## DiVA Tutorial

Interactive Selection

The main idea of interactive visualization in DiVA is that the system provides the possibility for users to highlight elements of the loaded dataset based on the information they are interested in through users selections. DiVA then automatically highlights the same elements in the other visualization displays, giving the user an overview of the selected data from different angles in fast and easy way.

Users can easily select data from any of the analysis displays (except for the profile plot) by selecting data in the tables, drag-select for the plots, direct mouse selections on the side and top clustering trees, or selection in the clustering heatmap, see **Figure 1**.
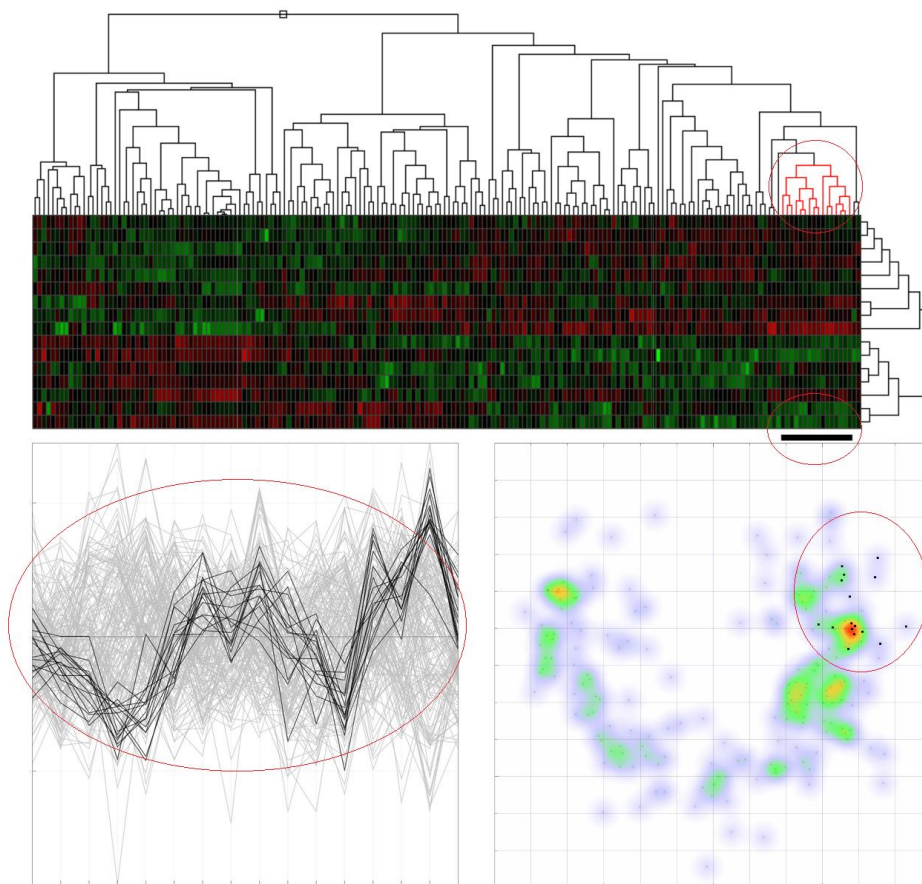

Figure 1: Interactive Selection.

Customized Colored Groups

DiVA supports the creation of customized row/column groups. The idea of using colored groups is to allow the users to label parts of the data and make it easy to separate data, helping users maximize the knowledge gained from the dataset.

