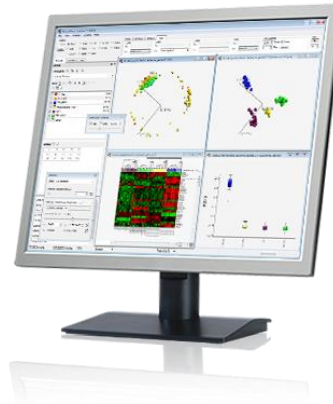


Qlucore Omics Explorer



Training

Making Sense of the Data: Visualization, Pathway and Enrichment Analysis

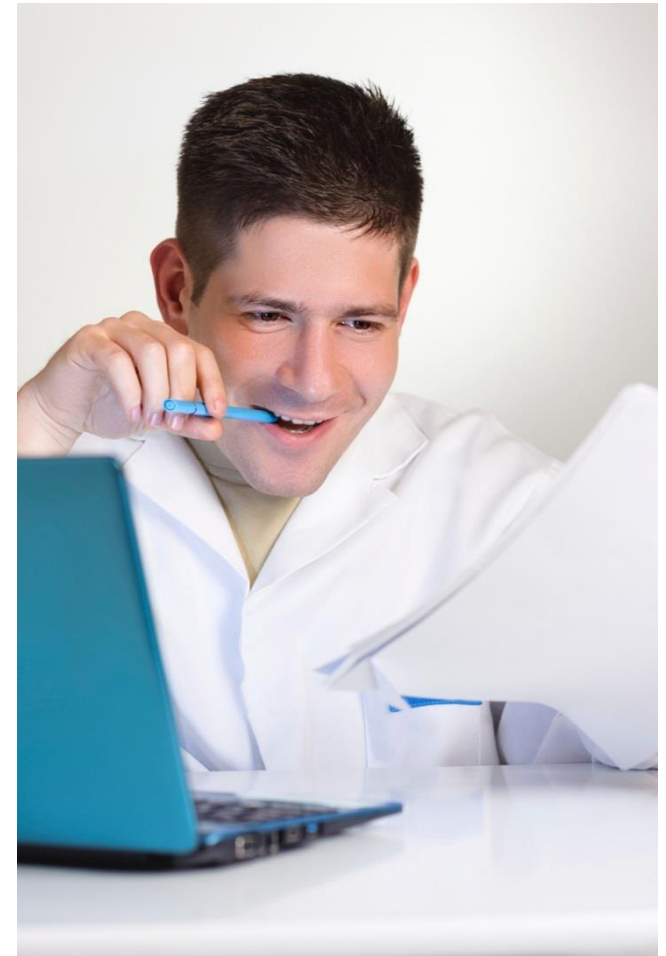
Carl Ivarsson
Qlucore

Prerequisites

The Training includes exercises on basic features and functionality in Qlucore Omics Explorer.

The training is intended for new users and does not require that you have previous experience with Qlucore Omics Explorer.

We assume that the attendants have a basic understanding of general statistical concepts.



Objectives: Qlucore Training

After the training you should be able to do the following using Qlucore Omics Explorer:

- Import data and annotations (from various sources including RNA-seq)
- Visualize data with different plot types (PCA, heat map, bar, box...)
- Identify discriminating variables using basic statistical tests
- Use Gene Set Enrichment analysis (GSEA)
- Use visualization to enhance analysis and interpret results
- Explore large data sets - find structure, patterns and subclusters in data
- Export variable lists and images

Background



Math. Sciences

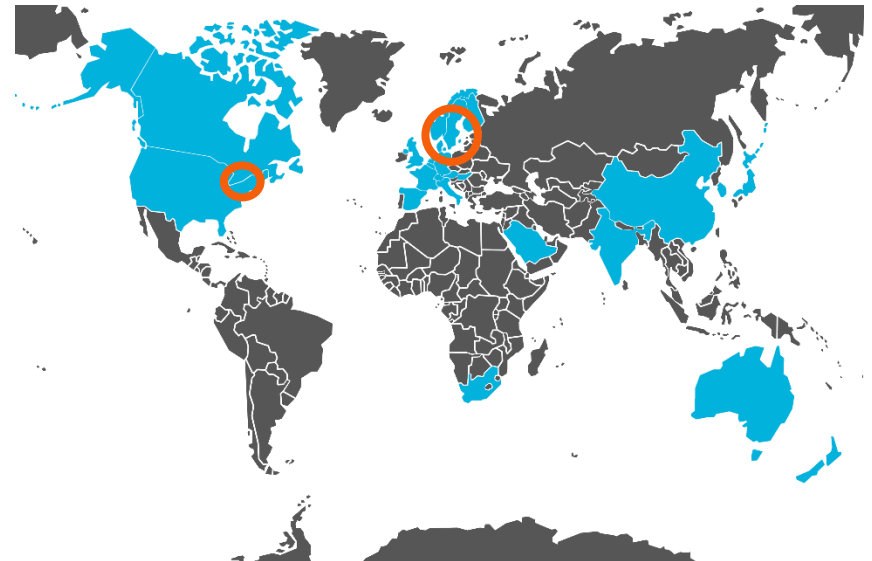


Prof. Magnus Fontes

Clinical Genetics



Prof. Thoas Fioretos



QLUCORE[®]

References and Front Covers

Qlucore "For me, one of the most compelling reasons for choosing Qlucore's Omics Explorer for the Human Protein Atlas program was its simplicity" *Mathias Uhlen, Professor, Royal Institute of Technology, Sweden*

HOME COMPANY PRODUCTS CONTACT SUPPORT DOWNLOADS LOGIN

OVERVIEW FEATURES VERSION 2.3 Q&A REFERENCES DOCUMENTATION

REFERENCES

Qlucore Omics Explorer has been used in:

JK Kilian, SY Kim, M Miettinen, C Smith, M Merino. Succinate dehydrogenase mutation underlies global epigenomic divergence in gastrointestinal stromal tumor. *Cancer Discovery*. 2013

Haglund F, Ma R, Huss M, Sulaiman L, Lu M, Nilsson IL, Hög A, Juhlin CC, Hartman J, Larsson. Evidence of a functional estrogen receptor in parathyroid adenomas. *J Clin Endocrinol Metab*. 2012

Misana K, Naicker N, Werner L, Roberts L, van Loggerenberg F, Baxter C, Passmore JA, Grobler AC, Sturm AW, Williamson C, Ronacher K, Venz G, Abdool Karim SS. Symptomatic vaginal discharge is a poor predictor of sexually transmitted infections and genital tract inflammation in high-risk women in South Africa. *L Infect Dis*. 2012

R Gonzalez, I Garitaonandia, T Abramkhina. Deriving dopaminergic neurons for clinical use. A practical approach. *Scientific Reports*. 2013

S Ørst-Nissen, KBV Dassing, M Rosting, C Lajer. Characterization of miRNA expression in human degenerative lumbar discs. *Connective Tissue*. 2013

D Jacobson, AR Monforte, AC Silva Ferreira. Untangling the chemistry of Port wine aging with the use of GC-FlD, multivariate statistics and network reconstruction. *Journal of Agricultural and Food Chemistry*. 2013

D Joshi, S Salehi, H Brenerton, M Arns, A Quaglia. Distinct MicroRNA profiles are associated with severity of HCV recurrence and acute cellular rejection after liver transplant. *Liver Transplantation*. 2013

+VERSION 2.3 NEWS
+HOW TO BUY
+GET THE VISUALISATION GUIDE

DATA
GENE EXPRESSION
PROTEIN ARRAY
MIRNA
DNA METHYLATION
PROTEOMICS

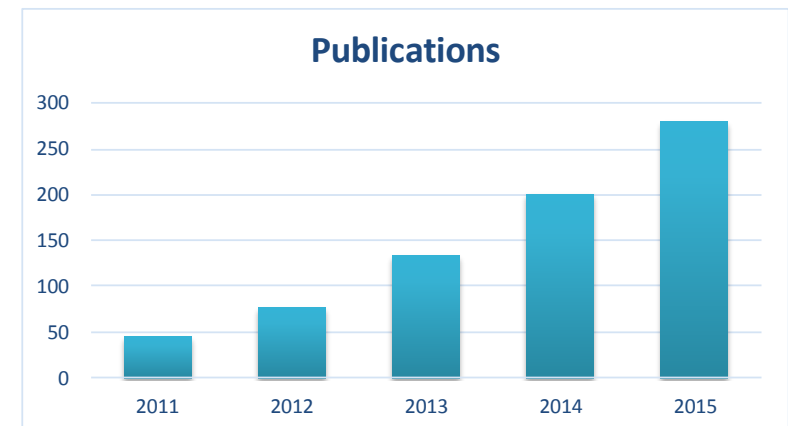
DISEASE AREA
CANCER
OBESITY
DIABETES

PERFORMANCE

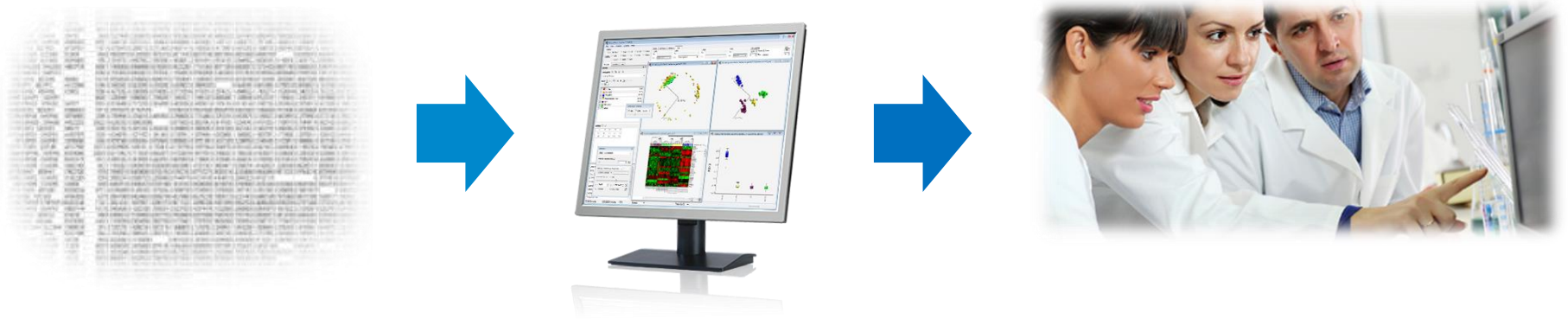
SCREENSHOTS



85% renewal rate



Qlucore Omics Explorer



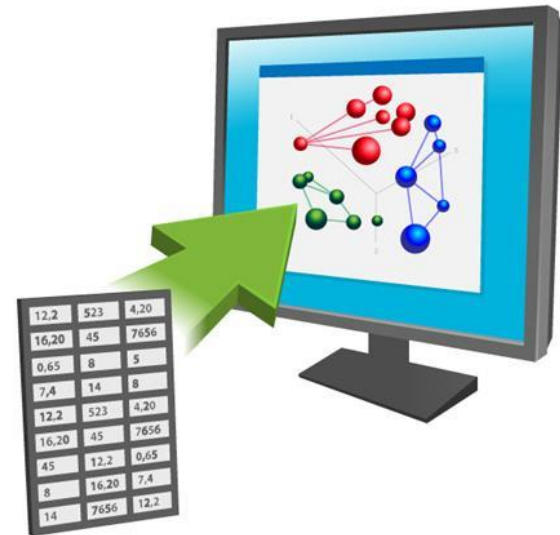
A very fast and easy to use tool to analyze and explore data
without being a statistical expert

Data types

Examples

- Gene Expression Data (array and NGS)
- DNA Methylation
- Proteomics
- Metabolomics
- Protein array data
- miRNA data
- qPCR data
- Flow Cytometry data

Any multivariate data

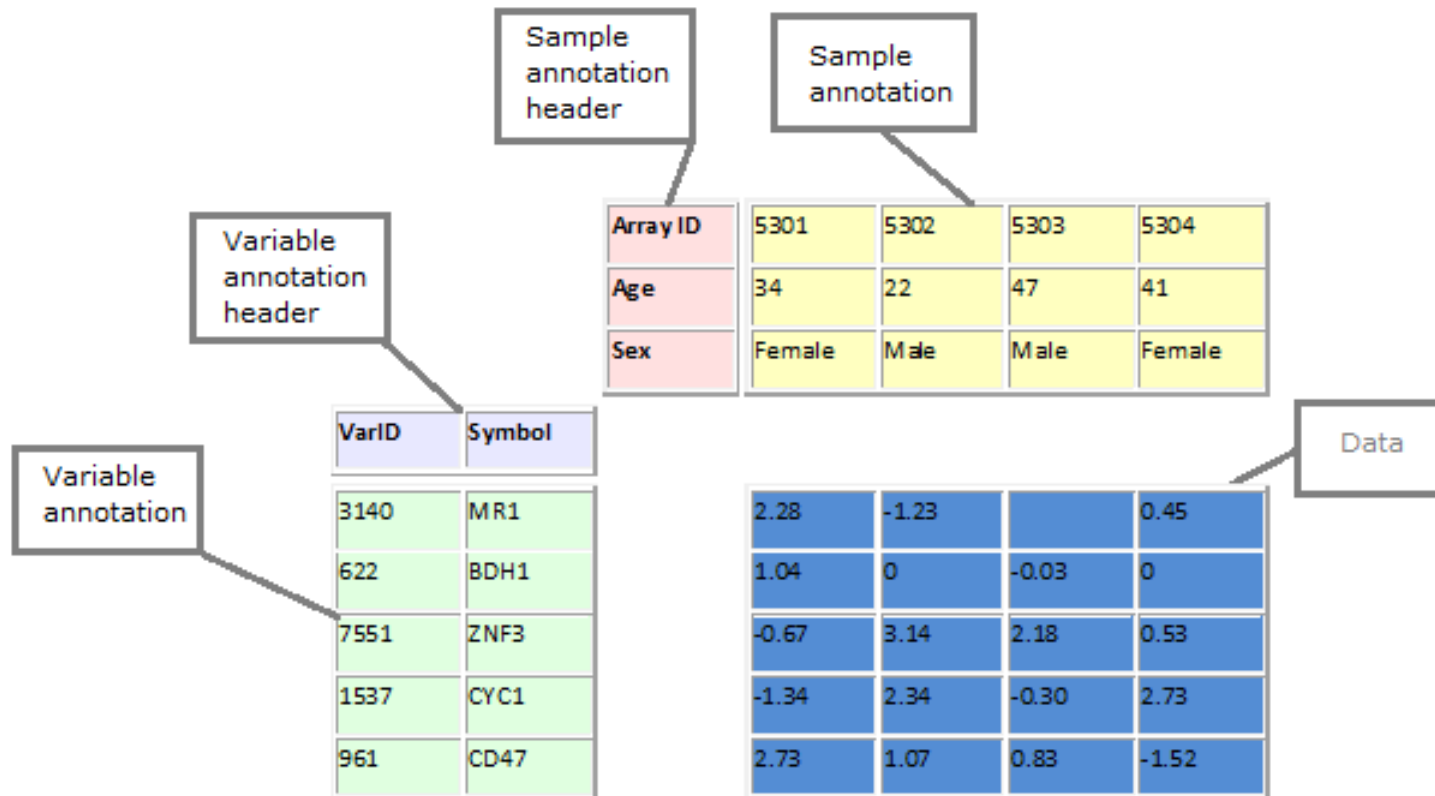


Supported File formats

- Affymetrix GeneChip compatible (.cel and .chp files)
- Agilent txt files
- BAM files (aligned for **RNA-seq**)
- GEO soft files
- Wizard (*.txt, *.csv)



Terminology



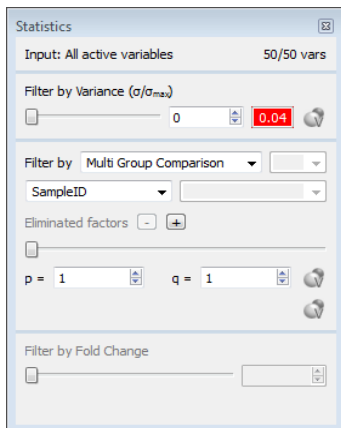
Work modes

Visualize and Explore

- Check quality
- Identify structures in the data
- Generate new hypotheses

Statistical Analysis

- t-test, ANOVA, Regression analysis.
- Open API to R
- Easy generation and export of documentation and plots

