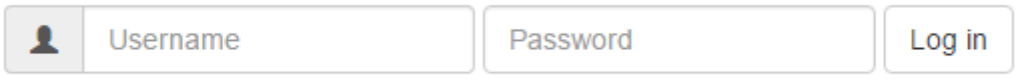
Quick Start Guide

Content

Getting started

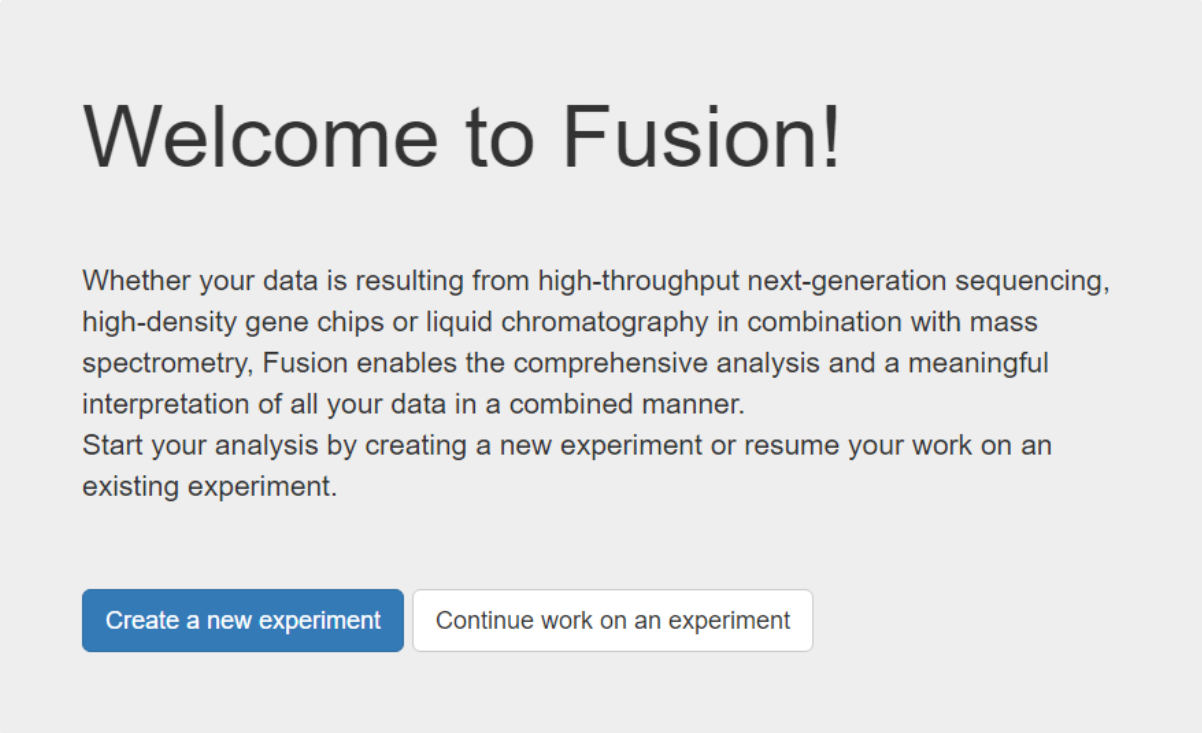
In the following, we would like to give you a quick introduction into *Omics Fusion*, a web based platform for polyomics data analysis and visualization. We will show you how to upload your data and how to start some basic analysis tools and visualization methods. Let's start with logging into *Omics Fusion*. Type in your **username** and **password** and click **Log in**:



If you don't have an account yet, simply request one at fusion-service@cebitec.uni-bielefeld.de. It's free of charge!

After you are logged in, you have to select a project. Usually, you will have access to one or multiple projects with your account, depending on what you requested when your account was created. If not, please contact the email address above.

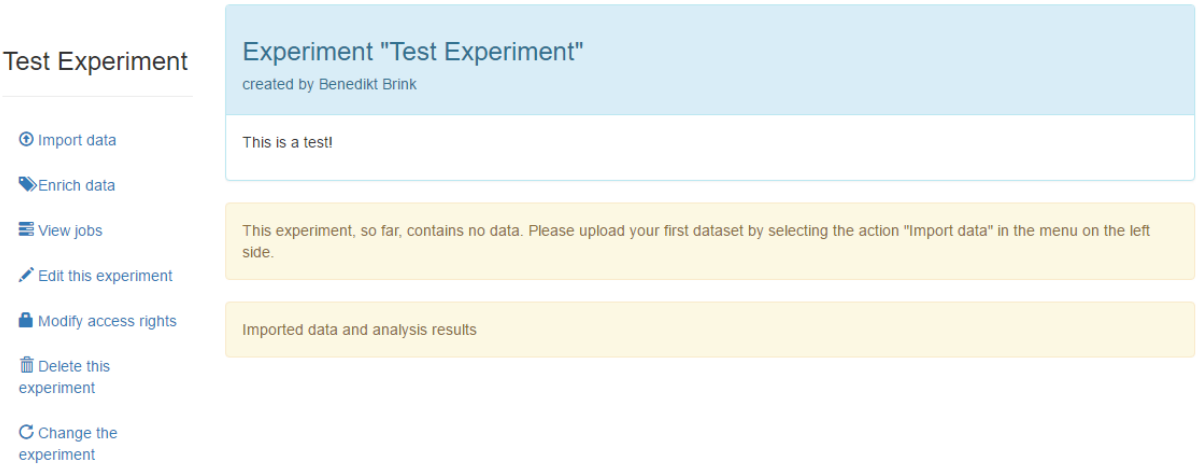
Now you have successfully logged into *Omics Fusion.* You can now choose if you wish to create a new experiment, or continue to work with an existing one:



Creating a new experiment

To create a new experiment, we have to choose a name and a description. Usually, we would want to choose a meaningful name and provide some basic information about the data, e.g. what kind of data, which organism it is from and under what conditions it was acquired.

