# **(Multi) omics data visualization in pathways**

## Visualisation of transcriptomics data

Please download the file **DEG\_F.txt** from the repository. PathVisio can read CSV files, too, but when opened in the program it added “” on the identifiers which makes them unreadable for PathVisio. The solution is simple, just open the DEG\_F.csv in Excel, and save it again in a different format - as tab delimited txt file. The DEG\_F.txt is exactly this, prepared for you.

### Step 1. Start PathVisio on your computer

Step 2. Check if gene and metabolite identifier mapping database is loaded (for transcriptomics, metabolite database is technically not needed but should be part of the preparation). If yes, the bottom of PathVisio should look like this:



Step 3. Open the 22q pathway. Go to Plugins -> WikiPathways -> Search **22q11.2** -> Open the human 22q pathway by double clicking when it appears in the result window.

Step 4. Upload the file with all the differentially expressed genes. Go to Data -> Import Expression Data. Use the Browse buttons to locate the files.

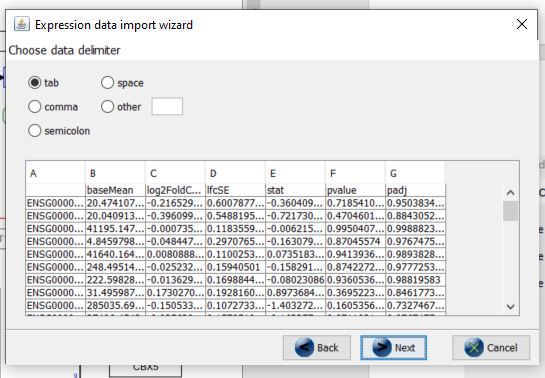
Input file: The experiment data file (**DEG\_F.txt**).

Output file: Will be filled in automatically after selecting the input file, you don't need to change this.

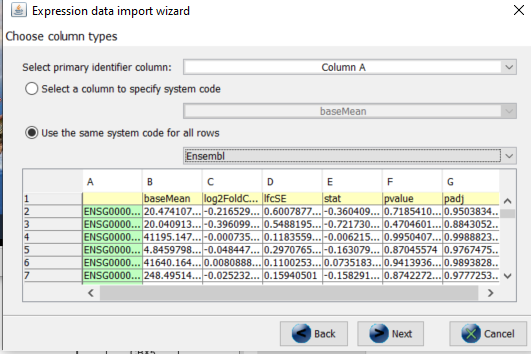
Gene database: Should be automatically filled in with the ID mapping database.

## 

Now follow the wizard to import the dataset (check the settings and then click on "Next").

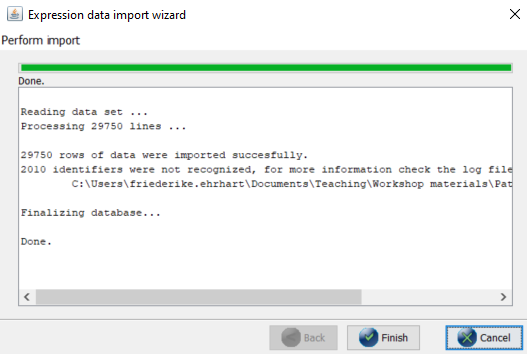


Select the columns that contain the identifiers (1) and identifier type (2). The database contains Ensembl IDs, so make sure Ensembl is selected in the system code drop-down box. The identifier column should be highlighted in green.

