

# Biogenity

#Demo / June 2022

## Bioinformatics Data Analysis

Tissue study

Your Global Partner in Omics

## A short introduction

---

This report contains an overview of the analysis made on your data. We have based the analysis on our state of the art analytic pipeline designed to give you the best foundation for understanding your data.

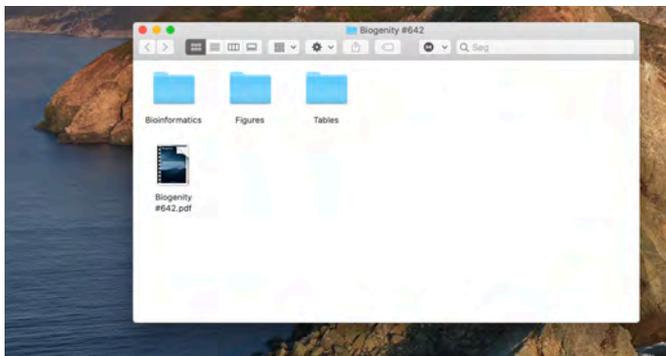
### Reading guide

A few tips to ensure the best understanding of the analysis:

1. The report functions as a walk-through, where every page shows a result of the analysis or pre-analysis.
2. Results are displayed by either a table or a figure followed by a thorough description.
3. You can always examine the original file. For each figure or table shown in the analysis, we have integrated the 'Pathfinder.' It lets you know where to find the shown object in the compressed file received. Example of such is displayed below:

"Figures/Mean log2 (Expression) across samples histogram.pdf"

Explained: The "Mean log2 (Expression) across samples histogram.pdf" can be located in the figures-folder provided in the compressed file.



Example of "compressed file"

### Thank you for choosing us

We appreciate the confidence you have placed in us, and we look forward to providing you with the best possible service in the future.

Best regards  
The Biogenity team

## Table of contents

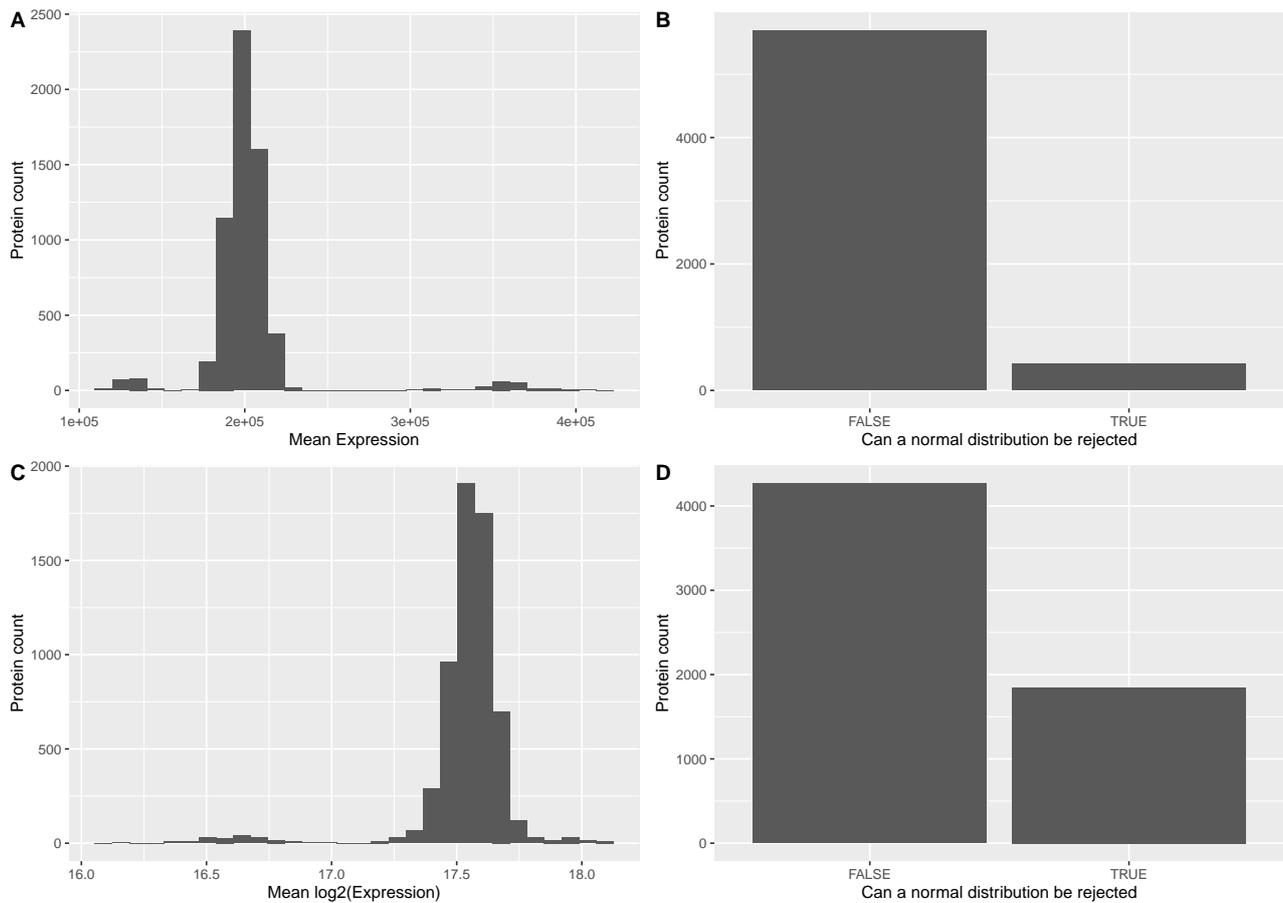
---

4	Pre-analysis
8	Cluster analysis
10	Statistical analysis
12	Bioinformatics
24	Reference list

**“The goal is to turn data into information,  
and information into insight.”**

**– Carly Fiorina**

## Pre-analysis



You can find the figure at:  
[Tissue\\_study analysis/Figures/Transformation overview.pdf](#)

### Data distribution and Transformation

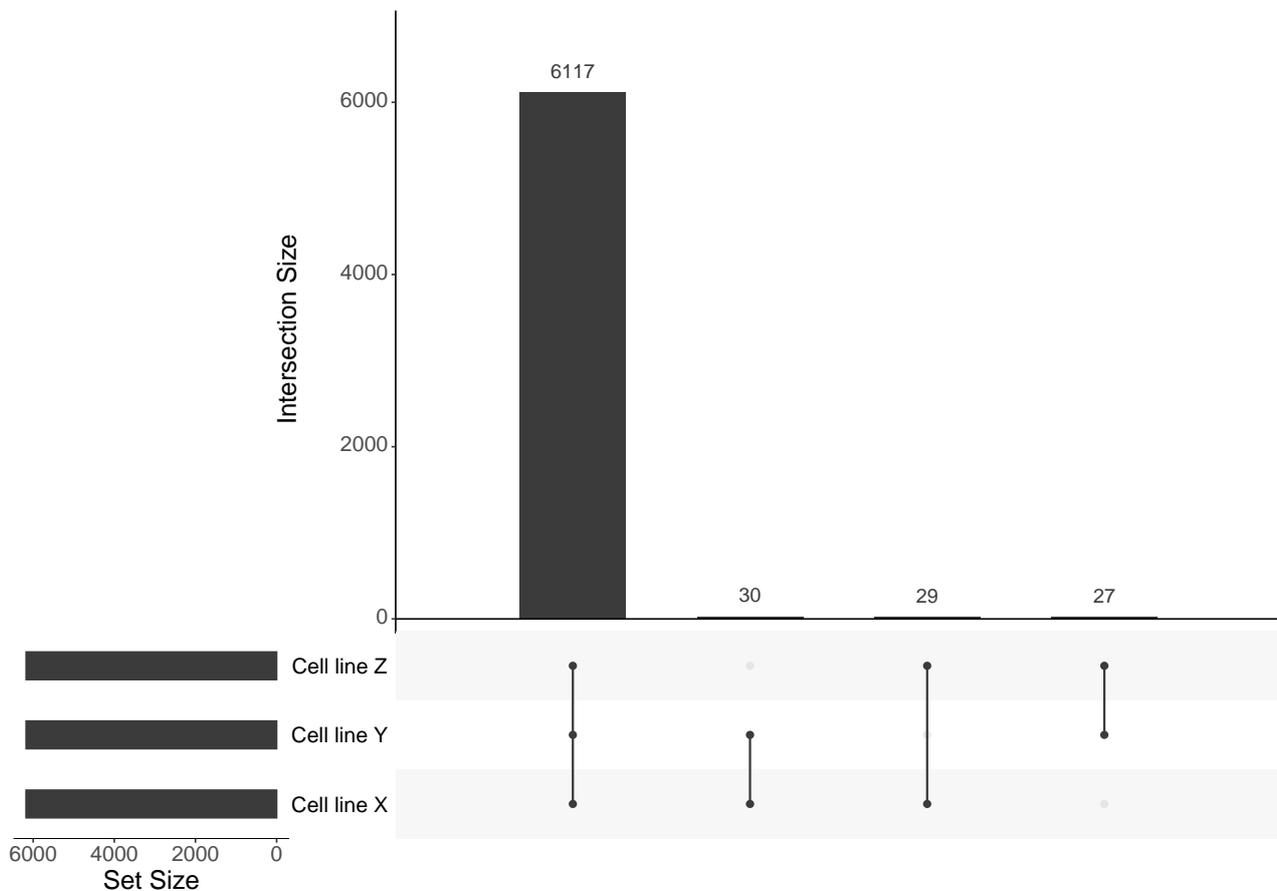
The data was filtered, so only proteins with at least 2 unique peptides were included. The majority of the protein expression should exhibit a normal distribution as it is a requirement for some statistical models. Often a log<sub>2</sub> transformation can be used to increase the number of expressions that demonstrate a normal distribution. Therefore, the effect of a log<sub>2</sub> transformation on the data distribution was investigated and illustrated by the histograms above (Figures A and C). A Shapiro Wilk normality test was used to count the number of proteins for which a normal distribution could be rejected.

A normal distribution could be rejected for 425 proteins before transformation and 1848 after transformation, as illustrated on bar charts above (Figures B and D).

As the number of expressions that rejected the normal

distribution hypothesis was higher after log<sub>2</sub> transformation, the data was not transformed before further processing. The individual plot and this figure (combined plots) can be found in the *Figures* folder.

## Pre-analysis



You can find the figure at:  
[Tissue\\_study analysis/Figures/UpSet diagram.pdf](#)

### Unique Identifications

The identification of unique proteins within a single group can potentially indicate actual differences between the groups. Owing to the nature of mass spectrometers, proteins can be missing due to chance. Thus, the evaluation of unique identifications should be based on their abundance, their biological function, and the missingness in the dataset as a whole. The identifications should be treated as potential targets, that are identified in a less stringent manner than the targets that are statistically significant.