As shown in **Figure 2**, although the tool was able to clearly separate the data in the PCA plot, it is still hard to separate the data in the profile plot. However, using colored groups enhances the separation, especially on the profile plot, allowing users to track the data and compare between the groups in an easy and clear way.
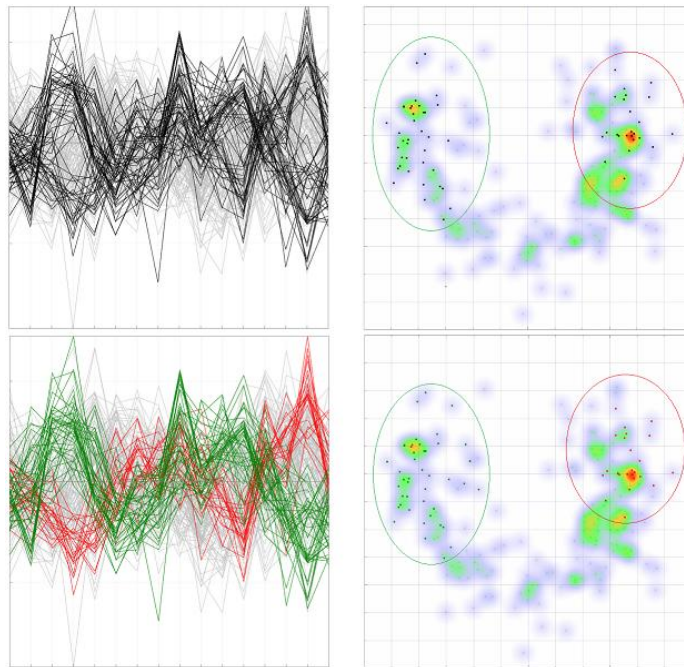
Figure 2: Customized Colored Groups.

Dataset Input Format

The input data format is a text file in a tabular format, where the first column represents the unique data identity key (e.g., protein accession number) and the first row represents the samples or conditions headers (**Figure 3**).

Uniprot	CT0	CT3	CT6	CT9	CT12
Q61838	-0.30023	-1.44339	1.210377	1.023491	0.056233
B1AWZ8;...	-0.37882	-0.40635	-0.80148	-1.16341	0.678253
A2A6R3;...	-0.35151	-1.01062	-0.83599	0.609558	1.645096
Q8K440-...	-0.03124	0.799212	-0.3481	0.395943	0.785763
Q8VI46;...	0.185595	-0.90542	-0.30063	-0.2662	0.72258
P50544;...	0.23632	-1.20918	-1.45858	0.317204	1.644904
Q5XG73-...	0.177394	0.24354	-1.3053	0.088746	0.42761
A2AKK5-...	-0.55615	0.997514	0.217695	-0.51129	-1.51211
Q91VA0-...	0.331171	-0.14051	-0.4445	0.827418	1.197661

Figure 3: Dataset Input Format Example.

The system supports adding data annotations using separate annotation text file. The first column in the annotation file has to be identical to the first column in the data file (the unique data identity column). Additionally, the annotation file must have the same file name as the dataset file name with the addition of "_Annotations.txt". For example if the dataset file name is "Proteomics Dataset Sample.txt", then the annotation file name should be "Proteomics Dataset Sample_Annotations.txt" (**Figure 4**).

Uniprot	Gene name	Protein...
Q61838	A2m	Alpha-2...
B1AWZ8;...	Abca1	ATP-bin...
A2A6R3;...	Abca6	ATP-bin...
Q8K440-...	Abca8b	ATP-bin...
Q8VI46;...	Abcc2	Canalic...
P50544;...	Acadvl	Very lo...
Q5XG73-...	Acbd5	Acyl-Co...
A2AKK5-...	Acnat1	Acyl-co...
Q91VA0-...	Acsm1	Acyl-co...

Figure 4: Annotation Input Format Example.

All the dataset and the annotation files should be placed in the "divaFiles" folder located on the server machine.

DiVA Demo Application

Before starting to use the DiVA omics demo application we advise users to go through the below tutorial and guidelines to familiarize themselves with the main system features.

Icons and Symbols

ℹ️	Information for users regarding how to use the DiVA components.
⤢	Maximize the given DiVA component.
⤧	Minimize the given DiVA component.
💾	Export image from chart or data in tables.
⚙️	Change the input parameters used for analysis calculations.
🔍	Reset the PCA scatter plot to the default zooming level.

Minimize/Maximize Component

DiVA consists of a number of visualization components, each component (except for the left panel tables) has two visualization states, minimized or maximized. Generally the maximized state provides better visualizations and makes better use of the available space in the browser.

Users can switch between the two states easily by clicking the maximize ⤢ or minimize icons ⤧ in the upper right corner of each of component.

Selection

Users can easily select data in any of the components (except for the profile plot) by drag-select data in the tables, drag-select in the plots, or by direct mouse selection for the side/top trees in the heatmap (**Figure 5**).