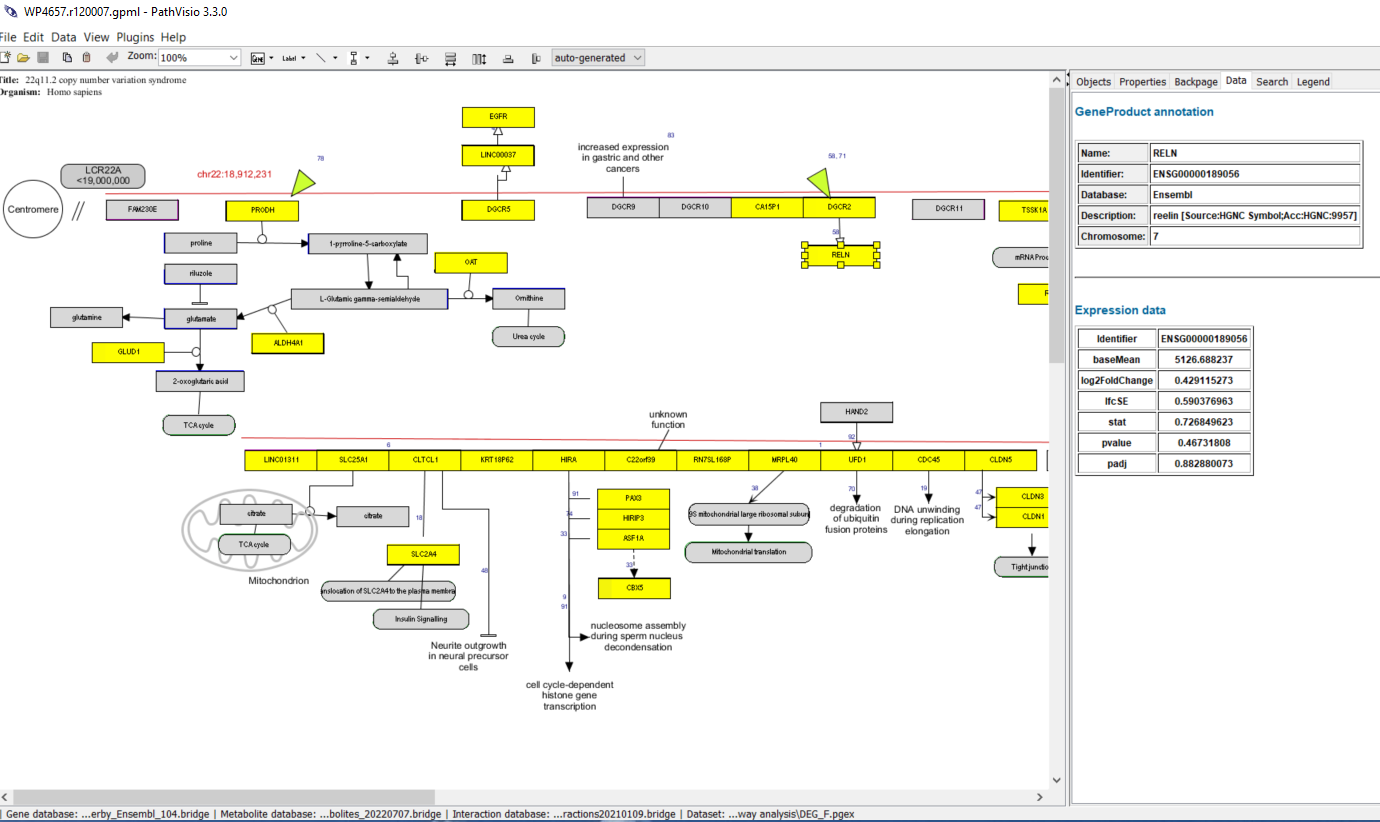


The data will now be imported into an expression dataset that is saved as a .pgex file on your hard disk (in the same directory as the dataset). Any exceptions will be reported to the file .pgex.ex (also in the same directory as the dataset). For example, a common exception occurs when an identifier is not found in the identifier mapping database and is therefore not included in the dataset. About 10% of identifiers not recognized is normal.

Please note that if you have an equal amount of "rows succesfully imported" and "identifiers not recognized" -> something went wrorng with importing the data !!



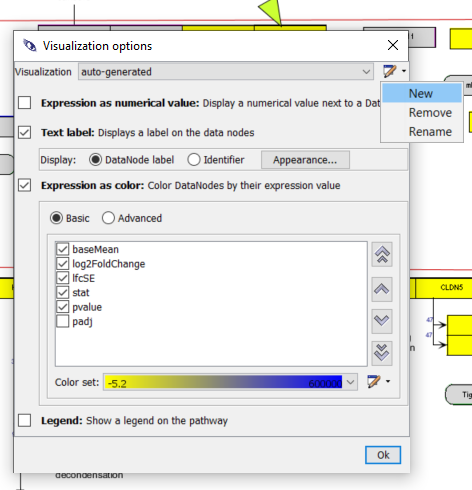
To test if the dataset has been created correctly, click on a colorful gene product in the pathway and the imported data will appear in the data panel (see tabs on the right side). If you don't see data for the gene product nodes, the import was not successful! Please redo the steps or ask an instructor for help.