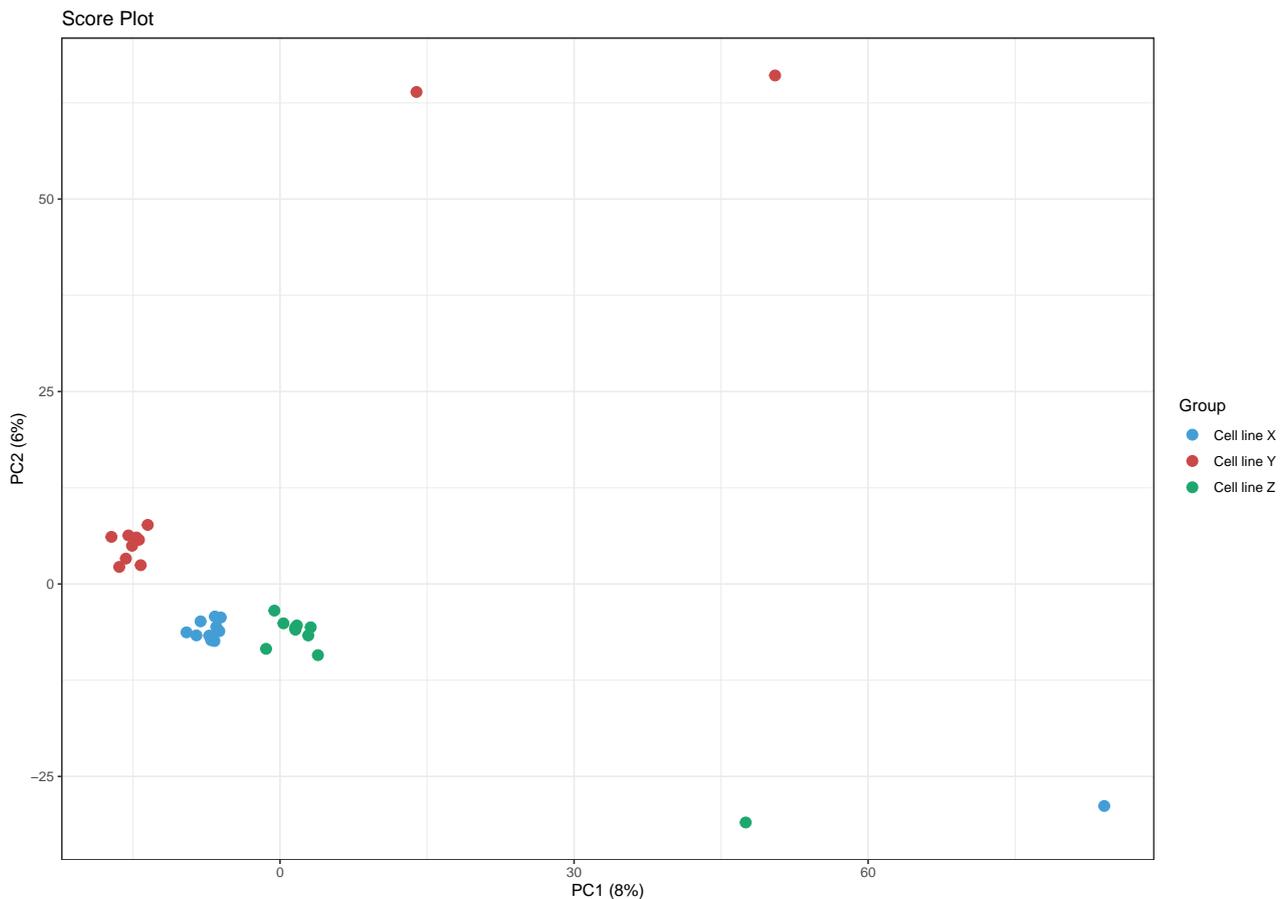
The unique identifications within groups were determined uniquely identified when the proteins were identified in at least half of the samples within a group, and completely missing in another group.

The findings are illustrated in the plot above. This plot is

known as an Upset plot which displays the various overlaps between different groups (The lower dots and lines), and how many proteins are in the different combinations of overlaps (The upper bar chart). It is an alternative to Venn diagrams as it provides an improved overview when comparing multiple groups. In this analysis 6203 identifications were found and 86 were identified as being unique.

Venn diagrams (color and greyscale) and Upset diagrams can be found in the *Figures* folder.

## Pre-analysis



You can find the figure at:  
[Tissue\\_study analysis/Figures/PCA of data with Group annotation.pdf](#)

### Data visualization with Principal Component Analysis (PCA)

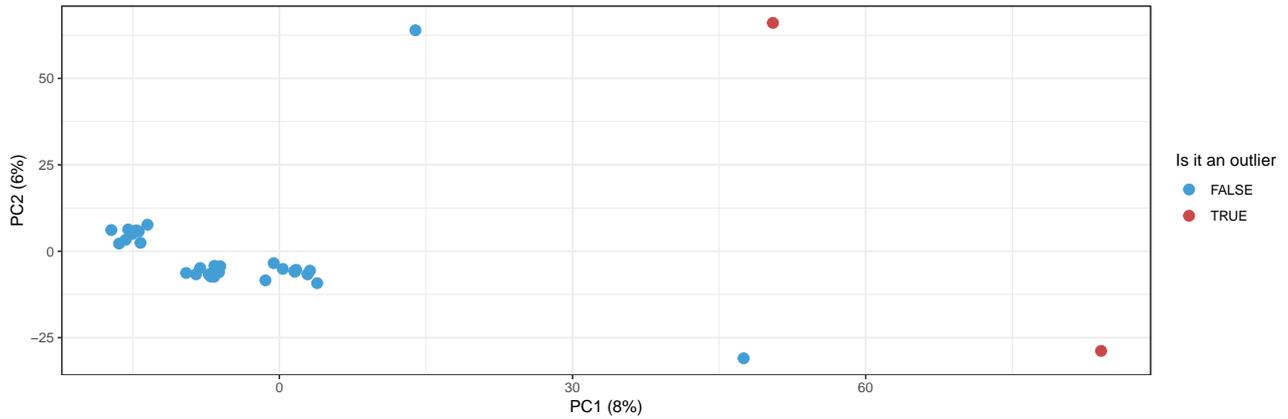
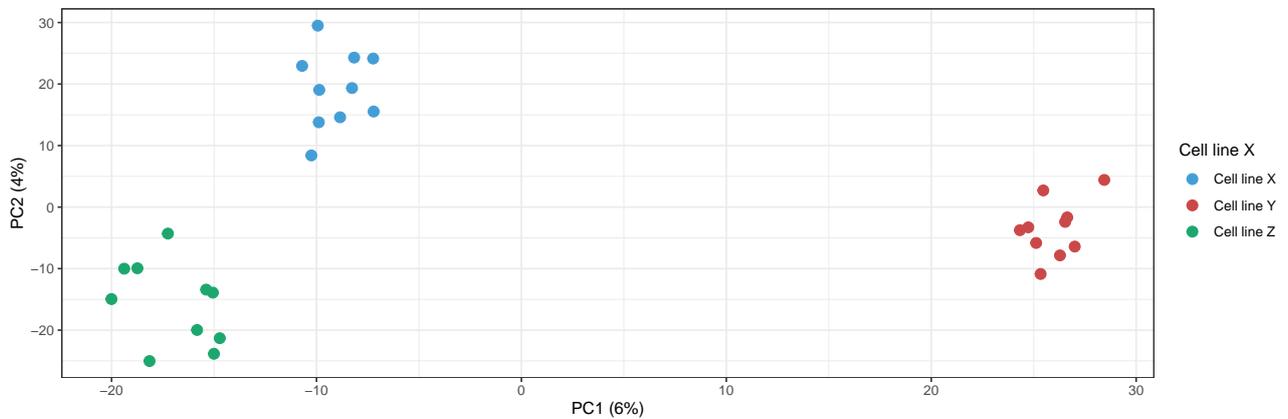
A Principal Component Analysis was performed to investigate and visualize the data. Principal components 1 and 2 were plotted to give a visualization of the data. Principal Component Analysis converts possibly correlated variables into a set of linearly uncorrelated variables, enabling visualization of high dimensional data in fewer dimensions with minimal information loss. It can be used to emphasize strong patterns in the data, and thereby potential outliers can be visually identified if such should be in the data.

This PCA plot was also constructed in an interactive form. Both plots can be found in the *Figures* folder.

Outlier analysis was made with a focus on finding sample error. The analysis was made under the assumption that the error would affect the total measurement of the

sample. Therefore, the sum of the intensity in each sample was calculated. Using a convenience function, potential outliers were detected based on their median absolute deviation, thus samples more than 3 median absolute deviations away from the median were considered an outlier. 2 outliers were found.

## Pre-analysis

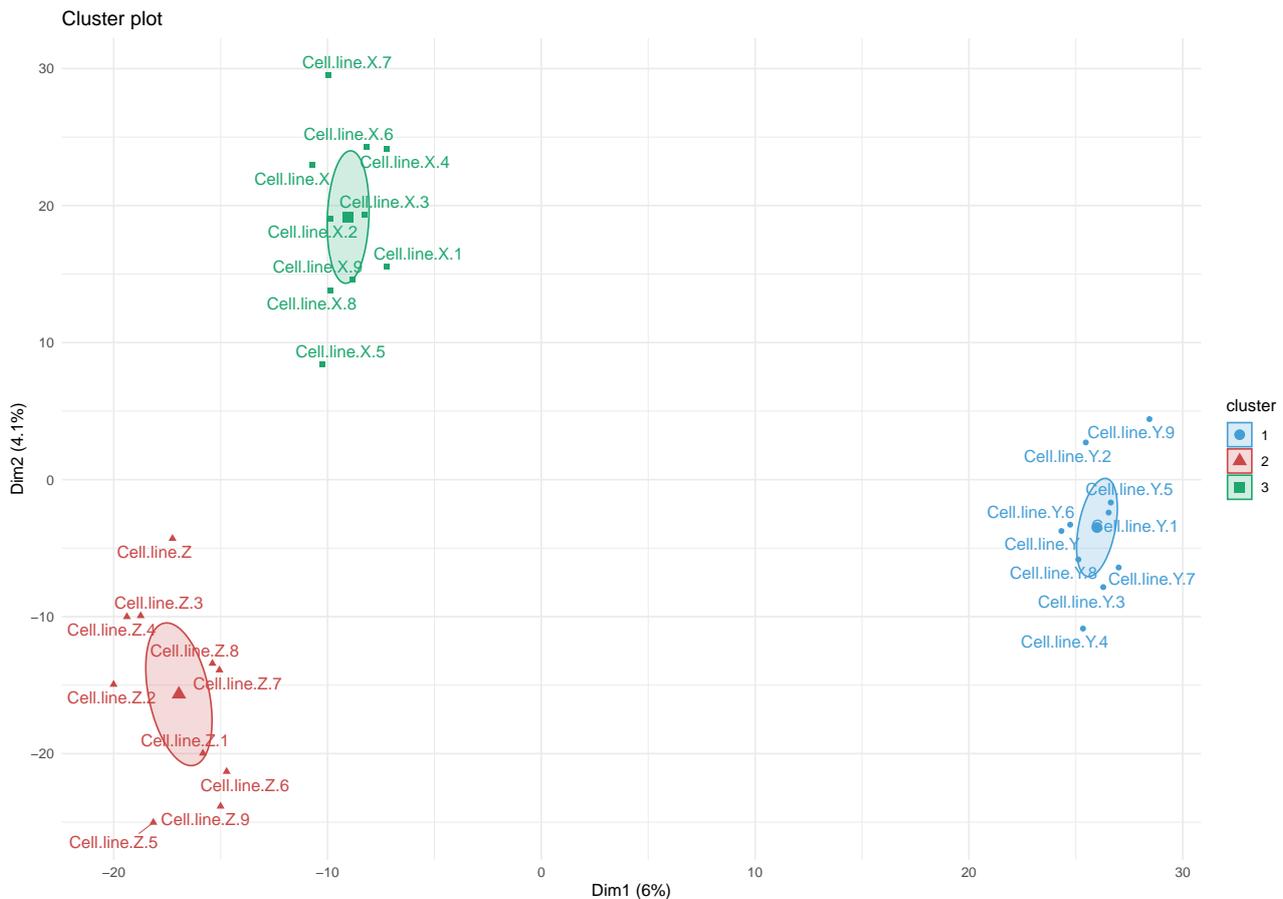
**A** Score Plot**B** Score Plot

You can find the figure at:  
[Tissue\\_study analysis/Figures/Outlier overview.pdf](#)

### Outlier analysis and filtration

All outliers were identified and plotted. The identified outliers were filtered before a new PCA analysis was performed. The plot with the outlier overview can be found in the *Figures* folder.