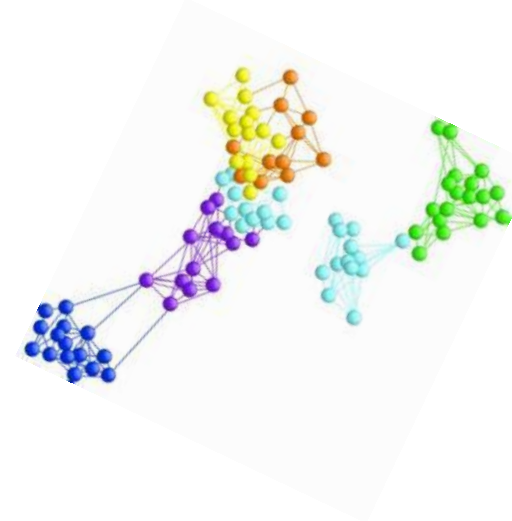
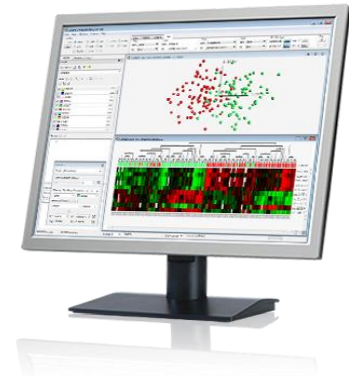


Achieve Biological Insight

- Explore annotations
- GO Browser
- GSEA – Gene Set Enrichment Analysis

Classify and Predict

- Build classifiers
- kNN, SVM, RT
- Use classifiers for prediction



Statistical tests

- **t-test (two group comparison)**
find variables that are statistically different between 2 groups.
- **F-test (ANOVA - Multigroup comparison)**
find variables that are statistically different between 2 or more groups.
- **Regression Analysis**
(used for numerical annotations e.g. time, dose, age etc)
find variables that increase or decrease over time, dose etc.
- **Open API to R**
(Welch, Limma, Mann Whitney - included)

Exercises

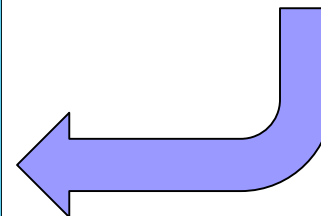
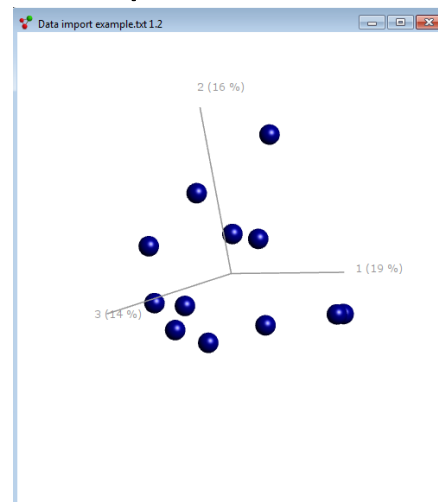
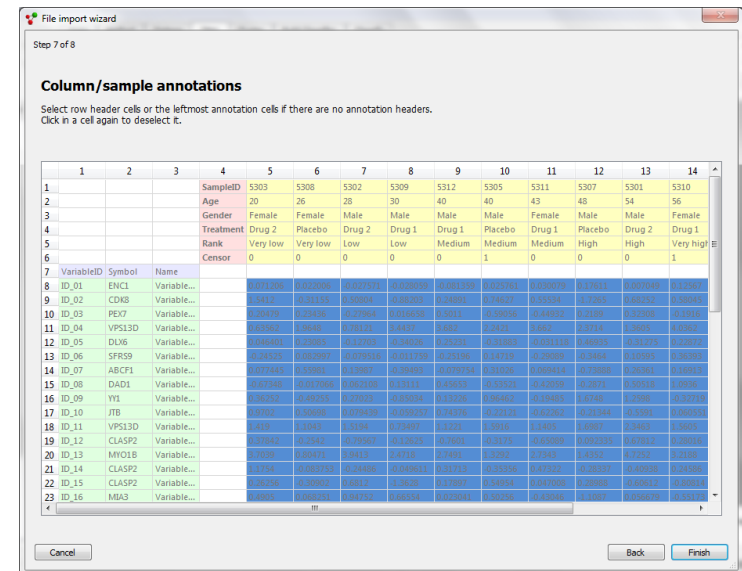
1. Import data with Wizard
2. User interface
3. Visualization with Heatmap
4. ANOVA (F-test) – Export Variable list
5. t-test – Visualization with synchronized plots
6. Box plot
7. GSEA – Gene Set Enrichment Analysis (Gene sets and pathways)
8. Compare lists – Set operations
9. Import and analyze GEO data
10. Exploratory Data Analysis using visualization
11. Import data (RNA-seq) and annotations
12. Explore and use variable annotations
13. GO Browser and set tool
14. *Linear Regression (not mandatory)*
15. *Rank Regression (not mandatory)*
16. *Bar Plot (not mandatory)*
17. *Clustering Bar Plot (not mandatory)*

First step

- Activate the training license
- Start QOE.
- Go to the license menu and select Import License file,
- Select the license file in the training folder

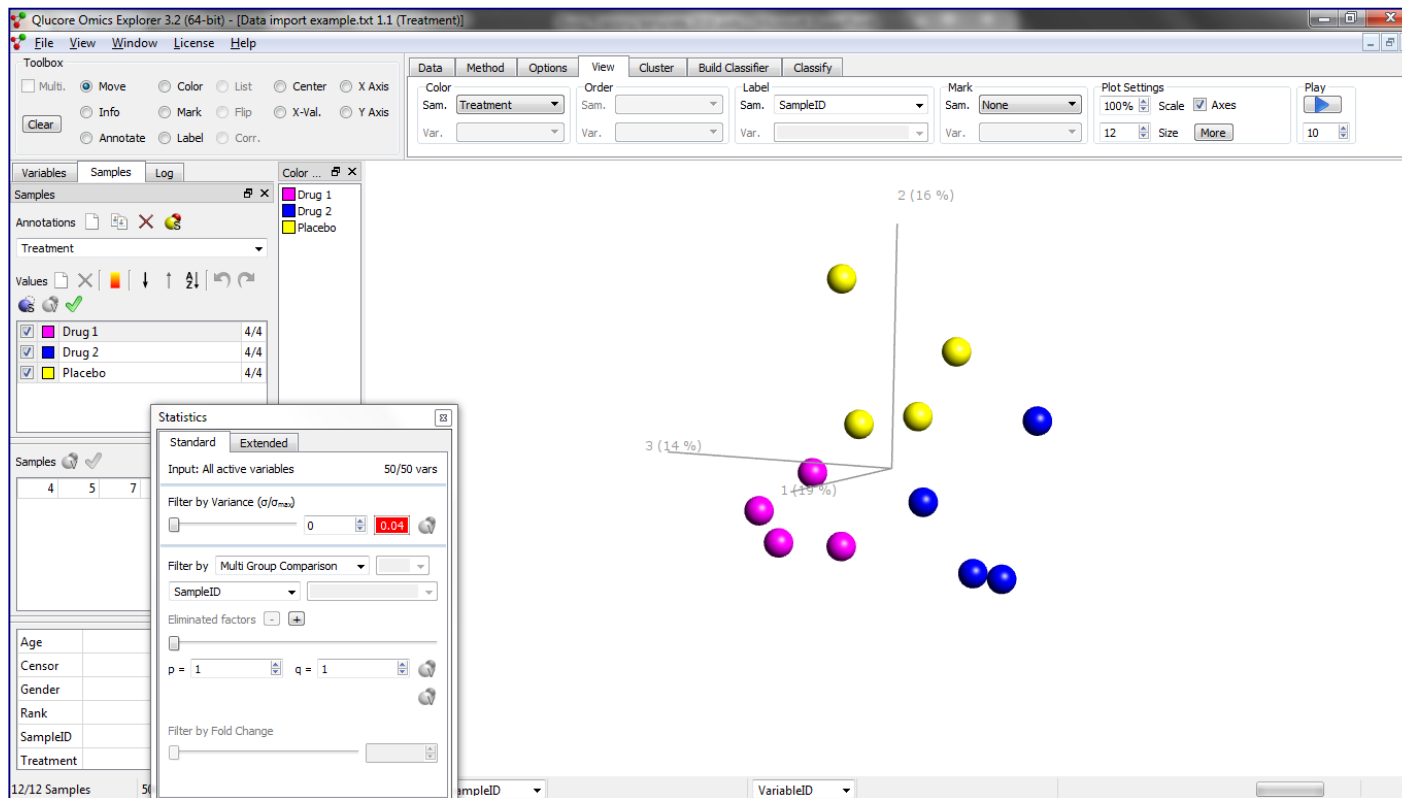
Exercise 1 – Import data with the Wizard

1. Select File/Open Wizard
2. Select the file **Data Import example.txt**
3. Select Wizard
4. Step through the Wizard
5. The dataset is loaded and a Sample PCA plot is displayed
6. Save the data set as a gedata-file (File/Save as) on your Desktop
7. Close the dataset

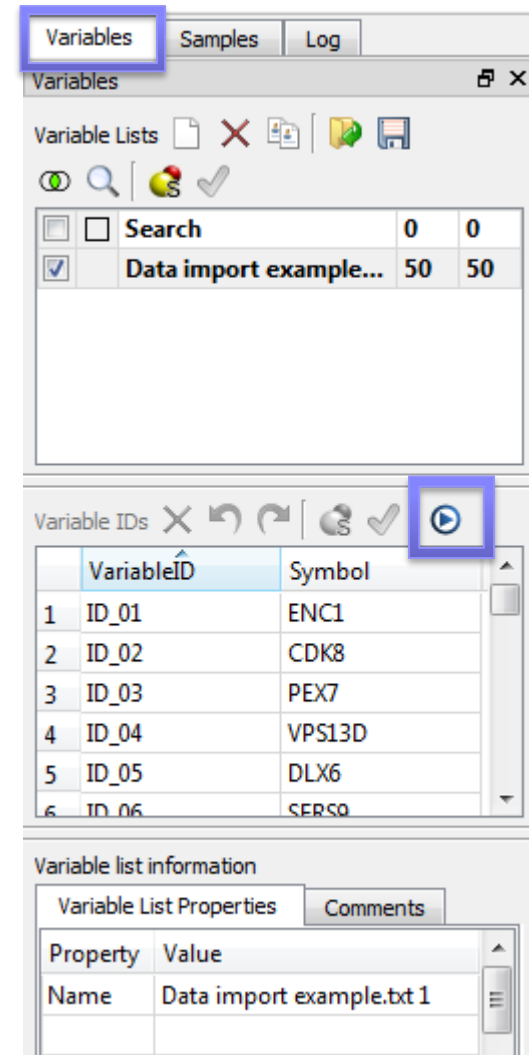
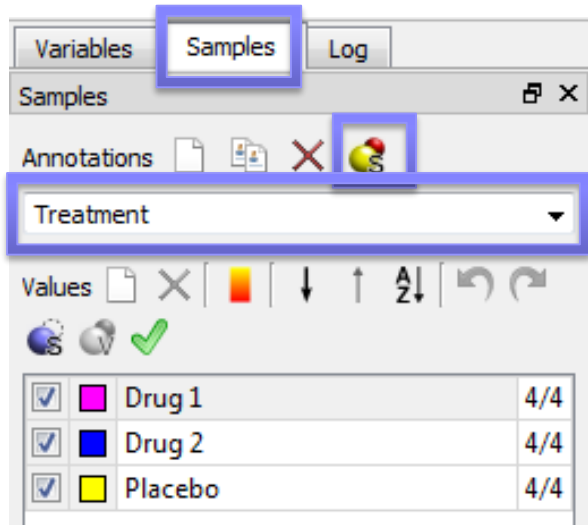
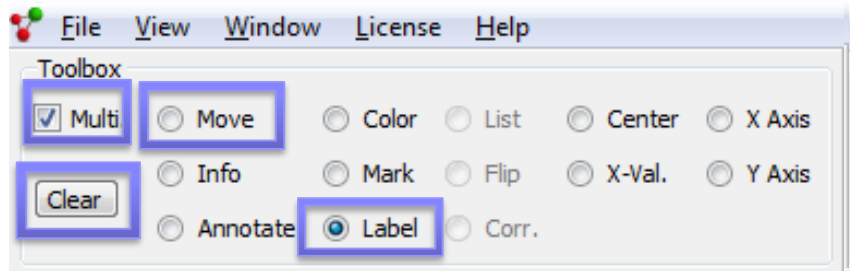


Exercise 2 – User interface

- Get familiar with the user interface

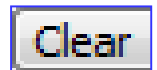
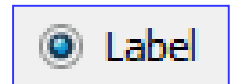
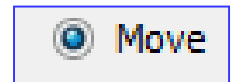


Exercise 2: Techniques



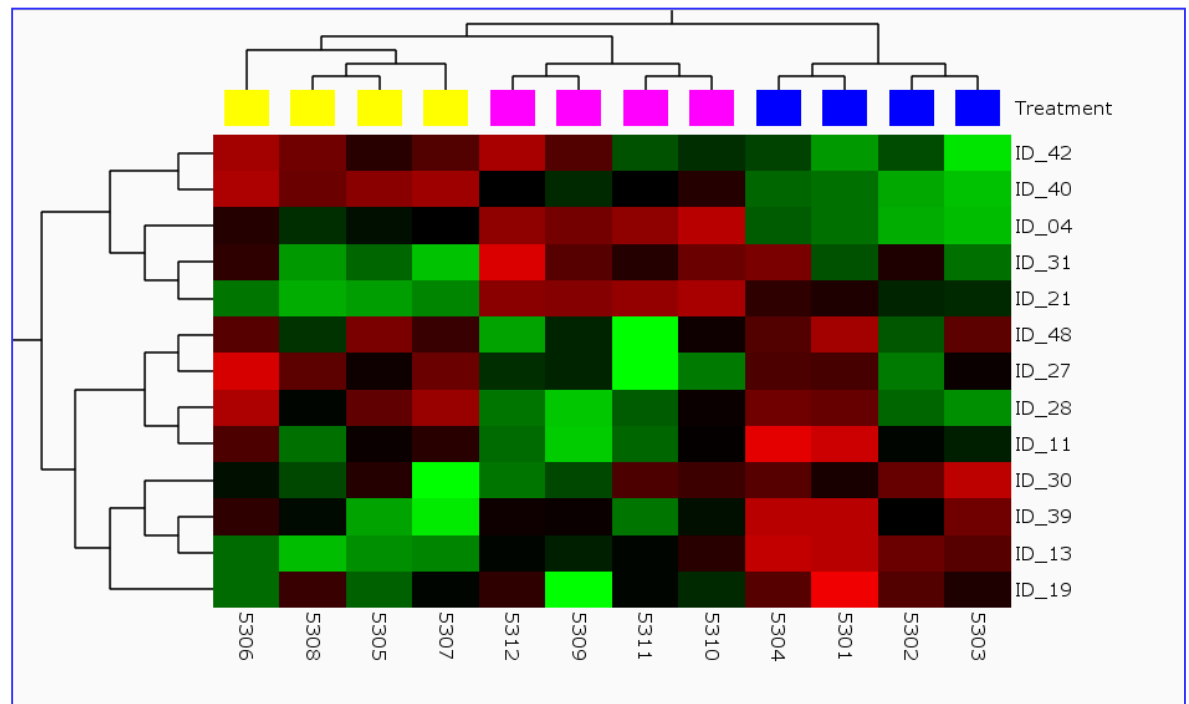
Exercise 2: Steps

- Open the Data Import Example gedata-file that you saved.
- Select Mouse tool Move (default). Rotate the PCA plot with the mouse to look at it from different angles
- Select mouse tool Label and label one/several samples by clicking on a sample/ circling several samples with the mouse (clockwise) (Note that the Multi check box must be marked)
- Undo by pressing Clear and reset the Mouse tool to Move
- Go to the Samples tab and colour the samples by Age, Gender and Treatment.
- See how the samples seem to cluster based on Treatment in the Sample PCA plot
- Go to the Variables tab.
- Click on the list and see the detailed list of the variables below.
- Add the column Symbol to the list using the column selector symbol
- Close the dataset



