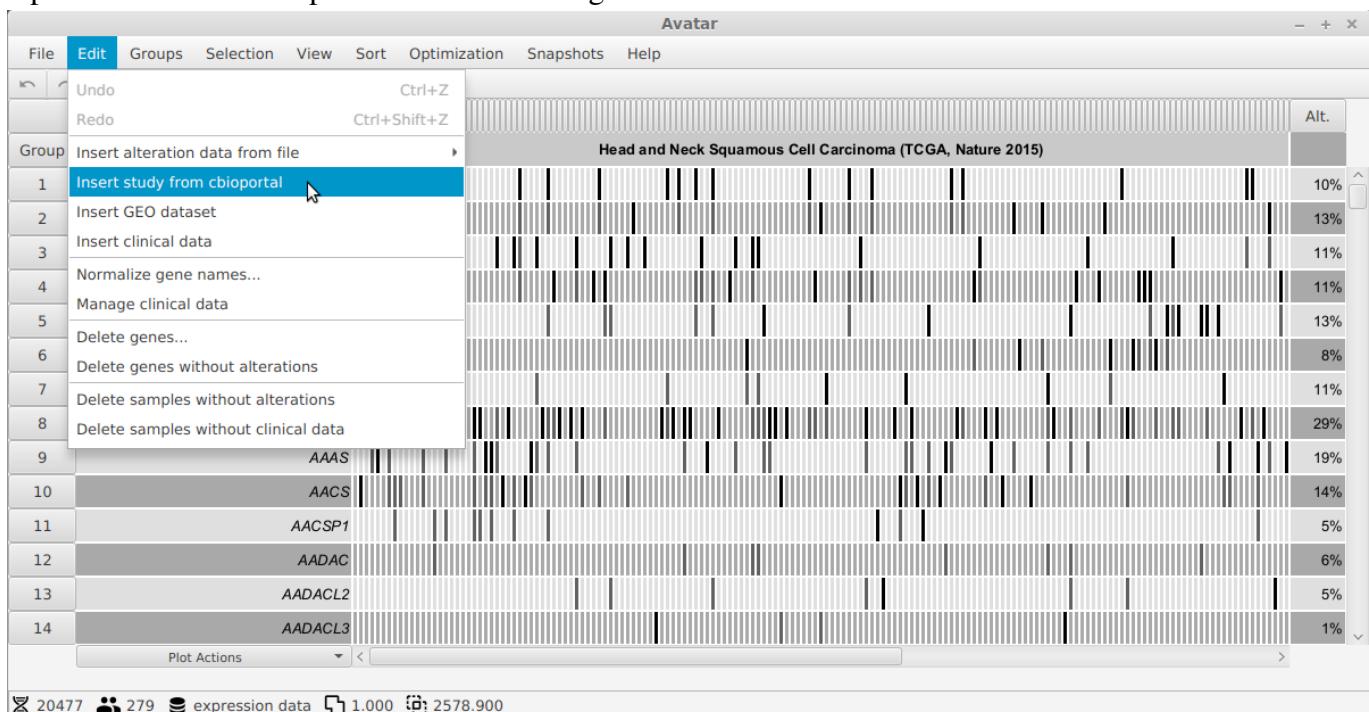
Scenario	Thresholds
<input type="radio"/> differentially expressed	upper: <input type="text" value="1.0"/> σ
<input checked="" type="radio"/> overexpressed	lower: <input type="text"/> σ
<input type="radio"/> underexpressed	

After the import is finished, the main window containing a plot of the imported data is shown.



1.2 HNSCC TCGA Provisional

Open the cBioPortal import menu for inserting the second dataset.



Type *head* into the search field to search for the Head and Neck Squamous Cell Carcinoma (HNSCC) studies. Select the *TCGA Provisional* study and click import.

After the download completed, select *expression data* as alteration data type to import.

Import data from cbiportal

Search: head

Study	Description
Head and Neck Squamous Cell Carcinoma (Broad, Science 2011)	Comprehensive profiling of 74 head and neck squamous cell carcinoma samples from the Broad Institute.
Head and Neck Squamous Cell Carcinoma (Johns Hopkins, Scienc...	Comprehensive profiling of 32 head and neck squamous cell carcinoma samples from Johns Hopkins.
Head and Neck Squamous Cell Carcinoma (TCGA, Nature 2015)	TCGA Head and Neck Squamous Cell Carcinoma, containing 279 samples; raw data at the https://tcga-data.nci.nih.gov/tcga/study/STU-0001/
Head and Neck Squamous Cell Carcinoma (TCGA, PanCancer Atlas)	PanCancer.gov/about-data/publications/pancanatlas data from 11 cancer types across multiple cancer types including melanoma, non-squamous, treatment resistant head and neck tumors
Head and Neck Squamous Cell Carcinoma (TCGA, Provisional)	raw data at the https://tcga-data.nci.nih.gov/
MSS Mixed Solid Tumors (Van Allen, 2018)	Mixed Solid Tumors across multiple cancer types including melanoma, non-squamous, treatment resistant head and neck tumors
Recurrent and Metastatic Head & Neck Cancer (MSKCC, JAMA Oncology)	Recurrence and metastasis of head and neck cancer

Select alteration data

Select alteration data types to import for study
Head and Neck Squamous Cell Carcinoma (TCGA, Provisional)

mutation data
(datamutations_extended.txt)

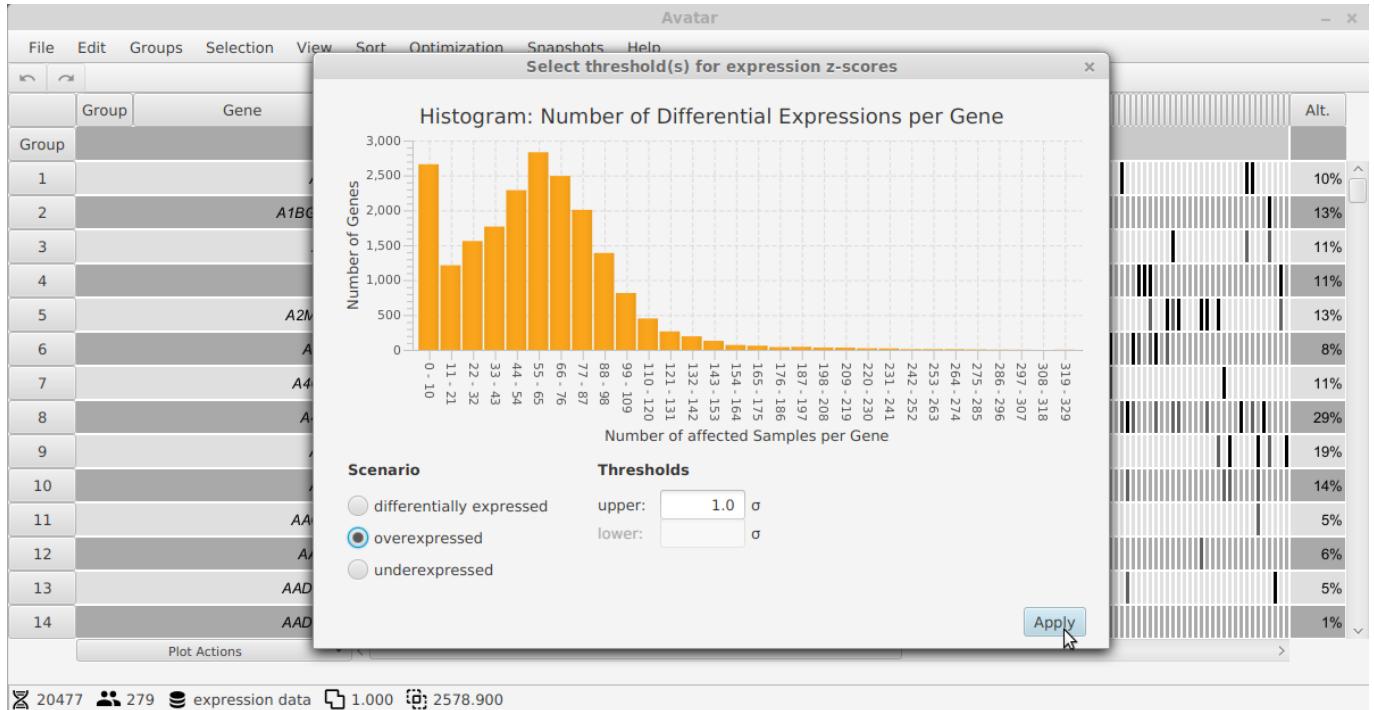
expression data
(dataRNA_Seq_v2_mRNA_median_Zscores.txt)

methylation data
(datamethylation_hm450.txt)

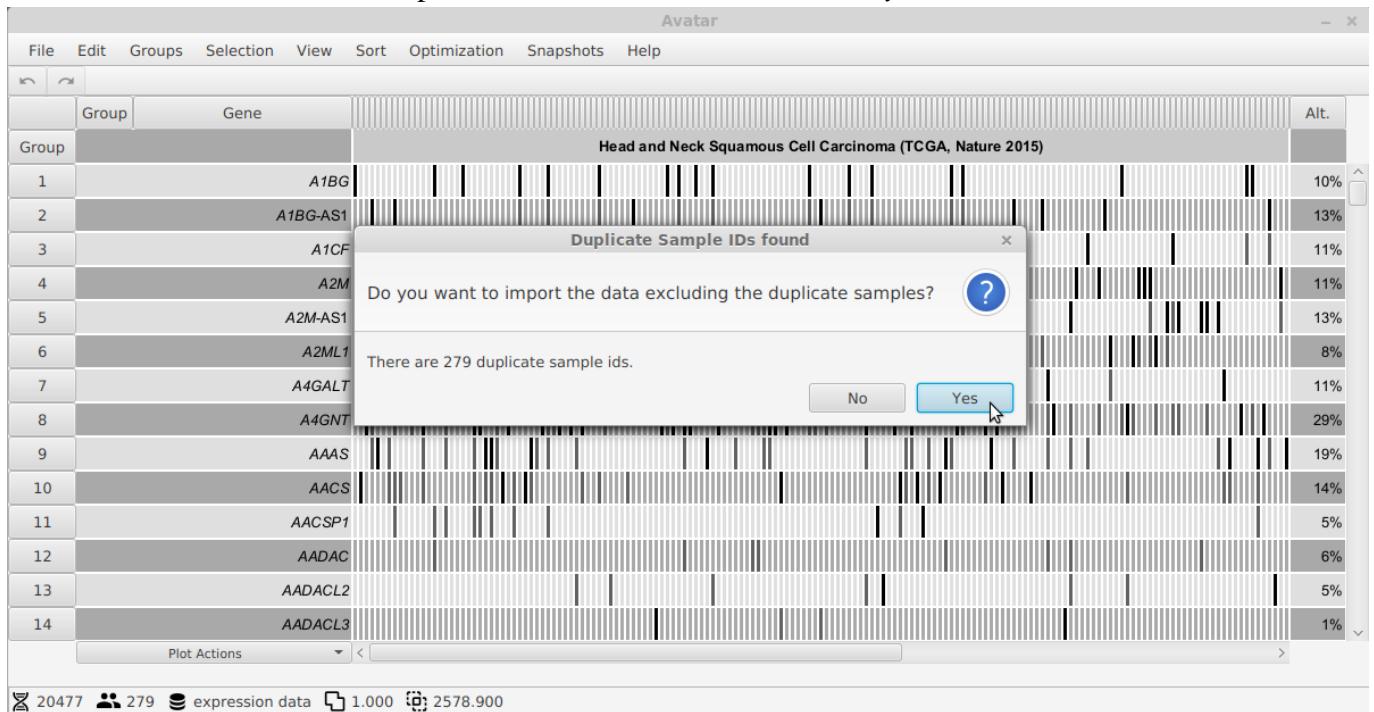
Found studies: 7

Import Cancel

Define an alteration as *overexpression* for a Z-score greater or equal to 1σ .



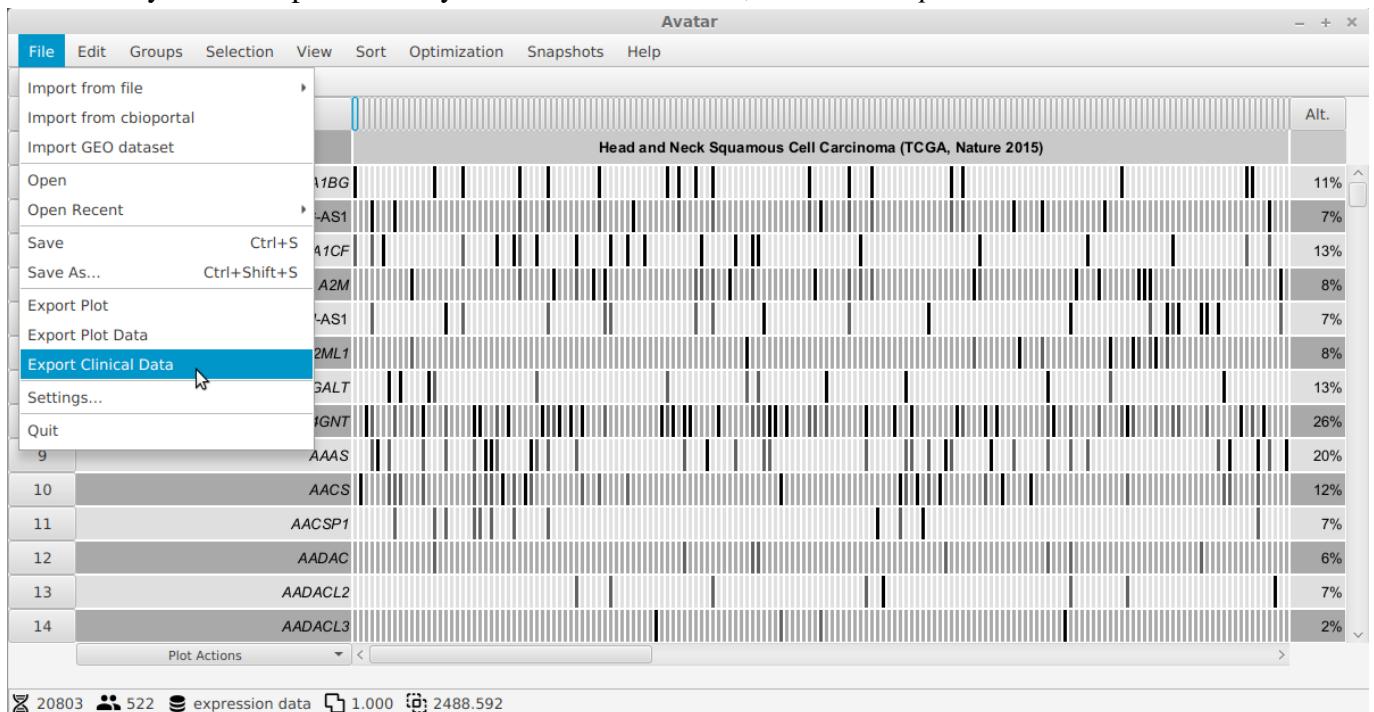
The *TCGA Provisional* dataset also contains all the samples from the *TCGA Nature 2015*. Avatar detects that and asks whether those samples shall be excluded. Answer with yes.



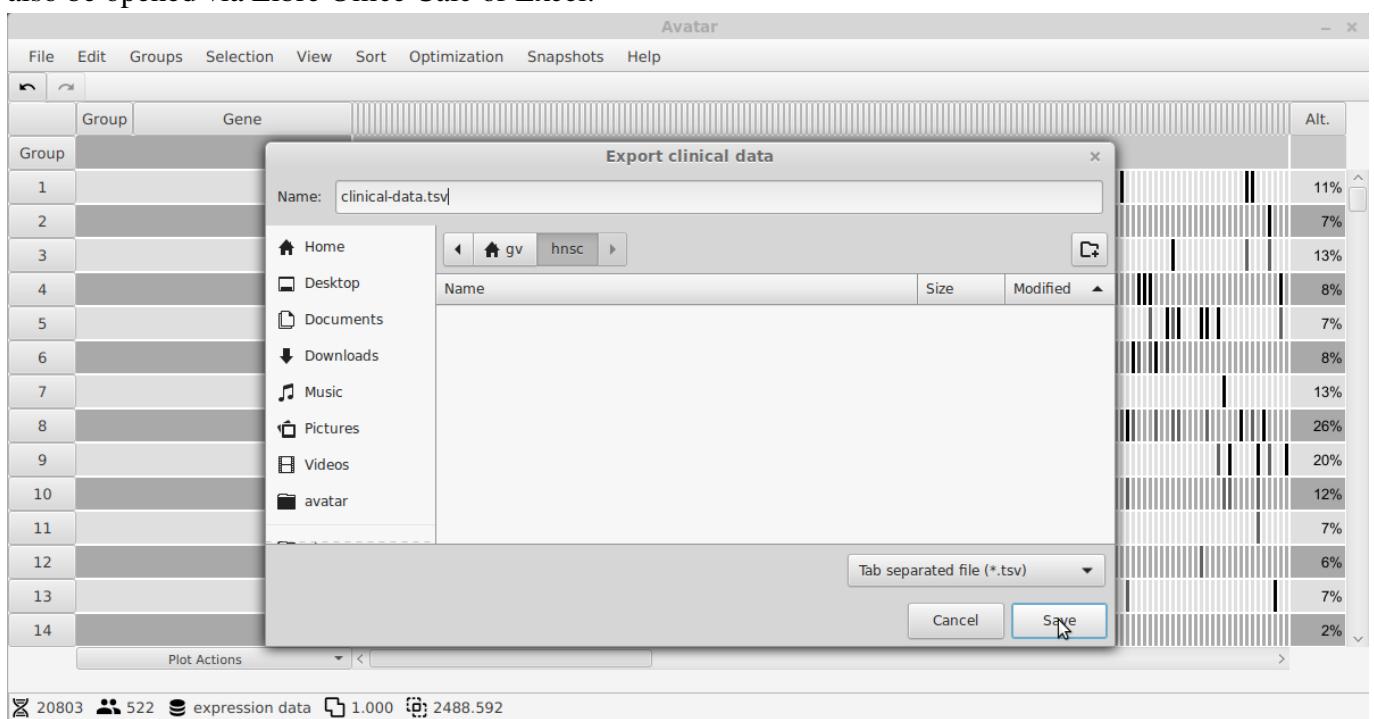
2 Data Preparation

2.1 Preprocess Clinical Data

Preparing the dataset for the optimization can be done in Avatar by deleting samples, grouping samples according to clinical attributes or filtering desired genes. More sophisticated preprocessing based on clinical data (e.g. harmonization of primary sites) can be accomplished by exporting the clinical data. The data can be manually edited or processed by other tools. Therefore, select the *Export Clinical Data* menu item.



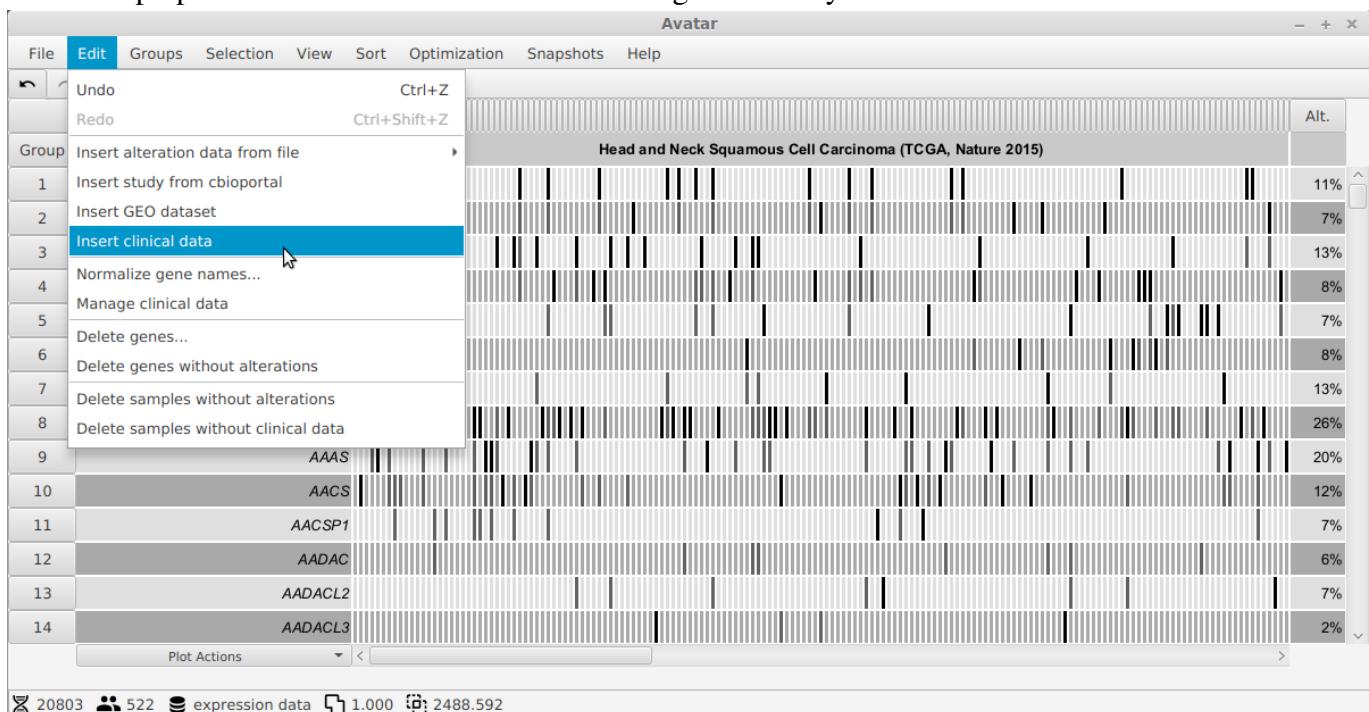
Specify a filename for the exported clinical data. The clinical data will be exported as a text file that can also be opened via Libre Office Calc or Excel.



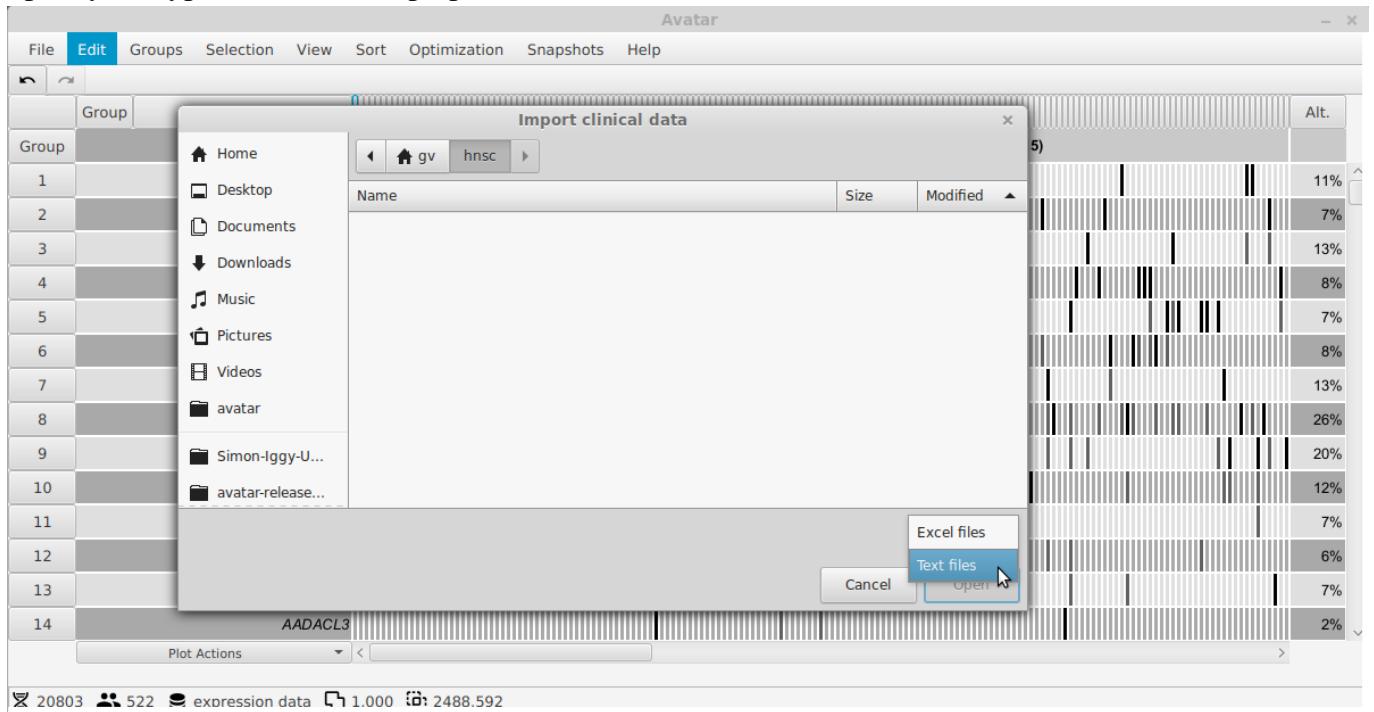
Exported clinical data shown as spreadsheet.

	A	B	C	D	E	F	
1	SAMPLE_ID	AGE	AJCC_METASTASIS_PATHOLOGIC_PM	AJCC_NODES_PATHOLOGIC_PN	AJCC_PATHOLOGIC_TUMOR_STAGE	AJCC_STAGING_EDITION	AJCC_TU
2	TCGA-BA-4074-01						
3	TCGA-MT-A7BN-01	74 [Not Available]	[Not Available]		Stage IVA	7th	[Not Available]
4	TCGA-QK-AA3J-01	69 M0	N0		Stage I	7th	T1
5	TCGA-UP-A6WW-01	58 [Not Available]	[Not Available]		[Not Available]	7th	[Not Available]
6	TCGA-CV-7446-01	66 [Not Available]		N2b	Stage IVA	7th	T2
7	TCGA-HD-7832-01						
8	TCGA-D6-6827-01	55 [Not Available]	N0		Stage I	7th	T1
9	TCGA-F7-A50G-01	66 M0	N1		Stage III	7th	T3
10	TCGA-HD-A610-01	56 MX	N1		Stage IVA	7th	T4a
11	TCGA-WA-A7H4-01	69 [Not Available]	N0		Stage II	7th	T2
12	TCGA-MT-A67D-01	55 M0	N0		Stage II	7th	T2
13	TCGA-BA-A6DI-01	62 M0	NX		[Not Available]	7th	TX
14	TCGA-KU-A66S-01	69 MX	N1		Stage III	7th	T1
15	TCGA-CV-7440-01						
16	TCGA-BB-7864-01	61 [Not Available]	N2c		Stage IVA	7th	T2
17	TCGA-CR-7369-01						
18	TCGA-BA-A6DB-01	24 M0	N0		Stage I	7th	T1
19	TCGA-CN-6021-01						
20	TCGA-CN-6089-01						
21	TCGA-CV-7247-01						
22	TCGA-CN-6018-01						
23	TCGA-CN-6997-01						
24	TCGA-QK-A6IG-01	69 M0	N1		Stage III	7th	T2

Insert the preprocessed clinical data via the following menu entry.