## Cluster analysis



You can find the figure at:  
[Tissue\\_study analysis/Figures/Kmeans clusters of data with Group annotation.pdf](#)

### Partitioning Around Medoid - K-means clustering

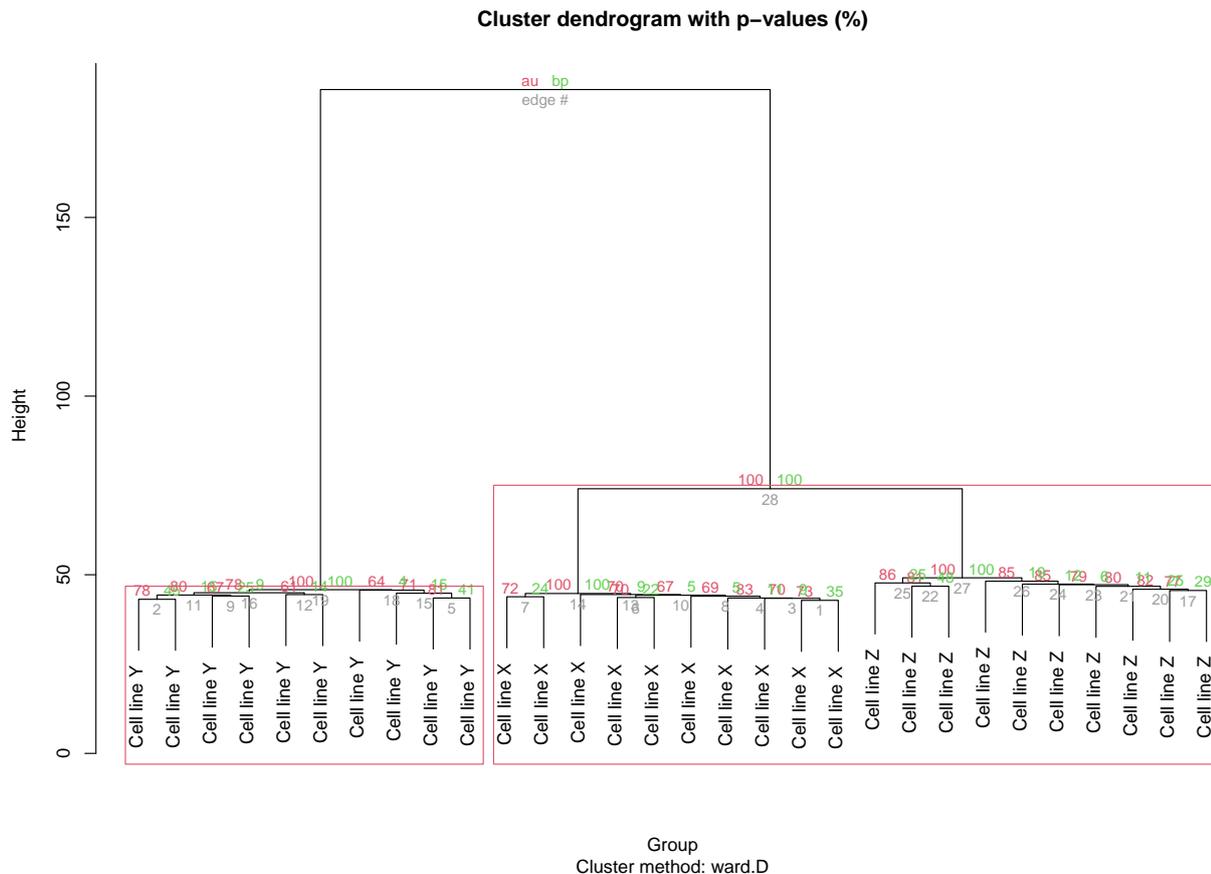
In order to investigate and visualize potential clusters in the data, clustering analysis was performed both using Partitioning Around Medoids (PAM), a more robust variant of K-means clustering, and the Ward method. PAM was applied to test if sample group related clusters could be found unsupervised, since this is an unsupervised machine learning algorithm. The first clustering analysis presented is K-means clustering, which is a divisive (top-down) approach to generate clusters.

Briefly, this clustering method initially considers all points in the dataset as belonging to one single cluster, which is then divided into two least similar clusters, etc. For the distance measure, the Euclidean method was utilized. Silhouette analysis was performed for 1-10 clusters to determine the best number of clusters, and then Partitioning Around Medoids (PAM) was used for clustering anal-

ysis. The data points inside the ellipses are the points that are assigned to that cluster with a 95% confidence interval for each cluster, if any.

K-means clusters (both with Group and Sample annotations) can be found in the *Figures* folder.

## Cluster analysis



You can find the figure at:

[Tissue\\_study analysis/Figures/Ward clusters of data with Group annotation.pdf](#)

### Ward clustering

A complementary clustering method is hierarchical clustering, such as the Ward method illustrated as a dendrogram above. This method is known as a bottom-up method which initially considers all datapoints to be a cluster of their own. From there it merges the points into clusters by the rule that it minimizes the within cluster variance as measured by the Euclidian distance. The length of the vertical branches indicates how similar the clusters are. Thus, clusters that have a short vertical line before merging with another cluster exhibit a higher degree of similarity compared to clusters that have a longer line prior to merging.

The plot contains two p-values: the red Approximately Unbiased (AU) and the green Bootstrap Probability (BP) values. Essentially, these two values are based on sampling the data in different ways 1000 times. Thus, a value

of 0.31 means that a specific cluster appeared in 310 of the 1000 samplings. The AU utilizes a different sampling and scaling version of BP which is more unbiased than the BP. Clusters with an AU equal to or higher than 0.95 are marked by a red square and are considered highly robust.

Ward clusters were built both with Group and Sample annotations and can be found in the *Figures* folder. Interactive Heat Maps were also constructed to graphically represent the data. These can be found in the *Figures* folder as well.

## Statistical analysis

Comparisons	p-value $\leq$ 0.05	Adj. p-value $\leq$ 0.05	Adj. p-value $\leq$ 0.05 + 30% regulation	Log2 fold change  $\geq$ 1
Cell line Y vs Cell line X	500	283	283	227
Cell line Z vs Cell line X	517	233	231	103
Cell line Z vs Cell line Y	390	209	207	178

You can find the figure at:  
[Tissue\\_study analysis/Tables/Statistics.txt](#)

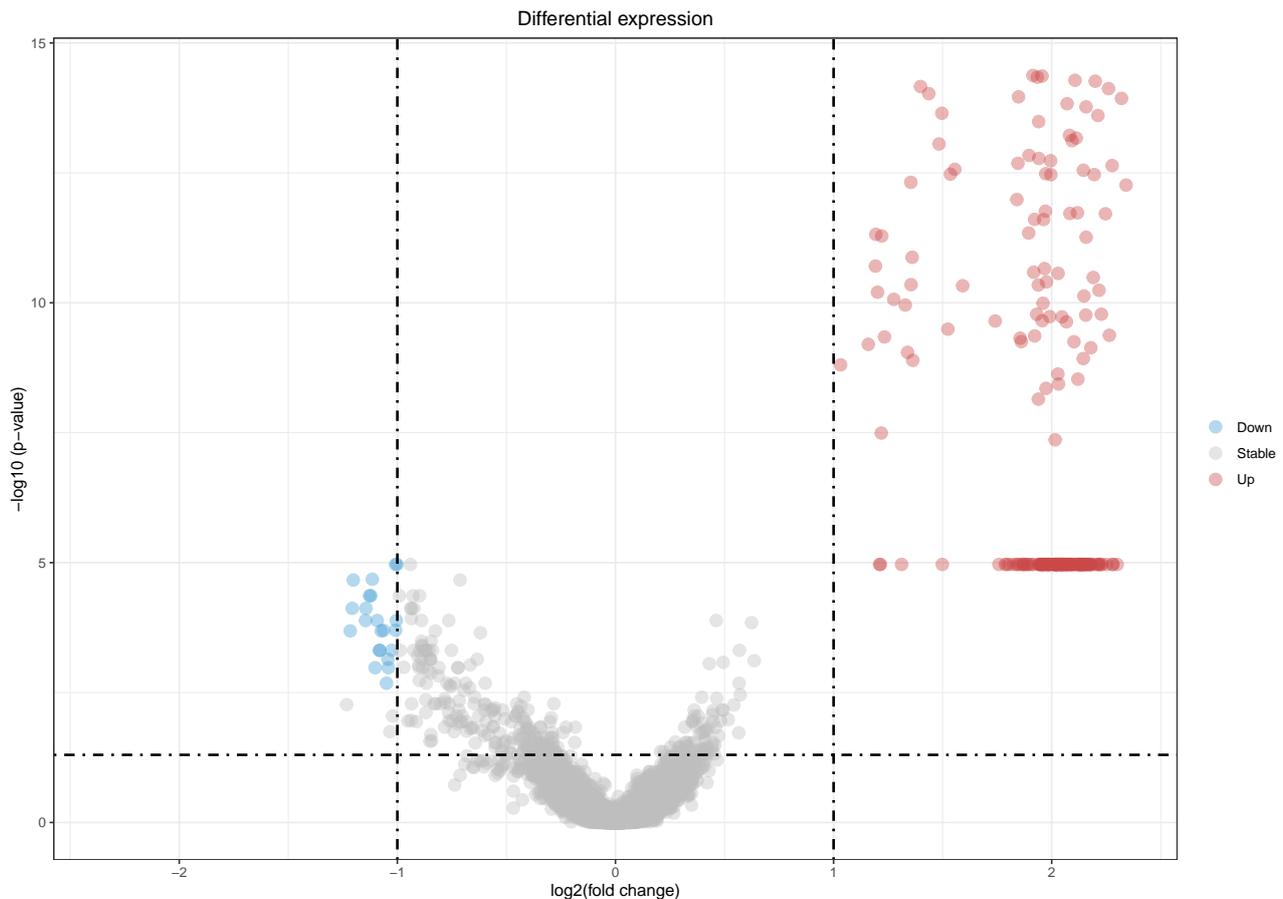
### Overview of regulations

To ensure high stringency, the proteins were filtered prior to statistical testing. Only proteins that were identified in at least half of the samples of a single sample group were included in the analysis. Each protein expression was tested for normal distribution by the Shapiro-Wilk test.

If the Shapiro-Wilk test returned a p-value below 0.05, which means that the residuals are normally distributed, then a parametric test such as Analysis of Variance (ANOVA) test, followed by a Tukey's post hoc test can be applied. Conversely, if the residuals were not normally distributed (p-value  $\leq$  0.05), then a non-parametric test such as the Wilcoxon-Mann-Whitney test (referred to as Wilcox test in the report) was applied. In the statistical result spreadsheets, provided in the *Tables* folder, the recommended test to use according to the p-values obtained in the Shapiro-Wilk test is noted in the "*Recommended statistical test*" column.