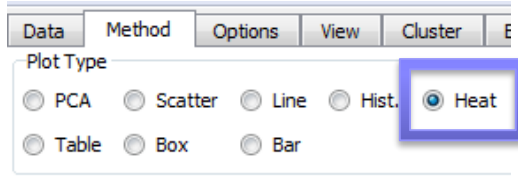
Exercise 3 – Visualization with heat map

- Heat map
- Export plot

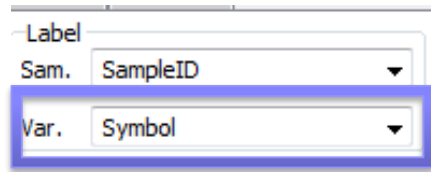


Exercise 3: Techniques

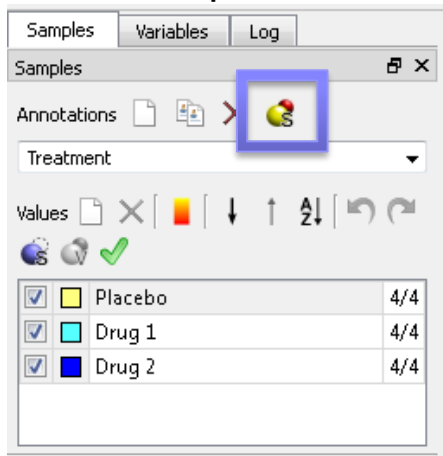
Select heat map



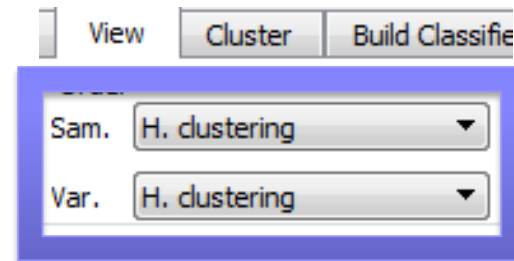
Select axis label



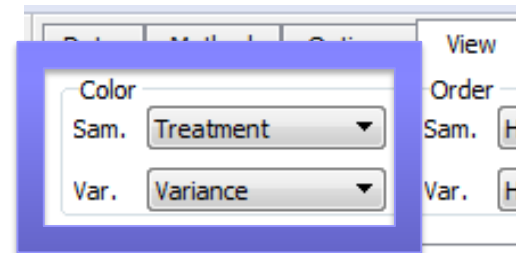
Color samples



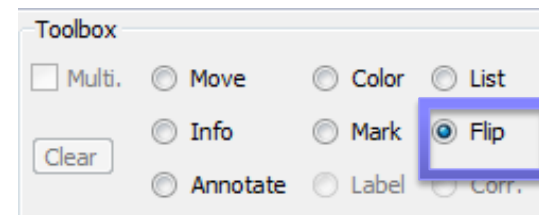
Order hierarchical clustering



Color samples and variables

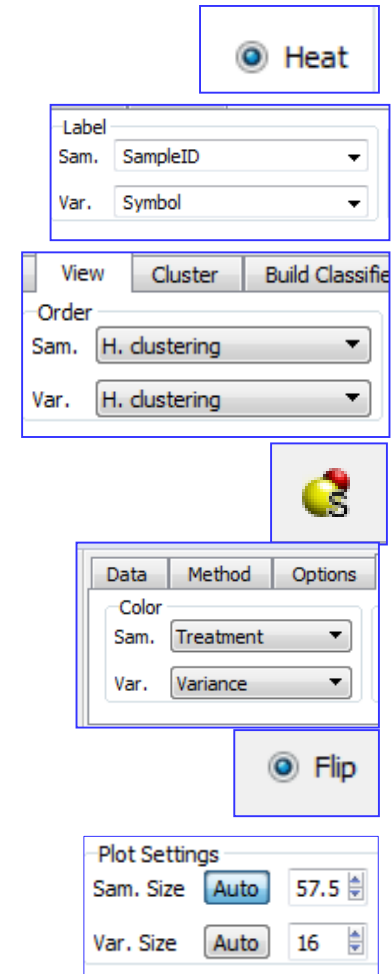


Flip the clusters



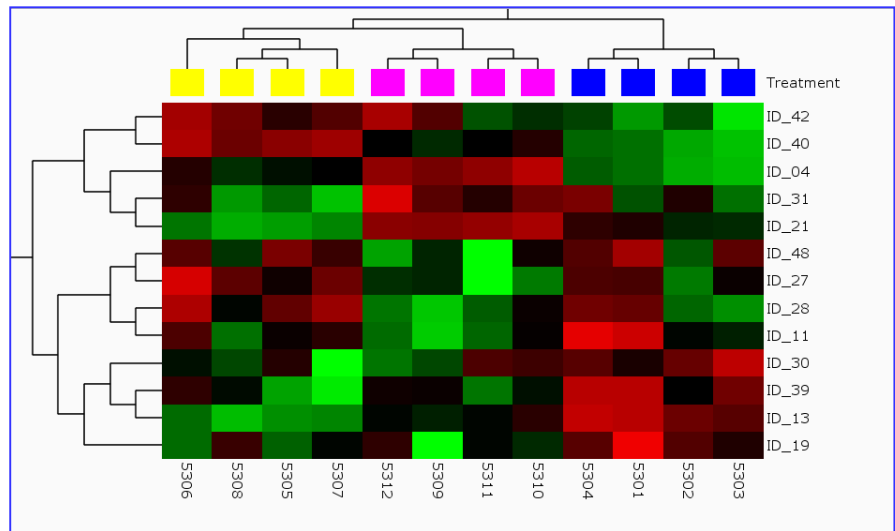
Exercise 3: Steps

1. Open the **Data Import Example** gedata-file
2. Select Heat map in the Method tab
3. Select the View Tab and change the y-axis label to Symbol (Label/Var)
4. Go to the View Tab and change Order to Hierarchical clustering for both the samples and the variables
5. Colour the samples according to the Treatment annotation.
6. Colour by additional annotations e.g. Age and Gender (View/Colour/Sam/Additional annotations)
7. Colour the variables by Variance (View/Colour/Var)
8. Use the Flip Mouse tool to flip the clusters
9. Zoom in the variables by using the arrows in View/Plot Settings. Reset by pressing auto.
10. Export the plot (File/Export/Image)
11. Open the exported file

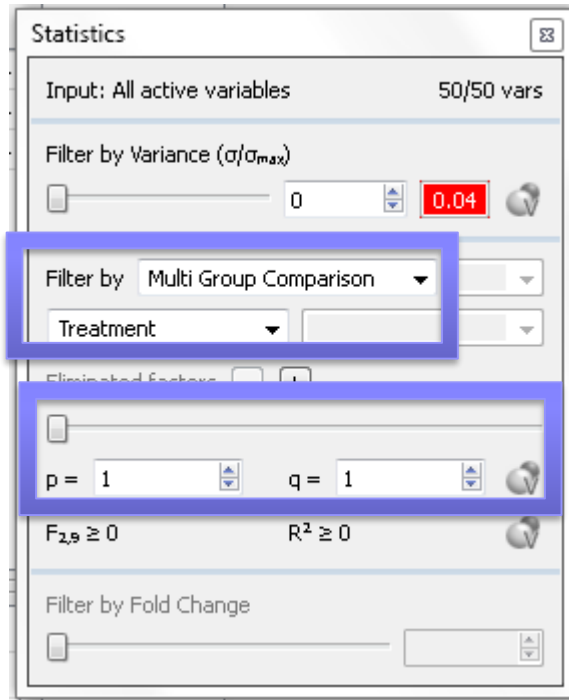


Exercise 4 - ANOVA

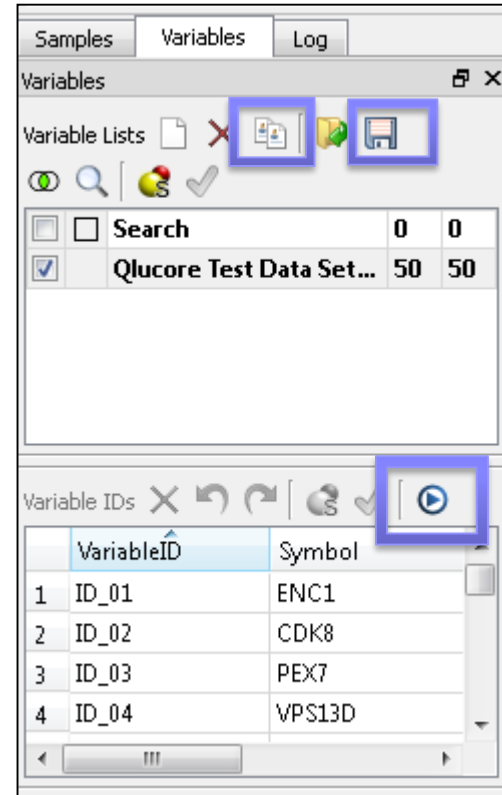
- Perform ANOVA - F-test (Multigroup comparison)
- Find discriminating variables separating the 3 Treatment groups
- Save variable list, Export list



Exercise 4: Techniques



ANOVA (Multi group comparison)
p-value slider



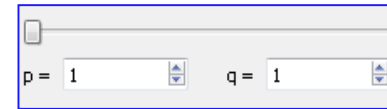
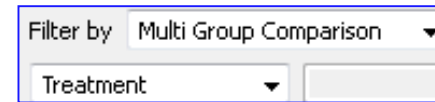
Copy the list

Add more info to the list

Save the list

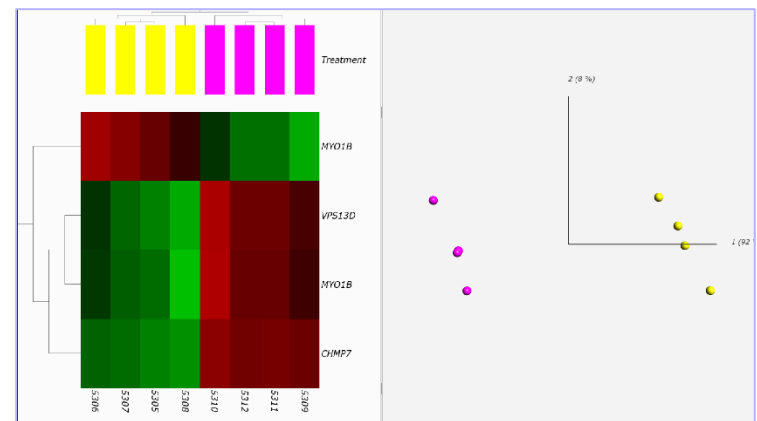
Exercise 4: Steps

1. Continue with the same data set and heat map as in the previous exercise
2. Select Multi group comparison in the Statistics dialogue.
3. Select the annotation Treatment.
4. Drag the p-value slider in the Statistics dialog. Monitor the p-values and q-values of the remaining variables.
5. Note that the heat map is instantly updated
6. Create a list of the variables by making a copy of the active variable list (Data Import Example) in the Variables tab.
7. Double click on the list to rename it.
8. Add more information to the list, click the column selector symbol. Add p-value and q-value. Also add a comment. (Mark the list and select the comment tab in the bottom part of the Window)
9. Export the list to your computer (select the Save icon) (include annotations)
10. Open the list and look at the content.

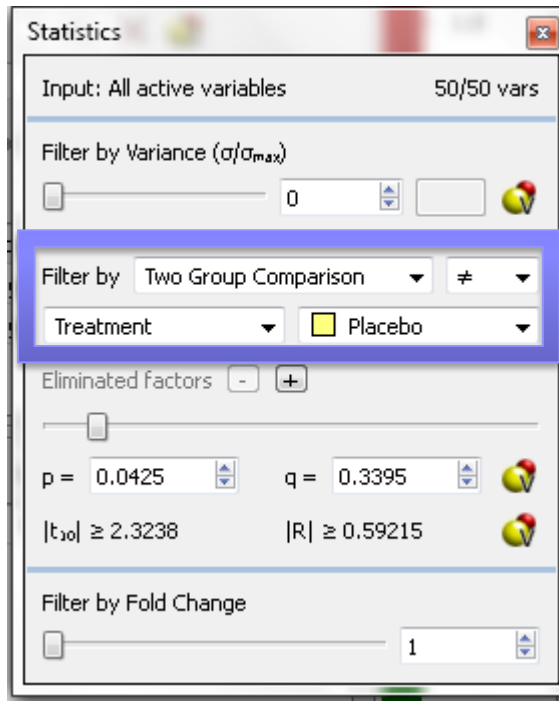


Exercise 5 – t-test

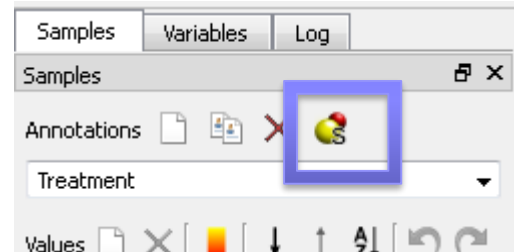
- Perform a t-test (Two Group comparison)
- Find discriminating variables separating the Placebo group from the treated group Drug 1.
- Show the samples and variables in synchronized plots - heat map and PCA
- Create a variable list from the heat map



Exercise 5: Techniques



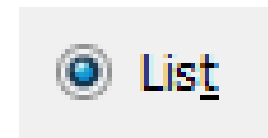
Two group comparison (t-test)



Colour the samples



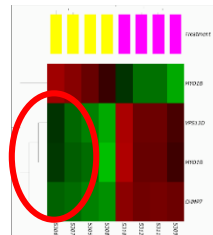
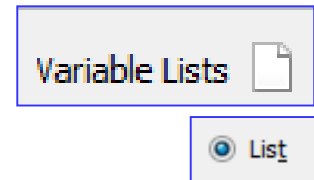
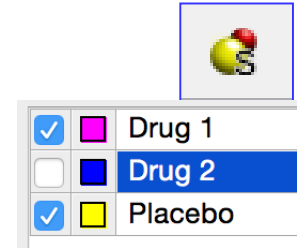
Create new variable list



The mouse List tool

Exercise 5: Steps

1. Continue with the same dataset
2. Select Window/New synchronized plot and Tile the Plots using CTRL+T or Cmd+T
3. Colour the new Sample PCA plot according to Treatment
4. Deselect Drug 2 under Treatment annotation in the sample tab - uncheck the check box.
5. Select Two Group comparison in the Statistics dialogue.
6. Select Treatment = Placebo
7. Drag the statistical slider to find discriminating variables, stop at a q-value = 0.10.
8. Observe that the same clusters can be viewed in both the sample PCA and the heatmap
9. Create New Variable list by clicking on the icon in the variables tab.
10. Change the Mouse tool to List
11. Select the down regulated genes for the placebo group with the list tool (select the node in the heatmap)
12. Rename the list and export it.