Once we click on **Create,** we will be automatically redirected to our new experiment:



The menu on the left provides us access to several important operations:

Import data	Import raw data into the current experiment (xls or xlsx files)
Enrich data	Enrich the data with information from UniprotKB, NCBI or KEGG
View jobs	List of all running, finished, and failed jobs
Edit this experiment	Change the name and the description of the experiment
Modify access rights	Give access to other *Omics Fusion* users
Delete this experiment	Delete this experiment
Change the experiment	Change to a different experiment

How to upload data

It is quite simple, really: basically, just select an Excel spreadsheet file (xls/xlsx) via the **Browse** button, click on **Import** and follow the instructions.

- Each type of data (transcripts, proteins, metabolites) should be in a separate file (multi-sheet documents are not yet supported).
- Please remove any frills and furbelows - first row should contain the headline, continuing with data from the second row on.
- Data should be formatted as numbers, formulas in cells might lead to complications.
- After file selection and upload, first, the type of each column will need to be specified, e.g. "name" (mandatory), "description", "ec number" or "value".
- Second, each "value"-column will need to be explained, e.g. by assigning the factor "time" with the levels "10min", "20min" and "30min" to three "value"-columns.
- All imported data will afterwards be found on the experiment's page as a new dataset. At this, we may click on **Analyze, visualize, transform...** to work with the data.

Specific characteristics and limitations:

- Replicates can simply be imported by specifying the same level for each column, e.g. "10min". At this, it makes no difference whether each replicate has its own sheet or all replicate values are in different columns of the same sheet.
- Annotation data such as EC numbers only needs to be imported once for each individual feature (transcript/protein/metabolite). This can be done directly in one file together with real data or in a separate file just containing the annotations (which most likely speeds up the import).
- It is possible and if applicable encouraged to assign not only one but also two (or even more) factors with any number of corresponding levels to individual columns. This can be set up in the following or alternatively be achieved by preparing the table header of our Excel sheet in a particular way. The following screenshot exemplarily describes a simple Excel file to set up a factor "strain" with the two levels "wild type" and "mutant" as well as a factor "temperature" with the two levels "30C" and "40C".

Example spreadsheet:

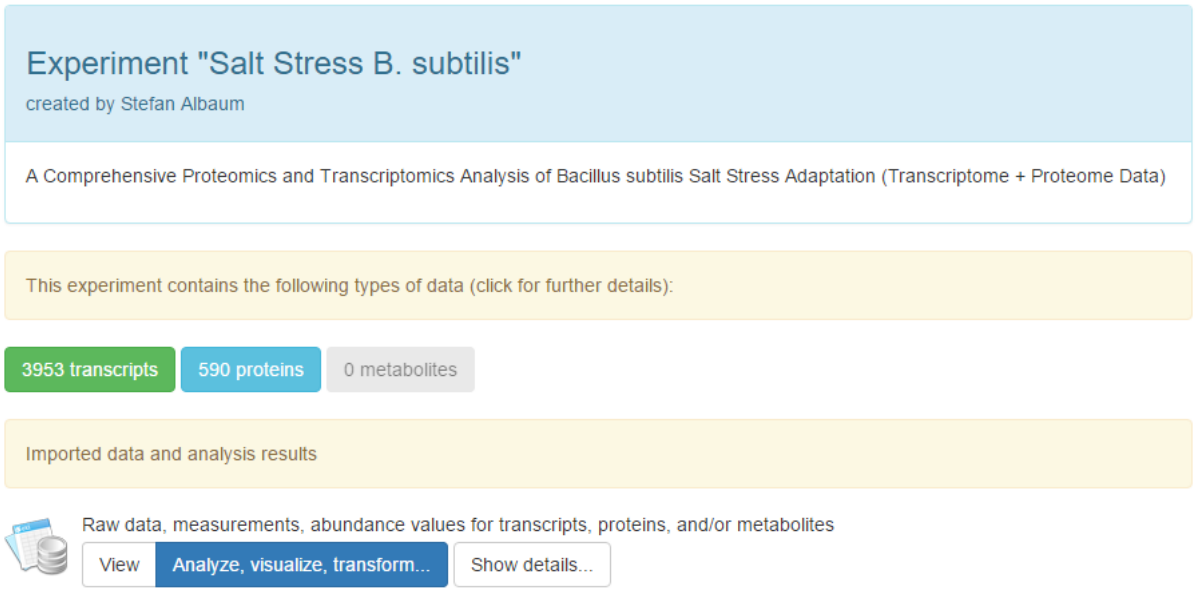
E15

	A	B	C	D	E
1	**Name**	**Strain:wt;Temperature:30C**	**Strain:wt;Temperature:40C**	**Strain:mutant;Temperature:30C**	**Strain:mutant;Temperature:40C**
2	Transcript001	0,533646472	-0,2255475362	0,6179142178	0,5183471143
3	Transcript002	0,2722488327	-0,9171337349	0,9457851515	0,1308158739
4	Transcript005	0,141056126	-1,0626481189	0,727637912	-0,5263924843
5	Transcript006	0,1616117459	1,9013418057	-0,2087163967	2,1857208418
6	Transcript010	-0,1759354827	-0,3209094766	-0,0657012296	-0,0108158562
7	Transcript012	-2,4433106618	-0,7376143731	-1,212476956	-0,051554186
8	Transcript014	0,3746966615	1,8524375522	0,2684520721	0,0148016107
9	Transcript018	1,1542848544	-2,5602048985	-8,1542691863	
10	Transcript020	-0,4469625074	-0,9857638814	-0,1401371662	-0,2564038685
11	Transcript024	-0,3909584779		-0,5123711182	-0,390319488
12	Transcript025	-0,7606879273	-1,1387581025	0,3434446972	-1,1117290468