In total, 18351 comparisons were made, 1407 of which had a p-value below or equal to 0.05. Due to the large number of comparisons, correction for multiple comparisons was performed using the Benjamini-Hochberg procedure. 725 proteins retained an adjusted p-value equal to or below 0.05, 722 of which were regulated by more than 30%, and distributed across 3 group comparisons (Cell line Y vs Cell line X, Cell line Z vs Cell line X, Cell line Z vs Cell line Y).

## Statistical analysis



You can find the figure at:  
[Tissue\\_study analysis/Figures/Volcano plot/Volcano plot of Cell line Y vs Cell line X statistical comparison.pdf](#)

### Volcano plot

Volcano plots were constructed to facilitate an easy overview of the fold changes and the p-values for the different comparisons. The plot above is for Cell line Y vs. Cell line X. On the Y axis is the  $-\log_{10}$  transformed p-values. The higher the value on the y axis, the lower the underlying p-value is. The x axis represents the  $\log_2$  transformed fold change.

The horizontal dashed line represents an unadjusted p-value of 0.05, and thus anything above the line has a p-value below 0.05. Similarly, the vertical lines represent a log fold change of 1 and -1 which corresponds to a fold change of 2 and -2 respectively.

Each dot represents a protein. The dots are colored if the adjusted p-value is below 0.05, and the color corre-

sponds to the direction of the regulation; blue if the  $\log_2$  fold change is below -1 and red if the  $\log_2$  fold change is above 1. Similar to the previous statistical results, the tests are corrected for multiple corrections using the Benjamini Hochberg procedure.

In the plot above, 205 proteins were significantly upregulated and 22 proteins were significantly downregulated.

The volcano plots for the different comparisons can be found in the *Figures/Volcano plot* folder.

## Bioinformatics

ID	Description	Gene Ratio	Bg Ratio	Count
R-HSA-8856828	Clathrin-mediated endocytosis	92/483	146/10867	92
R-HSA-199977	ER to Golgi Anterograde Transport	90/483	154/10867	90
R-HSA-6811442	Intra-Golgi and retrograde Golgi-to-ER traffic	99/483	202/10867	99
R-HSA-948021	Transport to the Golgi and subsequent modification	94/483	185/10867	94
R-HSA-9007101	Rab regulation of trafficking	71/483	124/10867	71
R-HSA-8876198	RAB GEFs exchange GTP for GDP on RABs	62/483	90/10867	62
R-HSA-8856825	Cargo recognition for clathrin-mediated endocytosis	65/483	106/10867	65
R-HSA-446203	Asparagine N-linked glycosylation	96/483	304/10867	96
R-HSA-204005	COPII-mediated vesicle transport	48/483	68/10867	48
R-HSA-6807878	COPI-mediated anterograde transport	54/483	101/10867	54
R-HSA-199992	trans-Golgi Network Vesicle Budding	46/483	72/10867	46
R-HSA-8856688	Golgi-to-ER retrograde transport	59/483	133/10867	59
R-HSA-2132295	MHC class II antigen presentation	46/483	123/10867	46
R-HSA-432722	Golgi Associated Vesicle Biogenesis	33/483	56/10867	33
R-HSA-5694530	Cargo concentration in the ER	26/483	33/10867	26

You can find the figure at:  
[Tissue\\_study analysis/Bioinformatics/Cell line Y vs Cell line X/Pathway/Enrichment data.txt](#)

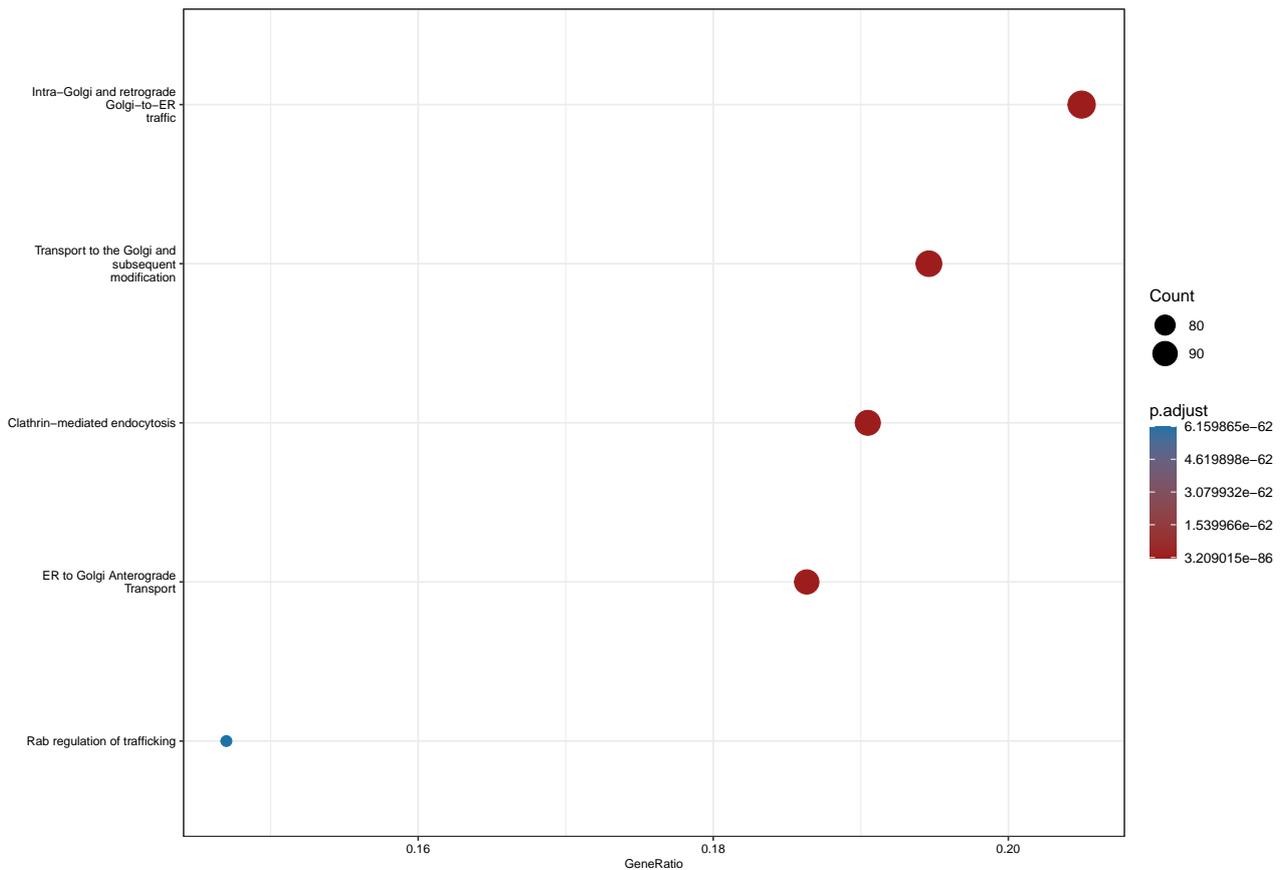
### Enrichment analysis

Pathway enrichments enable an easy overview of the biological pathways that are demonstrating some level of regulation. Here, the Reactome database [1] (please cite if results are used in a publication) was used for the pathway enrichment. In order to extract the most information on potentially regulated pathways, the proteins that had an unadjusted p-value below 0.05 in the group wise statistical test were used to populate the pathways. The enrichment results were controlled for multiple comparisons using the Benjamini Hochberg procedure and the summary is reported above.

The table is sorted by adjusted p-value from lowest to highest. For each pathway, the Gene Ratio, Bg Ratio and the Count is summarized. The Gene Ratio is the number of proteins out of the total number of proteins in the enrichment analysis that map to the pathway. Similarly, the Bg (background) Ratio is the number of proteins in that pathway out of the total number of proteins in the database. Essentially, these two numbers can yield information about the pathway coverage as the numerator of the Gene Ratio and the numerator of the Bg Ratio informs how many proteins were mapped, and how big the pathway is respectively. To simplify this, the column count is added, which directly informs how many proteins were mapped to the pathway. It is noteworthy to add that counts below 5 are considered low. Nevertheless, they are included to show where the proteins derive from.

Moreover, Enrichment analyses have been performed on the separate gene ontology (GO) terms; Molecular Functions, Cellular Components and Biological Processes. The combination of the pathway analyses and Gene Ontology terms should provide a broad overview of the biological differences between the tested sample groups. These analyses were performed using [2] and [3] (please cite if results are used in a publication). All of the tables and associated figures can be found in the *Bioinformatics* folder.

## Bioinformatics



You can find the figure at:  
[Tissue\\_study analysis/Bioinformatics/Cell line Y vs Cell line X/Pathway/Top 5 annotated potentially regulated Pathway.pdf](#)

### Pathway annotation of enrichment analysis illustrated by dot plot

The Top 5 enriched Reactome pathways with the lowest adjusted p-value are illustrated in the figure above for Cell line Y vs Cell line X. In total, 139 enriched pathways were found for Cell line Y vs Cell line X.

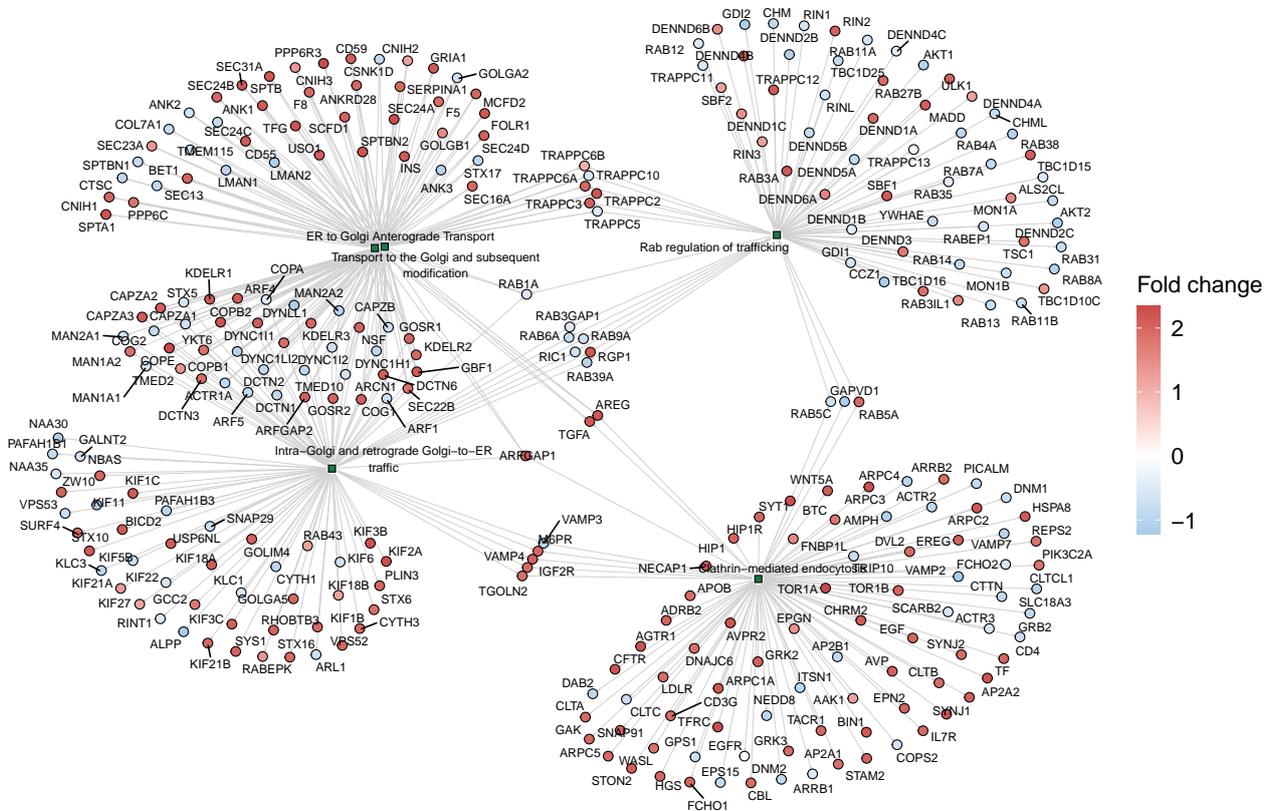
The Pathways are represented by dots which are colored by the adjusted p-value, and sized according to the count of proteins which were attributed to that pathway in the analysis. The Y axis represents the individual pathways, and the X axis represents the Gene ratio which is the ratio of protein regulations identified in that pathway out of the total number of protein regulations used in the test.