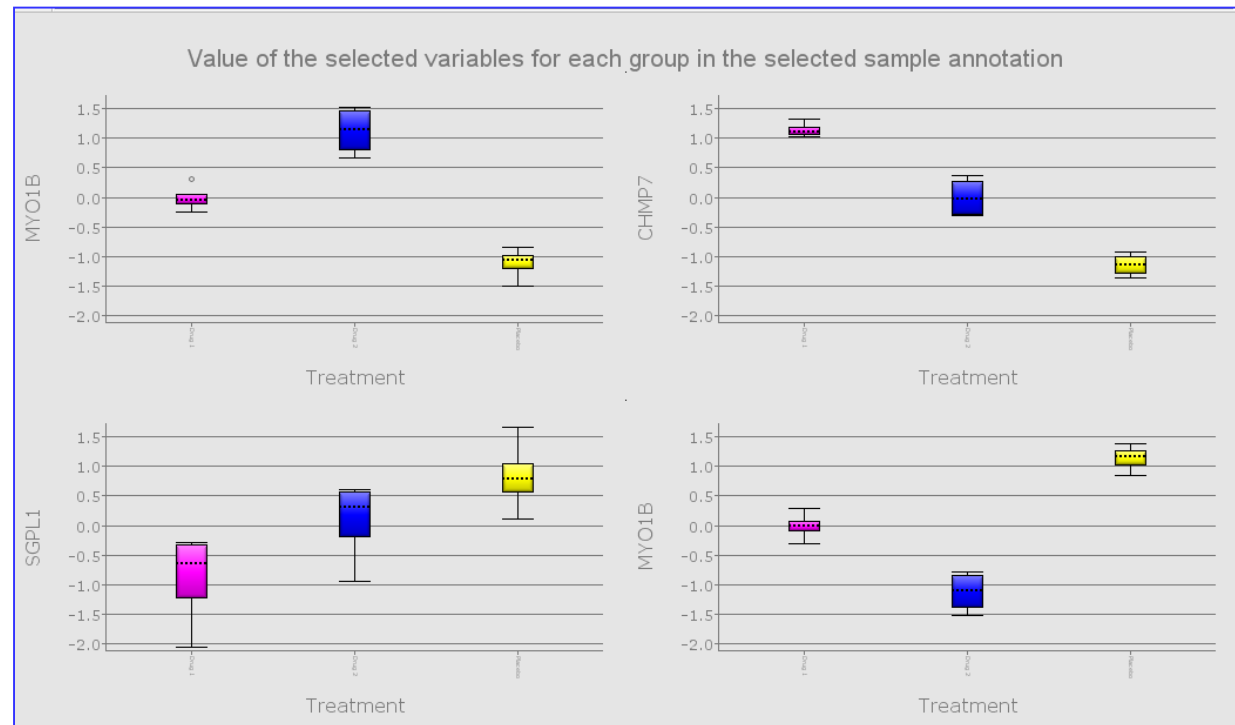


Extra Exercise – make a t-test of Placebo vs both the treated groups

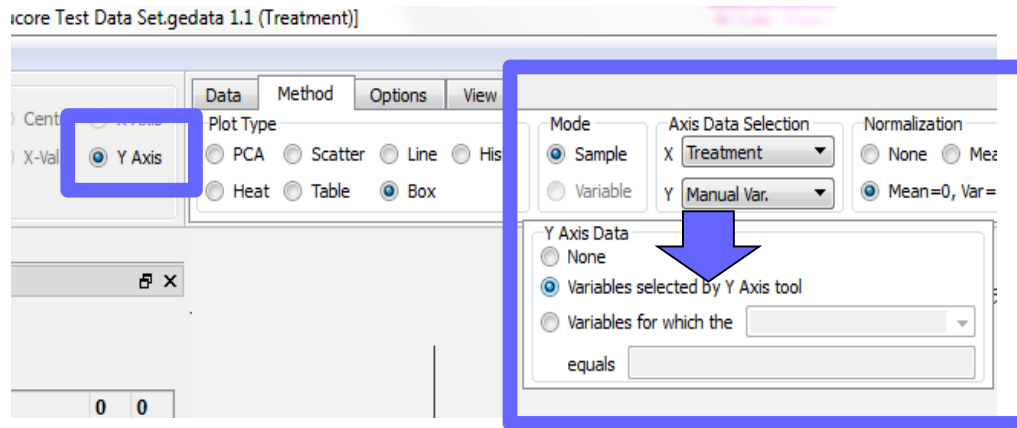
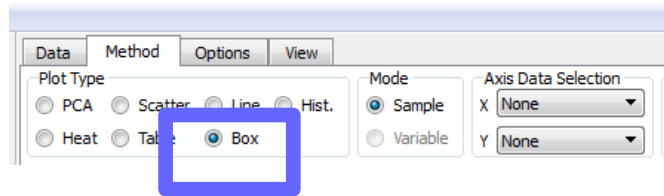
QLUCORE® (Activate all groups – select treatment Placebo and perform the test again)

Exercise 6 – Box plot

- Display discriminating variables in Box plots
- Export plots



Exercise 6: Techniques

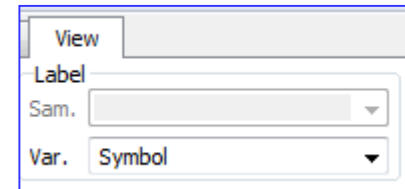
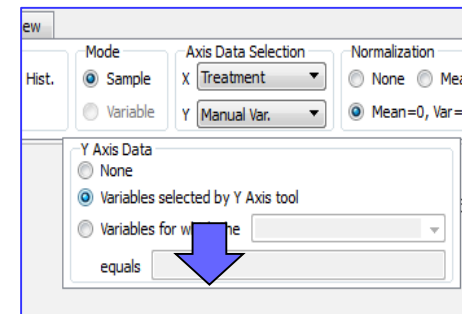
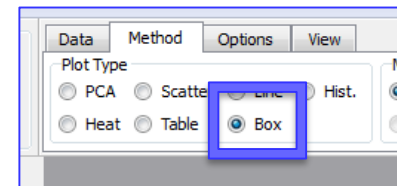
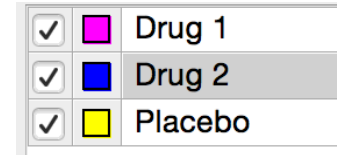


X- axis Sample annotation: Treatment
Y axis Variables selected by Y axis tool

Exercise 6: Steps

1. Continue with the same data set as in the previous exercise. Make sure to add Drug 2.
2. Change the Heat map to Box Plot in the Method Tab.
3. Select Treatment annotation on the x-axis.
4. Select “Variables selected by y-axis tool” on y-axis and select one or several variables from the active list with the mouse
5. Go to the View tab and change the y-axis label to Symbol (Label/Var)
6. Export the image – File/Export/Image.
7. Open the exported file.

Extra exercise: Make box plots of selected variables for the annotation Gender



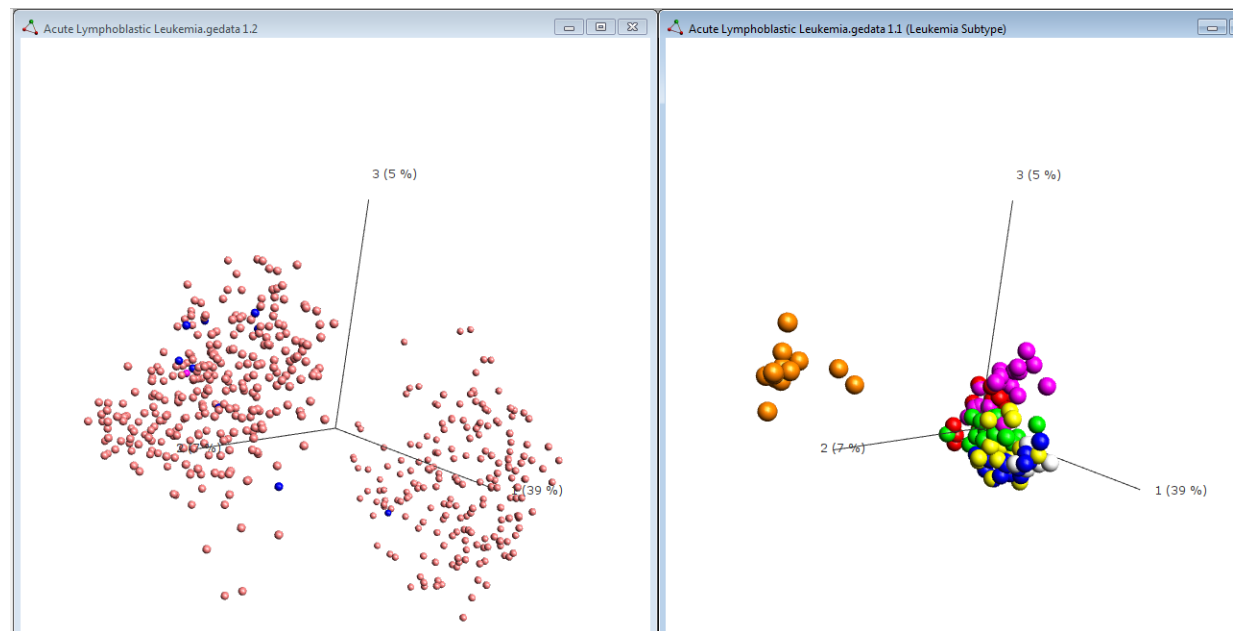
Exercise 7: Gene Set Enrichment Analysis

Match gene list from your data set with public gene sets (compilation of lists) e.g. from pathways and other biological contexts

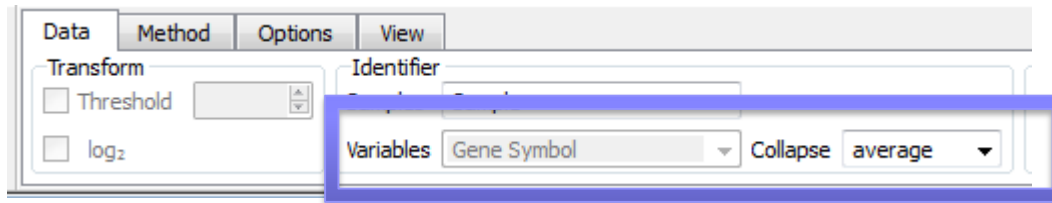


Exercise 7 - GSEA

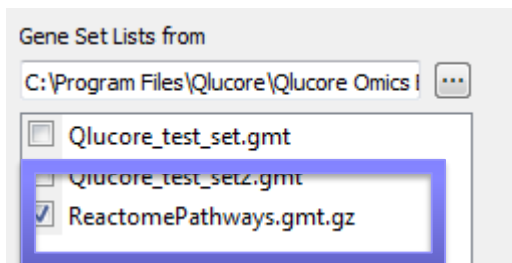
- Finding enriched gene sets/pathways for genes that discriminate T-ALL.
- Use the resulting gene sets for biological insights using visualization.



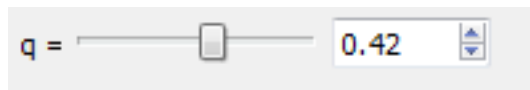
Techniques 7



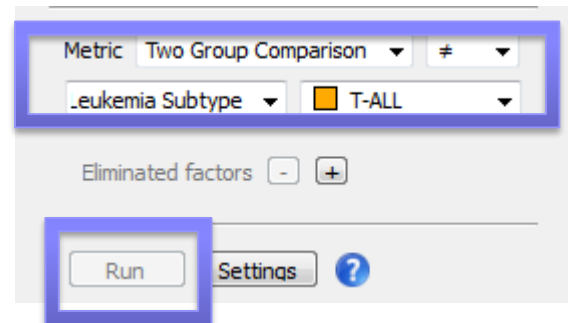
Collapse



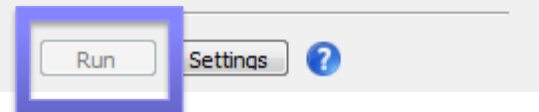
Select gene sets



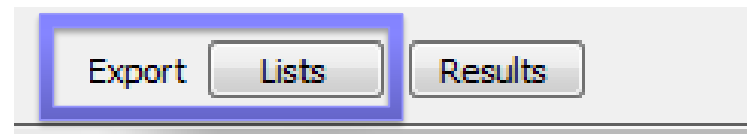
Filter on q-value



Select metrics



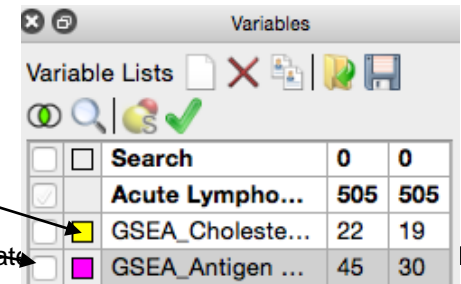
Press Run



Export

Steps 7

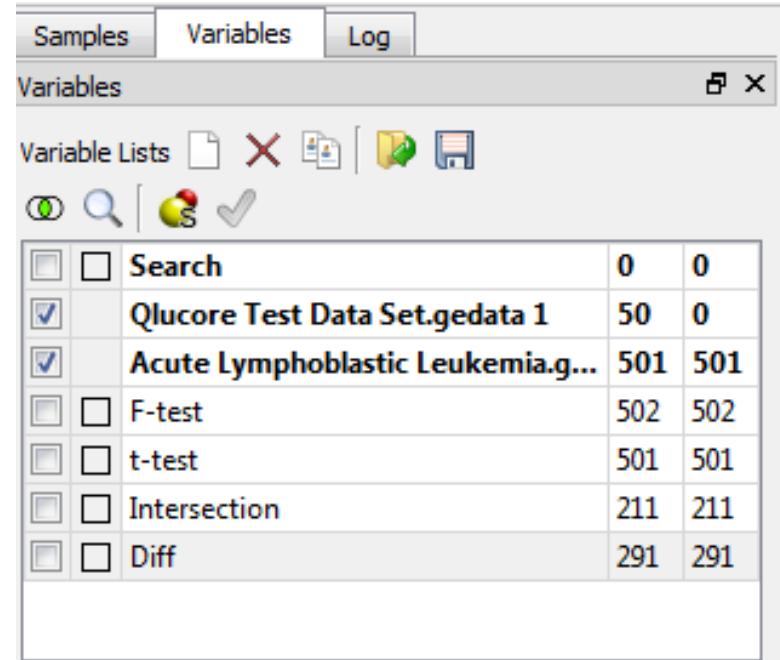
1. Load the **Acute Lymphoblastic Leukemia.gedata** from help menu.
2. In the Data tab, **Collapse** the dataset on Gene Symbol, select median.
3. Open up the GSEA workbench from the View Menu.
4. Select Reactome Pathways as input.
5. Select Metric Two Group Comparison, Leukemia Subtype and T-ALL.
6. Press Run.
7. Select the 2 pathways with the best Enrichment score and q-values and Export them.
8. Return to the normal workbench and open a synchronized variable PCA plot.
9. Perform the same test, Two Group Comparison on Leukemia Subtype and select T-ALL. Filter down to 500 genes.
10. Click the boxes next to the pathway lists in the **variable tab** to color the variable PCA plot according to these pathways.
11. Tick the leftmost box to only show the variables belonging to the selected pathway.



Variables			
Variable Lists			
<input type="checkbox"/>	Search	0	0
<input checked="" type="checkbox"/>	Acute Lympho...	505	505
<input type="checkbox"/>	GSEA_Choleste...	22	19
<input type="checkbox"/>	GSEA_Antigen ...	45	30

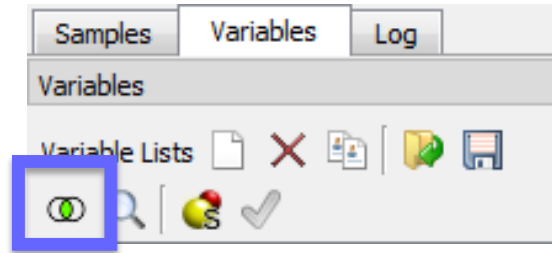
Exercise 8 – Set operations

- Compare lists
- Create intersection, unions, differences

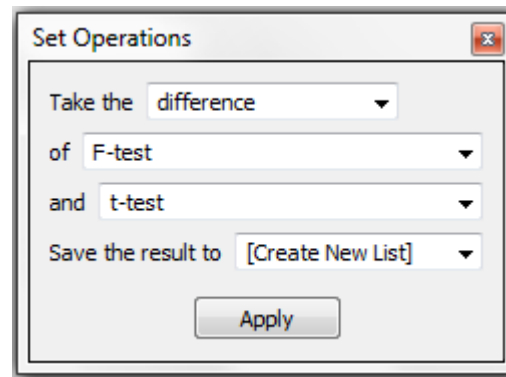


Variables			
Variable Lists			
<input type="checkbox"/>	<input type="checkbox"/>	Search	0 0
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Qlucore Test Data Set.gedata 1	50 0
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Acute Lymphoblastic Leukemia.g...	501 501
<input type="checkbox"/>	<input type="checkbox"/>	F-test	502 502
<input type="checkbox"/>	<input type="checkbox"/>	t-test	501 501
<input type="checkbox"/>	<input type="checkbox"/>	Intersection	211 211
<input type="checkbox"/>	<input type="checkbox"/>	Diff	291 291