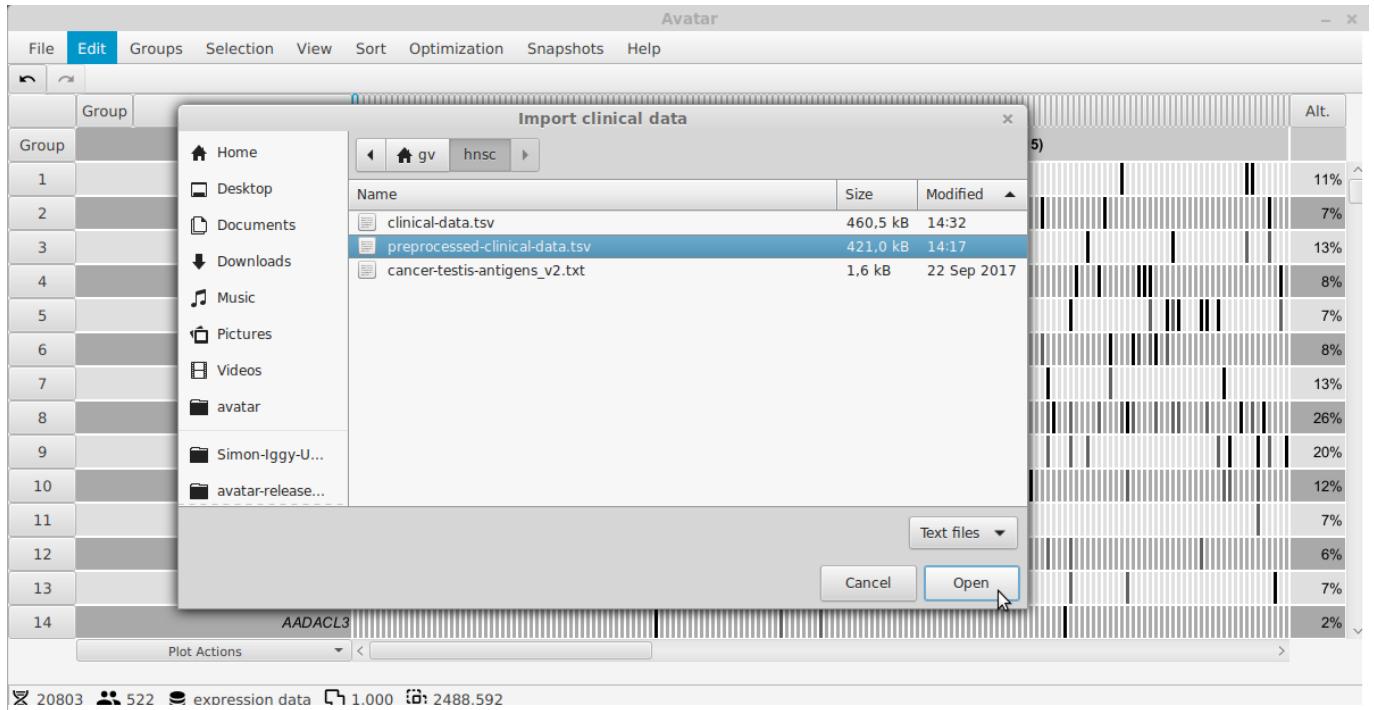


Specify the type of file that the preprocessed clinical data is stored in (Excel or Text).



For the HNSCC usecase a preprocessed clinical data file (text file) is available at: <https://github.com/sysbio-bioinf/avatar/raw/master/dist/preprocessed-clinical-data.tsv>.

Select the downloaded file to insert the clinical data.



Avatar offers a flexible import for clinical data. A clinical data file can contain additional information (e.g. explanations) before the table containing the clinical data starts. Select the index that corresponds to the row containing the column labels.

The screenshot shows the 'Import clinical data from CSV file' dialog. On the left, there's a sidebar with a tree view labeled 'Group' containing items 1 through 14. Below the tree is a status bar showing '20803'. In the center, there's a table with three columns: A, B, and C. The first row contains headers: SAMPLE_ID, HPV_STATUS, and PRIMARY_SITE. Rows 2 through 22 list various TCGA sample IDs along with their HPV status and primary site. To the right of the table is a vertical bar chart titled 'Alt.' showing the percentage distribution of different alleles. At the bottom right of the dialog are 'Import' and 'Cancel' buttons.

Choose the column that contains the sample ids.

This screenshot is similar to the previous one, but the 'Sample ID Column:' dropdown is now expanded, showing the options 'SAMPLE_ID', 'HPV_STATUS', and 'PRIMARY_SITE'. The 'SAMPLE_ID' option is highlighted with a blue background. The rest of the dialog, including the table and the bar chart, remains the same.

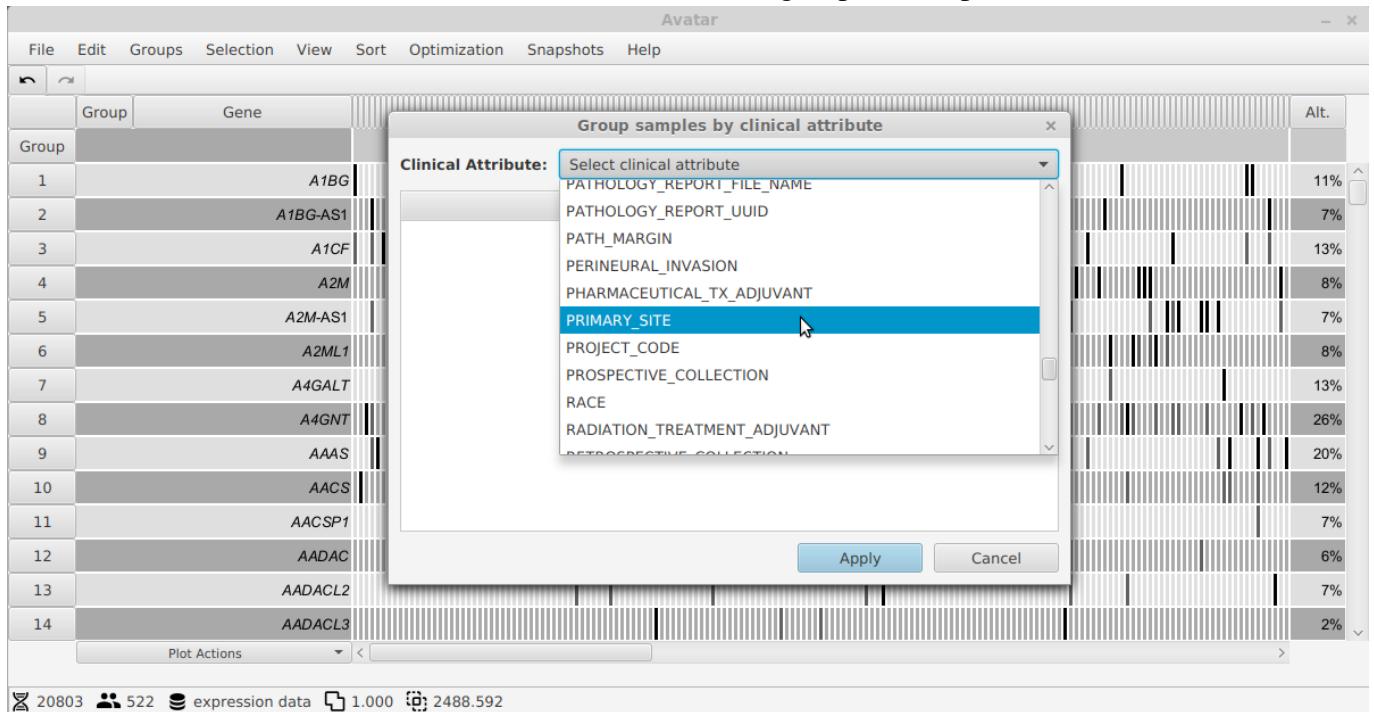
Select the columns that contain the clinical attributes that shall be inserted (here: *HPV_STATUS* and *PRIMARY_SITE*).

	A	B	C
1	SAMPLE ID	HPV_STATUS	PRIMARY_SITE
2	TCGA-BA-4074-01	HPV-	Oral Cavity
3	TCGA-QK-A6V9-01	HPV+	Oropharynx
4	TCGA-UF-A7JT-01	HPV-	Oral Cavity
5	TCGA-CV-A45Z-01	HPV-	Larynx
6	TCGA-IQ-7631-01	HPV-	Oral Cavity
7	TCGA-DQ-7594-01	HPV+	Oropharynx
8	TCGA-HD-A4C1-01	HPV-	Oral Cavity
9	TCGA-MZ-A5BI-01	HPV+	Oropharynx
10	TCGA-QK-A6II-01	HPV-	Oral Cavity
11	TCGA-BB-7862-01	HPV-	Larynx
12	TCGA-P3-A6T6-01	HPV-	Oral Cavity
13	TCGA-CX-7085-01	HPV-	Oral Cavity
14	TCGA-BB-A6UM-01	HPV+	Oropharynx
15	TCGA-CR-7371-01	HPV-	Larynx
16	TCGA-BA-A6DJ-01	HPV-	Oral Cavity
17	TCGA-CN-6022-01	HPV-	Larynx
18	TCGA-CN-6992-01	HPV-	Larynx
19	TCGA-CV-7250-01	HPV-	Larynx
20	TCGA-CN-6019-01	HPV-	Oral Cavity
21	TCGA-CN-6998-01	HPV-	Oral Cavity
22	TCGA-UF-A71D-01	HPV+	Larynx

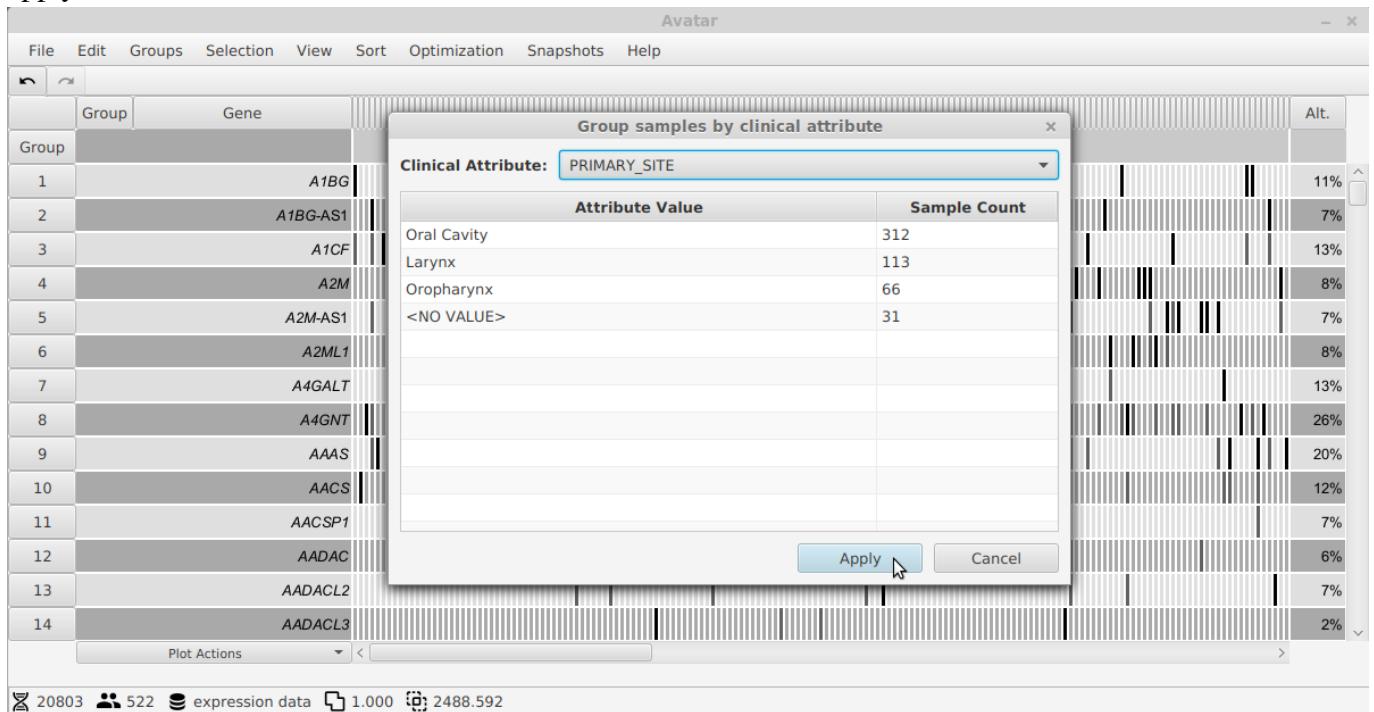
2.2 Filter Samples

Next, the samples will be grouped by primary site to delete the samples without information about the primary site. (Samples with original primary site lip or hypopharynx have no associated primary site in the preprocessed clinical data.) Select the following menu item.

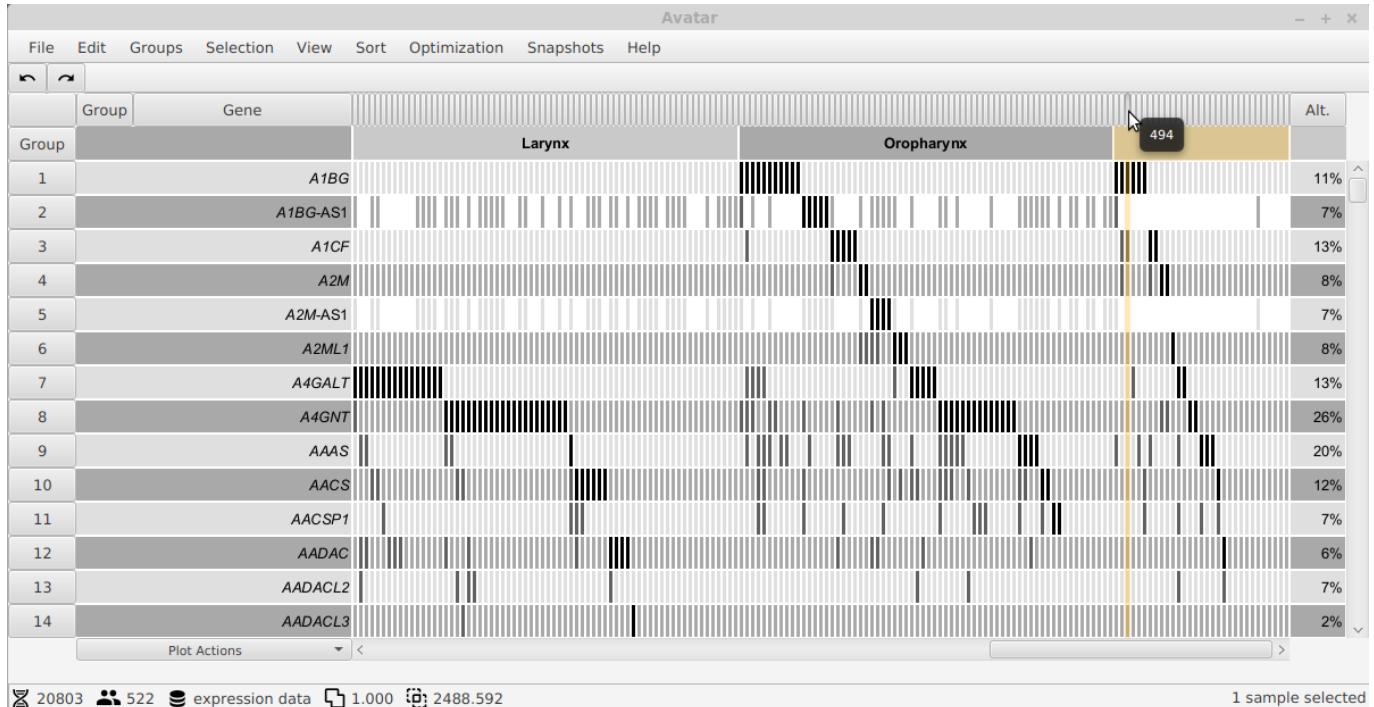
Choose the clinical attribute *PRIMARY_SITE* as criterion to group the samples.



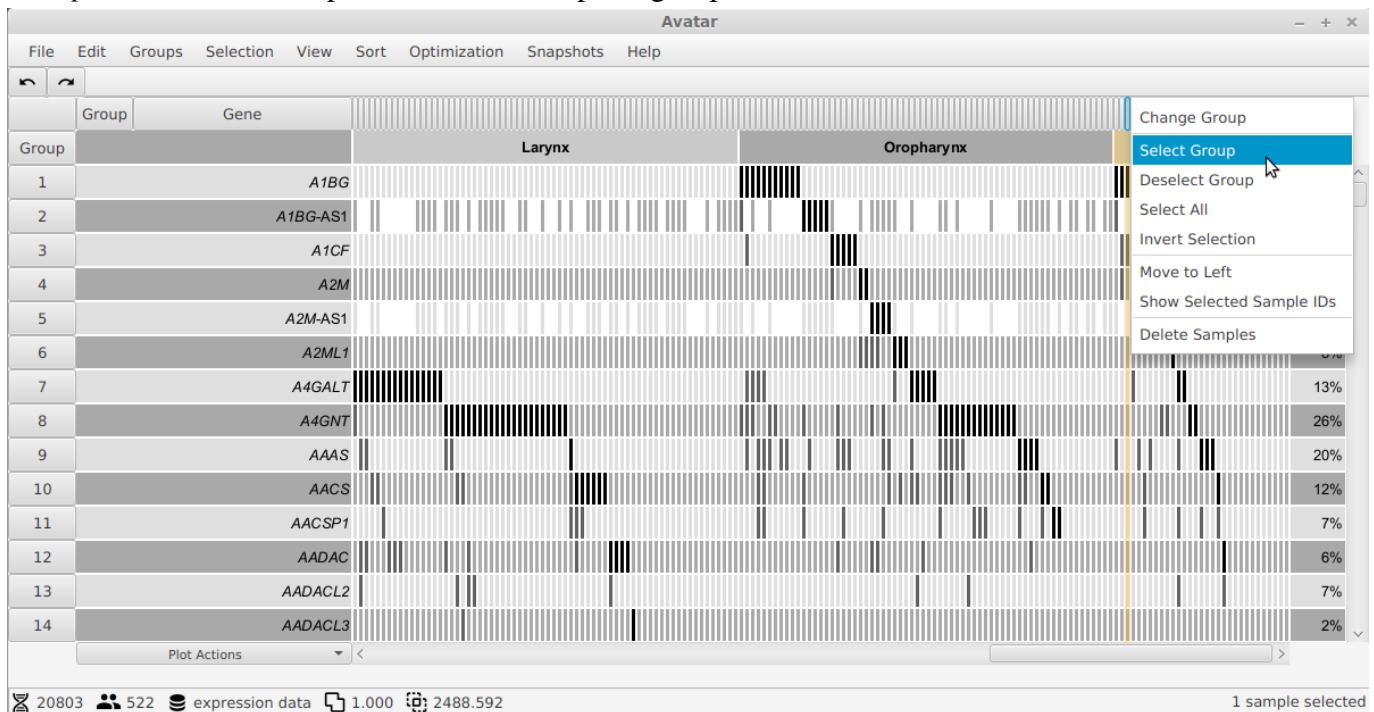
The number of samples per attribute value is shown. The samples with no value will be deleted next. Click apply.



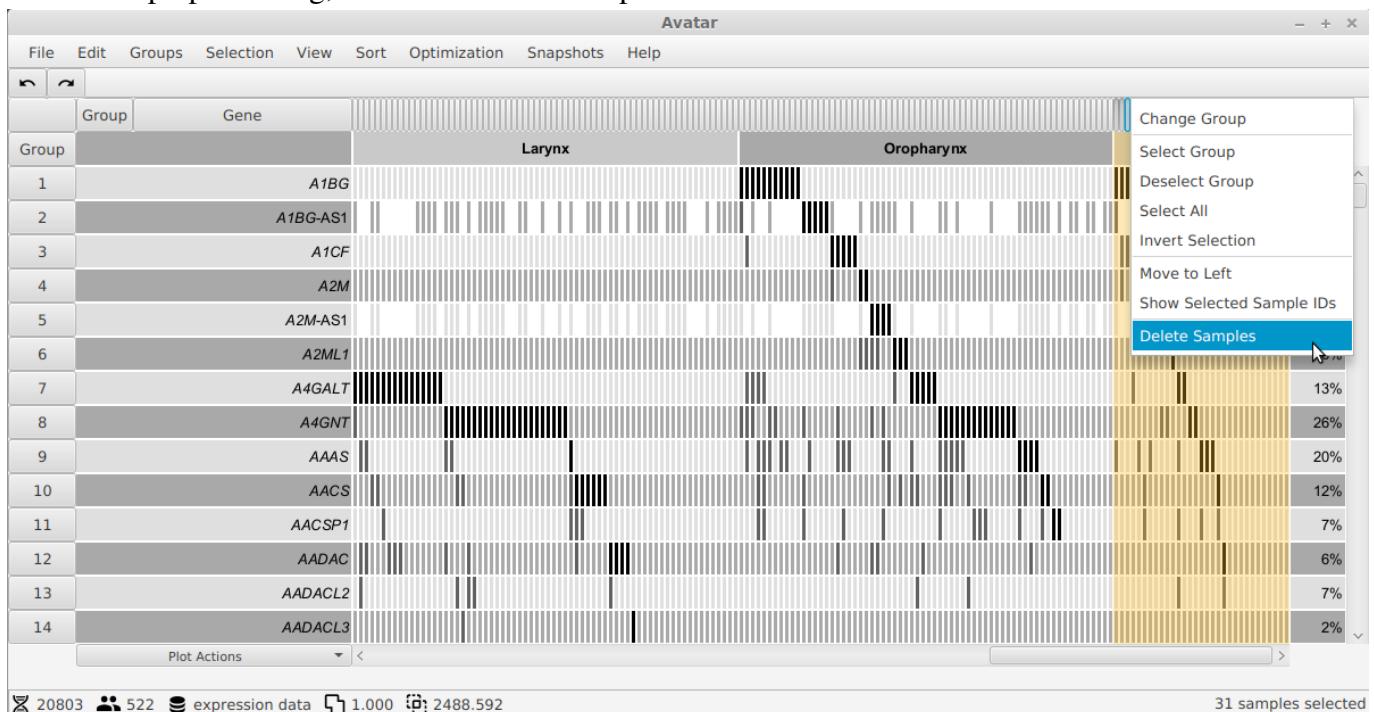
Scroll to the samples that have no group (topmost row in the plot). Click on the button above one of the samples without a group. The column of the sample will be highlighted as well as the cell in the group row.



Click on the selected button using the right mouse button to get to the context menu. A click on *Select Group* will select all samples that have no explicit group.

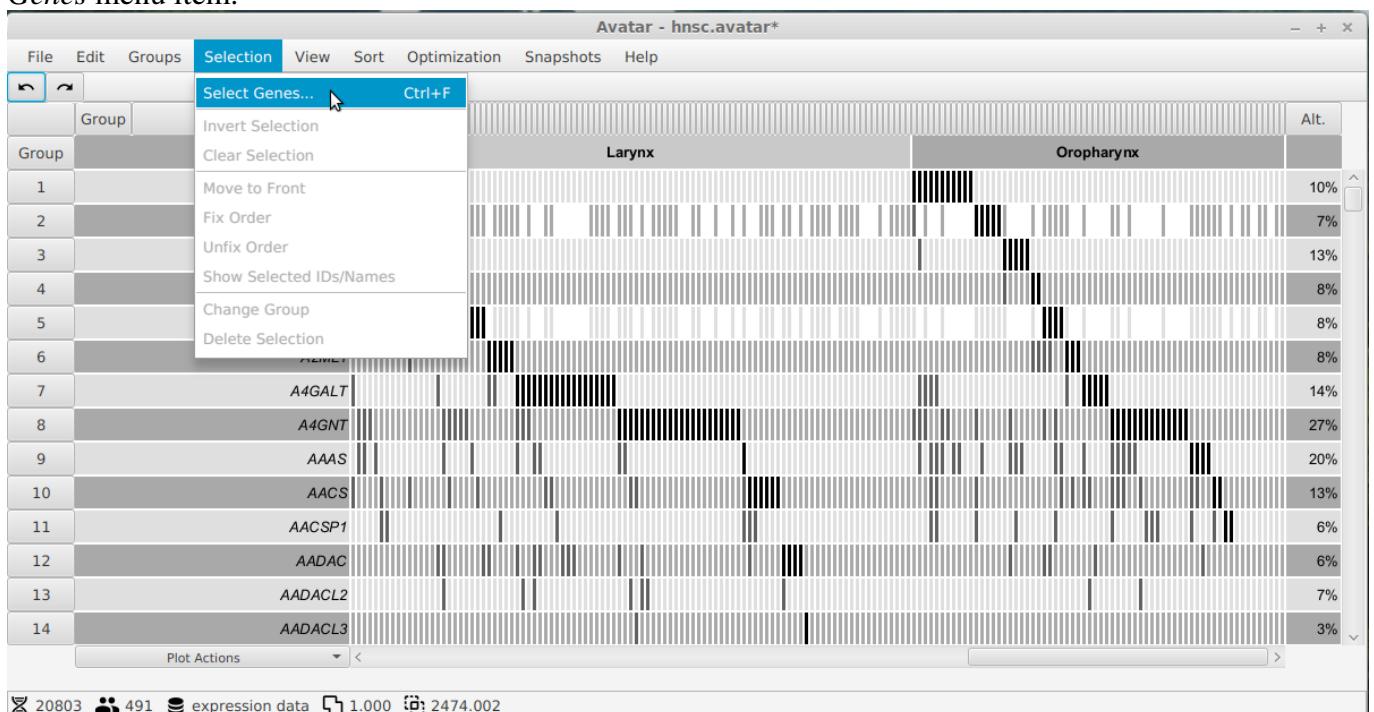


Bring up the context menu again by clicking on any of the selected buttons above the samples without a group. Click on *Delete Samples*. As a result only samples with one of the three desired primary sites remain. Due to the preprocessing, these are also the samples with an associated HPV status.