The proteins that make up the different pathways can be found in the *geneID* column in the excel file called *Enrichment data* which is in the pathways subfolder for the

group comparison Cell line Y vs Cell line X in the *Bioinformatics* folder.

the analysis and figures were made using [1] and [2] (please cite if results are used in a publication).

Cell line Y vs Cell line X  
Top 5 pathways; Cluster 1 / 1



You can find the figure at:  
[Tissue\\_study analysis/Bioinformatics/Cell line Y vs Cell line X/Pathway/Annotated top 5 potentially regulated Pathway - Protein interactions plot.pdf](#)

**Network plot depicting the linkages of proteins and pathways**

A network plot was made to visualize the associations between the proteins found in the enrichment analysis and biological concepts. It gives information on upregulation and downregulation of the proteins in the annotated pathways through the log2 fold change. To the right, a color scale represents the log2 fold change.

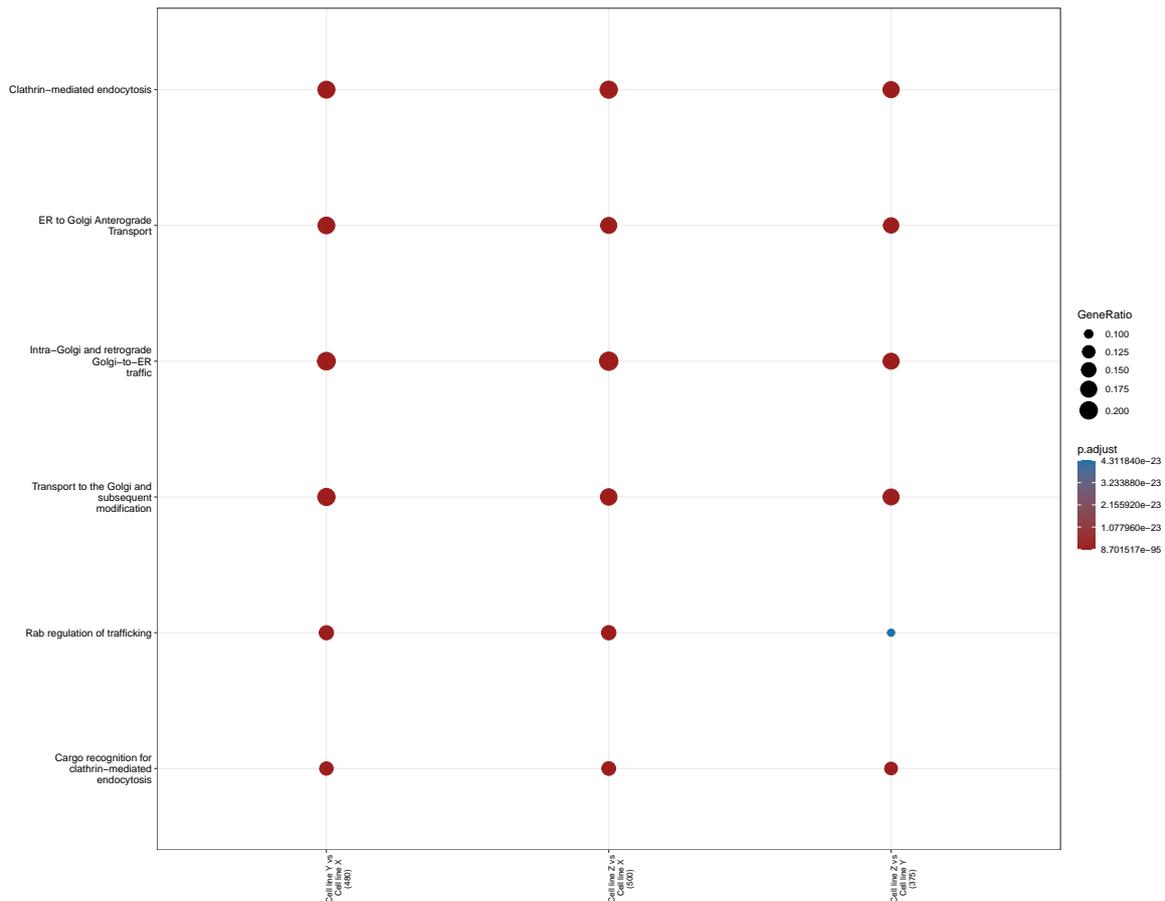
In the example above, the top 5 pathways found in the enrichment analysis are depicted. Each pathway is represented by green squares and is connected to their respective annotated proteins (colored circles) by grey lines

This illustration, showing the enriched pathways found for Cell line Y vs Cell line X, the proteins annotated to the pathways have been plotted as well, showing 234 anno-

tated proteins in these 3 pathways, 46 of which are annotated to multiple pathways and potentially could serve as links in-between the pathways. The top 5 annotated pathways are plotted in this example for Cell line Y vs Cell line X, but other plots with additional pathways can be found in the *Bioinformatics* folder.

Made using [1], [2] and [3] (please cite if results are used in a publication).

## Bioinformatics



You can find the figure at:

[Tissue\\_study analysis/Bioinformatics/Annotated potentially regulated Pathways across Groups.pdf](#)

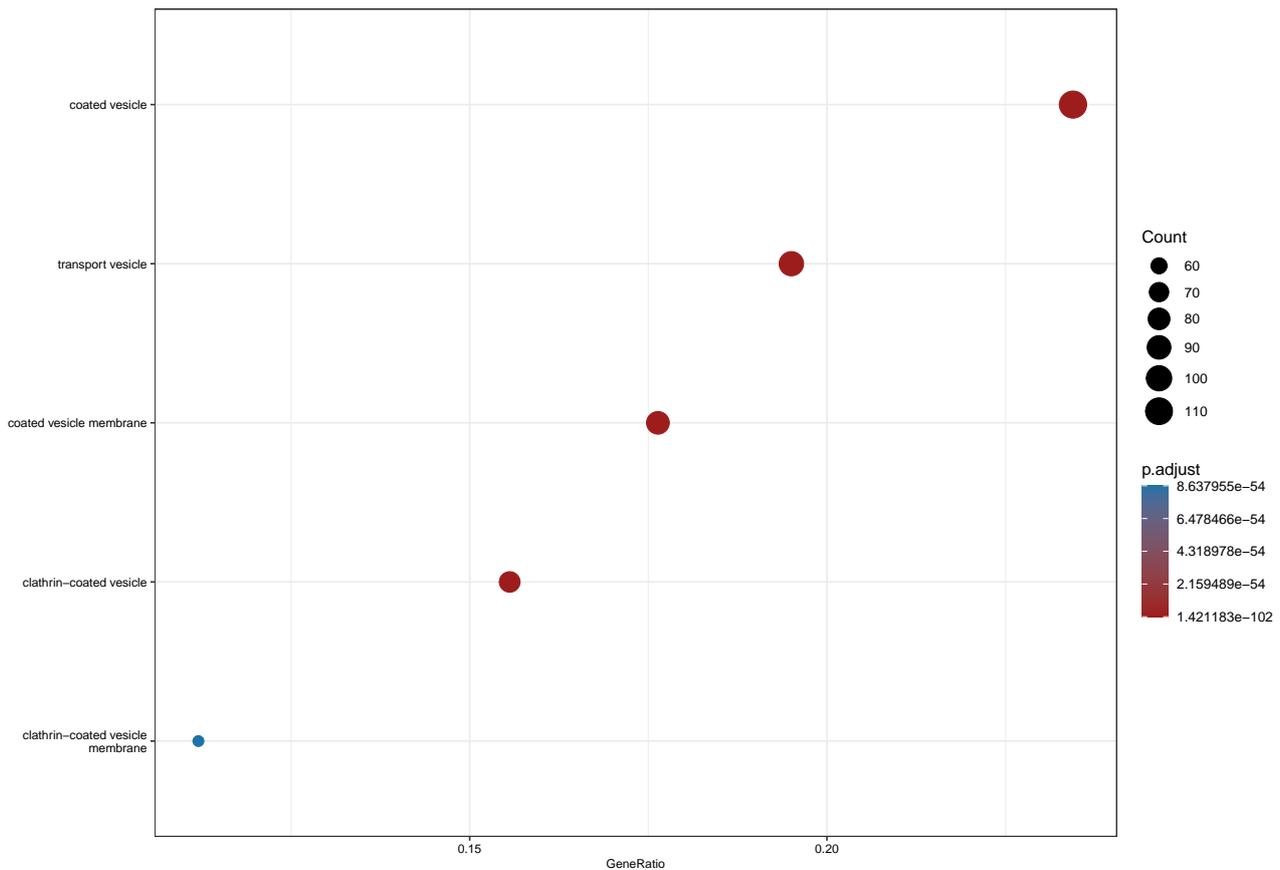
### Pathway annotation of enrichment analysis, comparison across sample group tests illustrated by dot plot

The comparison for the annotated pathways found within different sample group tests were also assessed and illustrated in a dot plot. The details from these enrichment analyses, such as the p-values, proteins and the ratios can be found in the *Bioinformatics* folder.

In this analysis, enriched terms were found in 3 comparisons (Cell line Y vs Cell line X, Cell line Z vs Cell line X, Cell line Z vs Cell line Y). The x-axis shows the different sample group tests as well as the number of regulations used for the enrichment analysis (within brackets). The y-axis shows the pathways that were annotated with the lowest adjusted p-value. The adjusted p-values are represented by the color of the dots and the gene ratio is represented by the dot size.

The figure was made using [1] and [2] (please cite if results are used in a publication).