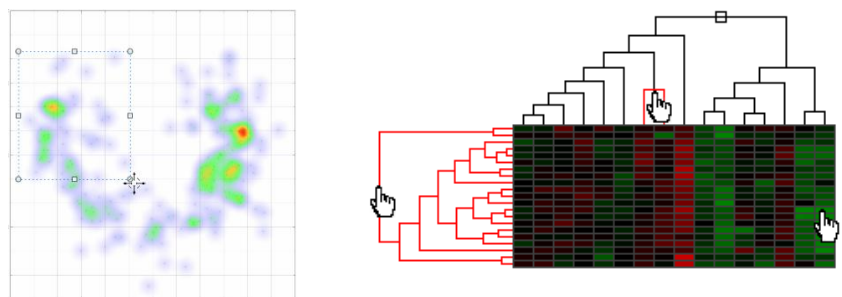Upon selection the other components will be updated and highlighted using the group color the data belongs to.

Figure 5: DiVA Data Selection.

Row and Column Groups

After data selection in any of the visualization components the user can create customized row colored groups using [Create Row Group]. The user can also create columns groups using [Create Column Group], after selecting the columns from the top tree in Hierarchical Clustering component or the column table. Please not that user can switch between the row and column tables using the control buttons (●) Rows ( ) Columns in the upper right corner in the left panel.

Export Data and Plots

Export the dataset (complete or row selection) as a tabular file using the export dataset button [Export Dataset ▼] located in the left panel.

Export the content in the Rank Product Table using the save icon [💾] in the upper right corner for the table.

Export plots as PDF using the save icon [💾] in right upper right corner for the maximized plots.

Zoom and Select

PCA: Invoke the zoom and select mode using the zoom checkbox, where the users can zoom to the desired level and then reactivate the selection mode by unchecking the zoom checkbox, **Figure 6**.
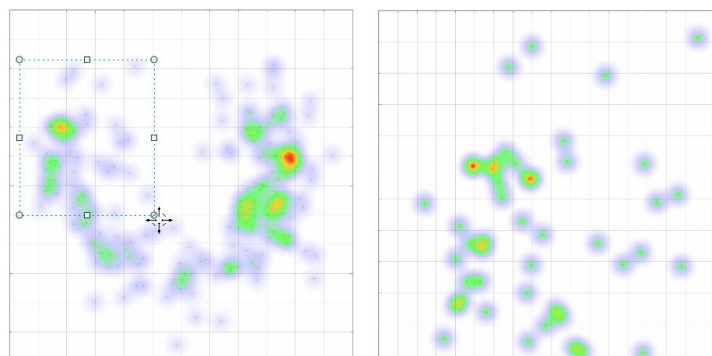
Figure 6: Zooming in the PCA plot.

Hierarchical Clustering: Zoom and select data using the zoom in/out scroll bar . Zoom to the desired level and then select data from the dendrograms or heatmap.

Searching for Proteins/Genes

The user can search for a specific protein/gene using keywords from any of the provided annotations (UniProt accession, proteins name or gene name in our case). This helps users find specific proteins/genes in the dataset and to directly get various displays for this protein/gene. For example the users can search for "albumin" as shown in **Figure 7**.
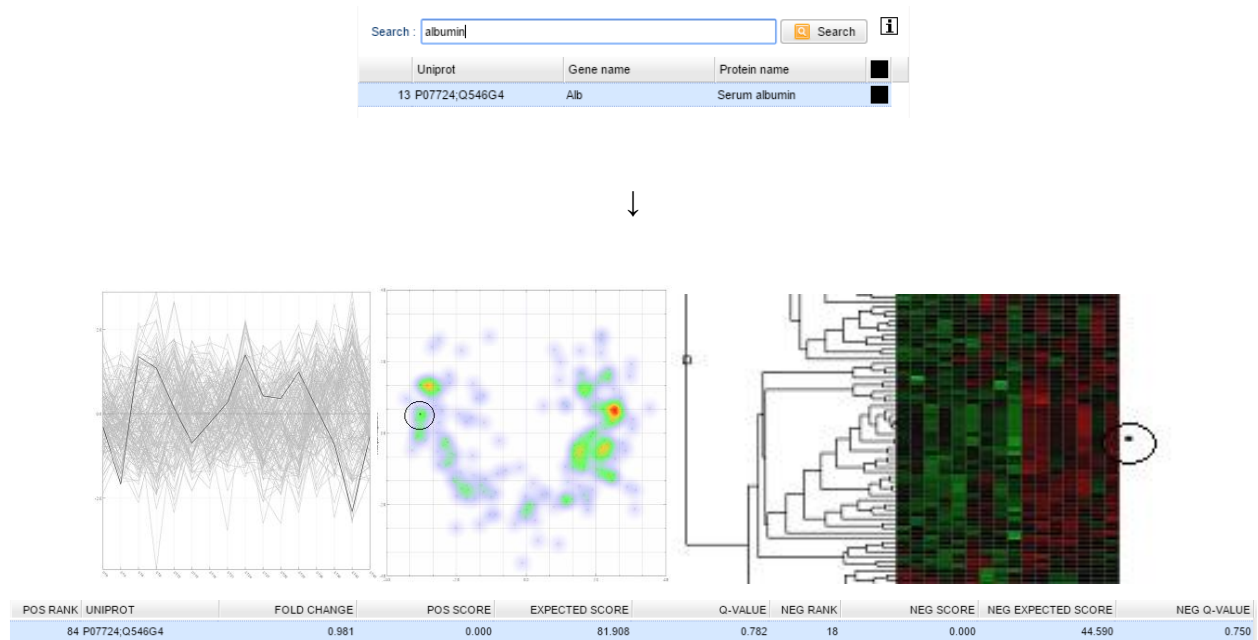
POS RANK	UNIPROT	FOLD CHANGE	POS SCORE	EXPECTED SCORE	Q-VALUE	NEG RANK	NEG SCORE	NEG EXPECTED SCORE	NEG Q-VALUE
84	P07724;Q546G4	0.981	0.000	81.908	0.782	18	0.000	44.590	0.750