Exercise 8: Techniques

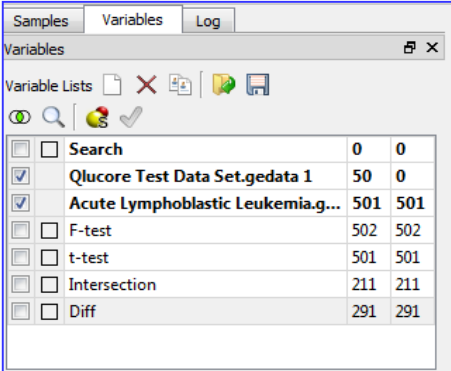


Set operations



Exercise 8: Steps

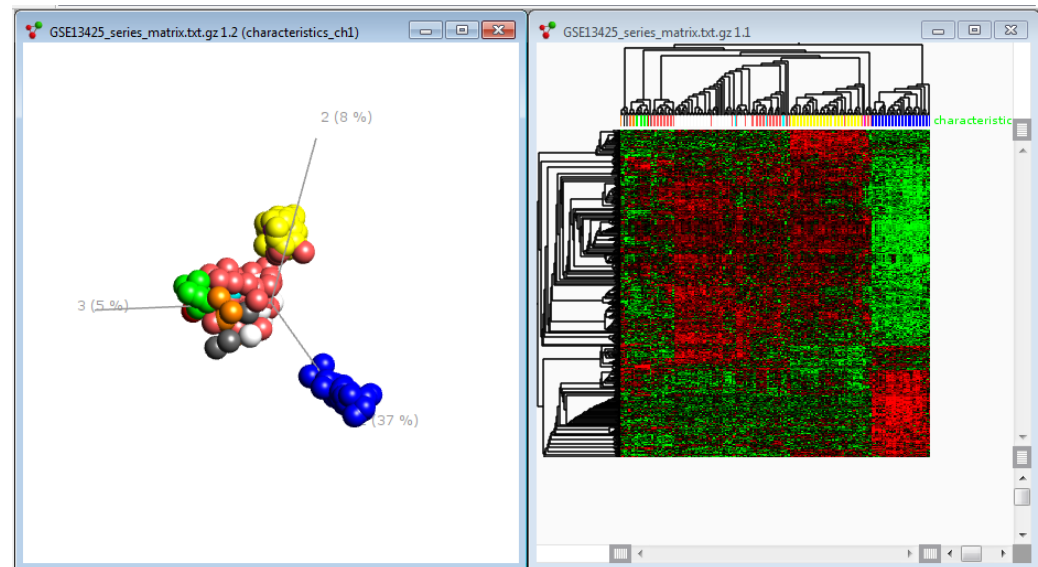
1. Load the **Leukemia dataset** from the Help menu
2. Perform a Multigroup comparison on Leukemia Subtype – use the p-value slider and filter to approx 500 variables
3. Copy the list and call it "F-test"
4. Perform a two group comparison on Leukemia subtype T-ALL and filter to app. 500 variables
5. Copy the list and call it "t-test"
6. Use the Set Operations tool and make an intersection of the lists – rename the list "Intersection"
7. Use the Set Operations tool and make a difference between F-test and t-test. Rename the list Diff.



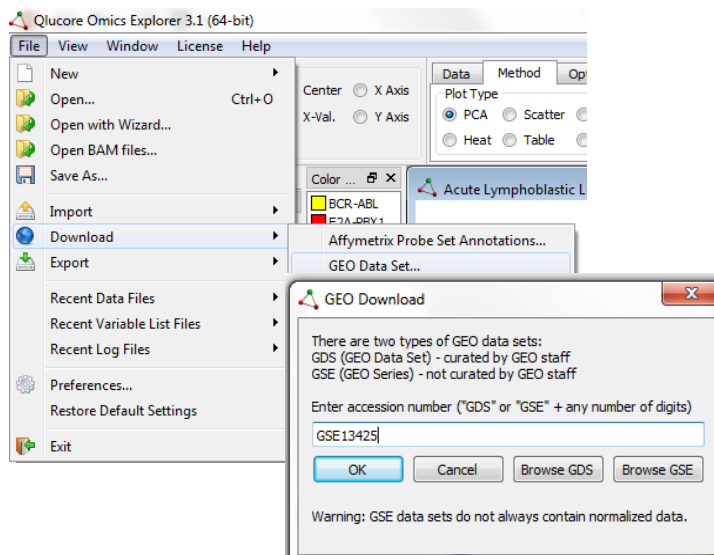
		0	0
<input checked="" type="checkbox"/>	Search	50	0
<input checked="" type="checkbox"/>	Qlucore Test Data Set.gedata 1	501	501
<input checked="" type="checkbox"/>	Acute Lymphoblastic Leukemia.g...	502	502
<input type="checkbox"/>	F-test	501	501
<input type="checkbox"/>	t-test	211	211
<input type="checkbox"/>	Intersection	291	291
<input type="checkbox"/>	Diff		

Exercise 9 – GEO data set

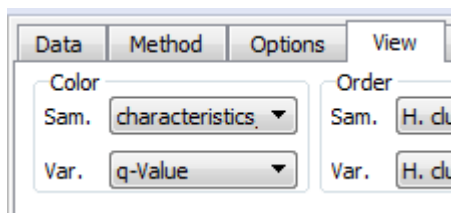
- Import data set from Gene Expression Omnibus (GEO)
- Perform ANOVA - F-test (Multigroup comparison)
- Find discriminating variables separating the different disease subtypes
- Present the result in synchronized plots



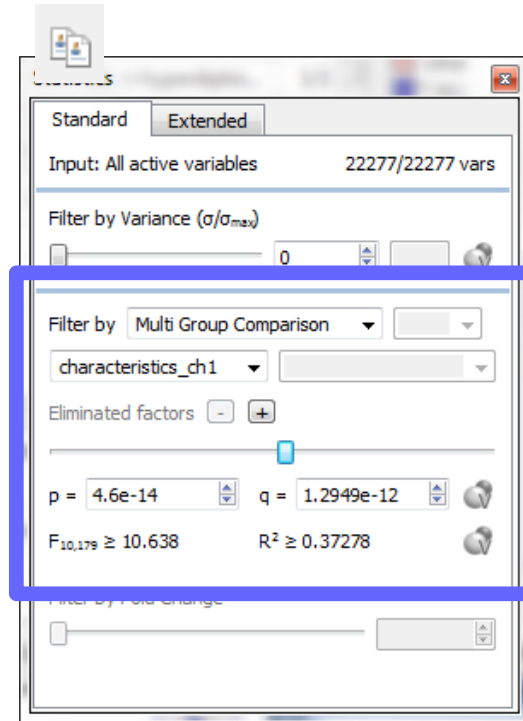
Exercise 9: Techniques



Download GEO dataset



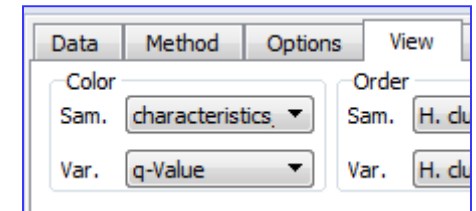
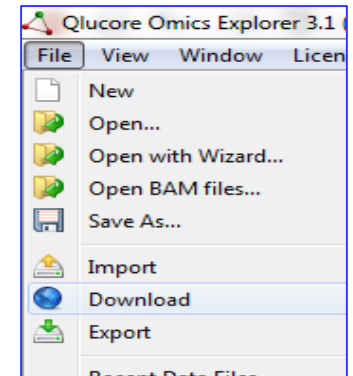
Colour samples and variables



Perform ANOVA (F-test)

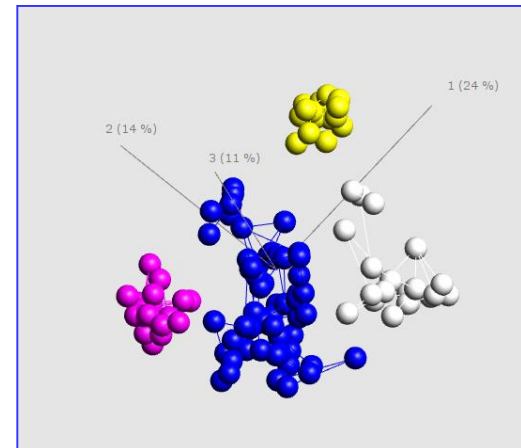
Exercise 9: Steps

1. Import a data set from GEO. Search for GSE13425 (childhood ALLs). (If no internet connection - find the data set in the training folder).
2. Colour the samples according to the annotation characteristics_ch1.
3. Select Multi group comparison in the Statistics dialog.
4. Select the annotation characteristics_ch1.
5. Drag the p-value slider in the Statistics dialog and filter to the 800 most significant variables that separates the disease subtypes.
6. Open up a new synchronized plot, and make it a heat map
7. Select hierarchical clustering for both samples and variables.
8. Colour the samples according to characteristics_ch1.
9. Show all the sample annotations in the heatmap (View/Color/Sam.)
10. Colour the variables according to q-value (View/Colour/Var.)

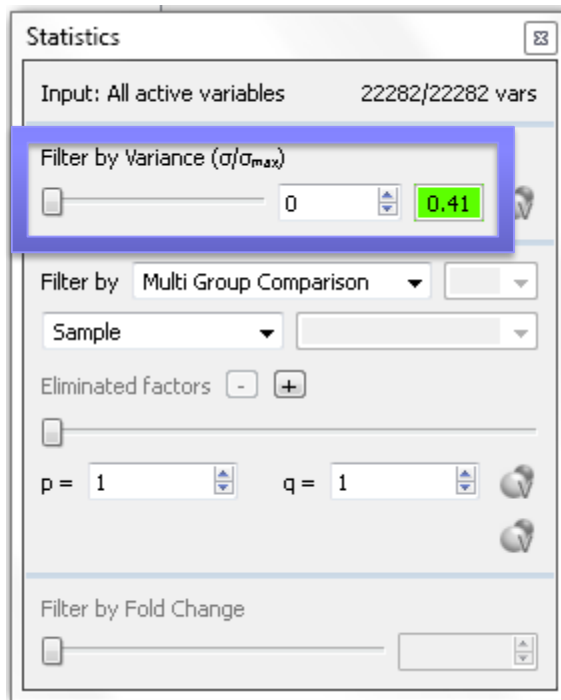


Exercise 10 – Exploratory Data Analysis

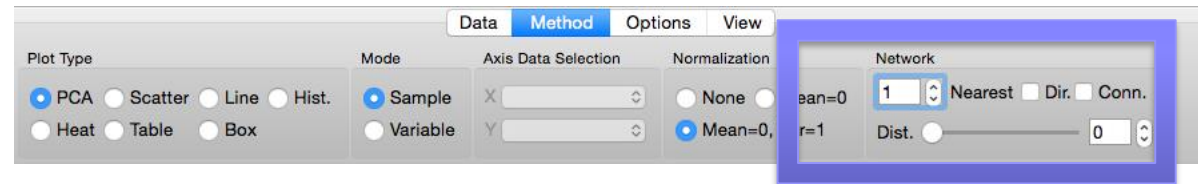
- Use the statistical and visual support in Qlucore Omics Explorer for exploratory analysis and hypothesis generation.
- Find new, previously unknown structures in data.
- Quality control – detect anomalies.
- Use the powerful pattern recognizing capabilities of the human brain.



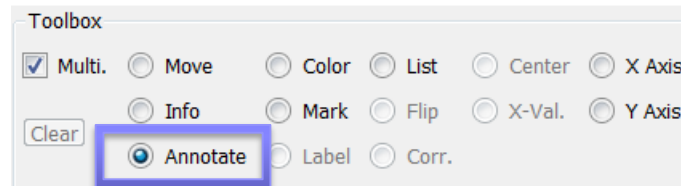
Exercise 10: Techniques



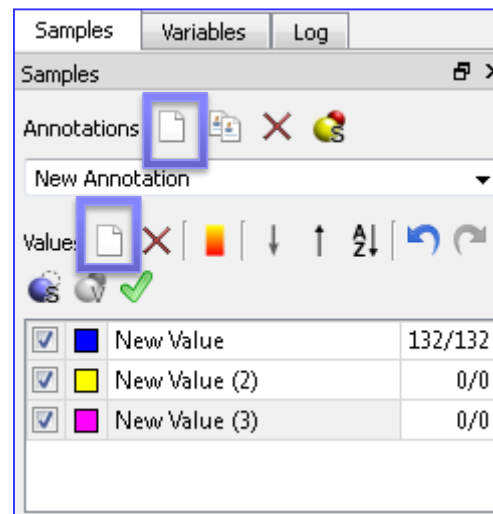
Variance filtering



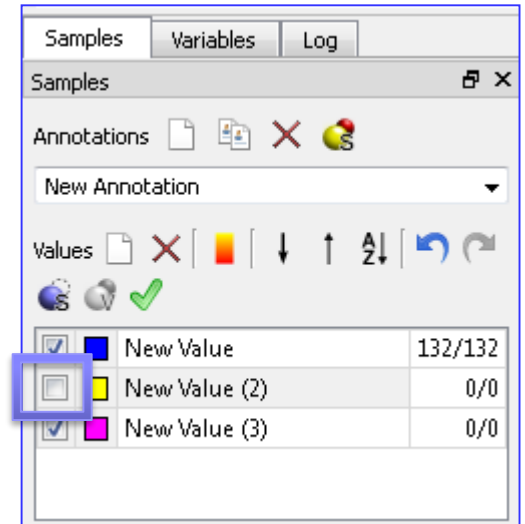
Network



Annotate



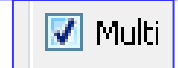
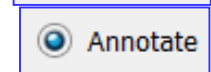
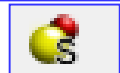
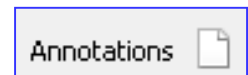
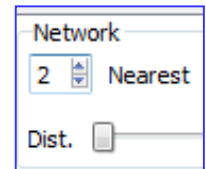
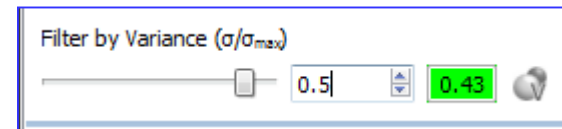
New annotation and values



Deselect group

Exercise 10: Steps

1. Load the **Acute Lymphoblastic Leukemia.gedata** data set from the Help menu.
2. Look at the data. Can you see any patterns?
3. Remove variables with low variance by moving the variance slider
4. Find the maximum Projection Score, 0.43.
5. Use networks to connect each sample to its closest neighbors.
6. Create a new sample annotation.
7. Color the data set according to the new annotation.
8. Create a new annotation value.
9. Select the Mouse tool Annotate and annotate the found group by circling the samples clockwise with the mouse
10. Deselect the found group.
11. Repeat steps 5-10.