If it looks like this, well done! You successfully imported the RNAseq dataset in PathVisio. The automatic visualization (in bright yellow) is unfortunately not very helpful. So, next we need to create a visualization that intuitively shows if transcripts are significantly up or downregulated in the dataset.

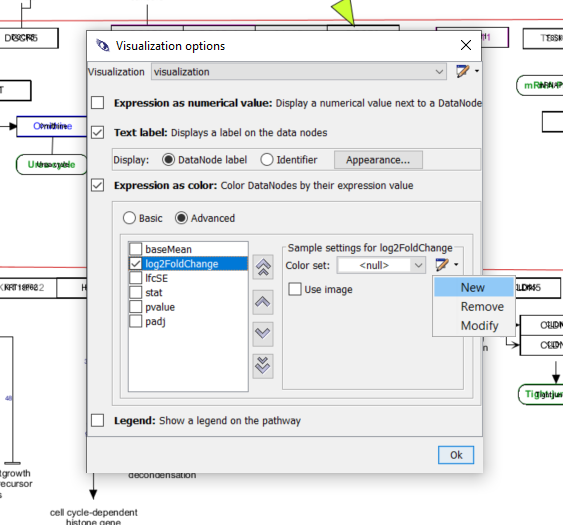
### Step 5: Visualize the differentially expressed genes on the pathway

Go to "Data -> Visualization options". The visualization options dialog will open. Create a new color set by clicking the button to the right of the color set dropdown box and select "New" to create a new color set.

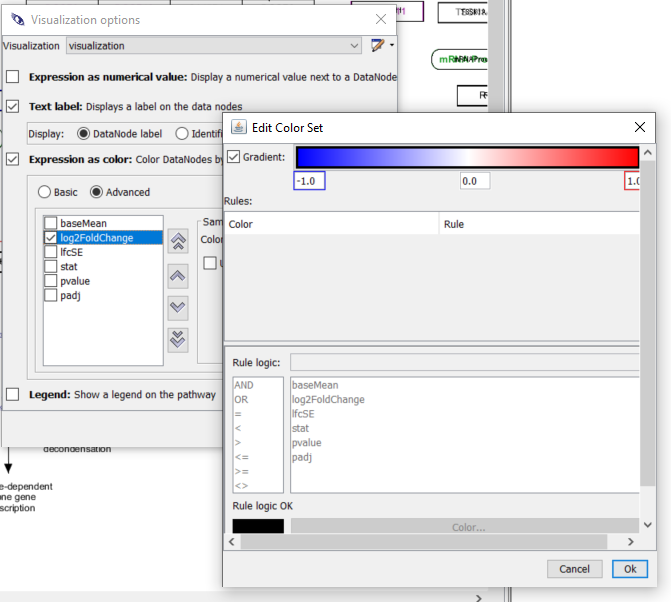


The color set dialog will now open. Click at “Text label”, “Expression as color” and “Advanced”.

Then, check the box “log2FoldChange” and click on “new” in order to create a new color set for log2FoldChange.