## Bioinformatics



You can find the figure at:  
[Tissue\\_study analysis/Bioinformatics/Cell line Y vs Cell line X/Cellular Components/Top 5 annotated potentially regulated Cellular Components.pdf](#)

### Cellular Components annotation of enrichment analysis illustrated by dot plot

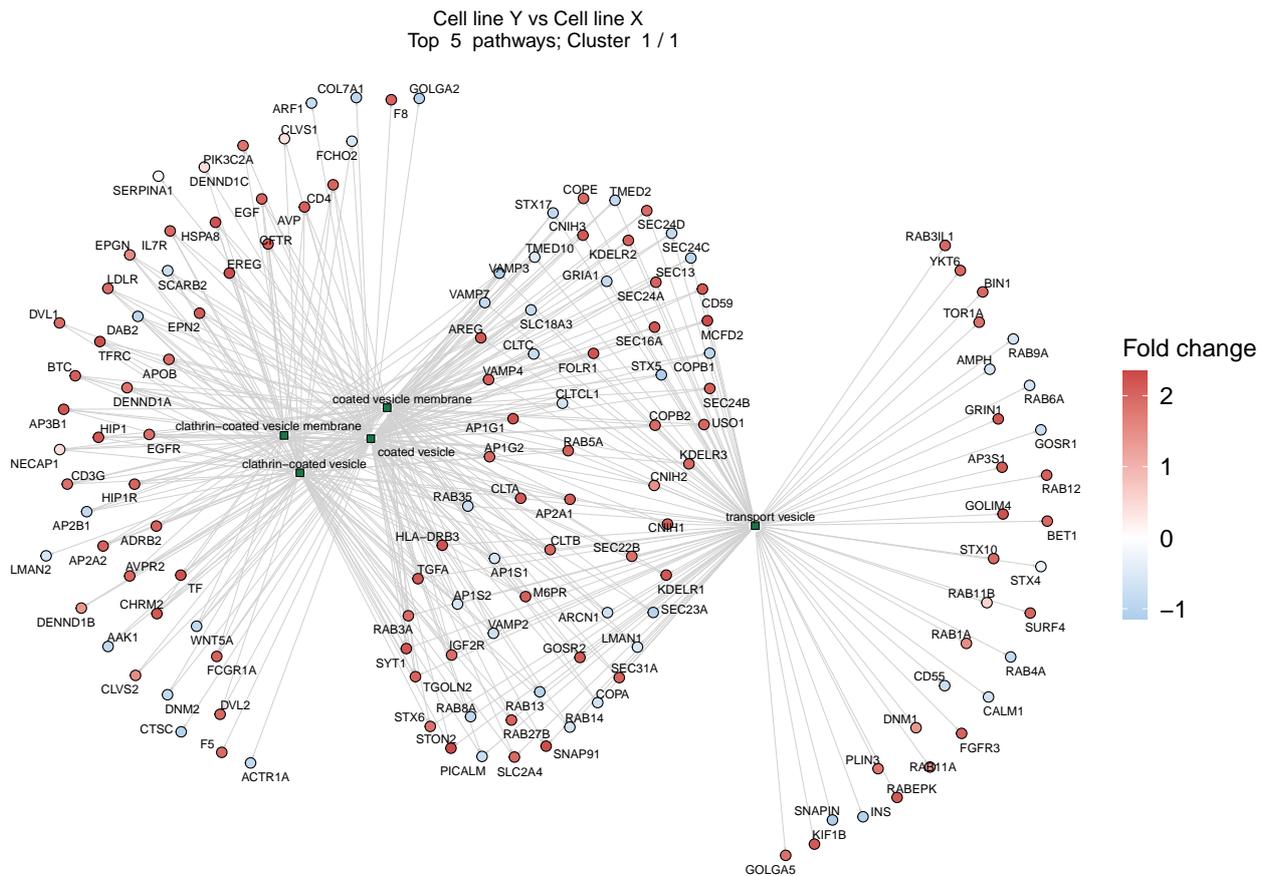
The enriched cellular components are illustrated in dot plots. The figure above is an example of the dot plots that can be found in the *Bioinformatics* folder.

This dot plot illustrates the Top 5 annotated cellular components found for Cell line Y vs Cell line X. In total, 192 enriched cellular components were found for Cell line Y vs Cell line X.

The y-axis shows the annotated cellular components with the lowest adjusted p-value. The color of the dots is scaled according to the adjusted p-value and the size of the dots is scaled according to the number of identifications in the components. The Gene Ratio is represented on the x-axis, which is the ratio between identifications annotated to the cellular components and the total of

identifications used for the analysis. The Top 5 annotated cellular components are plotted in this example for Cell line Y vs Cell line X, but other plots with additional cellular components are available in the *Bioinformatics* folder.

Made using [1] and [2] (please cite if results are used in a publication)



You can find the figure at:  
[Tissue\\_study analysis/Bioinformatics/Cell line Y vs Cell line X/Cellular Components/Annotated top 5 potentially regulated Cellular Components - Protein interactions plot.pdf](#)

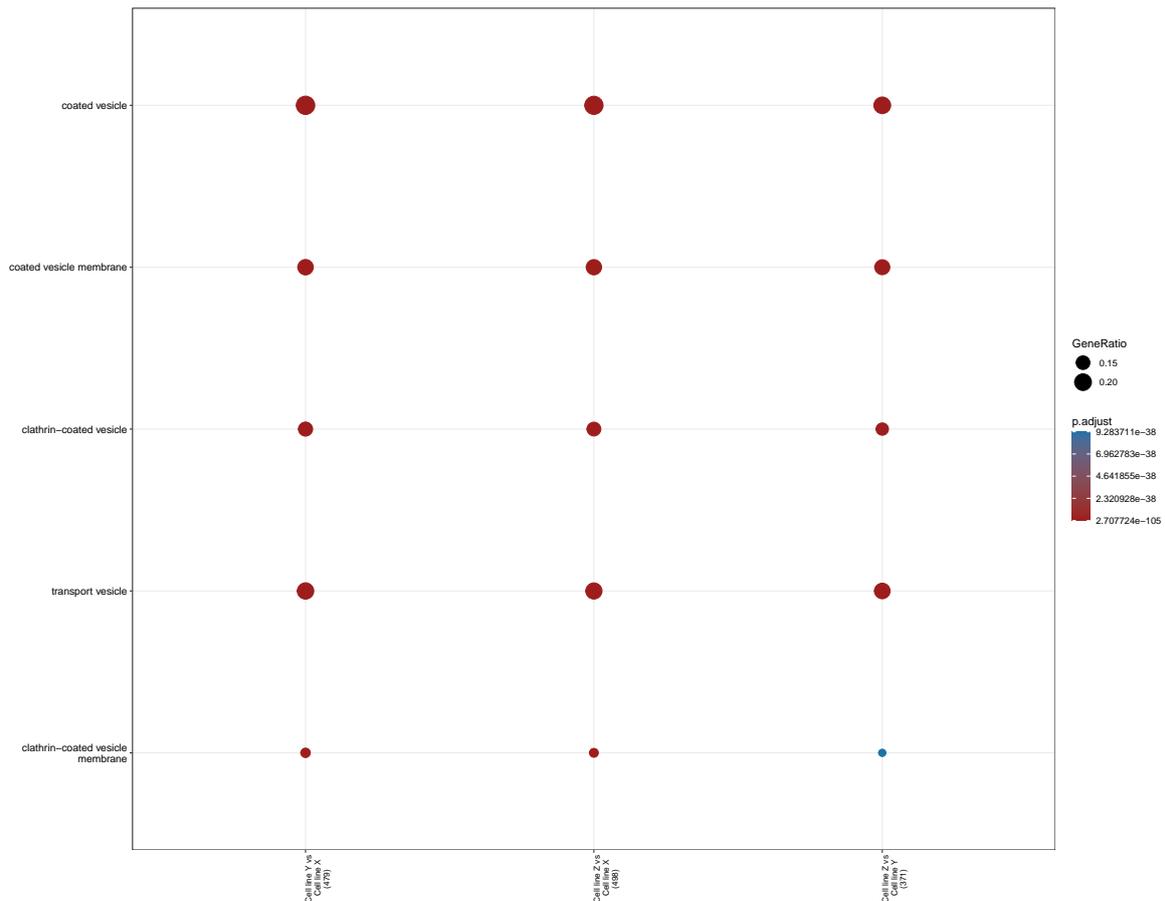
### Network plot depicting the linkages of proteins and Cellular Components

In this network illustration of the top 5 annotated cellular components, the enrichment terms are represented by green squares and the proteins are colored based on their log<sub>2</sub> fold change. The network plot of cellular components can help identify proteins that translocate between different compartments.

In this illustration, showing the enriched cellular components found for Cell line Y vs Cell line X, the proteins annotated to the cellular components have been plotted as well, showing 113 annotated proteins to these 5 cellular components, 104 of which are annotated to multiple cellular components and potentially could serve as links in-between the cellular compartments. The top 5 annotated cellular components are plotted in this example for

Cell line Y vs Cell line X, but other plots with additional cellular components are available in the *Bioinformatics* folder.

Made using [1], [2] and [3] (please cite if results are used in a publication).



You can find the figure at:  
[Tissue\\_study analysis/Bioinformatics/Annotated potentially regulated Cellular components across Groups.pdf](#)

### Cellular Components annotation of enrichment analysis, comparison across sample group tests illustrated by dot plot

The figure above compares enriched cellular components across the different group comparisons. The comparison of enriched cellular components across conditions yields a broad overview of protein localization to identify proteins or cellular components of interest. The proteins in the different pathways can be found in a spreadsheet in the *Bioinformatics* folder.

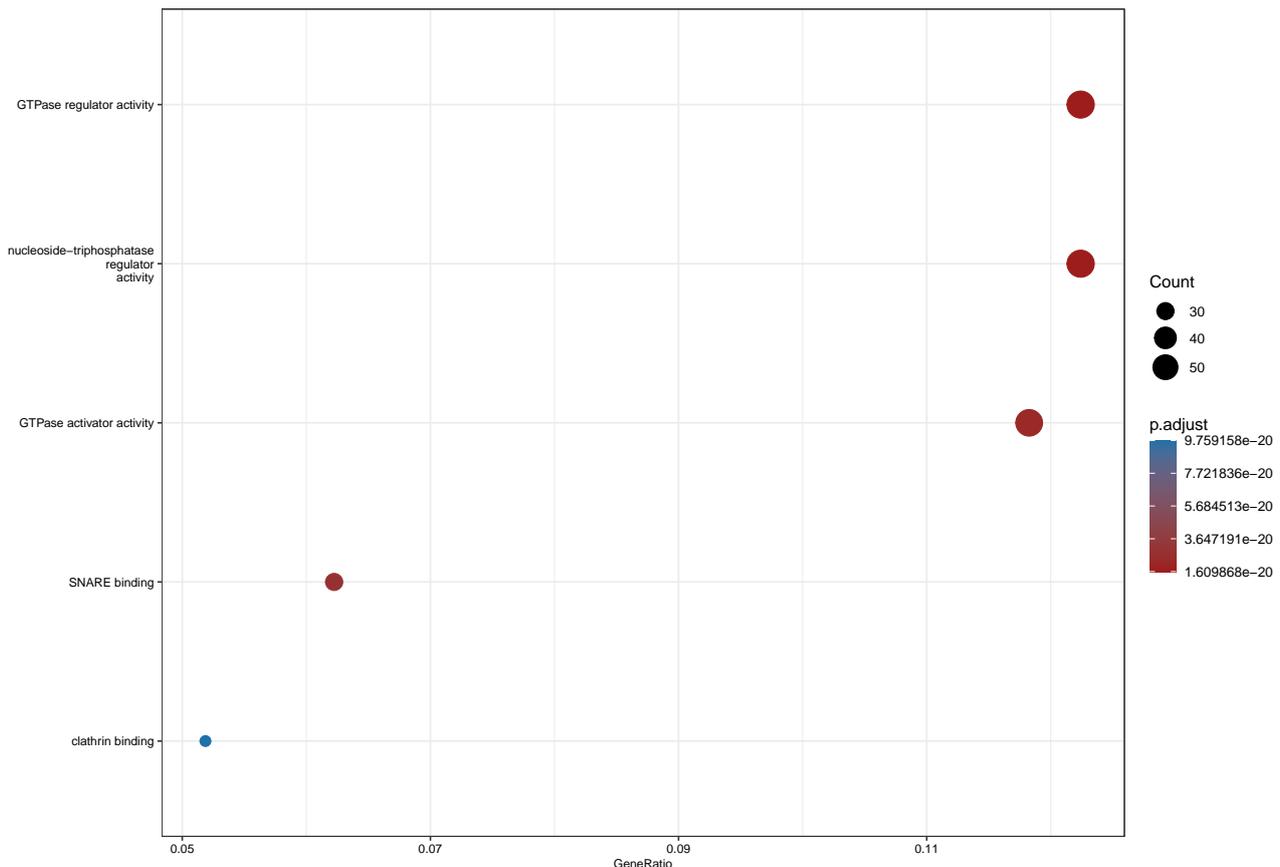
In this analysis, enriched terms were found in 3 comparisons (Cell line Y vs Cell line X, Cell line Z vs Cell line X, Cell line Z vs Cell line Y).