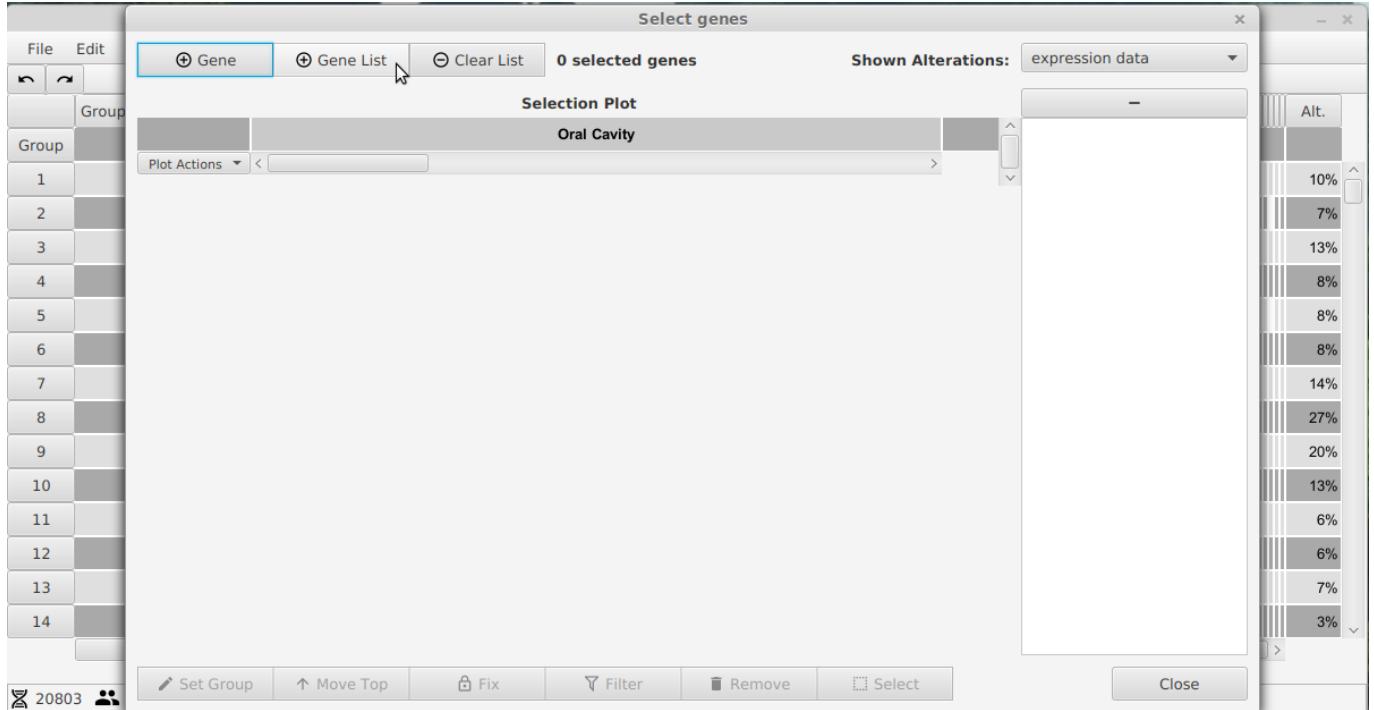


2.3 Filter CTA Genes

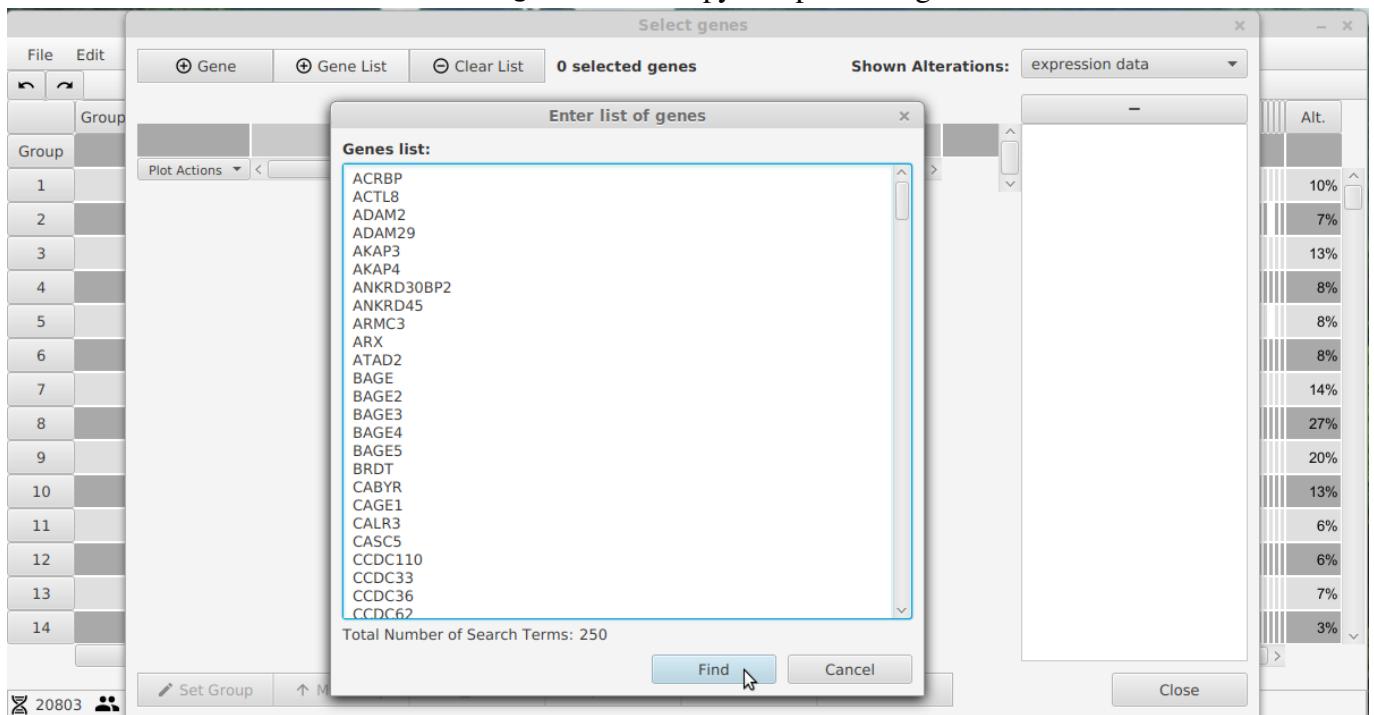
Now, the genes will be limited to the set of Cancer-Testis-Antigens (CTA). Therefore, click on the *Select Genes* menu item.



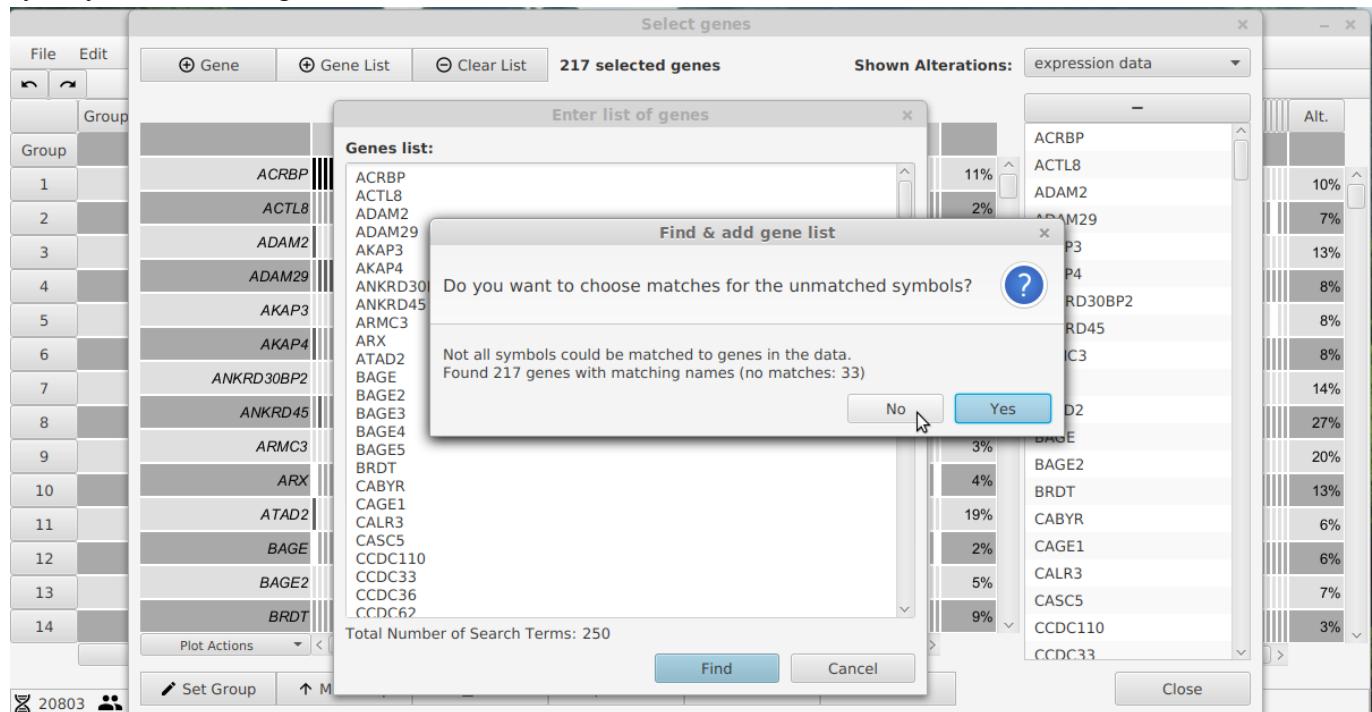
Click the Add Gene List button.



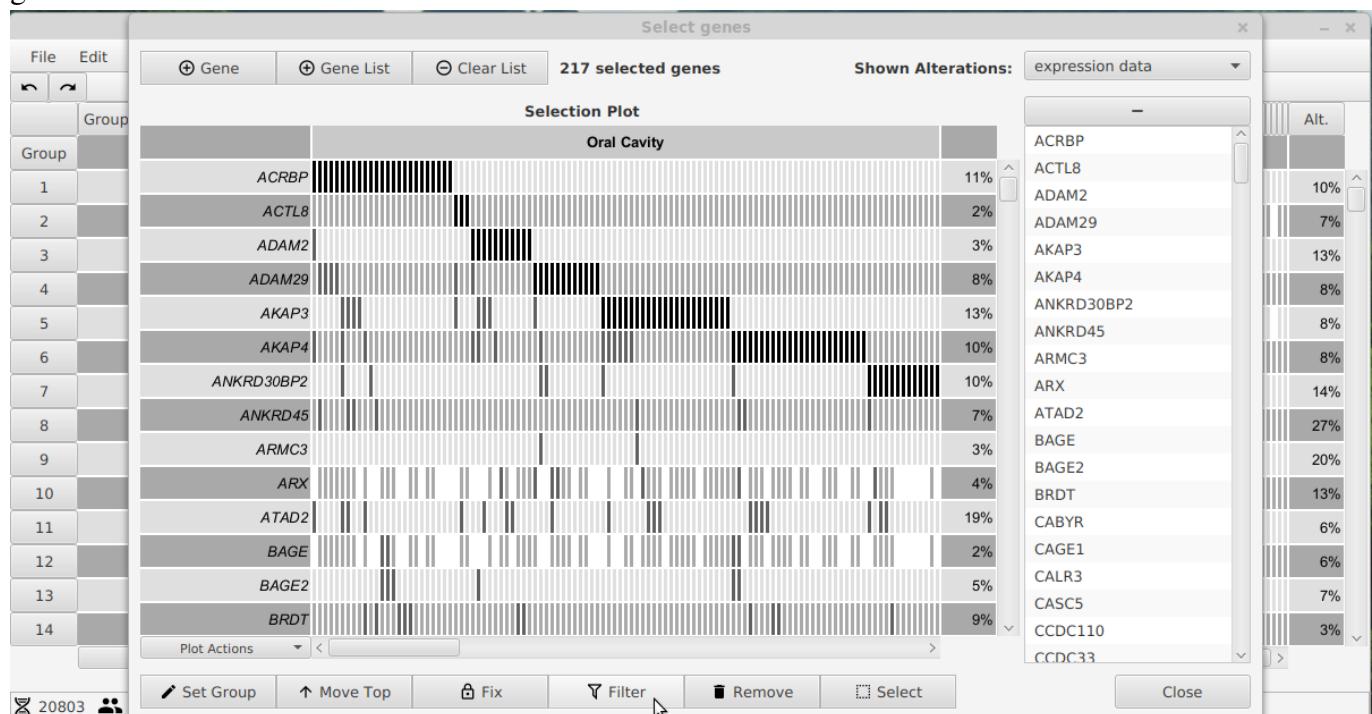
The used list of CTA genes is available at <https://github.com/sysbio-bioinf/avatar/raw/master/dist/cancer-testis-antigens.txt>. Copy and paste the gene names into the text box.



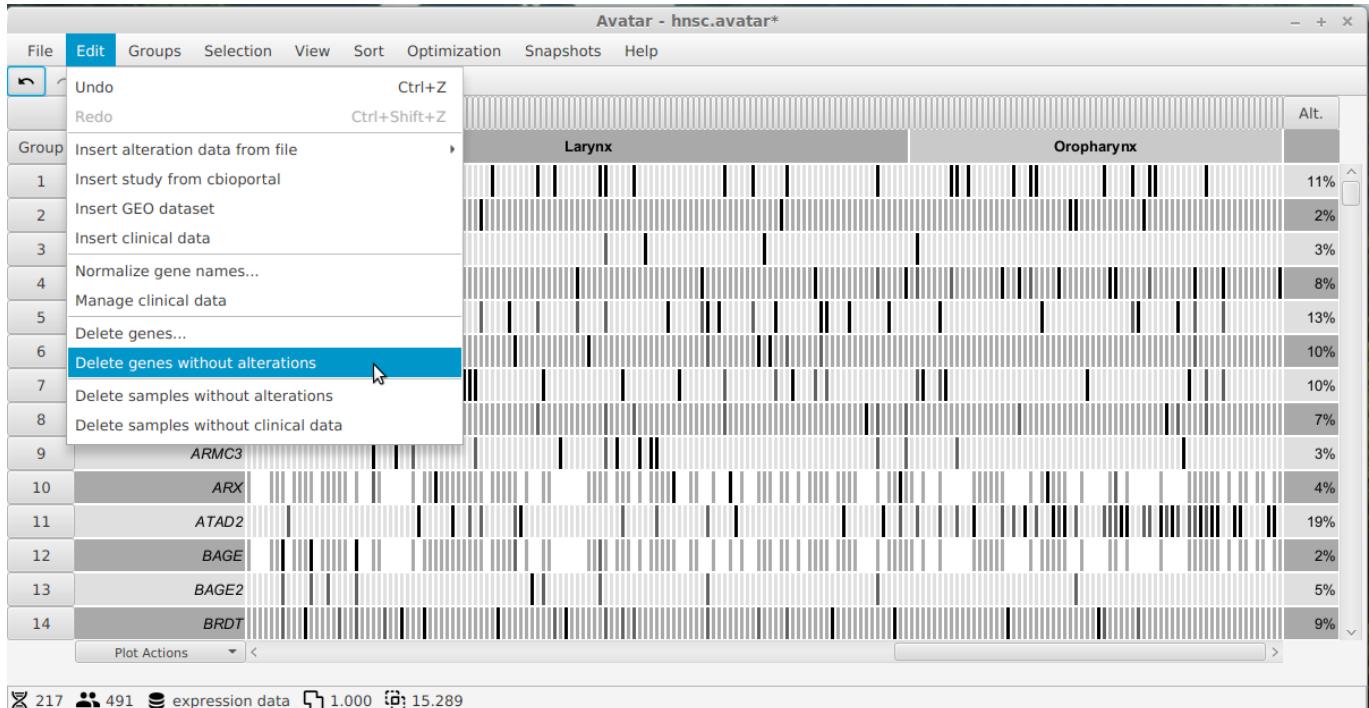
If genes are not found in the current data, Avatar will ask whether the user wants to select matches from synonyms or similar gene names.



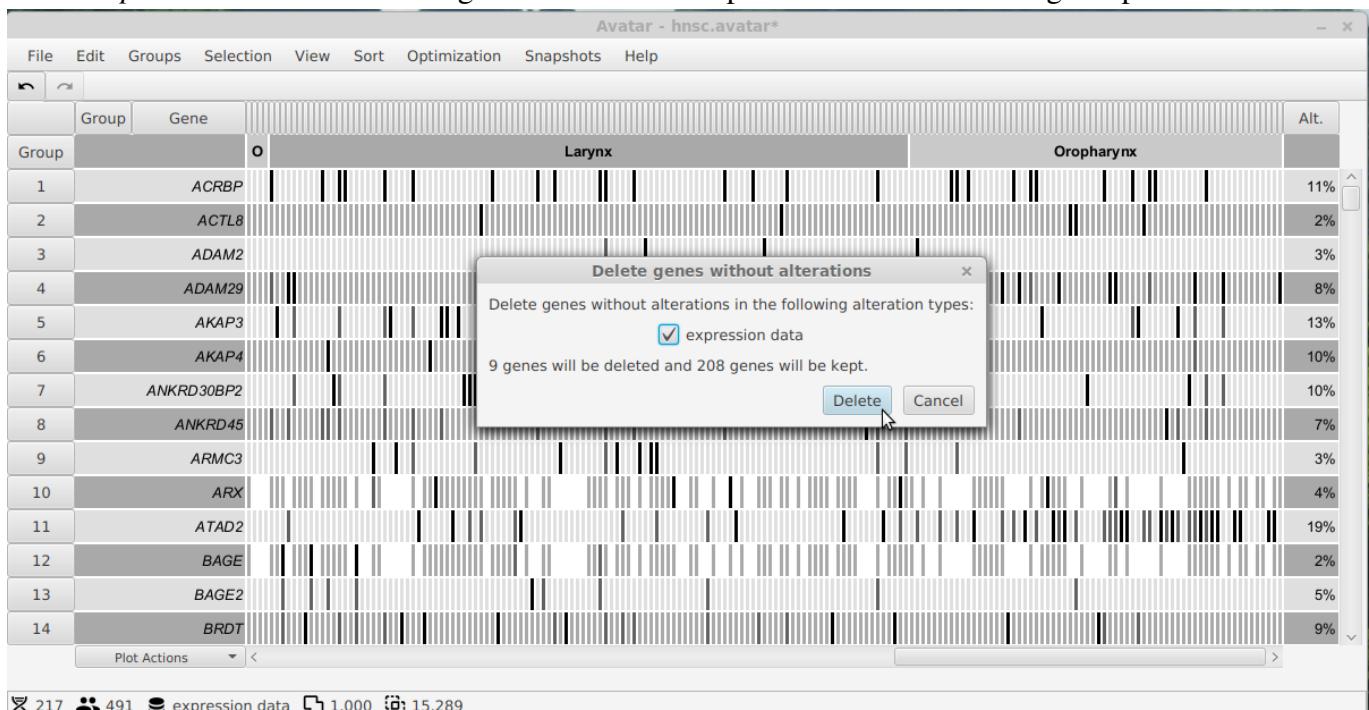
The found genes are display as alteration plot. From the possible actions select *Filter* to delete all non-CTA genes from the current data.



From the remaining CTA genes, delete those without alterations in the remaining samples.

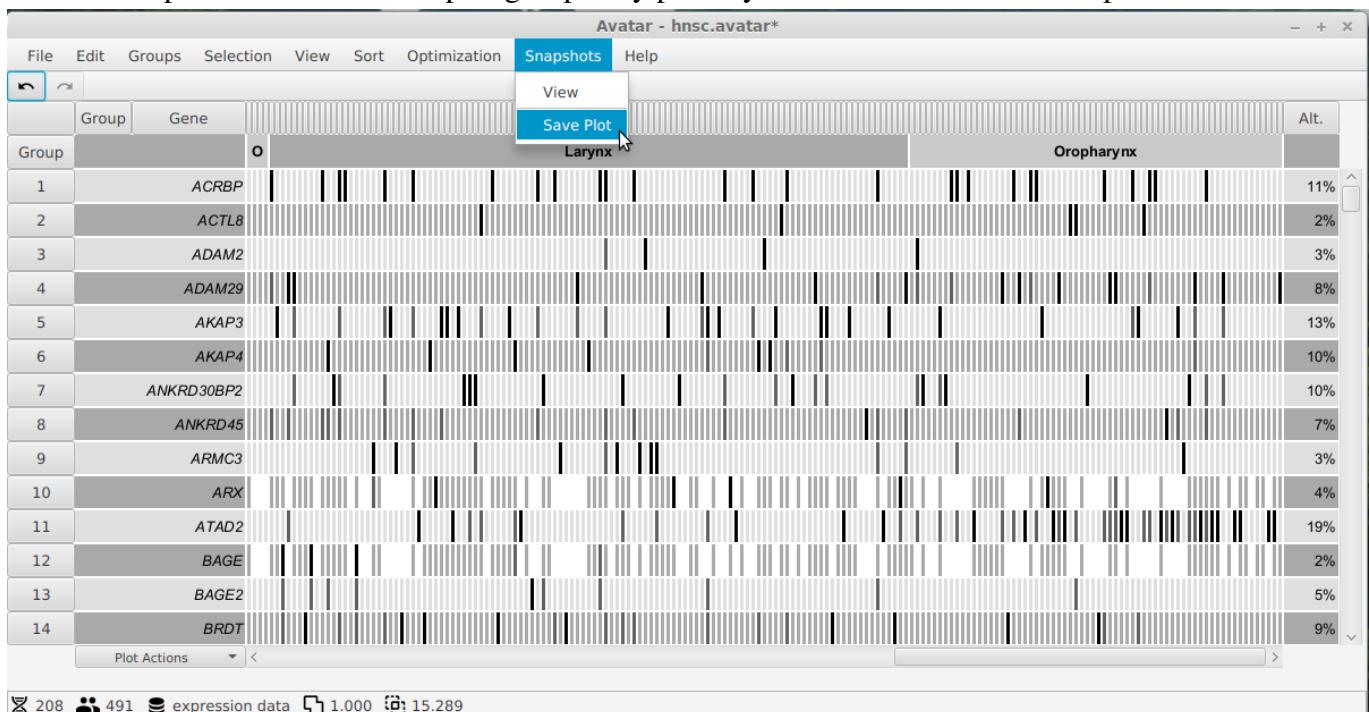


Select *expression data* to delete all genes without overexpressions in the remaining samples.

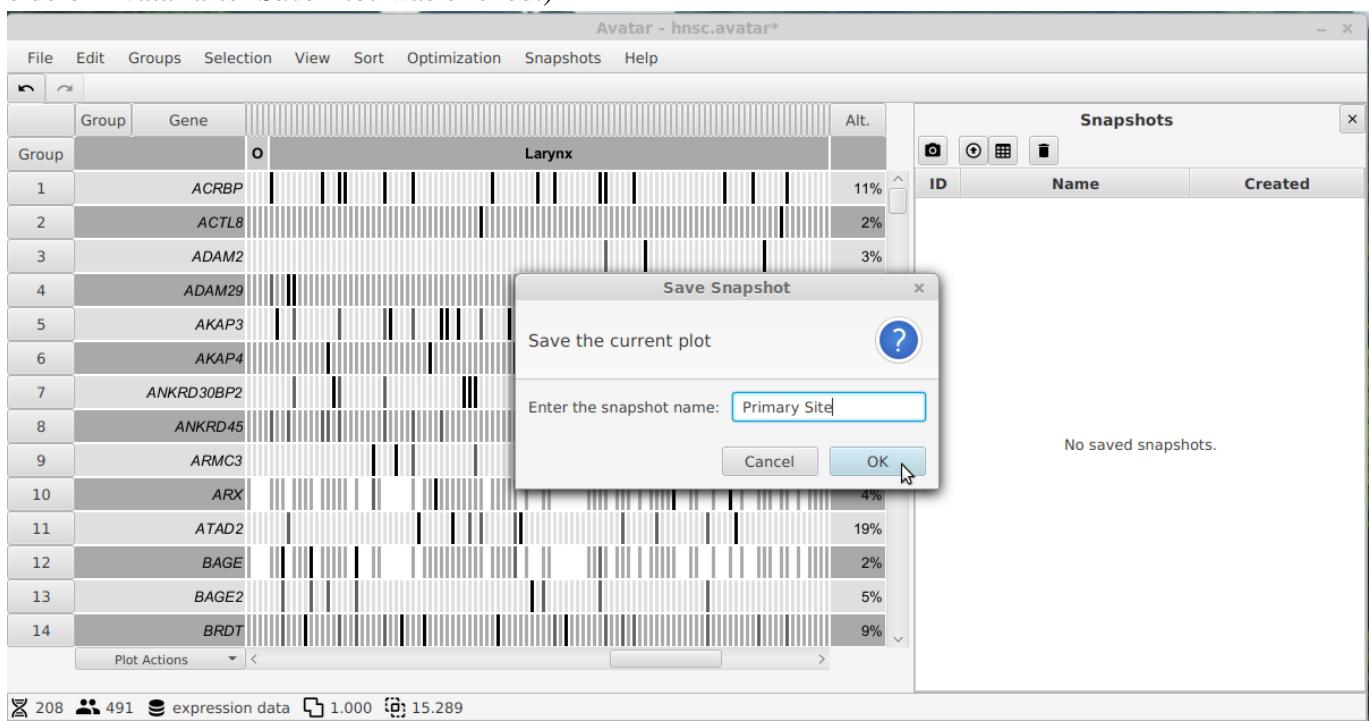


2.4 Create Snapshots

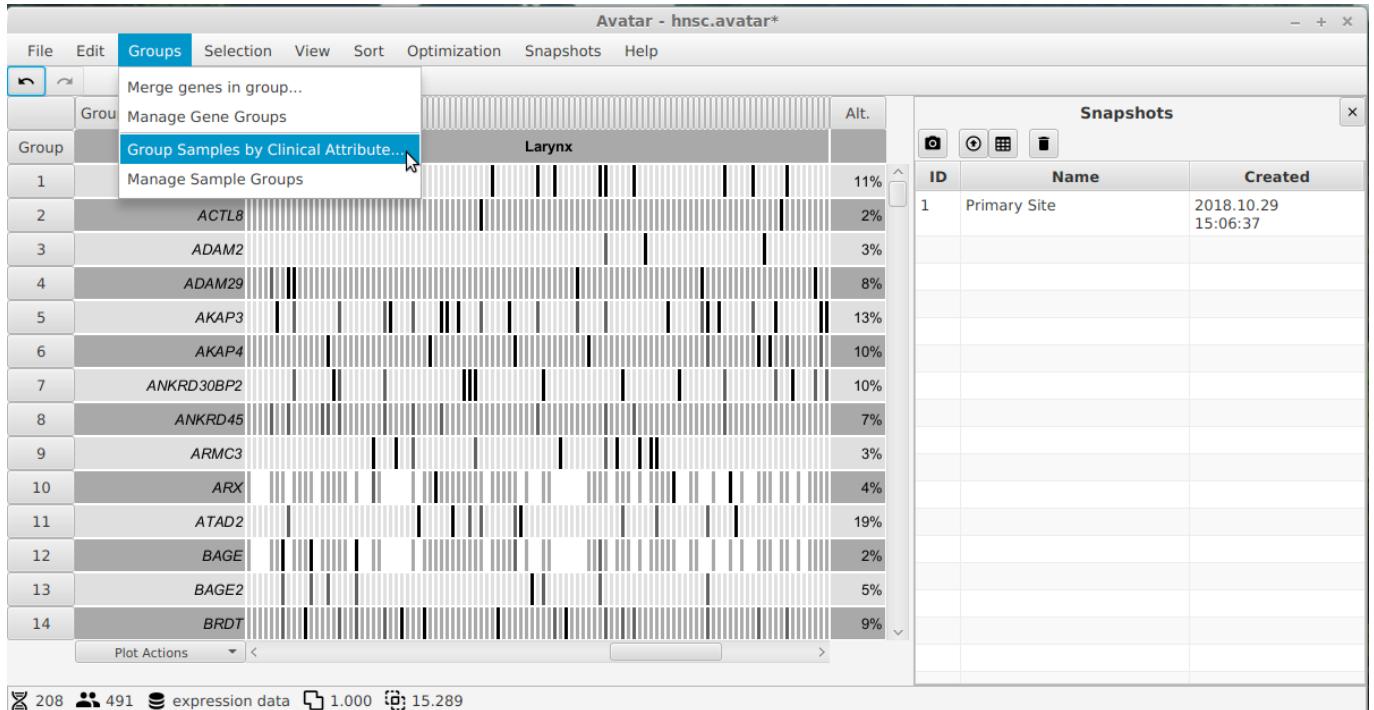
The current plot data with the samples grouped by primary site shall be saved as a snapshot.



