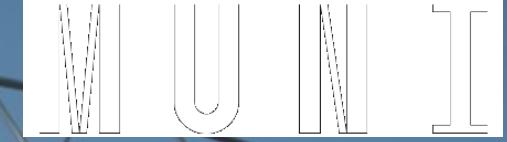




Central European Institute of Technology
BRNO | CZECH REPUBLIC



2019-11-20

Processing of proteomics data in **KNIME**



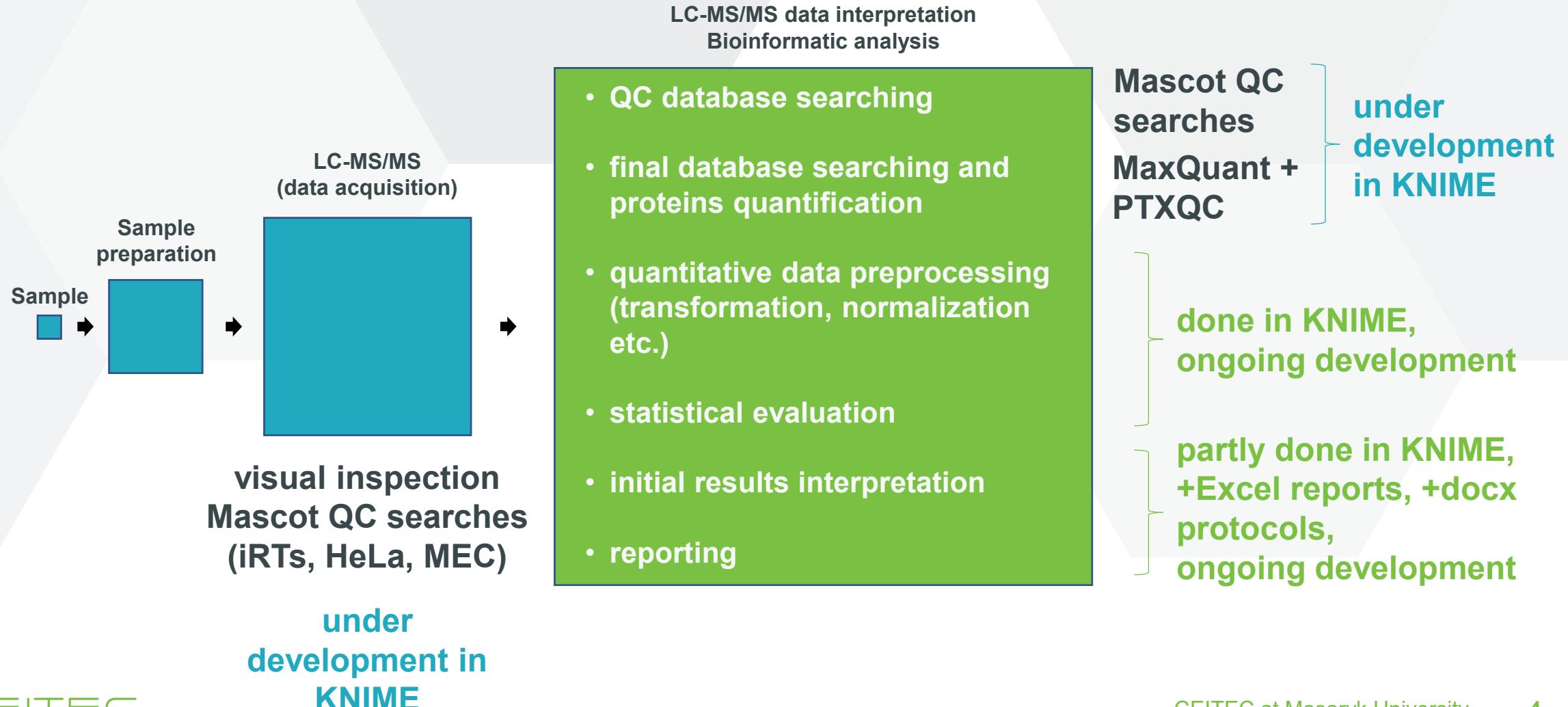
Workshop outline

- morning session – theoretical part

- 10:00 – 10:15 Opening and introduction
- 10:15 – 10:45 Software container running KNIME
- 10:45 – 11:00 Coffee break
- 11:00 – 11:30 Introduction to KNIME
- 11:30 – 11:45 Coffee break
- **11:45 – 12:30** **Practical applications, our KNIME metanodes**
- 12:30 – 13:30 Lunch break, visit of our laboratories for interested people

3) Practical applications, our KNIME metanodes

General proteomics study in our core facility - overview

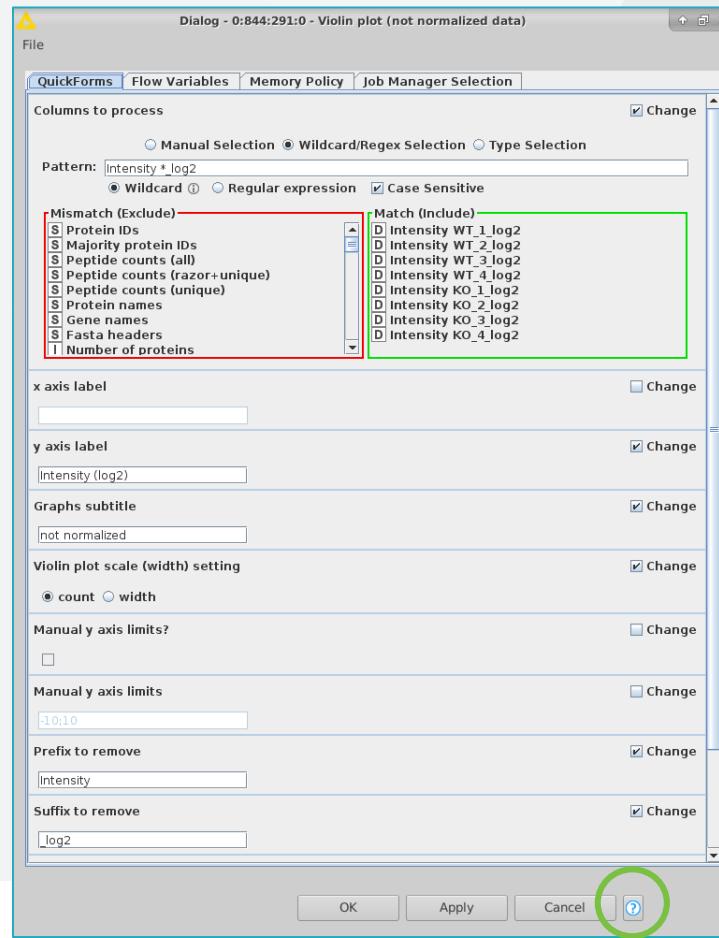
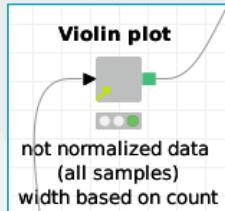




KNIME metanodes making concept

- utilize **in-built KNIME nodes** as much as possible
- well **documented**, including the script used
- the scripts are **completely accessible** and adjustable if needed, **no blackboxes**
- metanodes **versioning**
- metanodes **testing**
- easy to be used, no scripting needed
 - **you still have to know what you are doing and when to use the given tool!**

Metanodes example – Violin plot



Violin plot

Metanode to create Violin plot from selected columns of input table.

Note: any data preprocessing (like transformation, normalization) should be done prior the metanode usage!

=====

Used programs and tools and their respective licenses at the time of the metanode creation. Version numbers and the licenses might differ based on your local installation. Please inspect your local installation and contact us if you can not locate your local version and or license terms.

KNIME nodes (The KNIME nodes consists of the following GNU GPL 3.0 License. Licence terms are available here: <https://www.gnu.org/licenses/gpl.html>)
Python 3 (The Python consists of the following Python 3.6 License. Licence terms are available here: <https://docs.python.org/3.6/license.html>)
Python package Seaborn (The Seaborn consists of the following BSD License. Licence terms are available here: <https://opensource.org/licenses/BSD-3-Clause>)
Python package Matplotlib (The Matplotlib consists of the following Python Software Foundation License (BSD compatible). Licence terms are available here: <https://matplotlib.org/users/license.html>)
Python package Pandas (The Pandas consists of the following BSD License. Licence terms are available here: <https://opensource.org/licenses/BSD-3-Clause>)

The metanode was created in KNIME 3.7.1 running inside the docker image (<https://hub.docker.com/r/cfprot/knime/>), tag 3.7.1a.

=====

This version of metanode is available under the GNU GPL 3.0 License, unless stated otherwise. The full version of the license terms is available at <https://www.gnu.org/licenses/gpl.html>.
Version: 0.4.3 from 2019-03-20
Contact person: David Potesil (david.potesil@ceitec.muni.cz)
More information can be found at https://github.com/OmicsWorkflows/KNIME_metanodes

Dialog Options

Columns to process
select colomns to be processed

x axis label
how the graph x axis should be titled

y axis label
how the graph y axis should be titled

Graphs subtitle
additional information that should be present as the graphs subtitle

Violin plot scale (width) setting
sets the violin plot scale settings
- count - violin plot scale (width) will reflect the number of values
- width - all violin graphs will have the same width irrespective the number of values

Manual y axis limits?
whether to use manually set y axis limits (ckecked) or use automatic limits (unchecked)

manual y axis limits
limits of y axis in the form of two numbers separated by semicolon; use point(.) as decimal separator

Prefix to remove
common data columns prefix to be removed prior plotting

Suffix to remove
common data columns suffix to be removed prior plotting

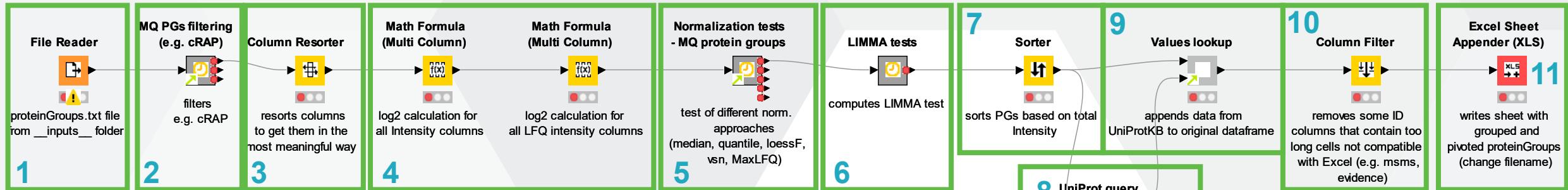
Our KNIME metanodes availability

- KNIME Metanodes templates GitHub repository
 - https://github.com/OmicsWorkflows/KNIME_metanodes
 - optimal utilization within the KNIME docker VNC container
 - work also independently in KNIME installation
 - utilized KNIME plugins and or python/R packages have to be installed



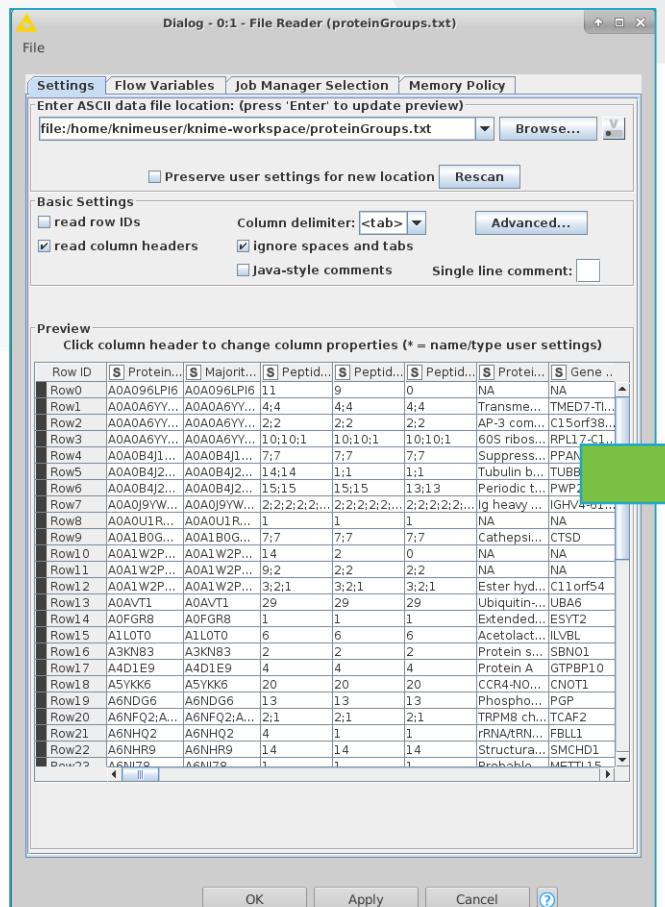
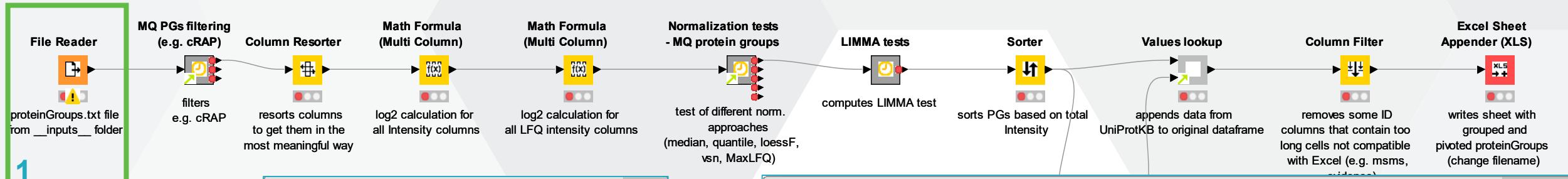
Example MaxQuant proteinGroups file processing workflow

Initial pipeline for the processing of LFQ bottom up MS data



1. table to be processed loading
2. filters out reverse hits and contaminants
3. resorts the samples/replicates ordering (figures)
4. data transformation (log2) – protein groups Intensities and LFQ intensities
5. data normalization approach selection and data check
6. statistical evaluation
7. resorts the table (important mainly for reporting purposes)
8. gets proteins annotation from UniProtKB
9. adds annotation info into the original table
10. removes not compatible columns
11. writes the table as the initial report

1. table to be processed loading



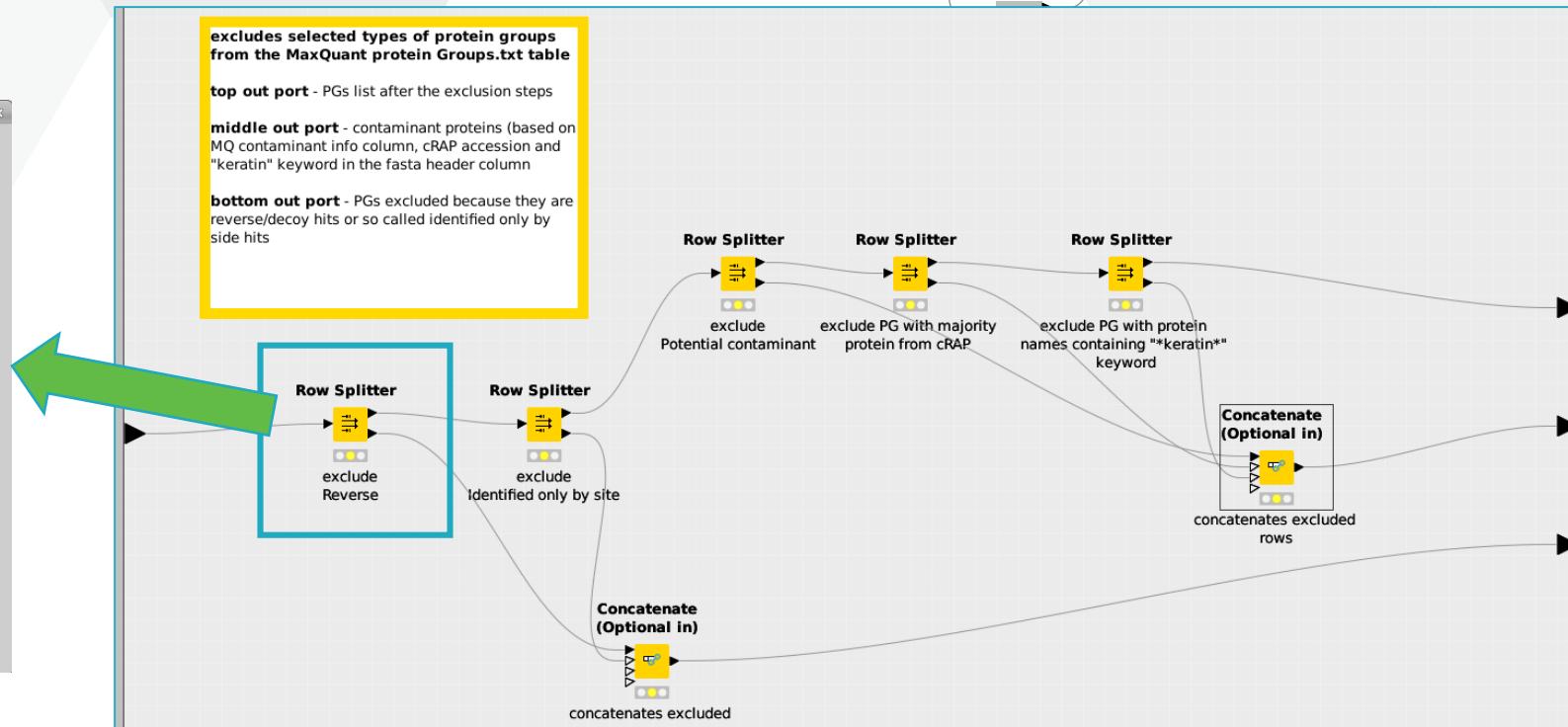
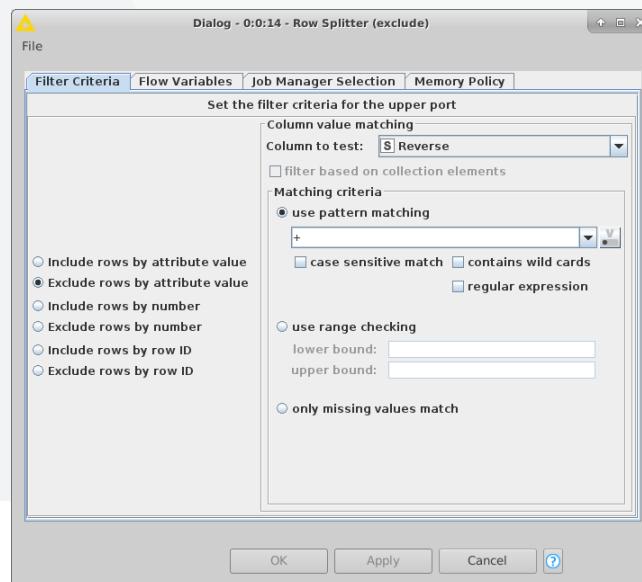
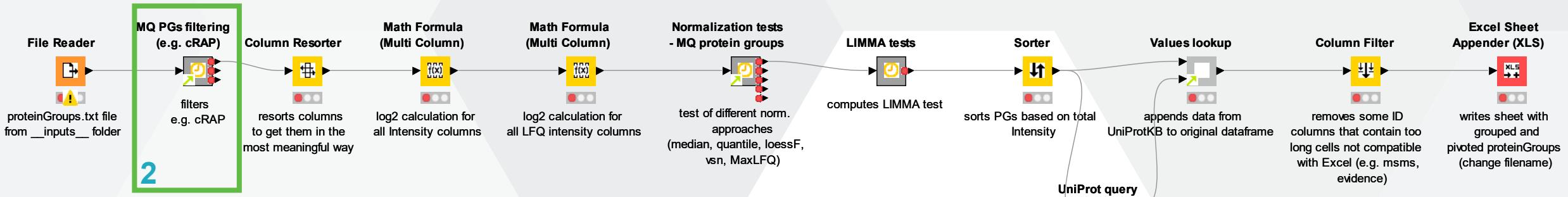
File Table - 0:1 - File Reader (proteinGroups.txt)

File Hilite Navigation View

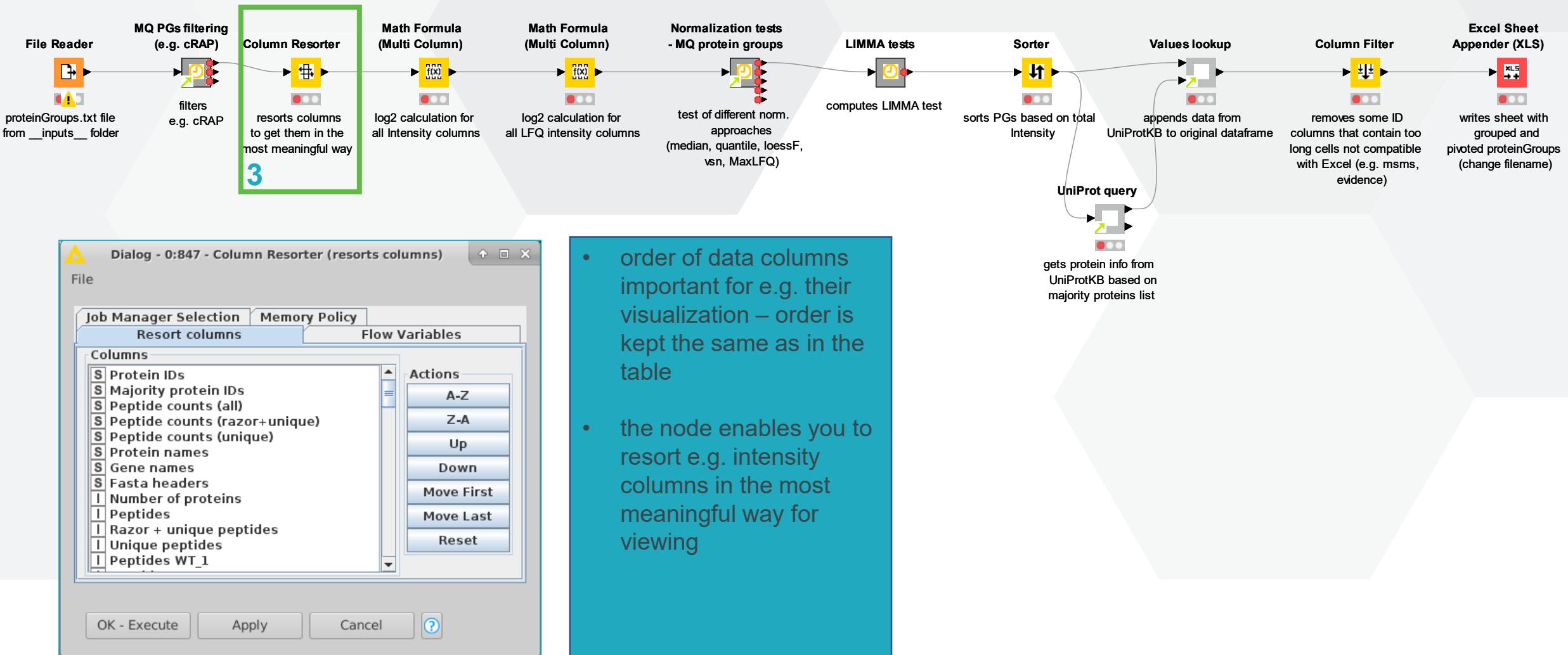
Table "proteinGroups.txt" - Rows: 3855 Spec - Columns: 81 Properties Flow Variables

| Row ID | \$ Protein... | \$ Majorit... | \$ Peptid... | \$ Peptid... | \$ Peptid... | \$ Protei... | \$ Gene ... | \$ Fasta headers |
|--------|----------------------------|----------------------------|---------------|---------------|---------------|---------------|------------------------------------------------------------|------------------------------------------------------|
| Row0 | A0A096LP16 | A0A096LP16 | 11 | 9 | 0 | NA | NA | Uncharacterized protein (Fragment) OS=Homo sa |
| Row1 | A0A0A6YY... A0A0A6YY... | A0A0A6YY... A0A0A6YY... | 4;4 | 4;4 | 4;4 | Transm... | Protein TMED7-TICAM2 OS=Homo sapiens GN=TM | |
| Row2 | A0A0A6YY... A0A0A6YY... | A0A0A6YY... A0A0A6YY... | 2;2 | 2;2 | 2;2 | AP-3 com... | C15orf38... Protein C15orf38-AP352 OS=Homo sapiens GN=C | |
| Row3 | A0A0A6YY... A0A0A6YY... | A0A0A6YY... A0A0A6YY... | 10;10;1 | 10;10;1 | 10;10;1 | 60S ribos... | RPL17-C1... | Protein RPL17-C18orf32 OS=Homo sapiens GN=R |
| Row4 | A0A0B4J1... A0A0B4J1... | A0A0B4J1... A0A0B4J1... | 7;7 | 7;7 | 7;7 | Suppress... | PPAN-P2... | HCG2039996 OS=Homo sapiens GN=PPAN-P2R1Y |
| Row5 | A0A0B4J2... A0A0B4J2... | A0A0B4J2... A0A0B4J2... | 14;14 | 1;1 | 1;1 | Tubulin b... | TUBB3 | Uncharacterized protein OS=Homo sapiens PE=1 |
| Row6 | A0A0B4J2... A0A0B4J2... | A0A0B4J2... A0A0B4J2... | 15;15 | 15;15 | 13;13 | Periodic t... | PWP2 | Uncharacterized protein OS=Homo sapiens GN=L |
| Row7 | A0A0J9YW... A0A0J9YW... | A0A0J9YW... A0A0J9YW... | 2;2;2;2;2;... | 2;2;2;2;2;... | 2;2;2;2;2;... | Ig heavy ... | IGHG4-61... | Uncharacterized protein (Fragment) OS=Homo sa |
| Row8 | A0A0U1R... | A0A0U1R... | 1 | 1 | 1 | NA | NA | Uncharacterized protein (Fragment) OS=Homo sa |
| Row9 | A0A1B0G... | A0A1B0G... | 7;7 | 7;7 | 7;7 | Cathepsi... | CTSD | Uncharacterized protein OS=Homo sapiens PE=1 |
| Row10 | A0A1W2P... | A0A1W2P... | 14 | 2 | 0 | NA | NA | Uncharacterized protein OS=Homo sapiens PE=4 |
| Row11 | A0A1W2P... | A0A1W2P... | 9;2 | 2;2 | 2;2 | NA | NA | RPS10-NUDT3 readthrough OS=Homo sapiens GN |
| Row12 | A0A1W2P... | A0A1W2P... | 3;2;1 | 3;2;1 | 3;2;1 | Ester hyd... | C11orf54 | Uncharacterized protein OS=Homo sapiens PE=4 |
| Row13 | A0AVT1 | A0AVT1 | 29 | 29 | 29 | Ubiquitin... | UBA6 | Ubiquitin-like modifier-activating enzyme 6 OS=Ho |
| | A0FGR8 | A0FGR8 | 1 | 1 | 1 | Extended... | ESYT2 | Extended synaptotagmin-2 OS=Homo sapiens GN |
| | A1L0TO | A1L0TO | 6 | 6 | 6 | Acetolact... | LLVBL | Acetolactate synthase-like protein OS=Homo sapi |
| Row16 | A3KN83 | A3KN83 | 2 | 2 | 2 | Protein s... | SBN01 | Protein strawberry notch homolog 1 OS=Homo sa |
| Row17 | A4D1E9 | A4D1E9 | 4 | 4 | 4 | Protein A | GTPBP10 | Protein A OS=Homo sapiens GN=GTPBP10 PE=1 |
| Row18 | A5YKK6 | A5YKK6 | 20 | 20 | 20 | CCR4-NOT... | CNOT1 | CCR4-NOT transcription complex subunit 1 OS=Ho |
| Row19 | A6NDG6 | A6NDG6 | 13 | 13 | 13 | Phospho... | PGP | Glycerol-3-phosphate phosphatase OS=Homo sap |
| Row20 | A6NFQ2;A... | A6NFQ2;A... | 2;1 | 2;1 | 2;1 | TRPM8 ch... | TCAF2 | TRPM8 channel-associated factor 2 OS=Homo sap |
| Row21 | A6NHQ2 | A6NHQ2 | 4 | 1 | 1 | rRNA/tRNA... | FBLL1 | rRNA/tRNA 2-O-methyltransferase fibrillarin-like pro |
| Row22 | A6NHR9 | A6NHR9 | 14 | 14 | 14 | Structura... | SMCHD1 | Structural maintenance of chromosomes flexible H |
| Row23 | A6NJ78 | A6NJ78 | 1 | 1 | 1 | Probable ... | METTL15 | Probable methyltransferase-like protein 15 OS=Ho |
| Row24 | A6NKT7;Q... | A6NKT7;Q... | 18;16 | 1;1 | 1;1 | RanBP2-li... | RGPD3;R... | RanBP2-like and GRIP domain-containing protein 3 |
| Row25 | A8CG34;Q... | A8CG34;Q... | 4;3 | 4;3 | 4;3 | Nuclear e... | POM121C... | Nuclear envelope pore membrane protein POM 12 |
| Row26 | A8MXV4;R... | A8MXV4 | 3;1;1 | 3;1;1 | 3;1;1 | Nucleosoi... | NUDT19 | Nucleoside diphosphate-linked moiety X motif 19 C |
| Row27 | A9UHW6 | A9UHW6 | 2 | 2 | 2 | MIF4G do... | MIF4G | MIF4G domain-containing protein OS=Homo sapi |
| Row28 | B01T2 | B01T2 | 39 | 39 | 38 | Unconve... | MYO1G | Unconventional myosin-Ig OS=Homo sapiens GN= |
| Row29 | B4DLN1;Q... | B4DLN1;Q... | 14;7 | 14;7 | 7;7 | Mitochon... | SLC25A10 | cDNA FLJ60124, highly similar to Mitochondrial dic |
| Row30 | P0CG08;B... | P0CG08;B... | 1;1 | 1;1 | 1;1 | Golgi pH ... | GPR89B;... | Golgi pH regulator B OS=Homo sapiens GN=GPR8 |
| Row31 | C9JAW5;Q... | C9JAW5;Q... | 1;1 | 1;1 | 1;1 | HIG1 dom... | HIGD1A | Uncharacterized protein OS=Homo sapiens PE=4 |
| Row32 | E7ENX8;P... | E7ENX8 | 11;2;1;1 | 11;2;1;1 | 2;2;1;1 | NA | NA | Uncharacterized protein (Fragment) OS=Homo sa |

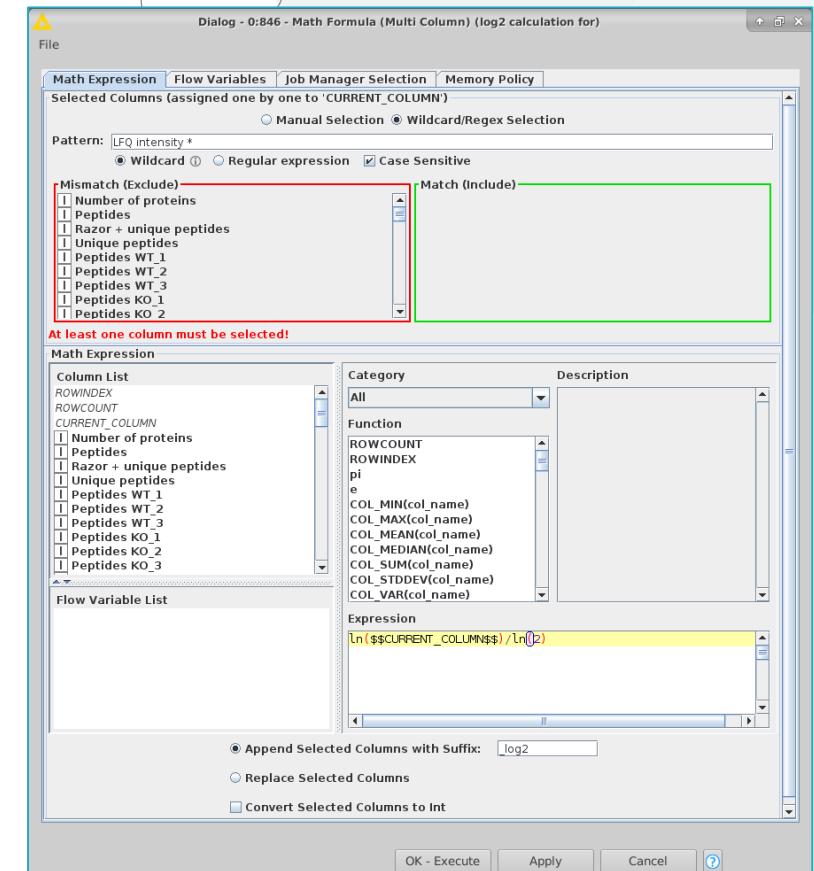
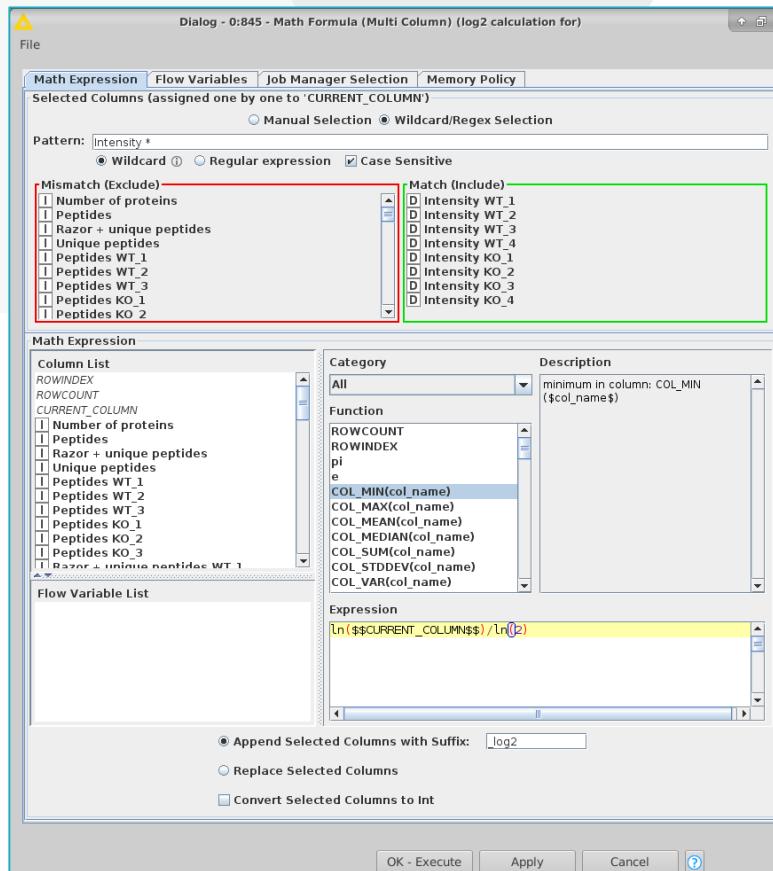
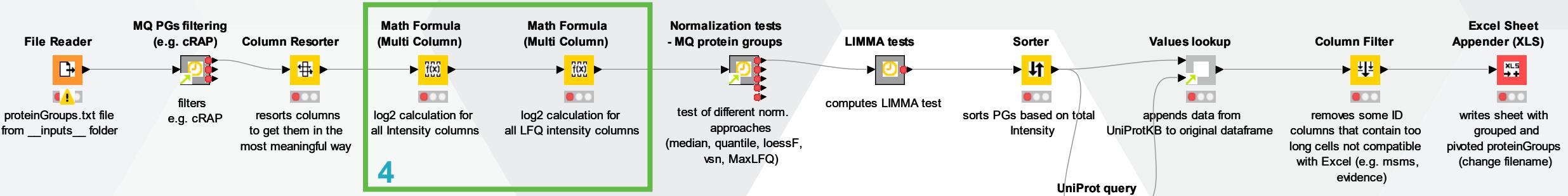
2. filters out reverse hits and contaminants



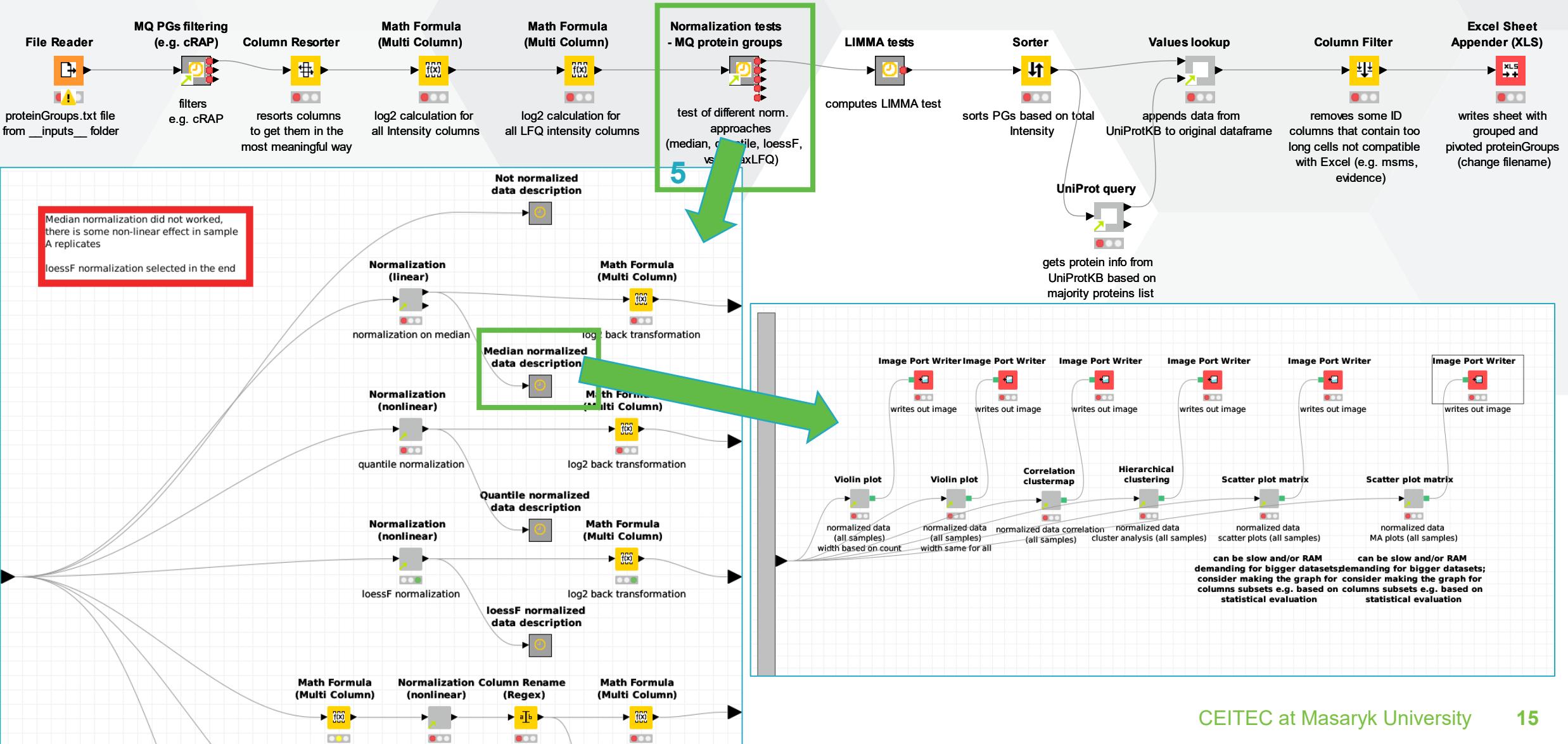
3. resorts the samples/replicates ordering (figures)



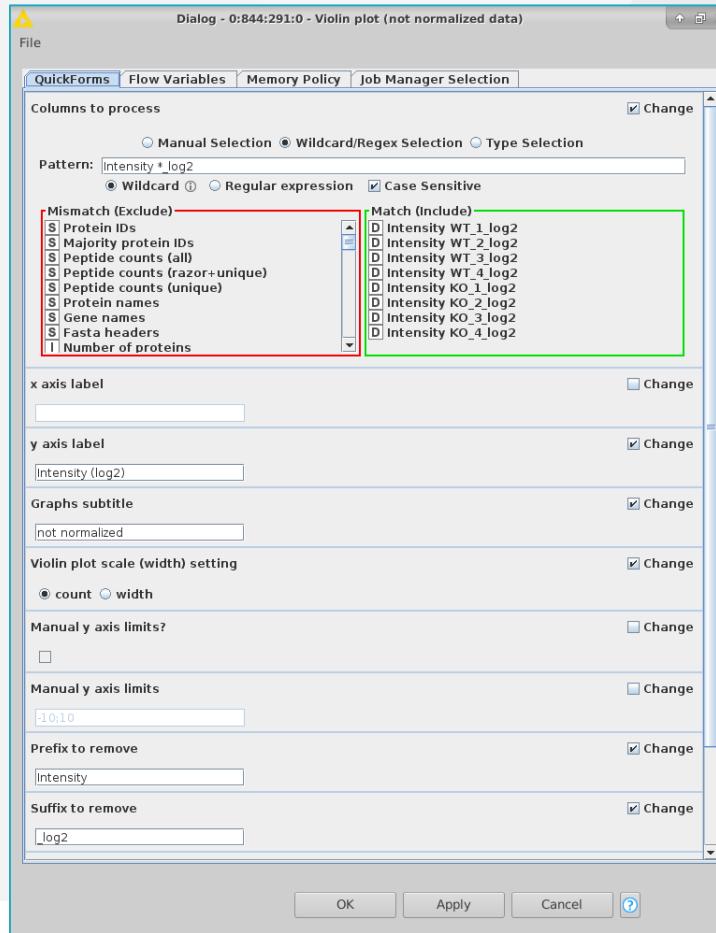
4. data transformation (log2) – protein groups Intensities and LFQ intensities



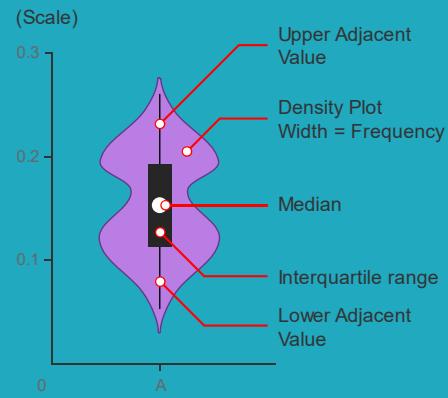
5. data normalization approach selection and data check



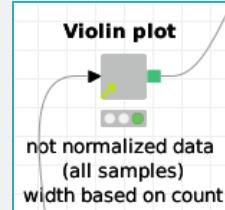
Metanodes examples – Violin plot



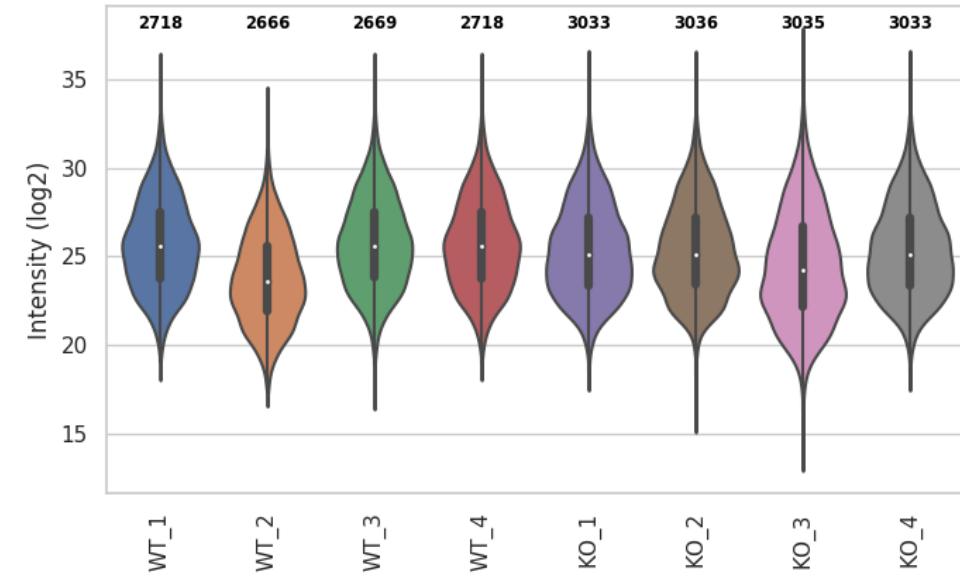
- creates violin plot based on all numerical values for the selected columns
- violin plot scale (width) can be selected