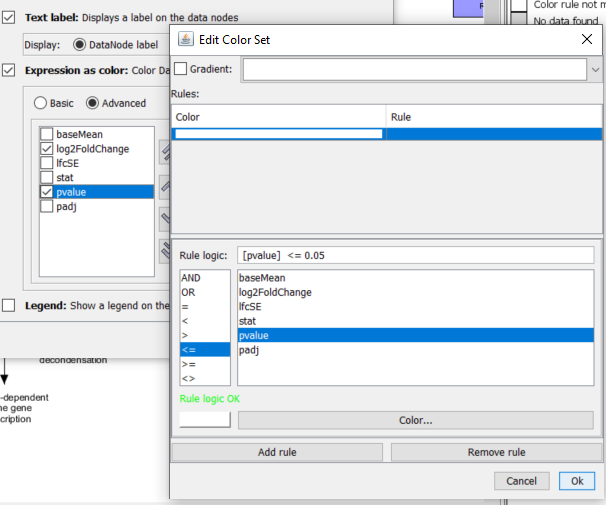
Select a gradient to visualize the logFC (blue to red over white) and click Ok. Strong colors indicate strong changes, blue indicates downregulation, red upregulation. For datasets with more or less strong expression changes you can adapt the minimum and maximum of the color scale accordingly – e.g. to ±2 or ±0.5 instead of 1.



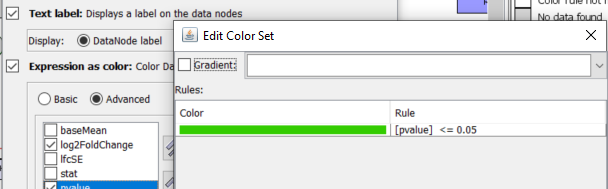
If you have done that correctly, the colors of the pathway should change to blue-white-red.

Thereafter, select pvalue and create a new visualisation. For the p-value, we are not so much interested in the gradient, but more whether it is below 0.05 or not. Therefore, we create a rule. If you are interested, you can later try using the padj (adjusted p-value) instead for a stricter selection.

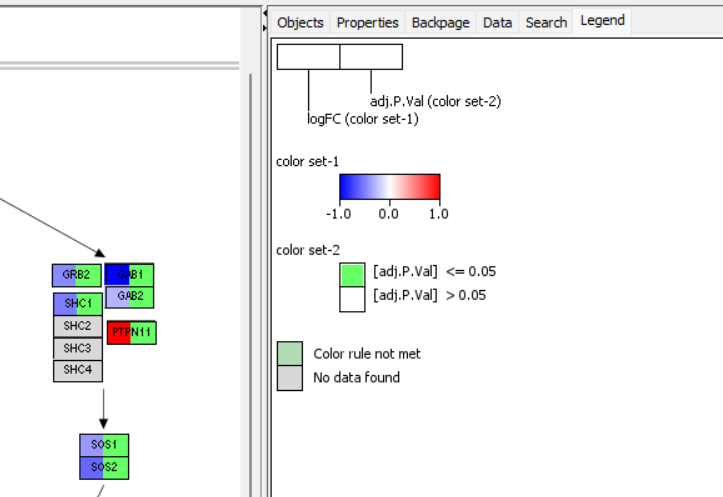
Click at “Add rule”, thereafter click at the pvalue and <= and type “0.05” in the rule logic field.