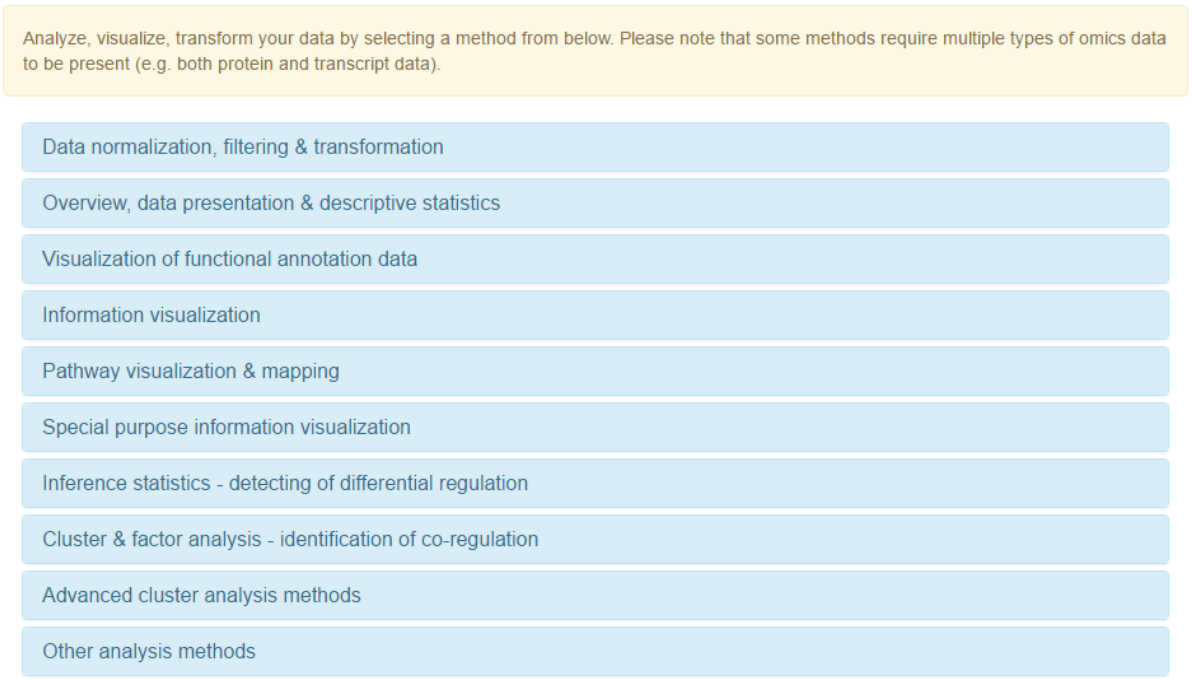
Example of a spreadsheet to import transcript data of an experiment comparing a wild type to a mutant at 30C as well as at 40C degrees.

Working with an experiment

After we imported data into our new experiment or if we chose to continue to work with an experiment that already had some data, the next page will give an overview over the experiment. It is also the starting point for all future analysis and visualizations.



We can see that this experiment contains both transcriptomic and proteomic data, but no metabolomics data. If we click on the respective buttons, we can browse through the raw data. The **Analyze, visualize, transform…** button will bring us to the control center for all operations in *Omics Fusion*:



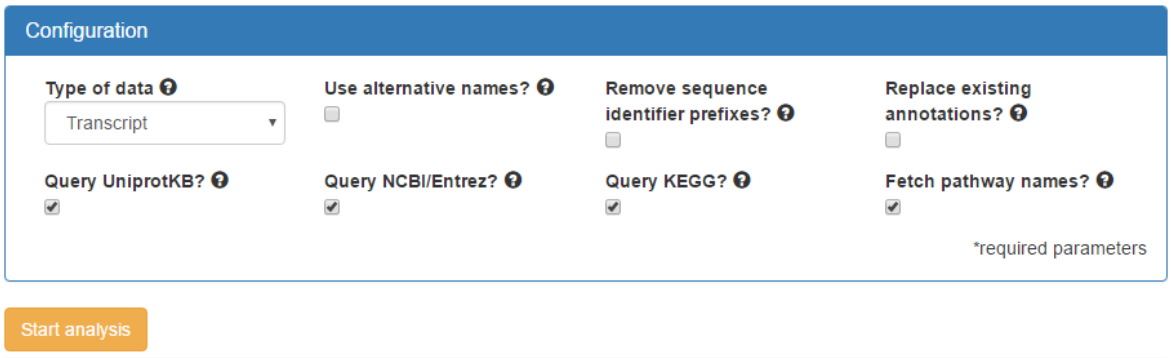
Each category contains a number of possible tasks, which we can apply to our data.