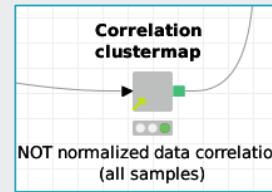
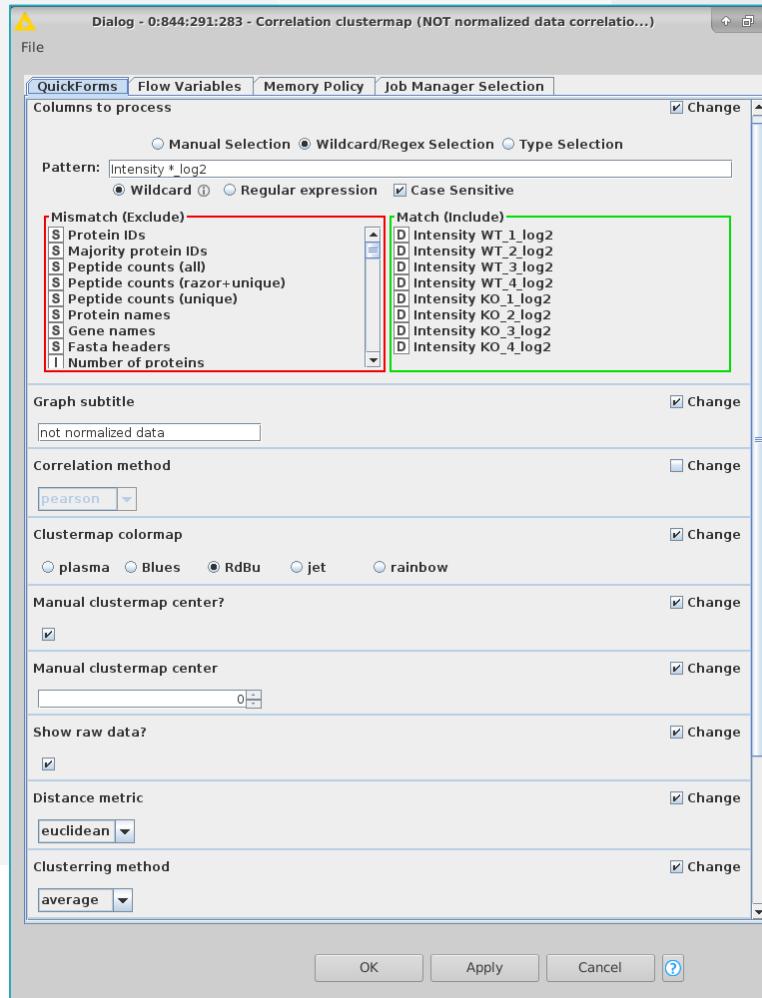
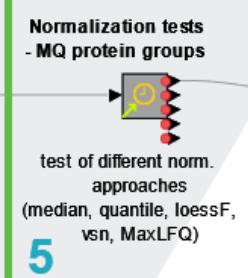
- Comparison of samples:** number of cases, distribution, basic statistics (min, max, median)



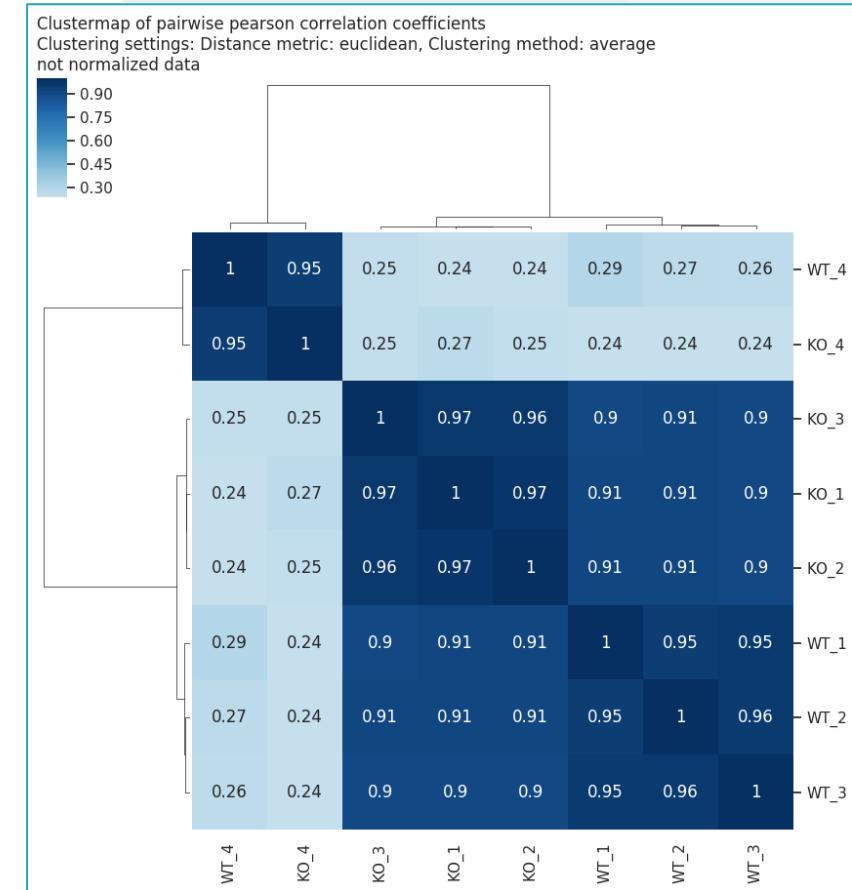
**Violin graphs and number of observed values:
violins width corresponds to the number of values
not normalized**



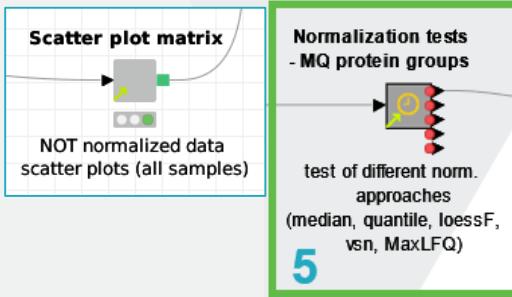
Metanodes examples – Correlation clustermap



- computes pairwise correlation coefficients and uses them for cluster analysis
- correlation method can be selected
- cluster analysis settings can be selected
- mainly used for the normalized data
- comparison of samples:** which samples tend to form clusters – are similar?
- how much are samples similar to each other (correlation coefficient)



Metanodes examples – Scatter plot matrix



QuickForms Flow Variables Memory Policy Job Manager Selection

Columns to process

Pattern: Intensity WT*_log2

○ Manual Selection ○ Wildcard/Regex Selection ○ Type Selection

Pattern: Intensity WT*_log2

● Wildcard ○ Regular expression Case Sensitive

Mismatch (Exclude)

Match (Include)

S Protein IDs
S Majority protein IDs
S Peptide counts (all)
S Peptide counts (razor+unique)
S Peptide counts (unique)
S Protein names
S Gene names
S Fasta headers
I Number of proteins

Graphs subtitle: not normalized data

Fit regression curve?

Type of regression: lowess

Show unity line?

Impute missing values?

Value to impute if requested: 0

Construct MA plot like graph?

manual x axis limits: None

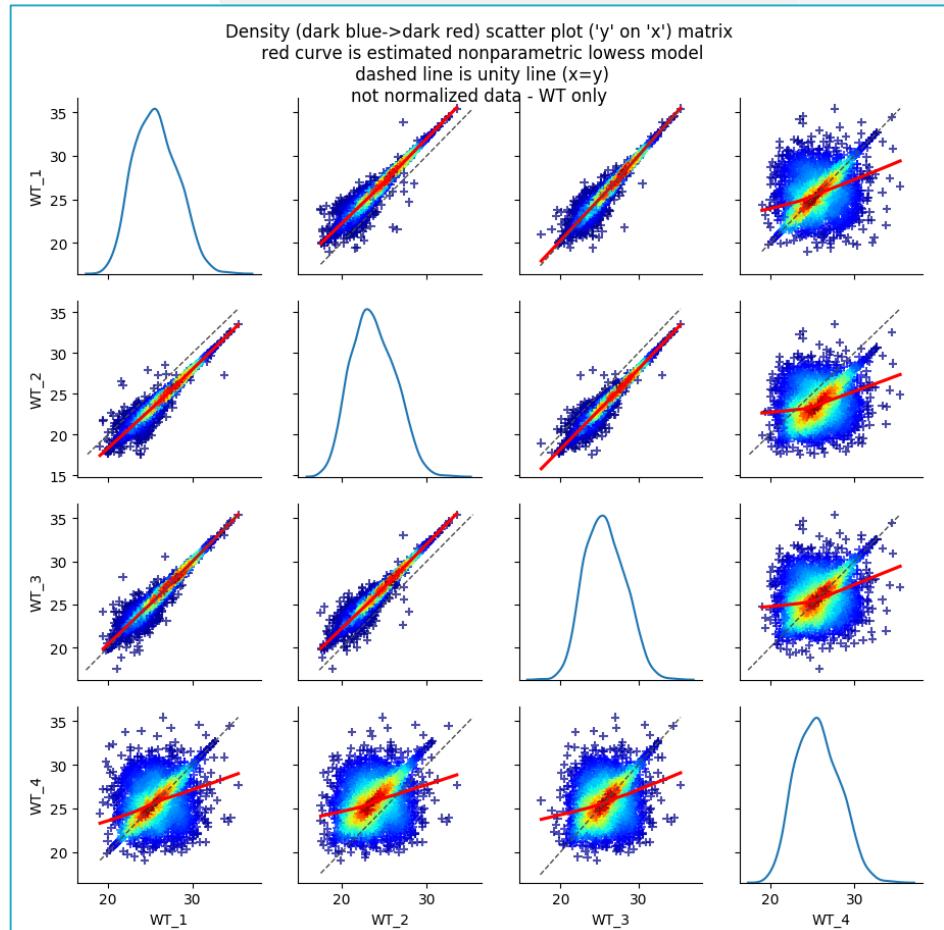
manual y axis limits: None

Prefix to remove: Intensity

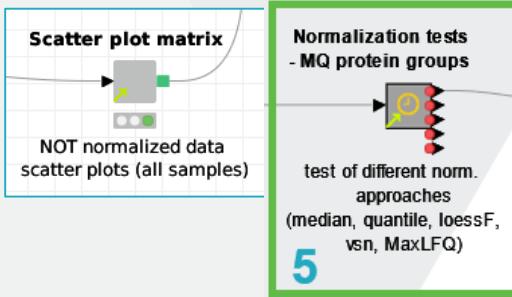
Suffix to remove: log2

Size: None

- Intensity correlation for all possible sample pairs
- Unity line ($y=x$) displayed as dashed line
- Regression curve (linear or loess) displayable as well
- Symbols colors provide the density information (i.e. blue - low density; red – high density)
- Including possible imputation of missing values by constant
- Possibility to create MA plot like graph style
- Information on how individual protein group intensities are similar among samples
- Normalization “quality check”



Metanodes examples – Scatter plot matrix



QuickForms Flow Variables Memory Policy Job Manager Selection

Columns to process Change

Pattern: Intensity WT*_log2

Manual Selection Wildcard/Regex Selection Type Selection

Mismatch (Exclude) Case Sensitive

Match (Include)

- Protein IDs
- Majority protein IDs
- Peptide counts (all)
- Peptide counts (razor+unique)
- Peptide counts (unique)
- Protein names
- Gene names
- Fasta headers
- Number of proteins

Graphs subtitle not normalized data Change

Fit regression curve?

Type of regression lowess Change

Show unity line?

Impute missing values? Change

Value to impute if requested 15 Change

Construct MA plot like graph?

manual x axis limits None Change

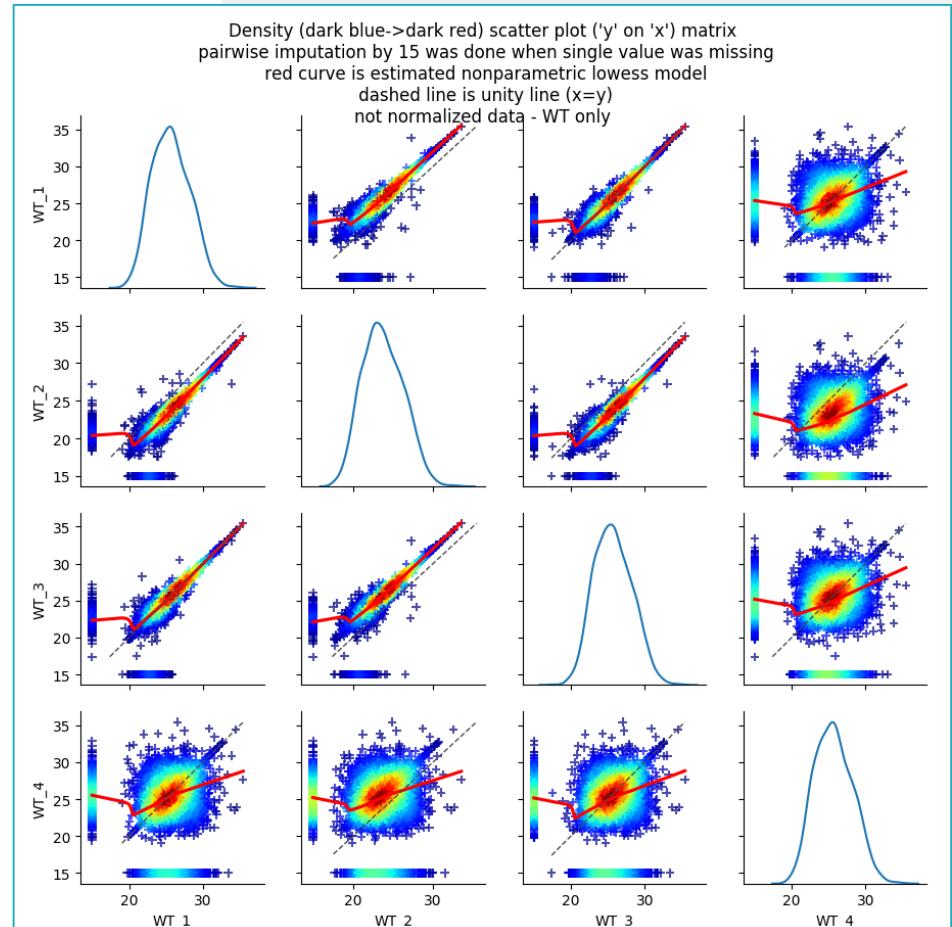
manual y axis limits None Change

Prefix to remove Intensity Change

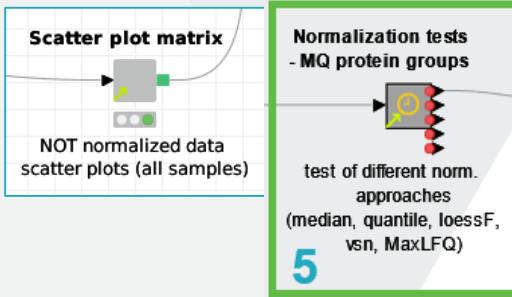
Suffix to remove log2 Change

Size None Change

- Intensity correlation for all possible sample pairs
- Unity line ($y=x$) displayed as dashed line
- Regression curve (linear or loess) displayable as well
- Symbols colors provide the density information (i.e. blue - low density; red – high density)
- Including possible imputation of missing values by constant
- Possibility to create MA plot like graph style
- Information on how individual protein group intensities are similar among samples
- Normalization “quality check”



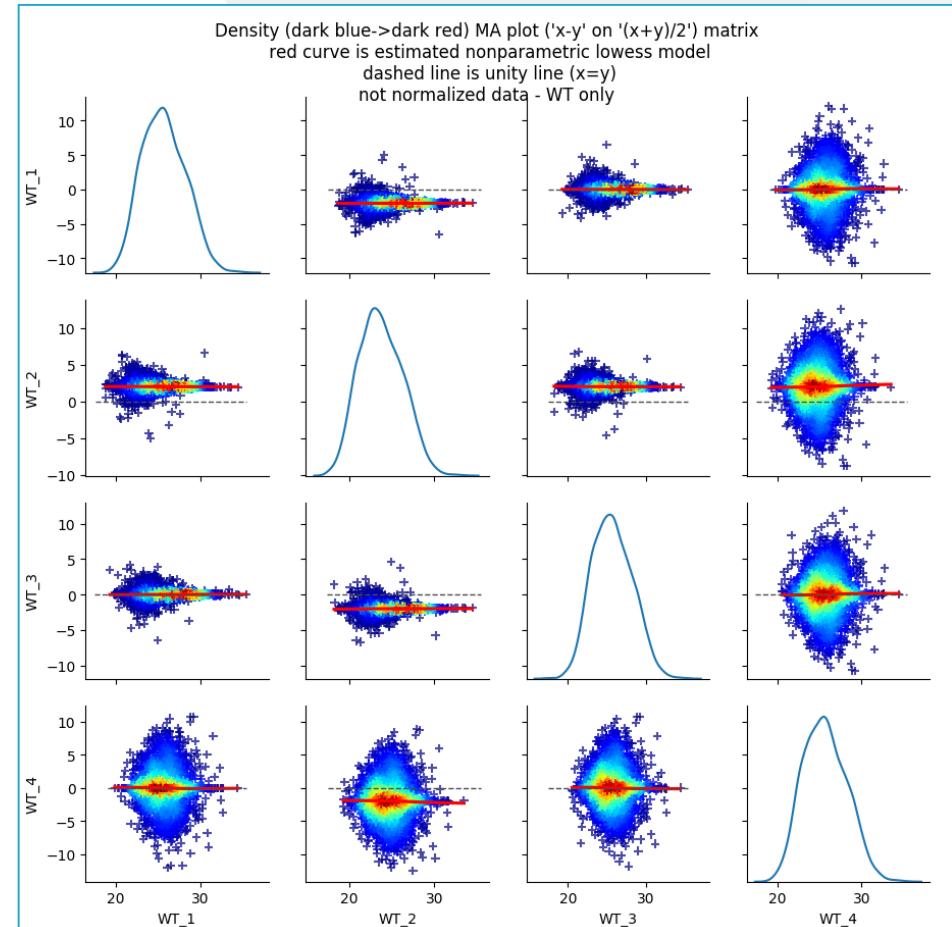
Metanodes examples – Scatter plot matrix



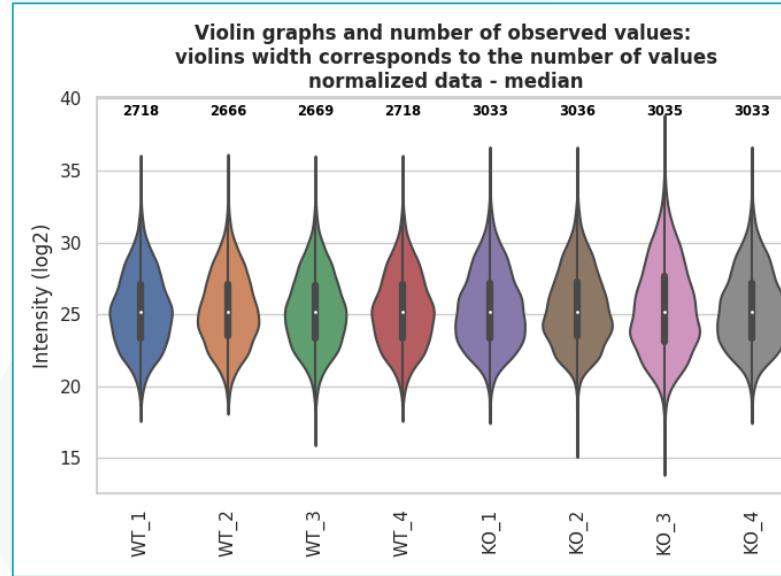
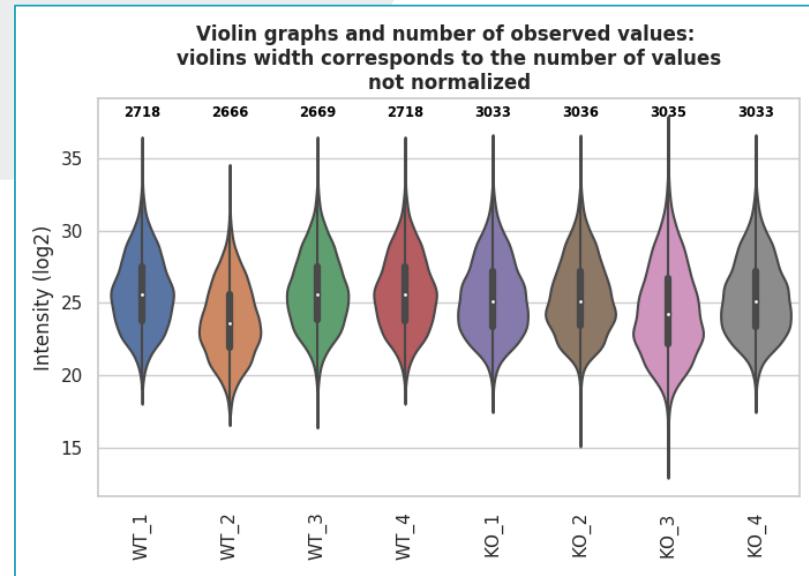
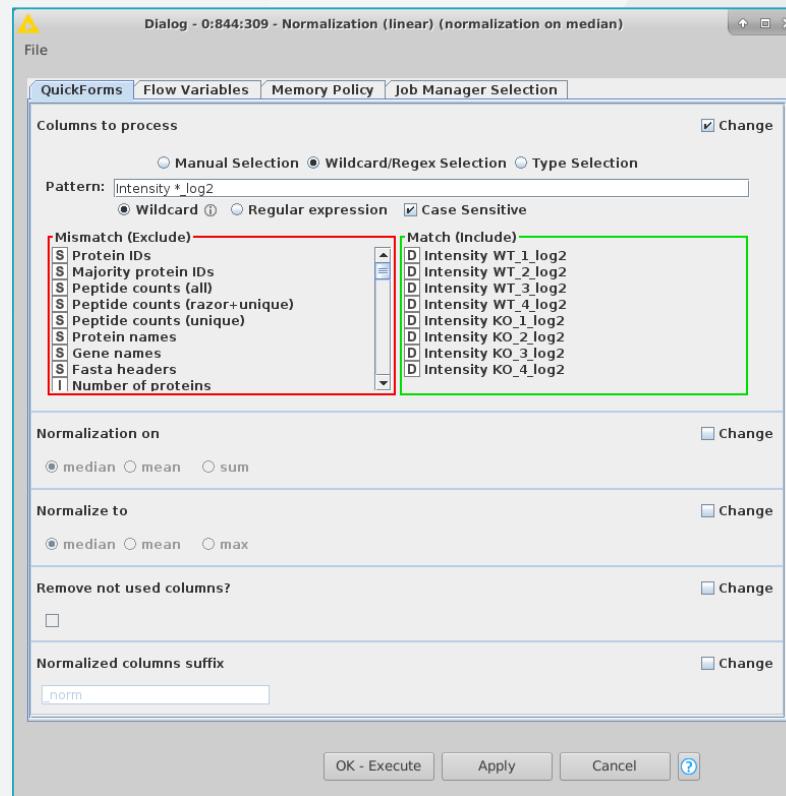
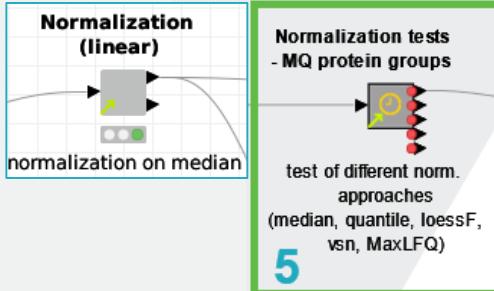
The screenshot shows the Metanodes software interface with various configuration options:

- Columns to process:**
 - Mismatch (Exclude):** Includes Protein IDs, Majority protein IDs, Peptide counts (all), Peptide counts (razor+unique), Peptide counts (unique), Protein names, Gene names, Fasta headers, and Number of proteins.
 - Match (Include):** Includes Intensity WT_1.log2, Intensity WT_2.log2, Intensity WT_3.log2, and Intensity WT_4.log2.
- Graphs subtitle:** not normalized data
- Fit regression curve?**: checked
- Type of regression:** loess
- Show unity line?**: checked
- Impute missing values?**: unchecked
- Value to impute if requested:** 0
- Construct MA plot like graph?**: checked
- manual x axis limits:** None
- manual y axis limits:** None
- Prefix to remove:** Intensity
- Suffix to remove:** .log2
- Size:** None

- Intensity correlation for all possible sample pairs
- Unity line ($y=x$) displayed as dashed line
- Regression curve (linear or lowess) displayable as well
- Symbols colors provide the density information (i.e. blue - low density; red – high density)
- Including possible imputation of missing values by constant
- Possibility to create MA plot like graph style
- Information on how individual protein group intensities are similar among samples
- Normalization “quality check”

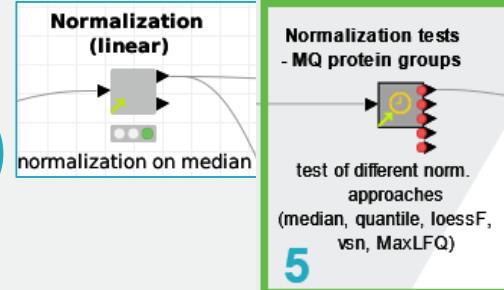


Metanodes examples – Normalization (linear)

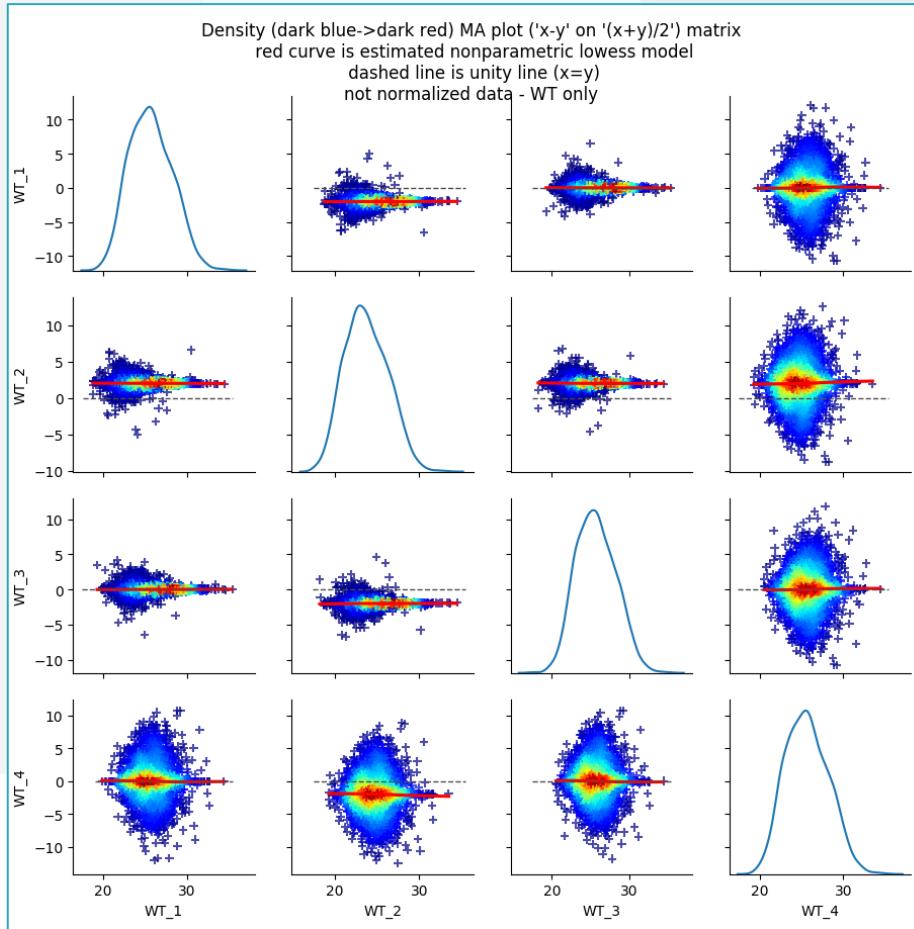


- Removal of **linear** technical processing artefacts from data
- Violin plots are just moved up or down to have the median value equal along all column selected for normalization
- subtracts e.g. column median ("Normalize on" option) from all columns and adds e.g. median ("Normalize to" option) of all medians prior the normalization step

Metanodes examples – Normalization (linear)

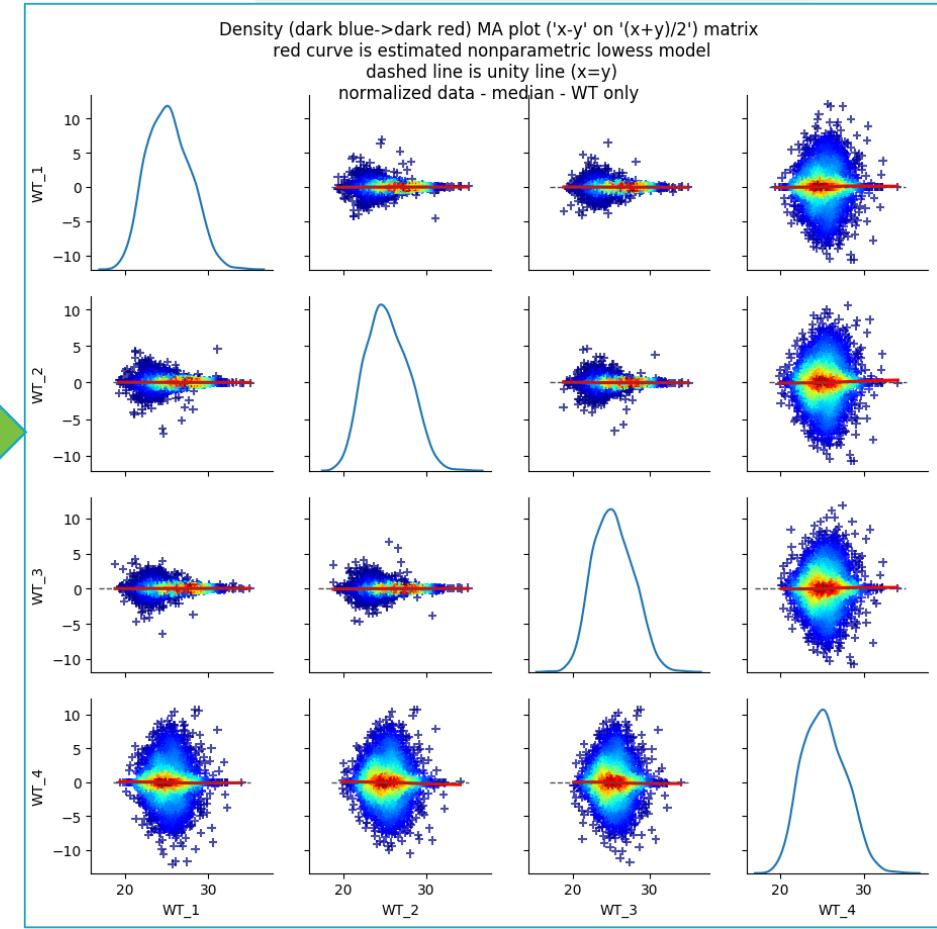


WT replicates only – prior normalization

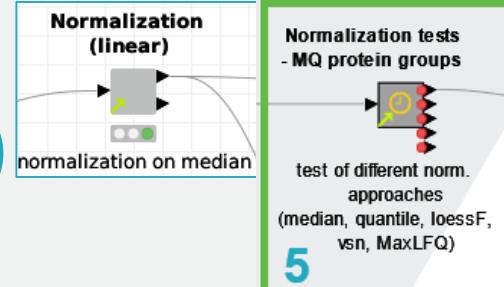


median
normalization

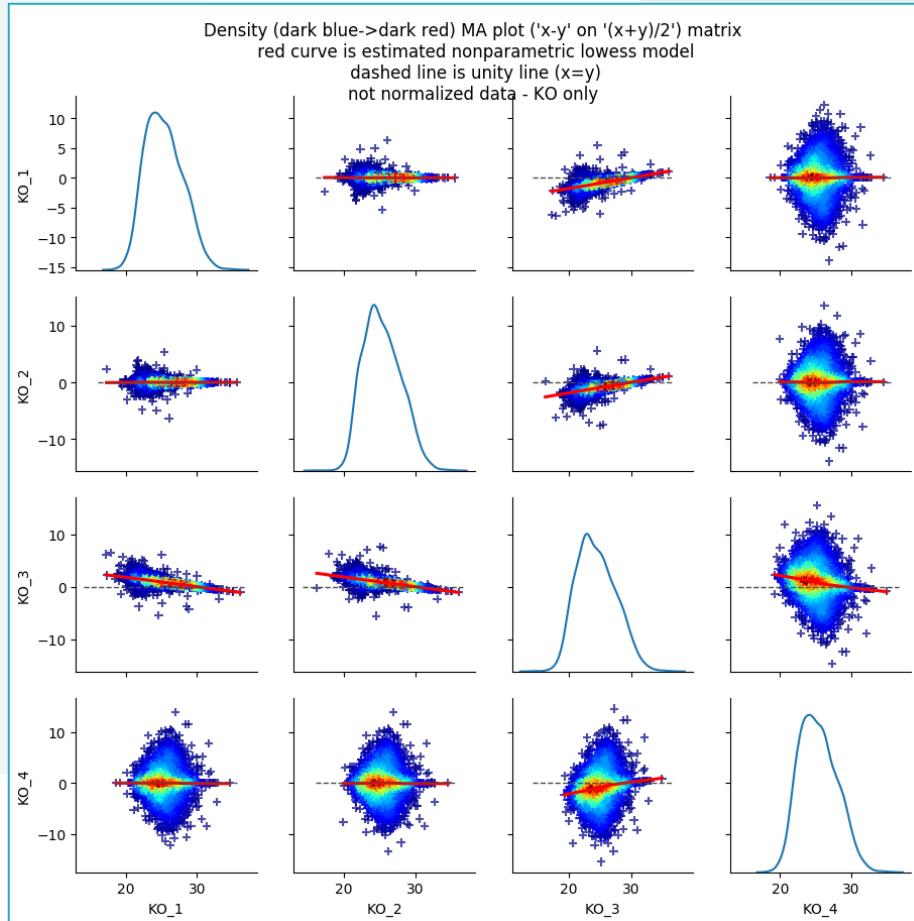
WT replicates only – after normalization



Metanodes examples – Normalization (linear)

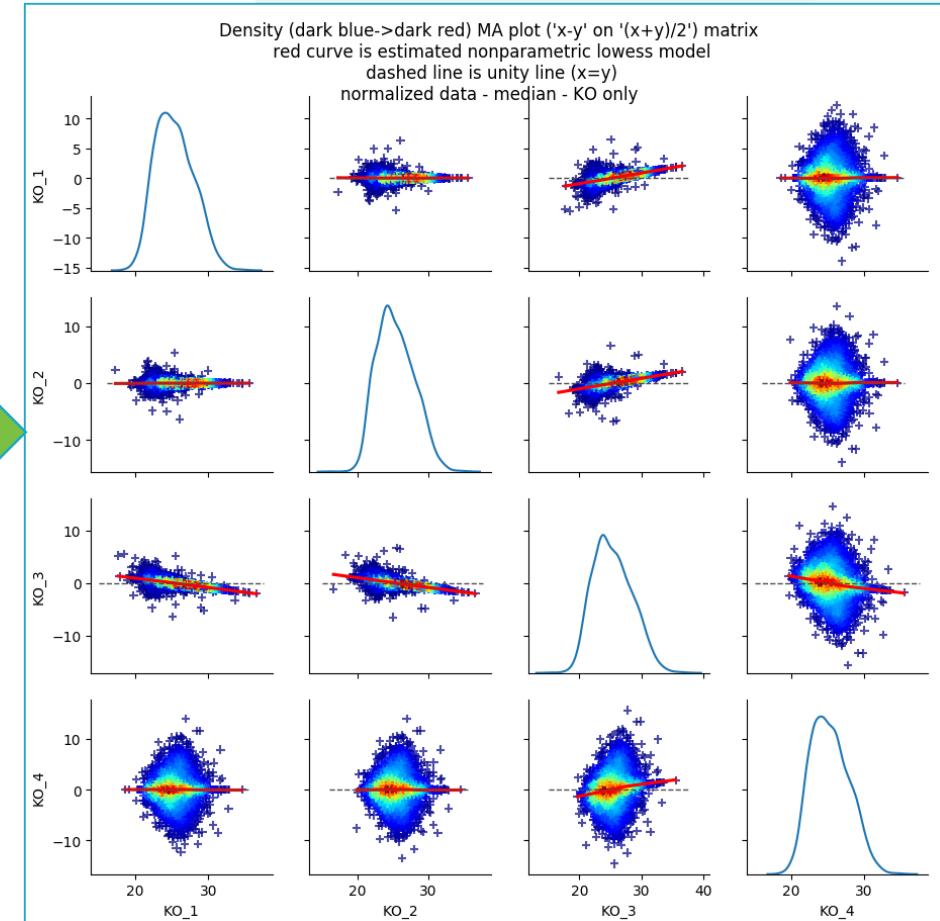


KO replicates only – prior normalization

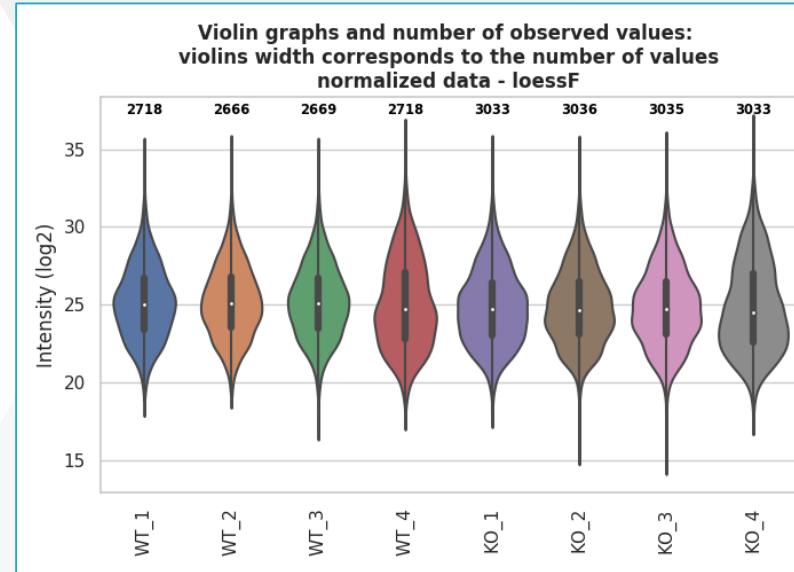
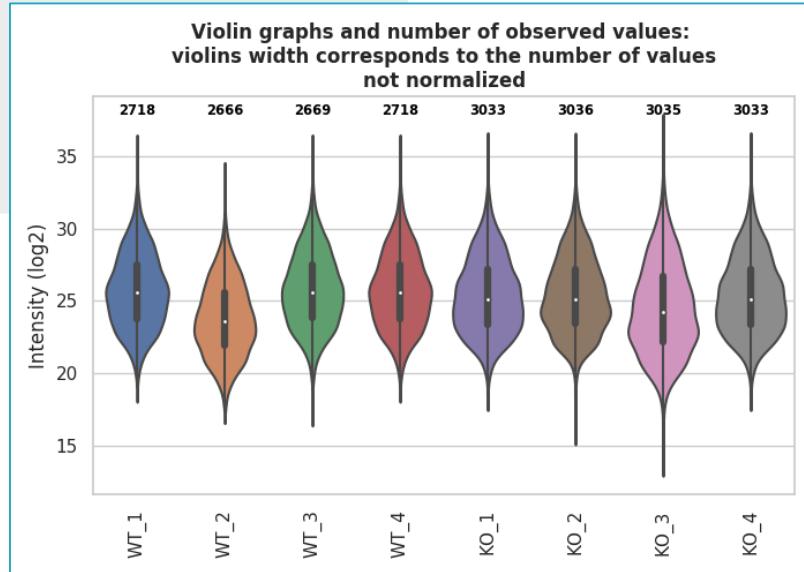
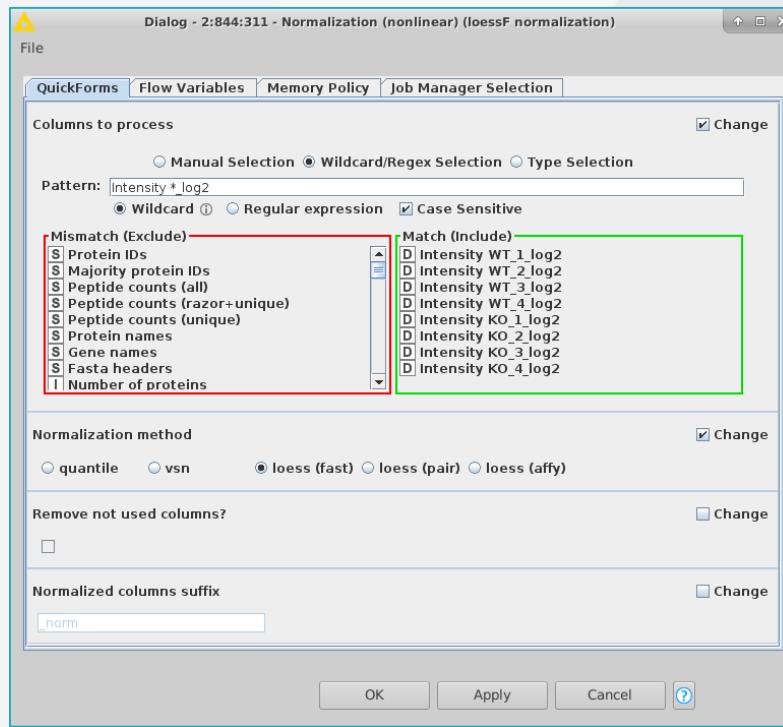
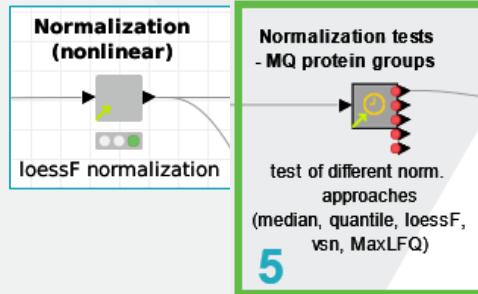


median
normalization

KO replicates only – after normalization

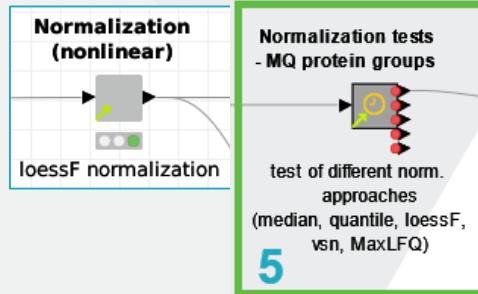


Metanodes examples – Norm. (nonlinear)

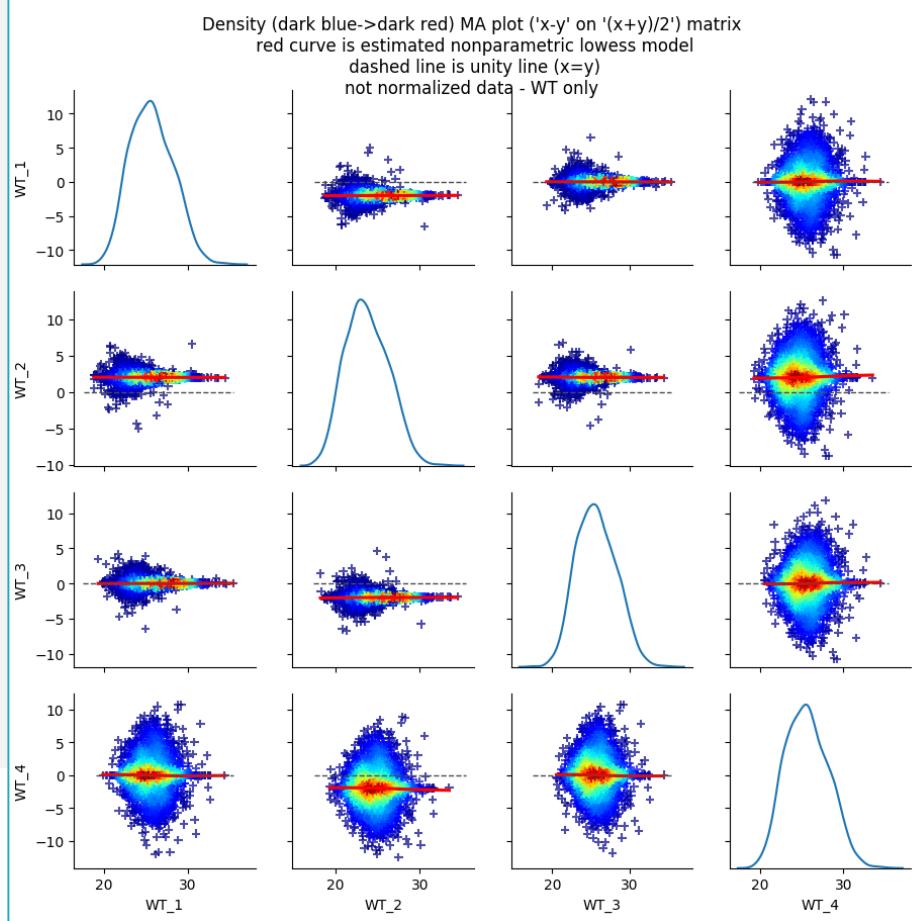


- Removal of **nonlinear** technical processing artefacts from data
- concrete data processing way depends on the selected normalization technique
- check the metanode description for implementation details and selected R packages for publications and algorithm details