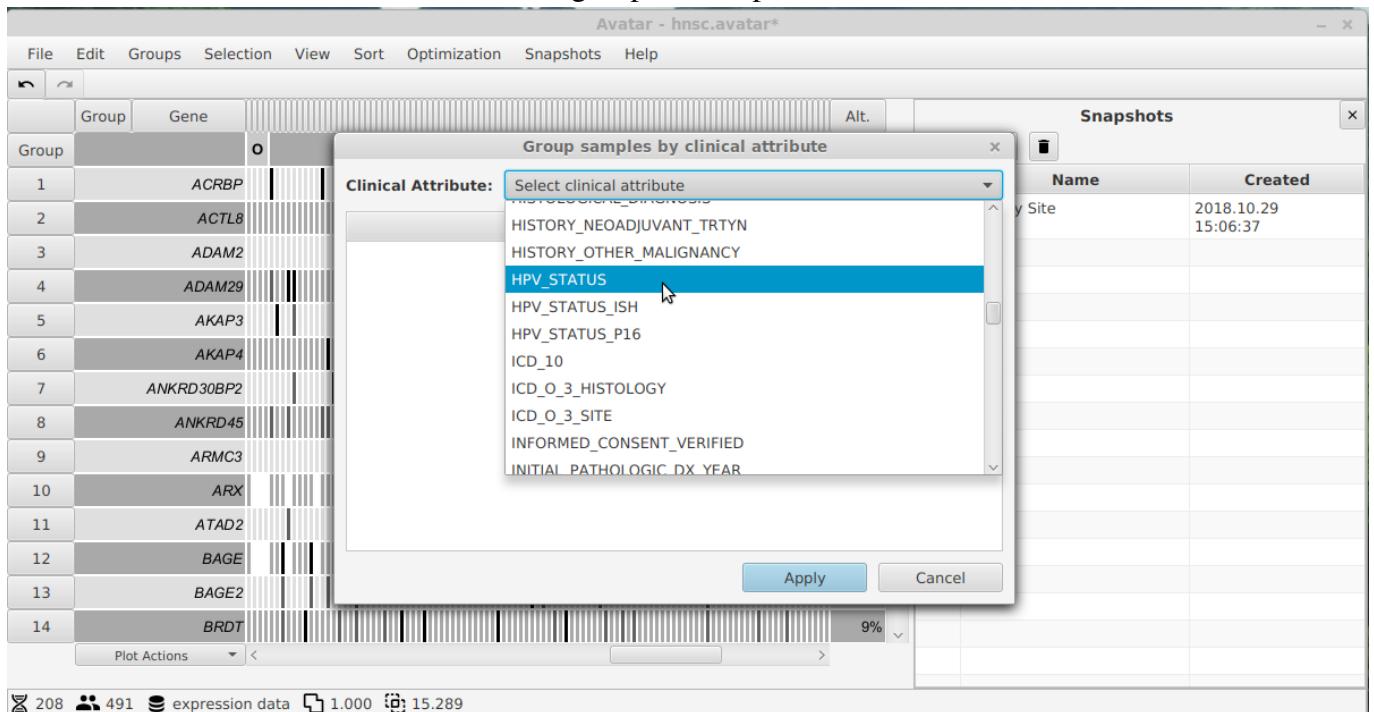
Specify a name for the snapshot, e.g. *Primary Site*. (The list of snapshots opened automatically on the right side of Avatar after *Save Plot* was clicked.)



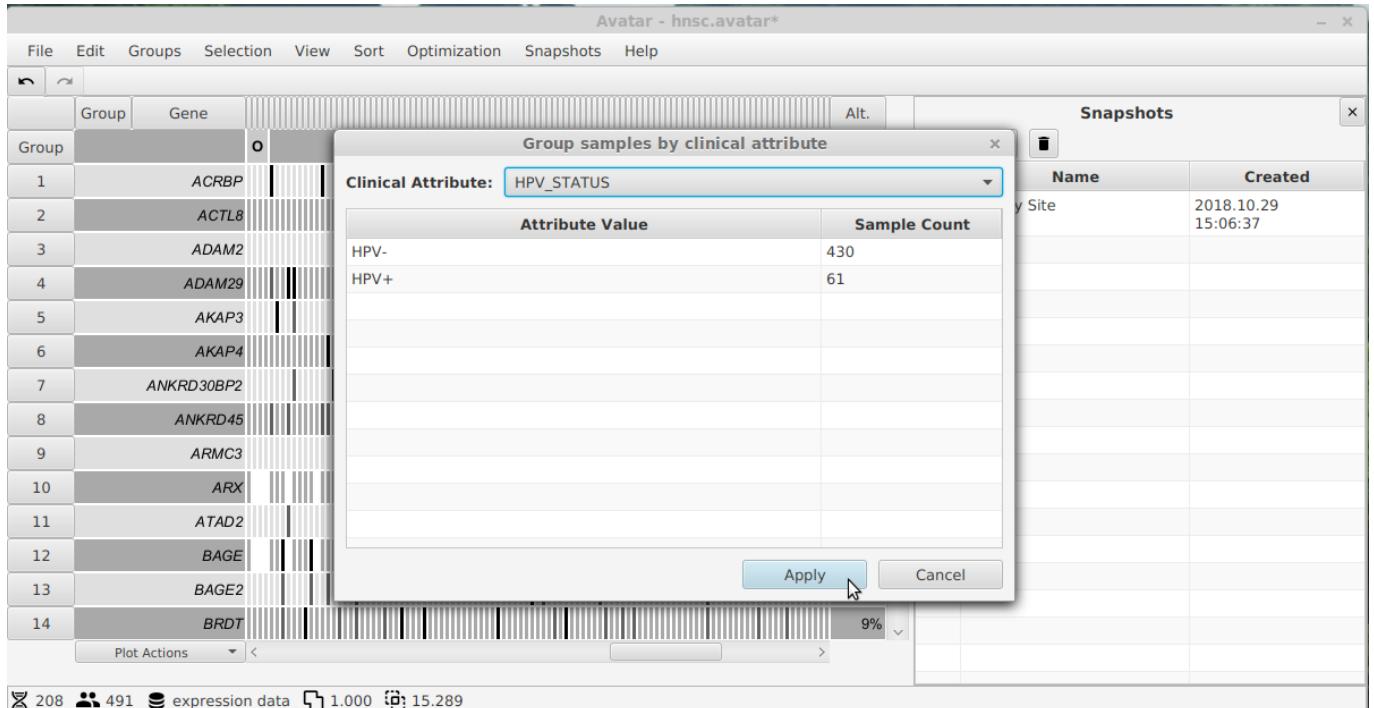
Now, the samples shall be grouped by HPV status. Click the corresponding menu item.



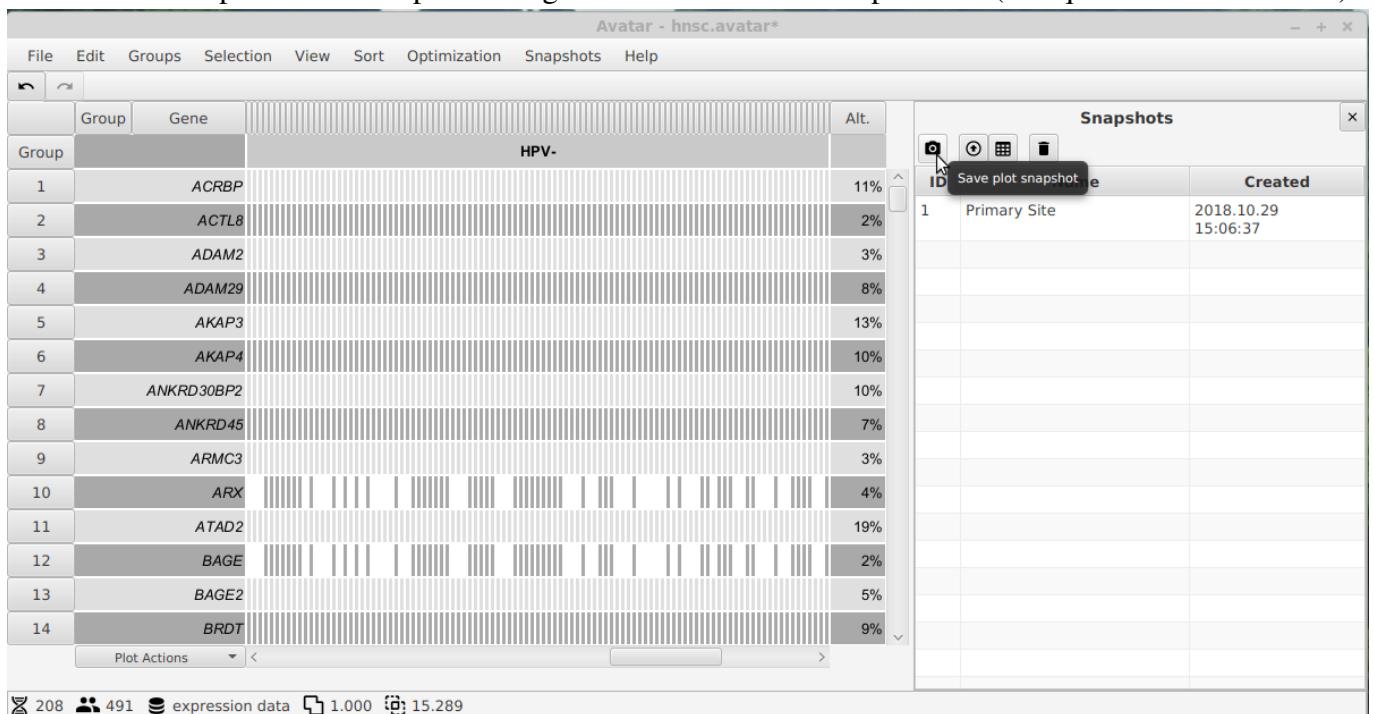
Select *HPV_STATUS* as clinical attribute to group the samples.



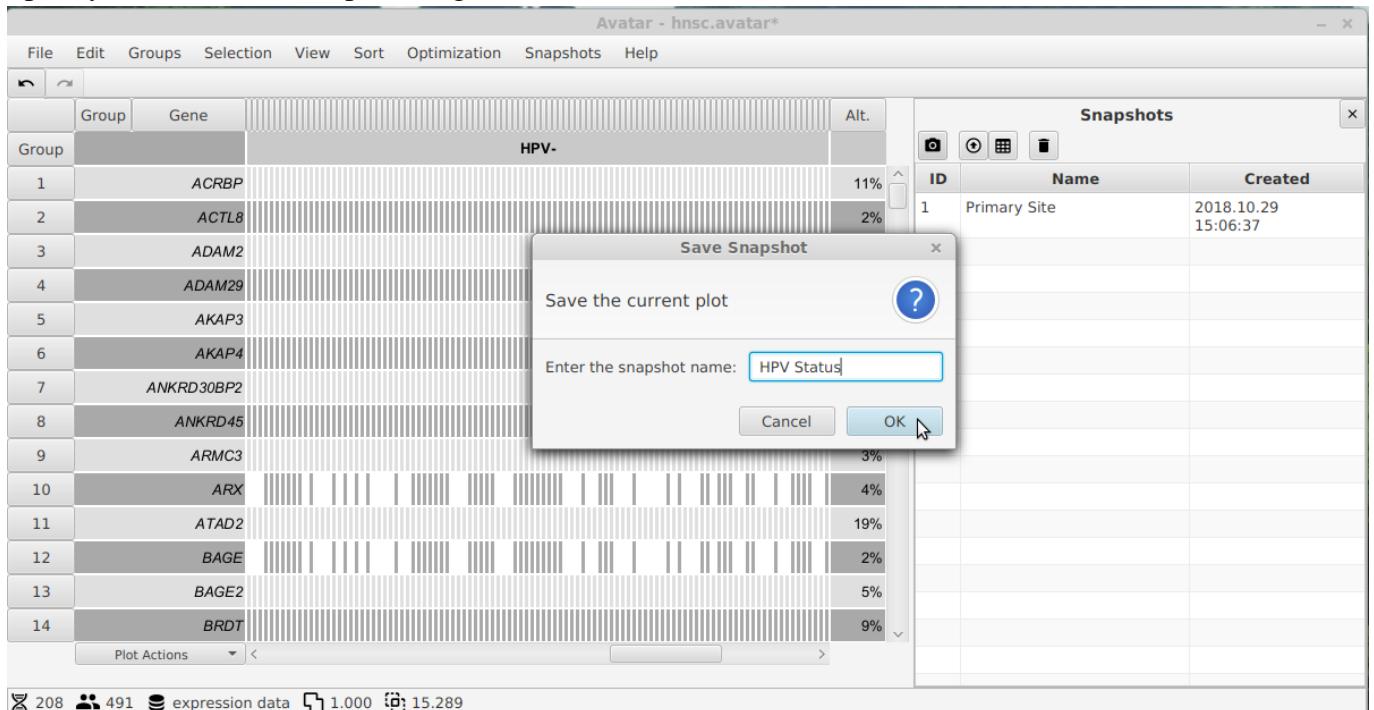
Confirm that the samples are either *HPV+* or *HPV-*. Click apply.



Save the current plot data as snapshot using the button above the snapshot list (it is quicker then the menu).

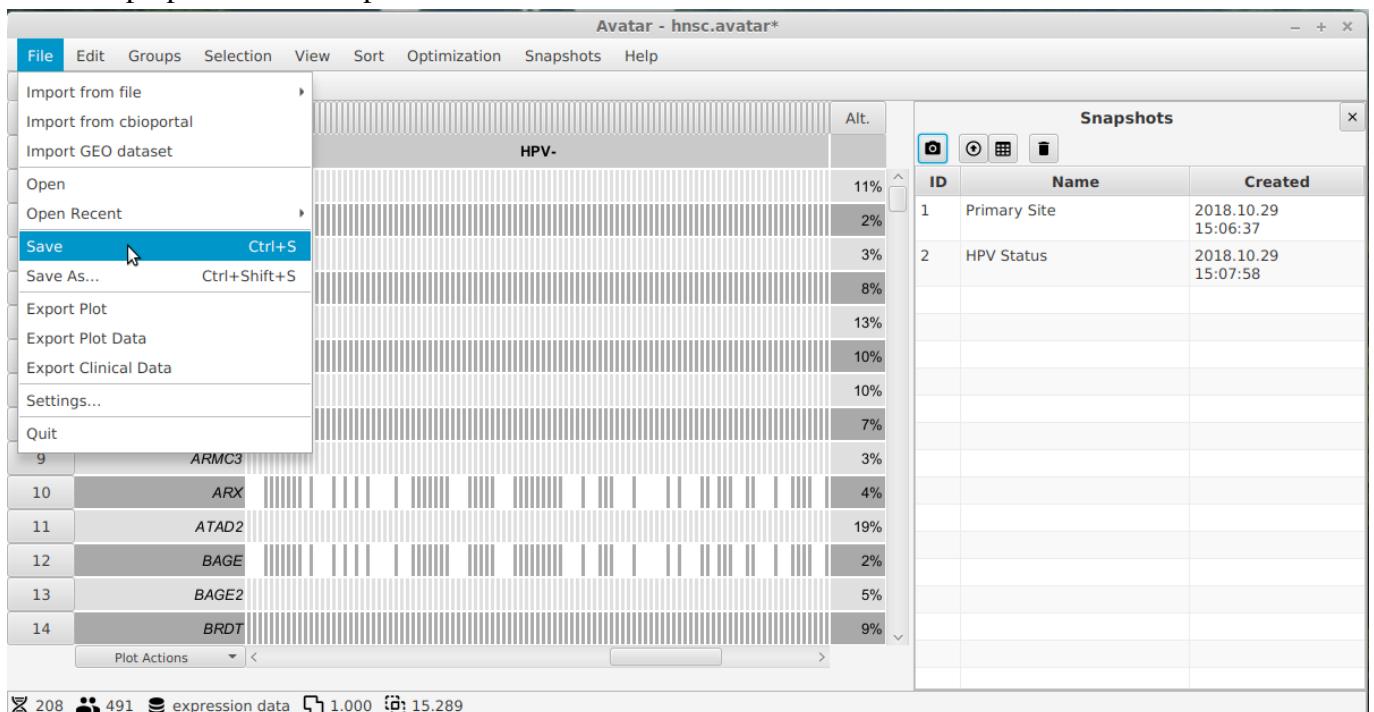


Specify a name for the snapshot, e.g. *HPV Status*.

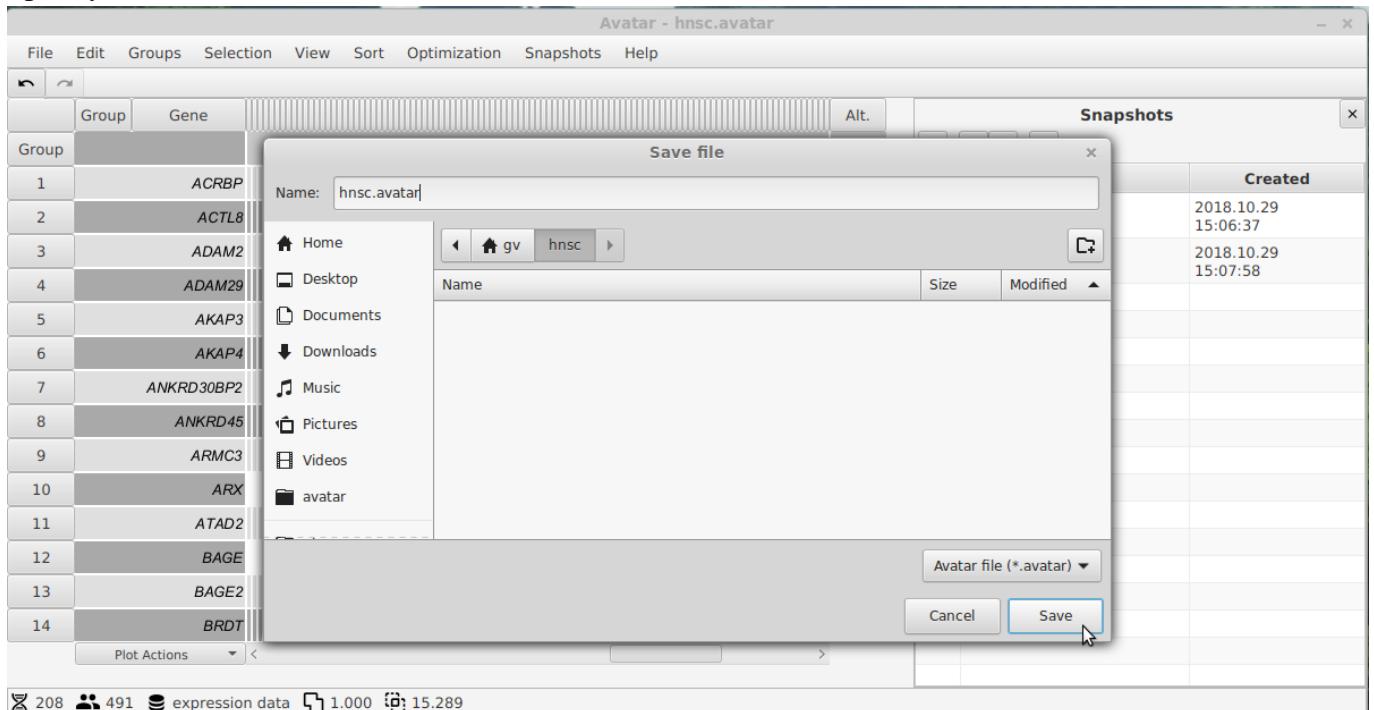


2.5 Save to File

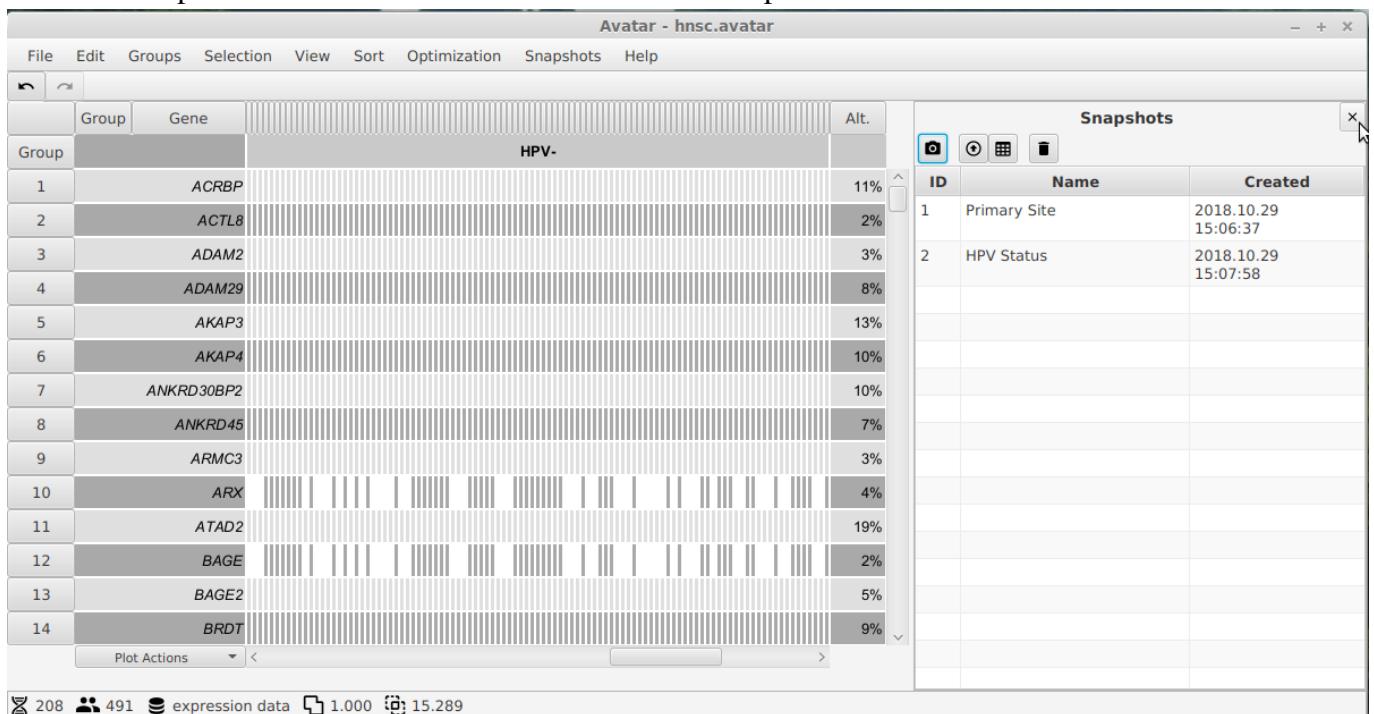
The data preparation is complete. Save the current data to a file.



Specify the filename where to store the data.



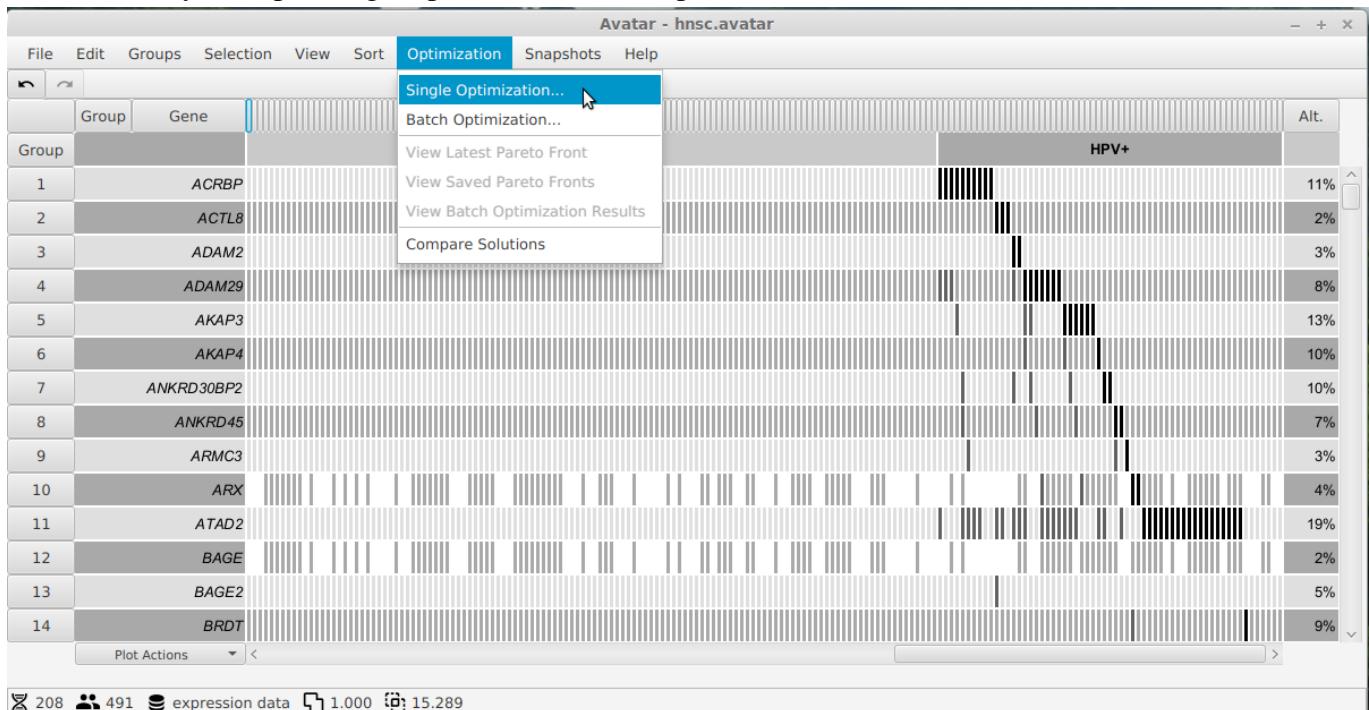
Close the snapshot list since it is not needed for the next steps.