Data manipulation

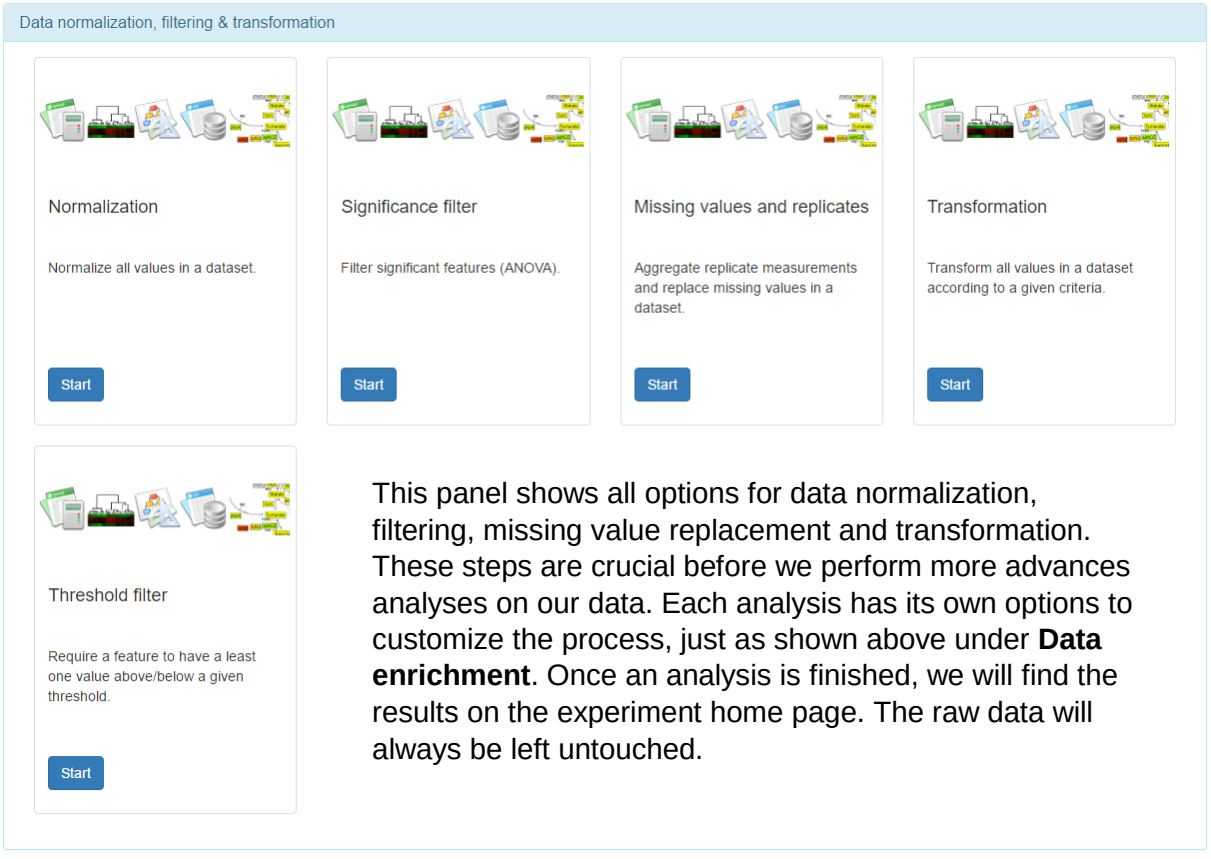
There are multiple tools available to manipulate data, ranging from simple but crucial normalization and filtering steps to transformation and missing value replacement. Data can also be enriched by querying other databases like KEGG, UniprotKb or NCBI/Entrez.

Data enrichment

To start the process of searching other databases for functional annotations of our data, we click on the **Enrich data** button, as shown on page 4. We select the type of data we want to enrich and customize which databases should be queried. Once we are satisfied with the setup, we click on **start analysis.** The process will probably take a couple of minutes - the progress can be seen in the job list.



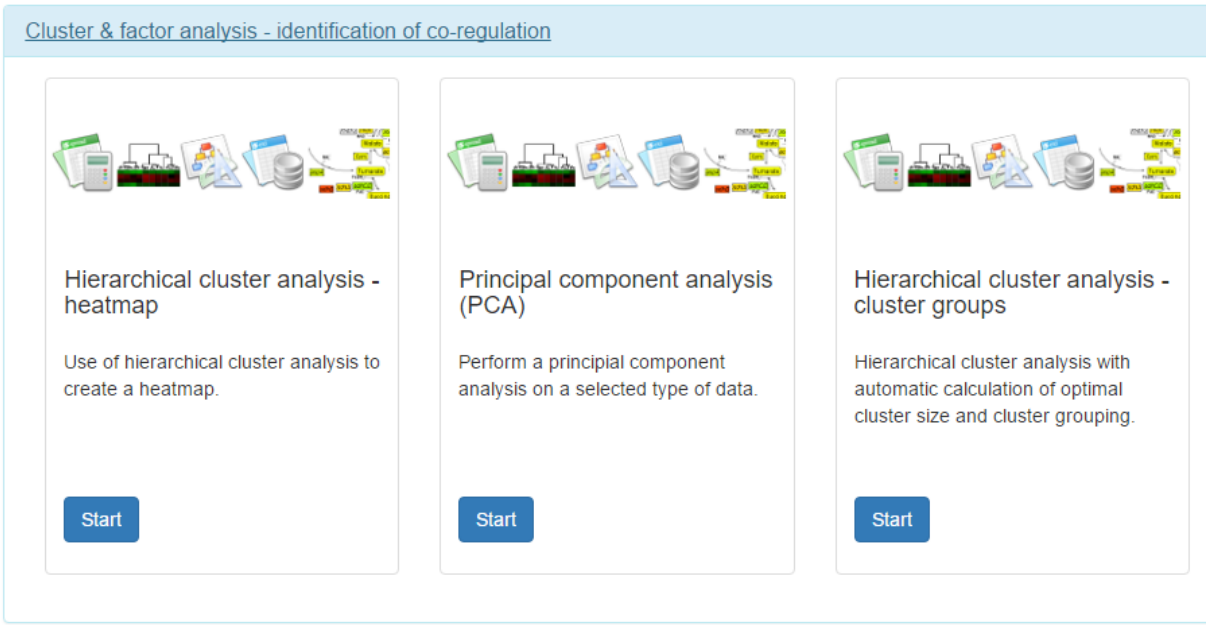
Data normalization, filtering & transformation