Metanodes examples – Norm. (nonlinear)

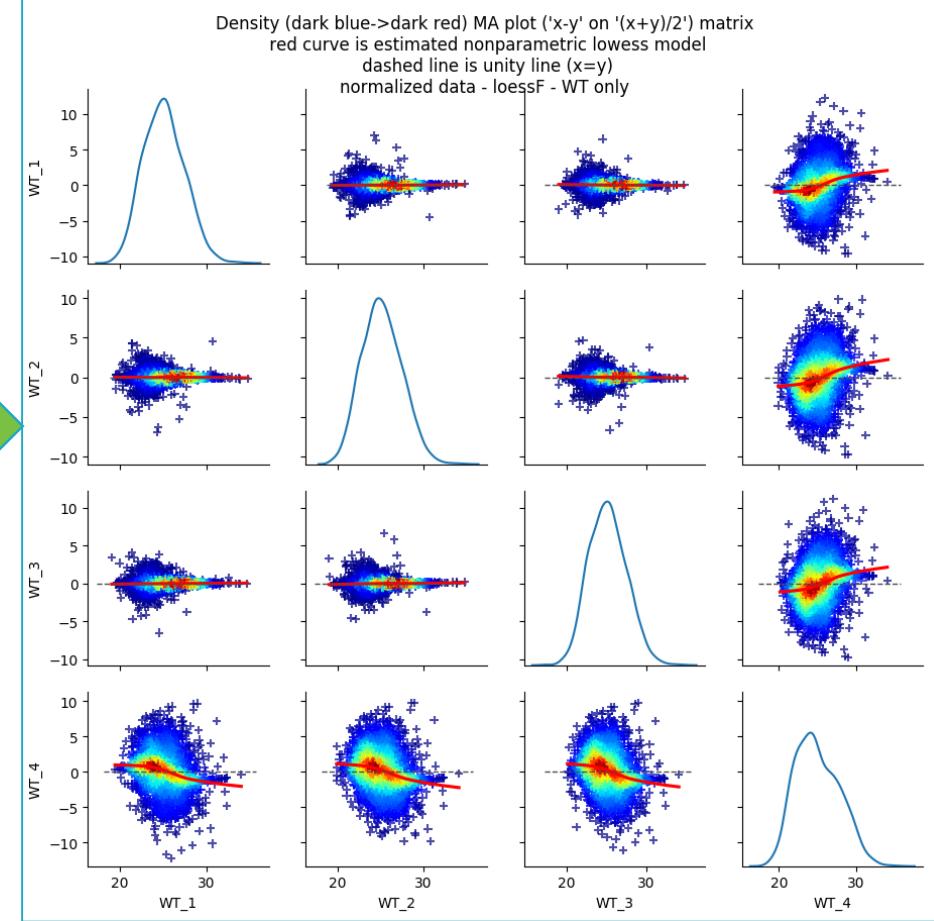


WT replicates only – prior normalization

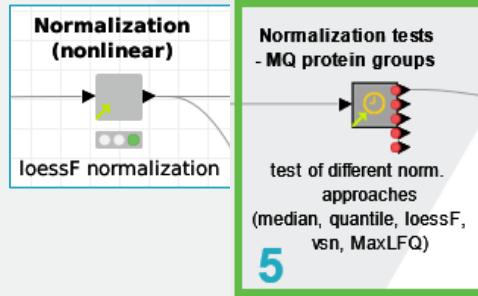


loess (fast)
normalization

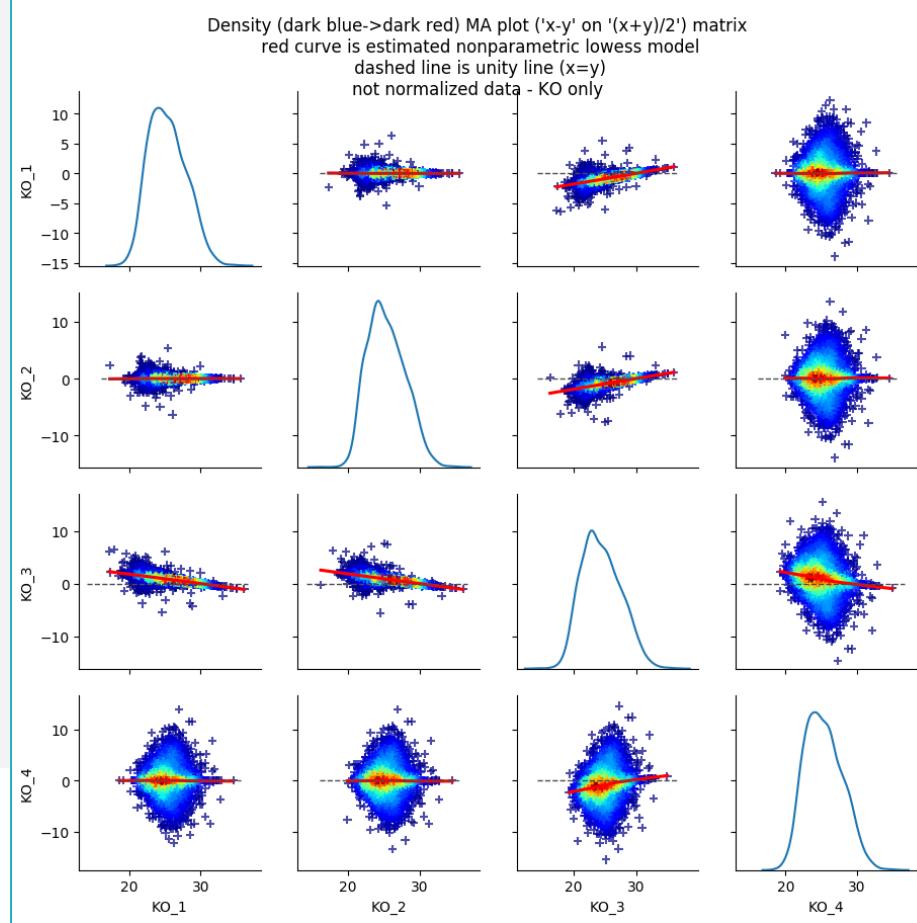
WT replicates only – after normalization



Metanodes examples – Norm. (nonlinear)

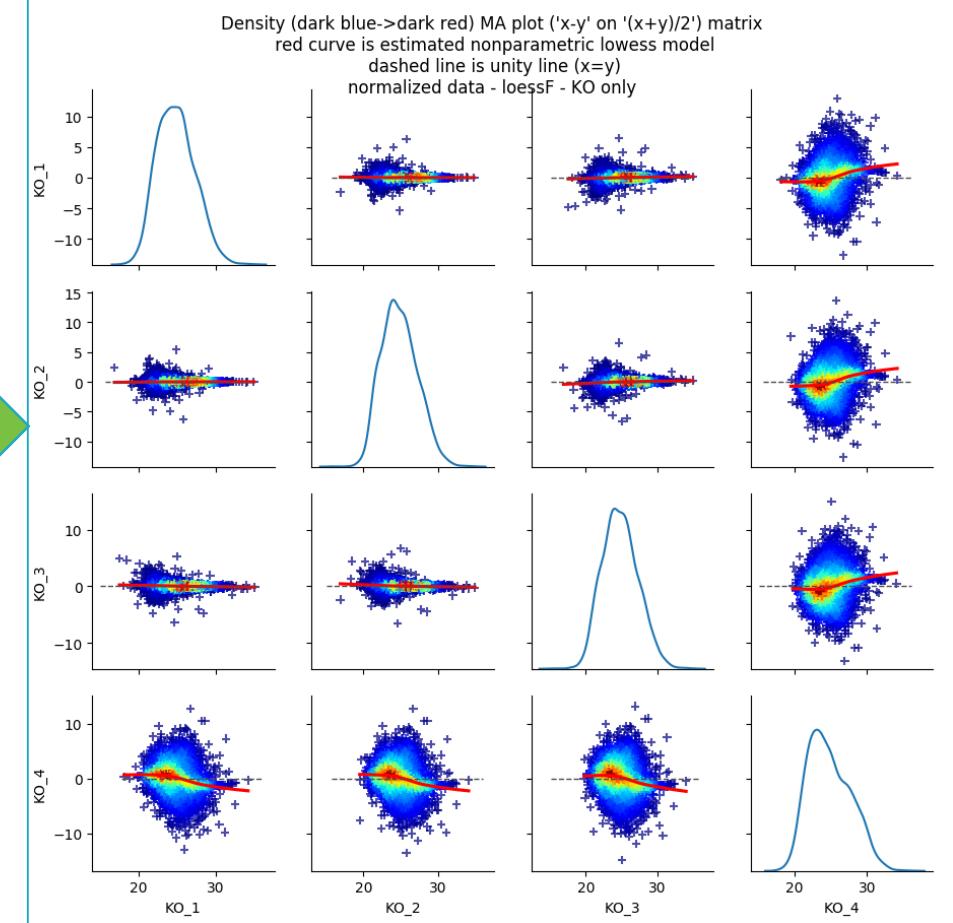


KO replicates only – prior normalization



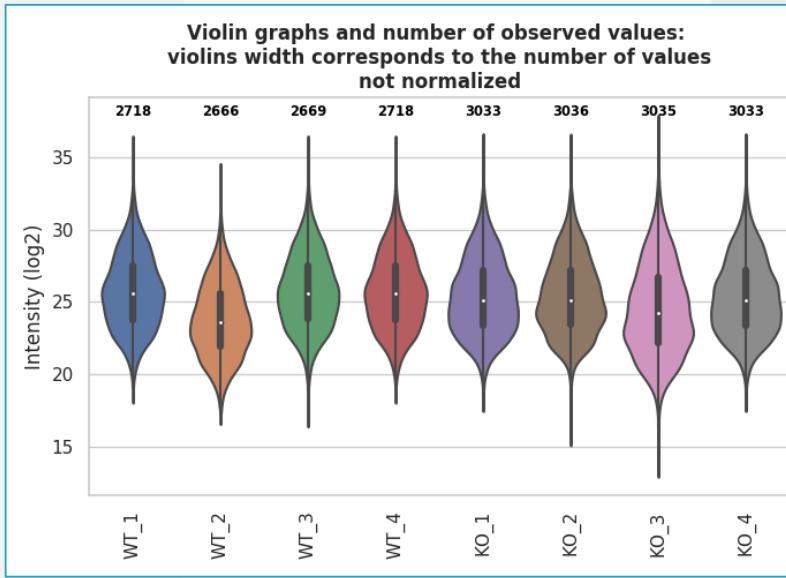
loess (fast)
normalization

KO replicates only – after normalization

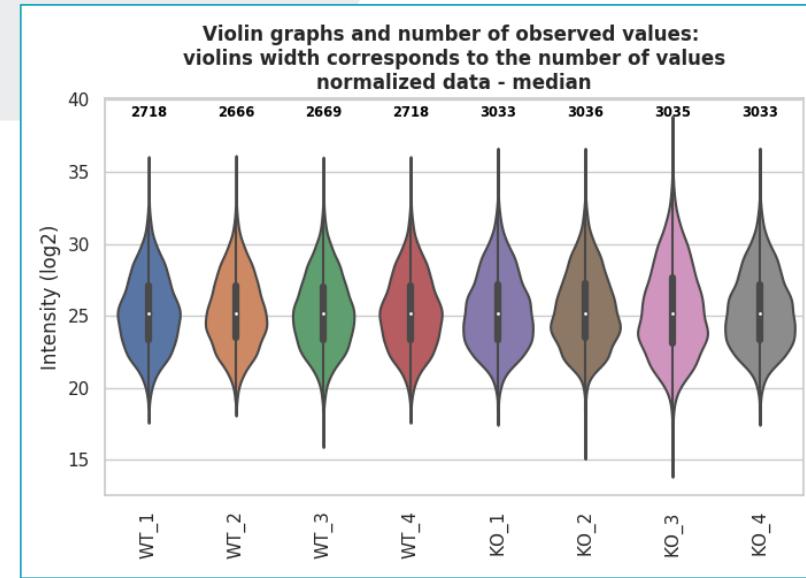


Normalization comparison – violin plots

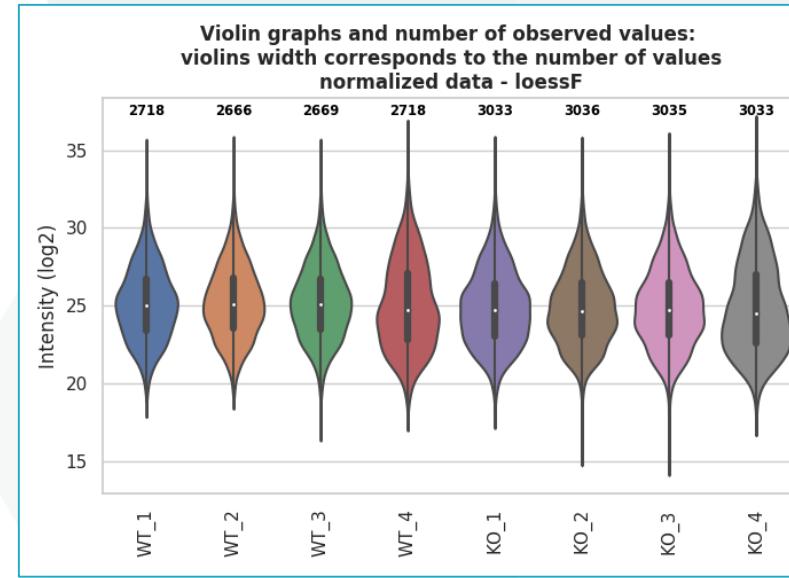
prior normalization



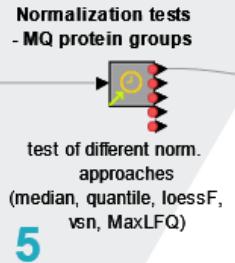
median normalization



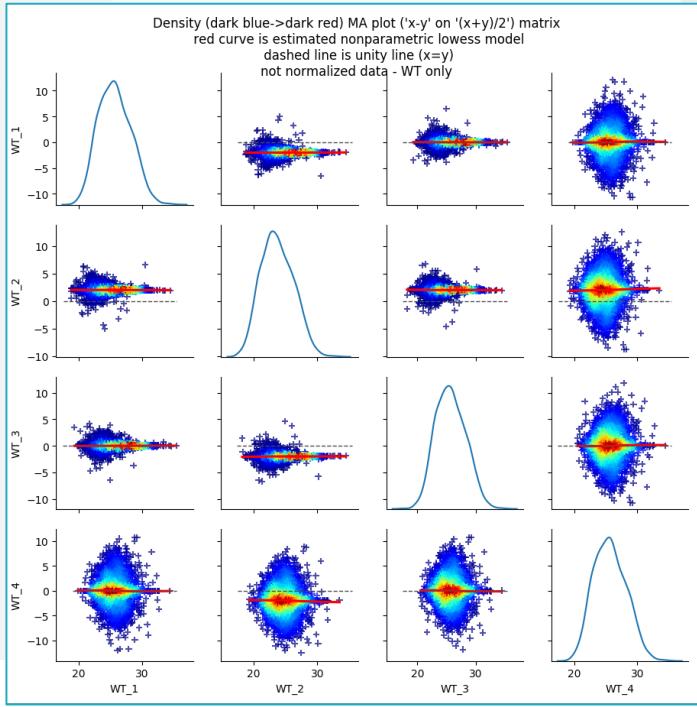
loess (fast) normalization



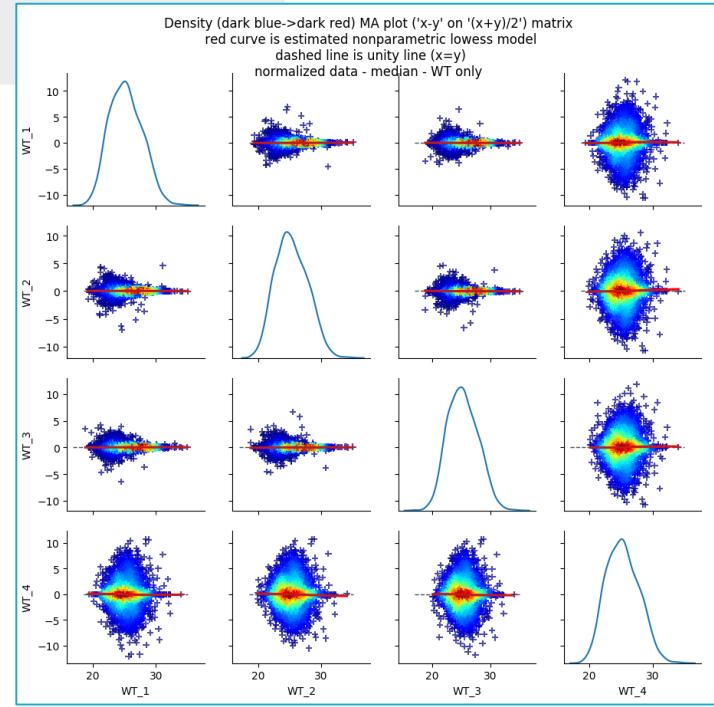
Normalization comparison – MA plots



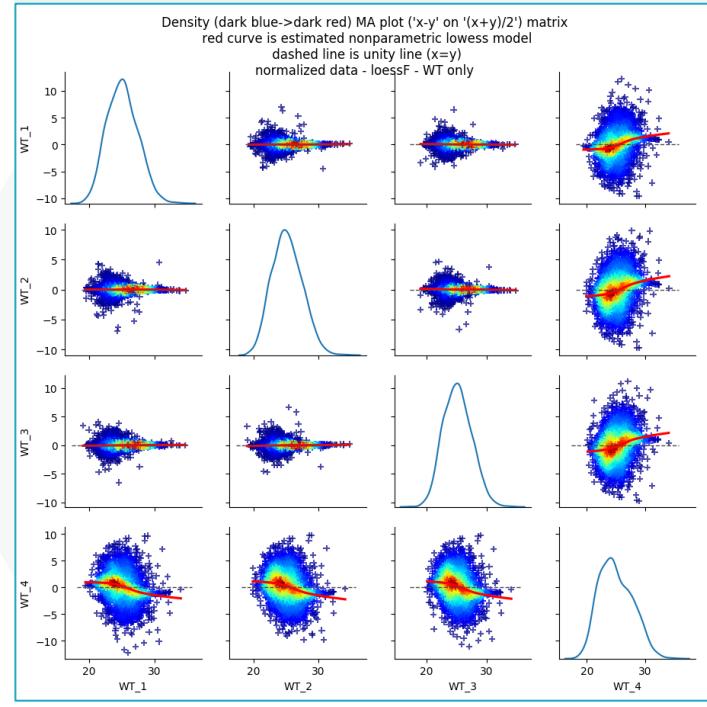
WT – prior normalization



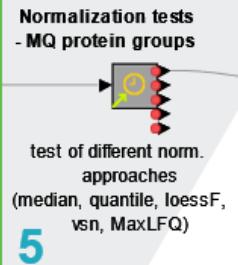
WT – median normalization



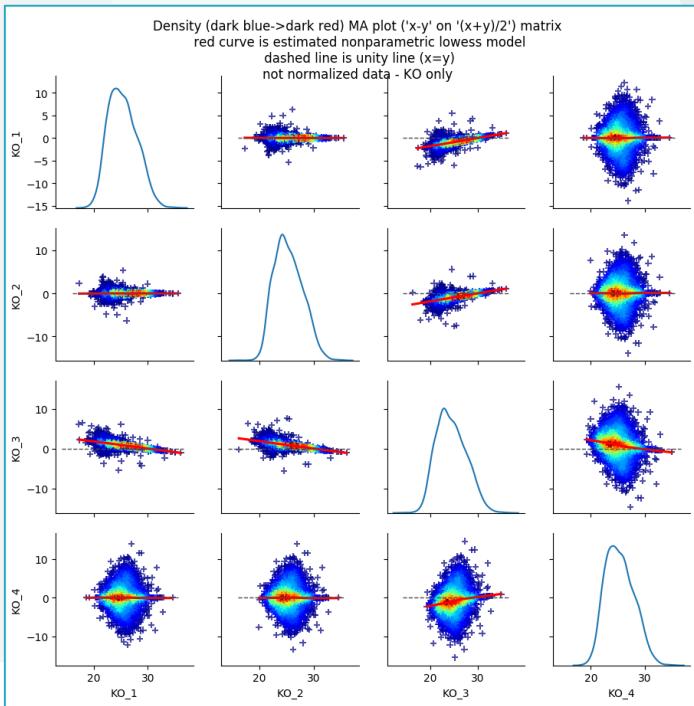
WT – loess (fast) normalization



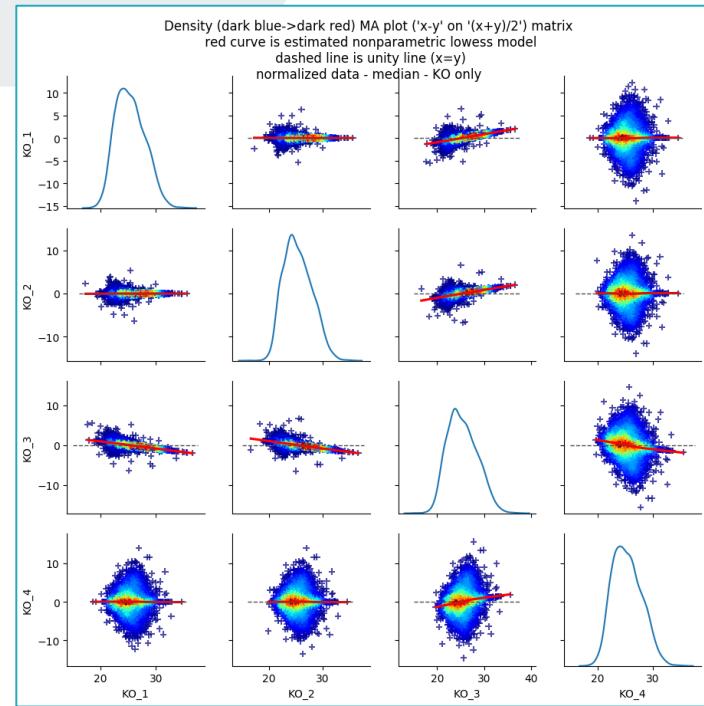
Normalization comparison – MA plots



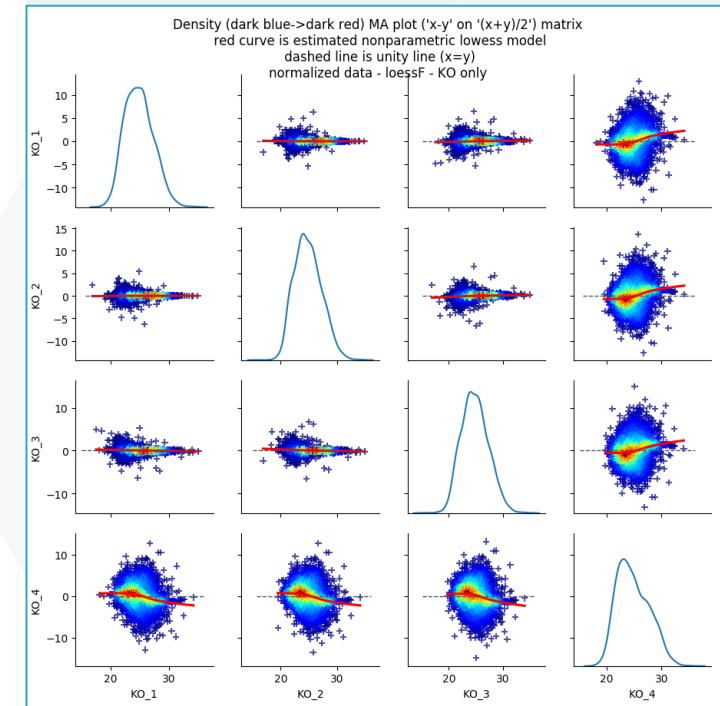
KO – prior normalization



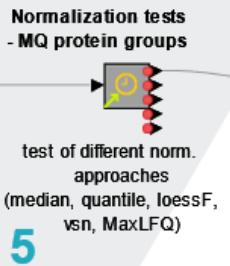
KO – median normalization



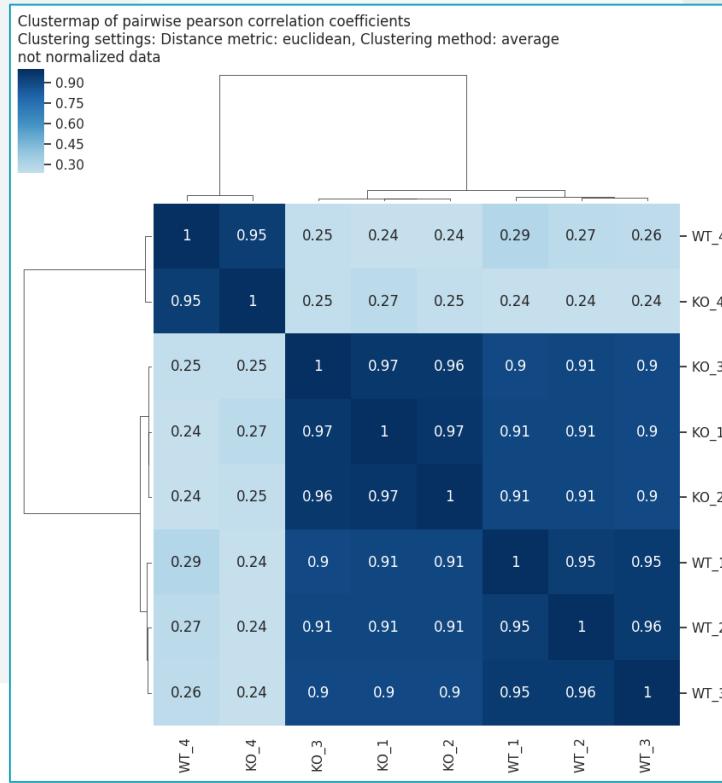
KO – loess (fast) normalization



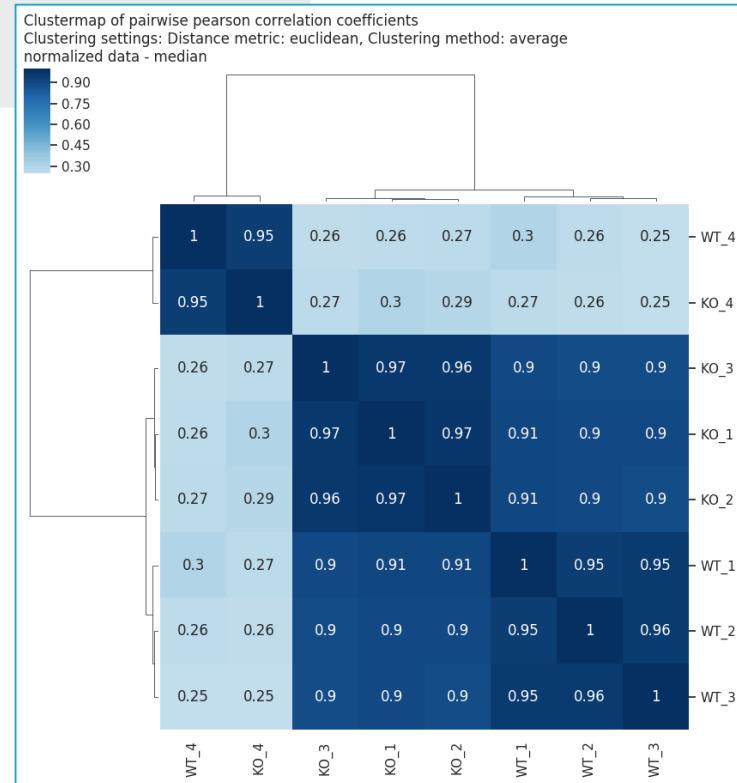
Normalization comparison – correlation clustermap



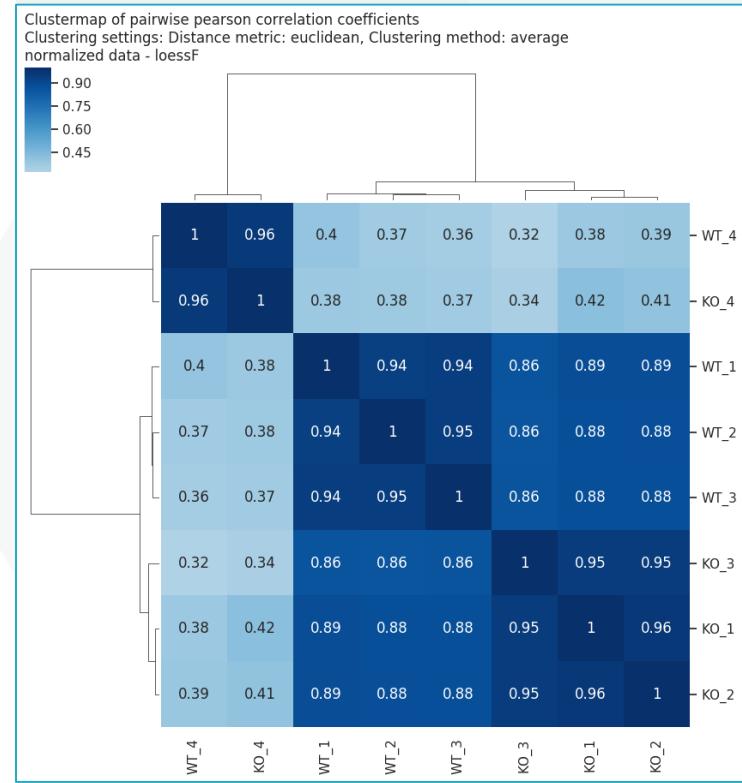
prior normalization



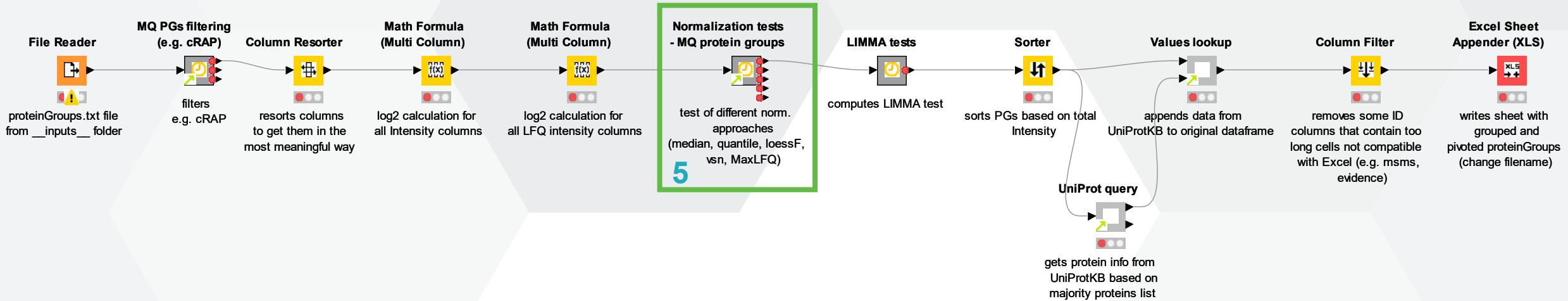
median normalization



loess (fast) normalization



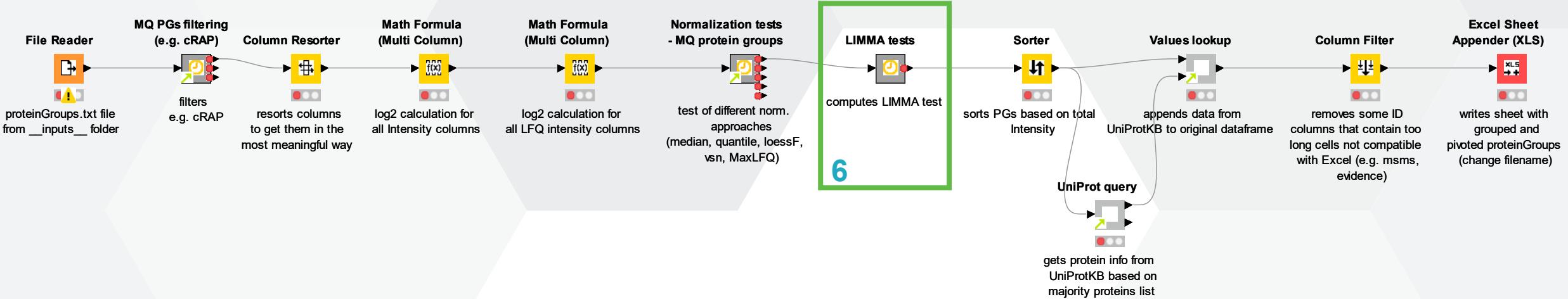
5. data normalization approach selection and data check



- selected outcomes of the normalization tests example

- relatively comparable protein numbers
- loessF preferred way of normalization
- better to remove the 4th replicates from both samples (MaxQuant processing step)
- different sample types cluster separately

6. statistical evaluation



Metanodes examples – LIMMA test

QuickForms Flow Variables Memory Policy Job Manager Selection

LIMMA test name Change
LIMMA_

Columns to process Change

Manual Selection Wildcard/Regex Selection Type Selection

Exclude Filter
 Intensity KO_2_log2
 Intensity KO_3_log2
 Intensity KO_4_log2
 Intensity WT_4_log2_norm
 Intensity WT_1_log2_norm
 Intensity WT_2_log2_norm
 Intensity WT_3_log2_norm
 Intensity WT_1_log2_norm_delog
 Intensity WT_2_log2_norm_delog
 Intensity WT_3_log2_norm_delog
 Enforce exclusion

Include Filter
 Intensity WT_1_log2_norm
 Intensity WT_2_log2_norm
 Intensity WT_3_log2_norm
 Intensity KO_1_log2_norm
 Intensity KO_2_log2_norm
 Intensity KO_3_log2_norm
 Enforce inclusion

Limma design Change
WT,WT,WT,
KO,KO,KO

Samples pairs Change
none

Samples blocks Change
b1,b2,b3,
b1,b2,b3

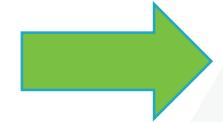
Comparisons (contrasts) Change
KO-WT

Treat comparisons separately?

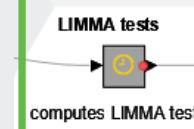
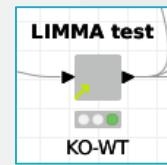
p-value adjustment method Benjamini & Hochberg

Remove not used columns?

- initial statistical evaluation
- proper design selection
 - paired design?
 - batches?
- does not handle missing values on its own, preferable to deal with them beforehand
 - needs at least 1 value in all sample types
- does not filter tested data based on e.g. number of numerical values in one sample type
 - good to consider before the final statistical evaluation
 - p-values of individual protein groups not affected; adjustment more/less stringent based on the number of tested protein groups
- important to remove any outlying sample beforehand

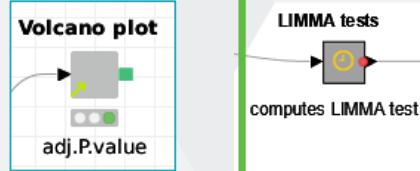


| D | LIMMA_K_O-WT.logFC | D | LIMMA_K_O-WT.AveExpr | D | LIMMA_K_O-WT.t | D | LIMMA_K_O-WT.P.Value | D | LIMMA_K_O-WT.... | D | LIMMA_K_O-V |
|--------|--------------------|---------|----------------------|-------|----------------|---|----------------------|---|------------------|---|-------------|
| -9.733 | 26.548 | -10.374 | 0 | 0 | 44.932 | | | | | | |
| 9.14 | 27.739 | 9.742 | 0 | 0 | 38.907 | | | | | | |
| -9.336 | 26.101 | -9.382 | 0 | 0 | 35.431 | | | | | | |
| -8.114 | 25.797 | -8.649 | 0 | 0 | 29.38 | | | | | | |
| 7.225 | 28.22 | 7.701 | 0 | 0 | 22.038 | | | | | | |
| 6.991 | 27.612 | 7.452 | 0 | 0 | 20.246 | | | | | | |
| 6.846 | 27.693 | 7.297 | 0 | 0 | 19.164 | | | | | | |
| -6.376 | 25.326 | -6.797 | 0 | 0 | 15.82 | | | | | | |
| 6.168 | 24.812 | 6.574 | 0 | 0 | 14.411 | | | | | | |
| 9.211 | 24.801 | 6.366 | 0 | 0 | 12.244 | | | | | | |
| 7.326 | 26.304 | 6.376 | 0 | 0 | 12.893 | | | | | | |
| -5.904 | 24.338 | -6.293 | 0 | 0 | 12.699 | | | | | | |
| -5.893 | 28.173 | -6.281 | 0 | 0 | 12.624 | | | | | | |
| -5.767 | 27.108 | -6.147 | 0 | 0 | 11.835 | | | | | | |
| 6.1 | 24.891 | 6.13 | 0 | 0 | 11.675 | | | | | | |
| 5.539 | 26.877 | 5.904 | 0 | 0 | 10.448 | | | | | | |
| -5.535 | 23.023 | -5.899 | 0 | 0 | 10.422 | | | | | | |
| -5.506 | 26.142 | -5.869 | 0 | 0 | 10.255 | | | | | | |
| -5.761 | 25.997 | -5.79 | 0 | 0 | 9.768 | | | | | | |
| 5.396 | 26.941 | 5.751 | 0 | 0 | 9.605 | | | | | | |
| -5.4 | 25.554 | -5.756 | 0 | 0 | 9.628 | | | | | | |
| 9.915 | 26.444 | 5.73 | 0 | 0 | 8.31 | | | | | | |
| 5.549 | 26.064 | 5.577 | 0 | 0 | 8.627 | | | | | | |
| -5.045 | 24.425 | -5.377 | 0 | 0 | 7.632 | | | | | | |
| -4.973 | 24.28 | -5.301 | 0 | 0 | 7.247 | | | | | | |
| -4.93 | 25.424 | -5.255 | 0 | 0 | 7.014 | | | | | | |
| 4.826 | 28.285 | 5.144 | 0 | 0 | 6.466 | | | | | | |
| -4.804 | 25.288 | -5.121 | 0 | 0 | 6.356 | | | | | | |
| 4.74 | 26.655 | 5.052 | 0 | 0 | 6.025 | | | | | | |
| 5.615 | 24.714 | 4.887 | 0 | 0 | 5.149 | | | | | | |
| 4.562 | 27.187 | 4.863 | 0 | 0 | 5.134 | | | | | | |
| -4.498 | 25.271 | -4.794 | 0 | 0 | 4.82 | | | | | | |
| 4.762 | 24.863 | 4.786 | 0 | 0 | 4.767 | | | | | | |
| 5.146 | 24.374 | 4.478 | 0 | 0.001 | 3.381 | | | | | | |
| 4.128 | 26.09 | 4.4 | 0 | 0.001 | 3.105 | | | | | | |
| 4.049 | 25.65 | 4.316 | 0 | 0.001 | 2.755 | | | | | | |
| -6.219 | 26.166 | -4.298 | 0 | 0.001 | 2.495 | | | | | | |
| 7.255 | 24.117 | 4.193 | 0 | 0.002 | 1.885 | | | | | | |
| 3.933 | 28.547 | 4.193 | 0 | 0.002 | 2.259 | | | | | | |
| -3.901 | 24.147 | -4.158 | 0 | 0.002 | 2.121 | | | | | | |
| 3.85 | 23.063 | 4.104 | 0 | 0.003 | 1.909 | | | | | | |
| 6.996 | 24.445 | 4.043 | 0 | 0.004 | 1.366 | | | | | | |



6

Metanodes examples – Volcano plot



QuickForms

Flow Variables

Log fold change column Change
LIMMA_KO-WT.logFC

p-value column selection Change
LIMMA_KO-WT.adj.P.Val

Apply -log10 on p-value column? Change

Zero p value replacement Change
0.001

x axis (logFC) threshold Change
1

y axis (p value) threshold Change
0.05

Graphs subtitle Change

x axis label Change

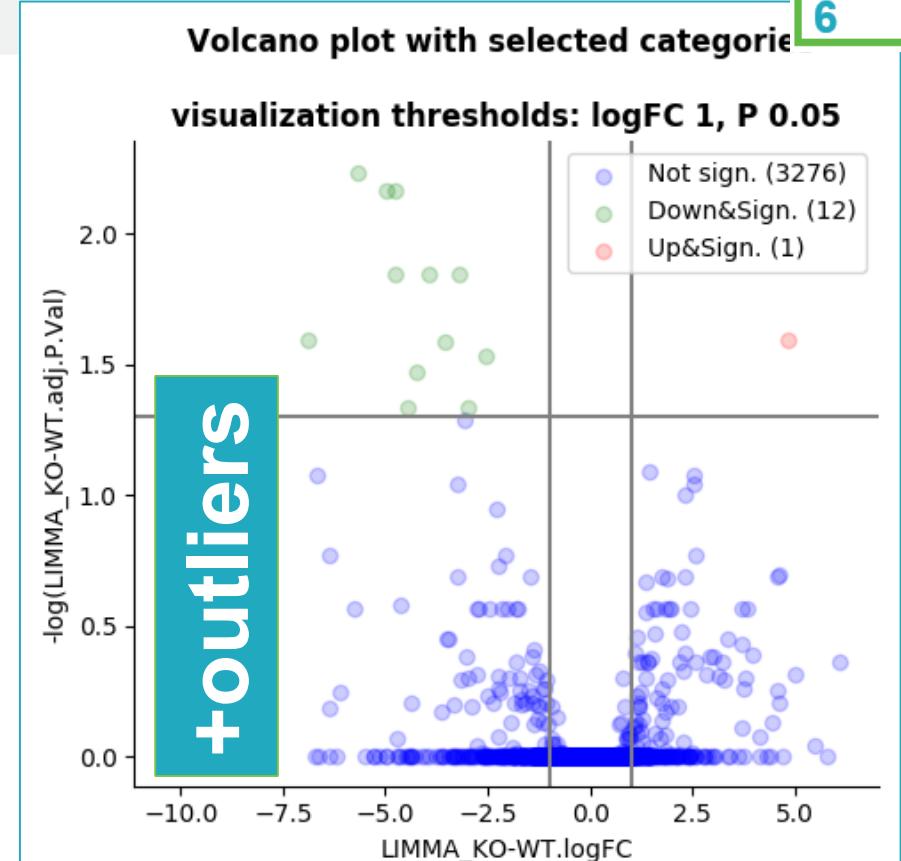
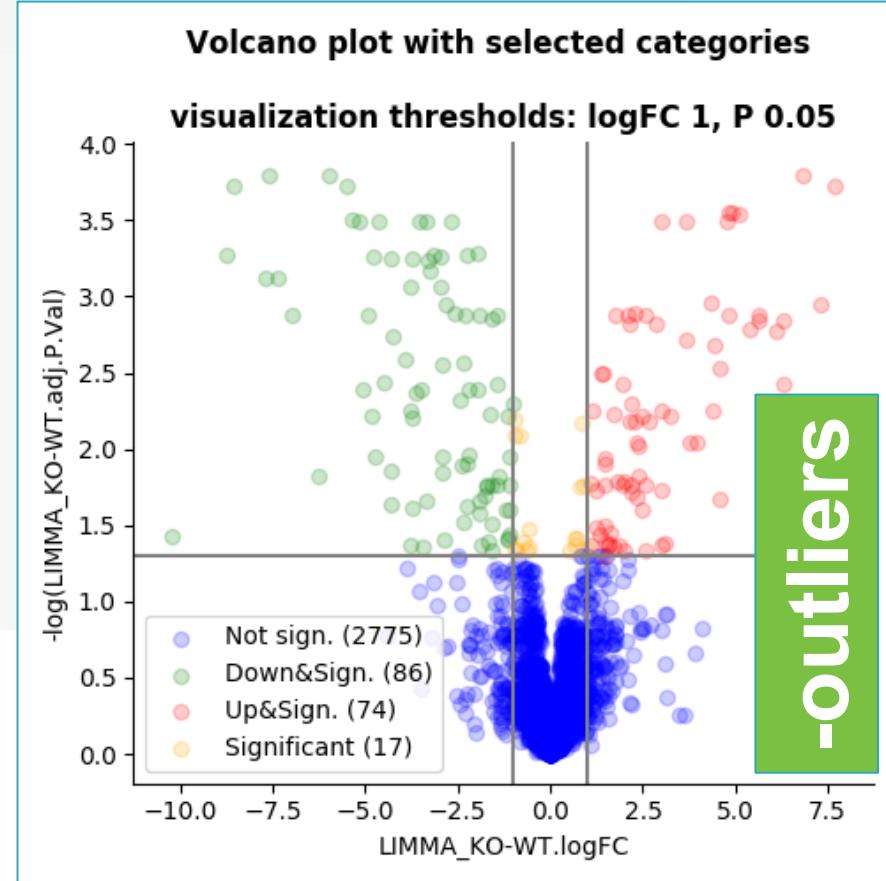
y axis label Change