Exercise 11 – Import RNA-seq data and annotations

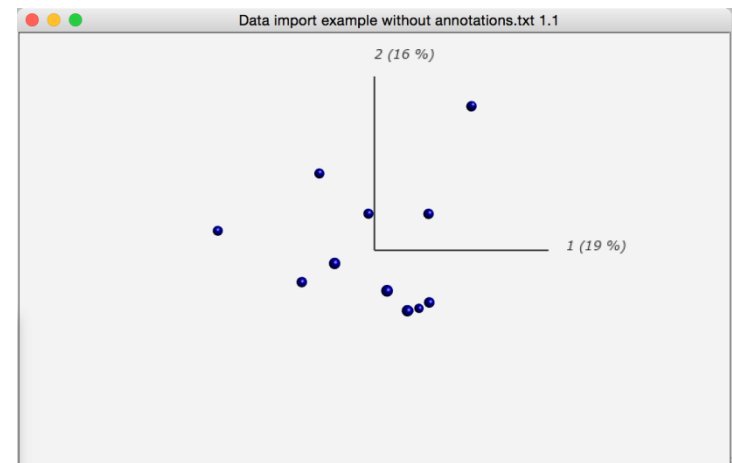
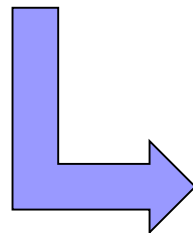
File import wizard

Column/sample annotations

Select row header cells or the leftmost annotation cells if there are no annotation headers.
Click in a cell again to deselect it.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1				SampleID	5303	5306	5302	5309	5312	5305	5311	5307	5301	5310
2	VariableID	Symbol	Name											
3	ID_01	ENC1	Variable 01											
4	ID_02	CKK8	Variable 02											
5	ID_03	PEX7	Variable 03											
6	ID_04	VPS13D	Variable 04											
7	ID_05	DLX8	Variable 05											
8	ID_06	SFRS9	Variable 06											
9	ID_07	ABCF1	Variable 07											
10	ID_08	DAD1	Variable 08											
11	ID_09	YY1	Variable 09											
12	ID_10	JTB	Variable 10											
13	ID_11	VPS13D	Variable 11											
14	ID_12	CLASP2	Variable 12											
15	ID_13	MYO1B	Variable 13											
16	ID_14	CLASP2	Variable 14											
17	ID_15	CLASP2	Variable 15											
18	ID_16	MA3	Variable 16											
19	ID_17	KIAA0746...	Variable 17											
20	ID_18	BCL2L1	Variable 18											
21	ID_19	NUP210	Variable 19											
22	ID_20	KIAA0746...	Variable 20											
23	ID_21	CHMP7	Variable 21											
24	ID_22	NUP210	Variable 22											
25	ID_23	TNP3	Variable 23											

Cancel Step 7 of 8 Back Finish

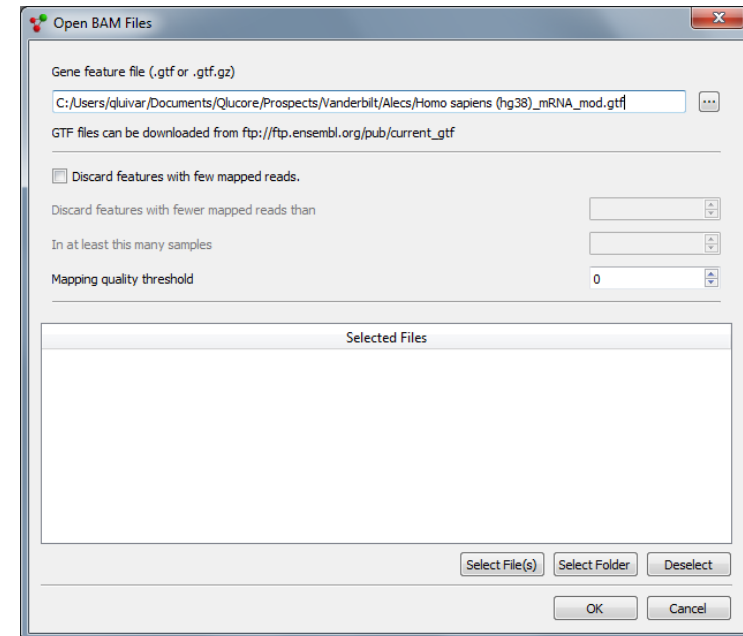


Exercise 11, step one: Import of RNA-seq BAM files

- Select File -> Open BAM files
- Define the GTF file (from the training folder)
- Define any filters and quality thresholds
- Select the BAM file folder (in the training folder)
- Press OK

The BAM files will be counted, normalized using TMM, log transformed and the data will be opened up in a PCA plot.

NOTE: At this step all functionality that you already have used and will continue to use is available.

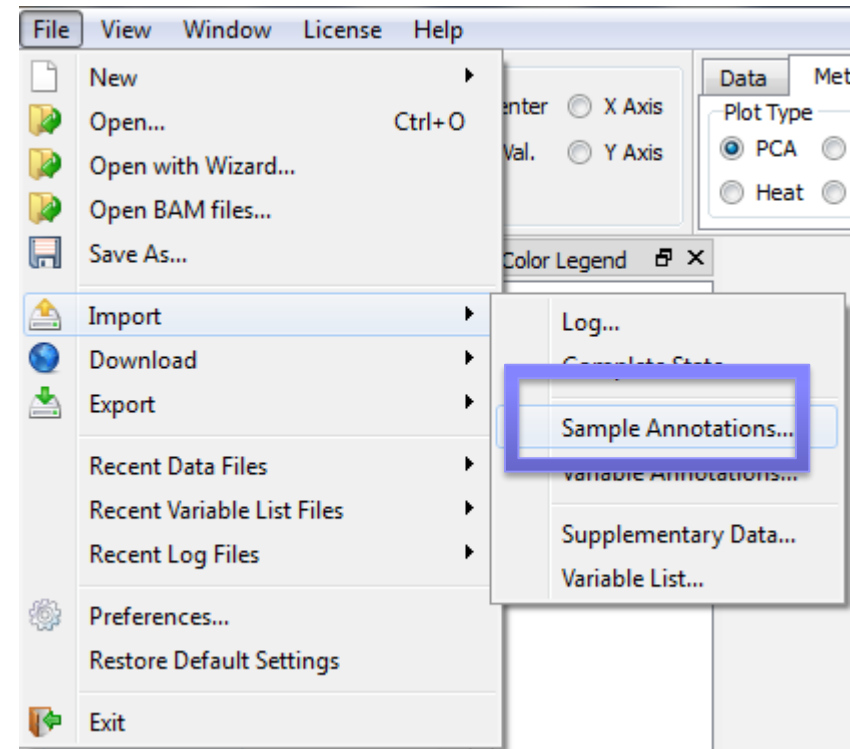


Note: The BAM files need to be aligned and sorted on coordinate.
The reference genome must be the same used for the alignment.

Exercise 11 – Add Annotations

Example content.

SampleID	Age	Gender	Treatment	Rank	Censor
5303	20	Female	Drug 2	Very low	1
5308	26	Female	Placebo	Very low	1
5302	28	Male	Drug 2	Low	1
5309	30	Male	Drug 1	Low	1
5312	40	Male	Drug 1	Medium	1
5305	40	Male	Placebo	Medium	0
5311	43	Female	Drug 1	Medium	1
5307	48	Male	Placebo	High	1
5301	54	Male	Drug 2	High	1
5310	56	Female	Drug 1	Very high	0
5306	56	Female	Placebo	Very high	1
5304	63	Female	Drug 2	Very high	1



- The unique sample/variable id in the first column must match the id of the imported data set.

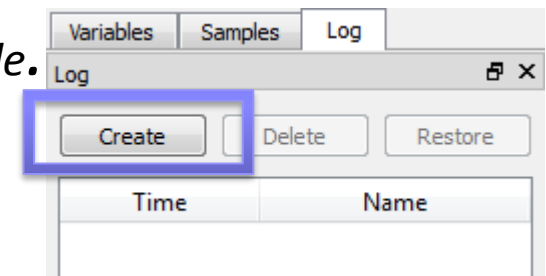
Exercise 11: Step 2 (import annotations)

1. Change the unique identifier for samples to Filename (Since this is the first column in the annotation file)
2. Select File/Import/Sample annotations
3. Select the file: **sample_annotations_RNA_seq.txt**
4. Import all the annotations
5. Check the sample annotations in the Sample tab and see that you now have all eight annotations

Save data and log points

- Once a data set is loaded you can save it as a .gedata file, which will contain all the imported data including annotations. (File/Save as)
- Save logpoint to save a snap shot of your analysis at a specific point in time.

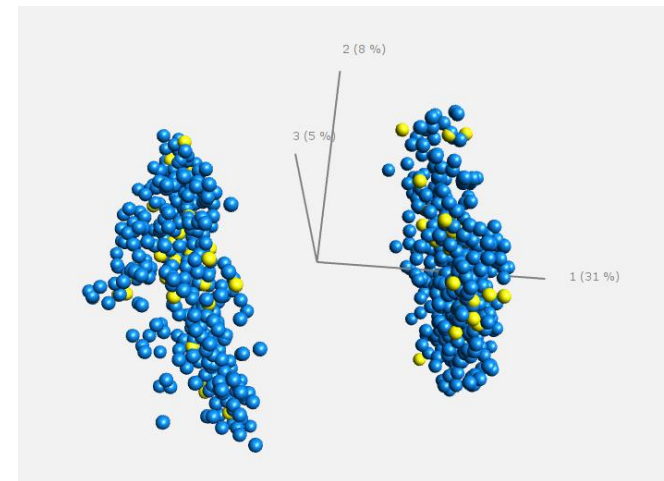
The log point requires that data is saved as a gedata file.



Exercise 12: Explore variable annotations

Three areas:

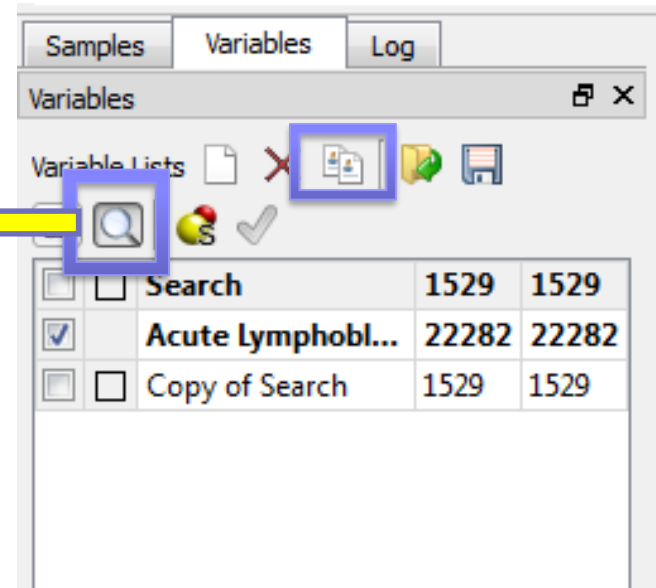
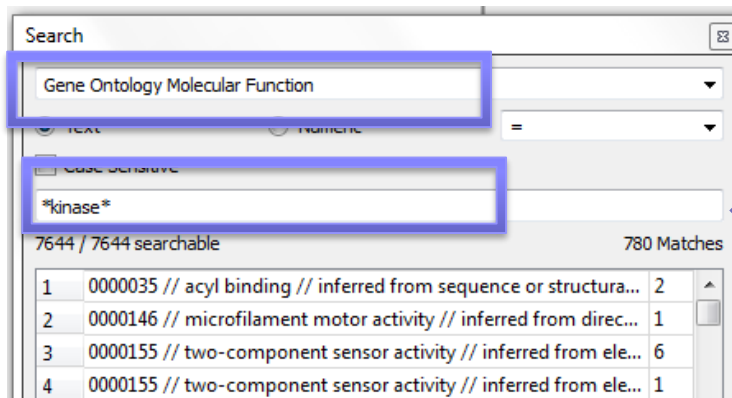
- Generate and use lists of genes with known functions
- Explore the data set using the lists
- Explore the data set interactively



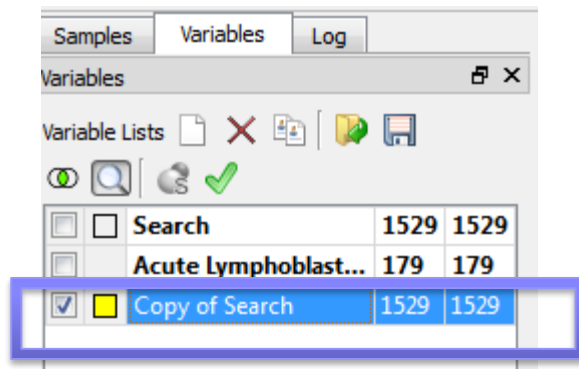
Exercise 12: Step one

1. Create a variable list with all kinases in the dataset
2. Colour the kinases
3. Use the list as input

Techniques12



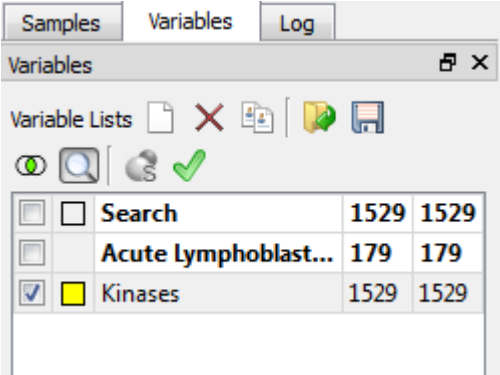
Search - Copy list



Colour and use list as input

Steps 12

1. Load the **Leukemia data set** from the Help menu.
2. Select the Variables tab.
3. Use the Search tool to find all kinases (Search the Gene Ontology Molecular Function for *kinase*)
4. Save the results by copying the search list to a new list called Kinases
5. Open up a synchronized plot and make it a variable PCA plot
6. Perform a t-test on Leukemia subtype – E2A-PBX1 and filter down to 500 genes
7. Colour the kinases
8. Use the kinases list as input
now you only see the kinases
9. To save the rel kinases – make a copy of the active list – "Kinases discr E2A-PBX1"

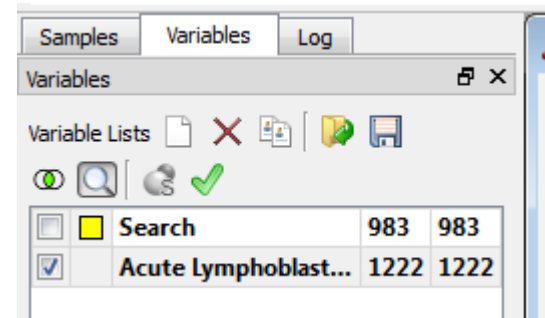


Variables			
Variable Lists			
<input type="checkbox"/>	Search	1529	1529
<input type="checkbox"/>	Acute Lymphoblast...	179	179
<input checked="" type="checkbox"/>	Kinases	1529	1529

Exercise 12 – Step two

Interactive Search:

1. Remove the use of the Kinase list
2. Mark the Acute Lymphoblastic Leukemia as the active list in the variable tab
3. Open the Search dialogue
4. Search chromosomal location for chr5*
5. Colour the variables by the Search list
6. Make a copy of the Search list and call it chr5
7. Change the Search to chr6*
8. Note the interactivity



Exercise 13: More ways of generating list

- Use and combine saved lists
- GO Browser

The screenshot shows the GO Browser interface with the search term 'kinase'. The 'Tree View' is selected, showing a hierarchy of GO terms. The 'kinase activity' term (GO:0016301) is selected, and its associated genes are listed in the 'Genes' column. The 'Flat View' is also visible, showing a list of genes and their associated GO terms. The 'Genes in Selected Terms' section at the bottom right lists the genes A1BG, A6NF36, and AAAS, along with their names and symbols.

ID	Name	Genes	Matches	Definition
GO:0003674	molecular function	685	1	Elemental activities, such as catalysis...
GO:0016301	kinase activity	128	1	Catalysis of the transfer of a phosph...

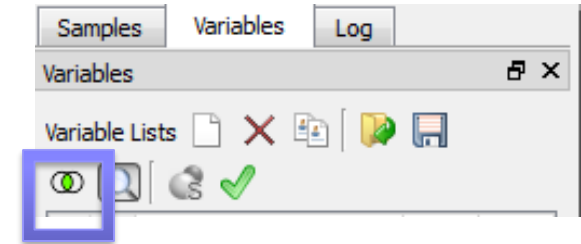
Symbol	Name
A1BG	Alpha-1B-glycoprotein
A6NF36	Coiled-coil domain-containing protein ENSP000002...
AAAS	Aladin

Genes in Selected Terms - 685

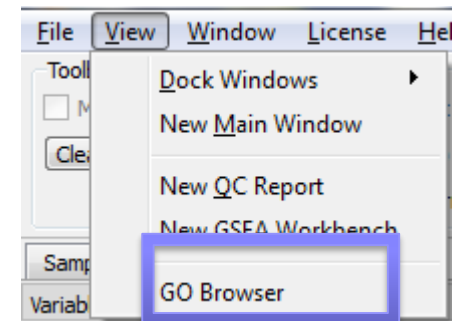
Symbol	Name
A1BG	Alpha-1B-glycoprotein
A6NF36	Coiled-coil domain-containing protein ENSP000002...
AAAS	Aladin

Exercise 13 steps

1. Use the set tool to combine lists, for instance combine the Kinase list and the Chr5 list.
2. If intersection is selected a new lists with the kinases on chr 5 is created.