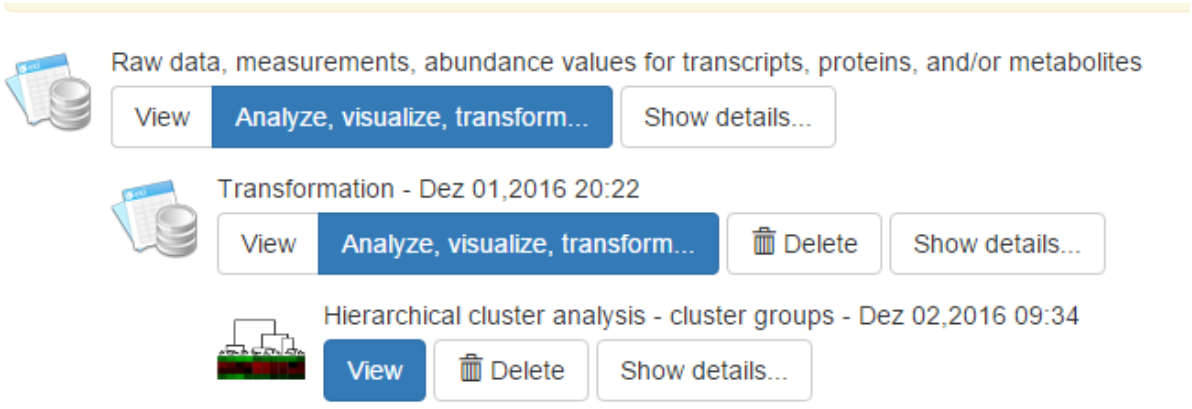
This panel shows all options for data normalization, filtering, missing value replacement and transformation. These steps are crucial before we perform more advances analyses on our data. Each analysis has its own options to customize the process, just as shown above under **Data enrichment**. Once an analysis is finished, we will find the results on the experiment home page. The raw data will always be left untouched.

Data analysis

Omics Fusion offers tools for descriptive statistics and distribution analysis to get an overview over the data, but also analysis of variance (ANOVA) for robust statistical testing. Furthermore, besides other classical methods like principial component analysis (PCA), *Omics Fusion* offers a hierarchical cluster analysis with automatic calculation of optimal cluster size and cluster grouping. This hierarchical clustering can be performed on data from multiple omics fields, grouping transcriptomic, proteomic and metabolomic data points with a similar signature. This facilitates the discovery of similar expression patterns throughout experiments from different omics fields. In order to perform a hierarchical clustering, click on **Analyze, visualize, transform...** on the experiment page and select the **Cluster & factor analysis** category.

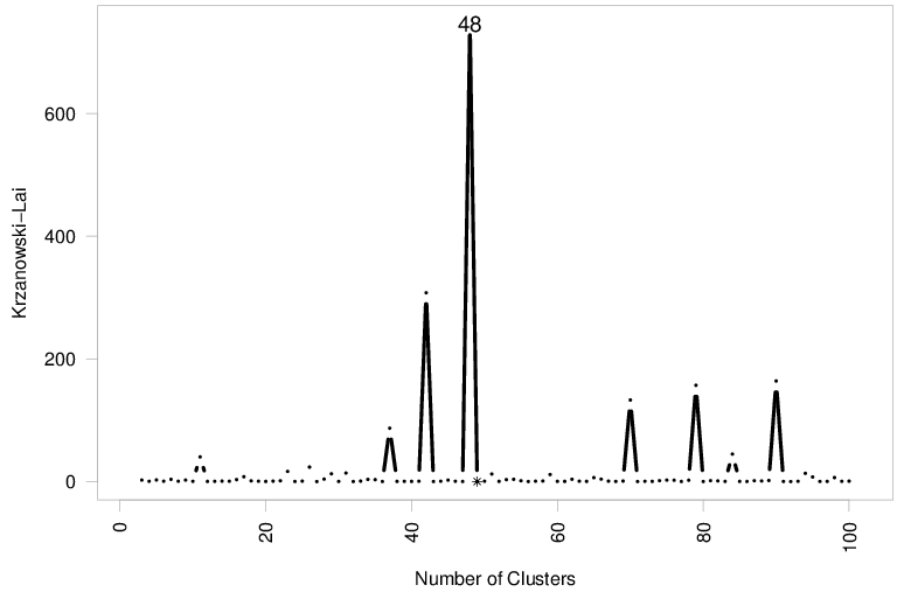


Let's start a hierarchical cluster analysis with automatic calculation of optimal cluster size and cluster grouping, as seen on the right. We will again have the option to change the parameters to fit our data. We leave the standard parameters for now and after the job is finished, the results can be found on the experiment page.