Manual x axis limits? Change

manual x axis limits -10:10

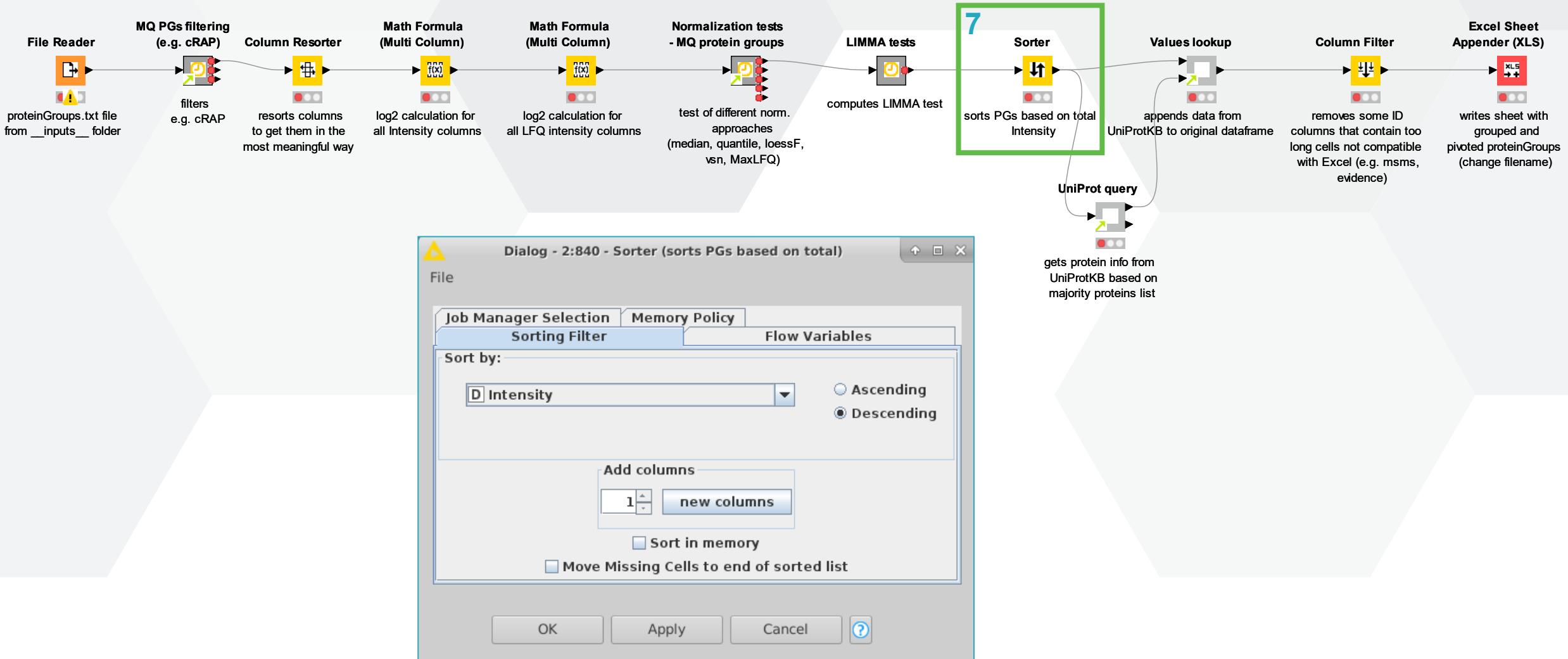
Manual y axis limits? Change

manual y axis limits -10:10

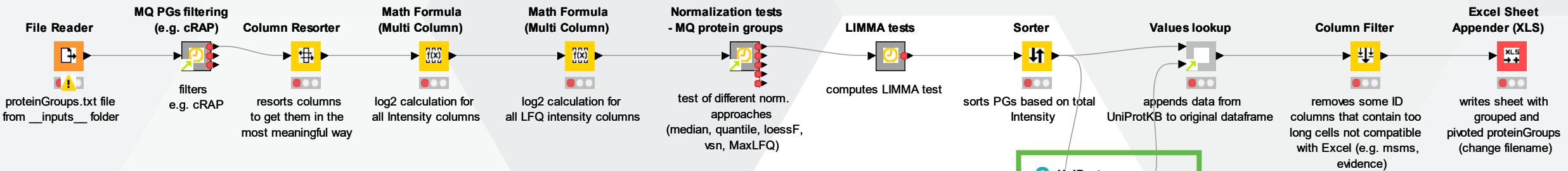


- visualize results of the statistical tests (logFC and P/adj.P values)
- categorize the results based on the specified threshold

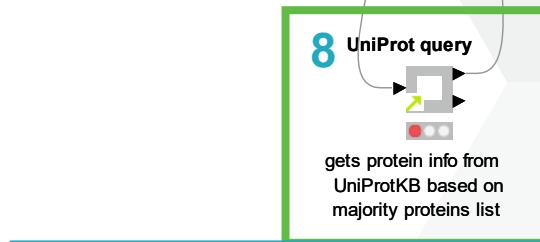
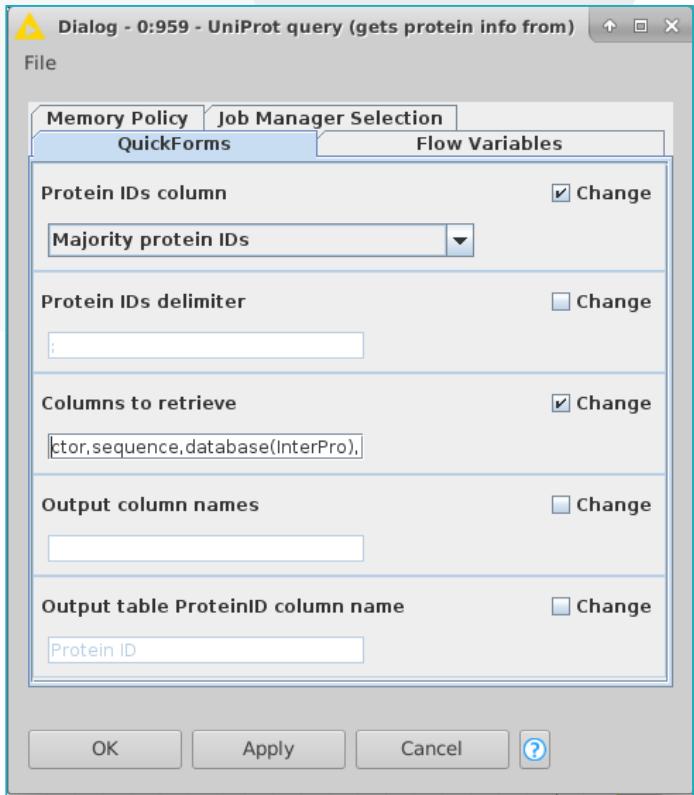
7. resorts the table (for reporting purposes only)



7. gets proteins annotation from UniProtKB



| Majority protein IDs |
|---------------------------------|
| A0A096LPI6 |
| A0A0A6YYA0;Q9Y3B3 |
| A0A0A6YYH1;P59780 |
| A0A0A6YVL6;P18621 |
| A0A0B4J1V8;Q9NQ55 |
| A0A0B4J269;Q13509 |
| A0A0B4J2E5;Q15269 |
| A0A0J9YWU9;P0DP08;P0DP06;A0A... |
| A0A0U1RQV1 |
| A0A1B0GU03;P07339 |
| A0A1W2PNV4 |
| A0A1W2PQS6 |
| A0A1W2PRB8;Q9H0W9 |
| A0AVT1 |
| A0FGR8 |
| A1L0T0 |
| A3KN83 |
| A4D1E9 |
| A5YK6 |
| A6NDG6 |
| A6NFQ2;A0A1B0GVM2 |
| A6NHQ2 |
| A6NHR9 |
| A6NJ78 |
| A6NKT7;Q7Z3J3 |
| A8CG34;Q96HA1 |



Data retrieved from UniProt

| Row ID | Gene names | Organism | Gene ontology (GO) | Gene ontology (biological process) | Gene ontology (molecular function) | Gene ontology (cellular component) | Subcellular location (CC) | Protein families |
|--------|------------------------------|----------------------------------------|------------------------------------|-----------------------------------------|------------------------------------|------------------------------------|---------------------------|---------------------------------|
| Row0 | PCMP_E22_4t2g02750 T20F6.11 | Arabidopsis thaliana (Mouse-ear cress) | RNA modification [GO:0009451] | RNA modification [GO:0009451] | | | | PPC family, PCMP-E subfamily |
| Row1 | THIUMP03 | Homo sapiens (Human) | cytosol [GO:0005829]; nucleo... | RNA methylation [GO:0030488] | | | | Methyltransferase superfamily |
| Row2 | PADI_PRC6A_A13p51260 F24M... | Arabidopsis thaliana (Mouse-ear cress) | cytoplasm [GO:0005737]; cyto... | proteasomal protein catabolic... | endopept... | | | Peptidase T1A family |
| Row3 | GPT1_GFAT_GFPT | Homo sapiens (Human) | cytosol [GO:0005829]; extra... | proteasomal regulation of gene e... | carboxyhy... | | | |
| Row4 | MDC1_KIAA0170_NFBDI | Homo sapiens (Human) | chromosome [GO:0005694]; fu... | double-strand break repair vi... | FHA dom... | chromos... | | |
| Row5 | DDX24 | Homo sapiens (Human) | membrane [GO:0016202]; nu... | RNA metabolic process [GO:0... | membran... | | | |
| Row6 | PRMT1_HMT2_HMTL2_I1B4 | Homo sapiens (Human) | cytoplasm [GO:0005737]; cyto... | cell surface receptor signalin... | enzymic b... | cell surface... | | |
| Row7 | SLC1AS ASCT2_M7V1_RDR RDRC | Homo sapiens (Human) | extracellular exosome [GO:00... | amino acid transport [GO:000... | amino aci... | extracell... | | |
| Row8 | AXS2_4t1g08200 T23G18.6 | Arabidopsis thaliana (Mouse-ear cress) | apoplast [GO:0048046]; cytop... | cell wall organization [GO:00... | catalytic ac... | apoplast... | | NAD(P)-dependent epimer... |
| Row9 | POLD3_KIAA0039 | Homo sapiens (Human) | ficolin-1-rich granule membra... | mitochondrial organization [GO... | | | | |
| Row10 | SLC3A2_MDU1 | Homo sapiens (Human) | amino acid transport complex... | amino acid transport [GO:000... | amino aci... | | | DEAD box helicase family, ... |
| Row11 | CLPPS_NCLP1_NCLP5_A1g102... | Arabidopsis thaliana (Mouse-ear cress) | chloroplast [GO:0009507]; ch... | protein quality control for mis... | ATPase bind... | chloropla... | | Class II SAM-binding me... |
| Row12 | EIF4EBP2 | Homo sapiens (Human) | cytoplasm [GO:0005737]; pos... | cAMP-mediated signalling [GO:... | chaperon... | cytoplas... | | Dicarboxylate/amino acid:c... |
| Row13 | AT5g45160_K18C1.4 | Arabidopsis thaliana (Mouse-ear cress) | cytoplasm [GO:0005737]; end... | endoplasmic reticulum membr... | cytoplas... | cytoplas... | | Peptidase S14 family |
| Row14 | HNRNPU_C1orf199_HNRNP_SAF... | Homo sapiens (Human) | catalytic step 2 spliceosome [... | adaptive thermogenesis [GO:... | actin bin... | catatytic... | | EIF4E-binding protein family |
| Row15 | ESD | Homo sapiens (Human) | cytoplasmic vesicle [GO:0031... | formaldehyde catabolic proce... | cytoplas... | cytoplas... | | TRAFAC class dynamin-like... |
| Row16 | ZNF638_NP220_ZFME | Homo sapiens (Human) | cytoplasm [GO:0005737]; intr... | RNA splicing [GO:0008380]; do... | double-str... | cytoplas... | | Esterase D family |
| Row17 | POLD3_KIAA0039 | Homo sapiens (Human) | cytoplasm [GO:0005737]; delt... | RNA damage response, detect... | cytoplas... | cytoplas... | | |
| Row18 | NTR2_NTRA_A1g17420 F5j6.18 | Arabidopsis thaliana (Mouse-ear cress) | mitochondrial matrix [GO:000... | removal of superoxide radical... | thioredox... | mitochon... | | |
| Row19 | ACCI_EM822_QK_PAS3_A1g3... | Arabidopsis thaliana (Mouse-ear cress) | cytosol [GO:0005829]; nucleo... | embryo development ending ... | acetyl co... | cytosol... | | Class-II pyridine nucleotide... |
| Row20 | FS1_P51g47770_MCA23.9 | Arabidopsis thaliana (Mouse-ear cress) | cytoplasm [GO:0005737]; cyto... | cholesterol biosynthetic proc... | dihydroxy... | | | Dicarboxylate/amino acid:c... |
| Row21 | CHLD_ALB1_PDE166_A1g0525... | Arabidopsis thaliana (Mouse-ear cress) | chloroplast [GO:0005737]; cyto... | cholesterol biosynthetic proc... | ATP bind... | chloropla... | | PPF/GPP synthase family |
| Row22 | NC_2014768 | Homo sapiens (Human) | cell body [GO:0044297]; chag... | neurophil degranulation [GO:... | cell body... | | | Mg-chelatase subunits Dif... |
| Row23 | CCT3_CCTG_TRIC5 | Homo sapiens (Human) | contractile ring [GO:0070938];... | binding of sperm to zona pell... | ATP bindin... | cell body... | | |
| Row24 | PROK2AP4 | Homo sapiens (Human) | cytoskeleton [GO:0002376]; act... | cell division [GO:0051301]; mi... | identical ... | contractin... | | TCP-1 chaperonin family |
| Row25 | CAP21 | Homo sapiens (Human) | actin filament [GO:0020757]; ac... | actin filament dynamics [GO:... | actin fil... | actin fil... | | MAP65/ASE1 family |
| Row26 | SCAMP1_SCAMP | Homo sapiens (Human) | clathrin-coated vesicle [GO:00... | neurophil degranulation [GO:... | clathrin-c... | clathrin-c... | | SCAMP family |
| Row27 | PPIC1_A1g47700_MCA23.2 | Arabidopsis thaliana (Mouse-ear cress) | cytosol [GO:0005829]; cytopla... | cytoplasmic translation [GO:... | protein ki... | cytosol... | | Eukaryotic ribosomal protei... |
| Row28 | BGLU42_AT5g36990_MLF18.1 | Arabidopsis thaliana (Mouse-ear cress) | cytosol [GO:0005829]; beta-gluc... | cellulose catabolic process [G... | beta-gluc... | cytosol... | | Glycosyl hydrolase 1 family |
| Row29 | At5g25754_At5g25757 | Arabidopsis thaliana (Mouse-ear cress) | cytosol [GO:0005829]; eukary... | formation of cytoplasmic tran... | mRNA bin... | cytosol... | | ElF-3 subunit L family |
| Row30 | NUM1_NMP22_NUMA | Homo sapiens (Human) | cell cortex [GO:0009388]; cell... | anastral spindle assembly [G... | actin fil... | cell conte... | | |
| Row31 | UBE2D3_UBC52_UBCH5C | Homo sapiens (Human) | cytosol [GO:0005829]; endopl... | disorderly process [GO:00069... | ATP bindi... | cytosol... | | ElF-3 subunit L family |
| Row32 | | Homo sapiens (Human) | cytosol [GO:0005829]; nucleu... | apoptotic process [GO:00069... | cytosol... | cytosol... | | |
| Row33 | EEF2_EF2 | Homo sapiens (Human) | nucleus [GO:0005634]; aggreg... | cell nucleus [GO:0005634]; cyt... | actin fil... | cytosol... | | Ubiquitin-conjugating enzym... |
| Row34 | UAP1_SPAG2 | Homo sapiens (Human) | cytosol [GO:0005829]; nucleo... | aggregation [GO:0007568]; cellular r... | actin fil... | cytosol... | | TRAFAC class translation fa... |
| Row35 | PSMC4_MP224_TB7P | Homo sapiens (Human) | cytosol [GO:0005829]; cytosol... | UDPG type 1 family | | | | |
| Row36 | MARCK1_CTA1_EMK2 | Homo sapiens (Human) | cytoplasm [GO:0005829]; cytosol... | anaphase-promoting complex...; AT... | ATP bindi... | cytosol... | | AAA ATPase family |
| Row37 | SRM2_KIAA0324_SRL300_SR... | Homo sapiens (Human) | Cajal body [GO:0015030]; cat... | intracellular signal transducti... | cytoplas... | mRNA splicing, via spliceosom... | | Protein kinase superfamily... |
| Row38 | A1A7829_E0K30_1.2 | Arabidopsis thaliana (Mouse-ear cress) | cell nucleus [GO:0040461]; cell... | enzymes involved in carbon metab... | carboxylat... | carboxylat... | | CWC21 family |

7. get

[BLAST](#) [Align](#) [Retrieve/ID mapping](#) [Peptide search](#)
[Help](#) [Contact](#)
File Reader

MQ F

(e)



proteinGroups.txt file
from __inputs__ folder

UniProtKB column names for programmatic access

Last modified August 8, 2019

This document lists the column names for programmatic (RESTful) access to tab-separated or Excel downloads of UniProtKB search results.

See also:

[How can I access resources on this website programmatically?](#)

[Customize display options](#)

[Customise and share your search results](#)

Names & Taxonomy

| Column names as displayed on website | Column names as displayed in URL |
|--------------------------------------|----------------------------------|
| Entry | id |
| Entry name | entry.name |
| Gene names | genes |
| Gene names (primary) | genes(PREFERRED) |
| Gene names (synonym) | genes(ALTERNATIVE) |
| Gene names (ordered locus) | genes(OLN) |
| Gene names (ORF) | genes(ORF) |
| Organism | organism |
| Organism ID | organism.id |
| Protein names | protein.names |
| Proteomes | proteome |
| Taxonomic lineage | lineage(ALL) |
| Virus hosts | virus.hosts |

Sequences

| Column names as displayed on website | Column names as displayed in URL |
|--------------------------------------|------------------------------------------|
| Fragment | fragment |
| Gene encoded by | encodedon |
| Alternative products | comment(ALTERNATIVE PRODUCTS) |
| Erroneous gene model prediction | comment(ERRONEOUS GENE MODEL PREDICTION) |
| Erroneous initiation | comment(ERRONEOUS INITIATION) |
| Erroneous termination | comment(ERRONEOUS TERMINATION) |

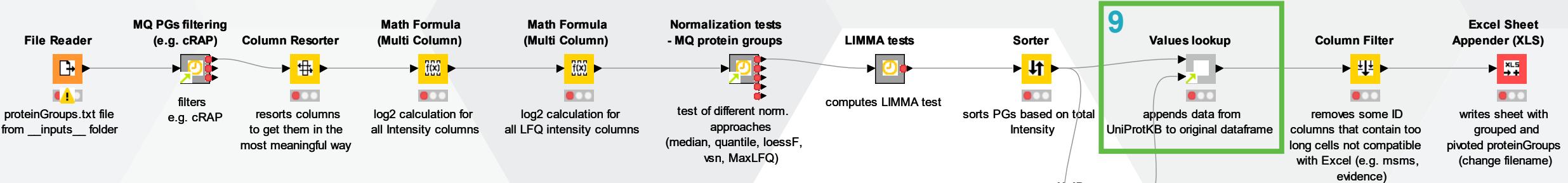
Excel Sheet Appender (XLS)


writes sheet with grouped and pivoted proteinGroups (change filename)

d from UniProt

| Gene ontology IDs | Subcellular location (CC) | Protein families |
|-------------------|------------------------------|---------------------------------|
| b-0009451 | PPR family; PCMP-E subfamily | Methyltransferase superfamily |
| b-0009372... | Subcellular location | Peptidase T1A family |
| b-000050... | SUBCELL... | |
| b-000436... | SUBCELL... | |
| b-000563... | SUBCELL... | |
| b-000372... | SUBCELL... | DEAD box helicase family, ... |
| b-000170... | SUBCELL... | Class-I like SAM-binding me... |
| b-000161... | SUBCELL... | Dicarboxylate/amino acid:c... |
| b-000382... | SUBCELL... | NAD(P)-dependent epimer... |
| b-000552... | SUBCELL... | AAA ATPase family |
| b-000372... | SUBCELL... | SLC3A transporter family |
| b-000417... | SUBCELL... | Peptidase S14 family |
| b-000573... | SUBCELL... | EIF4E-binding protein family |
| b-000392... | SUBCELL... | TRAFAC class dynamin-like .. |
| b-000012... | SUBCELL... | TRAFAC class dynamin-like .. |
| b-000578... | SUBCELL... | Esterase D family |
| b-000369... | SUBCELL... | |
| b-000072... | SUBCELL... | |
| b-000479... | SUBCELL... | Class-II pyridine nucleotide... |
| b-000398... | SUBCELL... | |
| b-000416... | SUBCELL... | PPP/GGPP synthase family |
| b-000552... | SUBCELL... | Mg-chelatase subunits D/F |
| b-000219... | SUBCELL... | TCP-1 chaperonin family |
| b-000002... | SUBCELL... | MAP65/ASE1 family |
| b-000371... | SUBCELL... | Protein kinase C protein alp... |
| b-000013... | SUBCELL... | SCAMP family |
| b-000218... | SUBCELL... | Eukaryotic ribosomal protei... |
| b-000582... | SUBCELL... | Glycosyl hydrolase 1 family |
| b-000173... | SUBCELL... | ElF-3 subunit L family |
| b-000003... | SUBCELL... | |
| b-000012... | SUBCELL... | Ubiquitin-conjugating enzy... |
| b-000534... | SUBCELL... | |
| b-000203... | SUBCELL... | TRAFAC class translation fa... |
| b-000397... | SUBCELL... | UDPGP type 1 family |
| b-000016... | SUBCELL... | AAA ATPase family |
| b-000016... | SUBCELL... | Protein kinase superfamily... |
| b-000039... | SUBCELL... | CWC21 family |
| b-000561... | SUBCELL... | |

9. adds annotation info into the original table



This screenshot shows the configuration of the 'Values lookup' metanode. It includes fields for 'Top Input table IDs column' (Majority protein IDs), 'Top Input table IDs delimiter' (e.g., comma), 'Bottom Input table ID column' (Protein ID), and 'Columns to add from bottom input table' (using Manual Selection). The 'Include' section lists various protein annotations like GO, Enzyme, and Subcellular location. The 'Added columns prefix' is set to 'UniProt_'. The 'Looked up values delimiter' is a colon (:). There are also checkboxes for 'Remove empty strings?', 'Remove duplicates?', and 'Remove not used columns?'.

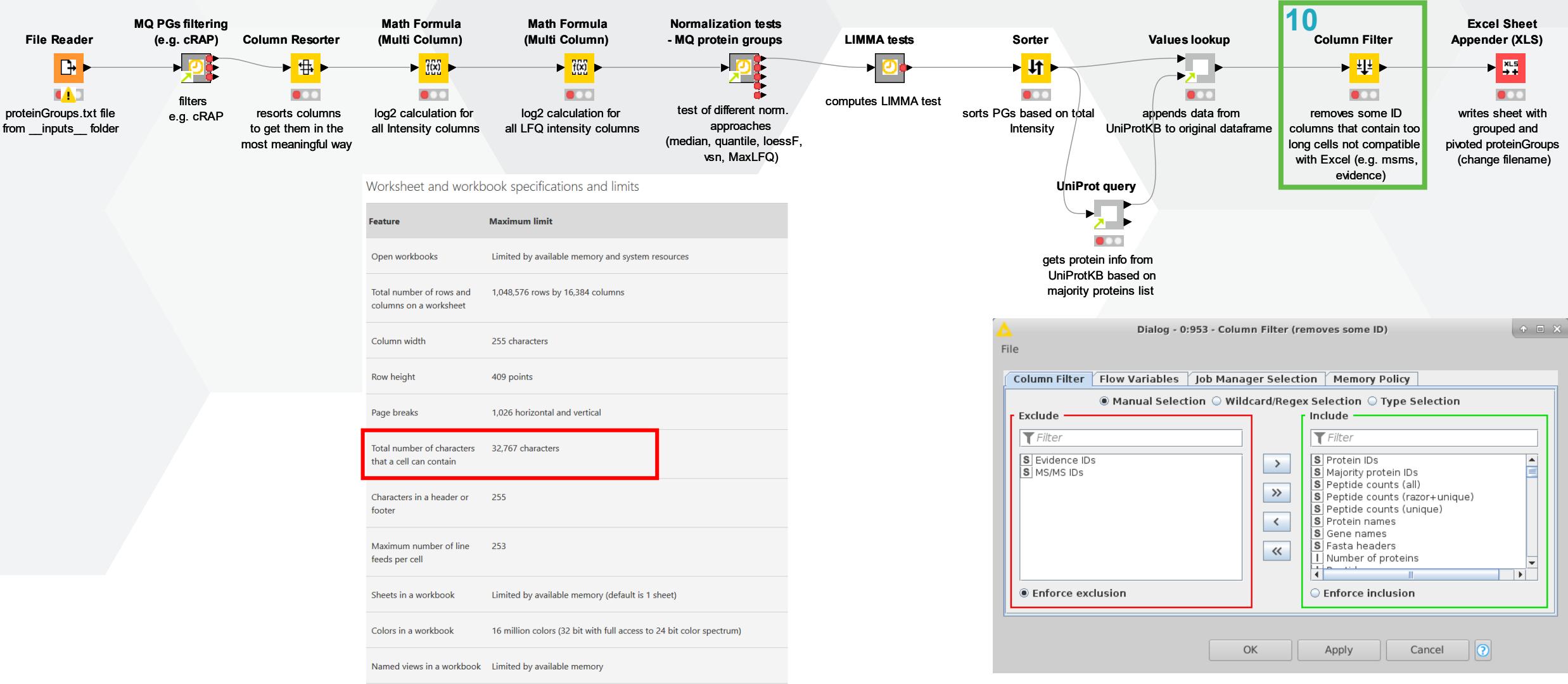
Values lookup metanode utilization

- adds info into the main table from e.g. annotation table
- merging of two tables
- one can alternatively use also KNIME in-built Joiner nodes, potential issue with e.g. multiple protein accessions in one cell (Majority protein IDs column, values separated by colon)

This screenshot shows a table of UniProt query results. The columns include 'Majority protein IDs', 'UniProt_Protein_families', 'UniProt_Sequence', 'UniProt_Cross-references (InterPro)', and 'UniProt_Cross-references (InAct)'. The data consists of many rows of protein accession numbers and their associated families and sequences. A green arrow points from the 'Values lookup' node in the workflow to this table, indicating the flow of data.

| Majority protein IDs | UniProt_Protein_families | UniProt_Sequence | UniProt_Cross-references (InterPro) | UniProt_Cross-references (InAct) |
|------------------------------|--------------------------|---------------------|-------------------------------------|----------------------------------|
| A0A096LP16 | XEV/TSSLPELVKEV/LEF... | IPR029062; IP... | PF04003; A0A096LP16; | |
| A0A06YVA00;P09Y3B3 | MPPRFGSA0TRWAAVAGR... | IPR009038; IP... | PF01105; & ... & 09Y3B3; | |
| A0A06YVH1L;P59780 | MSRIVHDGALRNKA/QSV... | IPR022775; IP... | PF01217; P... | |
| A0A06YVLYL;P18621 | MVRYSLDPEPNTKSKSR... | IPR001063; IP... | PF00237; & ... A0A06YV6... | |
| A0A084LV8;P9N055 | MGOSGRSRHOKRAARAOA... | IPR007109; IP... | PF00001; P... | |
| A0A084J269;P13506 | P002453; IP... | PF00001; P... | PF00001; P... | |
| A0A084J2E5;P15269 | MKFAVYRFNSNLGTVYRRG... | IPR020472; IP... | PF04003; P... | |
| A0A09YWW9;P0D0P08;P0D0P06... | MMLKWFVFLVVAAPRWL... | IPR007110; IP... | PF07686; & ... & & & ... | |
| A0A01J1QV1 | MVLFLQEMDKPCGCPGV... | | | |
| A0A1BOGU03;P07339 | MOPPSPLSLPALLCAAAPS... | IPR001461; IP... | PF07966; P... | |
| A0A1W2PN4 | MSLHQLLEPITPCCHAWN... | IPR030140; IP... | PF00400; & 07339; | |
| A0A1W2PQ55 | MLMPKKNRNIAVYELFKEGV... | IPR015797; IP... | PF00293; P... | |
| A0A1W2PB88;P9H0W9 | MOPPSPLSLPALLCAAAPS... | IPR0015021; IP... | PF08925; P... | |
| A0A1V1 | MEGSEPAVVAH;P0GE6AS... | IPR032420; IP... | PF16191; P... | |
| A0A1V1 | Q9BSJ5; A... | IPR006776; IP... | 0A0VT1; | |
| A0A1V1 | MTANRDAALSLSHRHPG... | IPR000008; IP... | PF0068; P... | |
| A0A1V1 | MTPEAAAPAGS1FPFSLL... | IPR029375; IP... | PF02775; A1LOTO; | |
| A0A1V1 | MVFPGDLLAALESQIS... | IPR027417; IP... | PF3729; P... | |
| A0A1V1 | TRAFAC class OBG... | IPR008379; Q5P59... | 0A3K9N3; | |
| A0A1V1 | MUHVSCAVFRKGNRIF... | IPR031167; IP... | PF01018; P... | |
| A0A1V1 | CNOT1; P... | IPR007302; Q9UCZ... | PF0001; P... | |
| A0A1V1 | MILDSLLSLSLQ51VDN... | IPR0071296; IP... | PF00154; P... | |
| A0A1V1 | MMAEEAAGDODAR;P0... | IPR03642; IP... | PF03344; A6NF02; | |
| A0A1V1 | MATIAAEAAEALMDQ;T... | IPR03642; IP... | PF03344; A6NF02; & | |
| A0A1V1 | MKSAASSLRGG;P0... | IPR000642; IP... | PF01269; A6NHQ2; | |
| A0A1V1 | MAAADGGPGEGAS;GTE... | IPR036890; IP... | PF06470; A6NHQ2; | |
| A0A1V1 | MRLRVYFCLLVAAPRWL... | IPR002903; IP... | PF01795; A6NHQ2; | |
| A0A1V1 | MCSCKAYGERYVAVQGS... | IPR032023; IP... | PF01465; P... | |
| A0A1V1 | MSSLAARRGGRG;P... | IPR026054; IP... | PF01465; P... | |
| A0A1V1 | Mudic hydrolyase fa... | IPR015797; IP... | ABC534; & Q... | |
| A0A1V1 | MIF4GD family | IPR053673; IP... | PF016024; P... | |
| A0A1V1 | TRAFAC class myo... | IPR036961; IP... | PF02854; A5UHW6; | |
| A0A1V1 | Mitochondrial car... | IPR036961; IP... | PF00063; P... | |
| A0A1V1 | G6AL162; P60... | IPR01624; P... | B01LT2; | |
| A0A1V1 | Mitochondrial car... | IPR036961; IP... | PF00063; P... | |
| A0A1V1 | Golgi ph regulator... | IPR020303; IP... | PF0053; P... | |
| A0A1V1 | & | IPR025959; IP... | B01LT2; & Q... | |
| A0A1V1 | MSTDGTGSPVLPSEYEDQGS... | IPR007667; &... | PF04588; & C9JAWS; & Q... | |
| A0A1V1 | MHPIDLRPWTSSLVDAAL... | IPR003595; IP... | PF00664; P... | |
| A0A1V1 | Q12982; Q96... | IPR018796; IP... | PF0231; P... | |
| A0A1V1 | MPLCAAAGRGRGAMV... | IPR018796; IP... | PF07866; P... | |
| A0A1V1 | E9PAV3 | IPR016641; IP... | PF01231; P... | |
| A0A1V1 | E9PL57;Q15843 | IPR016641; IP... | PF01849; P... | |
| A0A1V1 | E9PLD3;E9PRG8 | IPR03642; IP... | PF01849; P... | |
| A0A1V1 | E9PLNB8;Q9B0B6 | IPR03642; IP... | PF01849; P... | |
| A0A1V1 | F5H5P2;P12694 | IPR03642; IP... | PF01849; P... | |
| A0A1V1 | F8W031;Q9Y2B0 | IPR03642; IP... | PF01849; P... | |
| A0A1V1 | HOYAAO | IPR03642; IP... | PF01849; P... | |
| A0A1V1 | HOYC42 | IPR03642; IP... | PF01849; P... | |
| A0A1V1 | H3BN98 | IPR03642; IP... | PF01849; P... | |
| A0A1V1 | I3L0E3;Q9Y2R5 | IPR03642; IP... | PF01849; P... | |
| A0A1V1 | K7EQ62 | IPR03642; IP... | PF01849; P... | |
| A0A1V1 | K7ERE1;Q8N2U0 | IPR03642; IP... | PF01849; P... | |
| A0A1V1 | L7N2F9;Q15836;P63027 | IPR03642; IP... | PF01849; P... | |
| A0A1V1 | MOQY70 | IPR03642; IP... | PF01849; P... | |
| A0A1V1 | MQQZD8 | IPR03642; IP... | PF01849; P... | |

10. removes not compatible columns

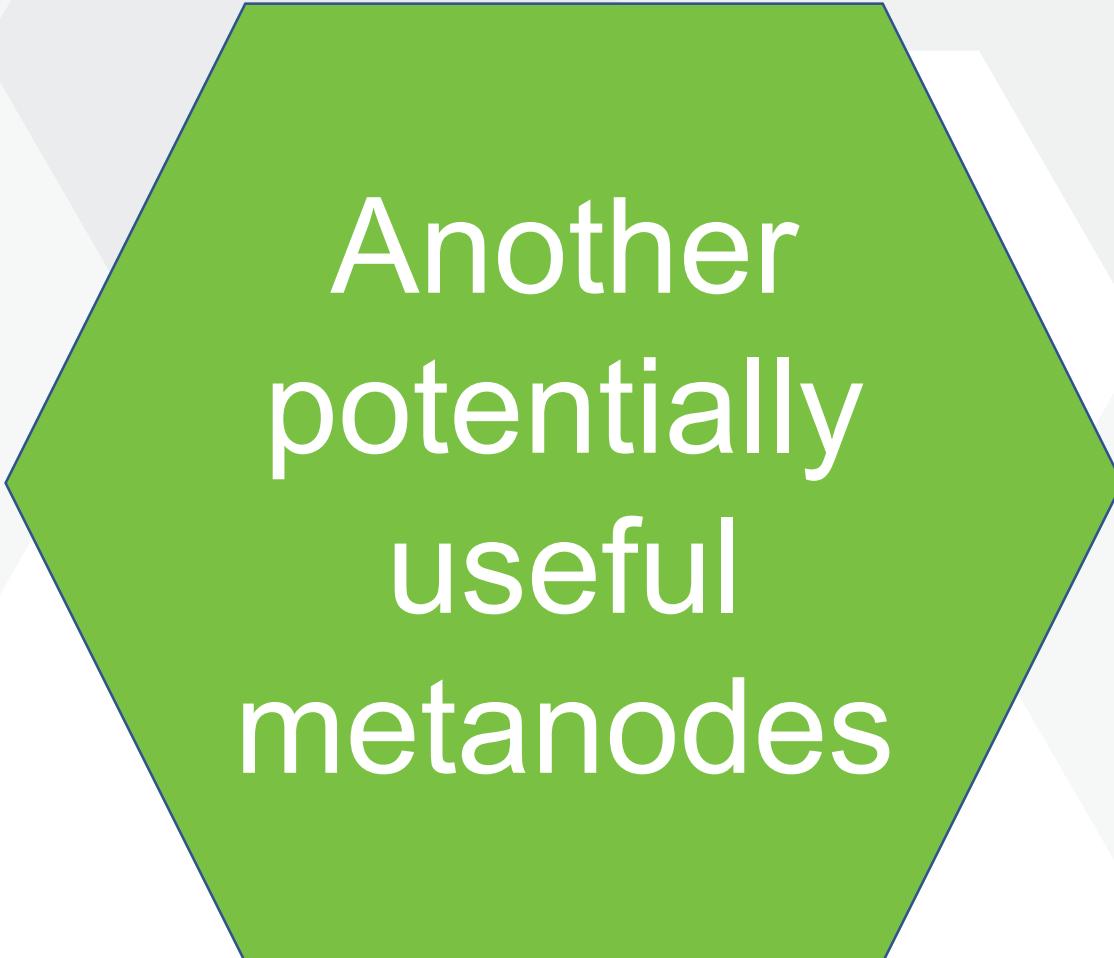


11. writes the table as the initial report



How the final report looks like? Excel file...