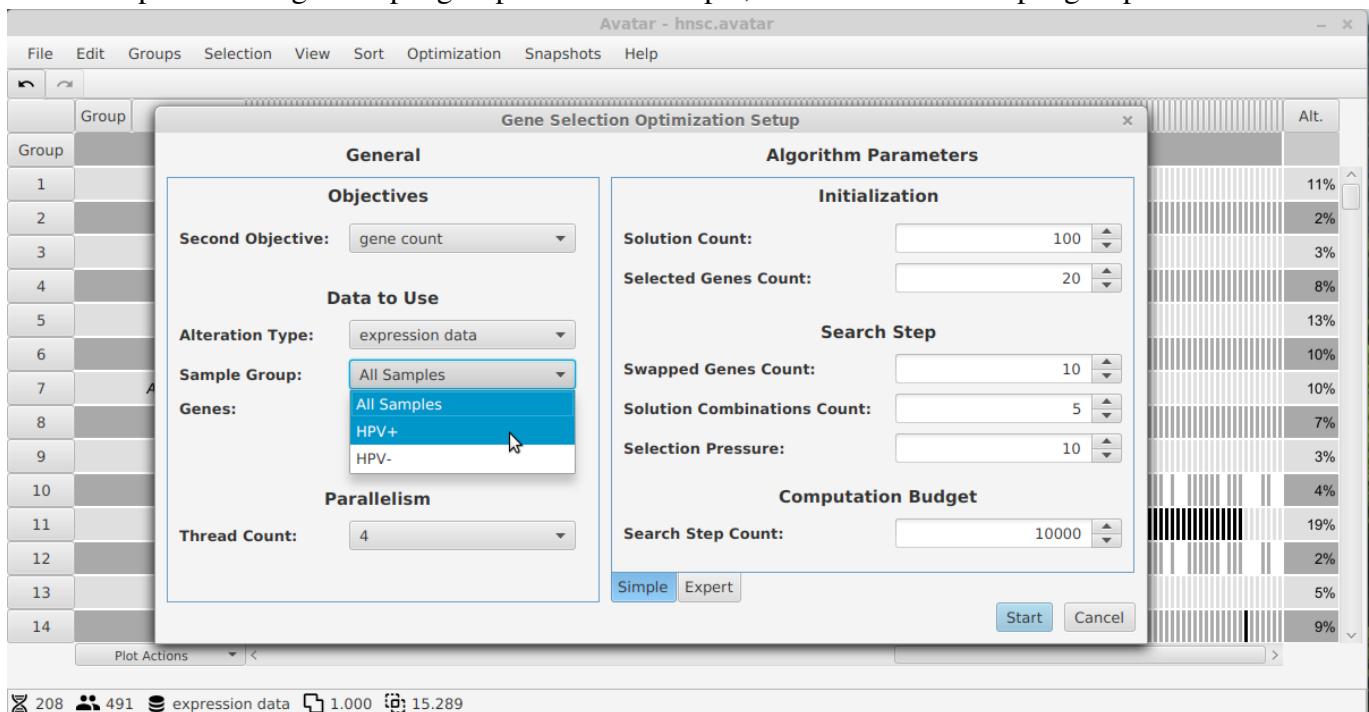
3 Optimization

3.1 Single Optimization

As a first analysis step, a single optimization will be performed. Click on the shown menu item.

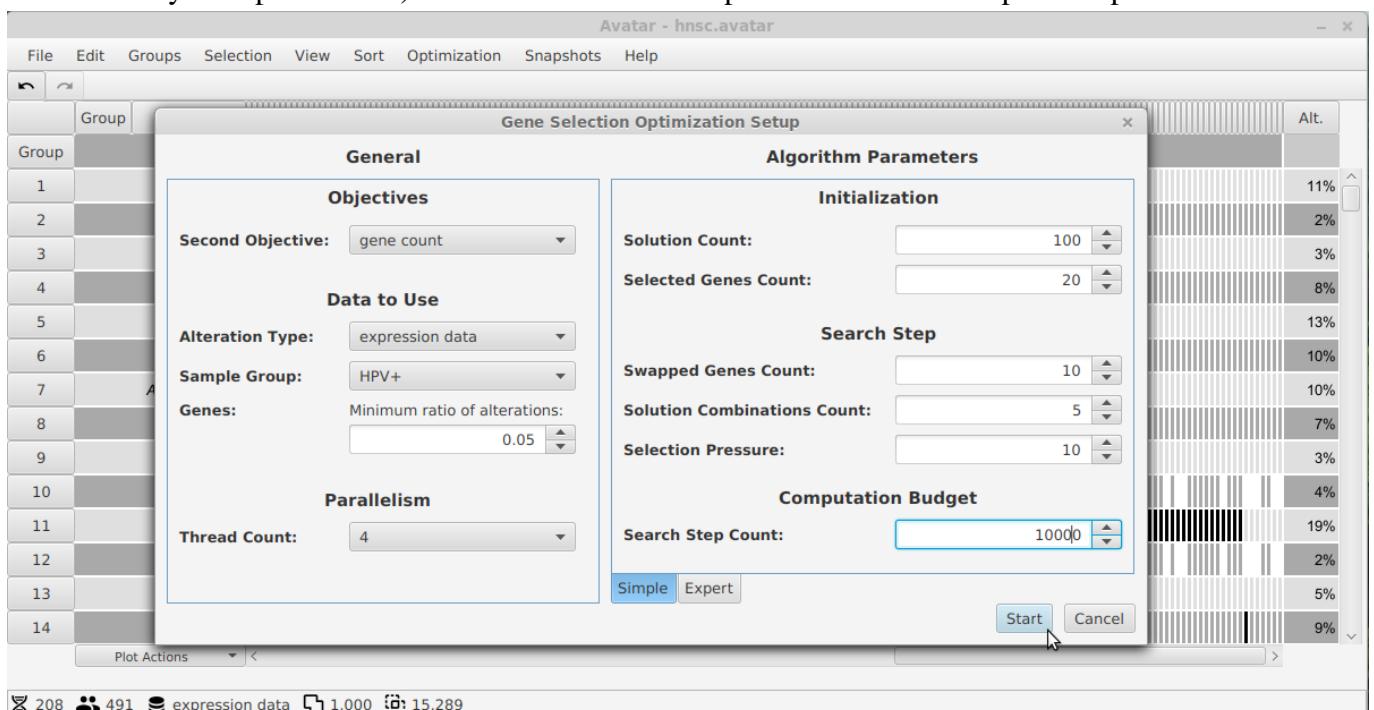


The setup dialog for single optimization specifies two aspects: the details of the optimization problem (on the left) and the settings for the evolutionary algorithm (on the right). Coverage of a gene selection will always be maximized. The second objective can be selected. Either the overlap within the gene selection or the gene count of the selection will be minimized. When the data contains more than one alteration type, the alteration type that will be used in the optimization needs to be specified. The optimization can be restricted to the samples of a single sample group. For this example, select *HPV+* as sample group.

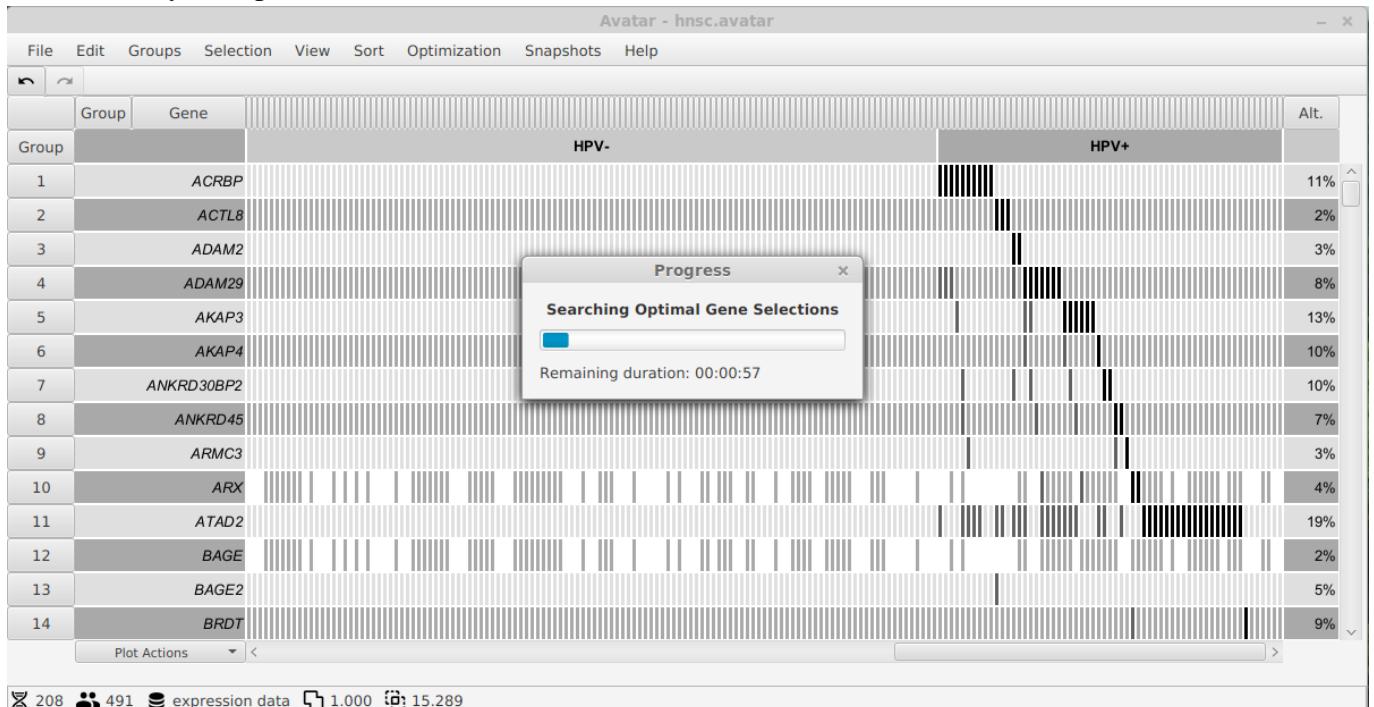


The genes can be limited to the ones that have a minimum number of alterations in the samples selected for optimization (here: *HPV+*). In this case a minimal relative number of 5% has been selected. So genes with less than 5% alterations in the samples of the *HPV+* group are not considered by the optimization. Parallelization is used to speed up the computation with the four CPU cores (the number depends on the hardware Avatar is running on) which have been detected automatically.

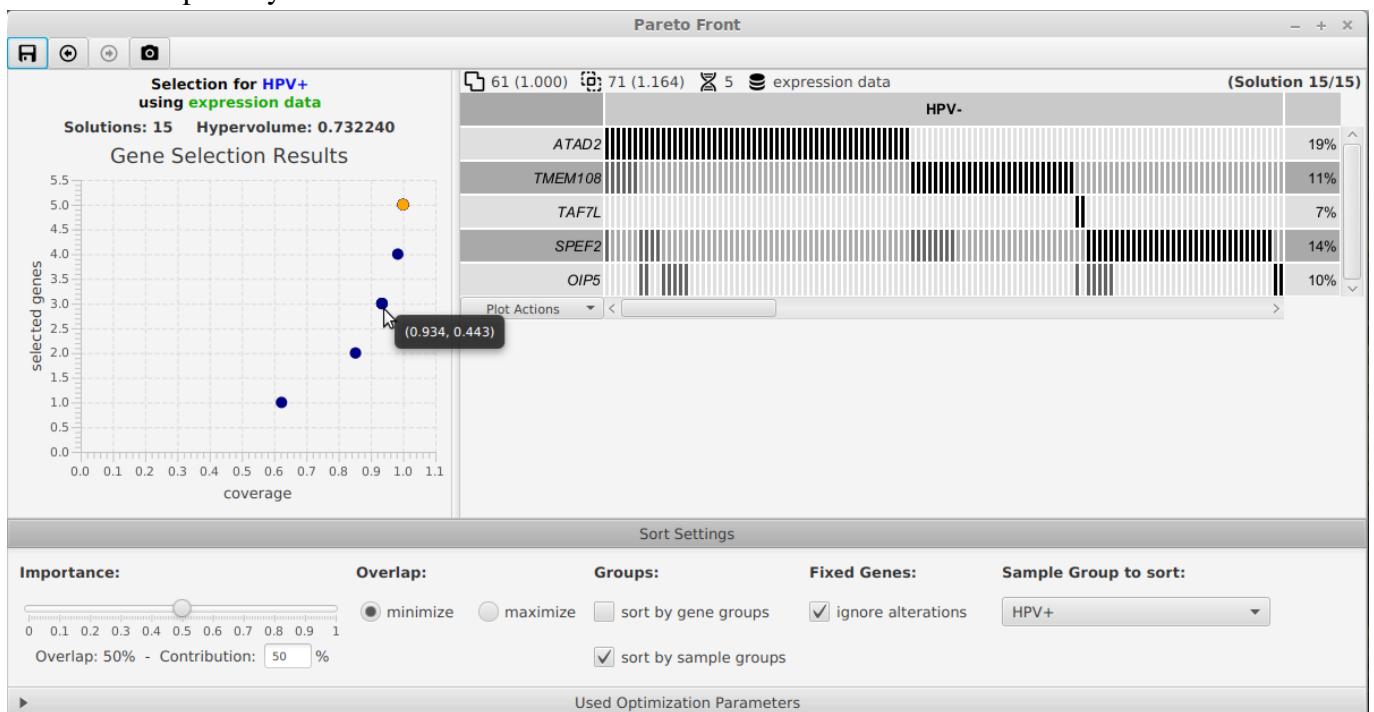
The default parameters for the evolutionary algorithm in *Simple* mode should be suitable for many cases. The parameters are explained in the Online Materials of the article. A brief comment on two important parameters: The *Solution Count* is the maximal number of solutions that the algorithm maintains at the same time throughout the algorithm. The resulting Pareto front (Pareto set) is limited to this maximal number of solutions. Together with the *Search Step Count* it determines the duration of the optimization. The larger the search step count the longer the duration. Broadly speaking, on the other hand better solutions are found for larger search step counts. The duration is also influenced by the data (number of genes and samples considered by the optimization). Click *Start* to run the optimization with the specified parameter values.



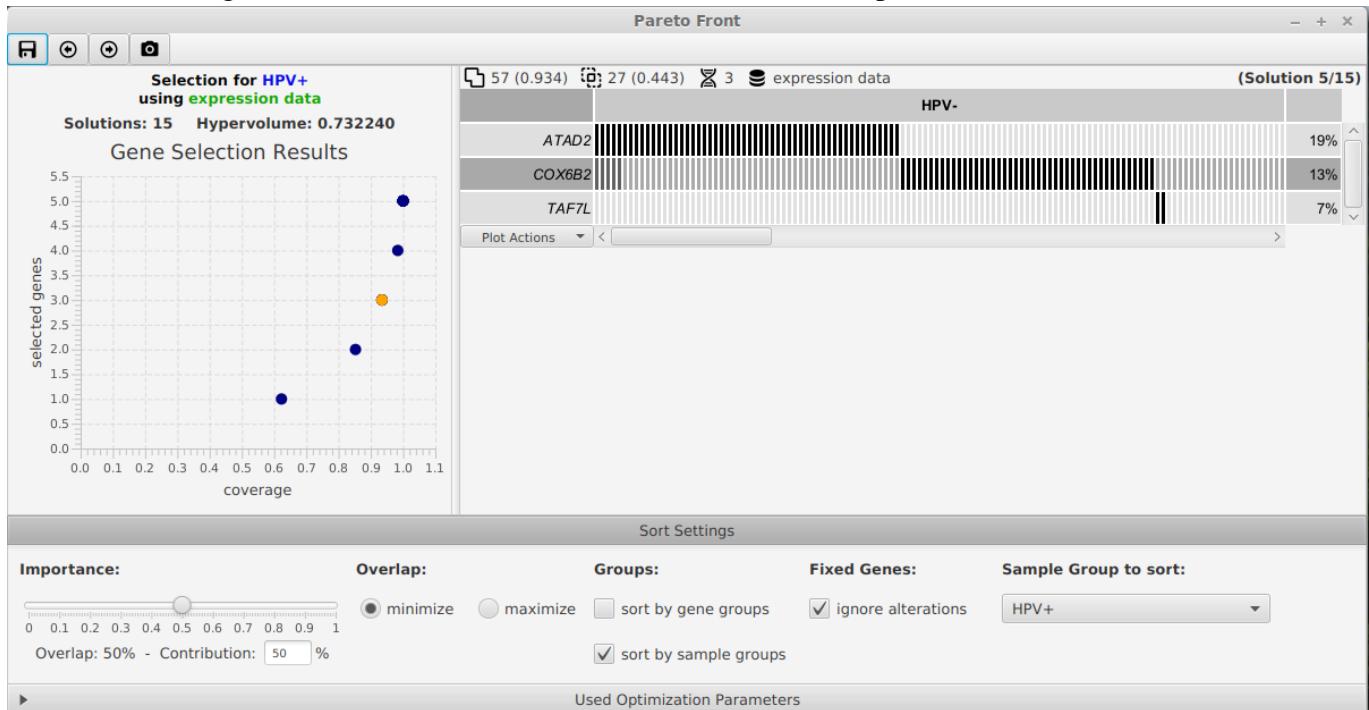
The progress of the optimization algorithm is shown including an estimated remaining duration. In this case the total duration is rather small for 10^4 search steps since only 208 genes and the 61 *HPV+* samples are considered by the optimization.



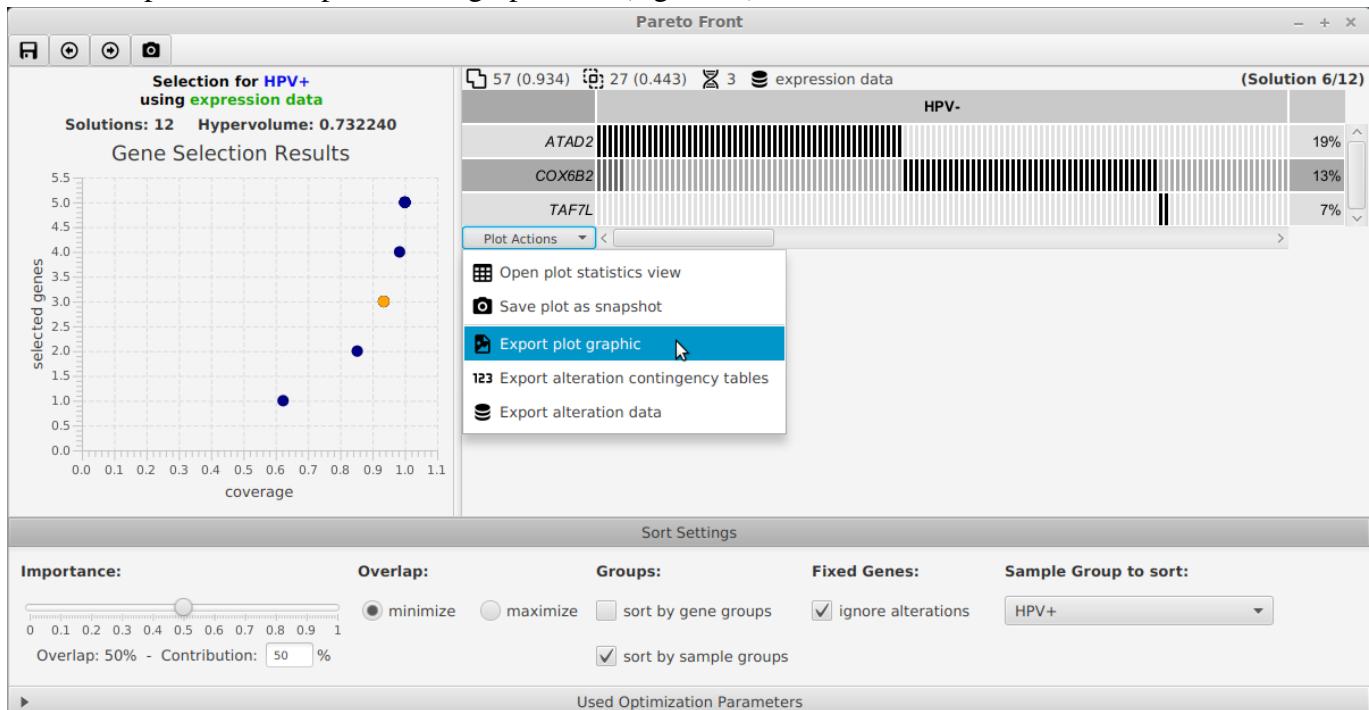
When the optimization is finished, a dialog showing the Pareto front is opened. The Pareto front is shown on the left. Each solution is represented as a circle in a scatter plot with the coverage on the x-axis and the second objective on the y-axis (here: gene count). Double click a circle to display the corresponding solution in an alteration plot on the right. This plot can be sorted using the settings on the bottom. Navigating through the solutions is also possible using the arrow buttons in the toolbar at the top. Sometimes there are multiple solutions with the same objective values but different genes as in this example. Then, the arrow buttons are especially useful.



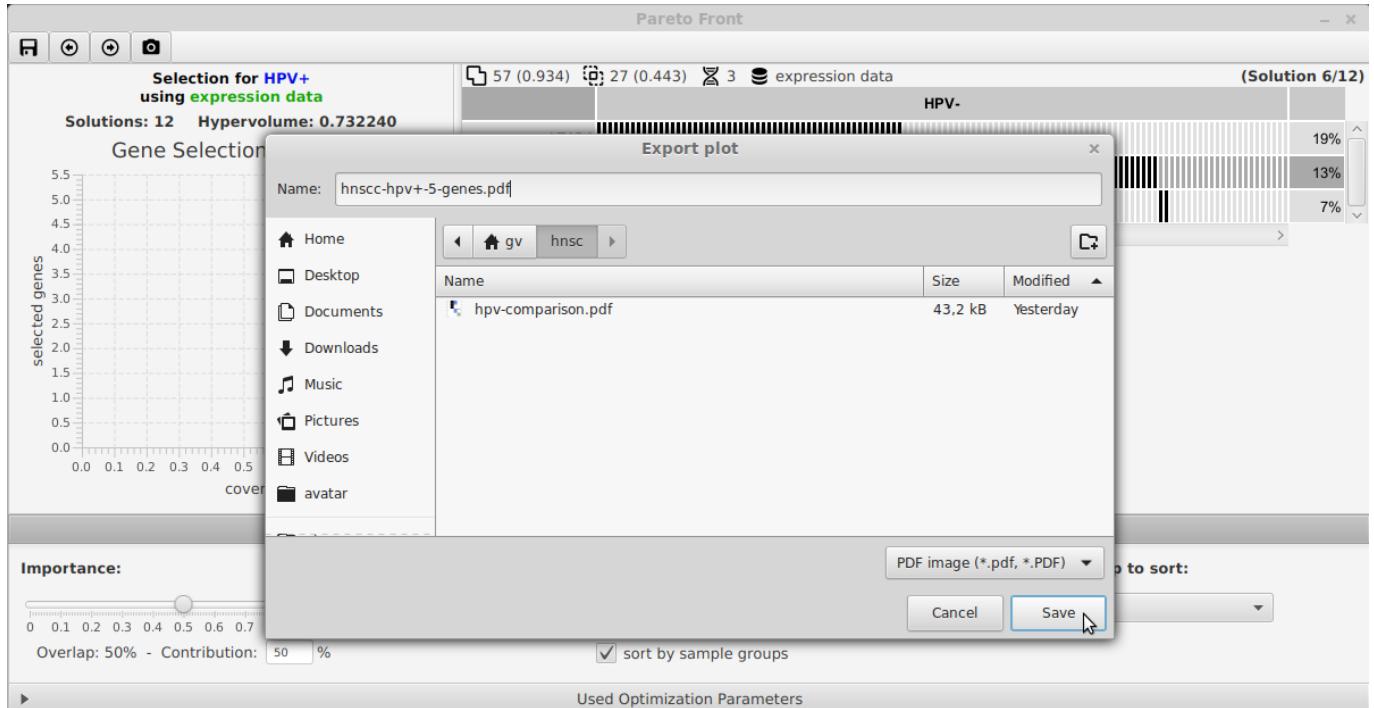
One of the three genes solutions is shown after a double click in the plot.



Most alteration plots in Avatar have the *Plot Actions* menu on the bottom left. Among other possibilities, the shown plot can be exported as a graphic file (e.g. PDF).

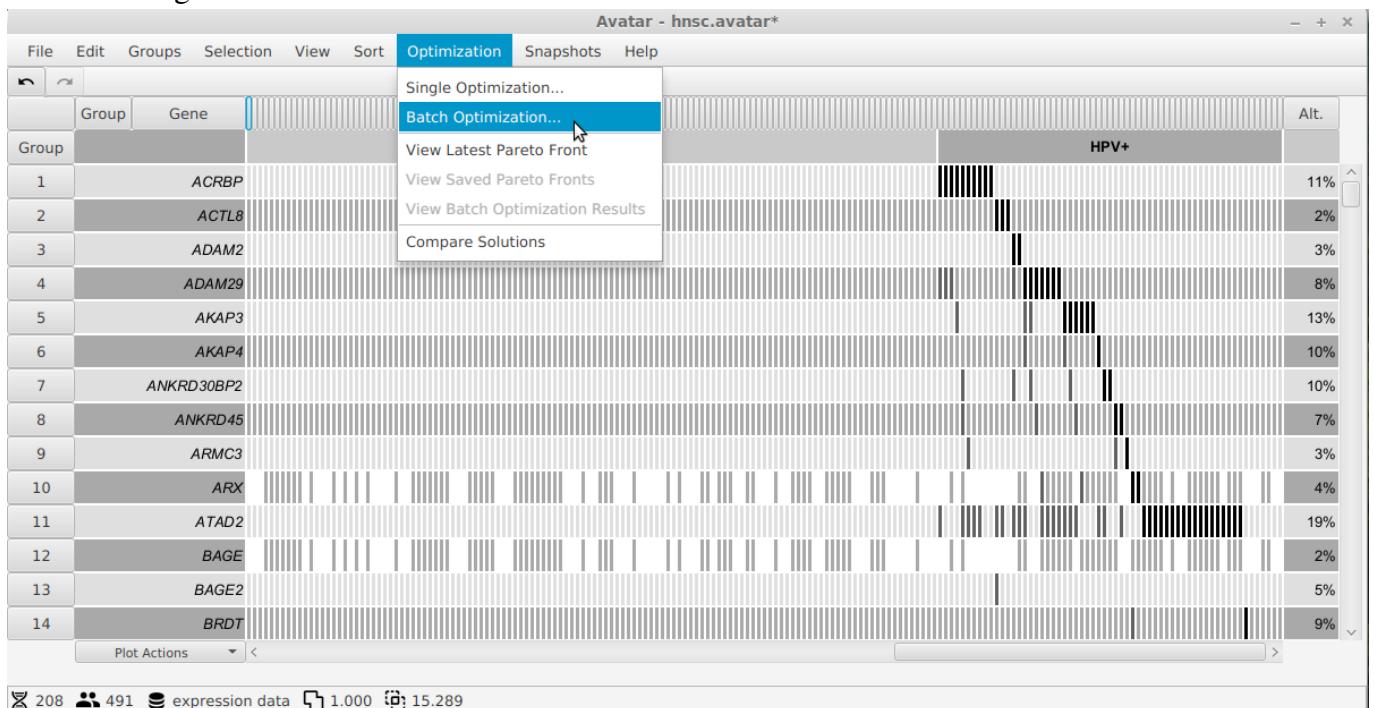


Choose a file name where to store the plot graphic.