The sample group comparisons are along the X axis with the enriched cellular components on the Y axis. The dots

are colored by the adjusted p-value, and sized by the gene ratio, which is the number of proteins in that pathway out of the total number of tested proteins in that comparison.

The figure was made using [1], [2] and [3] (please cite if results are used in a publication).

## Bioinformatics



You can find the figure at:  
[Tissue\\_study analysis/Bioinformatics/Cell line Y vs Cell line X/Molecular functions/Top 5 annotated potentially regulated Molecular functions.pdf](#)

### Molecular functions annotation of enrichment analysis illustrated by dot plot

Dot plots were constructed as a means to visualize the enrichment terms. An example is displayed in the figure above for Cell line Y vs Cell line X, where the adjusted p-value (p-adjust) and gene ratio are depicted.

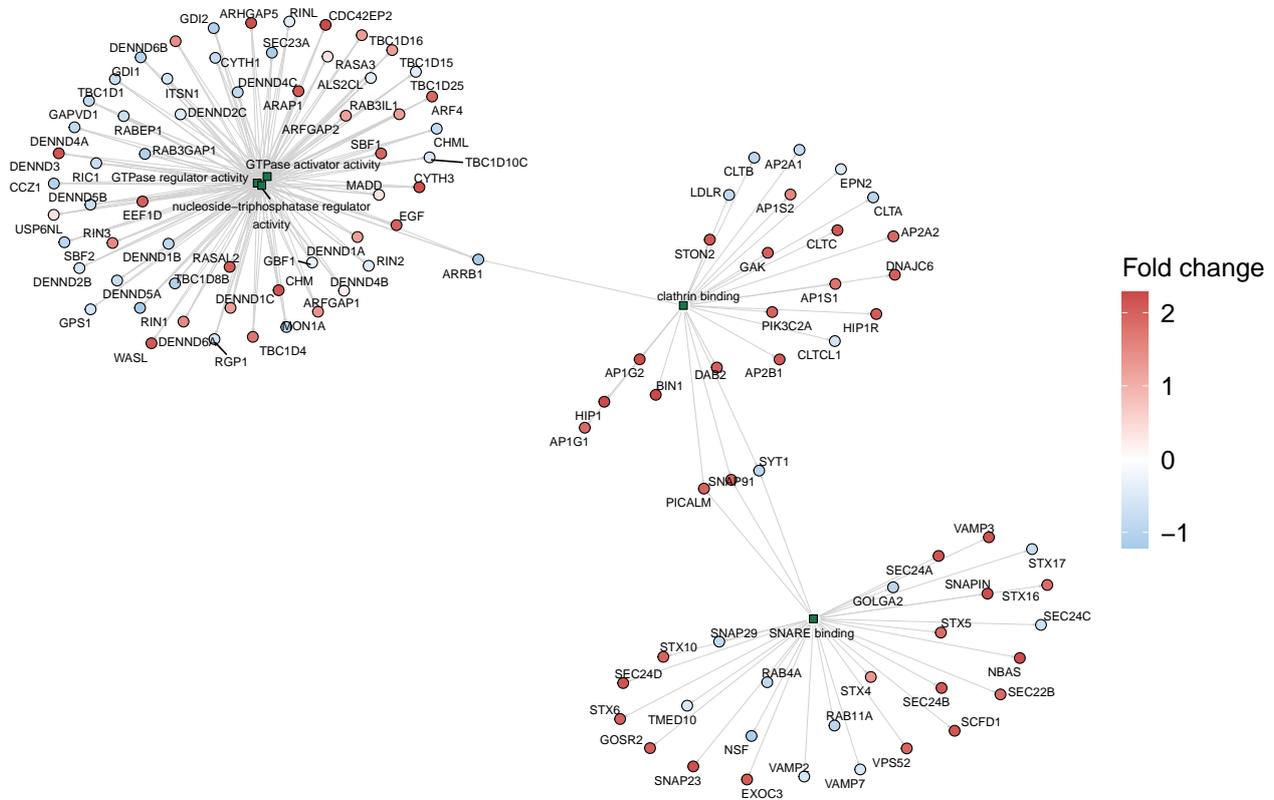
For Cell line Y vs Cell line X, the molecular functions enrichment analysis detected 77 molecular functions. The y-axis shows the annotated molecular functions with the lowest adjusted p-value and the x-axis shows the Gene Ratio (ratio between identifications annotated to the molecular functions, and the total of identifications used for the analysis). The color of the dots is scaled according to the adjusted p-value, and the size of the dots is scaled according to the number of identifications in the molecular functions.

The Top 5 annotated molecular functions are plotted in this example for Cell line Y vs Cell line X, but other plots with additional molecular functions are available in the *Bioinformatics* folder.

Made using [1], [2] and [3] (please cite if results are used in a publication).

## Bioinformatics

Cell line Y vs Cell line X  
Top 5 pathways; Cluster 1 / 1



You can find the figure at:  
[Tissue\\_study analysis/Bioinformatics/Cell line Y vs Cell line X/Molecular functions/Annotated top 5 potentially regulated Molecular functions - Protein interactions plot.pdf](#)

### Network plot depicting the linkages of proteins and Molecular functions

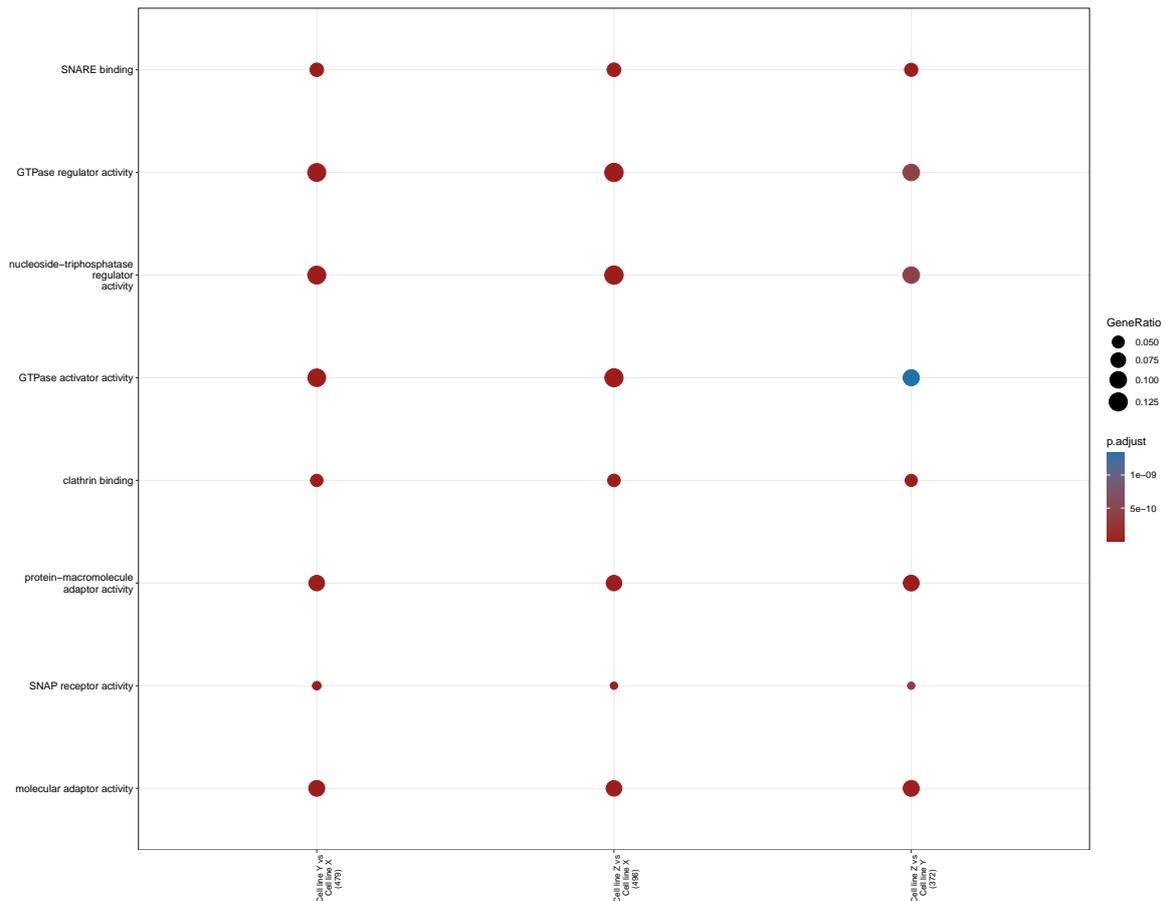
The annotations found in the molecular functions enrichment analysis are illustrated in network plots. The image above is an example of the network plots constructed for this analysis. The color of the proteins (represented as circles) is scaled based on the log<sub>2</sub> fold changes of their expression between the two sample groups. The molecular functions are indicated by green squares and the proteins are connected to the molecular functions they are associated with by grey lines.

In this example of the top 5 annotated molecular functions for Cell line Y vs Cell line X, 59 proteins were identified, 59 of which are annotated to multiple molecular functions and potentially could serve as links in-between functions.

The top 5 annotated molecular functions are plotted in this example for Cell line Y vs Cell line X, but other plots with additional molecular functions can be found in the *Bioinformatics* folder.

Made using [1], [2] and [3] (please cite if results are used in a publication).

## Bioinformatics



You can find the figure at:  
[Tissue\\_study analysis/Bioinformatics/Annotated potentially regulated Molecular functions across Groups.pdf](#)

### Molecular functions annotation of enrichment analysis, comparison across sample group tests illustrated by dot plot

The dot plot above illustrates the comparison for the annotated molecular functions in the different sample group tests.