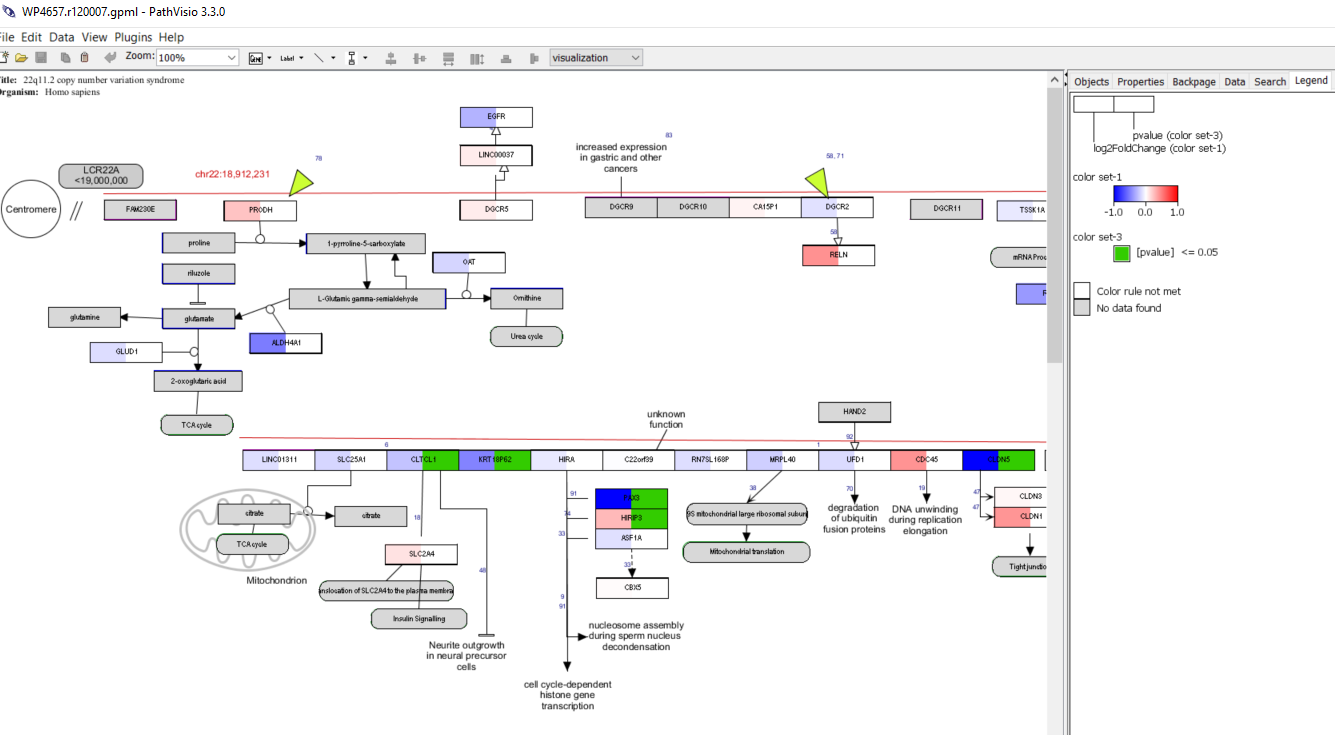


Click at color (the big field at the bottom of the window) and select a bright green. Click at ok.

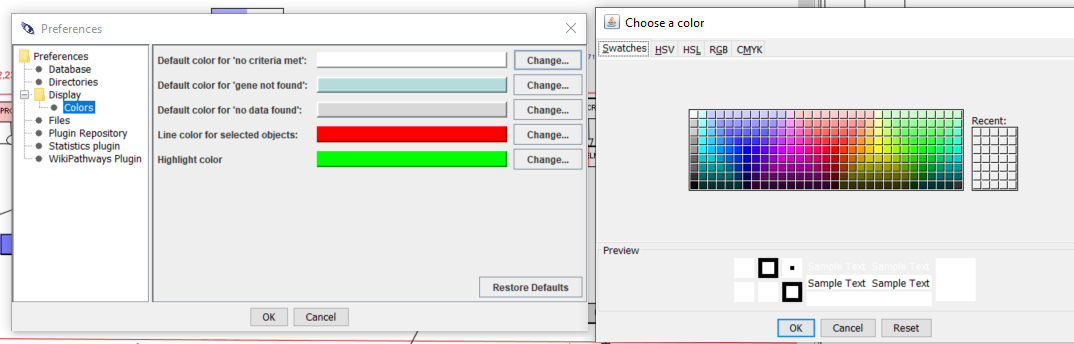


Now the differentially expressed genes are visualized in an intuitive way on the 22q11.2 pathway: red and blue indicates up or downregulation on the transcript level, strong colors indicate strong changes, pale colors smaller changes. Grey indicates that for this gene – or other node type like metabolite or pathway – no matching identifier could be found in the dataset. The bright green color in the second part of the gene box indicates if the change is significant on the p-value level 0.05. At the Legend Tab at the right hand side it is shown how the boxes are split.





The standard color for “rule not met” is a pale grey-green. If you want to change that to e.g. white, go to “Edit” – “Preferences” – press the + at “Display”- “Colors” – select a different color, e.g. white – press OK. The changes will only become apparent after restarting PathVisio!



**Q1: What do the blue and red colors indicate?**

**Q2: How is, in the 22q pathway, the gene expression changed?**

### Step 6: Open another pathway in PathVisio and investigate the changes in gene expression. Select pathways that was found overrepresented in the analysis earlier!