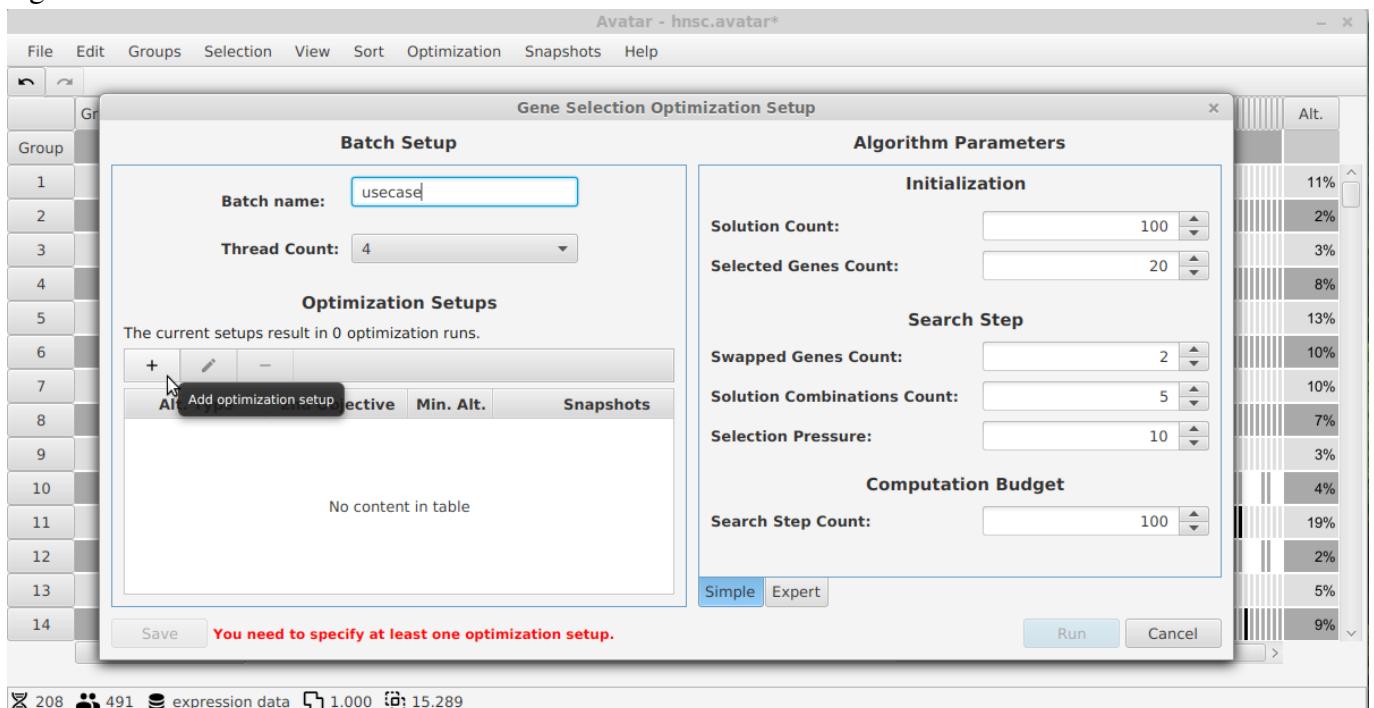


3.2 Batch Optimization

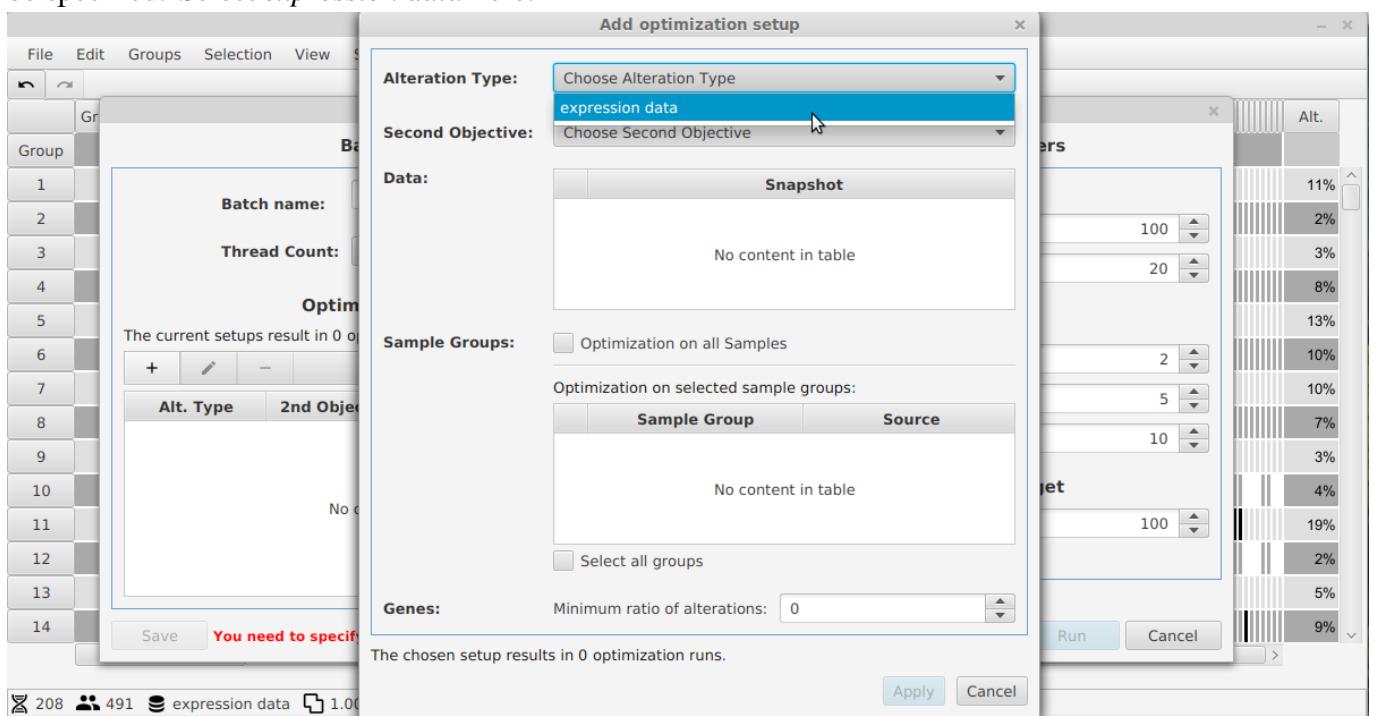
Performing optimization individually for a larger number of sample groups will be cumbersome. Therefore, Avatar offers the *Batch Optimization* mode. Here, several optimization problems can be specified and executed altogether afterwards. Click on the shown menu item.



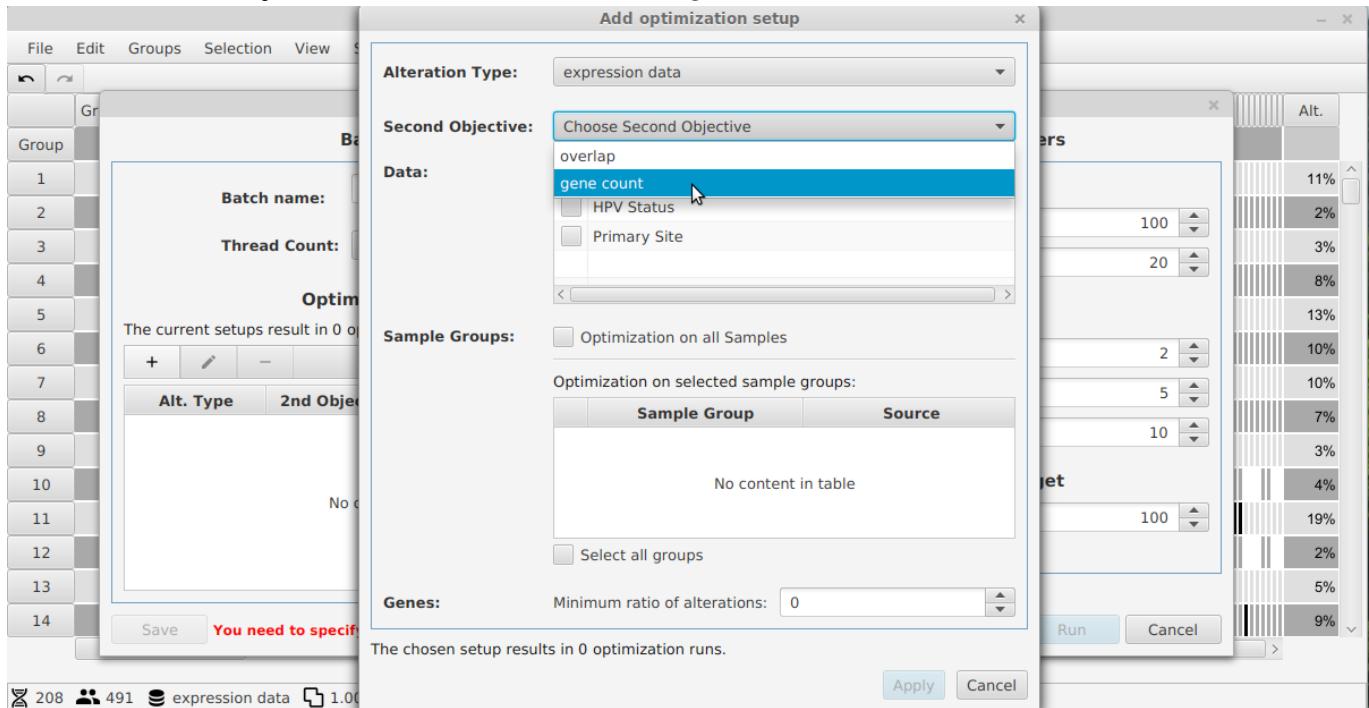
The batch setup dialog opens. The batch of optimization problem is specified on the left. Algorithm parameters (same as in single optimization) are configured on the right. Specify a name for the batch setup, e.g. *usecase*. Click on the + button above the table on the left to add a batch.



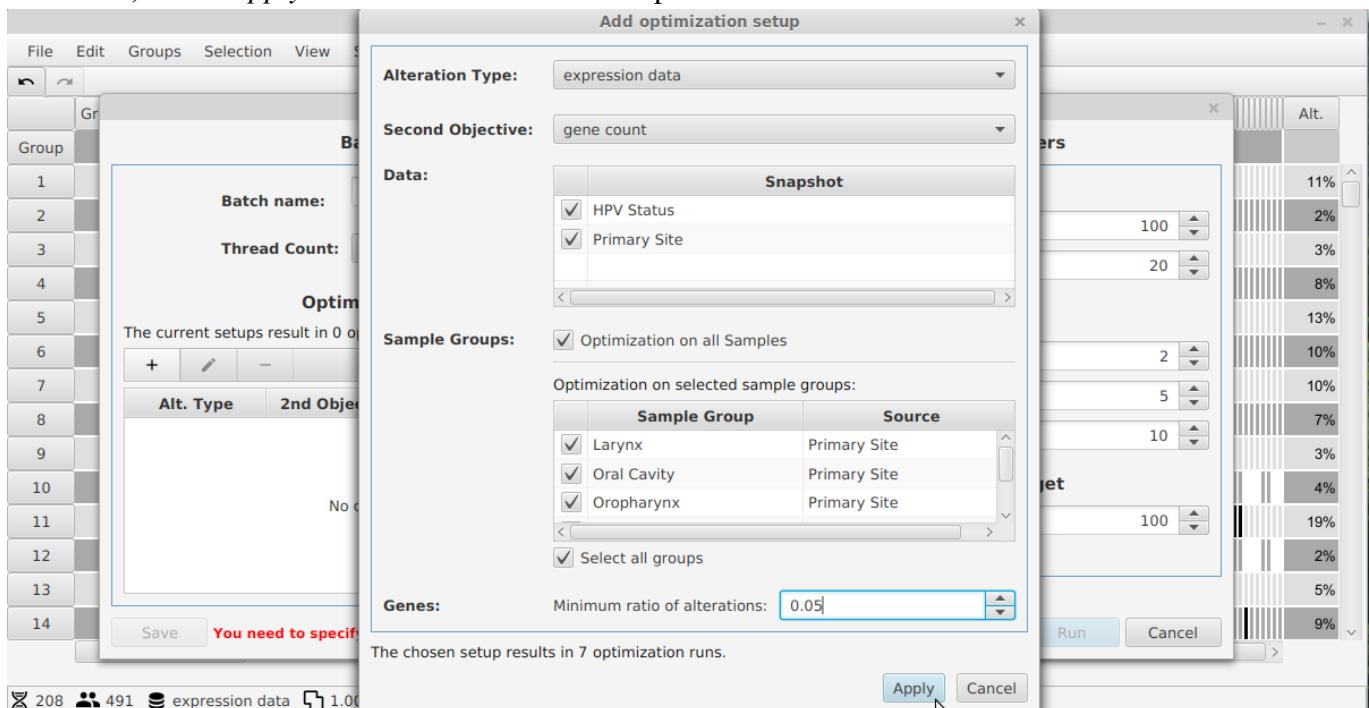
This opens the configuration dialog for a batch. First, the alteration type to use in the optimization needs to be specified. Select *expression data* here.



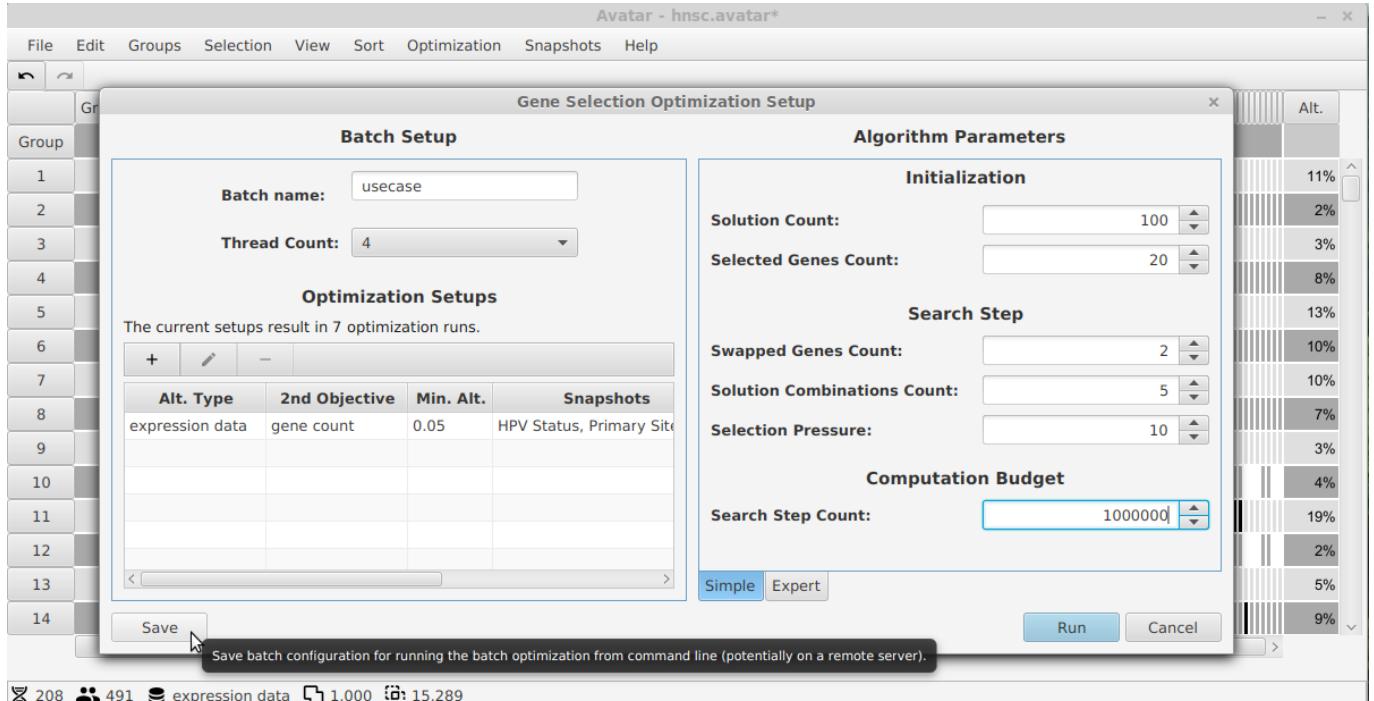
Then, the second objective must be selected. Choose *gene count* here.



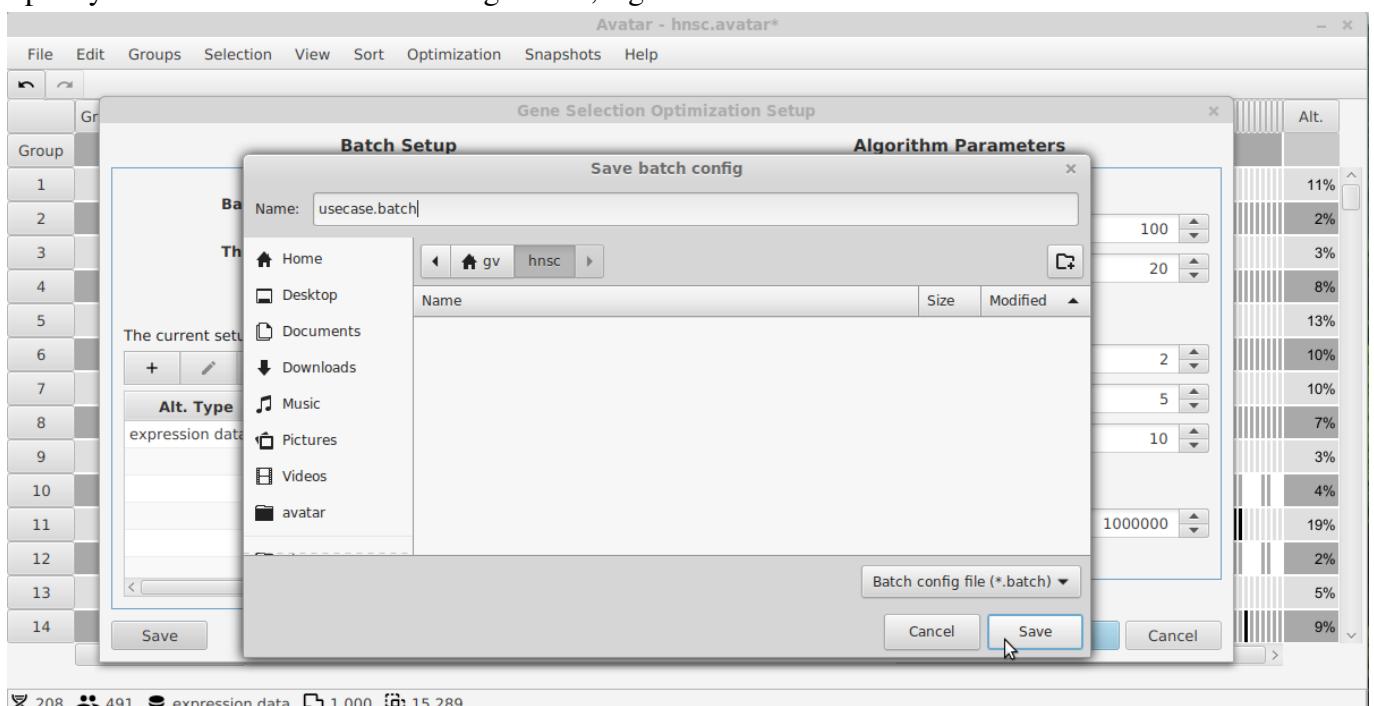
Choose both previously created snapshots from the list of compatible snapshots (labeled *Data*). Select all sample groups and check *Optimization on all Samples*. Specify a *minimum ratio of alterations* of 5% (for an explanation see Single Optimization). Avatar informs that this batch results in seven optimization runs: three primary site sample groups, two HPV status sample groups and all samples twice. (This latter redundancy arises in some cases due the easy-to-use setup process and could be avoided in future versions of Avatar.) Click *Apply* to add the batch to the setup.



Back on the batch setup dialog, the add batch can be seen in the corresponding table on the left. More batches can be added by following the previous description. On the right, specify a *search step count* of 10^6 to reproduce the HNSCC usecase. When only a quick demonstration is desired, use a smaller number, e.g. 10^3 . For the quick demonstration just click *Run*. To reproduce the usecase, you should save the batch setup as shown in the following. This allows to execute the batch optimization on a computation server with many CPU cores.



Specify a filename for the batch configuration, e.g. *usecase.batch*.



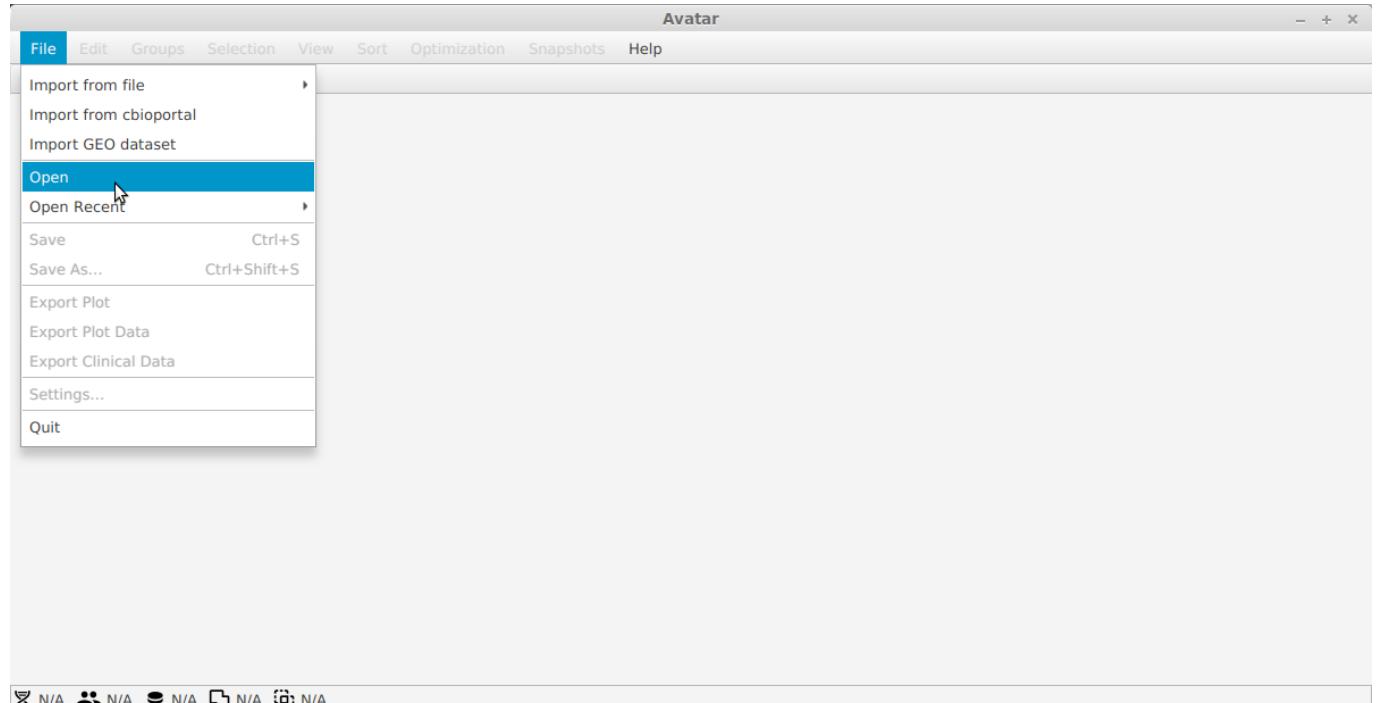
Now, copy the Avatar software (e.g. `avatar-0.4.4-SNAPSHOT.jar`), the saved Avatar file (e.g. `hnsc.avatar`) and the batch configuration file (e.g. `usecase.batch`) to an available computation server. The batch optimization is started via the following command. Additionally, the command specifies that seven optimization runs shall be executed in parallel. (Alternatively, ask a technician to do these steps.)

```
$ javall -jar avatar-0.4.4-SNAPSHOT.jar -f hnsc.avatar -o usecase.batch -t 7
```

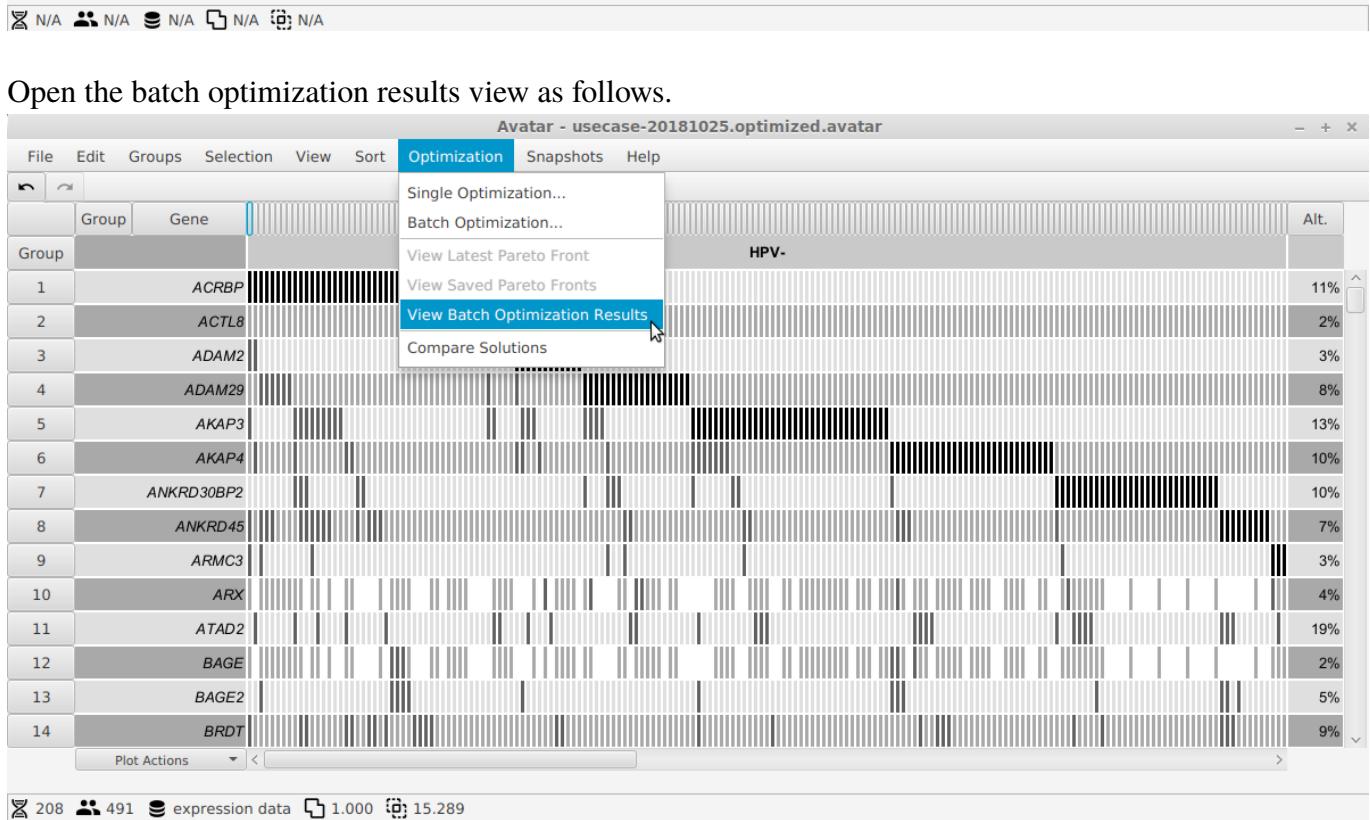
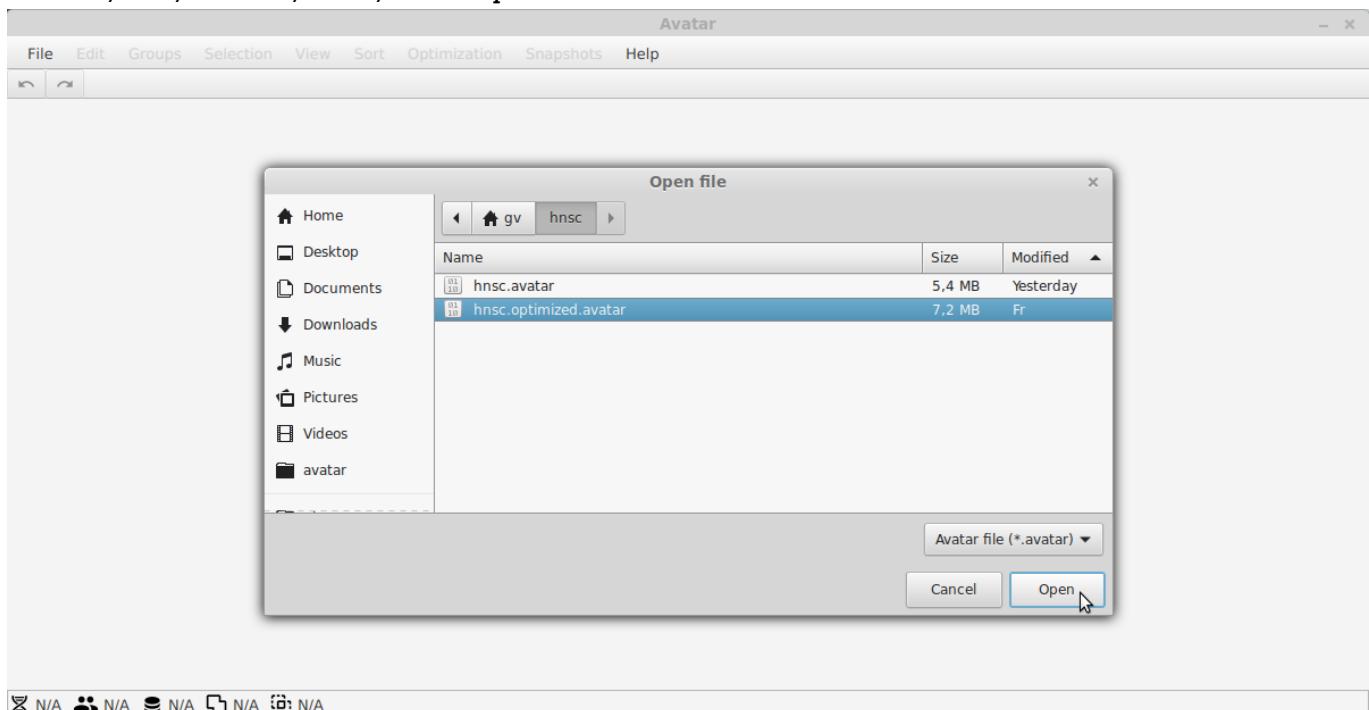
Avatar shows information about the progress of the batch optimization including an estimation of the remaining duration. The result is saved in a separate file containing `.optimized.` in the file name, e.g. `hnsc.optimized.avatar` for `hnsc.avatar`. Copy the result file from the server to continue working with it.

```
Runs: 0/7 - Progress: 0.243% - Estimated Duration: 05:42:54
```

Start Avatar and open the batch optimization results file.