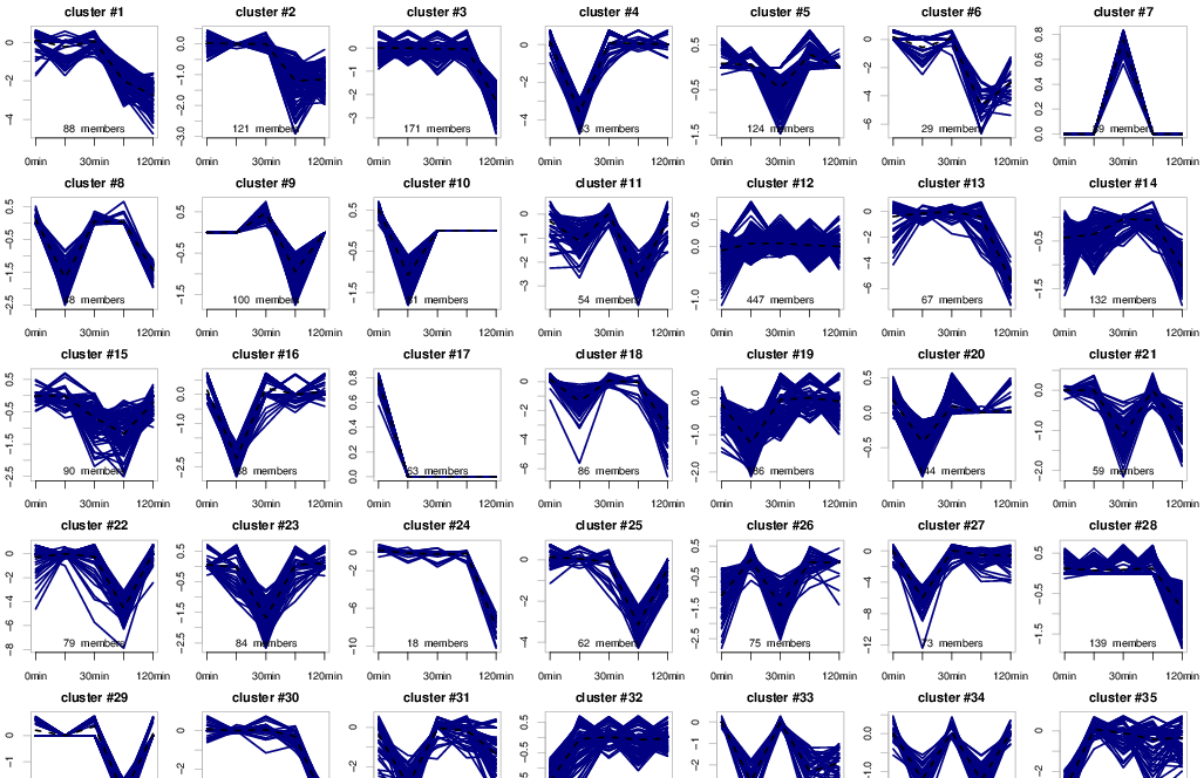


This view also represents the workflow: the raw data has been transformed and then a hierarchical clustering has been performed on the transformed data.

The hierarchical clustering automatically estimates an appropriate number of clusters based on the Krzanowski-Lai index.



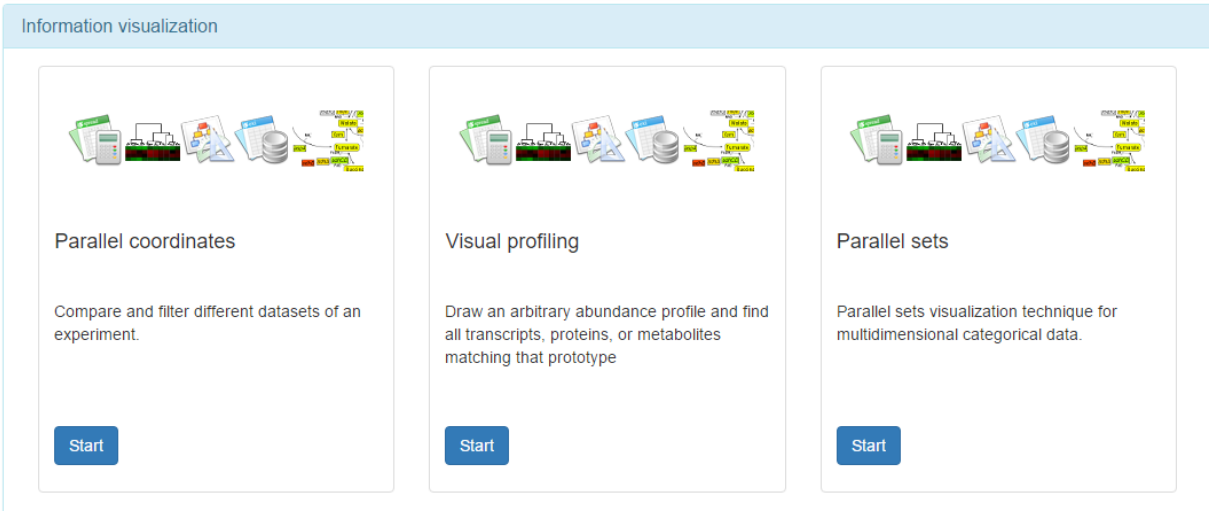
In case of the *B. subtilis* salt stress example dataset, the optimal number of clusters is calculated to be 48. The results can be explored within *Omics Fusion*, as well as exported and downloaded to our local computer using the respective button in the menu on the left.