| Header | | Proteins | | Identification details in Original Repo | | Quantitative Information | | Statistical evaluation | | Ratios Categorical Evaluation - Stat.ratio | | Protein Info from UniProtK | | | | | | | | | | | | | | | | | | | |
|--------|---------|-----------------------------------------------|-----------------------------------------------------------------------------------------|-----------------------------------------|--|--------------------------|----|------------------------|--|--------------------------------------------|--|----------------------------|-------|--------------------|-------|-------------------|-------|--------------------|-------|--------------------|-------|-------------|--------|-------------|-------|------------|-----|------------|------------------|------------|---|
| PGs | | proteins | | Protein ID | | Rep prots | | Qual Data | | PG Score, Q | | Peptides | | Sequence coverage | | Data | | Statistical values | | KO-WT | | Ratio Cat.E | | thresholds | | KO-WT | | Annotation | | | |
| total | | 3522 | | 3751 | | thresholds | | #replicates | | Details | | Data | | Statistical values | | KO-WT | | Ratio Cat.E | | thresholds | | KO-WT | | Annotation | | | | | | | |
| shown | | 132 | | 136 | | Protein ID | | Rep prots | | Qual Data | | PG Score, Q | | Peptides | | Sequence coverage | | Data | | Statistical values | | KO-WT | | Ratio Cat.E | | thresholds | | KO-WT | | Annotation | |
| PG I | Access. | Description | | | | Summa | Y1 | KO | | | | WT | WT | WT | KO | KO | KO | KO r | WT r | logFC | FC | P.Value | adj.P. | Summary | E | T | P.x | adj.P. | Protein families | Y | |
| 679 | 348 | P09326 | CD46 antigen OS=Homo sapiens GN=CD48 PE=1 SV=2 | | | YYYY YYYY | 3 | 3 | | | | 24.33 | 24.86 | 24.70 | 27.08 | 26.34 | 27.37 | 3 | 3 | 2.28 | 4.85 | 0.000 | 0.001 | U Y Y | U | Y | Y | Y | N | | |
| 699 | 3045 | Q86U42 | Polyadenylate-binding protein 2 OS=Homo sapiens GN= | | | YYYY YYYY | 3 | 3 | | | | 22.32 | - | - | 26.81 | 26.88 | 26.92 | 3 | 2 | 3.78 | 13.74 | 0.000 | 0.003 | U Y Y | U | Y | Y | Y | N | | |
| 721 | 1708 | P43720 | Proteasome subunit beta type-3 OS=Homo sapiens GN= | | | YYYY YYYY | 3 | 3 | | | | 25.08 | 24.63 | 24.35 | 27.13 | 26.86 | 27.03 | 3 | 1 | 4.60 | 24.23 | 0.000 | 0.003 | U Y Y | U | Y | Y | Y | Y | | |
| 748 | 4013 | Q9BZE4 | Nucleolar GTP-binding protein 1 OS=Homo sapiens GN= | | | YYYY YYYY | 3 | 3 | | | | 25.33 | 25.01 | 25.64 | 26.92 | 26.65 | 26.34 | 3 | 3 | 2.10 | 4.28 | 0.000 | 0.001 | U Y Y | U | Y | Y | Y | Y | | |
| 754 | 1311 | P26062 | Proteasome subunit beta type-6 OS=Homo sapiens GN= | | | YYYY YYYY | 3 | 3 | | | | - | - | - | 20.93 | 21.43 | 26.66 | 25.05 | 27.23 | 3 | 2 | 5.62 | 43.28 | 0.000 | 0.001 | U Y Y | U | Y | Y | Y | Y |
| 759 | 2597 | Transferrin OS=Homo sapiens GN=TFNS PE=1 SV=1 | | | | YYYY YYYY | 3 | 3 | | | | 23.41 | 23.56 | 23.67 | 26.40 | 26.58 | 26.63 | 3 | 3 | 3.01 | 8.03 | 0.000 | 0.000 | U Y Y | U | Y | Y | Y | Y | | |
| 817 | 1871 | P55145 | Mesencephalic astrocyte-derived neurotrophic factor C OS=Homo sapiens GN=MNDN PE=1 SV=1 | | | YYYY YYYY | 3 | 3 | | | | 21.62 | - | - | 22.68 | 26.44 | 26.76 | 3 | 2 | 4.39 | 21.00 | 0.000 | 0.006 | U Y Y | U | Y | Y | Y | Y | | |
| 819 | 1475 | P36871 | Phosphoglucomutase-1 OS=Homo sapiens GN=PGM1P | | | YYYY YYYY | 3 | 3 | | | | - | 21.70 | 22.48 | 26.45 | 26.61 | 26.47 | 26.47 | 3 | 2 | 4.42 | 21.41 | 0.000 | 0.002 | U Y Y | U | Y | Y | Y | Y | |
| 847 | 598 | P75400 | Pre-mRNA-processing factor 10 homolog A OS=Homo sapiens | | | YYYY YYYY | 3 | 3 | | | | - | - | - | - | - | - | 2 | 0 | NsN | NsN | NsN | U I - | U I - | - | - | - | N | | | |
| 819 | 2193 | P98111 | Rho GTPase-activating protein 4 OS=Homo sapiens GN= | | | NNNN NYYY | 0 | 2 | | | | - | - | - | - | - | - | 22.24 | 21.32 | - | - | - | - | U Y Y | U | Y | Y | Y | N | | |
| 934 | 3827 | Q93442 | Translocation protein SEC62 OS=Homo sapiens GN=S1 | | | YYYY YYYY | 3 | 3 | | | | 24.46 | 23.56 | 24.50 | 26.08 | 26.22 | 26.06 | 3 | 3 | 1.35 | 3.85 | 0.001 | 0.017 | U Y Y | U | Y | Y | Y | Y | | |
| 939 | 3877 | Q9BPX5 | Actin-related protein 2/3 complex subunit 5-like protein | | | YYYY YYYY | 3 | 3 | | | | - | - | - | 27.01 | 25.70 | 25.43 | 3 | 0 | NsN | NsN | NsN | U I - | U I - | - | - | - | Y | | | |
| 985 | 2234 | Q86257 | Quinone oxidoreductase OS=Homo sapiens GN=CRY2 | | | YYYY YYYY | 3 | 3 | | | | 24.34 | 25.01 | 24.74 | 26.15 | 26.03 | 25.37 | 3 | 3 | 1.15 | 2.23 | 0.000 | 0.006 | U Y Y | U | Y | Y | Y | Y | | |
| 1022 | 2031 | P63165 | Small ubiquitin-related modifier 1 OS=Homo sapiens GN= | | | YYYY YYYY | 3 | 3 | | | | 24.67 | 24.32 | 24.60 | 26.12 | 26.05 | 26.22 | 3 | 3 | 1.40 | 2.63 | 0.000 | 0.003 | U Y Y | U | Y | Y | Y | Y | | |
| 1026 | 5319 | Q9Y636 | Chloride intracellular channel protein 4 OS=Homo sapiens | | | YYYY YYYY | 3 | 3 | | | | 23.26 | 23.90 | 24.05 | 26.11 | 25.76 | 26.24 | 3 | 3 | 2.30 | 4.92 | 0.000 | 0.007 | U Y Y | U | Y | Y | Y | Y | | |
| 1065 | 2533 | Q9C622 | Condensin complex subunit 2 OS=Homo sapiens GN=SEC62 | | | YYYY YYYY | 3 | 3 | | | | 23.32 | 22.31 | 23.18 | 25.71 | 25.87 | 25.86 | 3 | 3 | 2.45 | 5.48 | 0.000 | 0.006 | U Y Y | U | Y | Y | Y | Y | | |
| 1088 | 1076 | P14735 | Insulin-degrading enzyme OS=Homo sapiens GN=IDE PE=1 SV=1 | | | YYYY YYYY | 3 | 3 | | | | 23.64 | 22.78 | 23.54 | 25.54 | 24.59 | 26.04 | 3 | 3 | 2.37 | 5.16 | 0.000 | 0.010 | U Y Y | U | Y | Y | Y | Y | | |
| 1092 | 1111 | P16402 | Histone H1.3 OS=Homo sapiens GN=H1TH1HD PE=1 SV=1 | | | YYYY YYYY | 3 | 3 | | | | - | - | - | 25.70 | 25.61 | 25.76 | 3 | 0 | NsN | NsN | NsN | U I - | U I - | - | - | - | Y | | | |
| 1111 | 5323 | Q9Y6A9 | Signal peptidase complex subunit 1 OS=Homo sapiens GN= | | | YYYY YYYY | 3 | 3 | | | | 24.21 | 24.52 | 24.55 | 25.81 | 25.83 | 25.89 | 3 | 3 | 1.42 | 2.67 | 0.000 | 0.003 | U Y Y | U | Y | Y | Y | Y | | |
| 1120 | 2325 | Q9T675 | Secernin-1 OS=Homo sapiens GN=SCRN1 PE=1 SV=2 | | | YYYY YYYY | 3 | 3 | | | | 23.17 | - | - | 25.53 | 24.56 | 25.69 | 3 | 1 | 2.39 | 5.24 | 0.001 | 0.015 | U Y Y | U | Y | Y | Y | Y | | |
| 1158 | 653 | P07964 | ATP synthase subunit g, mitochondrial OS=Homo sapiens | | | YYYY YYYY | 3 | 3 | | | | - | - | - | 25.36 | 25.34 | 25.41 | 3 | 0 | NsN | NsN | NsN | U I - | U I - | - | - | - | Y | | | |
| 1200 | 3292 | Q8NI36 | WD repeat-containing protein 36 OS=Homo sapiens GN= | | | YYYY YYYY | 3 | 3 | | | | 18.32 | - | - | 25.02 | 25.05 | 25.50 | 3 | 1 | 6.28 | 77.53 | 0.000 | 0.004 | U Y Y | U | Y | Y | Y | Y | | |
| 1205 | 632 | Q75717 | WD repeat and HMG-box DNA-binding protein 1 OS=H | | | YYYY YYYY | 3 | 3 | | | | 20.89 | 20.53 | 20.51 | 24.37 | 24.50 | 25.52 | 3 | 3 | 4.35 | 20.43 | 0.000 | 0.001 | U Y Y | U | Y | Y | Y | Y | | |
| 1268 | 3345 | Q8DTN6 | Ribosome biogenesis protein BRX1 homolog OS=Hom | | | YYYY YYYY | 3 | 3 | | | | - | - | - | 25.26 | 25.43 | 24.21 | 3 | 0 | NsN | NsN | NsN | U I - | U I - | - | - | - | Y | | | |
| 1291 | 714 | P04804 | Serine/threonine-protein kinase 10 OS=Homo sapiens G | | | YYYY YYYY | 3 | 3 | | | | 22.31 | 21.78 | 21.90 | 25.36 | 25.19 | 25.73 | 3 | 3 | 3.25 | 5.52 | 0.000 | 0.006 | U Y Y | U | Y | Y | Y | Y | | |
| 1319 | 2328 | Q12788 | Transducin beta-like protein 3 OS=Homo sapiens GN=ET | | | YYYY YYYY | 3 | 3 | | | | - | - | - | 25.18 | 25.11 | 25.41 | 3 | 0 | NsN | NsN | NsN | U I - | U I - | - | - | - | Y | | | |
| 1244 | 233 | P00764 | Pyridoxal kinase OS=Homo sapiens GN=PDXK PE=1 SV=2 | | | YYYY YYYY | 3 | 3 | | | | 22.11 | 23.33 | 22.46 | 25.14 | 25.21 | 24.63 | 3 | 3 | 2.36 | 5.13 | 0.001 | 0.020 | U Y Y | U | Y | Y | Y | Y | | |
| 1292 | 3915 | Q9BT23 | LIM domain-containing protein 2 OS=Homo sapiens GN= | | | YYYY YYYY | 3 | 3 | | | | - | 13.68 | 13.63 | 24.33 | 24.86 | 24.85 | 24.85 | 3 | 2 | 5.03 | 34.17 | 0.000 | 0.000 | U Y Y | U | Y | Y | Y | Y | |
| 1418 | 3640 | Q96U07 | Proteasome assembly chaperone 2 OS=Homo sapiens C | | | YYYY YYYY | 3 | 3 | | | | - | - | - | 24.78 | 24.38 | 24.32 | 3 | 0 | NsN | NsN | NsN | U I - | U I - | - | - | - | Y | | | |
| 1437 | 2160 | P84085 | ADP-ribosylation factor 5 OS=Homo sapiens GN=ARF5 | | | YYYY YYYY | 3 | 3 | | | | 21.44 | 22.08 | 21.70 | 25.20 | 23.81 | 25.27 | 3 | 3 | 3.02 | 8.10 | 0.001 | 0.013 | U Y Y | U | Y | Y | Y | Y | | |
| 1450 | 1 | DAA067Y | Protein TMED7-1 OS=Homo sapiens GN=TMED7 | | | YYYY YYYY | 2 | 3 | | | | - | - | - | 25.01 | 24.64 | 24.58 | 3 | 0 | NsN | NsN | NsN | U I - | U I - | - | - | - | Y | | | |
| 1466 | 4666 | Q9WV53 | Pyridoxine-5'-phosphate oxidase OS=Homo sapiens GN= | | | YYYY YYYY | 3 | 3 | | | | - | - | - | 24.36 | 25.00 | 24.64 | 3 | 0 | NsN | NsN | NsN | U I - | U I - | - | - | - | Y | | | |
| 1472 | 3771 | Q9K6R1 | Zinc finger RNA-binding protein OS=Homo sapiens GN= | | | YYYY YYYY | 3 | 3 | | | | - | - | - | 24.24 | 24.75 | 24.80 | 3 | 0 | NsN | NsN | NsN | U I - | U I - | - | - | - | Y | | | |
| 1509 | 2828 | Q5TJY8 | Torsin-1A-interacting protein 1 OS=Homo sapiens G | | | YYYY YYYY | 3 | 3 | | | | 23.02 | 22.96 | - | 24.58 | 24.39 | 24.49 | 3 | 2 | 1.63 | 3.10 | 0.002 | 0.036 | U Y Y | U | Y | Y | Y | Y | | |
| 1597 | 1257 | P25490 | Transcriptional repressor protein YY1 OS=Homo sapiens | | | YYYY YYYY | 3 | 3 | | | | 20.77 | - | 20.78 | 24.39 | 24.47 | 24.49 | 3 | 2 | 3.68 | 12.78 | 0.000 | 0.000 | U Y Y | U | Y | Y | Y | Y | | |
| 1614 | 642 | Q5T844 | CAAX prenyl protease 1 homolog OS=Homo sapiens G | | | YYYY YYYY | 3 | 3 | | | | 23.72 | 22.93 | 22.77 | 24.83 | 24.57 | 24.57 | 3 | 3 | 1.43 | 2.82 | 0.002 | 0.041 | U Y Y | U | Y | Y | Y | Y | | |
| 1640 | 3712 | P062310 | U6 snRNA-associated Sm-like protein LSm3 OS=Homo | | | YYYY YYYY | 2 | 3 | | | | 22.47 | 22.68 | 23.43 | 24.84 | 24.23 | 24.74 | 3 | 3 | 1.72 | 3.23 | 0.002 | 0.044 | U Y Y | U | Y | Y | Y | Y | | |
| 1660 | 2044 | P2815 | Sigma intracellular receptor 2 OS=Homo sapiens GN=TI | | | YYYY YYYY | 2 | 3 | | | | | | | | | | | | | | | | | | | | | | | |



Another
potentially
useful
metanodes

Missing values imputation step

- Missing values in LC-MS/MS experiments are frequent (50-70 %)
- Problem for differential expression analysis
- Several approaches, often used global minimum, kNN, MinProb...
- Imputation done on peptide/protein level?
- Implemented algorithms
 - global minimum
 - imp4p (R package)
 - proDA

How does protein A look like in our data table?

| Row ID | Protein names | Gene names | Fasta headers | Intensity WT_1_log2_norm | Intensity WT_2_log2_norm | Intensity WT_3_log2_norm | Intensity KO_1_log2_norm | Intensity KO_2_log2_norm | Intensity KO_3_log2_norm |
|--------|-----------------------------------|-------------|---------------------------------------------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Row6 | Periodic t... | PWP2 | Uncharacterized protein OS=Homo sapiens GN=LOC102724159 P... | 24.853 | 23.993 | 25.746 | 25.156 | 24.823 | 25.147 |
| Row7 | Ig heavy ... | IGHV4-61... | Uncharacterized protein (Fragment) OS=Homo sapiens PE=1 SV=1 | 24.91 | 25.216 | 24.202 | 24.859 | 25.393 | 24.894 |
| Row9 | Cathepsin... | CTSD | Uncharacterized protein OS=Homo sapiens PE=1 SV=1;Cathepsin... | 22.975 | 23.137 | 24.321 | 23.228 | 22.998 | 22.798 |
| Row10 | NA | NA | Uncharacterized protein OS=Homo sapiens PE=4 SV=... | ? | ? | ? | ? | 21.47 | 21.257 |
| Row12 | Ester hyd... | C11orf54 | Uncharacterized protein OS=Homo sapiens PE=4 SV=1;Ester hyd... | 22.163 | 22.37 | 22.056 | 22.57 | 21.549 | 21.798 |
| Row13 | Ubiquitin... Acetyltransferase | UVR8 | Ubiquitin-modifier-activating enzyme 6 OS=Homo sapiens GN=... | 25.155 | 25.264 | 25.486 | 25.264 | 25.785 | 25.474 |
| Row15 | NA | NA | Acetyl-CoA acyltransferase 1 OS=Homo sapiens GN=UVR8 P... | 21.699 | ? | ? | 23.402 | 23.072 | 23.906 |
| Row17 | Protein A | GTPBP10 | Protein A OS=Homo sapiens SN=GTPBP10 PE=1 SV=1 | ? | 22.294 | 22.698 | ? | ? | ? |
| Row18 | CCR4-NOT... | CNOT1 | CCR4-NOT transcription complex subunit 1 OS=Homo sapiens GN=... | 25.129 | 25.064 | 24.704 | 24.617 | 24.366 | 24.186 |
| Row19 | Phospho... PGP | PGP | Glyceral-3-phosphate phosphatase OS=Homo sapiens GN=PGP P... | 25.593 | 25.677 | 25.894 | 25.938 | 25.979 | 25.637 |
| Row20 | TRPM9 ch... | TCAF2 | TRPM9-channel-associated factor 2 OS=Homo sapiens GN=TCAF2... | 19.119 | ? | ? | 19.369 | 20.047 | 19.483 |
| Row21 | rRNA-RN... | FBL11 | rRNA/rRNA 2'-O-methyltransferase fibrillarin-like protein 1 OS=... | ? | 25.075 | 25.415 | 25.005 | ? | 25.013 |
| Row22 | Structura... | SMCHD1 | Structural maintenance of chromosomes flexible hinge domain-co... | 25.201 | 23.944 | 25.116 | 21.908 | 23.824 | 22.899 |
| Row23 | Probable e... | METTL15 | Probable methyltransferase-like protein 15 OS=Homo sapiens GN=... | 25.913 | 25.648 | 25.682 | 20.375 | ? | 20.449 |
| Row24 | RanBP2-1... RGPD3;Ran... | RanBP2 | RanBP2-like and GRIP domain-containing protein 3 OS=Homo sapi... | 22.053 | ? | ? | 22.421 | 22.393 | 22.006 |
| Row25 | Nuclear e... | POM121C | Nuclear envelope pore membrane protein POM121 OS=Homo s... | ? | ? | ? | 21.311 | 22.211 | 21.958 |
| Row26 | Nucleos... NUDT19 | NUDT19 | Nucleoside diphosphate-linked moiety X motif 19 OS=Homo sapi... | 22.331 | ? | ? | 21.739 | 21.86 | 22.895 |
| Row27 | MIF4G do... MIF4GD | MIF4G | MIF4G domain-containing protein OS=Homo sapiens GN=MIF4GD... | 25.954 | 25.665 | 25.773 | 20.581 | 20.653 | ? |
| Row28 | Unconverte... | MYOLG | Unconventional myosin-Ig OS=Homo sapiens GN=MYOLG PE=1 SV=... | 28.501 | 28.366 | 28.464 | 28.5 | 28.586 | 28.729 |
| Row29 | Mitochon... | SLC25A10 | cDNA FL60124, highly similar to Mitochondrial dicarboxylate carrier | 27.161 | 27.024 | 26.914 | 26.513 | 26.903 | 26.791 |
| Row32 | NA | NA | Uncharacterized protein (Fragment) OS=Homo sapiens PE=3 SV=1 | 25.294 | 25.14 | 25.584 | 25.292 | 25.438 | 25.448 |
| Row33 | Kinesin ... | KLC1 | Uncharacterized protein OS=Homo sapiens PE=4 SV=1;Kinesin lig... | 23.18 | 23.458 | 24.392 | 24.344 | 24.388 | 24.419 |
| Row34 | Nascent ... | NACA | Nascent polypeptide-associated complex subunit alpha; muscle-e... | 26.584 | 26.81 | 26.841 | 26.348 | 26.545 | 26.52 |
| Row35 | NEDD8... NEDD8-MDP1 | NEDD8 | NEDD8-MDP1 readthrough (Fragment) OS=Homo sapiens GN=NED... | 25.697 | 25.806 | 25.74 | 25.585 | 25.509 | 25.433 |
| Row36 | Uncharact... | C11orf98 | Uncharacterized protein OS=Homo sapiens PE=4 SV=1;Uncharact... | 24.292 | 23.748 | 24.074 | 23.846 | 23.663 | 23.492 |
| Row37 | Vitamin K... VKORC1 | VKORC1 | Uncharacterized protein (Fragment) OS=Homo sapiens PE=4 SV=... | 23.502 | 23.253 | 22.974 | 23.489 | 23.359 | 23.433 |
| Row38 | 2-oxoisov... | BCKDHA | Uncharacterized protein OS=Homo sapiens PE=4 SV=1;2-oxoisov... | 20.691 | 20.704 | 20.1 | 20.754 | 20.72 | 20.413 |
| Row39 | Protein C... | CNPY2 | Uncharacterized protein (Fragment) OS=Homo sapiens PE=1 SV=... | 23.699 | 22.932 | 23.818 | 23.602 | 23.889 | 23.4 |
| Row41 | NA | NA | Uncharacterized protein OS=Homo sapiens PE=4 SV=2 | 26.639 | 26.401 | 26.025 | 25.747 | 26.439 | 26.415 |
| Row42 | NA | NA | Uncharacterized protein (Fragment) OS=Homo sapiens PE=4 SV=... | 20.925 | 22.817 | ? | 22.622 | 21.491 | 24.078 |
| Row43 | 28S ribos... | hCG_198... | hCG1984214, isoform CRA_a OS=Homo sapiens GN=hCG_198421... | 23.026 | 23.152 | 23.814 | 24.138 | 22.847 | 22.726 |
| Row46 | Vesicle-a... | VAMP3;... | Uncharacterized protein (Fragment) OS=Homo sapiens PE=4 SV=... | 22.919 | 22.85 | 22.873 | 22.896 | 22.726 | 22.343 |
| Row47 | NA | NA | Uncharacterized protein (Fragment) OS=Homo sapiens PE=1 SV=1 | 26.691 | 26.511 | 26.652 | 24.68 | 24.677 | 24.6 |
| Row48 | NA | NA | Uncharacterized protein OS=Homo sapiens GN=Or40n469 PE=4 | 26.255 | 25.809 | 25.901 | 26.377 | 26.46 | 26.316 |

Different types of missing values... E.g.:

- left censored
- missing completely at random
- **global minimum vs. imp4p missing values imputation**

proDA – another statistical analysis next to LIMMA

proDA: Probabilistic Dropout Analysis for Identifying Differentially Abundant Proteins in Label-Free Mass Spectrometry

Constantin Ahlmann-Eltze, Simon Anders

doi: <https://doi.org/10.1101/661496>

This article is a preprint and has not been certified by peer review [what does this mean?].

Abstract Full Text Info/History Metrics

Preview PDF

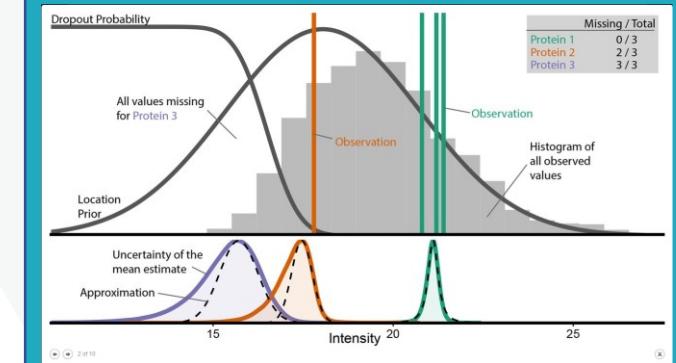
Abstract

Protein mass spectrometry with label-free quantification (LFQ) is widely used for quantitative proteomics studies. Nevertheless, well-principled statistical inference procedures are still lacking, and most practitioners adopt methods from transcriptomics. These, however, cannot properly treat the principal complication of label-free proteomics, namely many non-randomly missing values.

We present *proDA*, a method to perform statistical tests for differential abundance of proteins. It models missing values in an intensity-dependent probabilistic manner. *proDA* is based on linear models and thus suitable for complex experimental designs, and boosts statistical power for small sample sizes by using variance moderation. We show that the currently widely used methods based on ad hoc imputation schemes can report excessive false positives, and that *proDA* not only overcomes this serious issue but also offers high sensitivity. Thus, *proDA* fills a crucial gap in the toolbox of quantitative proteomics.

Availability The *proDA* method is implemented as an open-source R package, available on <https://github.com/const-ae/proDA>.¹

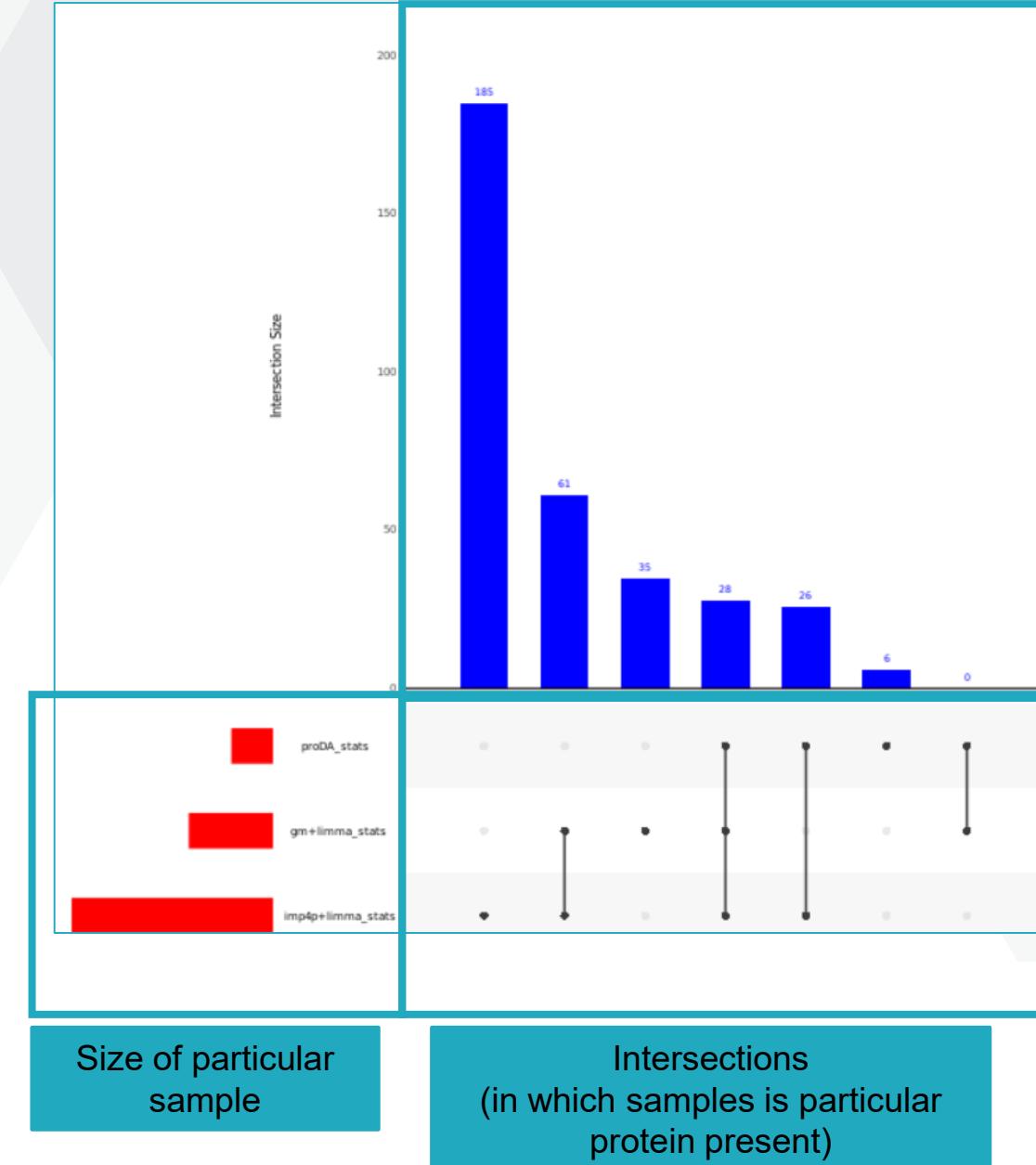
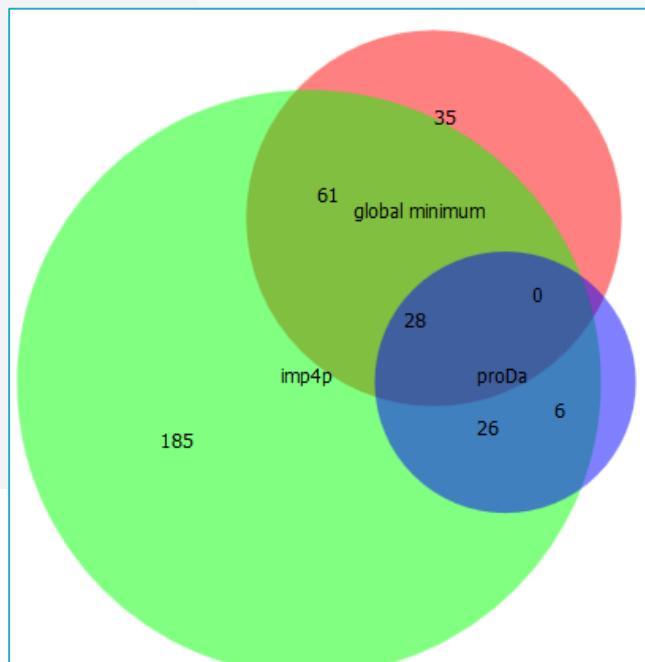
- statistical analysis with missing values still present
- similar results to global minimum imputation + LIMMA strategy
- tends to be more conservative compared to imp4p + LIMMA



(“Almost”) implemented in the form of metanode! Present, fully understood?

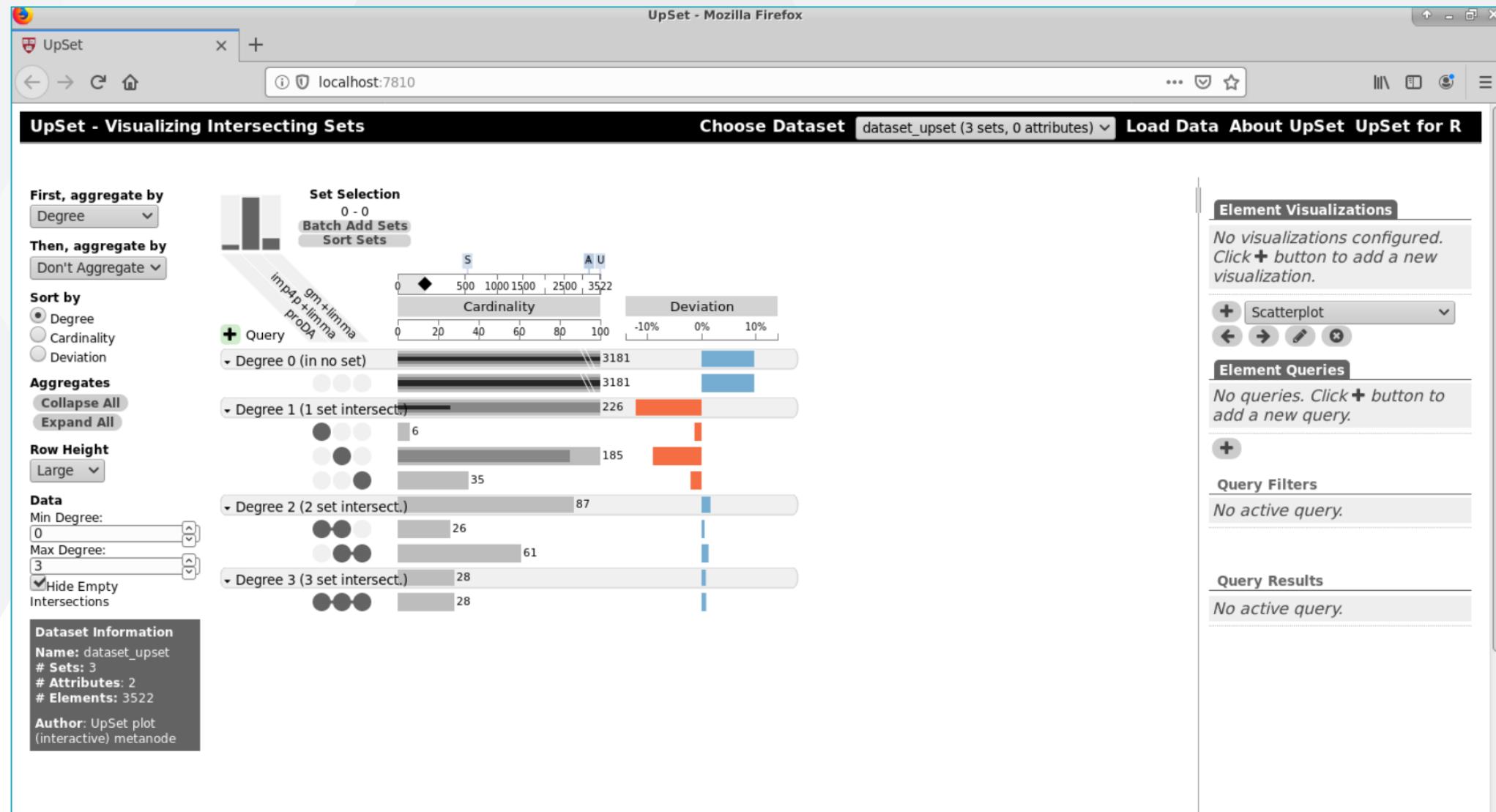
UpSet plot

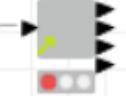
- Alternative to Venn diagram for more samples
- Shows intersections between samples (what do and don't samples have in common)



How many proteins are present in the particular intersection
(e.g. 2745 protein groups are present in all three samples; 852 only in 1-1 sample, etc.)

UpSet plot (interactive)





Reactome query

Dialog - 4:956 - Reactome query

File

Memory Policy **Job Manager Selection**

QuickForms **Flow Variables**

Protein IDs column Change
Majority protein IDs

Protein IDs delimiter Change
;

Include interactors? Change

Resource Change
UNIPROT

OK Apply Cancel ?

Reactome query

Metanode that makes api call to Reactome.org to retrieve the analysis results. So far only overrepresentation analysis is implemented using the protein IDs present in the selected protein IDs column. Protein IDs have to be in supported format (e.g. UniProtKB identified; see <https://reactome.org/dev/analysis> for their full list). Multiple protein IDs separated by delimiter (e.g. semicolon) are supported.

You can select whether to include also interactors during the analysis and what resource should be used during the analysis,

Metanode generates 4 outputs in total

1) pathways table
table of pathways returned by the analysis (similarly to downloading the pathway analysis results in csv format via web interface)

2) not mapped identifiers
identifiers that were not found/mapped during the Reactome analysis

3) summary table
summary information about the analysis including date of running, database version and more

4) complete analysis results
complete analysis results in json format for further data mining if needed

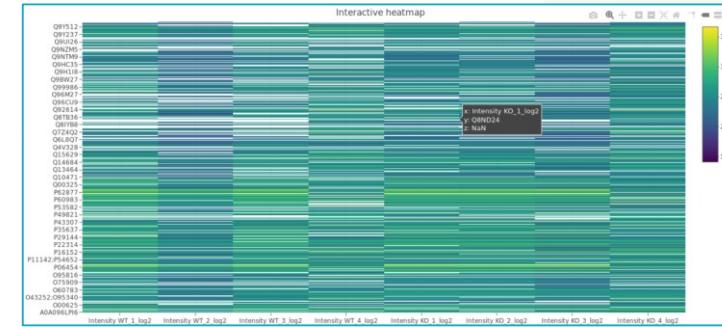
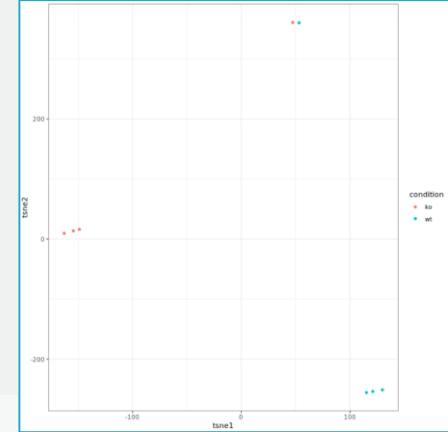
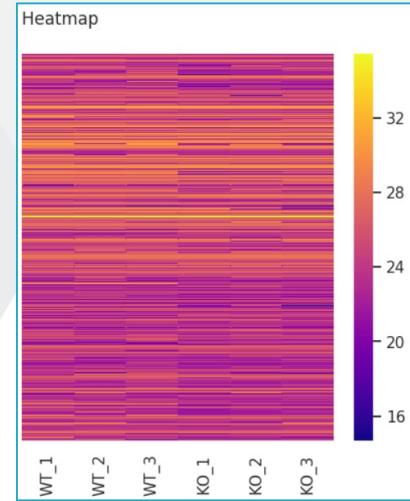
More information on the used API can be found here
<https://reactome.org/AnalysisService/#/identifiers> and here <https://reactome.org/dev/analysis>.

Reactome query

| Table "default" - Rows: 1733 Spec - Columns: 15 Properties Flow Variables | | | | | | | | | | | | | | |
|---------------------------------------------------------------------------|--------------------|-----------------------|-----------------|-----------------|----------------|-----------------|--------------|------------------|------------------|-----------------|--------------------|--------------|--------------------------------------------------|--|
| Row ID | Pathway identifier | Pathway name | #Entities found | #Entities total | Entities ratio | Entities pValue | Entities FDR | #Reactions found | #Reactions total | Reactions ratio | Species identifier | Species name | Submitted entities found | |
| Row0 | R-HSA-8953854 | Metabolism of RNA | 515 | 675 | 0.06 | 0 | 0 | 150 | 187 | 0.016 | 9606 | Homo sapiens | Q92979;075940;P67809;Q9NRW3;Q06265;Q96PZ0;0002 | |
| Row1 | R-HSA-72766 | Translation | 254 | 294 | 0.026 | 0 | 0 | 92 | 99 | 0.008 | 9606 | Homo sapiens | 075821;075822;P62917;Q9Y5M8;Q8IXM3;Q04637;Q9NX2 | |
| Row2 | R-HSA-72203 | Processing of Ca... | 206 | 245 | 0.022 | 0 | 0 | 32 | 32 | 0.003 | 9606 | Homo sapiens | Q16629;Q7RTV0;Q9BY77;075940;P67809;P35658;P1398 | |
| Row3 | R-HSA-72163 | mRNA Splicing - ... | 159 | 180 | 0.016 | 0 | 0 | 9 | 9 | 0.001 | 9606 | Homo sapiens | Q43395;Q16629;Q7RTV0;075940;P67809;P52272;Q86V8 | |
| Row4 | R-HSA-8868773 | rRNA processing ... | 167 | 193 | 0.017 | 0 | 0 | 14 | 15 | 0.001 | 9606 | Homo sapiens | Q92979;P23396;P62917;Q12788;P62081;000541;Q0626 | |
| Row5 | R-HSA-72312 | rRNA processing | 173 | 203 | 0.018 | 0 | 0 | 17 | 21 | 0.002 | 9606 | Homo sapiens | Q92979;P23396;Q7L0Y3;P62917;Q12788;P62081;00054 | |
| Row6 | R-HSA-6791226 | Major pathway of... | 158 | 183 | 0.016 | 0 | 0 | 7 | 7 | 0.001 | 9606 | Homo sapiens | Q92979;P23396;P62917;Q12788;P62081;000541;Q0626 | |
| Row7 | R-HSA-72172 | mRNA Splicing | 159 | 188 | 0.017 | 0 | 0 | 14 | 14 | 0.001 | 9606 | Homo sapiens | Q43395;Q16629;Q7RTV0;075940;P67809;P52272;Q86V8 | |
| Row8 | R-HSA-72613 | Eukaryotic Transl... | 112 | 120 | 0.011 | 0 | 0 | 21 | 21 | 0.002 | 9606 | Homo sapiens | P23396;075821;075822;P62917;P62081;P18124;00030 | |
| Row9 | R-HSA-72737 | Cap-dependent T... | 112 | 120 | 0.011 | 0 | 0 | 18 | 18 | 0.002 | 9606 | Homo sapiens | P23396;075821;075822;P62917;P62081;P18124;00030 | |
| Row10 | R-HSA-72706 | GTP hydrolysis a... | 106 | 113 | 0.01 | 0 | 0 | 3 | 3 | 0 | 9606 | Homo sapiens | P23396;075821;075822;P62917;P62081;P18124;00030 | |
| Row11 | R-HSA-156827 | L13a-mediated tr... | 105 | 112 | 0.01 | 0 | 0 | 3 | 3 | 0 | 9606 | Homo sapiens | P23396;075821;075822;P62917;P62081;P18124;00030 | |
| Row12 | R-HSA-168254 | Influenza Infection | 145 | 171 | 0.015 | 0 | 0 | 21 | 58 | 0.005 | 9606 | Homo sapiens | P07900;P23396;P62917;P62081;P18124;P35658;P1398 | |
| Row13 | R-HSA-168255 | Influenza Life Cycle | 137 | 160 | 0.014 | 0 | 0 | 15 | 50 | 0.004 | 9606 | Homo sapiens | P07900;P23396;P62917;P62081;P18124;P35658;P1398 | |
| Row14 | R-HSA-1799339 | SRP-dependent c... | 104 | 113 | 0.01 | 0 | 0 | 5 | 5 | 0 | 9606 | Homo sapiens | P23396;P62917;P62081;Q9Y5M8;P18124;P62750;P4791 | |
| Row15 | R-HSA-72689 | Formation of a p... | 95 | 102 | 0.009 | 0 | 0 | 2 | 2 | 0 | 9606 | Homo sapiens | Q9Y3U8;P23396;075821;Q71UM5;075822;P62917;P620 | |
| Row16 | R-HSA-9010553 | Regulation of ex... | 142 | 172 | 0.015 | 0 | 0 | 6 | 19 | 0.002 | 9606 | Homo sapiens | P23396;P62917;P62081;P18124;Q04637;P62750;P4791 | |
| Row17 | R-HSA-168273 | Influenza Viral RN... | 128 | 151 | 0.014 | 0 | 0 | 6 | 13 | 0.001 | 9606 | Homo sapiens | P07900;P23396;P62917;P62081;P18124;P35658;P1398 | |
| Row18 | R-HSA-975956 | Nonsense Media... | 89 | 96 | 0.009 | 0 | 0 | 1 | 1 | 0 | 9606 | Homo sapiens | Q9Y3U8;P23396;Q71UM5;P62917;P62081;P62280;P181 | |
| Row19 | R-HSA-156842 | Eukaryotic Transl... | 87 | 95 | 0.008 | 0 | 0 | 9 | 9 | 0.001 | 9606 | Homo sapiens | Q9Y3U8;P23396;Q71UM5;Q5VTE0;P62917;P29692;P620 | |
| Row20 | R-HSA-162909 | Host Interactions | 120 | 144 | 0.013 | 0 | 0 | 46 | 54 | 0.005 | 9606 | Homo sapiens | P35658;Q9HC16;P62195;060563;P62877;Q9BW27;P632 | |
| Row21 | R-HSA-72764 | Eukaryotic Transl... | 86 | 94 | 0.008 | 0 | 0 | 5 | 5 | 0 | 9606 | Homo sapiens | Q9Y3U8;P23396;Q71UM5;P62917;P62081;P62280;P181 | |
| Row22 | R-HSA-156902 | Peptide chain el... | 83 | 90 | 0.008 | 0 | 0 | 5 | 5 | 0 | 9606 | Homo sapiens | Q9Y3U8;P23396;Q71UM5;P62917;P62081;P62280;P181 | |
| Row23 | R-HSA-2408522 | Selenoamino aci... | 102 | 118 | 0.011 | 0 | 0 | 18 | 23 | 0.002 | 9606 | Homo sapiens | P23396;P00390;P62917;P62081;P18124;P62750;P4791 | |
| Row24 | R-HSA-927802 | Nonsense-Media... | 100 | 117 | 0.01 | 0 | 0 | 6 | 6 | 0.001 | 9606 | Homo sapiens | P23396;P62917;P62081;P18124;Q04637;P62750;P4791 | |
| Row25 | R-HSA-975957 | Nonsense Media... | 100 | 117 | 0.01 | 0 | 0 | 5 | 5 | 0 | 9606 | Homo sapiens | P23396;P62917;P62081;P18124;Q04637;P62750;P4791 | |
| Row26 | R-HSA-2408557 | Selenocysteine s... | 83 | 94 | 0.008 | 0 | 0 | 4 | 7 | 0.001 | 9606 | Homo sapiens | Q9Y3U8;P23396;Q71UM5;P62917;P62081;P62280;P181 | |
| Row27 | R-HSA-450531 | Regulation of mR... | 78 | 88 | 0.008 | 0 | 0 | 26 | 26 | 0.002 | 9606 | Homo sapiens | P0CG47;P0DMV8;P51665;Q06265;P35658;Q07352;POCG | |
| Row28 | R-HSA-72649 | Translation initia... | 57 | 59 | 0.005 | 0 | 0.001 | 2 | 2 | 0 | 9606 | Homo sapiens | P23396;075821;Q71UM5;075822;P62081;P62280;0003 | |
| Row29 | R-HSA-72702 | Ribosomal scann... | 57 | 59 | 0.005 | 0 | 0.001 | 2 | 2 | 0 | 9606 | Homo sapiens | P23396;075821;Q71UM5;075822;P62081;P62280;0003 | |
| Row30 | R-HSA-376176 | Signaling by ROB... | 160 | 218 | 0.019 | 0 | 0.001 | 21 | 59 | 0.005 | 9606 | Homo sapiens | P23396;P62917;P62081;P12931;P18124;P13861;Q0463 | |
| Row31 | R-HSA-72662 | Activation of the ... | 57 | 60 | 0.005 | 0 | 0.001 | 6 | 6 | 0.001 | 9606 | Homo sapiens | P23396;075821;Q71UM5;075822;P62081;P62280;0003 | |
| Row32 | R-HSA-450408 | AUF1 (hnRNP D0)... | 53 | 56 | 0.005 | 0 | 0.002 | 4 | 4 | 0 | 9606 | Homo sapiens | P0CG47;P0DMV8;P51665;POCG48;Q04637;P62195;Q925 | |
| Row33 | R-HSA-68875 | Mitotic Prophase | 95 | 119 | 0.011 | 0 | 0.002 | 32 | 34 | 0.003 | 9606 | Homo sapiens | P06899;P28482;Q9H8Y8;P14635;P35658;Q71UI9;Q9HO | |
| Row34 | R-HSA-72695 | Formation of the ... | 50 | 52 | 0.005 | 0 | 0.002 | 3 | 3 | 0 | 9606 | Homo sapiens | P23396;075821;Q71UM5;075822;P62081;P62280;0003 | |
| Row35 | R-HSA-192823 | Viral mRNA Trans... | 83 | 101 | 0.009 | 0 | 0.002 | 2 | 2 | 0 | 9606 | Homo sapiens | Q9Y3U8;P23396;Q71UM5;P62917;P62081;P62280;P181 | |
| Row36 | R-HSA-5578749 | Transcriptional r... | 69 | 80 | 0.007 | 0 | 0.002 | 3 | 5 | 0 | 9606 | Homo sapiens | P06899;P35658;Q71UI9;P52434;P52435;Q99567;P5705 | |
| Row37 | R-HSA-6790901 | rRNA modificatio... | 55 | 60 | 0.005 | 0 | 0.003 | 7 | 8 | 0.001 | 9606 | Homo sapiens | Q9H0A0;P35658;Q71UI9;P52434;P52435;Q99567;P5705 | |
| Row38 | R-HSA-150221 | Transport of Mat | 45 | 46 | 0.004 | 0 | 0.002 | 2 | 2 | 0 | 9606 | Homo sapiens | P52209;P8CG34;Q96V8;Q10PRTV1;P0TEM1;P57740;Q1272 | |

Further development

- **Visualization metanodes:**
 - Heatmap (static) ✓
 - Heatmap (interactive) ✓
 - tSNE ✓
 - PCA
 - Volcano plot (interactive) ✓
 - Plotly based interactive visualizations?
 - **Databases**
 - DAVID
 - String
 - iRefIndex
 - **Cytoscape**
 - **BLAST**
 - **OpenMS utilization for raw MS data processing**





Thank you
for your
attention

Workshop outline

- morning session – theoretical part
 - 10:00 – 10:15 Opening and introduction
 - 10:15 – 10:45 Software container running KNIME
 - 10:45 – 11:00 Coffee break
 - 11:00 – 11:30 Introduction to KNIME
 - 11:30 – 11:45 Coffee break
 - 11:45 – 12:30 Practical applications, our KNIME metanodes
 - **12:30 – 13:30 Lunch break, visit of our laboratories for interested people**

Other metanodes – overviews

Data processing

Normalization (apply factors)

Normalization (linear)

Normalization (non-linear)

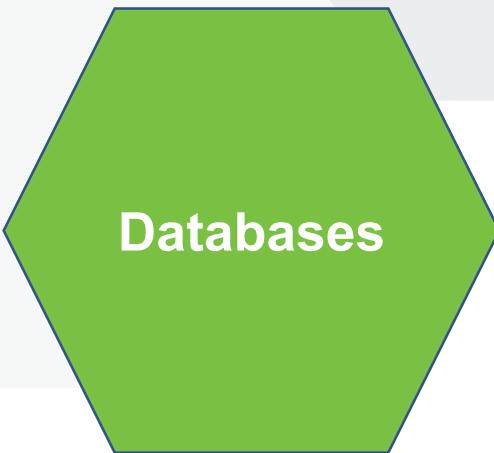
Ratio calculator

Rows normalization (linear)

Signal-to-inference correction

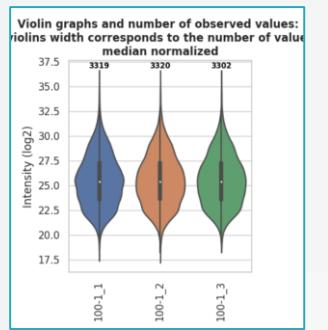
Split and summarize

Values imputation (imp4p)