Select the results file. Our results file can be downloaded from <https://github.com/sysbio-bioinf/avatar/raw/master/dist/hnsc.optimized.avatar>.

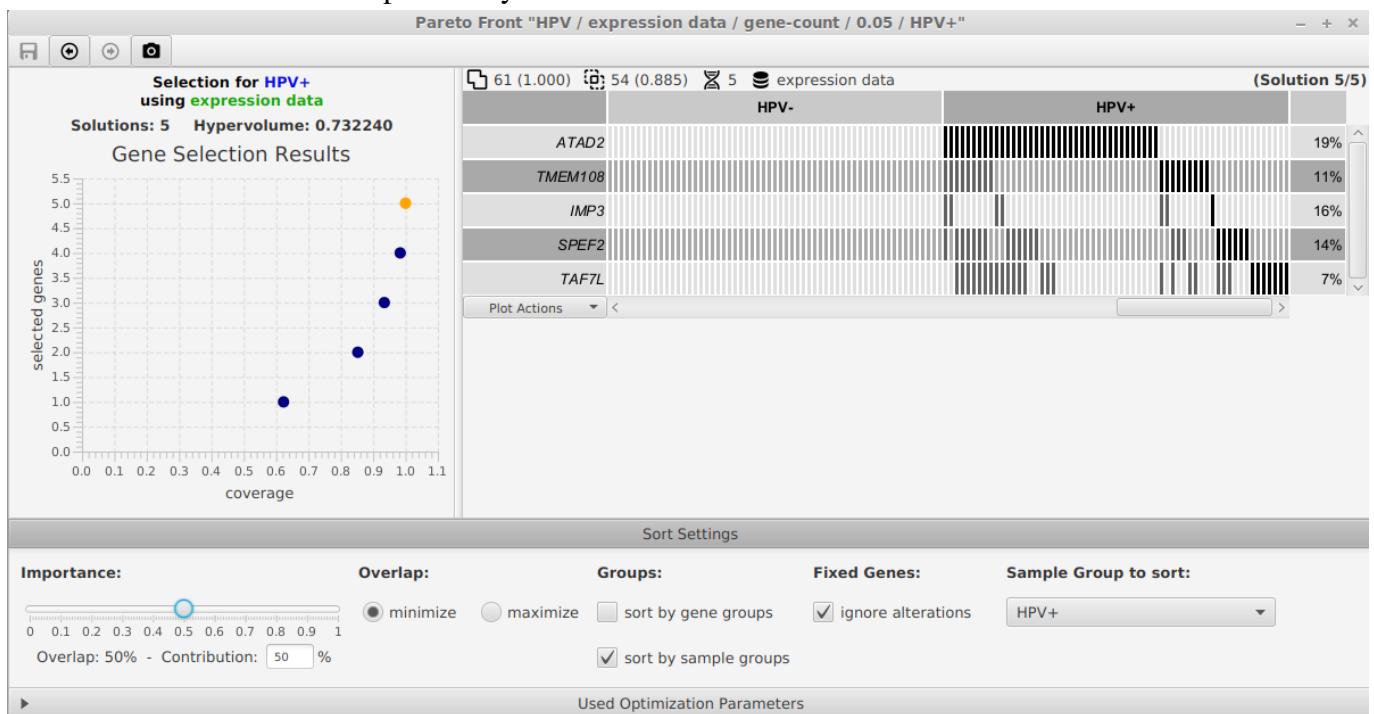


This view shows a table containing the batch results (solutions) along with the most important parameters. Double click on a solution row (e.g. the one with sample group *HPV+*) to open the Pareto front view (described earlier).

Batch Optimization Pareto Fronts							
Group	Batch	Snapshot	Alteration Type	Second Objective	Min. Alt.	Sample Group	Hypervolume
1							11%
2	usecase	HPV	expression data	gene count	0.05		0.831806
3	usecase	HPV	expression data	gene count	0.05	HPV+	0.732240
4	usecase	HPV	expression data	gene count	0.05	HPV-	2%
5	usecase	PS	expression data	gene count	0.05	Larynx	3%
6	usecase	PS	expression data	gene count	0.05	Oral Cavity	8%
7	usecase	PS	expression data	gene count	0.05	Oropharynx	13%
8							10%
9							7%
10							3%
11							4%
12							19%
13							2%
14	BRDT						5%
							9%

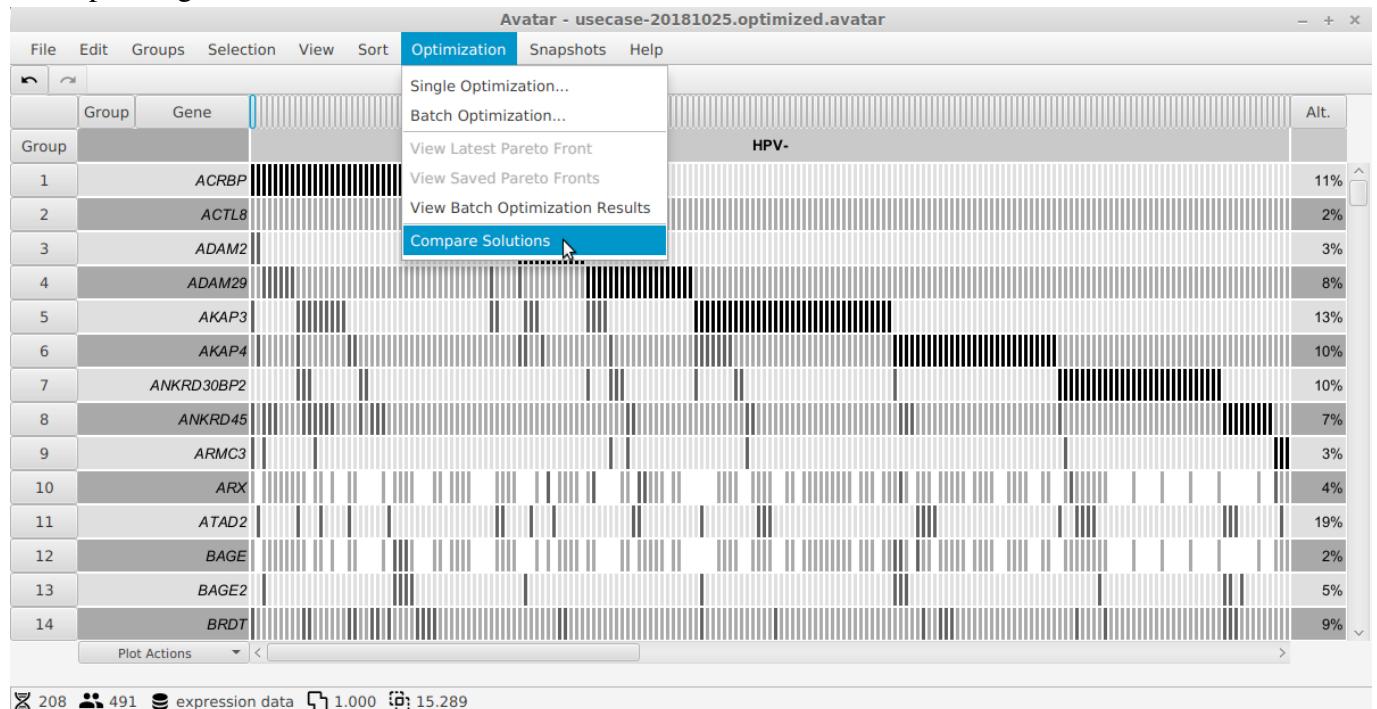
208 491 expression data 1.000 15.289

The Pareto front view for the previously selected solution.



3.3 Solution Comparison

In addition to the Pareto front view, solutions can be compared with respect to the selected genes. Next, it is described how to create a comparison of the solutions with at most 10 genes for the optimization runs on all samples, the *HPV-* samples and the *HPV+* samples. Open the solution comparison view using the corresponding menu.



The solution comparison view consists of three parts: a tree view (left) to select the solutions to compare, a comparison table (right) which shows the selected solutions and a settings part (bottom) to customize the comparison table. Snapshots can also be selected for comparison. Expand the subtree at *Batches*, *usecase* and then *1 HPV/.../ALL* as shown.

Solution Comparison

Name				
▶ Snapshots	<input type="checkbox"/>			
Saved Pareto Fronts	<input type="checkbox"/>			
▼ Batches	<input type="checkbox"/>			
▼ usecase	<input type="checkbox"/>			
▶ 1 HPV / expression data / gene-count / 0.05 / ALL	<input checked="" type="checkbox"/>			
▶ 2 HPV / expression data / gene-count / 0.05 / HPV+	<input type="checkbox"/>			
▶ 3 HPV / expression data / gene-count / 0.05 / HPV-	<input type="checkbox"/>			
▶ 4 PS / expression data / gene-count / 0.05 / ALL	<input type="checkbox"/>			
▶ 5 PS / expression data / gene-count / 0.05 / Larynx	<input type="checkbox"/>			
▶ 6 PS / expression data / gene-count / 0.05 / Oral Cavity	<input type="checkbox"/>			
▶ 7 PS / expression data / gene-count / 0.05 / Oropharynx	<input type="checkbox"/>			

No gene sets selected.

Nothing to display.

Columns Style Objectives

Select the 10-gene solution.

Solution Comparison

Name		X			
▶ Snapshots					
Saved Pareto Fronts					
▼ Batches					
▼ usecase					
▼ 1 HPV / expression data / gene-count / 0.05 / ALL					
usecase/1 #1		20	491	1.000	1.79
usecase/1 #2		19	491	0.998	1.78
usecase/1 #3		18	491	0.996	1.77
usecase/1 #4		17	491	0.992	1.70
usecase/1 #5		16	491	0.988	1.61
usecase/1 #6		15	491	0.982	1.48
usecase/1 #7		14	491	0.974	1.43
usecase/1 #8		13	491	0.967	1.30
usecase/1 #9		12	491	0.955	1.17
usecase/1 #10		11	491	0.941	1.06
usecase/1 #11		10	491	0.923	1.02
usecase/1 #12		9	491	0.902	0.87
usecase/1 #13		8	491	0.876	0.79
usecase/1 #14		7	491	0.841	0.72
usecase/1 #15		6	491	0.804	0.59

No gene sets selected.

Nothing to display.

Columns Style Objectives

The solution is displayed in the table on the right. The rows contain the genes of the solutions and a black rectangle marks the genes contained in the solution. Edit the caption of the selected solution to *All* in the table at the bottom.

Name						
▶	Snapshots	<input type="checkbox"/>				
	Saved Pareto Fronts	<input type="checkbox"/>				
▼	Batches	<input type="checkbox"/>				
▼	usecase	<input type="checkbox"/>				
▼	1 HPV / expression data / gene-count / 0.05 / ALL	<input type="checkbox"/>				
usecase/1 #1		<input type="checkbox"/>	20	491	1.000	1.79
usecase/1 #2		<input type="checkbox"/>	19	491	0.998	1.78
usecase/1 #3		<input type="checkbox"/>	18	491	0.996	1.77
usecase/1 #4		<input type="checkbox"/>	17	491	0.992	1.70
usecase/1 #5		<input type="checkbox"/>	16	491	0.988	1.61
usecase/1 #6		<input type="checkbox"/>	15	491	0.982	1.48
usecase/1 #7		<input type="checkbox"/>	14	491	0.974	1.43
usecase/1 #8		<input type="checkbox"/>	13	491	0.967	1.30
usecase/1 #9		<input type="checkbox"/>	12	491	0.955	1.17
usecase/1 #10		<input type="checkbox"/>	11	491	0.941	1.06
usecase/1 #11		<input checked="" type="checkbox"/>	10	491	0.923	1.02
usecase/1 #12		<input type="checkbox"/>	9	491	0.902	0.87
usecase/1 #13		<input type="checkbox"/>	8	491	0.876	0.79
usecase/1 #14		<input type="checkbox"/>	7	491	0.841	0.72
usecase/1 #15		<input type="checkbox"/>	6	491	0.804	0.59

Solution Comparison

Gene	All
CCDC36	
CEP290	
DCAF12	
DKKL1	
KDM5B	
LY6K	
MAGEA11	
ODF2	
SPAG1	
TSSK6	

Caption *

Name usecase/1 #11

Coverage 0.9226

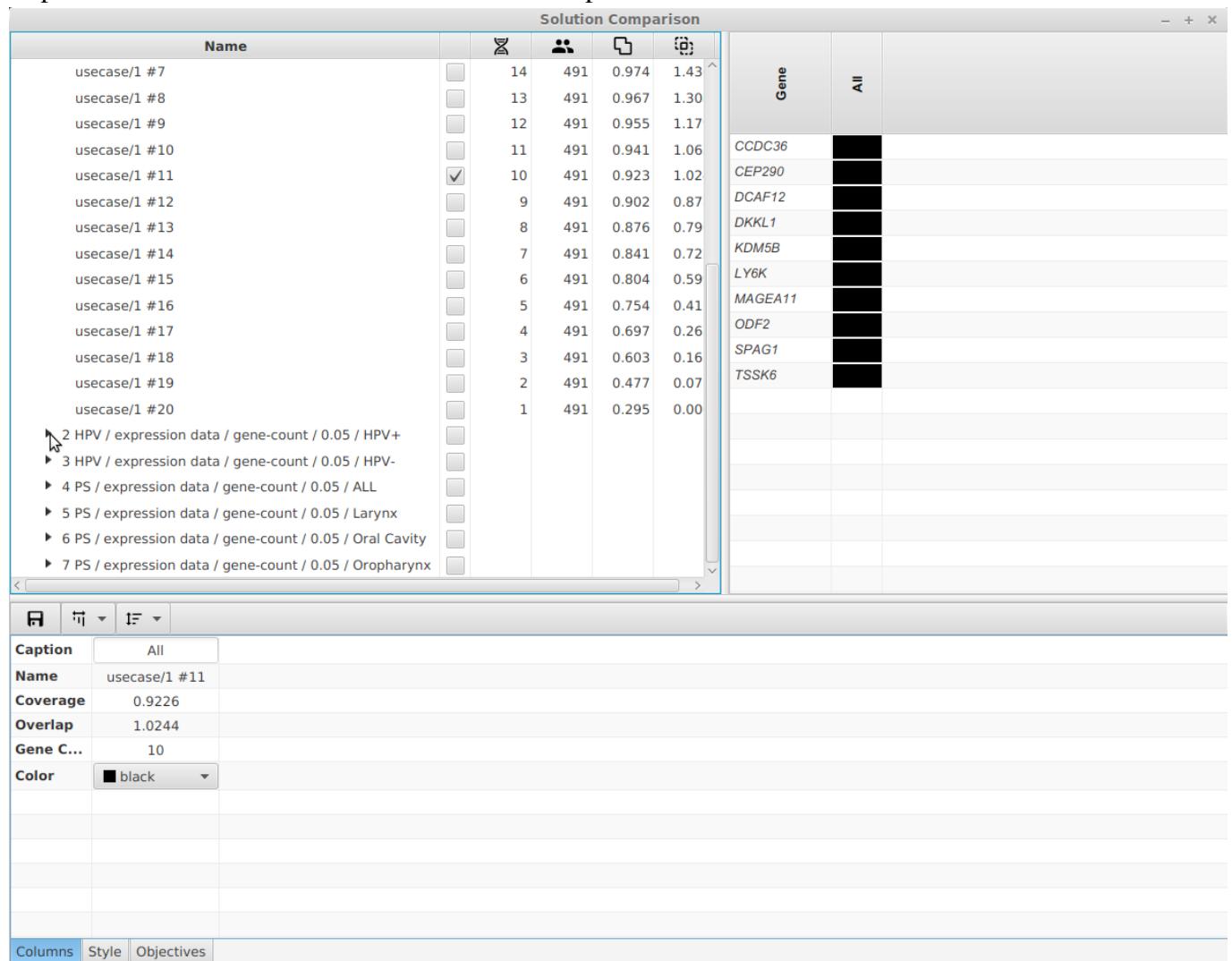
Overlap 1.0244

Gene C... 10

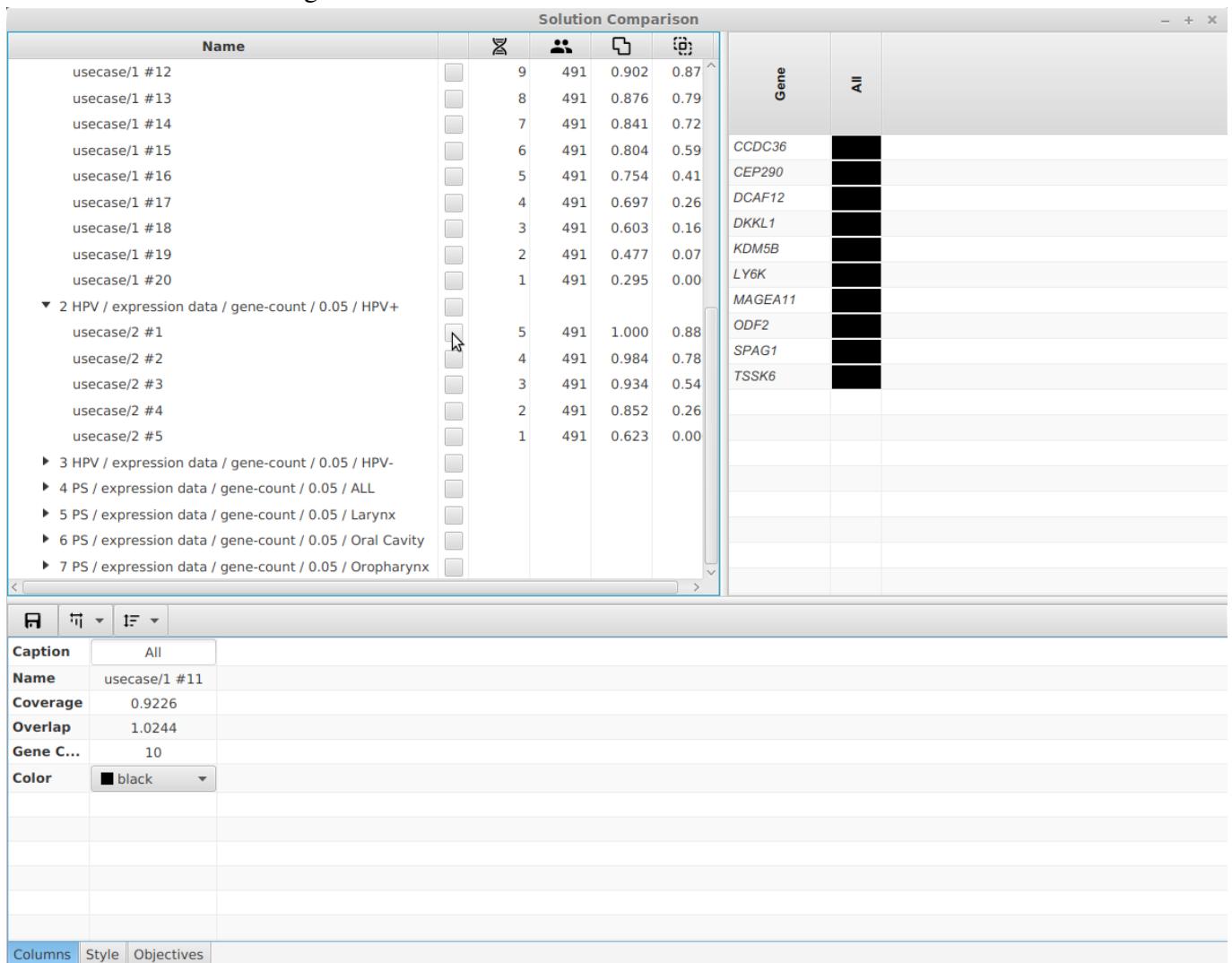
Color ■ black ▼

Columns
Style
Objectives

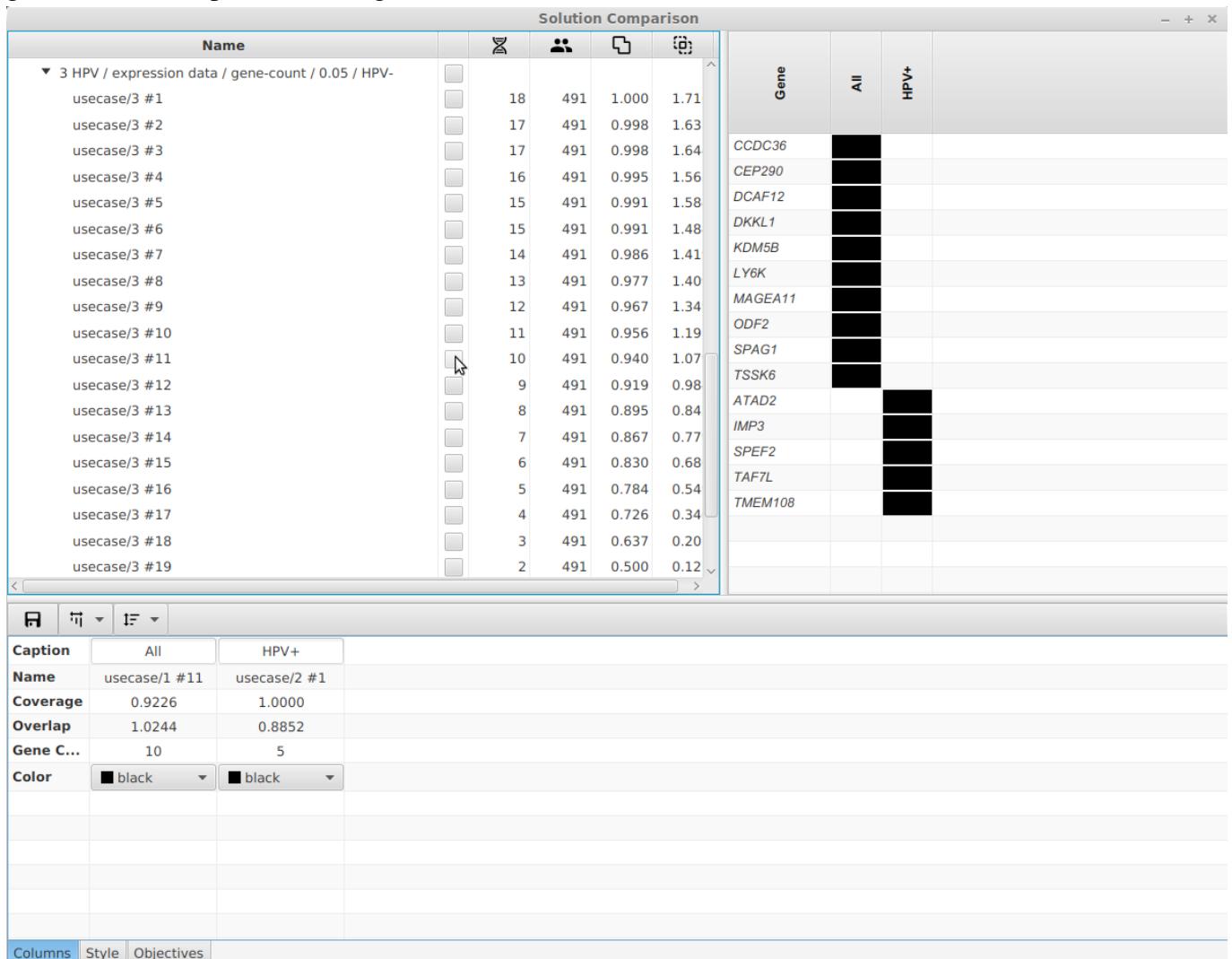
Expand the subtree of the batch for the *HPV+* samples.



Select the solution with 5 genes.



Now, both solution are shown in the comparison table. In this case, the two solutions are disjoint sets of genes. Edit the caption of the 5-gene solution to *HPV+*.



Repeat the previous steps for the 10-genes solution of the *HPV-* samples. After editing the caption, reorder the solutions in the comparison table by dragging the *HPV+* column to the right.