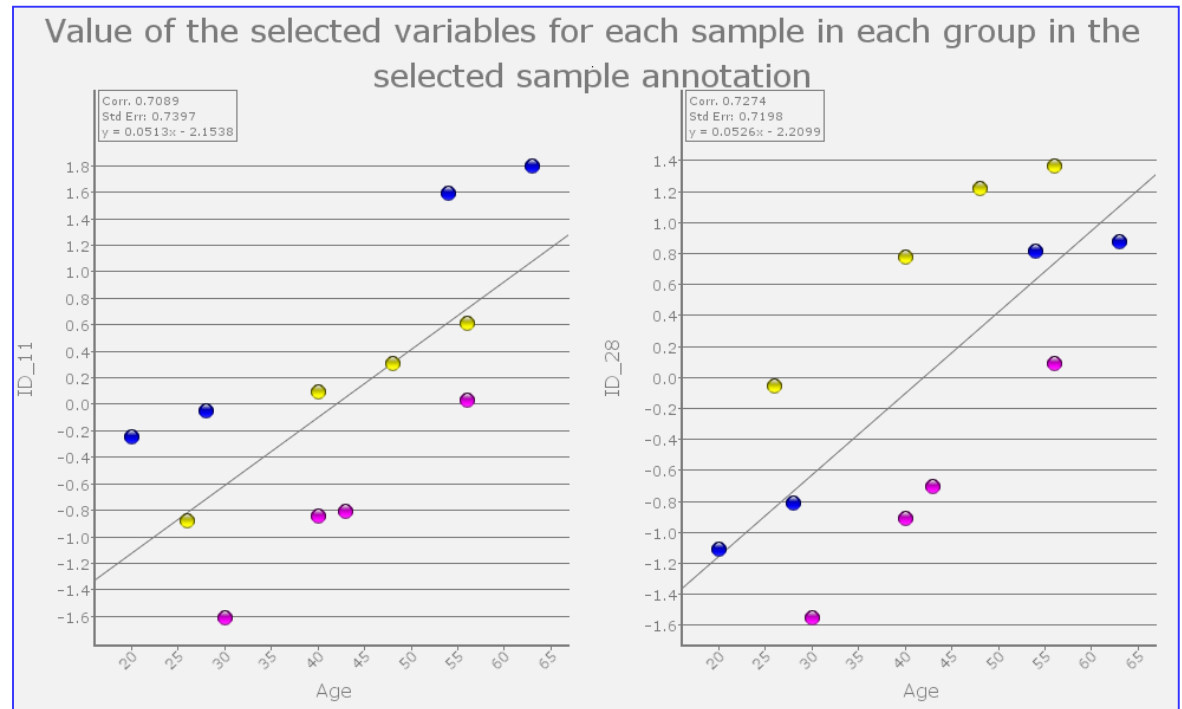


1. Start the GO Browser (View)
2. Use Find to identify terms matching with "kinase"
3. Select the kinase activity term by checking the check box
4. The select Export and the list is available in the main window Variable panel.

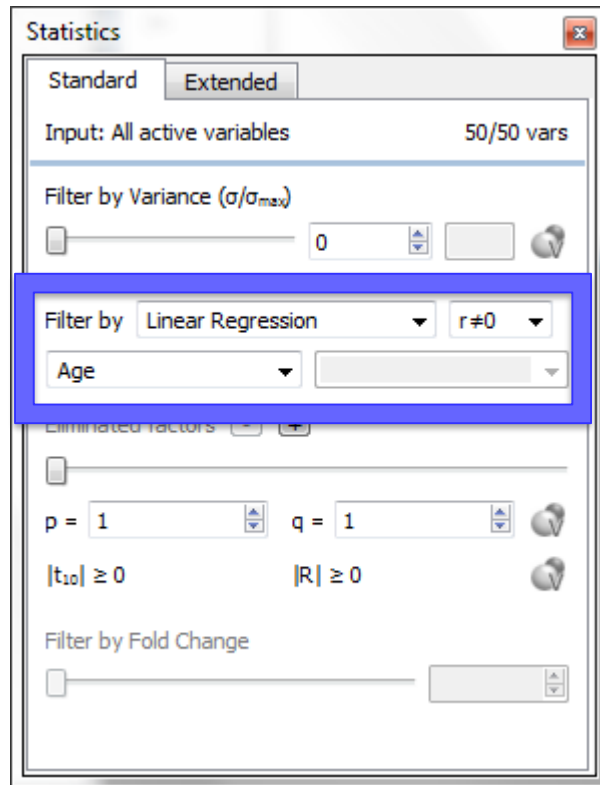


Exercise 14 – Linear regression

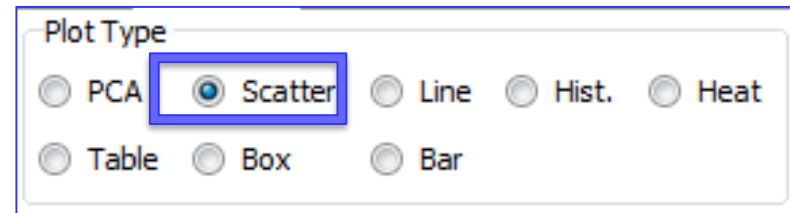
- Perform Linear regression
- Find statistically significant variables
- Show the variables in a scatter plot



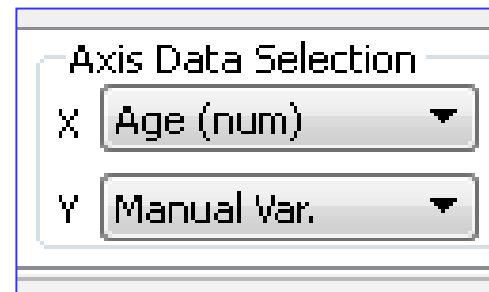
Exercise 14: Techniques



Linear regression



Select Scatter plot

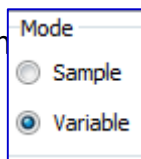
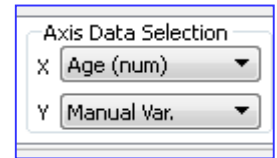
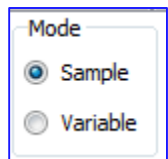
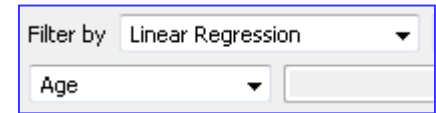


Define the axes

Exercise 14: Steps

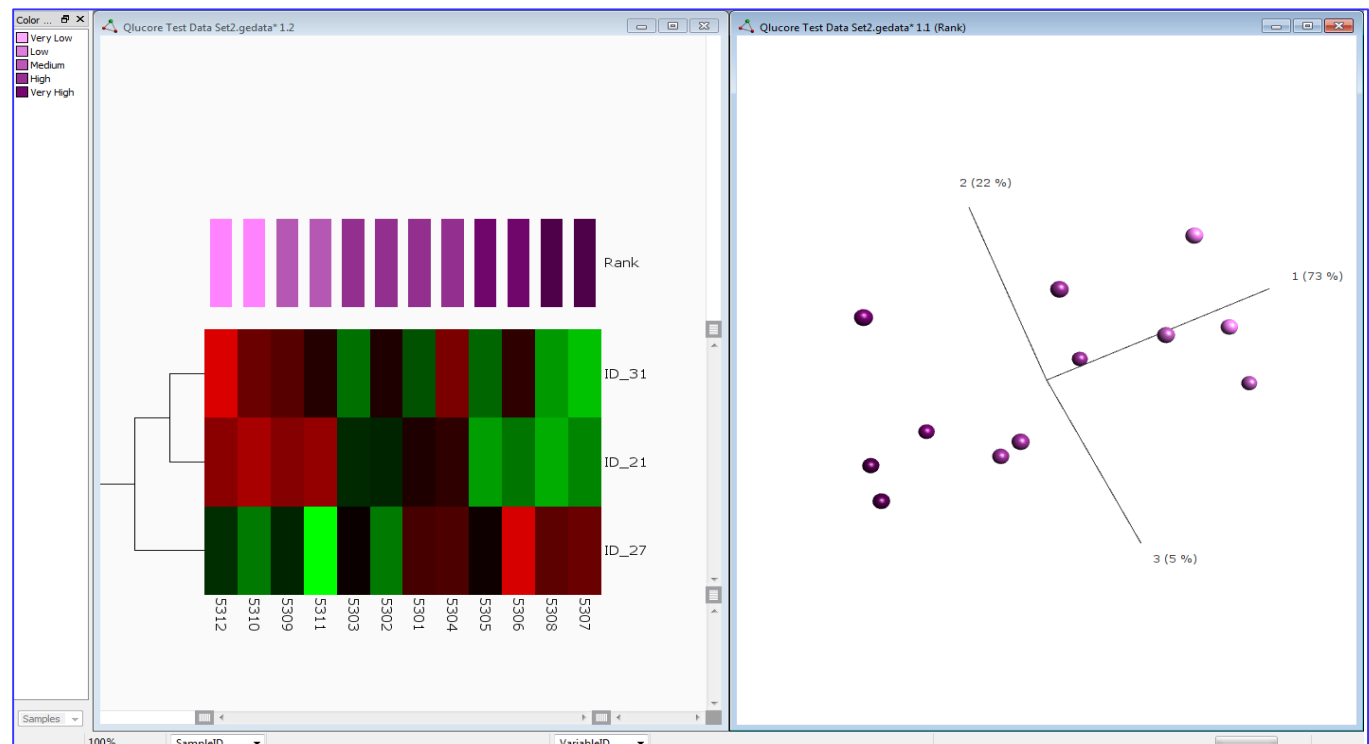
1. Start with **Qlucore test data set**
2. Start with one Sample PCA plot
3. Select Linear Regression in the Statistics dialogue
4. Select the annotation Age
5. Drag the statistical slider to find discriminating variables
6. Show the variables in Scatter plots.
 - Select Scatter Plot in the Method Tab, Sample Mode
 - Select the dropdown menu for the x-axis in the Axes Data Selection in the Method tab
 - Select Sample annotation Age and ***Numerical scale!***
 - Select the dropdown menu for the y-axis in the Axes Data Selection in the Method tab
 - Select Variables selected by the Y-axis tool
 - Click on your list of variables in the variables tab or select individual variables in the variable list

Extra exercise: Select Mode/Variable and make a scatter plot with sample 5301 on the x-axis and sample 5302 on the y-axis to show the variables.

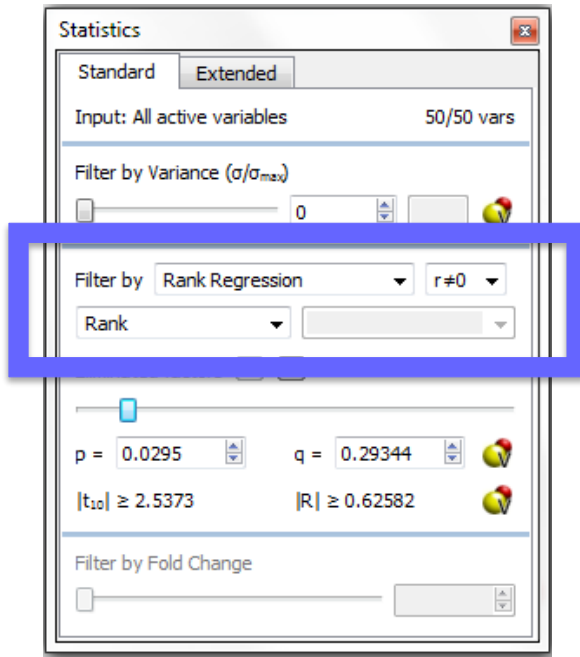


Exercise 15 – Rank regression

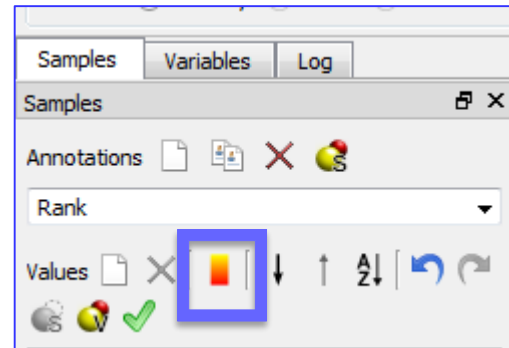
- Perform Rank regression
- Work with Color gradient



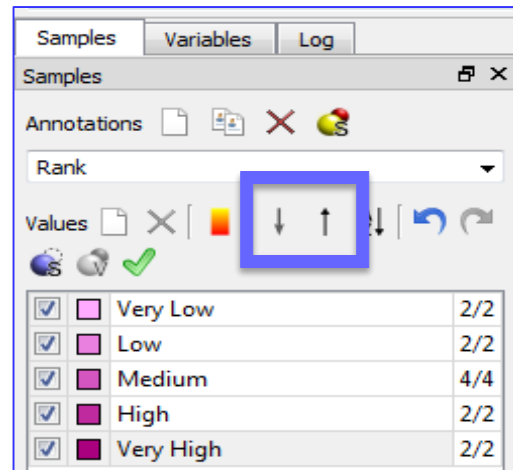
Exercise 15: Techniques



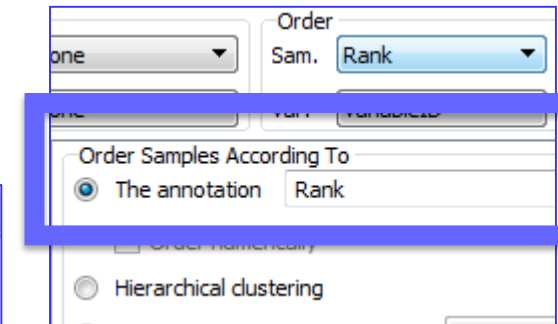
Rank regression



Color gradient



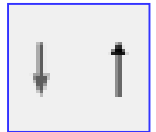
Sort values in ranked order



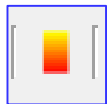
Sort heatmap on Rank

Exercise 15: Steps

1. Close the open dataset and load the **Data Import Example.gedata** file again.
2. Open a new synchronized plot. Make it a heat map.
Tile the plots
3. Select the Rank annotation(Sample tab). Sort the rows in rank order if they are not correctly sorted.
Use the up and down arrows.
4. Create a color gradient for the Rank annotation:
 - Double click the colored box next to the "very low" value to change color to a light color
 - Do the same thing on the "very high" value but select a dark color. Pick the colors you like.
 - Now click the gradient icon and follow the instructions.



<input checked="" type="checkbox"/>	Very Low	2/2
<input checked="" type="checkbox"/>	Low	2/2
<input checked="" type="checkbox"/>	Medium	4/4
<input checked="" type="checkbox"/>	High	2/2
<input checked="" type="checkbox"/>	Very High	2/2



Exercise 15: Steps – continued

5. Color your PCA plot and your heatmap according to the Rank annotation.
6. In the Statistics box select Rank regression and select the annotation Rank.
7. Drag the statistical slider until you have only 3 variables left.
8. Notice that the samples line up in the rank order
9. Sort the heat map according to Rank (View/Order/The annotation: Rank)