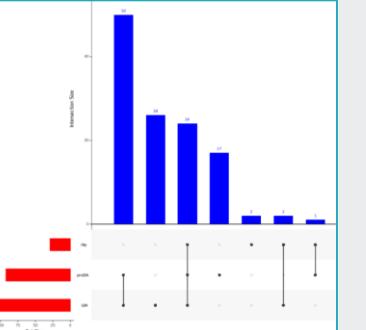


UniProt query

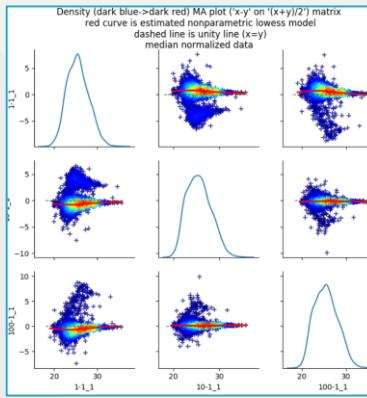
Reactome query



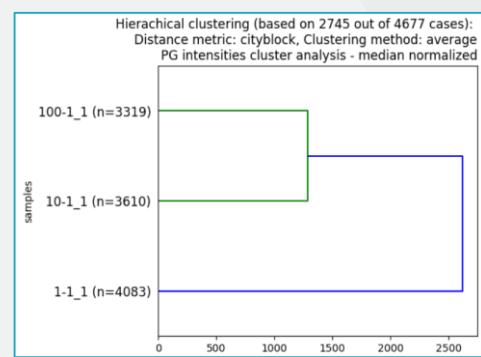
Violin plot



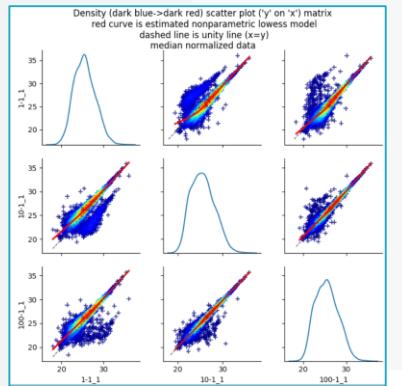
UpSet plot



MA plot

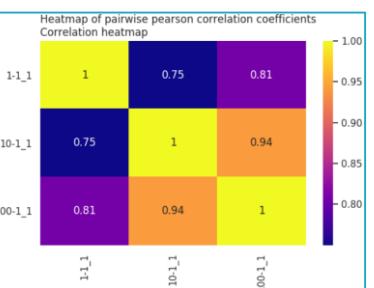


Hierarchical clustering

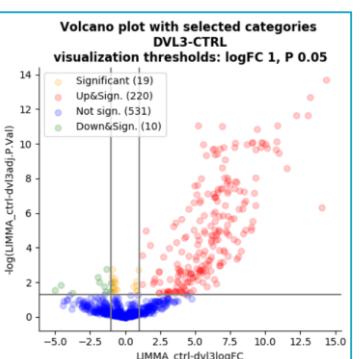


Scatter plot matrix

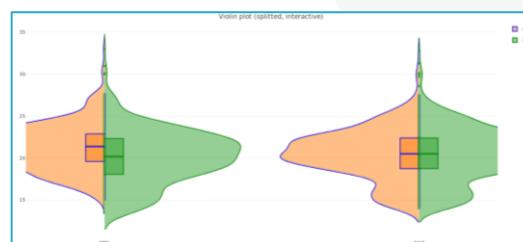
Graphs



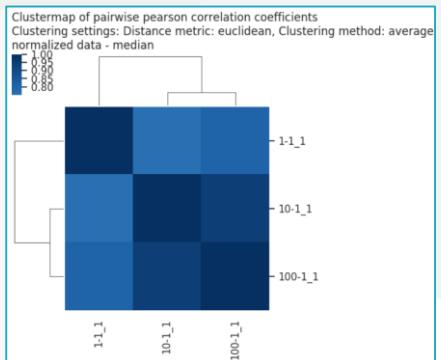
Correlation heatmap



Volcano plot



Splitted violin plot



Correlation clustermap

Manipu- lation

Binary matrix calculator

FASTA to table

Table to FASTA

Values lookup

Substring extraction



Column descriptive statistics

LIMMA test

proDA

Other
metanodes
– individual
metanodes
info

Data processing

Normalization (apply factors)

Normalization (linear)

Normalization (non-linear)

Ratio calculator

Rows normalization (linear)

Signal-to-inference correction

Split and summarize

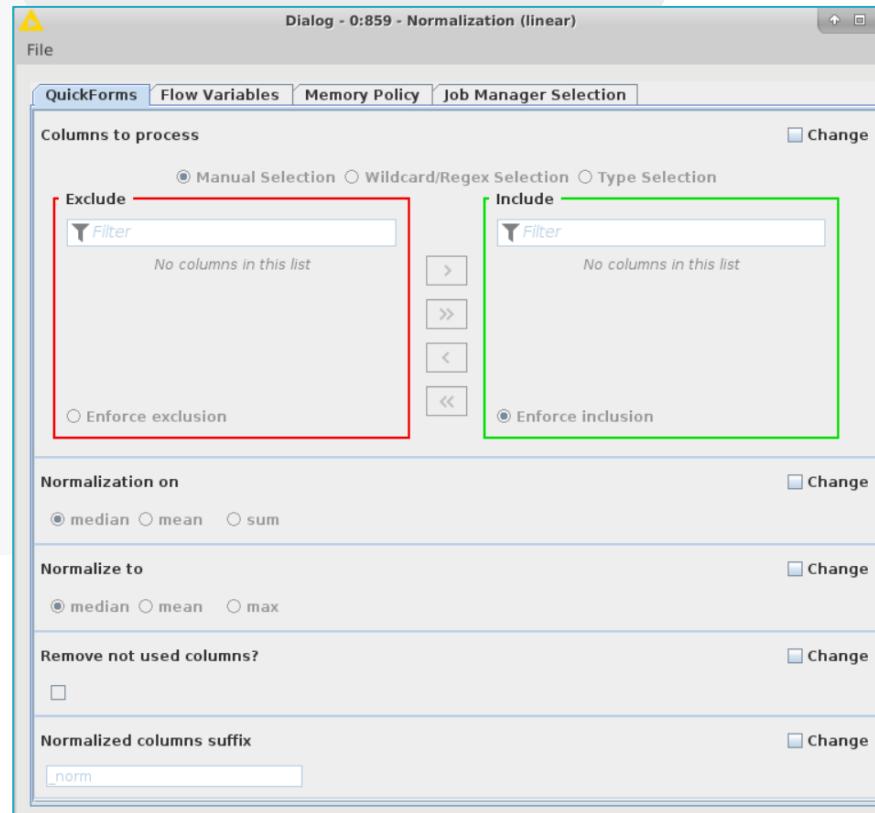
Values imputation (imp4p)

Normalization

Normalization (apply factors)

Normalization (linear)

Normalization (non-linear)



Normalization (linear)

Data normalization metanode using linear normalization methods (on median/mean/sum).

You can select what columns should be processed by the metanode, no need to preselect them before the node.

The node can return original dataframe with normalized columns or just selected columns after applied normalization.

Normalization factors are returned as well next to the normalized data. They have the same orientation as selected data columns.

Normalized columns suffix can be selected.

ATTENTION!: if you select blank normalized column suffix and not used columns will be returned, original data in processed columns will be rewritten by the normalized ones! This is desired behavior.

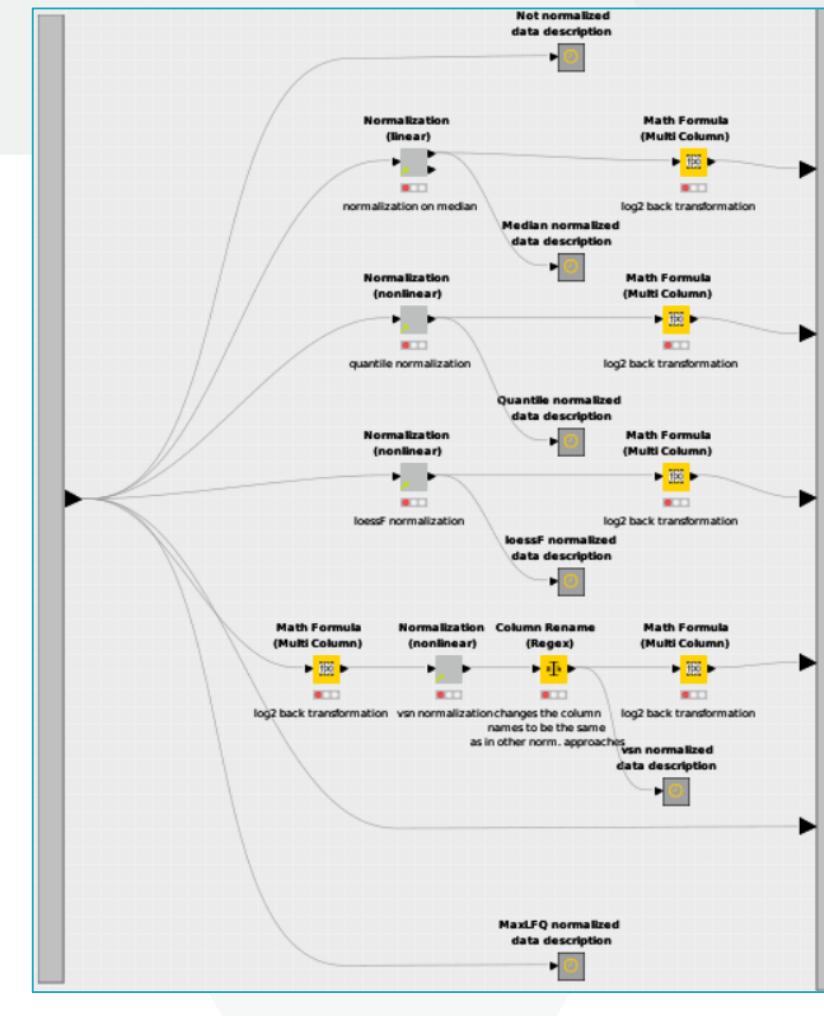
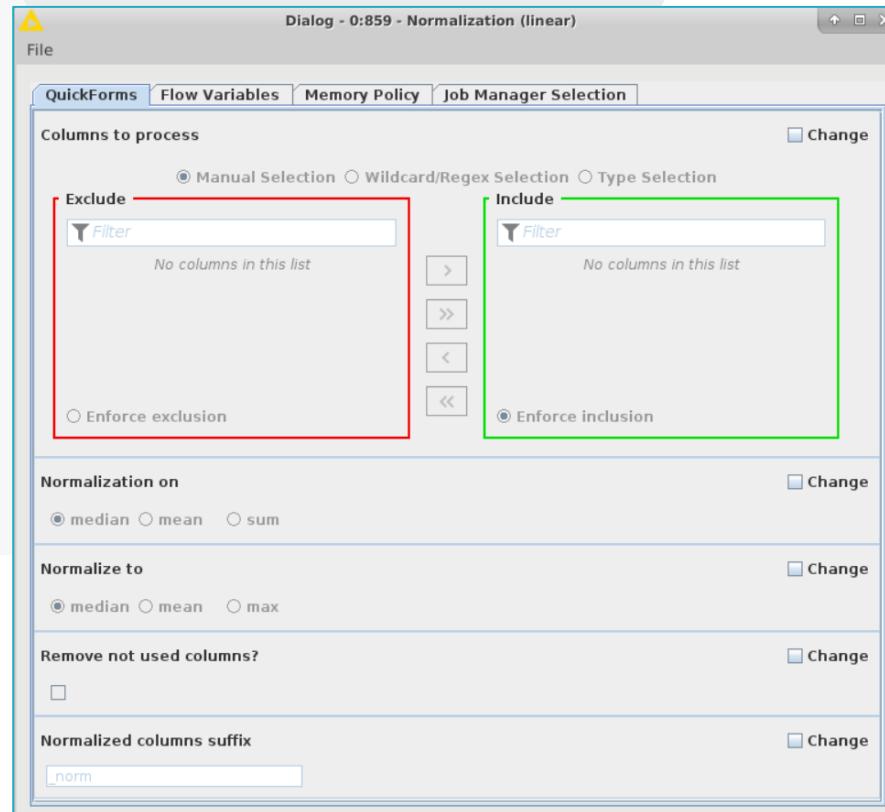
- Removal of technical processing artefacts from data

Normalization

Normalization (apply factors)

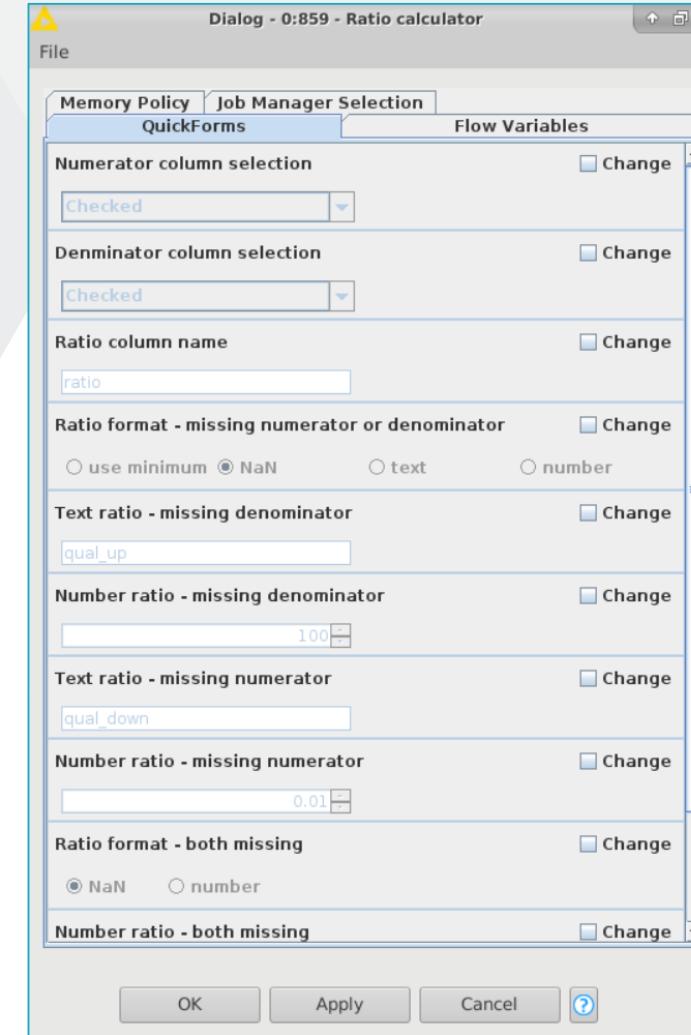
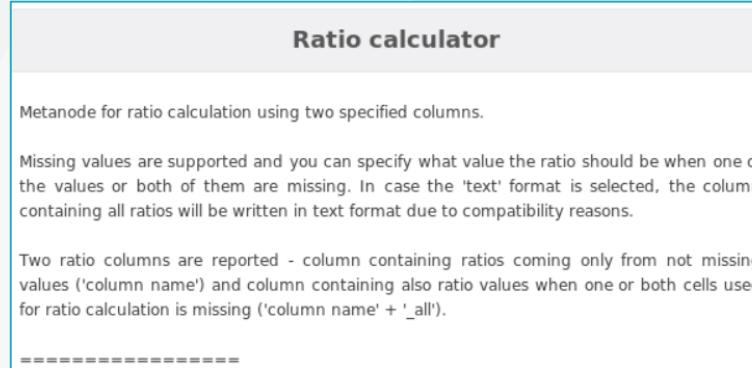
Normalization (linear)

Normalization (non-linear)



- Removal of technical processing artefacts from data

Ratio calculator



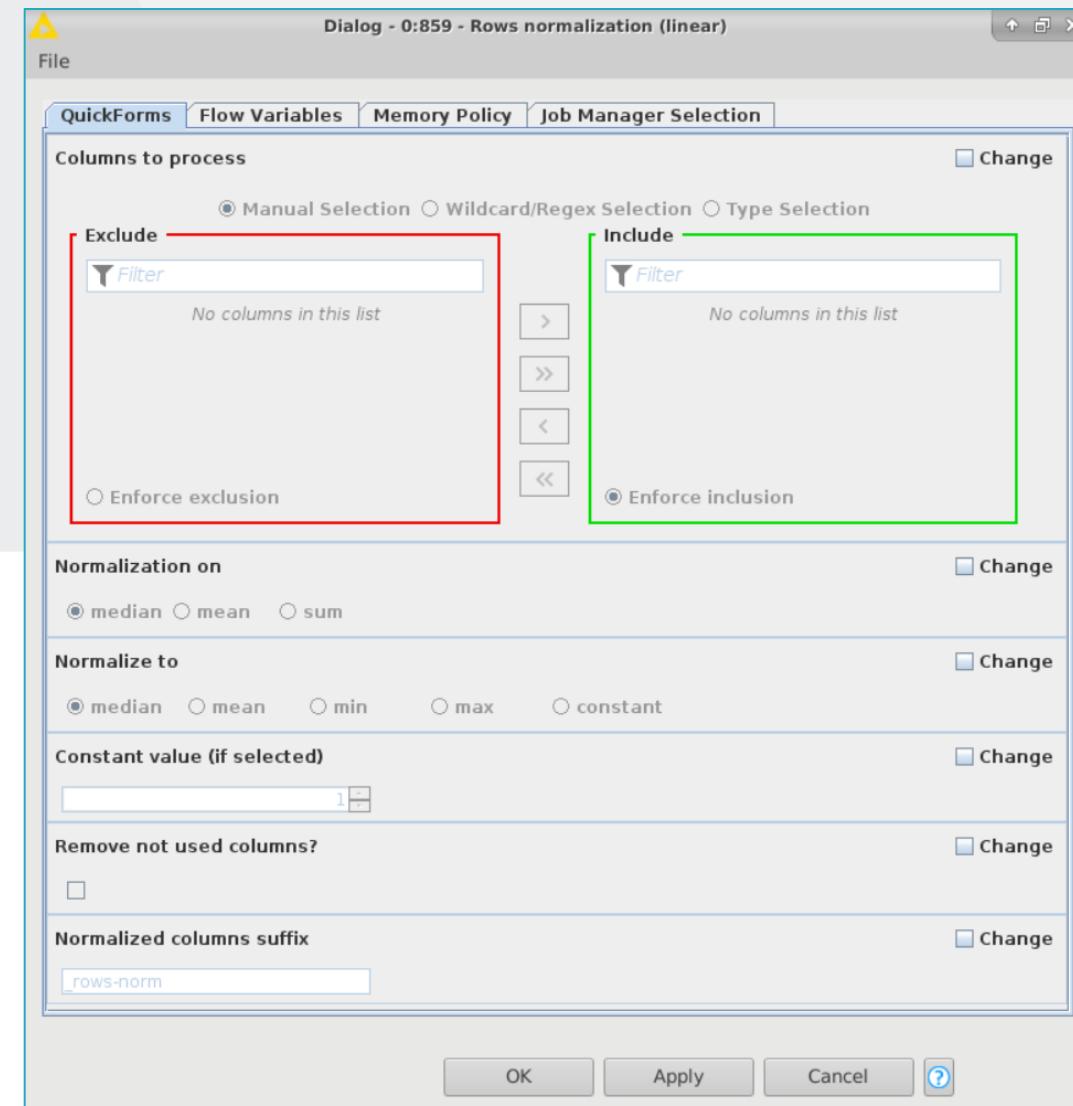
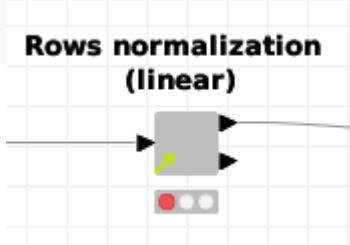
- Calculation of samples ratios (Up/Down-regulated)
- In case no replicates would be provided, so statistics (e.g. LIMMA) is not possible
- E.g. in case of individual patients measurements

Rows normalization

Rows normalization (linear)

Rows normalization metanode using selected normalization methods. Individual rows are normalized separately based on the selection technique.

ATTENTION! if you select blank normalized column suffix and not used columns will be returned, original data in processed columns will be rewritten by the normalized ones! This is desired behavior.



- Row-wise data normalization

Signal-to-interference correction

Signal-to-interference (S2I) correction

Calculates corrected values from the selected columns using selected column containing signal to interference (S2I) ratio (1 = no interference; 0 = complete interference).

Individual rows in data columns are at first normalized/scaled so their sum is equal to 1 (each value is divided by the sum across the whole row of data columns).

Median values in all columns are used as the estimate of interference values. In case grouping column(s) is selected, median is calculated groupwise.

Estimated interference values are normalized to give 1.000 (they are divided by their sum).

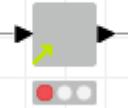
Correction is done by subtracting median interference value for each channel multiplied by the (1-S2I) factor for the given PSMs.

Negative values after the correction are replaced by specified value.

Calculations based on S2I correction as published here: Savitski, M. M. et al. Measuring and Managing Ratio Compression for Accurate iTRAQ/TMT Quantification. *J. Proteome Res.* 12, 3586-3598 (2013).

Metanode returns full data matrix with corrected values in selected columns or processed columns only with grouping column(s) if used - based on the setting selected.

Signal-to-interference (S2I) correction



- Signal-to-interference (S2I) correction based on the given publication
- for removal of expected interference signal intensities from e.g. TMT reporter ion intensities

Dialog - 0:950 - Signal-to-interference (S2I) correction

File

QuickForms Flow Variables Memory Policy Job Manager Selection

Signal-to-interference (S2I) column

Pattern: S2I

Change

Manual Selection Wildcard/Regex Selection Type Selection

Wildcard Regular expression Case Sensitive

Mismatch (Exclude) Match (Include)

Columns to process

Exclude

Filter No columns in this list

Include

Filter No columns in this list

Enforce exclusion Enforce inclusion

Grouping column(s)

Exclude

Filter No columns in this list

Include

Filter No columns in this list

Enforce exclusion Enforce inclusion

Corrected columns suffix

S2Icorr

Minimal S2Icorr value (0-1)

0

Replacement method

zero NaN

Remove not used columns?

OK - Execute Apply Cancel ?

CEITEC - Centre of Macademy of Sciences of the Czech Republic

Split and summarize

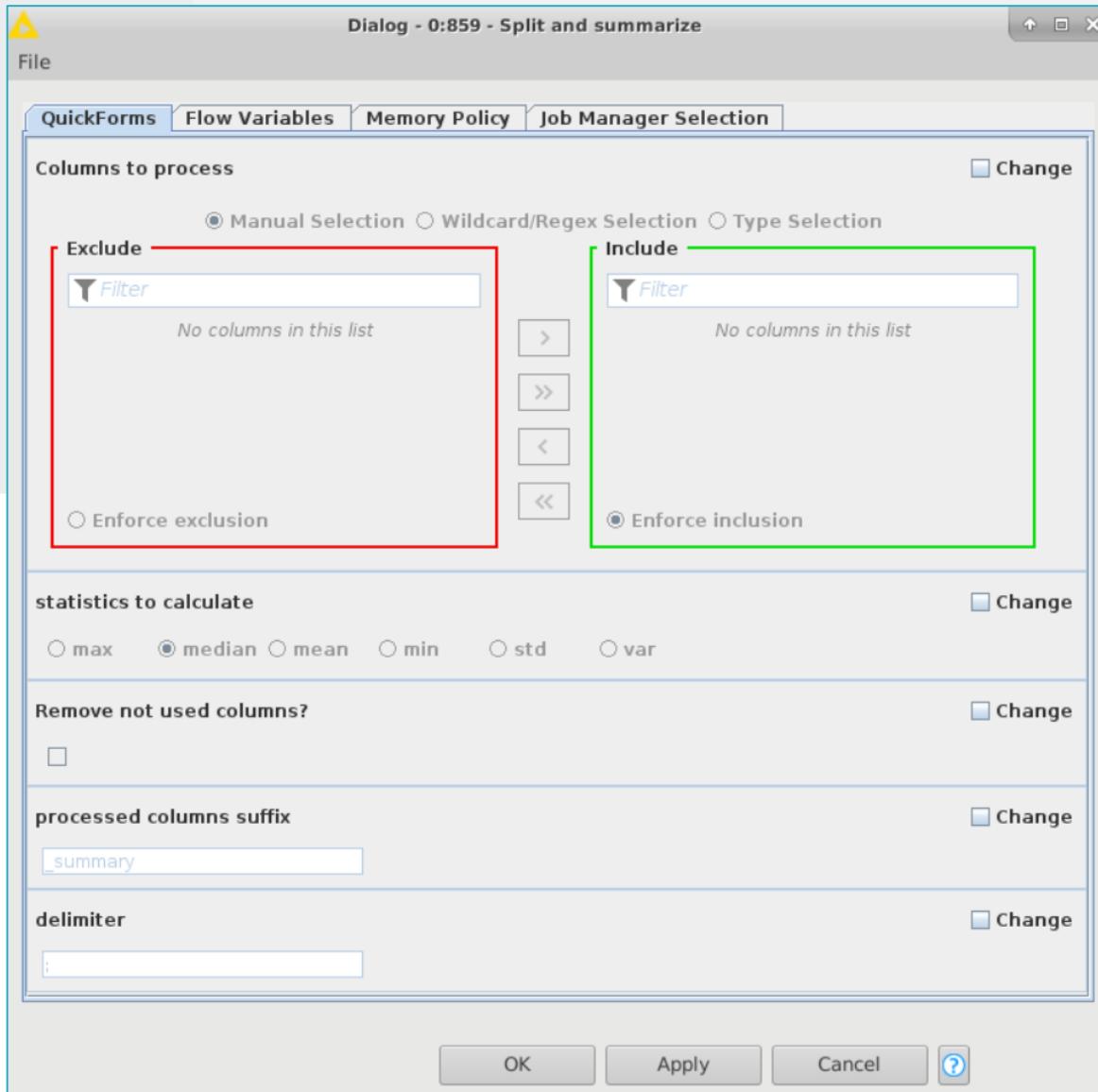
Split and summarize

Metanode for processing content of individual column cells containing one or several numerical values separated by delimiter.

Individual values from a single column cell are extracted using the provided delimiter and converted to numbers (text and missing values are ignored). Selected descriptive statistics (e.g. median) is calculated in the end for each column cell, ignoring any NaN value.

These descriptive statistics are implemented using numpy (<https://docs.scipy.org/doc/numpy/reference/routines.statistics.html>):
max - maximal value
mean - average value
median - median
min - minimal value
std - standard deviation
var - variance

Please note that all methods are implemented in their "nan" variant (e.g. nanmax), i.e. ignoring any NaNs.



- If more numbers are in one cell separated by particular delimiter, statistics (max/median/mean...) resulting in one number is provided

Values imputation (imp4p)

- Missing values in LC-MS/MS experiments are frequent (50-70 %)
- Problem for differential expression analysis
- Several approaches, often used global minimum, kNN, MinProb...
- Imputation done on peptide/protein level?
- **Imp4p R package**

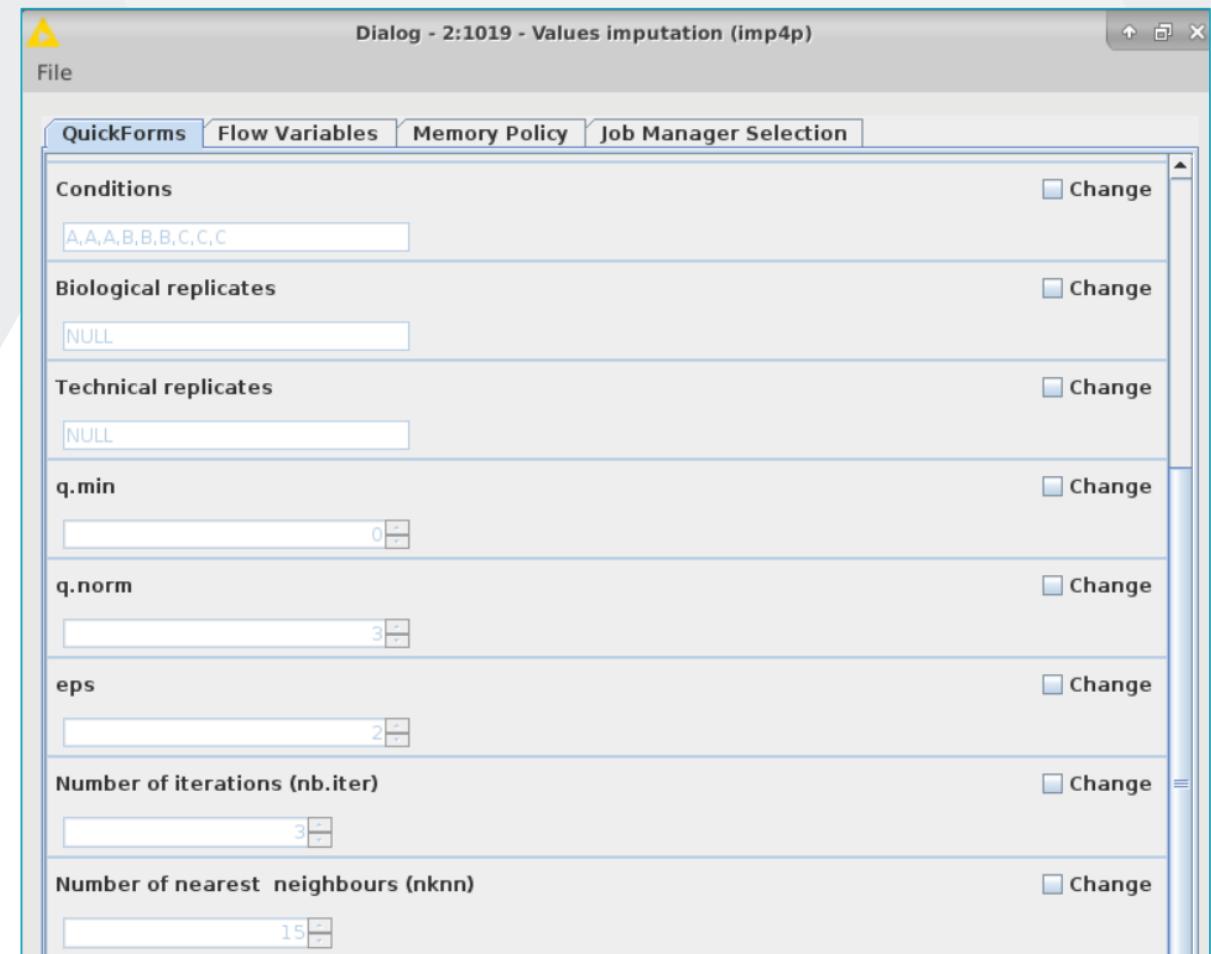
Values imputation (imp4p)

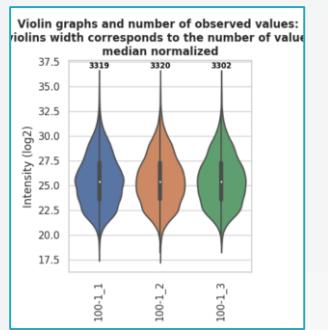
Node for the imputation of missing values in bottom-up MS-based quantitative proteomic data using R package imp4p.

The node imputes the missing values under both missing not completely at random (MNAR) and missing completely at random (MCAR) assumptions. The imputed values are provided as new columns of input dataframe with the user-specified suffix. Additionally, user is provided on the second output a binary matrix, where 0 stands for non-imputed values, 1 for the values which were imputed. Rows, where all values were missing contain "?".

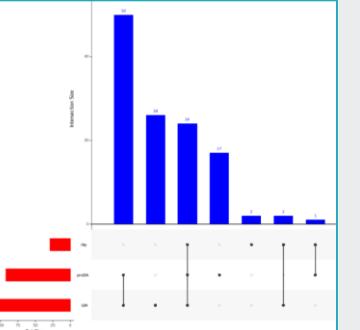
For the input user should provide log2 transformed normalized protein or peptide intensities (in a format of e.g. MaxQuant software output).

For further information please refer to the imp4p package description: <https://cran.r-project.org/web/packages/imp4p/imp4p.pdf>

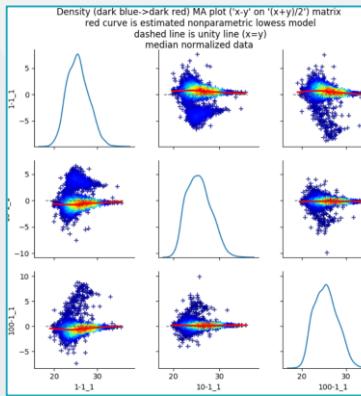




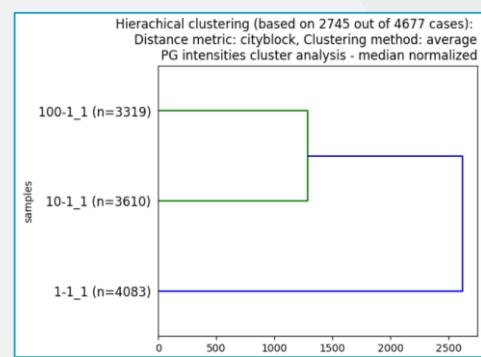
Violin plot



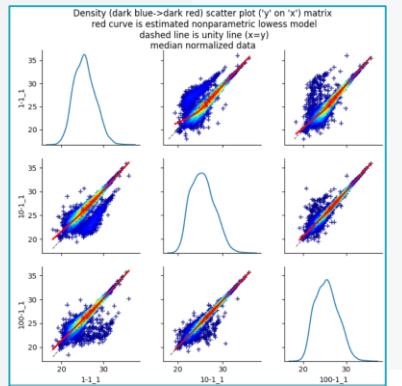
UpSet plot



MA plot

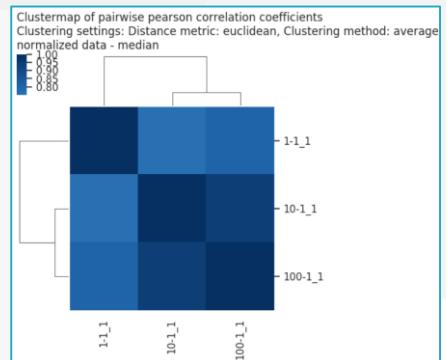


Hierarchical clustering

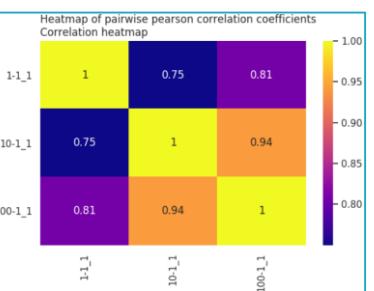


Scatter plot matrix

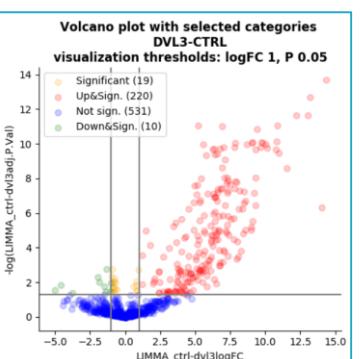
Graphs



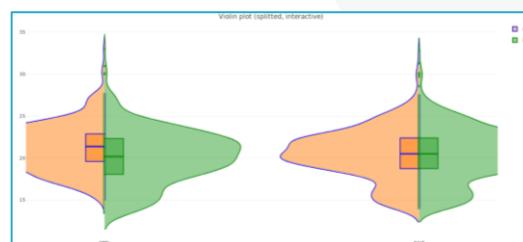
Correlation clustermap



Correlation heatmap



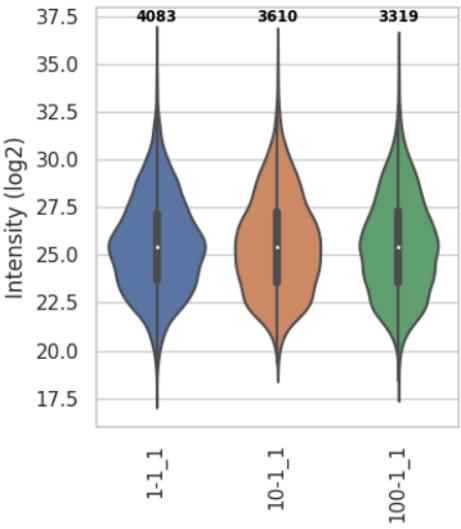
Volcano plot



Splitted violin plot

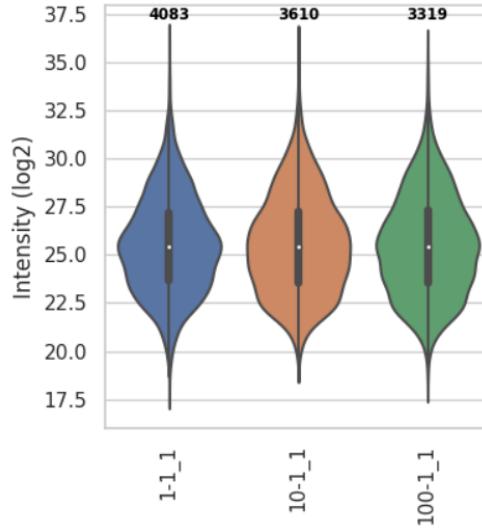
Violin plot

Violin graphs and number of observed values:
violins width corresponds to the number of values
median normalized

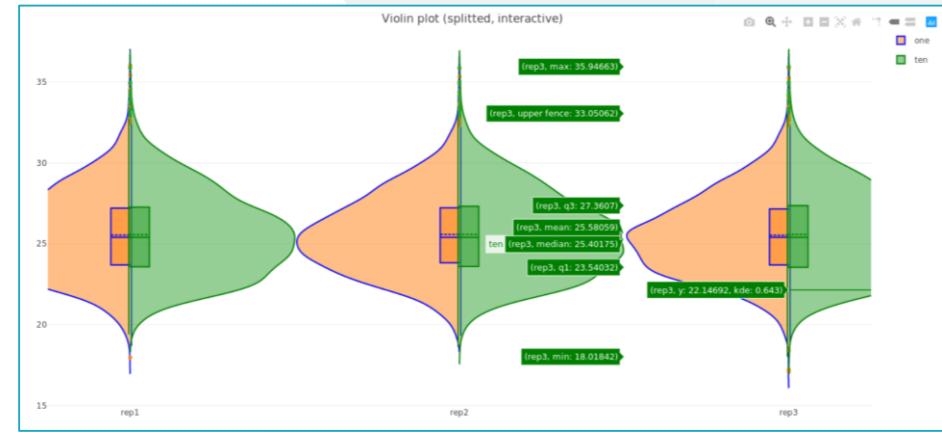
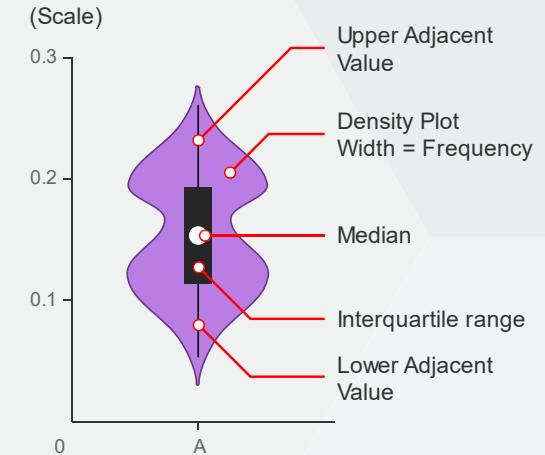


Width based on count

Violin graphs and number of observed values:
all violins have the same width
median normalized



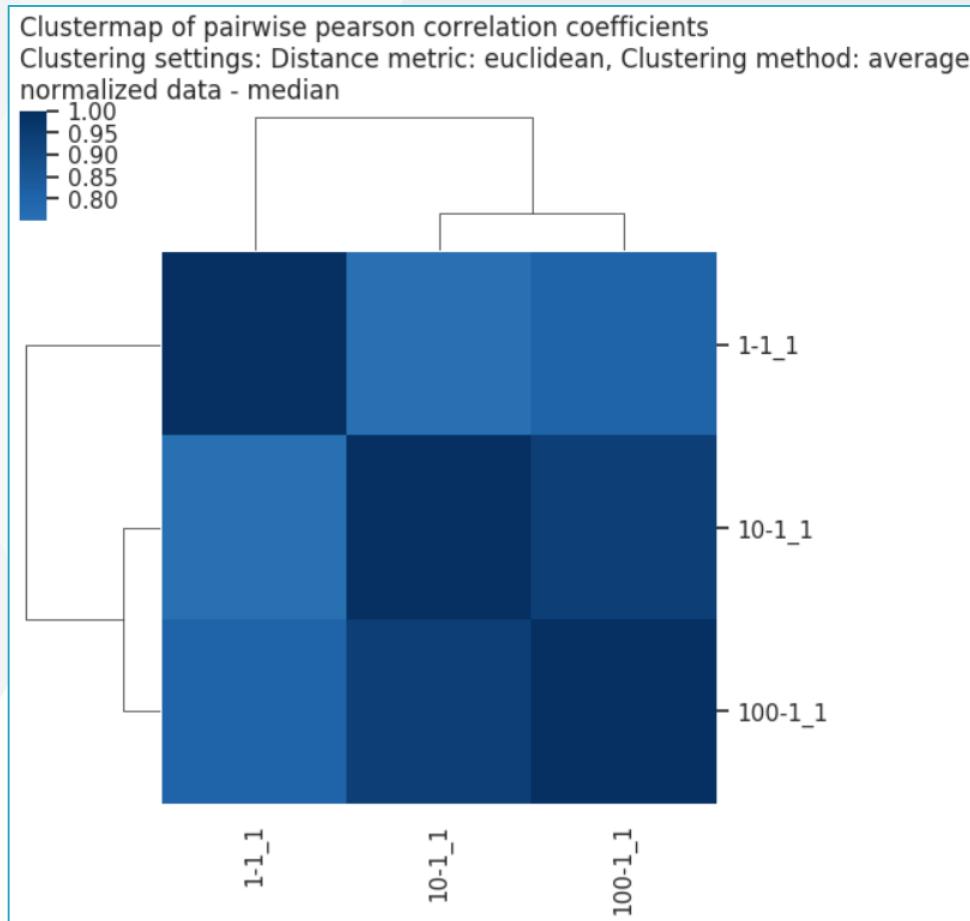
Width same for all plots



Splitted violin plots (interactive)

- **Comparison of samples:** number of cases, distribution, basic statistics (min, max, median)
- Splitted violin plot: before/after, pair comparisons

Correlation clustermap



Correlation method

pearson

Clustermap colormap

plasma Blues RdBu jet rainbow

Manual clustermap center?

Manual clustermap center

Show raw data?