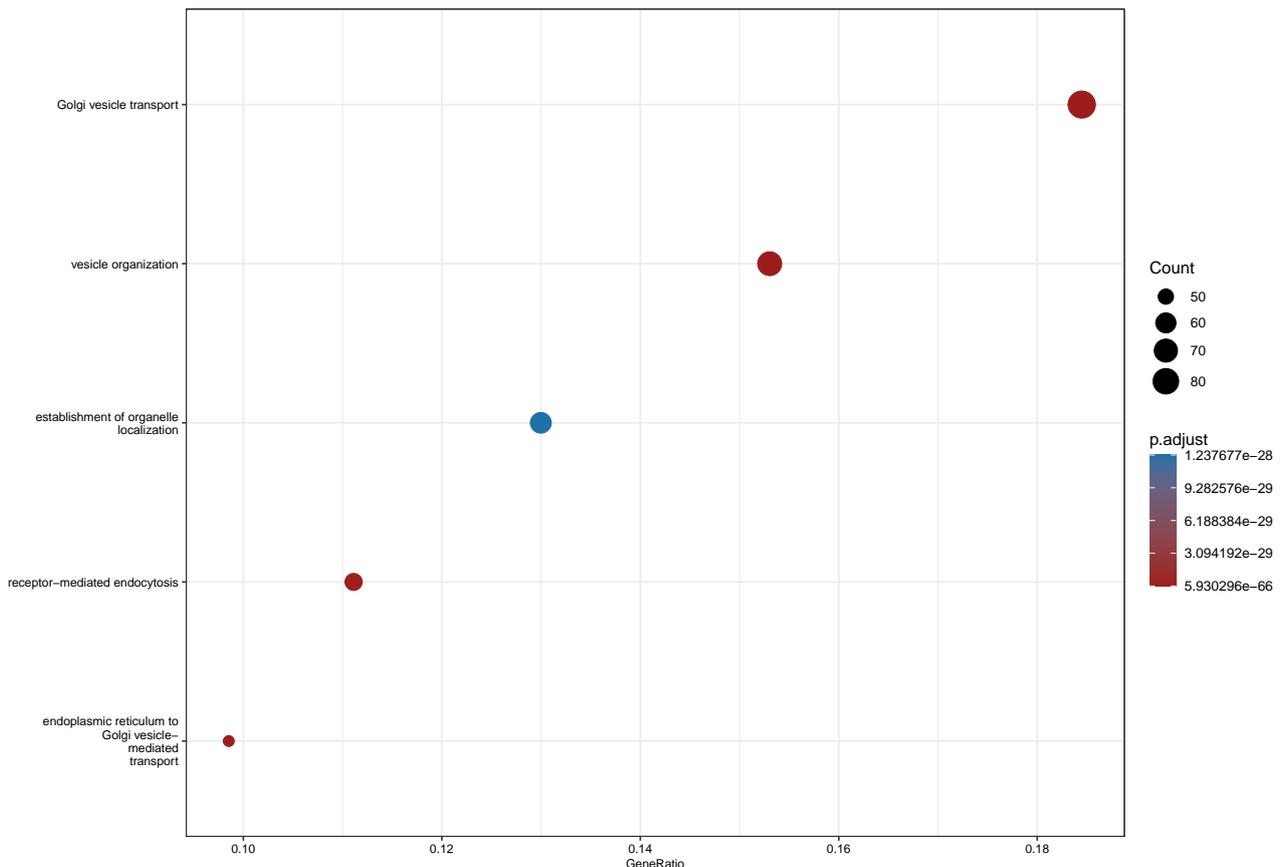
In this analysis, 3 comparisons (Cell line Y vs Cell line X, Cell line Z vs Cell line X, Cell line Z vs Cell line Y), were found to have molecular functions enriched terms.

The x-axis shows the different sample group tests (with the number of proteins used for each enrichment analysis within brackets). The y-axis shows annotated molecular functions, ranked by the lowest adjusted p-value. Moreover, the color of the dots is scaled by the adjusted p-value from the enrichment analysis, and the size of the

dots is scaled by the ratio of regulated expressions annotated to the molecular functions. All protein regulations with an unadjusted p-value below 0.05 were utilized for this analysis.

This figure is available in the *Bioinformatics* folder. Made using [1], [2] and [3] (please cite if results are used in a publication).

## Bioinformatics



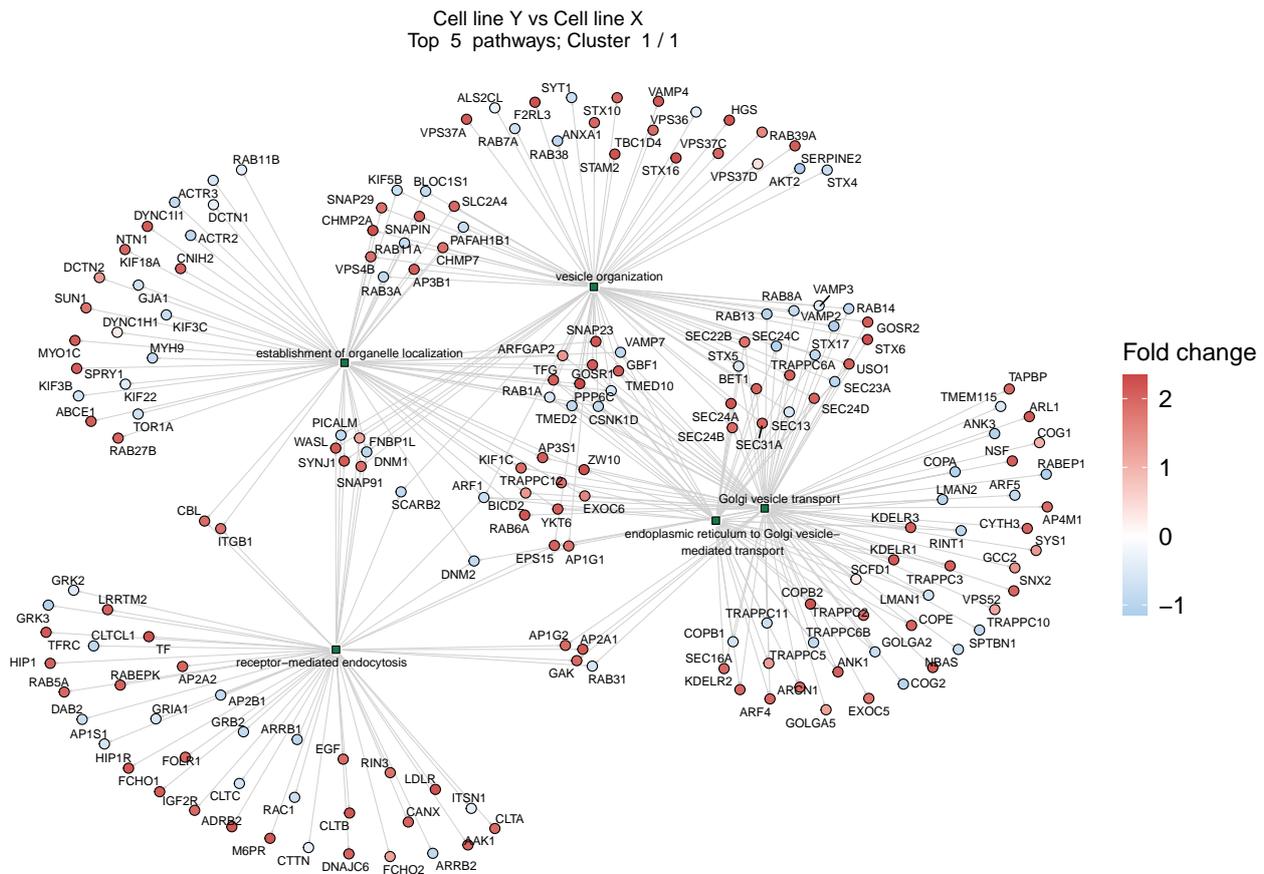
You can find the figure at:  
[Tissue\\_study analysis/Bioinformatics/Cell line Y vs Cell line X/Biological Processes/Top 5 annotated potentially regulated Biological Processes.pdf](#)

### Biological Processes annotation of enrichment analysis illustrated by dot plot

In this illustration of the annotated biological processes, the y-axis shows the annotated biological processes with the lowest adjusted p-value. The color of the dots is scaled according to the adjusted p-value, and the size of the dots is scaled according to the number of identifications in the biological processes. The Gene Ratio is represented on the x-axis, which is the ratio between identifications annotated to the processes and the total of identifications used for the analysis. In this analysis, 641 enriched biological processes were found.

The Top 5 annotated biological processes are plotted in this example for Cell line Y vs Cell line X, but other plots with additional processes can be found in the *Bioinformatics* folder.

Made using [1], [2] and [3] (please cite if results are used in a publication)



You can find the figure at:  
[Tissue\\_study analysis/Bioinformatics/Cell line Y vs Cell line X/Biological Processes/Annotated top 5 potentially regulated Biological Processes - Protein interactions plot.pdf](#)

### Network plot depicting the linkages of proteins and Biological Processes

In this network illustration of the annotated biological processes, the terms are represented by green squares and the proteins (circles) are colored based on their log<sub>2</sub> fold change. The biological processes can help identify additional regulated proteins that belong to a particular process of interest.

In this illustration, showing the enriched biological processes found for Cell line Y vs Cell line X, the proteins annotated to the biological processes have been plotted as well, showing 127 annotated proteins in these 5 biological processes, 58 of which are annotated to multiple biological processes and potentially could serve as links in-between processes.

The top 5 annotated biological processes are plotted in

this example for Cell line Y vs Cell line X, but other plots with additional biological processes can be found in the *Bioinformatics* folder.

Made using [1], [2] and [3] (please cite if results are used in a publication).

## Reference list

---

**Please remember to cite if results are used in a publication.**

- [1] G Yu, QY He\*. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Molecular BioSystems* 2016, 12(2):477- 479. doi: 10.1039/C5MB00663E.
- [2] Guangchuang Yu, Li-Gen Wang, Yanyan Han, and Qing-Yu He. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology* 2012, 16(5):284-287.
- [3] Guangchuang Yu, Li-Gen Wang, Yanyan Han, and Qing-Yu He. DOSE: an R/Bioconductor package for Disease Ontology Semantic and Enrichment analysis. *Bioinformatics* 2015, 31(4):608-609

# Thank you for choosing Biogenity

We appreciate the confidence you have placed in us by choosing Biogenity for your omics study. We are donating 15 trees in your name to the Amazon rainforest to acknowledge and celebrate our collaboration.

## Partnering with One Tree Planted

The Amazon rainforest is the host for the largest variety of species in the world, but it is threatened by deforestation. Deforestation, in general, contributes to increased CO<sub>2</sub> emissions, and 80,000 hectares of forest disappear from the Earth every day. More than 28,000 animal species will be extinct in 25 years if we continue this development, and the Amazon rainforest will be gone. We need to act now to try to rescue Earth's evolutionary catalyzer.

Biogenity believes that we must counteract deforestation by planting new and preserving existing forests, which also counteract some of the arising factors of the climate crisis. We know that we will not solve all climate problems or secure the Amazon rainforest by ourselves. However, we believe that many kind-hearted acts will form a greener future and a hope for the Earth's ecosystem.

This is why we donate trees for reforestation of the Amazon rainforest for every complete work order using our partner One Tree Planted. Every time you receive a report on your data, we will attach a certificate showing how many trees we planted in the Amazon rainforest in honor of your project.

Let's help each other preserve the Earth to create a better future for generations to come.

Kindly,  
The Biogenity Team

## 15 Trees Planted in the Amazon Rainforest

**THE PRESENT CERTIFICATE IS AWARDED TO PRIVILEGED AND DISTINGUISHED PARTNERS OF ONE TREE PLANTED WHOSE CONTRIBUTIONS HAVE BEEN AND CONTINUE TO BE, ESSENTIAL TO THE REFORESTATION, CONSERVATION AND PROTECTION OF FORESTS AROUND THE WORLD.**

Kenneth Kastaniegaard

PRESENTED BY  
CEO of Biogenity



Certificate id: #Phospho DEMO

June 17, 2022

DATE  
YOU CHANGED THE WORLD

# Biogenity

Your Global Partner in Omics