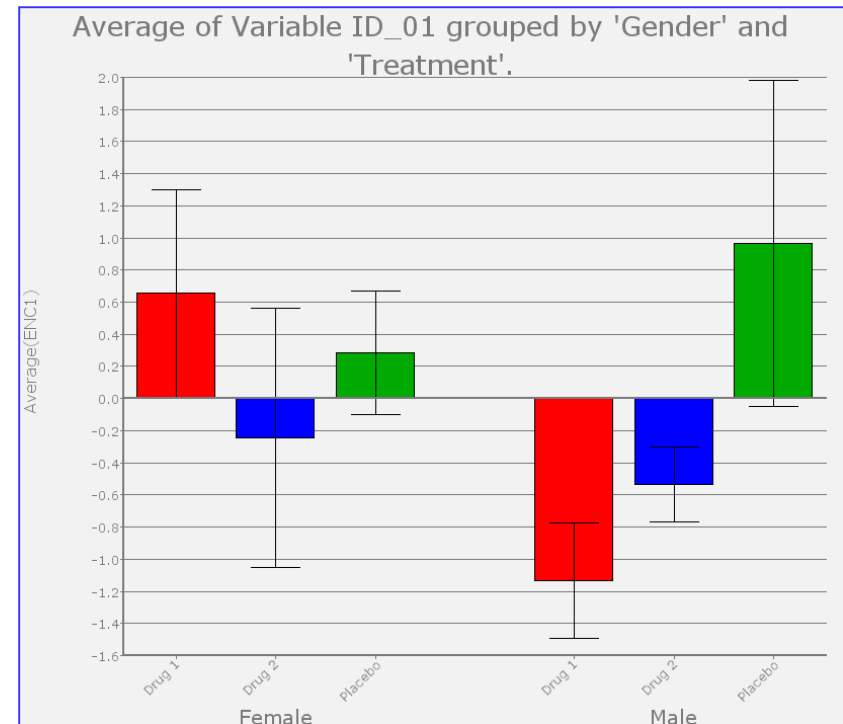
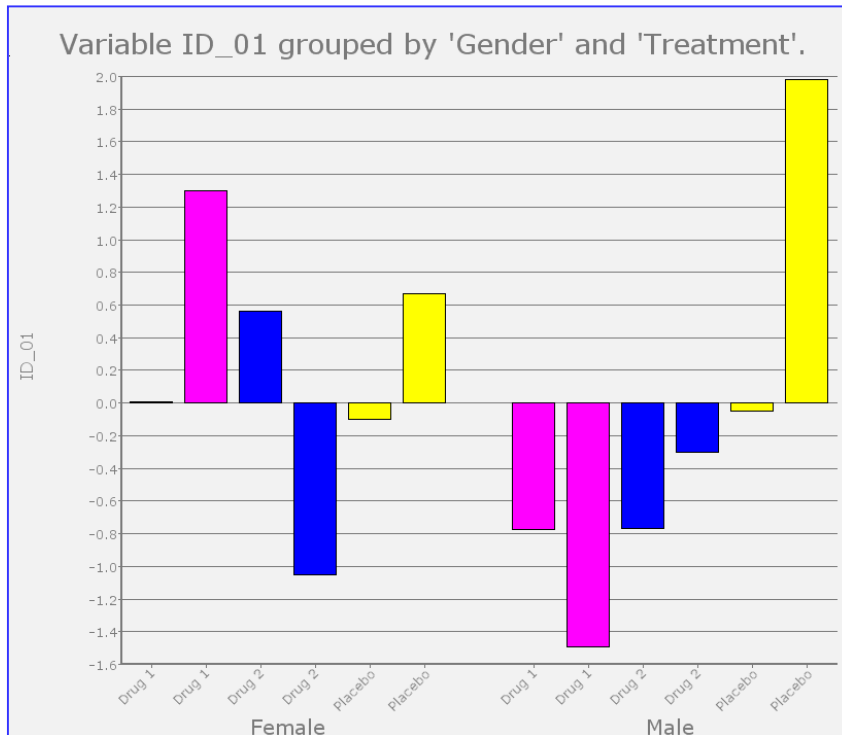


Filter by	Rank Regression	r≠0
Rank		

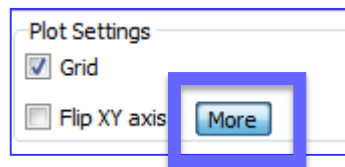
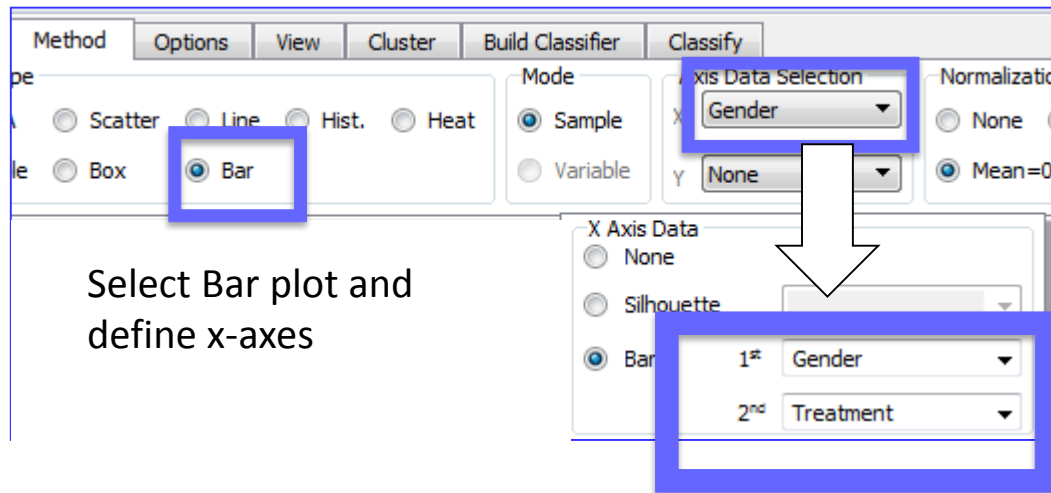
Order Samples According To	
<input checked="" type="radio"/> The annotation	Rank

Exercise 16 – Bar plot

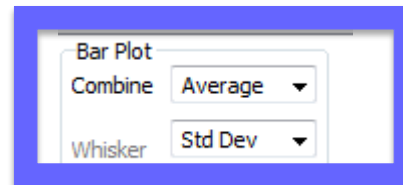
- Display results in Bar plots



Exercise 16: Techniques



View tab – More plot settings

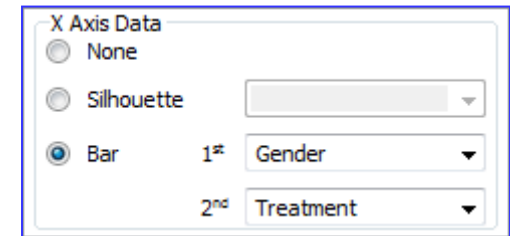


Options tab – for average bars

Exercise 16: Steps

1. Load the **Qlucore Test Data Set** from the Help menu.
 2. Select Bar plot and Select Gender and Treatment for the x-axes.
 3. Select one (or several) variable(s) for the y-axis
 4. Colour the bars according to Treatment
 5. Create average bars by selecting Average from Bar options – to the far right in the Options tab
 6. Select Whiskers: Std Dev
 7. Change the y-axis label to Symbol
- Continue on next page....

☒ Bar



X Axis Data

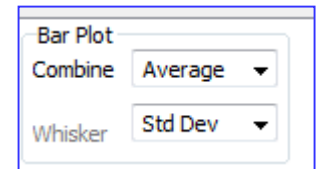
☐ None

☐ Silhouette

☒ Bar

1st Gender

2nd Treatment



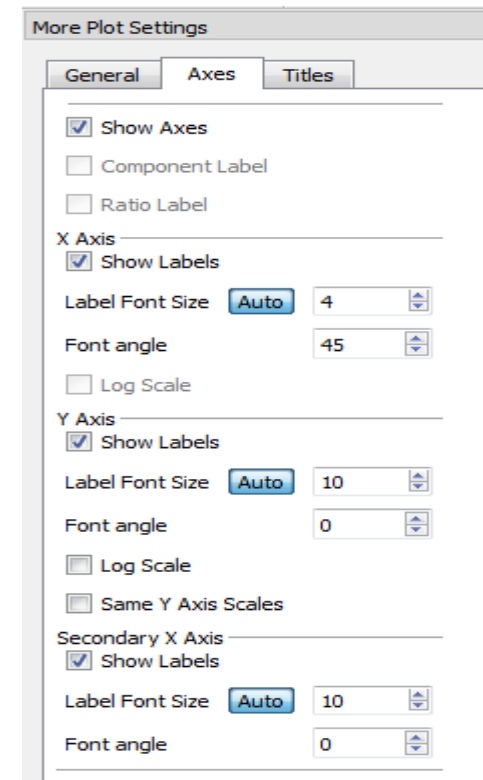
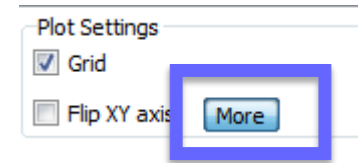
Bar Plot

Combine Average

Whisker Std Dev

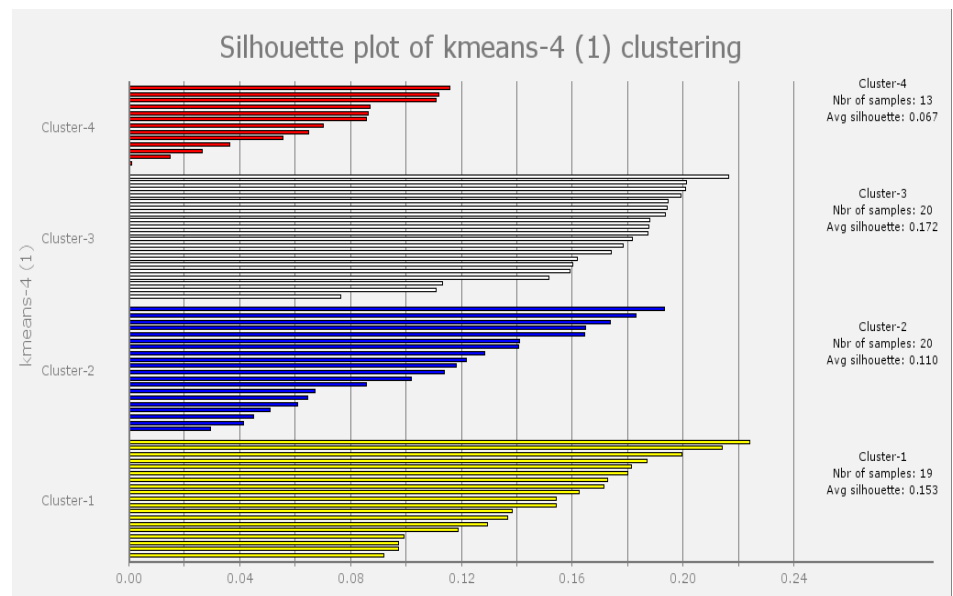
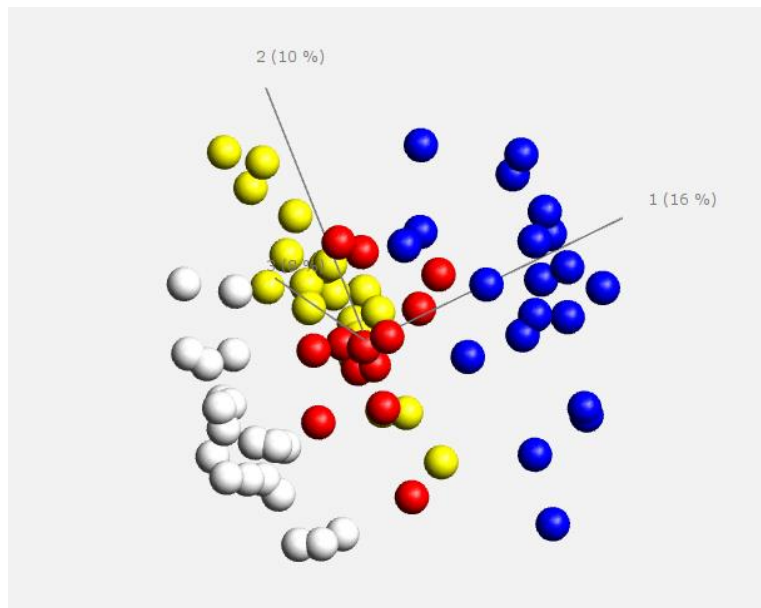
Exercise 16: Steps - continued

8. Click the More button in the View tab and open the More plot settings Window
9. Set the size of the 2nd x-axis label to 15
10. Change the distance between the individual bars and the groups
11. Change the angle of the first x-axis label

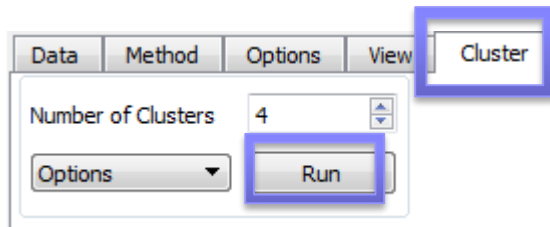


Exercise 17 - Clustering

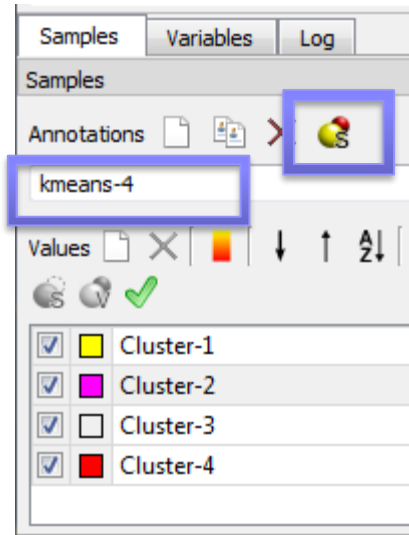
- Use k-means clustering to find structures
- Use Silhouette plots to evaluate clusters



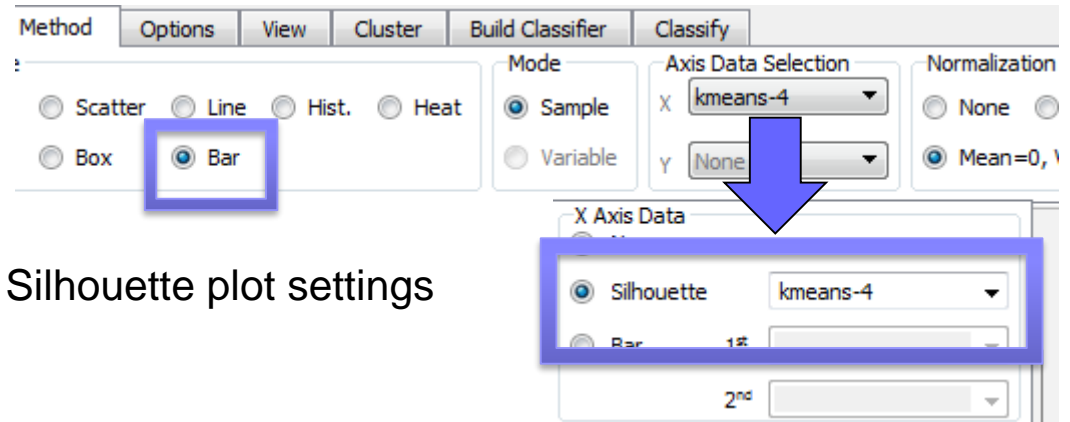
Exercise 17: Techniques



Cluster



Colour the clusters



Silhouette plot settings

Exercise 17: Steps

1. Load the **Leukemia Example Clustering** dataset
2. Perform variance filtering so that you keep approx 1400 of the variables
3. Go to the Cluster tab, set no of Clusters to 3 and press Run
4. Colour the samples according to the new clustering annotation (k-means-3)
5. Open a New synchronized plot and make it a Silhoutte plot
6. Inspect the clusters in a silhouette plot
(Select Bar, X-axis, Silhoutte, k-means 3)
7. Use the mouse tool Mark to mark the sample with a negative Silhoutte value
8. Repeat step 3-6 with 4 clusters. Which cluster option seems most relevant?

More info – www.qlucore.com

- Qlucore Omics Explorer
 - Tutorial
 - Reference manual
 - List of References of Implemented Algorithms and Methods
- Homepage
 - Films
 - How to documents
 - FAQ
- Monthly webinars
 - Register on the homepage

The screenshot shows the Qlucore Omics Explorer website. The header is blue with the QLUCORE logo and a quote: "Qlucore Omics Explorer is adding more creativity to our research than any other software I have used." by David Gisselsson Nord, MD, PhD, Associate Professor, Department of Clinical Genetics Lund University. The navigation bar includes links for HOME, COMPANY, PRODUCTS, CONTACT, SUPPORT, DOWNLOADS, and a LOGIN button. Below the navigation bar, there are links for OVERVIEW, FEATURES, VERSION 2.3, Q&A, REFERENCES, and DOCUMENTATION. The main content area is titled "DOCUMENTATION" and contains a paragraph about finding documents and video clips. Below this is a section for "MOVIES AND INTRODUCTIONS" with a list of items: "Introduction To Qlucore Omics Explorer" (Video), "Introduction to 2D plots" (Video, Note: No sound), "Qlucore Omics Explorer Features" (.pdf), "Qlucore Omics Explorer - a very brief introduction" (Video), "Working with fewer samples." (Video), "Work with subgroups and find discriminating variables. Generate lists with p- and q-values." (Video), and "Qlucore Omics Explorer Tutorial" (.pdf). On the right side, there are three buttons: "VERSION 2.3 NEWS", "HOW TO BUY", and "GET THE VISUALISATION GUIDE". Below these are sections for "DATA" (GENE EXPRESSION, PROTEIN ARRAY, MIRNA, DNA METHYLATION, PROTEOMICS), "DISEASE AREA" (CANCER, OBESITY, DIABETES), "PERFORMANCE" (with a bar chart), and "SCREENSHOTS" (with several small images).

Contact/Support

- Please feel free to contact us if you have any questions.

glucoreinfo@glucore.com

support@glucore.com