Investigating the gene expression visualized in the overrepresented pathways can help you to understand which parts of the pathway are affected and if you can expect the pathway may be up or down regulated.

Go to “Plugins” – “WikiPathways” – “Search” – type in a keyword from the pathway you are interested in. Select a pathway from the result list by double clicking.

The pathway will open while keeping the expression data and visualization. You can check different pathway this way.

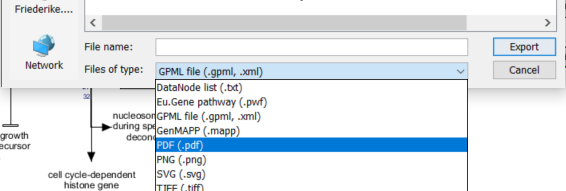
**Q3: Which of the other pathways are affected and how? Interpret the influence of differential expression in COS vs. control data of FB neurons on this pathway.**

### Step 7: Export and save

In order to export pathways with visualization data go to “File”- “Export” – navigate to the folder in which you want to save the file.

Select a file type: for figures pdf and svg are production quality, png or tiff are also good for presentations etc.

Datanode list will only give you a list of data nodes from this pathway, and .gpml and .mapp files are pathway format files.



## Visualisation of multi omics data

Download the file **Multiomics.txt** from the online repository. This datafile is merged file containing the top 10 000 most significantly changed transcripts from our FB neuron COS vs control RNAseq dataset and metabolomics data from a different experiment (on Dravet syndrome – there was no good COS data available). For good multi omics experiments, the different omics samples should come from the same sample – the same experiment, the same patient, tissue etc.

Open the file **Multiomics.txt** in Excel. You see now, there was an additional column added “System code” that indicates the database from which the identifiers are from. PathVisio works with bridgeDb identifier mappings and bridgeDb uses the system code listed on this page <https://www.bridgedb.org/pages/system-codes.html>. So, you can merge data from any identifier system listed there and visualize this in PathVisio. If you scroll a bit down you will find the other type of data.

**Q4: Which databases are the identifiers from?**

When you close Excel, please close without saving to avoid changes in the dataset!

### Step 1. Close and restart PathVisio on your computer

Step 2. Check if gene and metabolite identifier mapping database is loaded as before.

Step 3. Open the 22q pathway. Go to Plugins -> WikiPathways -> Search **22q11.2** -> Open the human 22q pathway when it appears in the result window.

Step 4. Upload the file with the multi omics data. Go to Data -> Import Expression Data. Use the Browse buttons to locate the files.

Input file: The experiment data file (**Multiomics.txt**).

Output file: Will be filled in automatically after selecting the input file, you don't need to change this.

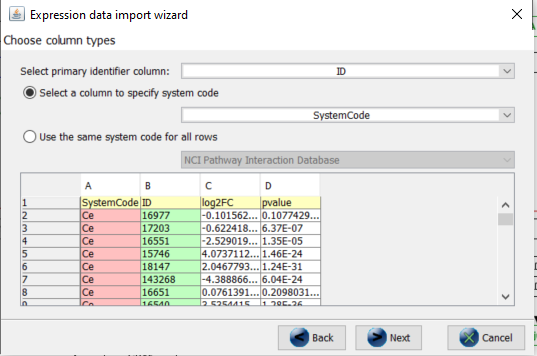
Gene database: Should be automatically filled in.

Follow the wizard to import the dataset (check the settings and then click on "Next").

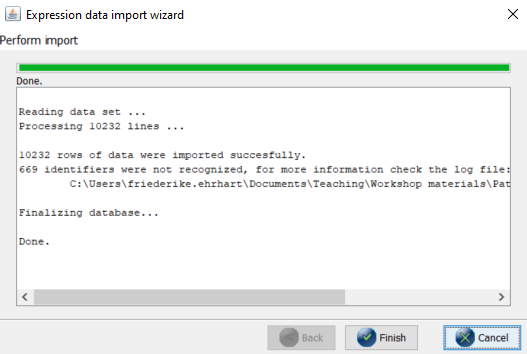
Now, we have two different identifier systems – so we need to tell PathVisio what is the identifier and what is the database.