Distance metric

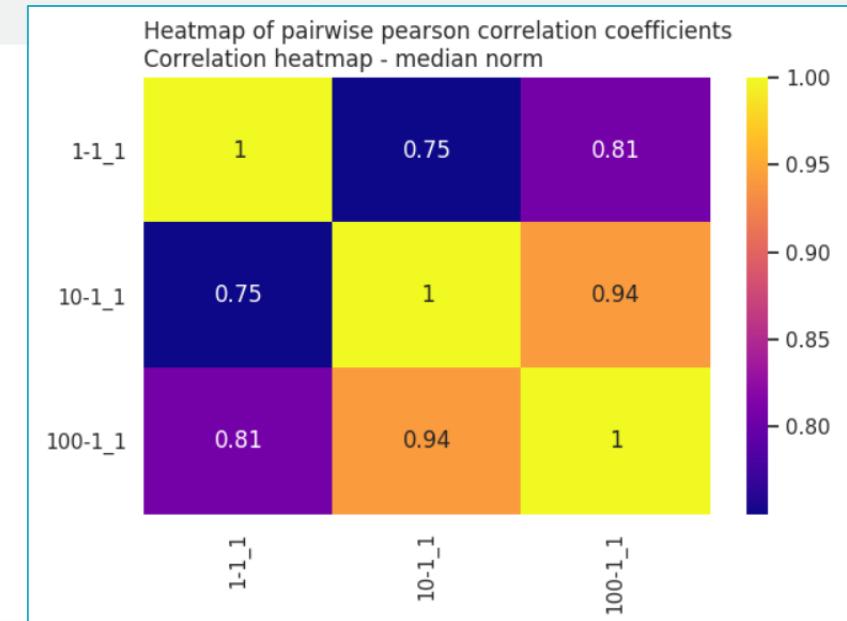
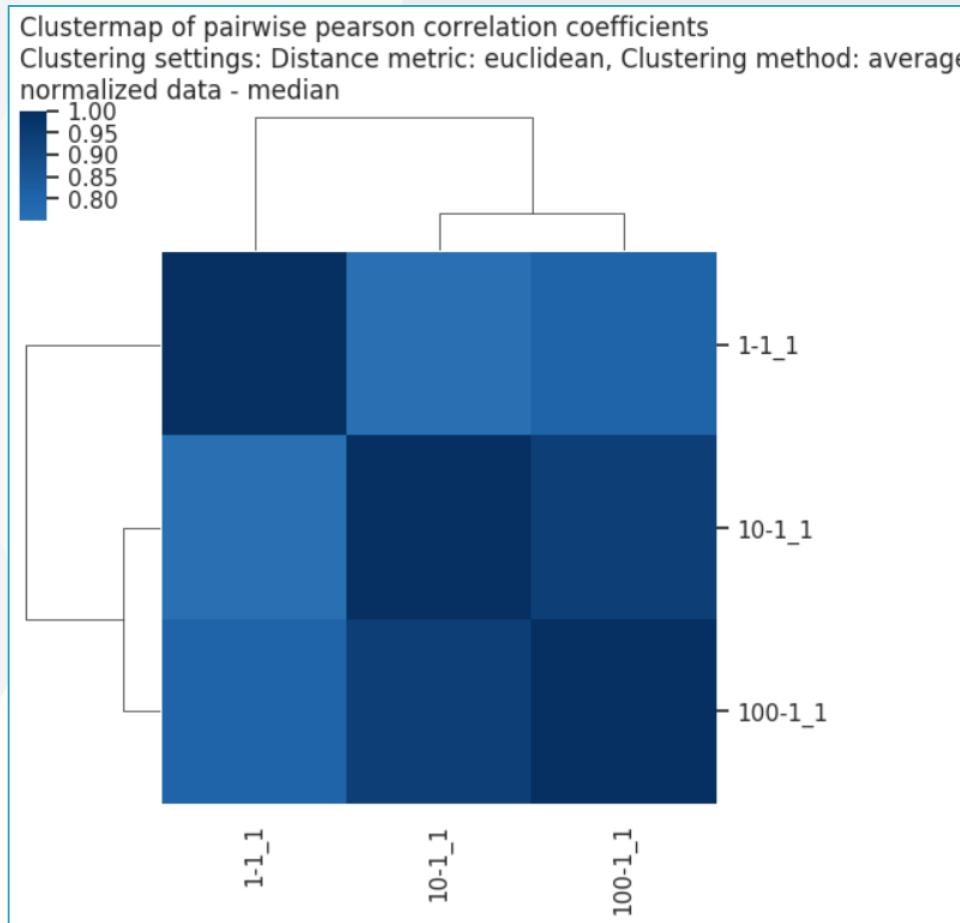
euclidean

Clustering method

average

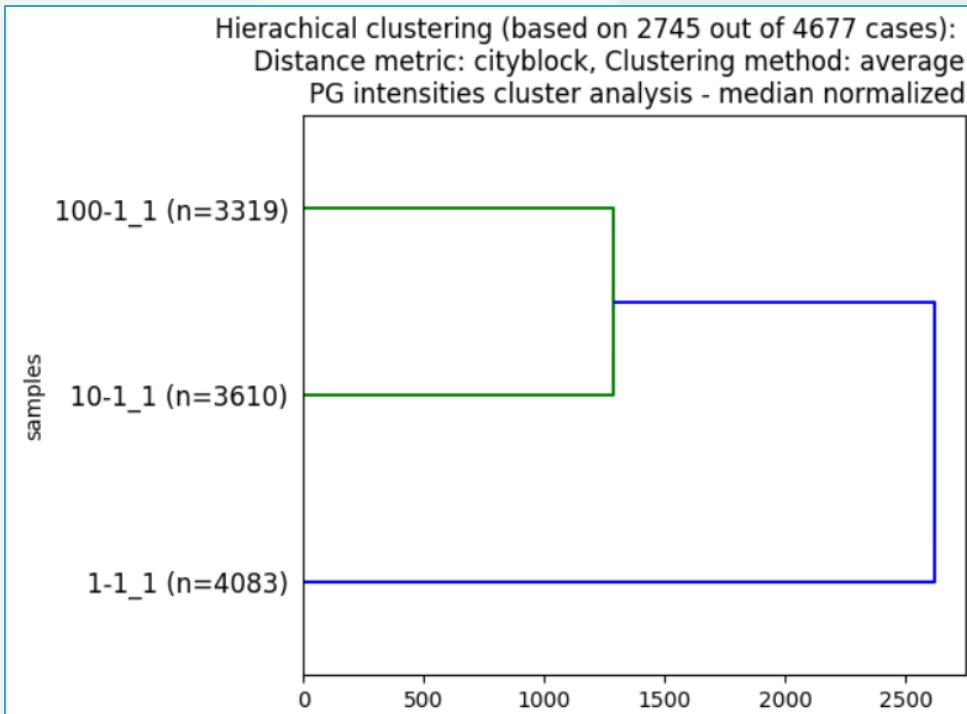
- **Comparison of samples:** which samples tend to form clusters
- How much are samples similar to each other (Pearson's coefficient)

Correlation clustermap (heatmap)

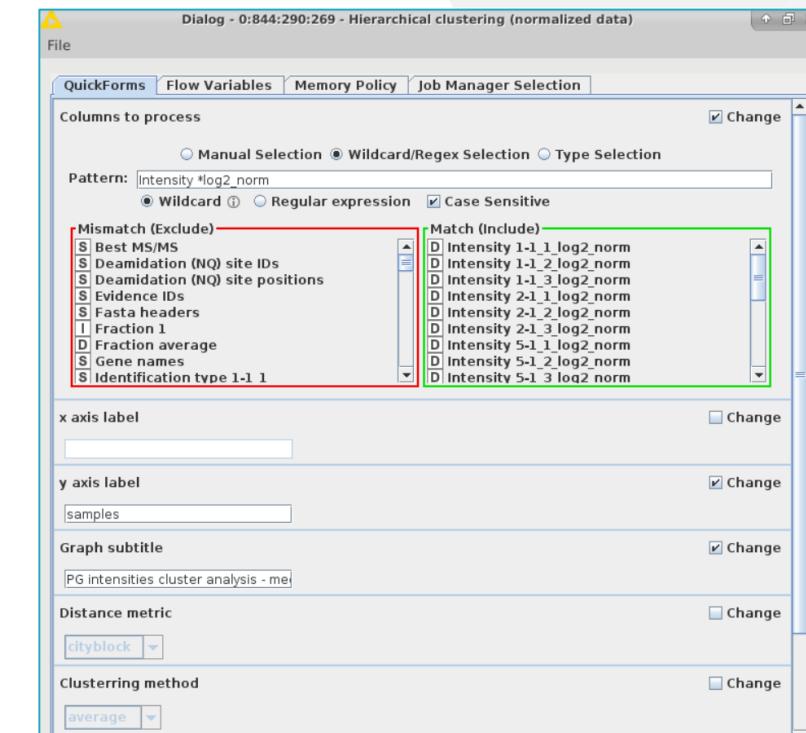


- **Comparison of samples:** which samples tend to form clusters
- Another graphical representation of how samples are similar to each other

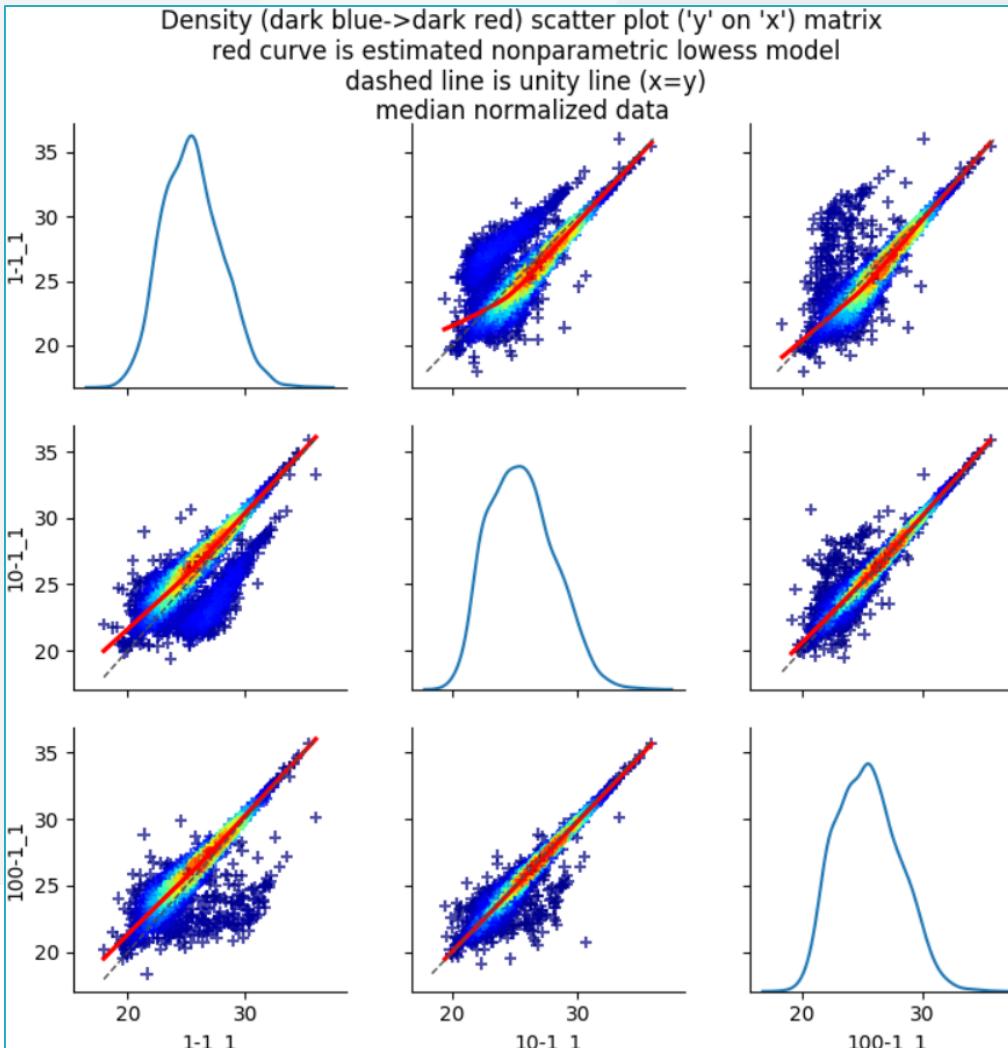
Hierarchical clustering



- **Comparison of samples**
- More similar samples tend to form clusters
- Only rows with non-NA values in all conditions are considered!

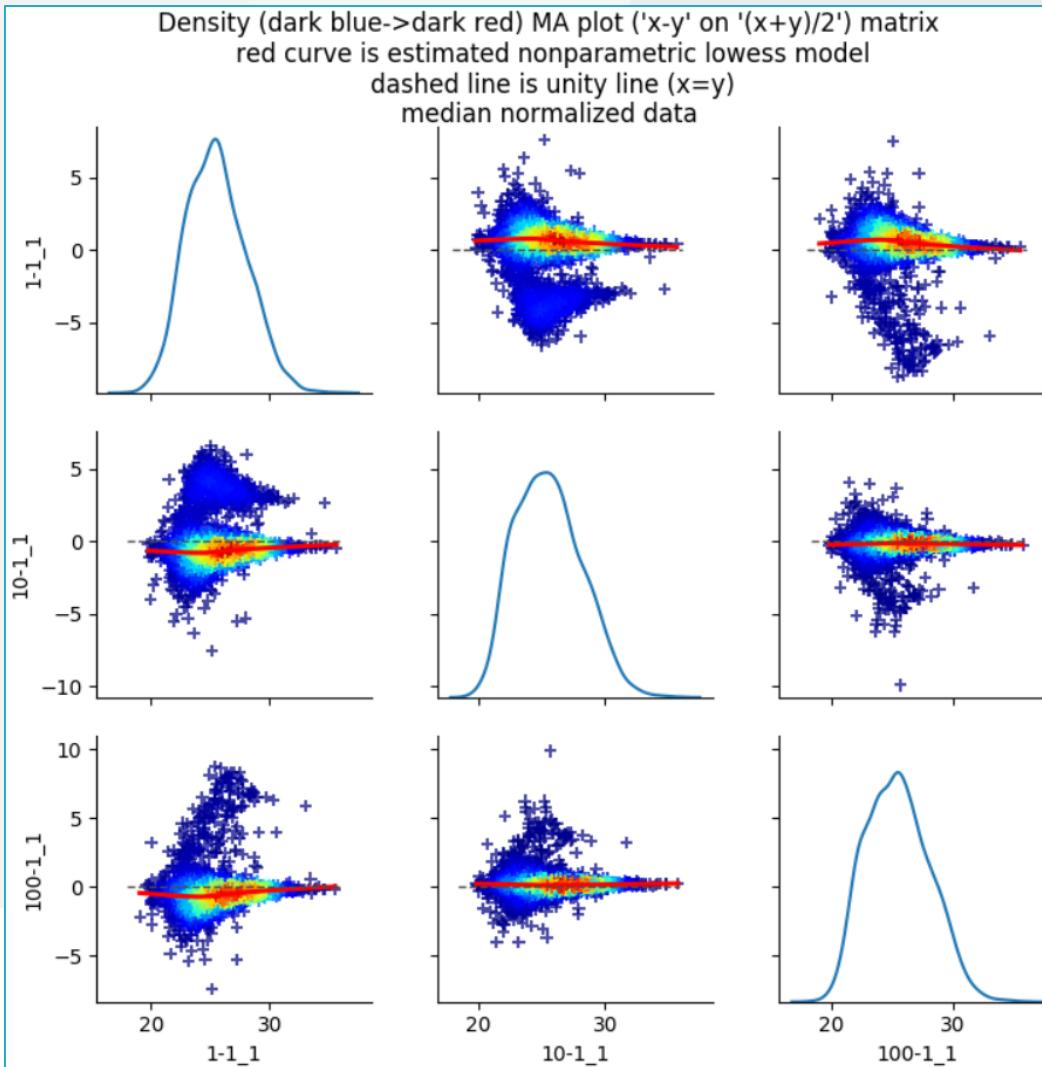


Scatter plot matrix

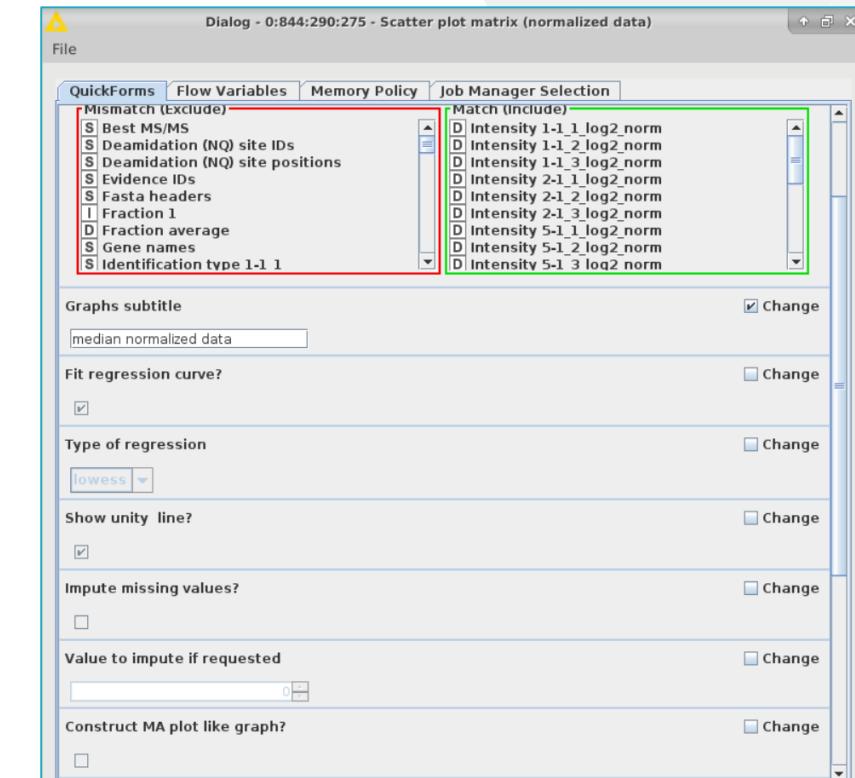


- Intensity correlation for all possible sample pairs
- Dashed line $\rightarrow y=x$
- Colors provide the density information (i.e. blue – low density; red – high density)
- Information about how samples are similar
- Normalization “quality check”

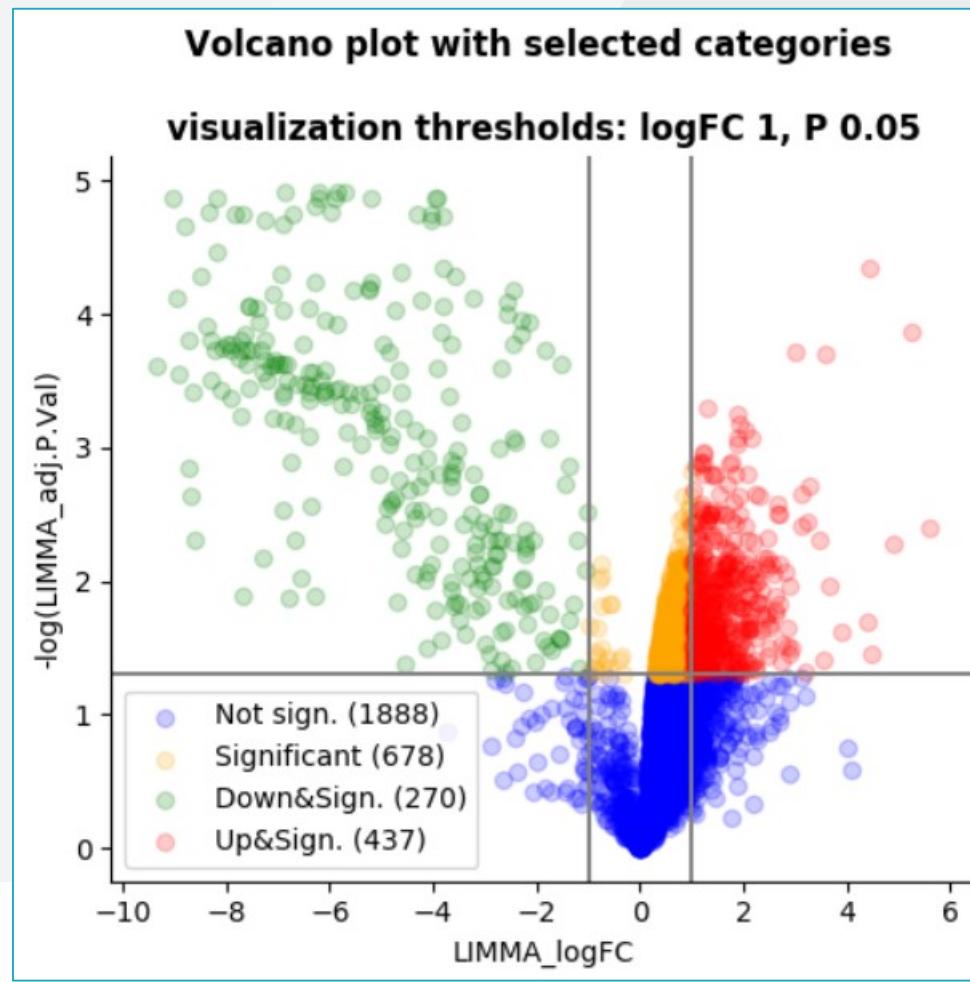
MA plot



- Difference vs average (i.e. $x-y \sim (x+y)/2$)
- Dashed line: $x=y$
- How much do the sample differ from the average value

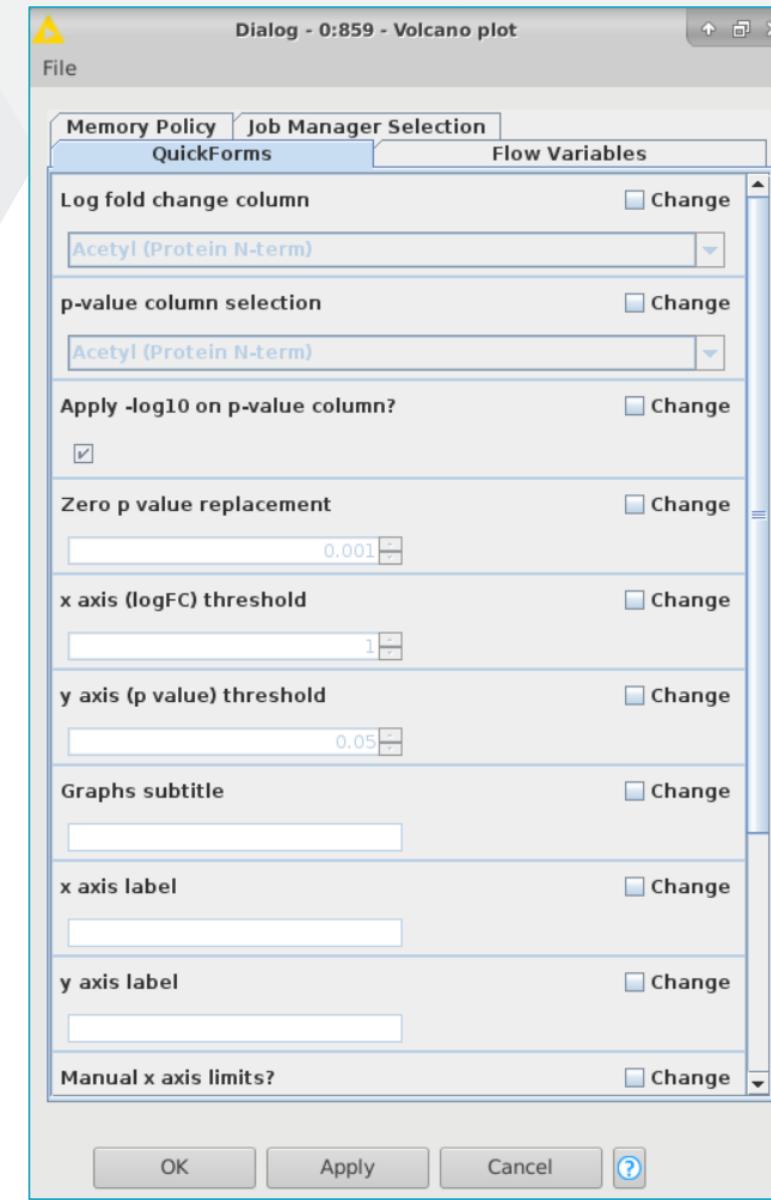
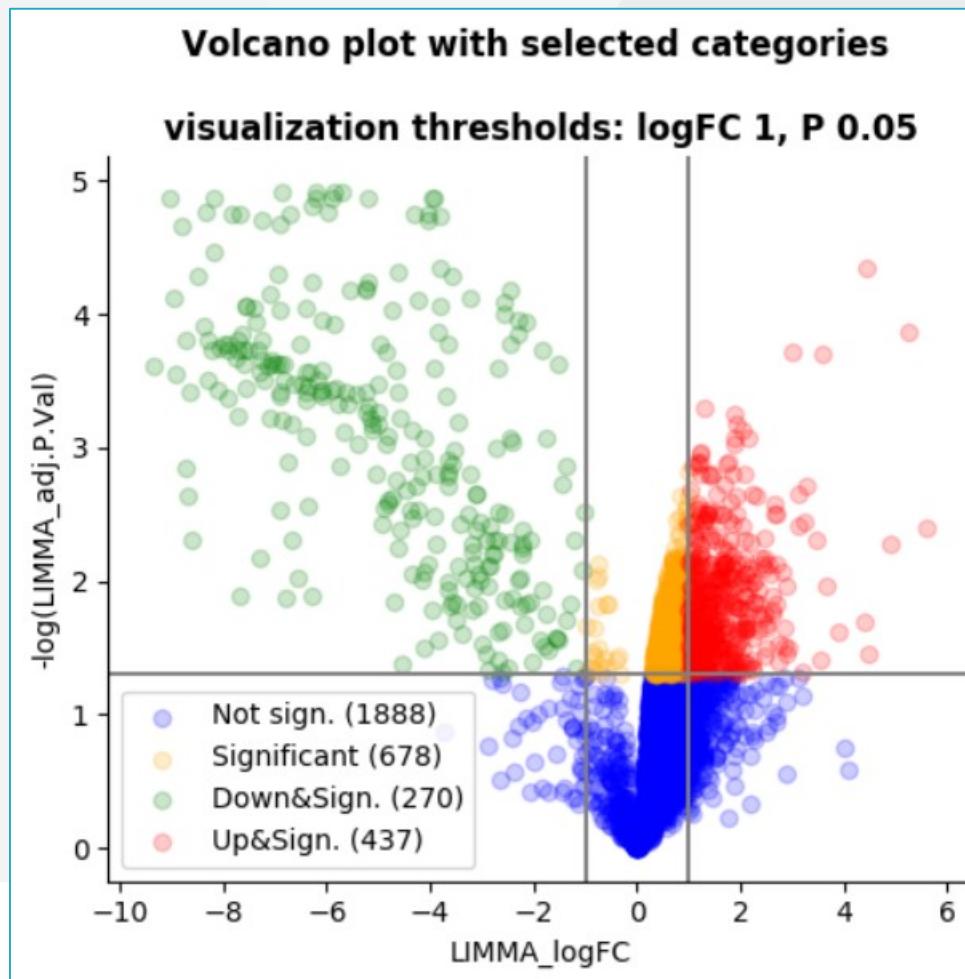


Volcano plot

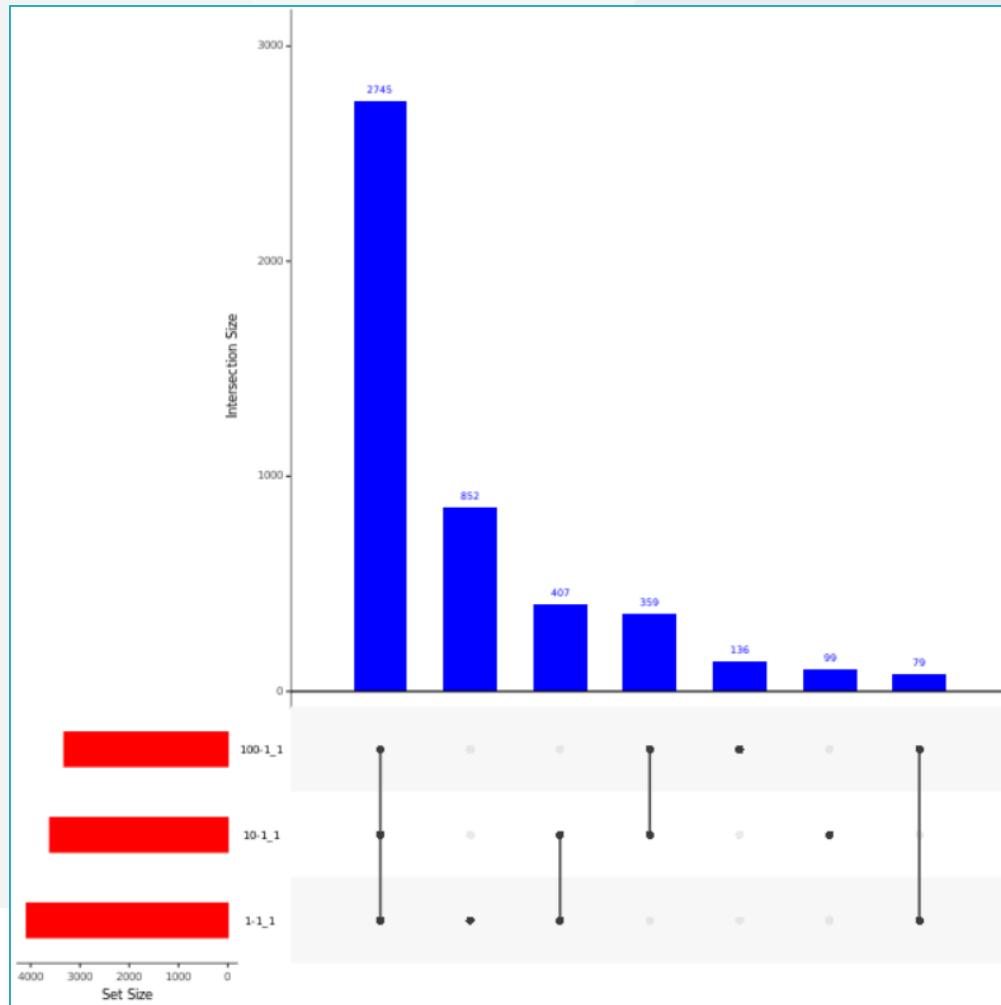


- Displays the **result of statistical test** (e.g. LIMMA, proDA)
- x axis: fold change (\log_2 transformed)
- y axis: p value (or adjusted p value) ($-\log_{10}$ transformed)
- Based on threshold ($\log FC = 1$; p value = 0.05) divides the proteins into groups:
 - Not significant
 - Significant
 - Upregulated and significant
 - Downregulated and significant
- We can detect **changes of protein expression between two samples**

Volcano plot



UpSet plot



- Alternative to Venn diagram for more samples
- Shows intersections between samples (what do and don't samples have in common)

UpSet plot based on frequency or degree Change

freq degree

Logical operators for comparisons in binary matrix creation Change

== != > < >= <=

Threshold value for binary matrix creation Change

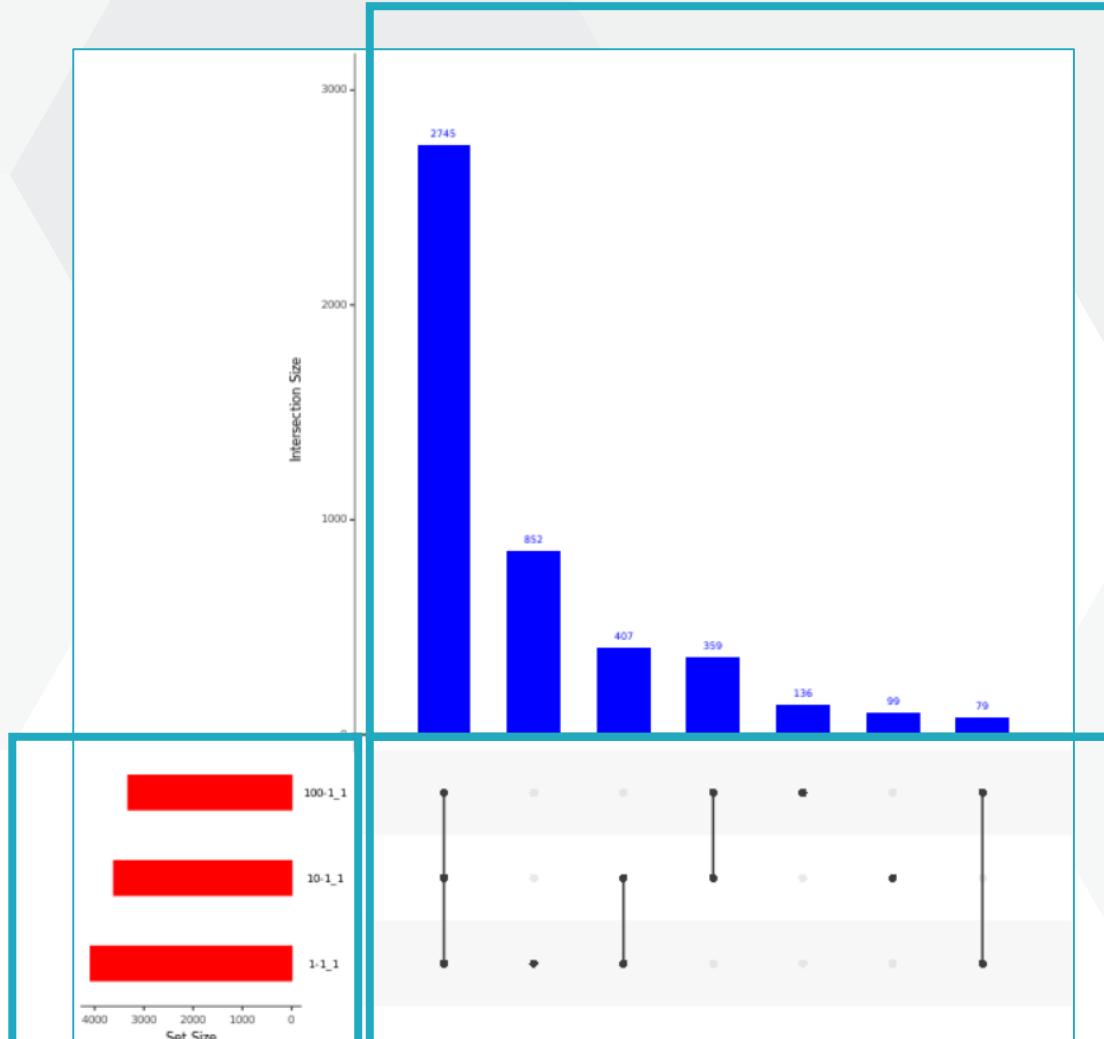
Number of sets Change

Number of intersections to be displayed Change

Show empty intersections? Change

on off

UpSet plot

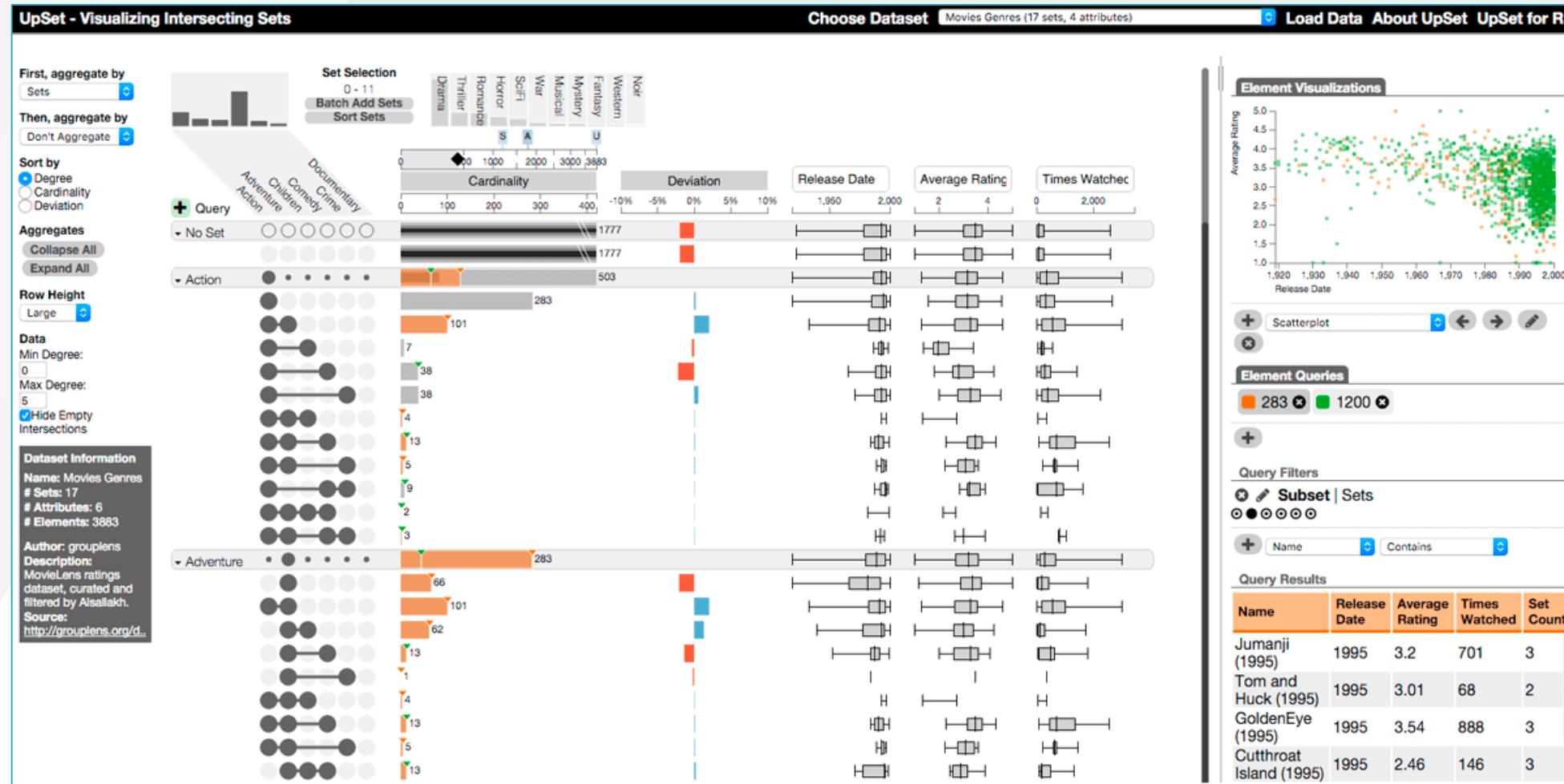


Size of particular sample

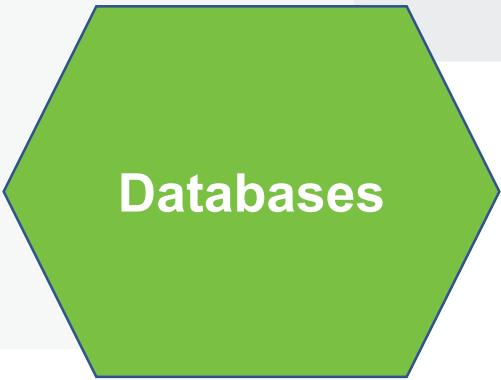
Intersections
(in which samples is particular protein present)

How many proteins are present in the particular intersection
(e.g. 2745 protein groups are present in all three samples; 852 only in 1-1 sample, etc.)

UpSet plot (interactive)



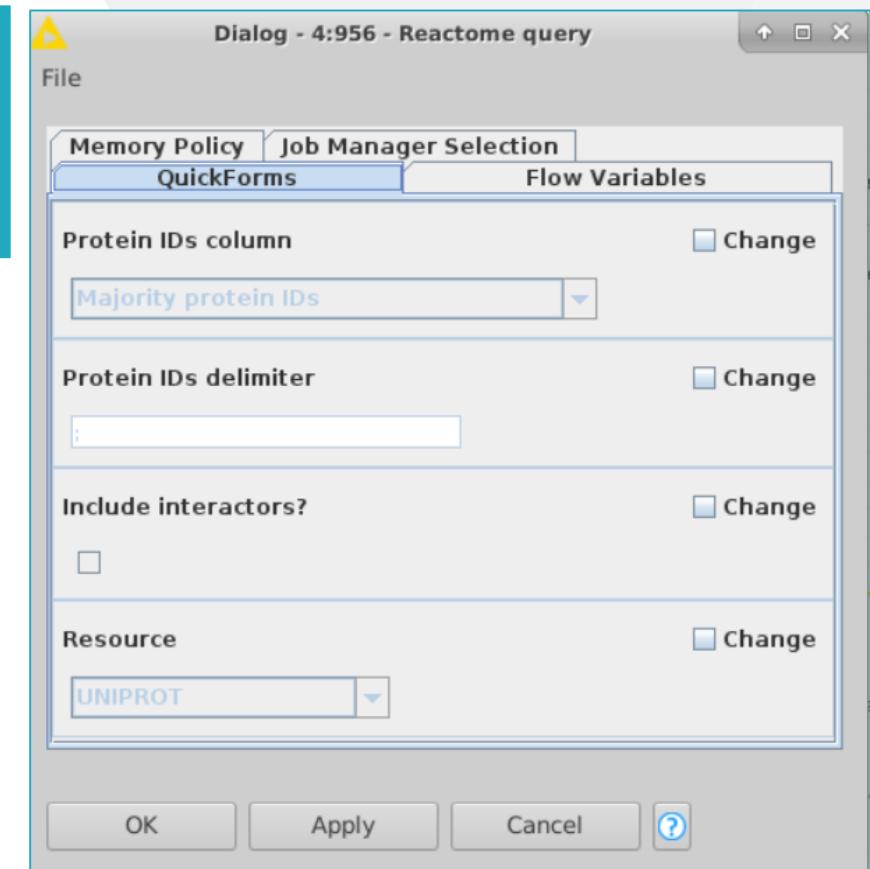
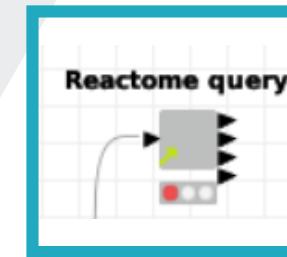
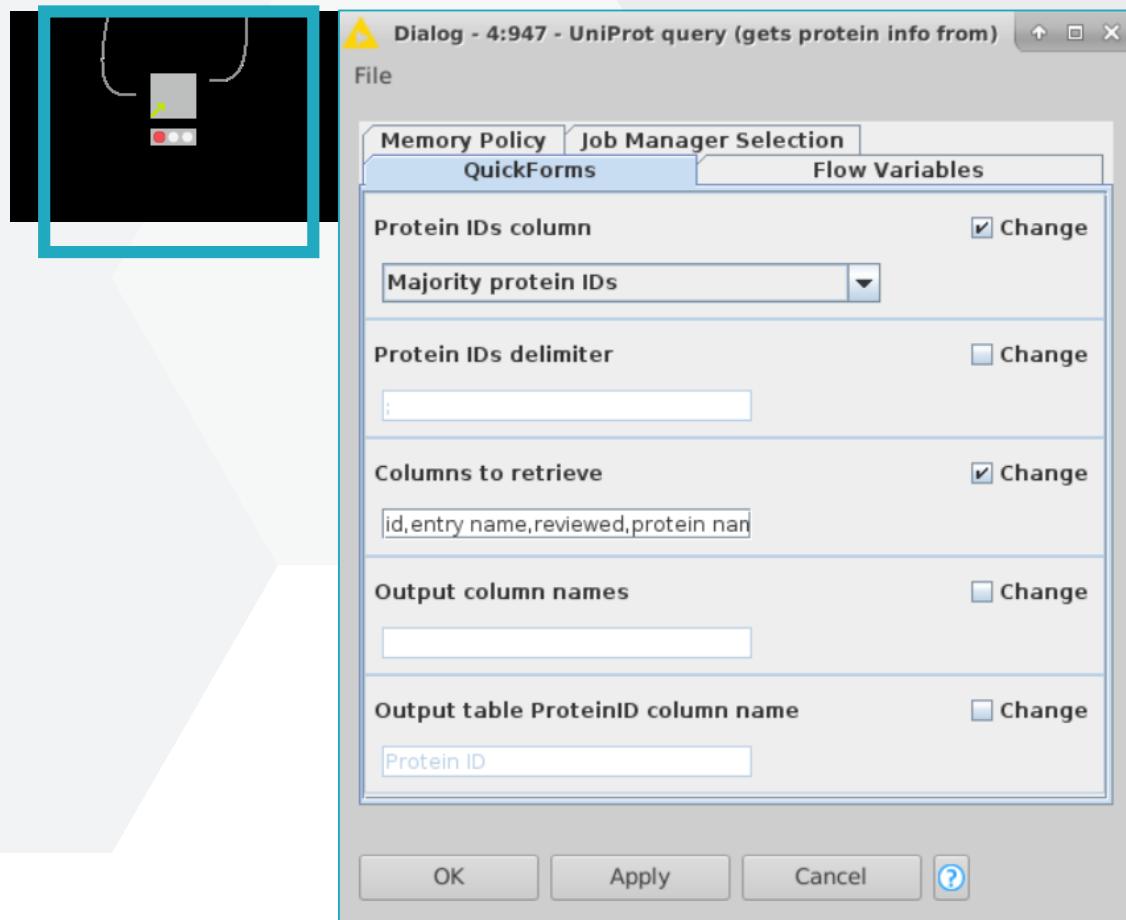
Based on <https://caleydo.org/tools/upset/>



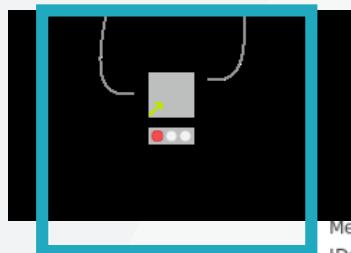
UniProt query

Reactome query

UniProt and Reactome query



UniProt and Reactome query



UniProt query

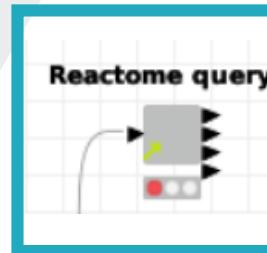
Metanode that makes api calls to UniProtKB to retrieve selected data based on protein IDs. Protein IDs have to be in UniProt format, e.g. P12345 (https://www.uniprot.org/help/accession_numbers). Multiple protein IDs separated by delimiter (e.g. semicolon) are supported.