Figure 7: Search for a specific protein/gene.

The DiVA omics demo

Start using the system by visiting http://diva-omics-demo.googlecode.com and select any of the available datasets from the dataset drop menu (**Figure 8**).

Figure 8: Datasets drop down list.

Loading the Selected Dataset

Upon the dataset selection from the drop down menu, the system starts loading the dataset, performs the different analyses for the data and visualizes the generated results (**Figure 9**). Currently the system supports Profile Plot, PCA, Rank Product analysis and Hierarchical Clustering, in addition to the interactive annotations table.
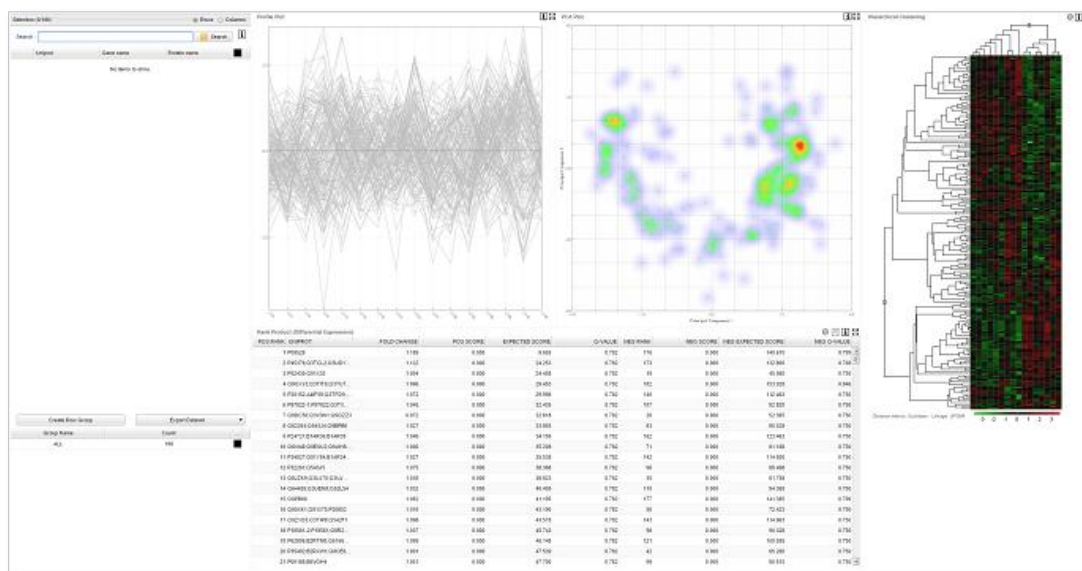

Figure 9: Loading the Proteomics Dataset Sample.

Now you are ready to work with DiVA!