Select primary identifier column (the column where the Ensembl and the ChEBI IDs are) -> ID (should appear in green).

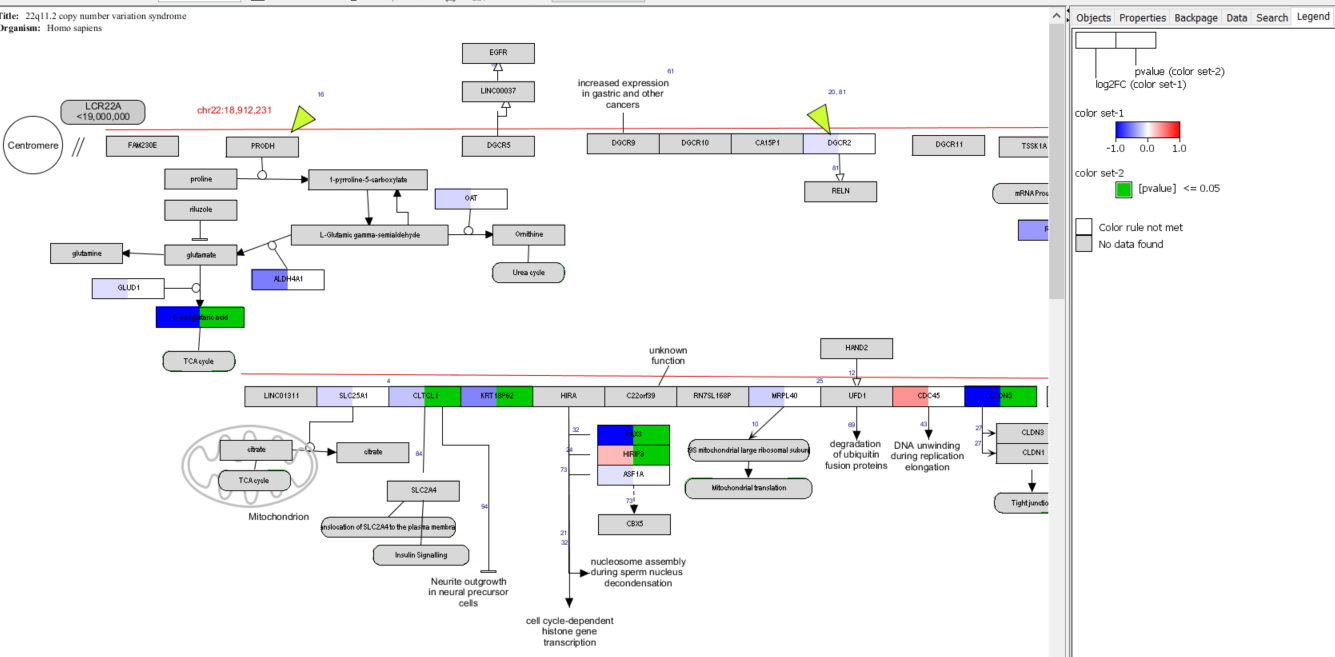
Check “Select a column to specify system code” and select the column with the system code -> SystemCode (should appear in red).



The import should work like before – this dataset is a bit smaller, and should be a bit faster.



Create a new visualization as before with a gradient for log2FC and rule for pvalue.



**Q5: Do you see any metabolite changes?**

### Step 5. Open another pathway in PathVisio and investigate the changes in gene expression and amounts of metabolites!

## Literature

1. Kutmon, M, et al. "PathVisio 3: an extendable pathway analysis toolbox." PLoS Comput Biol 11.2 (2015): e1004085. doi: 10.1371/journal.pcbi.1004085.
2. Martens M, et al. WikiPathways: connecting communities. NAR. D613–D621 (2021).

doi: 10.1093/nar/gkaa1024

Adapted procedure from the intro to pathway analysis and visualization from WikiPathways:

<https://github.com/gladstone-institutes/Bioinformatics-Workshops/tree/master/intro-pathway-analysis-visualization>