Metanode retrieves query in tabular output form where one should select columns to be returned specifically. Please check UniProt help (http://www.uniprot.org/help/uniprotkb_column_names) for the full list of available columns and use "Columns to retrieve" setting to adjust accordingly. Not correctly spelled column names are ignored and not present in the output table.

Two tables are returned by the metanode - top one containing individual protein IDs and specified columns retrieved from UniProt and bottom table containing accessions for which empty query result was retrieved (e.g. missing accessions).

You can specify output column names (as strings separated by comma ","), UniProt column names will be used as default values.

=====



Reactome query

Metanode that makes api call to Reactome.org to retrieve the analysis results. So far only overrepresentation analysis is implemented using the protein IDs present in the selected protein IDs column. Protein IDs have to be in supported format (e.g. UniProtKB identified; see <https://reactome.org/dev/analysis> for their full list). Multiple protein IDs separated by delimiter (e.g. semicolon) are supported.

You can select whether to include also interactors during the analysis and what resource should be used during the analysis.

Metanode generates 4 outputs in total

1) pathways table

table of pathways returned by the analysis (similarly to downloading the pathway analysis results in csv format via web interface)

2) not mapped identifiers

identifiers that were not found/mapped during the Reactome analysis

3) summary table

summary information about the analysis including date of running, database version and more

4) complete analysis results

complete analysis results in json format for further data mining if needed

More information on the used API can be found here <https://reactome.org/AnalysisService/#/identifiers> and here <https://reactome.org/dev/analysis>.

=====

UniProt and Reactome query

| Row ID | Gene names | Organism | Gene ontology (GO) | Gene ontology (biological process) | Gene ontology (molecular function) | Gene ontology (cellular component) | Gene ontology IDs | Subcellular location [CC] | Protein families |
|--------|-------------------------------|----------------------------------------|-----------------------------------|--------------------------------------|----------------------------------------|----------------------------------------|-------------------|---------------------------|-----------------------------------|
| Row0 | PCMP-E22 At2g02750 T20F6.11 | Arabidopsis thaliana (Mouse-ear cress) | RNA modification [GO:0009451] | RNA modification [GO:0009451] | | | GO:0009451 | | PPR family, PCMP-E subfamily |
| Row1 | THUMPD3 | Homo sapiens (Human) | cytosol [GO:0005829]; nucleol... | tRNA methylation [GO:0030488] | RNA binding [GO:000372...] | | | | Methyltransferase superfamily |
| Row2 | PAD1 PRC6A At3g51260 F24M... | Arabidopsis thaliana (Mouse-ear cress) | cytoplasm [GO:0005737]; cyto... | proteasomal protein catabolic... | endopeptidase activity [GO:00... | cytoplasm [GO:000050...] | SUBCELL... | | Peptidase T1A family |
| Row3 | GFPT1 GFAT GFPT | Homo sapiens (Human) | cytosol [GO:0005829]; extrac... | circadian regulation of gene e... | carbohydrate binding [GO:0004... | cytosol [GO:000436...] | | | |
| Row4 | MDC1 KIAA0170 NFBD1 | Homo sapiens (Human) | chromosome [GO:0005694]; f... | double-strand break repair vi... | FHA domain binding [GO:0005... | chromosome [GO:000563...] | SUBCELL... | | |
| Row5 | DDX24 | Homo sapiens (Human) | membrane [GO:0016020]; nu... | RNA metabolic process [GO:00... | ATP binding [GO:000552...] | membrane [GO:000372...] | | | DEAD box helicase family, ... |
| Row6 | PRMT1 HMT2 HRMT1L2 IR1B4 | Homo sapiens (Human) | cytoplasm [GO:0005737]; cyto... | cell surface receptor signalin... | enzyme b... [GO:000170...] | cytoplasm [GO:000170...] | SUBCELL... | | Class I-like SAM-binding me... |
| Row7 | SLC1A5 ASCT2 M7V1 RDR RDRC | Homo sapiens (Human) | extracellular exosome [GO:00... | amino acid transport [GO:000... | amino acid: organic acid exch... | extracellular [GO:000161...] | SUBCELL... | | Dicarboxylate/amino acid:c... |
| Row8 | AXS2 At1g08200 T23G18.6 | Arabidopsis thaliana (Mouse-ear cress) | apoplast [GO:0048046]; cytop... | cell wall organization [GO:007... | catalytic activity [GO:000382...] | apoplast [GO:000382...] | SUBCELL... | | NAD(P)-dependent epimer... |
| Row9 | ATAD3B KIAA1273 TOB3 | Homo sapiens (Human) | ficolin-1-rich granule membra... | mitochondrion organization [G... | ATP binding [GO:000552...] | ficolin-1-rich granule membra... | SUBCELL... | | AAA ATPase family |
| Row10 | SLC3A2 MDU1 | Homo sapiens (Human) | amino acid transport complex... | amino acid transport [GO:000... | cadherin activity [GO:000372...] | amino acid: organic acid exch... | SUBCELL... | | SLC3A transporter family |
| Row11 | CLPP5 NCLPP1 NCLPP5 At1g02... | Arabidopsis thaliana (Mouse-ear cress) | chloroplast [GO:0009507]; chl... | protein quality control for mis... | ATPase b... [GO:000417...] | chloroplast [GO:000417...] | SUBCELL... | | Peptidase S14 family |
| Row12 | EIF4EBP2 | Homo sapiens (Human) | cytoplasm [GO:0005737]; pos... | cAMP-mediated signaling [GO:00... | eukaryotic translation initiation... | cytoplasm [GO:000573...] | | | EIF4E-binding protein family |
| Row13 | At5g45160 K18C1.4 | Arabidopsis thaliana (Mouse-ear cress) | cytoplasm [GO:0005737]; end... | endoplasmic reticulum membran... | GTPase activity [GO:000392...] | cytoplasm [GO:000392...] | SUBCELL... | | TRAFAC class dynamin-like .. |
| Row14 | HNRNPU C1orf199 HNRPU SAF... | Homo sapiens (Human) | catalytic step 2 spliceosome [... | adaptive thermogenesis [GO:00... | actin binding protein activity [... | catalytic activity [GO:000012...] | SUBCELL... | | |
| Row15 | ESD | Homo sapiens (Human) | cytoplasmic vesicle [GO:0031... | formaldehyde catabolic proce... | carboxylic acid ester hydrolase... | cytoplasm [GO:000578...] | SUBCELL... | | Esterase D family |
| Row16 | ZNF638 NP220 ZFML | Homo sapiens (Human) | cytoplasm [GO:0005737]; intr... | RNA splicing [GO:0008380] | double-stranded RNA binding [GO:... | cytoplasm [GO:000369...] | SUBCELL... | | |
| Row17 | POLD3 KIAA0039 | Homo sapiens (Human) | cytoplasm [GO:0005737]; delte... | DNA damage response, detection... | DNA-directed RNA polymerase acti... | cytoplasm [GO:000072...] | SUBCELL... | | |
| Row18 | NTR2 NTRA At2g17420 F5J6.18 | Arabidopsis thaliana (Mouse-ear cress) | mitochondrial matrix [GO:0000... | removal of superoxide radical... | thioredoxin: oxidoreductase ac... | mitochondrion [GO:000479...] | SUBCELL... | | Class-II pyridine nucleotide-t... |
| Row19 | ACC1 EMB22 GK PAS3 At1g36... | Arabidopsis thaliana (Mouse-ear cress) | cytosol [GO:0005829]; nucleo... | embryo development ending in ... | acetyl-CoA: organic acid exch... | cytosol [GO:000398...] | SUBCELL... | | |
| Row20 | FPS1 At5g47770 MCA23.9 | Arabidopsis thaliana (Mouse-ear cress) | cytoplasm [GO:0005737]; cyto... | cholesterol biosynthetic proce... | dimethylallyl pyrophosphate: org... | cytosol [GO:000416...] | SUBCELL... | | FPP/GGPP synthase family |
| Row21 | CHLD ALB1 PDE166 At1g0852... | Arabidopsis thaliana (Mouse-ear cress) | chloroplast [GO:0009507]; chl... | chlorophyll biosynthetic proce... | ATP binding [GO:000552...] | chloroplast [GO:000552...] | SUBCELL... | | Mg-chelatase subunits D/I f... |
| Row22 | hCG 2014768 | Homo sapiens (Human) | | | | | | | |
| Row23 | CCT3 CCTG TRIC5 | Homo sapiens (Human) | cell body [GO:0044297]; chap... | binding of sperm to zona pellu... | ATP binding [GO:000219...] | cell body [GO:000219...] | SUBCELL... | | TCP-1 chaperonin family |
| Row24 | PRC1 | Homo sapiens (Human) | contractile ring [GO:0070938];... | cell division [GO:0051301]; mi... | identical sequence DNA binding [... | contractile ring [GO:000002...] | SUBCELL... | | MAP65/ASE1 family |
| Row25 | CAPZA1 | Homo sapiens (Human) | actin cortical patch [GO:0030... | actin cytoskeleton organizatio... | actin binding protein activity [... | actin cortex [GO:000377...] | SUBCELL... | | F-actin-capping protein alp... |
| Row26 | SCAMP1 SCAMP | Homo sapiens (Human) | clathrin-coated vesicle [GO:00... | neutrophil degranulation [GO:00... | clathrin-mediated endocytosis [G... | clathrin-coated vesicle [GO:000013...] | SUBCELL... | | SCAMP family |
| Row27 | RPP1C At5g47700 MCA23.2 | Arabidopsis thaliana (Mouse-ear cress) | cytosol [GO:0005829]; cytosol... | cytoplasmic translation [GO:00... | protein kinase activity [GO:000218...] | cytosol [GO:000218...] | | | Eukaryotic ribosomal protein |
| Row28 | BGLU42 At5g36890 MLF18.1 | Arabidopsis thaliana (Mouse-ear cress) | cytosol [GO:0005829]; beta-gl... | cellulose catabolic process [G... | beta-glucosidase activity [GO:00... | cytosol [GO:000582...] | | | Glycosyl hydrolase 1 family |
| Row29 | At5g25754 At5g25757 | Arabidopsis thaliana (Mouse-ear cress) | cytosol [GO:0005829]; eukary... | formation of cytoplasmic trans... | mRNA binding [GO:000173...] | cytosol [GO:000173...] | SUBCELL... | | EIF-3 subunit L family |
| Row30 | NUMA1 NMP22 NUMA | Homo sapiens (Human) | cell cortex [GO:0005938]; cell... | anastral spindle assembly [G... | disorderly cell cortex [GO:000013...] | cell cortex [GO:000013...] | SUBCELL... | | |
| Row31 | UBE2D3 UBC5C UBCH5C | Homo sapiens (Human) | cytosol [GO:0005829]; endoso... | apoptotic process [GO:00069... | ATP binding [GO:000012...] | cytosol [GO:000012...] | SUBCELL... | | Ubiquitin-conjugating enzyme |
| Row32 | | Homo sapiens (Human) | nucleus [GO:0005634] | | | nucleus [GO:0005634] | | | |
| Row33 | EEF2 EF2 | Homo sapiens (Human) | aggresome [GO:0016235]; cyt... | aging [GO:0007568]; cellular r... | 5S rRNA binding [GO:000203...] | aggresome [GO:000203...] | SUBCELL... | | TRAFAC class translation fa... |
| Row34 | UAP1 SPAG2 | Homo sapiens (Human) | cytosol [GO:0005829]; nucleo... | UDP-N-acetylglucosamine biosy... | carbohydrate binding [GO:000397...] | cytosol [GO:000397...] | SUBCELL... | | UDPGP type 1 family |
| Row35 | PSMC4 MIP224 TBP7 | Homo sapiens (Human) | cytosol [GO:0005829]; cytosol... | anaphase-promoting complex [G... | ATP binding [GO:000016...] | cytosol [GO:000016...] | SUBCELL... | | AAA ATPase family |
| Row36 | MARK3 CTAK1 EMK2 | Homo sapiens (Human) | cytoplasm [GO:0005737]; cyto... | intracellular signal transducti... | ATP binding [GO:000016...] | cytosol [GO:000016...] | SUBCELL... | | Protein kinase superfamily.. |
| Row37 | SRRM2 KIAA0324 SRL300 SRM... | Homo sapiens (Human) | Cajal body [GO:0015030]; cat... | mRNA splicing, via spliceosom... | C2H2 zinc finger [GO:000039...] | Cajal body [GO:000039...] | SUBCELL... | | CWC21 family |
| Row38 | At1g27892 At5g20122 | Arabidopsis thaliana (Mouse-ear cress) | apoplast [GO:0048046]; cell w... | carbohydrate binding [GO:0000561...] | carbohydrate binding [GO:0000561...] | apoplast [GO:0000561...] | SUBCELL... | | |

UniProt and Reactome query

| Table "default" - Rows: 1733 Spec - Columns: 15 Properties Flow Variables | | | | | | | | | | | | | | |
|---------------------------------------------------------------------------|--------------------|-----------------------|-----------------|-----------------|----------------|-----------------|--------------|------------------|------------------|-----------------|--------------------|--------------|--------------------------------------------------|--|
| Row ID | Pathway identifier | Pathway name | #Entities found | #Entities total | Entities ratio | Entities pValue | Entities FDR | #Reactions found | #Reactions total | Reactions ratio | Species identifier | Species name | Submitted entities found | |
| Row0 | R-HSA-8953854 | Metabolism of RNA | 515 | 675 | 0.06 | 0 | 0 | 150 | 187 | 0.016 | 9606 | Homo sapiens | Q92979;075940;P67809;Q9NRW3;Q06265;Q96PZ0;0002 | |
| Row1 | R-HSA-72766 | Translation | 254 | 294 | 0.026 | 0 | 0 | 92 | 99 | 0.008 | 9606 | Homo sapiens | 075821;075822;P62917;Q9Y5M8;Q8IXM3;Q04637;Q9NX2 | |
| Row2 | R-HSA-72203 | Processing of Ca... | 206 | 245 | 0.022 | 0 | 0 | 32 | 32 | 0.003 | 9606 | Homo sapiens | Q16629;Q7RTV0;Q9BY77;075940;P67809;P35658;P1398 | |
| Row3 | R-HSA-72163 | mRNA Splicing - ... | 159 | 180 | 0.016 | 0 | 0 | 9 | 9 | 0.001 | 9606 | Homo sapiens | Q43395;Q16629;Q7RTV0;075940;P67809;P52272;Q86V8 | |
| Row4 | R-HSA-8868773 | rRNA processing ... | 167 | 193 | 0.017 | 0 | 0 | 14 | 15 | 0.001 | 9606 | Homo sapiens | Q92979;P23396;P62917;Q12788;P62081;000541;Q0626 | |
| Row5 | R-HSA-72312 | rRNA processing | 173 | 203 | 0.018 | 0 | 0 | 17 | 21 | 0.002 | 9606 | Homo sapiens | Q92979;P23396;Q7L0Y3;P62917;Q12788;P62081;00054 | |
| Row6 | R-HSA-6791226 | Major pathway of... | 158 | 183 | 0.016 | 0 | 0 | 7 | 7 | 0.001 | 9606 | Homo sapiens | Q92979;P23396;P62917;Q12788;P62081;000541;Q0626 | |
| Row7 | R-HSA-72172 | mRNA Splicing | 159 | 188 | 0.017 | 0 | 0 | 14 | 14 | 0.001 | 9606 | Homo sapiens | Q43395;Q16629;Q7RTV0;075940;P67809;P52272;Q86V8 | |
| Row8 | R-HSA-72613 | Eukaryotic Transl... | 112 | 120 | 0.011 | 0 | 0 | 21 | 21 | 0.002 | 9606 | Homo sapiens | P23396;075821;075822;P62917;P62081;P18124;00030 | |
| Row9 | R-HSA-72737 | Cap-dependent T... | 112 | 120 | 0.011 | 0 | 0 | 18 | 18 | 0.002 | 9606 | Homo sapiens | P23396;075821;075822;P62917;P62081;P18124;00030 | |
| Row10 | R-HSA-72706 | GTP hydrolysis a... | 106 | 113 | 0.01 | 0 | 0 | 3 | 3 | 0 | 9606 | Homo sapiens | P23396;075821;075822;P62917;P62081;P18124;00030 | |
| Row11 | R-HSA-156827 | L13a-mediated tr... | 105 | 112 | 0.01 | 0 | 0 | 3 | 3 | 0 | 9606 | Homo sapiens | P23396;075821;075822;P62917;P62081;P18124;00030 | |
| Row12 | R-HSA-168254 | Influenza Infection | 145 | 171 | 0.015 | 0 | 0 | 21 | 58 | 0.005 | 9606 | Homo sapiens | P07900;P23396;P62917;P62081;P18124;P35658;P1398 | |
| Row13 | R-HSA-168255 | Influenza Life Cycle | 137 | 160 | 0.014 | 0 | 0 | 15 | 50 | 0.004 | 9606 | Homo sapiens | P07900;P23396;P62917;P62081;P18124;P35658;P1398 | |
| Row14 | R-HSA-1799339 | SRP-dependent c... | 104 | 113 | 0.01 | 0 | 0 | 5 | 5 | 0 | 9606 | Homo sapiens | P23396;P62917;P62081;Q9Y5M8;P18124;P62750;P4791 | |
| Row15 | R-HSA-72689 | Formation of a p... | 95 | 102 | 0.009 | 0 | 0 | 2 | 2 | 0 | 9606 | Homo sapiens | Q9Y3U8;P23396;075821;Q71UM5;075822;P62917;P620 | |
| Row16 | R-HSA-9010553 | Regulation of ex... | 142 | 172 | 0.015 | 0 | 0 | 6 | 19 | 0.002 | 9606 | Homo sapiens | P23396;P62917;P62081;P18124;Q04637;P62750;P4791 | |
| Row17 | R-HSA-168273 | Influenza Viral RN... | 128 | 151 | 0.014 | 0 | 0 | 6 | 13 | 0.001 | 9606 | Homo sapiens | P07900;P23396;P62917;P62081;P18124;P35658;P1398 | |
| Row18 | R-HSA-975956 | Nonsense Media... | 89 | 96 | 0.009 | 0 | 0 | 1 | 1 | 0 | 9606 | Homo sapiens | Q9Y3U8;P23396;Q71UM5;P62917;P62081;P62280;P181 | |
| Row19 | R-HSA-156842 | Eukaryotic Transl... | 87 | 95 | 0.008 | 0 | 0 | 9 | 9 | 0.001 | 9606 | Homo sapiens | Q9Y3U8;P23396;Q71UM5;Q5VTE0;P62917;P29692;P620 | |
| Row20 | R-HSA-162909 | Host Interactions | 120 | 144 | 0.013 | 0 | 0 | 46 | 54 | 0.005 | 9606 | Homo sapiens | P35658;Q9HC16;P62195;060563;P62877;Q9BW27;P632 | |
| Row21 | R-HSA-72764 | Eukaryotic Transl... | 86 | 94 | 0.008 | 0 | 0 | 5 | 5 | 0 | 9606 | Homo sapiens | Q9Y3U8;P23396;Q71UM5;P62917;P62081;P62280;P181 | |
| Row22 | R-HSA-156902 | Peptide chain el... | 83 | 90 | 0.008 | 0 | 0 | 5 | 5 | 0 | 9606 | Homo sapiens | Q9Y3U8;P23396;Q71UM5;P62917;P62081;P62280;P181 | |
| Row23 | R-HSA-2408522 | Selenoamino aci... | 102 | 118 | 0.011 | 0 | 0 | 18 | 23 | 0.002 | 9606 | Homo sapiens | P23396;P00390;P62917;P62081;P18124;P62750;P4791 | |
| Row24 | R-HSA-927802 | Nonsense-Media... | 100 | 117 | 0.01 | 0 | 0 | 6 | 6 | 0.001 | 9606 | Homo sapiens | P23396;P62917;P62081;P18124;Q04637;P62750;P4791 | |
| Row25 | R-HSA-975957 | Nonsense Media... | 100 | 117 | 0.01 | 0 | 0 | 5 | 5 | 0 | 9606 | Homo sapiens | P23396;P62917;P62081;P18124;Q04637;P62750;P4791 | |
| Row26 | R-HSA-2408557 | Selenocysteine s... | 83 | 94 | 0.008 | 0 | 0 | 4 | 7 | 0.001 | 9606 | Homo sapiens | Q9Y3U8;P23396;Q71UM5;P62917;P62081;P62280;P181 | |
| Row27 | R-HSA-450531 | Regulation of mR... | 78 | 88 | 0.008 | 0 | 0 | 26 | 26 | 0.002 | 9606 | Homo sapiens | P0CG47;P0DMV8;P51665;Q06265;P35658;Q07352;POCG | |
| Row28 | R-HSA-72649 | Translation initia... | 57 | 59 | 0.005 | 0 | 0.001 | 2 | 2 | 0 | 9606 | Homo sapiens | P23396;075821;Q71UM5;075822;P62081;P62280;0003 | |
| Row29 | R-HSA-72702 | Ribosomal scann... | 57 | 59 | 0.005 | 0 | 0.001 | 2 | 2 | 0 | 9606 | Homo sapiens | P23396;075821;Q71UM5;075822;P62081;P62280;0003 | |
| Row30 | R-HSA-376176 | Signaling by ROB... | 160 | 218 | 0.019 | 0 | 0.001 | 21 | 59 | 0.005 | 9606 | Homo sapiens | P23396;P62917;P62081;P12931;P18124;P13861;Q0463 | |
| Row31 | R-HSA-72662 | Activation of the ... | 57 | 60 | 0.005 | 0 | 0.001 | 6 | 6 | 0.001 | 9606 | Homo sapiens | P23396;075821;Q71UM5;075822;P62081;P62280;0003 | |
| Row32 | R-HSA-450408 | AUF1 (hnRNP D0)... | 53 | 56 | 0.005 | 0 | 0.002 | 4 | 4 | 0 | 9606 | Homo sapiens | P0CG47;P0DMV8;P51665;POCG48;Q04637;P62195;Q925 | |
| Row33 | R-HSA-68875 | Mitotic Prophase | 95 | 119 | 0.011 | 0 | 0.002 | 32 | 34 | 0.003 | 9606 | Homo sapiens | P06899;P28482;Q9H8Y8;P14635;P35658;Q71UI9;Q9HO | |
| Row34 | R-HSA-72695 | Formation of the ... | 50 | 52 | 0.005 | 0 | 0.002 | 3 | 3 | 0 | 9606 | Homo sapiens | P23396;075821;Q71UM5;075822;P62081;P62280;0003 | |
| Row35 | R-HSA-192823 | Viral mRNA Trans... | 83 | 101 | 0.009 | 0 | 0.002 | 2 | 2 | 0 | 9606 | Homo sapiens | Q9Y3U8;P23396;Q71UM5;P62917;P62081;P62280;P181 | |
| Row36 | R-HSA-5578749 | Transcriptional r... | 69 | 80 | 0.007 | 0 | 0.002 | 3 | 5 | 0 | 9606 | Homo sapiens | P06899;P35658;Q71UI9;P52434;P52435;Q99567;P5705 | |
| Row37 | R-HSA-6790901 | rRNA modificatio... | 55 | 60 | 0.005 | 0 | 0.003 | 7 | 8 | 0.001 | 9606 | Homo sapiens | Q9H0A0;P35658;Q71UI9;P52434;P52435;Q99567;P5705 | |
| Row38 | R-HSA-150221 | Transport of Mat... | 45 | 46 | 0.004 | 0 | 0.002 | 2 | 2 | 0 | 9606 | Homo sapiens | P52208;P8CG34;Q96F8;Q10PRTV1;P0TEM1;P57740;Q1272 | |

Manipulation

Binary matrix calculator

FASTA to table

Table to FASTA

Values lookup

Substring extraction

Binary matrix calculator

- Calculates the binary matrix according to some threshold
- Useful for “qualitative changes mapping”
- Calculates also sample type summary
- Useful for questions such as in How many replicates is my protein present?

Binary matrix calculator

This metanode calculates the binary matrix from selected columns based on the threshold and operator and appends transformed columns to the end of the dataframe with user-defined suffix.

On top of the binary matrix columns there are also summary columns added if requested. Summary columns can be provided on two levels:

A) the sample type level - sums binary matrix values on the sample type level, i.e. sums all binary matrix columns for individual sample types based on the provided design.

B) the biological replicate level - sums binary matrix values for individual sample type and biological replicate level, i.e. sums all binary matrix columns having identical sample type and biological repl. identifier.

Sample types in experimental design and biological replicate numbers should be provided as comma separated list of values in the same order is the selected set of columns used for the binary matrix calculation.

You can specify suffix for the binary matrix and also prefix and suffix for the summary columns. Summary column names use also design (both summary column types) and biological replicate (biol. repl summary only) inputs to get sample type and biological replicate names, respectively.

=====

The screenshot shows the configuration interface for the Binary matrix calculator. At the top, there are tabs: QuickForms, Flow Variables, Memory Policy, and Job Manager Selection. The QuickForms tab is active.

Below the tabs, there are two radio buttons: "Enforce exclusion" (unchecked) and "Enforce inclusion" (checked). To the right of these buttons are three small arrows: a double arrow pointing up, a left arrow, and a double arrow pointing down.

Next, there is a section titled "Logical operators for comparisons in binary matrix creation" with the following options: $\circ ==$, $\circ !=$, $\circ >$, $\circ <$, $\circ >=$, and $\circ <=$. To the right of this section is a "Change" button.

Then, there is a section titled "Threshold value for binary matrix creation" with a text input field containing "0.1". To the right of this section is a "Change" button.

Following that is a section titled "Calculate sample type summary?" with a checkbox. To the right of this section is a "Change" button.

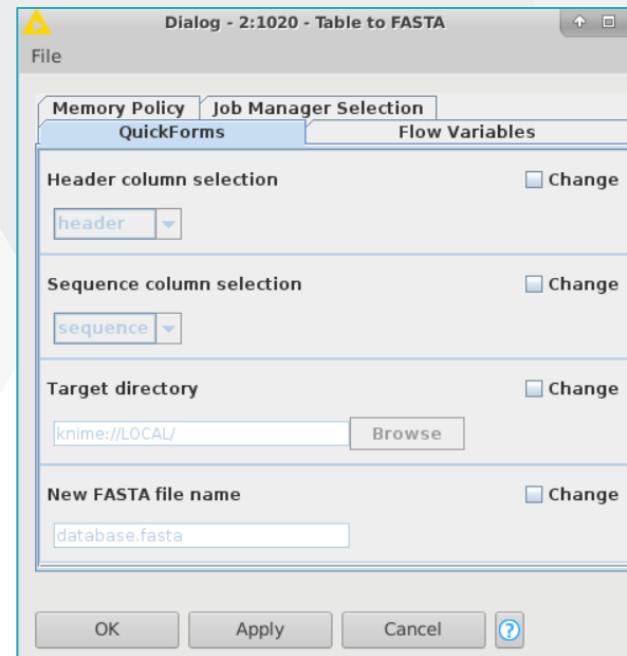
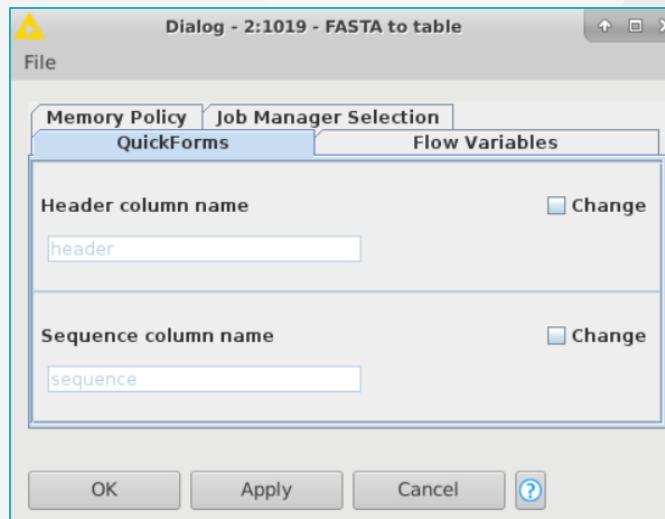
Next is a section titled "experimental design" with a dropdown menu showing "none". To the right of this section is a "Change" button.

Then, there is a section titled "Calculate biological replicate summary?" with a checkbox. To the right of this section is a "Change" button.

Finally, there is a section titled "biological replicates" with a dropdown menu showing "none". To the right of this section is a "Change" button.



Table to FASTA / FASTA to table



FASTA to table

Metanode for FASTA files conversion to tabular format.

Single table is the metanode output with sequence header in its 1st column and sequence without spaces in the second column. You can specify the names of the header and sequence columns.

There is currently limitation in size of the database related to the RAM usage.

=====

Table to FASTA

Metanode for creation of FASTA file based on the two specified columns containing header and sequence.

Sequences strings are wrapped to have 60 letters on single sequence line.

There might be limitation in the processable tabular database size because it is read into RAM and on some systems the RAM may not be sufficient.

=====

- Conversion of FASTA files to table or vice versa

Values lookup

- Looks up values from our table in the second table
- Adds values from the bottom table to the top input table
- In case there are more values, are separated then by a delimiter
- Split and summarize metanode is an useful follow-up node

Values lookup

Metanode for data lookup in selected columns of input table 2 (bottom input) using IDs from input table 1 (top input).

IDs column in input table 1 can contain multiple values, delimiter should be specified if this is the case.

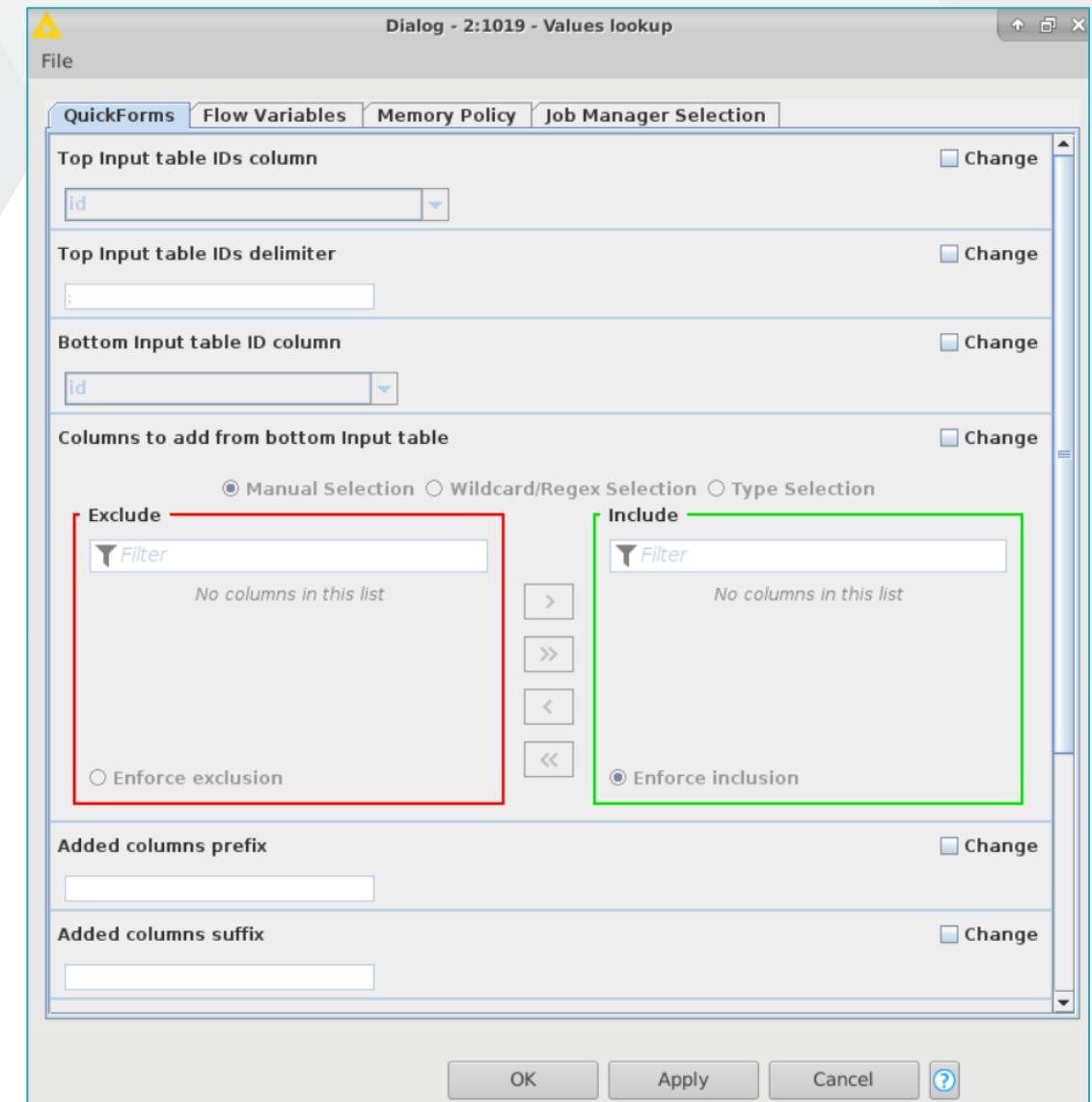
The metanode can return original input table 1 with additional columns or only input table 1 IDs column with additional columns with looked up values.

Name of the added columns can be modified using specified prefix/suffix.

IDs in both tables does not have to be of the same data type (e.g. string), they are converted to string internally if needed.

Please note that the order of looked up values separated by the delimiter will not be the same as the order of IDs in the top table IDs column if you select to remove empty strings or remove duplicate values!

=====



Substring extraction

- Allows processing of columns where values are separated by a delimiter into two columns?

Substring extraction

The metanode process the "column to process" this way ("sp|P68133|ACTS_HUMAN;sp|P68032|ACTC_HUMAN;sp|P63267|ACTH_HUMAN" is used as an example below)

1) splits the content of the column to process based on the specified delimiter so each part can be processed separately
(example string will be splitted using ";" into 3 texts: "sp|P68133|ACTS_HUMAN", "sp|P68032|ACTC_HUMAN" and "sp|P63267|ACTH_HUMAN")

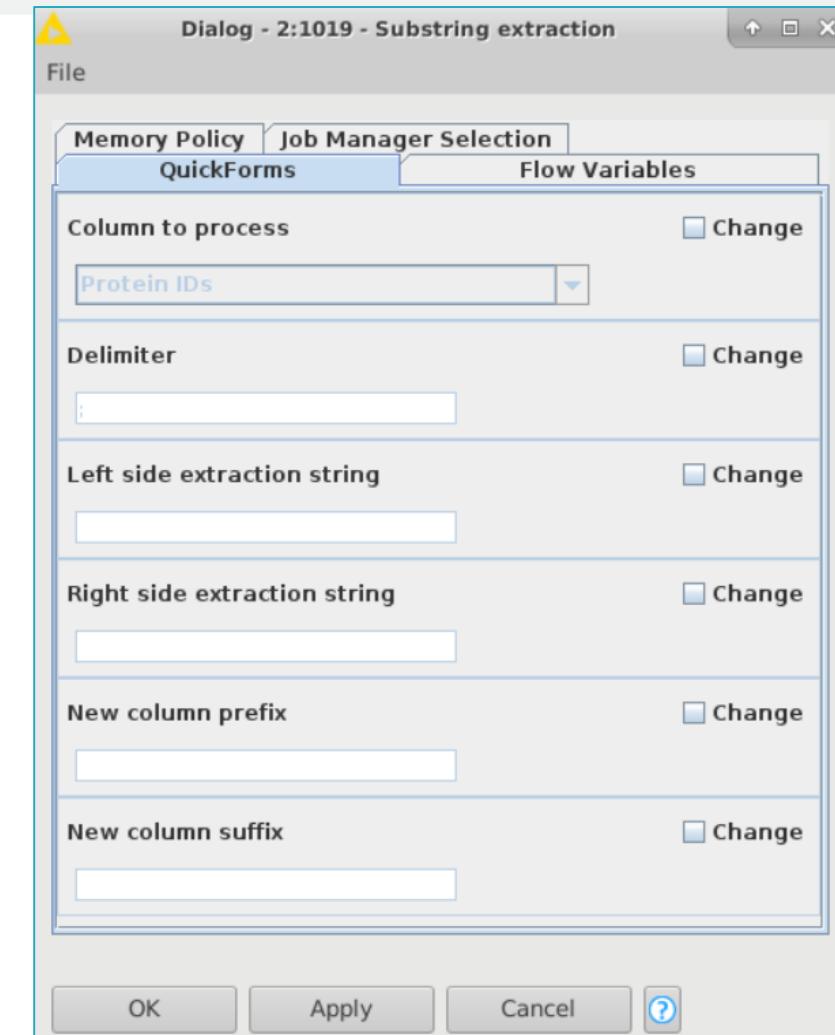
2) removes left side of the string up to the specified leftside extraction string - operation is done on each part of the string got using the delimiter in the step 1:
(this would result in the following texts in our example when "|" would be the lefside extraction string: "P68133|ACTS_HUMAN", "P68032|ACTC_HUMAN" and "P63267|ACTH_HUMAN"; i.e. 3 first letters would be removed)

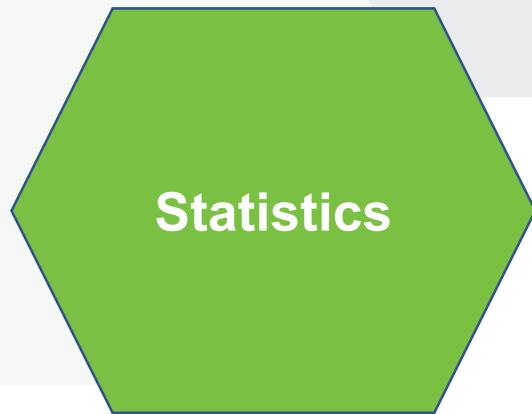
3) removes right side of the string from the specified rightside extraction string - operation is done on each part of the string got using the delimiter in the step 1:
(this would result in the following texts in our example when "|" would be the lefside extraction string: "P68133", "P68032" and "P63267"; i.e. everything after "|" would be removed)

4) combines text parts got in the step 1 into one text again using the same delimiter:
(this would result in the following text in our example: "P68133;P68032;P63267")

Name of the new column can be modified using specified prefix/suffix. If you will not change the column name, it will be replaced!

=====





Column descriptive statistics

LIMMA test

proDA?

Column descriptive statistics

- Compute particular statistics from the selected columns in the table

Columns descriptive statistics

Metanode for computation of selected descriptive statistics for selected columns.

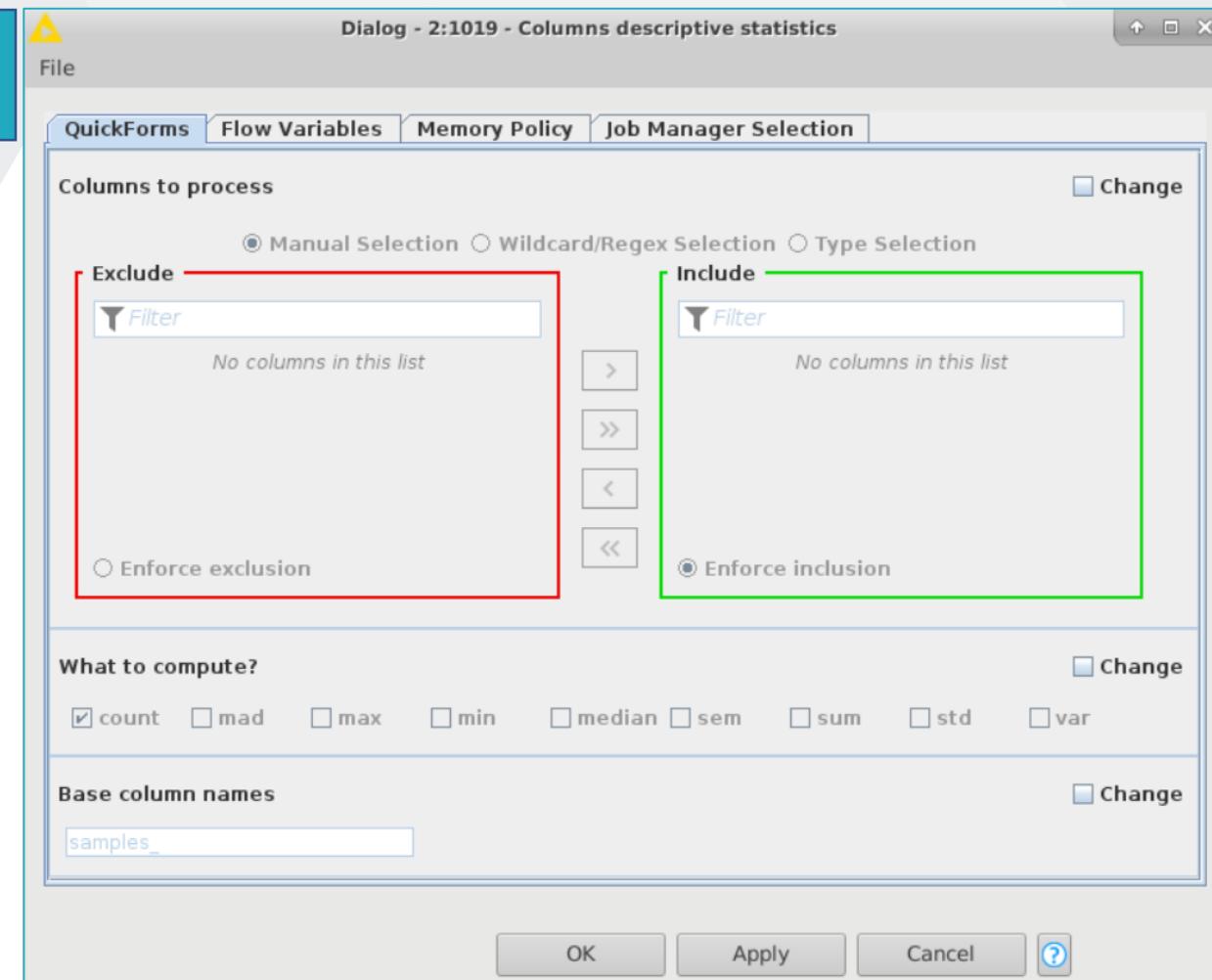
Selected descriptive statistics are:

- count - count of numerical values
- mad - median absolute deviation
- max - maximum
- min - minimum
- median - median
- sem - standard error of the mean
- sum - sum
- std - standard deviation
- var - variance

Statistics are calculated using pandas DataFrame statistics, more details can be found e.g. here <https://pandas.pydata.org/pandas-docs/stable/reference/frame.html#computations-descriptive-stats>

You can select what descriptive statistics should be calculated and common base name for newly added columns. New columns will be named as "base name" "statistics name" (i.e. without any space or delimiter).

=====



LIMMA test

- Statistical test for differential expression

LIMMA test

LIMMA (Linear Models for Microarray Data) test for selected experimental design usecases. Expected usecase is 'single-channel' design using lmFit with least square method fitting. Only selected scenarios of experimental designs are implemented as well, namely: several groups, paired samples, one samples block design and combination of paired and single block design (yet to be fully implemented in this version).

See e.g. LIMMA useguide for more details on selected settings explanation (<https://www.bioconductor.org/packages/release/bioc/vignettes/limma/inst/doc/usersguide.pdf>).

Missing values are treated row-wise, e.g. protein group wise. In case there is no non-missing value in any sample type name (i.e. in A, B or C columns), the whole row is excluded from the statistical evaluation during lmFit step!

Several more or less connected inputs are needed for proper results:

- data columns:

data columns to be used during the LIMMA test (e.g. columns A1,A2,A3,B1,B2,B3,C1,C2,C3); already normalized and log (any base relevant for further processing) transformed data are expected!

- experimental design:

should be provided in the form of sample type names in the same order as data columns without replicate number specification (e.g. "A,A,A,B,B,B,C,C,C")

- samples pairs:

information about the pairs of samples (e.g. different tissues coming from a single patient) where the pairs of samples have to be specified using identical identifier, e.g. "1,2,3,1,2,3" for 3 pairs of samples 1, 2 and 3

The screenshot shows the 'LIMMA test' configuration interface. At the top, there are tabs for QuickForms, Flow Variables, Memory Policy, and Job Manager Selection. Below the tabs, the 'LIMMA test name' is set to 'LIMMA'. The 'Columns to process' section includes 'Exclude' and 'Include' filters, both currently empty. Under 'Limma design', the sample types are listed as 'A,A,A,B,B,B,C,C,C'. The 'Samples pairs' field contains 'none'. The 'Samples blocks' field also contains 'none'. In the 'Comparisons (contrasts)' field, 'B-A, C-A' is listed. The 'Treat comparisons separately?' checkbox is checked. The 'p-value adjustment method' dropdown is set to 'Benjamini & Hochberg'. The 'Remove not used columns?' checkbox is unchecked.