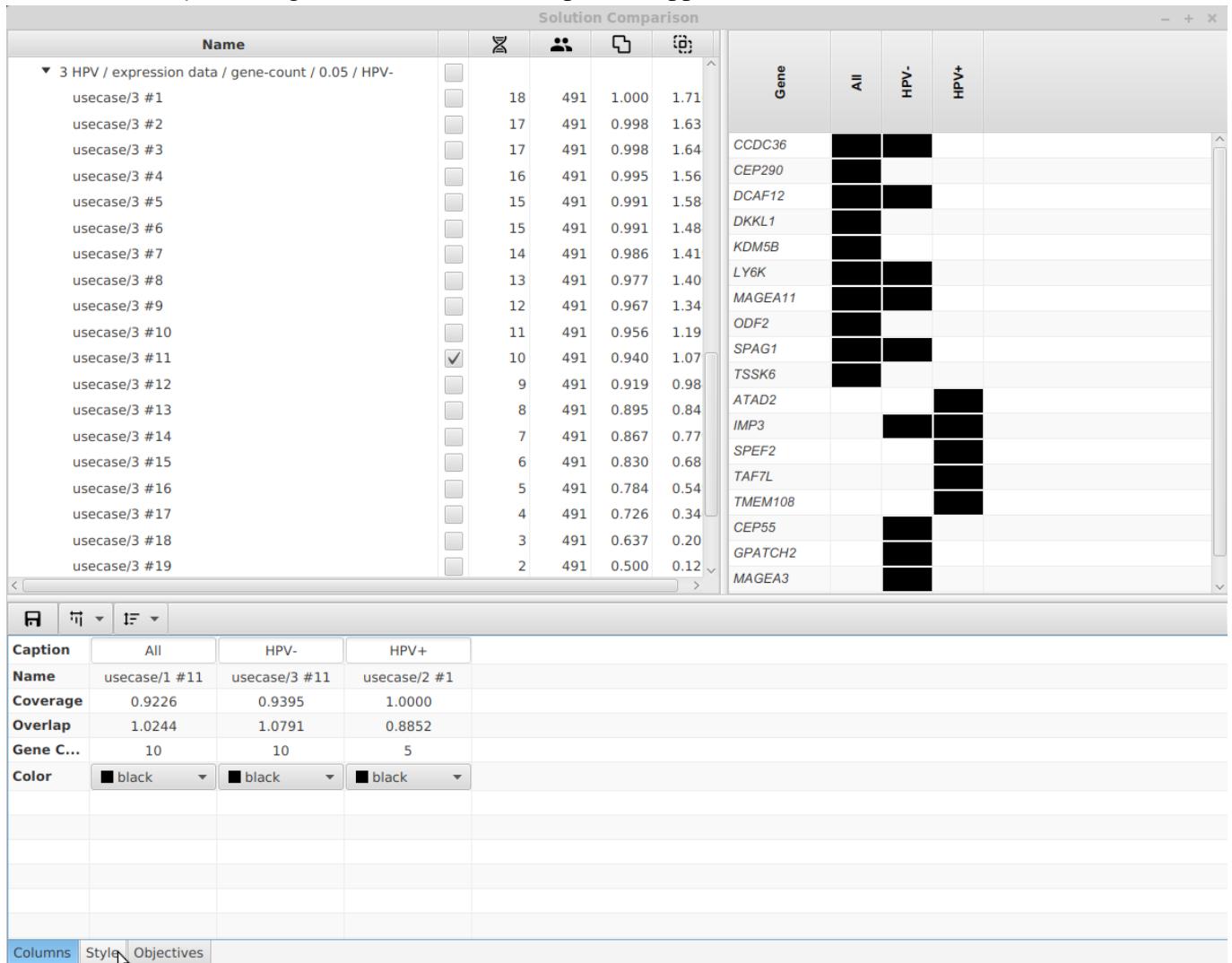
Solution Comparison

Name	All	HPV+	HPV-
usecase/3 #1	18	491	1.000
usecase/3 #2	17	491	0.998
usecase/3 #3	17	491	0.998
usecase/3 #4	16	491	0.995
usecase/3 #5	15	491	0.991
usecase/3 #6	15	491	0.991
usecase/3 #7	14	491	0.986
usecase/3 #8	13	491	0.977
usecase/3 #9	12	491	0.967
usecase/3 #10	11	491	0.956
usecase/3 #11	10	491	0.940
usecase/3 #12	9	491	0.919
usecase/3 #13	8	491	0.895
usecase/3 #14	7	491	0.867
usecase/3 #15	6	491	0.830
usecase/3 #16	5	491	0.784
usecase/3 #17	4	491	0.726
usecase/3 #18	3	491	0.637
usecase/3 #19	2	491	0.500

Caption	All	HPV+	HPV-
Name	usecase/1 #11	usecase/2 #1	usecase/3 #11
Coverage	0.9226	1.0000	0.9395
Overlap	1.0244	0.8852	1.0791
Gene C...	10	5	10
Color	black	black	black

Columns Style Objectives

Switch to the *Style* settings to customize the comparison appearance.



Turn off the rotation of the solution captions.

Solution Comparison

Name					
3 HPV / expression data / gene-count / 0.05 / HPV-					
usecase/3 #1		18	491	1.000	1.71
usecase/3 #2		17	491	0.998	1.63
usecase/3 #3		17	491	0.998	1.64
usecase/3 #4		16	491	0.995	1.56
usecase/3 #5		15	491	0.991	1.58
usecase/3 #6		15	491	0.991	1.48
usecase/3 #7		14	491	0.986	1.41
usecase/3 #8		13	491	0.977	1.40
usecase/3 #9		12	491	0.967	1.34
usecase/3 #10		11	491	0.956	1.19
usecase/3 #11	<input checked="" type="checkbox"/>	10	491	0.940	1.07
usecase/3 #12		9	491	0.919	0.98
usecase/3 #13		8	491	0.895	0.84
usecase/3 #14		7	491	0.867	0.77
usecase/3 #15		6	491	0.830	0.68
usecase/3 #16		5	491	0.784	0.54
usecase/3 #17		4	491	0.726	0.34
usecase/3 #18		3	491	0.637	0.20
usecase/3 #19		2	491	0.500	0.12

Gene	All	HPV-	HPV+
CCDC36			
CEP290			
DCAF12			
DKKL1			
KDM5B			
LY6K			
MAGEA11			
ODF2			
SPAG1			
TSSK6			
ATAD2			
IMP3			
SPEF2			
TAF7L			
TMEM108			
CEP55			
GPATCH2			
MAGEA3			

General

column captions: rotate

header height: 100

row height: 25

Gene Column

column width: 100

label: Gene

label font: Gene

gene font: TP53

Solution Columns

column width: 50

caption font: Solution 1

Objective Rows

label font: Coverage (all)

value font: 27.1%

Columns Style Objectives

Adjust table header height, gene column width and solution column width. Then, switch to the *Objectives* settings.

The figure shows a software interface titled "Solution Comparison". The main window displays a grid of data. The columns are labeled "Name", "Gene", "All", "HPV-", and "HPV+". The "Name" column lists various use cases from "usecase/3 #1" to "usecase/3 #19". The "Gene" column lists genes such as CCDC36, CEP290, DCAF12, DKKL1, KDM5B, LY6K, MAGEA11, ODF2, SPAG1, TSSK6, ATAD2, IMP3, SPEF2, TAF7L, TMEM108, CEP55, GPATCH2, MAGEA3, and TDRD6. The "All" column contains numerical values ranging from 0.12 to 1.71. The "HPV-" and "HPV+" columns contain black bars representing binary or categorical data. Below the main grid is a configuration panel with tabs for "General", "Gene Column", "Solution Columns", and "Objective Rows". Each tab has specific settings like column widths, font styles, and color definitions.

Add a coverage column to the comparison table by selecting the snapshot based on which the coverage is calculated for the genes. Since both snapshots have identical alteration data in this example, it does not matter which snapshot is chosen.

Solution Comparison

Name					Gene	All	HPV-	HPV+
3 HPV / expression data / gene-count / 0.05 / HPV-		18	491	1.000	CCDC36			
usecase/3 #1		17	491	0.998	CEP290			
usecase/3 #2		17	491	0.998	DCAF12			
usecase/3 #3		16	491	0.995	DKKL1			
usecase/3 #4		15	491	0.991	KDM5B			
usecase/3 #5		15	491	0.991	LY6K			
usecase/3 #6		14	491	0.986	MAGEA11			
usecase/3 #7		13	491	0.977	ODF2			
usecase/3 #8		12	491	0.967	SPAG1			
usecase/3 #9		11	491	0.956	TSSK6			
usecase/3 #10		10	491	0.940	ATAD2			
usecase/3 #11	<input checked="" type="checkbox"/>	9	491	0.919	IMP3			
usecase/3 #12		8	491	0.895	SPEF2			
usecase/3 #13		7	491	0.867	TAF7L			
usecase/3 #14		6	491	0.830	TMEM108			
usecase/3 #15		5	491	0.784	CEP55			
usecase/3 #16		4	491	0.726	GPATCH2			
usecase/3 #17		3	491	0.637	MAGEA3			
usecase/3 #18		2	491	0.500	TDRD6			
usecase/3 #19								

Coverage Column

Alteration Type: expression data

Objective Rows

Measure	Samples
No measure rows added.	

data snapshot: NONE

width: NONE

label: HPV

label font: Coverage

value font: 15.3%

Columns Style Objectives

Set the coverage column width to 50 and the label to *Cov..* Then, click on the *Add* button above the *Objective Rows* table to add a new objective row.

Name					
3 HPV / expression data / gene-count / 0.05 / HPV-					
usecase/3 #1		18	491	1.000	1.71
usecase/3 #2		17	491	0.998	1.63
usecase/3 #3		17	491	0.998	1.64
usecase/3 #4		16	491	0.995	1.56
usecase/3 #5		15	491	0.991	1.58
usecase/3 #6		15	491	0.991	1.48
usecase/3 #7		14	491	0.986	1.41
usecase/3 #8		13	491	0.977	1.40
usecase/3 #9		12	491	0.967	1.34
usecase/3 #10		11	491	0.956	1.19
usecase/3 #11	<input checked="" type="checkbox"/>	10	491	0.940	1.07
usecase/3 #12		9	491	0.919	0.98
usecase/3 #13		8	491	0.895	0.84
usecase/3 #14		7	491	0.867	0.77
usecase/3 #15		6	491	0.830	0.68
usecase/3 #16		5	491	0.784	0.54
usecase/3 #17		4	491	0.726	0.34
usecase/3 #18		3	491	0.637	0.20
usecase/3 #19		2	491	0.500	0.12

Gene	All	HPV-	HPV+	Cov.
CCDC36				9.4%
CEP290				16.7%
DCAF12				17.7%
DKKL1				12.2%
KDM5B				16.1%
LY6K				29.5%
MAGEA11				19.8%
ODF2				26.1%
SPAG1				29.1%
TSSK6				18.1%
ATAD2				18.7%
IMP3				15.9%
SPEF2				13.8%
TAF7L				6.5%
TMEM108				10.8%
CEP55				14.5%
GPATCH2				20.6%
MAGEA3				20.8%
TDRD6				13.4%

Alteration Type:

Coverage Column

data snapshot:

width:

label:

label font:

value font:

Objective Rows

Measure	Samples
No measure rows added.	

Select Coverage as objective.

Solution Comparison

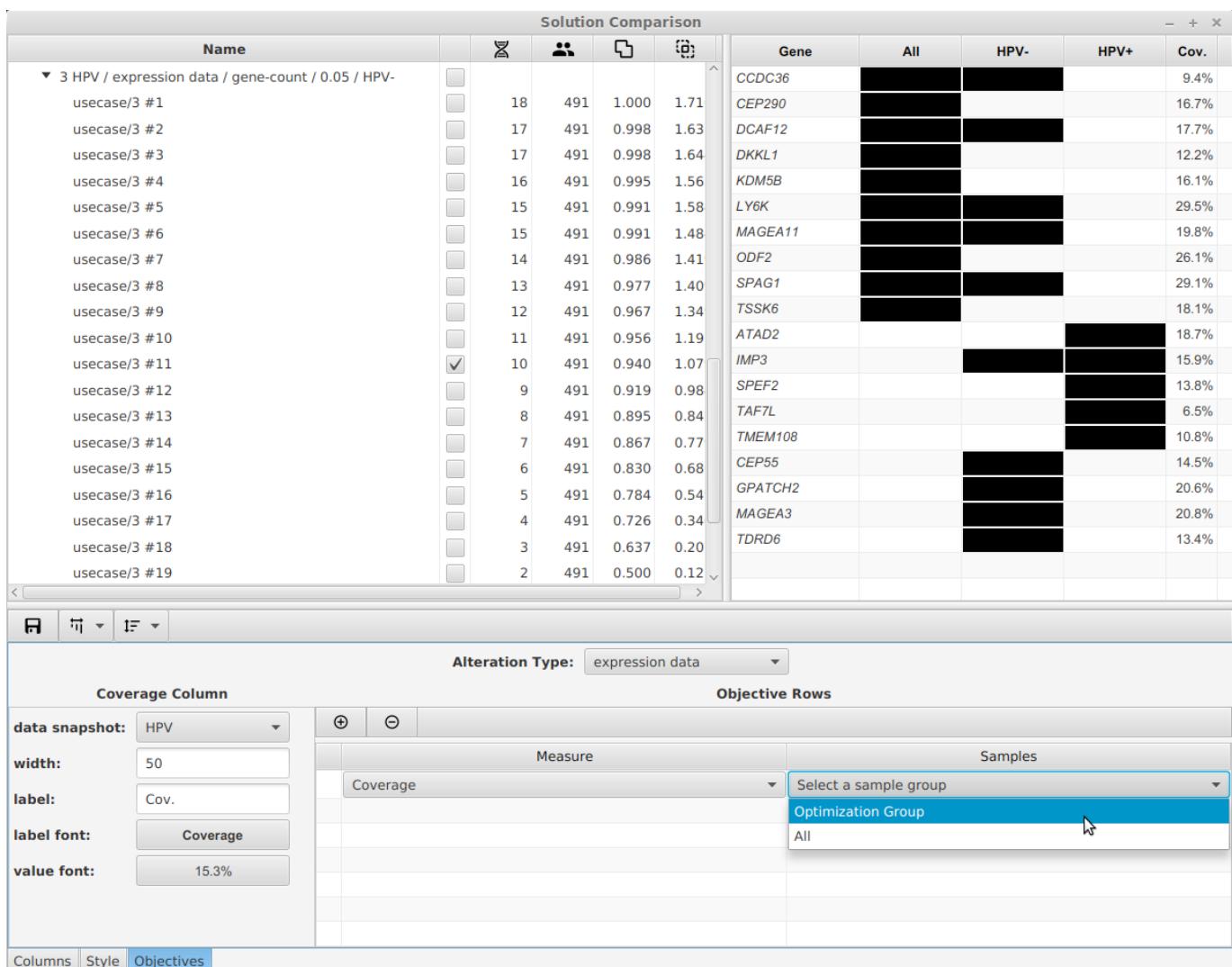
Name					Gene	All	HPV-	HPV+	Cov.	
3 HPV / expression data / gene-count / 0.05 / HPV-					CCDC36				9.4%	
usecase/3 #1		18	491	1.000	1.71	CEP290				16.7%
usecase/3 #2		17	491	0.998	1.63	DCAF12				17.7%
usecase/3 #3		17	491	0.998	1.64	DKKL1				12.2%
usecase/3 #4		16	491	0.995	1.56	KDM5B				16.1%
usecase/3 #5		15	491	0.991	1.58	LY6K				29.5%
usecase/3 #6		15	491	0.991	1.48	MAGEA11				19.8%
usecase/3 #7		14	491	0.986	1.41	ODF2				26.1%
usecase/3 #8		13	491	0.977	1.40	SPAG1				29.1%
usecase/3 #9		12	491	0.967	1.34	TSSK6				18.1%
usecase/3 #10		11	491	0.956	1.19	ATAD2				18.7%
usecase/3 #11	✓	10	491	0.940	1.07	IMP3				15.9%
usecase/3 #12		9	491	0.919	0.98	SPEF2				13.8%
usecase/3 #13		8	491	0.895	0.84	TAF7L				6.5%
usecase/3 #14		7	491	0.867	0.77	TMEM108				10.8%
usecase/3 #15		6	491	0.830	0.68	CEP55				14.5%
usecase/3 #16		5	491	0.784	0.54	GPATCH2				20.6%
usecase/3 #17		4	491	0.726	0.34	MAGEA3				20.8%
usecase/3 #18		3	491	0.637	0.20	TDRD6				13.4%
usecase/3 #19		2	491	0.500	0.12					

Alteration Type: expression data

Coverage Column		Objective Rows	
data snapshot:	HPV	Measure	Samples
width:	50	Select an objective	Select a sample group
label:	Cov.	Coverage	
label font:	Coverage	Overlap	
value font:	15.3%		

Columns Style Objectives

For the samples select *Optimization Group*. The added objectives row is now setup to show the coverage per solution column of the corresponding genes within the samples group that has been used in the optimization.



Add a second objectives row displaying the coverage on all samples. Then, sort the gene rows via the shown menu. Choose the sort order *left → right*.

Solution Comparison

Name						Gene	All	HPV-	HPV+	Cov.
3 HPV / expression data / gene-count / 0.05 / HPV-						CCDC36				9.4%
usecase/3 #1		18	491	1.000	1.71	CEP290				16.7%
usecase/3 #2		17	491	0.998	1.63	DCAF12				17.7%
usecase/3 #3		17	491	0.998	1.64	DKKL1				12.2%
usecase/3 #4		16	491	0.995	1.56	KDM5B				16.1%
usecase/3 #5		15	491	0.991	1.58	LY6K				29.5%
usecase/3 #6		15	491	0.991	1.48	MAGEA11				19.8%
usecase/3 #7		14	491	0.986	1.41	ODF2				26.1%
usecase/3 #8		13	491	0.977	1.40	SPAG1				29.1%
usecase/3 #9		12	491	0.967	1.34	TSSK6				18.1%
usecase/3 #10		11	491	0.956	1.19	ATAD2				18.7%
usecase/3 #11	✓	10	491	0.940	1.07	IMP3				15.9%
usecase/3 #12		9	491	0.919	0.98	SPEF2				13.8%
usecase/3 #13		8	491	0.895	0.84	TAF7L				6.5%
usecase/3 #14		7	491	0.867	0.77	TMEM108				10.8%
usecase/3 #15		6	491	0.830	0.68	CEP55				14.5%
usecase/3 #16		5	491	0.784	0.54	GPATCH2				20.6%
usecase/3 #17		4	491	0.726	0.34	MAGEA3				20.8%
usecase/3 #18		3	491	0.637	0.20	TDRD6				13.4%
usecase/3 #19		2	491	0.500	0.12	Coverage (group)	92.3%	94.0%	100.0%	

Objectives

Alteration Type: expression data

Objective Rows

Measure	Samples
Coverage	Optimization Group
Coverage	All

Columns Style Objectives

Co

Lexicographical left -> right

Lexicographical right -> left

Alphabetical ascending

Alphabetical descending

data snapshot

width: 50

label: Cov.

label font: Coverage

value font: 15.3%

The gene rows are now sorted lexicographically. In a nutshell, this means that the genes included in all solutions are before the genes included only in the first two solutions and so on. This scheme is continued for the genes that are not part of the first solution where the genes with alterations in the last two solutions are before the genes that are only included in the second solution. Go back to the columns settings to add some colors to the comparison.

Solution Comparison

Name						Gene	All	HPV-	HPV+	Cov.
3 HPV / expression data / gene-count / 0.05 / HPV-						LY6K				29.5%
usecase/3 #1		18	491	1.000	1.71	SPAG1				29.1%
usecase/3 #2		17	491	0.998	1.63	MAGEA11				19.8%
usecase/3 #3		17	491	0.998	1.64	DCAF12				17.7%
usecase/3 #4		16	491	0.995	1.56	CCDC36				9.4%
usecase/3 #5		15	491	0.991	1.58	ODF2				26.1%
usecase/3 #6		15	491	0.991	1.48	TSSK6				18.1%
usecase/3 #7		14	491	0.986	1.41	CEP290				16.7%
usecase/3 #8		13	491	0.977	1.40	KDM5B				16.1%
usecase/3 #9		12	491	0.967	1.34	DKKL1				12.2%
usecase/3 #10		11	491	0.956	1.19	IMP3				15.9%
usecase/3 #11	✓	10	491	0.940	1.07	MAGEA3				20.8%
usecase/3 #12		9	491	0.919	0.98	GPATCH2				20.6%
usecase/3 #13		8	491	0.895	0.84	CEP55				14.5%
usecase/3 #14		7	491	0.867	0.77	TDRD6				13.4%
usecase/3 #15		6	491	0.830	0.68	ATAD2				18.7%
usecase/3 #16		5	491	0.784	0.54	SPEF2				13.8%
usecase/3 #17		4	491	0.726	0.34	TMEM108				10.8%
usecase/3 #18		3	491	0.637	0.20	TAF7L				6.5%
usecase/3 #19		2	491	0.500	0.12	Coverage (group)	92.3%	94.0%	100.0%	

Coverage Column

Alteration Type: expression data

Objective Rows

Coverage Column

Objective Rows

Objectives

Select a color for the *HPV-* solution column in the settings table on the bottom.

Solution Comparison									
Name					Gene	All	HPV-	HPV+	Cov.
3 HPV / expression data / gene-count / 0.05 / HPV-	<input type="checkbox"/>	18	491	1.000	LY6K				29.5%
usecase/3 #1	<input type="checkbox"/>	17	491	0.998	<i>SPAG1</i>				29.1%
usecase/3 #2	<input type="checkbox"/>	17	491	0.998	<i>MAGEA11</i>				19.8%
usecase/3 #3	<input type="checkbox"/>	16	491	0.995	<i>DCAF12</i>				17.7%
usecase/3 #4	<input type="checkbox"/>	15	491	0.991	<i>CCDC36</i>				9.4%
usecase/3 #5	<input type="checkbox"/>	15	491	0.991	<i>ODF2</i>				26.1%
usecase/3 #6	<input type="checkbox"/>	14	491	0.986	<i>TSSK6</i>				18.1%
usecase/3 #7	<input type="checkbox"/>	13	491	0.977	<i>CEP290</i>				16.7%
usecase/3 #8	<input type="checkbox"/>	12	491	0.967	<i>KDM5B</i>				16.1%
usecase/3 #9	<input type="checkbox"/>	11	491	0.956	<i>DKKL1</i>				12.2%
usecase/3 #10	<input type="checkbox"/>	10	491	0.940	<i>IMP3</i>				15.9%
usecase/3 #11	<input checked="" type="checkbox"/>	9	491	0.919	<i>MAGEA3</i>				20.8%
usecase/3 #12	<input type="checkbox"/>	8	491	0.895	<i>GPATCH2</i>				20.6%
usecase/3 #13	<input type="checkbox"/>	7	491	0.867	<i>CEP55</i>				14.5%
usecase/3 #14	<input type="checkbox"/>	6	491	0.830	<i>TDRD6</i>				13.4%
usecase/3 #15	<input type="checkbox"/>	5	491	0.784	<i>ATAD2</i>				18.7%
usecase/3 #16	<input type="checkbox"/>	4	491	0.726	<i>SPEF2</i>				13.8%
usecase/3 #17	<input type="checkbox"/>	3	491	0.637	<i>TMEM108</i>				10.8%
usecase/3 #18	<input type="checkbox"/>	2	491	0.500	<i>TAF7L</i>				6.5%
					Coverage (group)	92.3%	94.0%	100.0%	



Caption All

Name usecase/1 #11

Coverage 0.9226

Overlap 1.0244

Gene C... 10

Color █ black █ black █ black

Columns

Style

Objectives

Select a color for the *HPV+* solution.

Solution Comparison

Name			All	HPV-	HPV+	Cov.
3 HPV / expression data / gene-count / 0.05 / HPV-		18	491	1.000	1.71	29.5%
usecase/3 #1		17	491	0.998	1.63	29.1%
usecase/3 #2		17	491	0.998	1.64	19.8%
usecase/3 #3		16	491	0.995	1.56	17.7%
usecase/3 #4		15	491	0.991	1.58	9.4%
usecase/3 #5		15	491	0.991	1.48	26.1%
usecase/3 #6		14	491	0.986	1.41	18.1%
usecase/3 #7		13	491	0.977	1.40	16.7%
usecase/3 #8		12	491	0.967	1.34	16.1%
usecase/3 #9		11	491	0.956	1.19	12.2%
usecase/3 #10		10	491	0.940	1.07	15.9%
usecase/3 #11	✓	9	491	0.919	0.98	20.8%
usecase/3 #12		8	491	0.895	0.84	20.6%
usecase/3 #13		7	491	0.867	0.77	14.5%
usecase/3 #14		6	491	0.830	0.68	13.4%
usecase/3 #15		5	491	0.784	0.54	18.7%
usecase/3 #16		4	491	0.726	0.34	13.8%
usecase/3 #17		3	491	0.637	0.20	10.8%
usecase/3 #18						TAF7L
usecase/3 #19						6.5%
		91	0.500	0.12		Coverage (group)
						92.3%
						94.0%
						100.0%

Color palette:

black	#6680e	black
-------	--------	-------

Legend:

All	HPV-
usecase/1 #11	usecase/3 #11
0.9226	0.9395
1.0244	1.0791
10	10

Buttons: Columns, Style, Objectives

Export a graphic of the solution comparison via the shown button. This concludes the walkthrough.

Solution Comparison

Name						Gene	All	HPV-	HPV+	Cov.
3 HPV / expression data / gene-count / 0.05 / HPV-						LY6K				29.5%
usecase/3 #1		18	491	1.000	1.71	SPAG1				29.1%
usecase/3 #2		17	491	0.998	1.63	MAGEA11				19.8%
usecase/3 #3		17	491	0.998	1.64	DCAF12				17.7%
usecase/3 #4		16	491	0.995	1.56	CCDC36				9.4%
usecase/3 #5		15	491	0.991	1.58	ODF2				26.1%
usecase/3 #6		15	491	0.991	1.48	TSSK6				18.1%
usecase/3 #7		14	491	0.986	1.41	CEP290				16.7%
usecase/3 #8		13	491	0.977	1.40	KDM5B				16.1%
usecase/3 #9		12	491	0.967	1.34	DKKL1				12.2%
usecase/3 #10		11	491	0.956	1.19	IMP3				15.9%
usecase/3 #11	<input checked="" type="checkbox"/>	10	491	0.940	1.07	MAGEA3				20.8%
usecase/3 #12		9	491	0.919	0.98	GPATCH2				20.6%
usecase/3 #13		8	491	0.895	0.84	CEP55				14.5%
usecase/3 #14		7	491	0.867	0.77	TDRD6				13.4%
usecase/3 #15		6	491	0.830	0.68	ATAD2				18.7%
usecase/3 #16		5	491	0.784	0.54	SPEF2				13.8%
usecase/3 #17		4	491	0.726	0.34	TMEM108				10.8%
usecase/3 #18		3	491	0.637	0.20	TAF7L				6.5%
usecase/3 #19		2	491	0.500	0.12	Coverage (group)	92.3%	94.0%	100.0%	

Export the current table as graphic.

Name	usecase/1 #11	usecase/3 #11	usecase/2 #1
Coverage	0.9226	0.9395	1.0000
Overlap	1.0244	1.0791	0.8852
Gene C...	10	10	5
Color	black	#6680e	#80b38

Columns Style Objectives