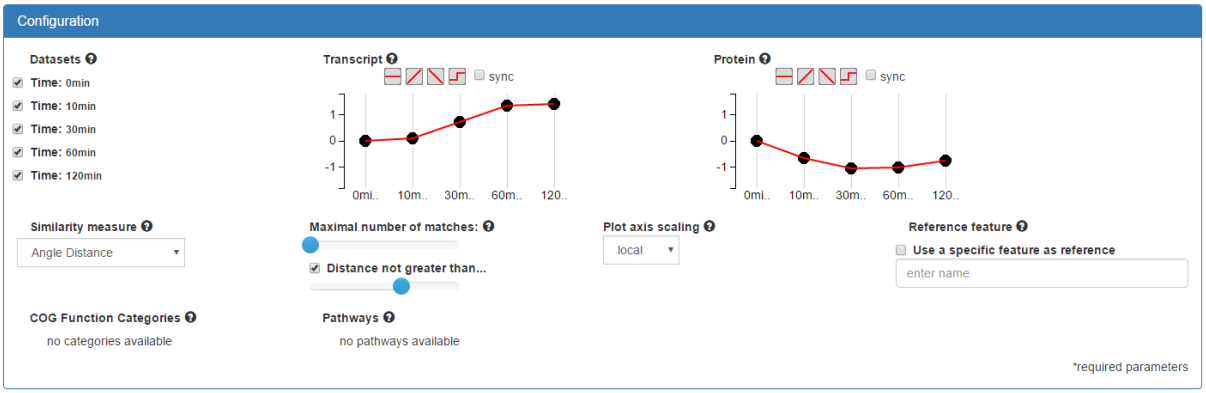
Detail of the clustering overview of the *B. subtilis* salt stress example dataset within *Omics Fusion*.

Visualization methods

Omics Fusion offers an increasing number of ways to explore and visualize data. A few examples are box plots, scatter plots, parallel coordinates or parallel sets. Beyond that, users can choose from a number of custom visualizations that introduce new ways to look at data from different omics disciplines. An example for that is a method termed “visual profiling”, which allows users to manually draw an arbitrary abundance profile and find all transcripts, proteins, or metabolites matching that prototype. To find this method, again click **Analyze, visualize, transform...** on the experiment page and select the **Information visualization** category.



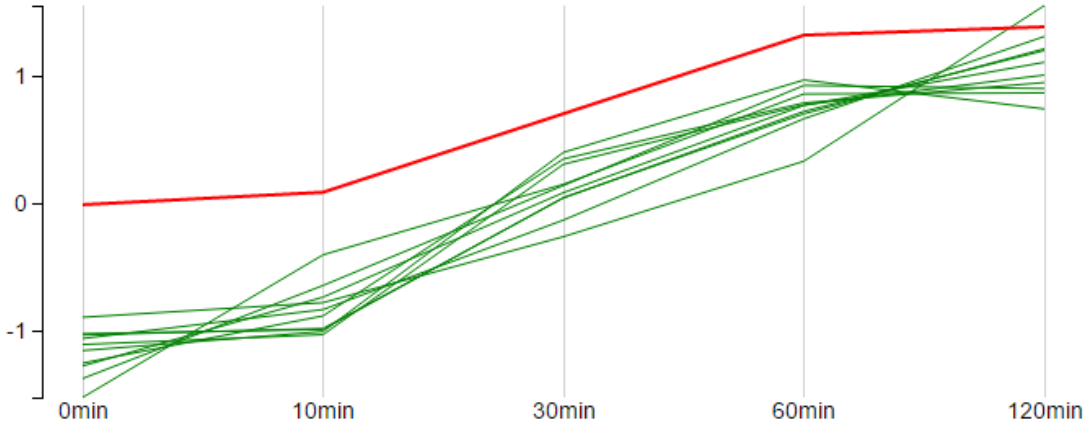
We can then look for example for transcripts that are upregulated and proteins that are downregulated in the dataset. We can draw the respective expression pattern in the configuration panel for this analysis.



For all information visualization methods, the results are shown on the same page to facilitate interactive exploration. We can directly change the configuration in the panel above, if we are not happy with the result. To save a visualization, we can click on the respective button:

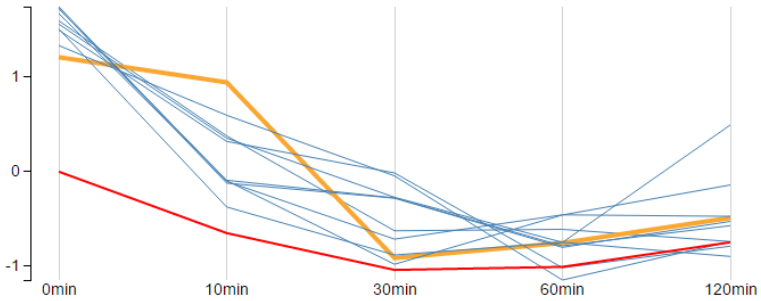
The result of our visual profiling is shown below.

Transcripts



We also get a list of the transcripts and proteins that match the pattern and additional information is shown when we hover the mouse over the respective entry in the list.

Proteins



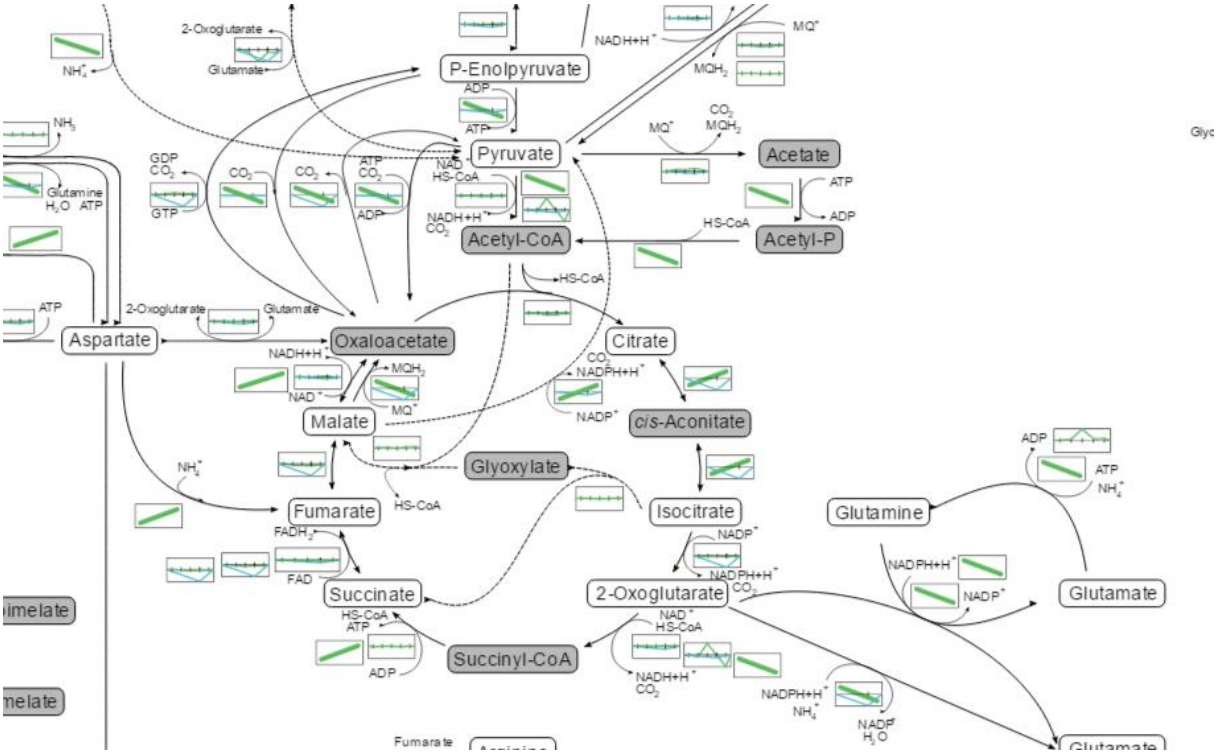
Values: 1.21061, 0.94512534, -0.91326714, -0.75397646, -0.48849174

10 best match 558 proteins

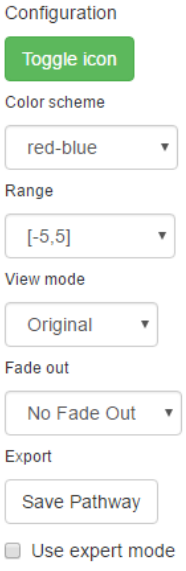
Name	Description	AltId	Dist.	Annot.
purL	...	/	1.1226854	info
fliY	...	/	1.1528409	info
rpsD	...	/	1.1651363	info
thiC	...	/	1.1889869	info
rpmC	...	/	1.2237095	info

Pathway map

The pathway viewer component implemented within the *Omics Fusion* framework enables the mapping of complete polyomics datasets on metabolic pathway images. Customized pathway maps can be easily imported as SVG-files and the interactive visualization provides different levels of highlighting important aspects of the data, including stylized icons for different expression patterns or a heatmap representation.



This shows the metabolic pathway of *C. glutamicum* with experimental data from a heatshock experiment mapped onto it. Such a pathway has to be imported as a SVG-file, containing additional attributes for the correct assignment of data points according to the Systems Biology Markup Language (SBML) standard. We have several options to customize the pathway view:



Change the icon from  to 

Change the color scheme to increase readability in some cases, for example color-blindness

Adjust the range of displayed data

Change the view mode (**Improved** removes unnecessary lines, **Black** makes the background dark, **Regulation** highlights regulation patterns)

Different levels of fading out text and line elements to improve visibility of the mapped data

Save the pathway to our local computer

Additional options to change thresholds for correlations, colours, etc.

Thank you for your interest in *Omics Fusion*.

If you have any questions, please don't hesitate to contact us:

fusion-service@cebitec.uni-bielefeld.de

Written by: Benedikt G. Brink, Bielefeld University, bbrink@cebitec.uni-bielefeld.de

Omics Fusion lead developer: Stefan P. Albaum, Bielefeld University, alu@cebitec.uni-bielefeld.de

This work is partially funded by the German-Canadian DFG International Research Training Group GRK 1906 and received financial support from the BMBF in the frame of the de.NBI initiative [FKZ 031A533A]. We further wish to thank the BRF system administrators for expert technical support and Annika Seidel, Nils Kleinbölting, Sonja Klingberg, Daniel Blume, Peter Belmann, Denis Kramer, Julia Gierens, Stefan Biermann, Yannic Kerkhoff, Carina Wenzel and Ruben Christian Hamann for their invaluable contributions.