



Central European Institute of Technology
BRNO | CZECH REPUBLIC

MUNI

2019-11-20

Processing of proteomics data in **KNIME**



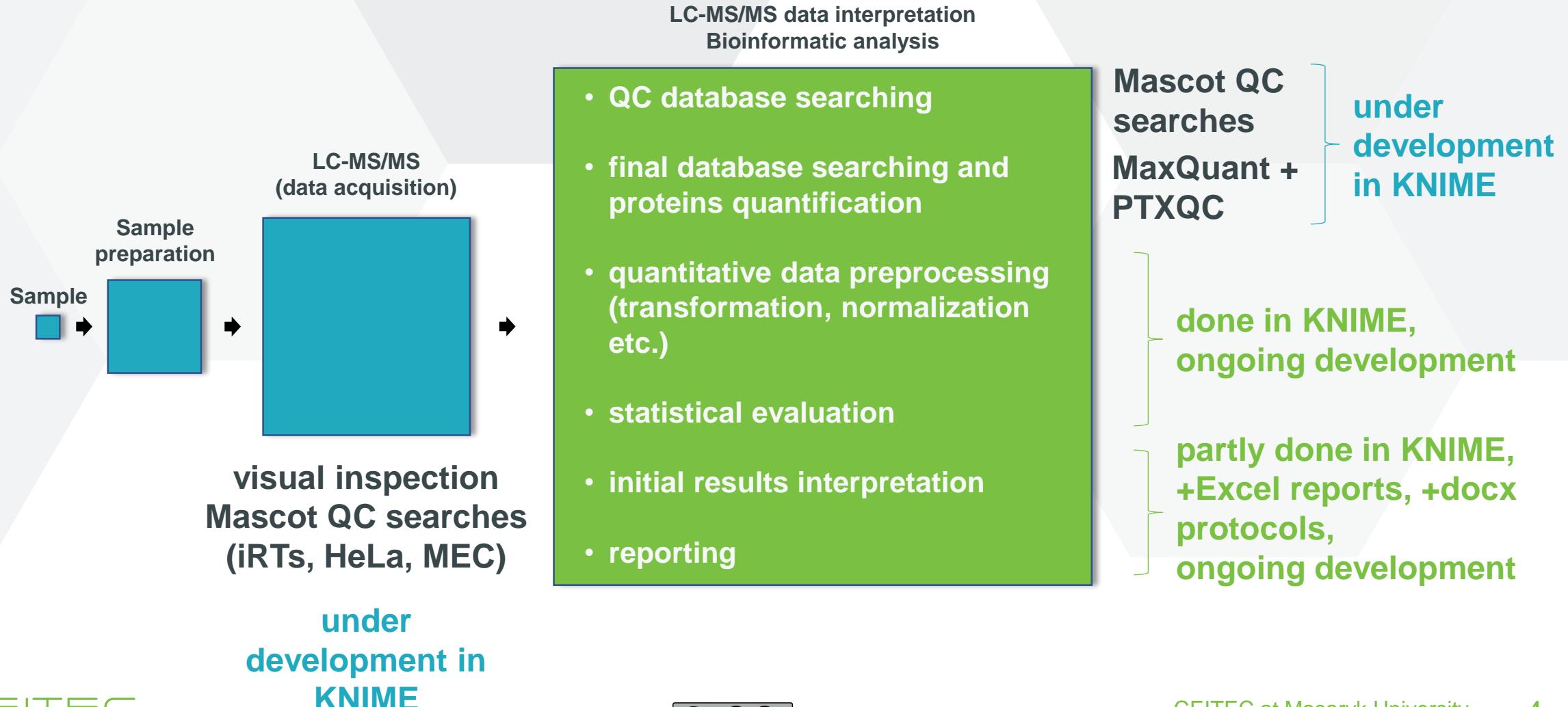
Workshop outline

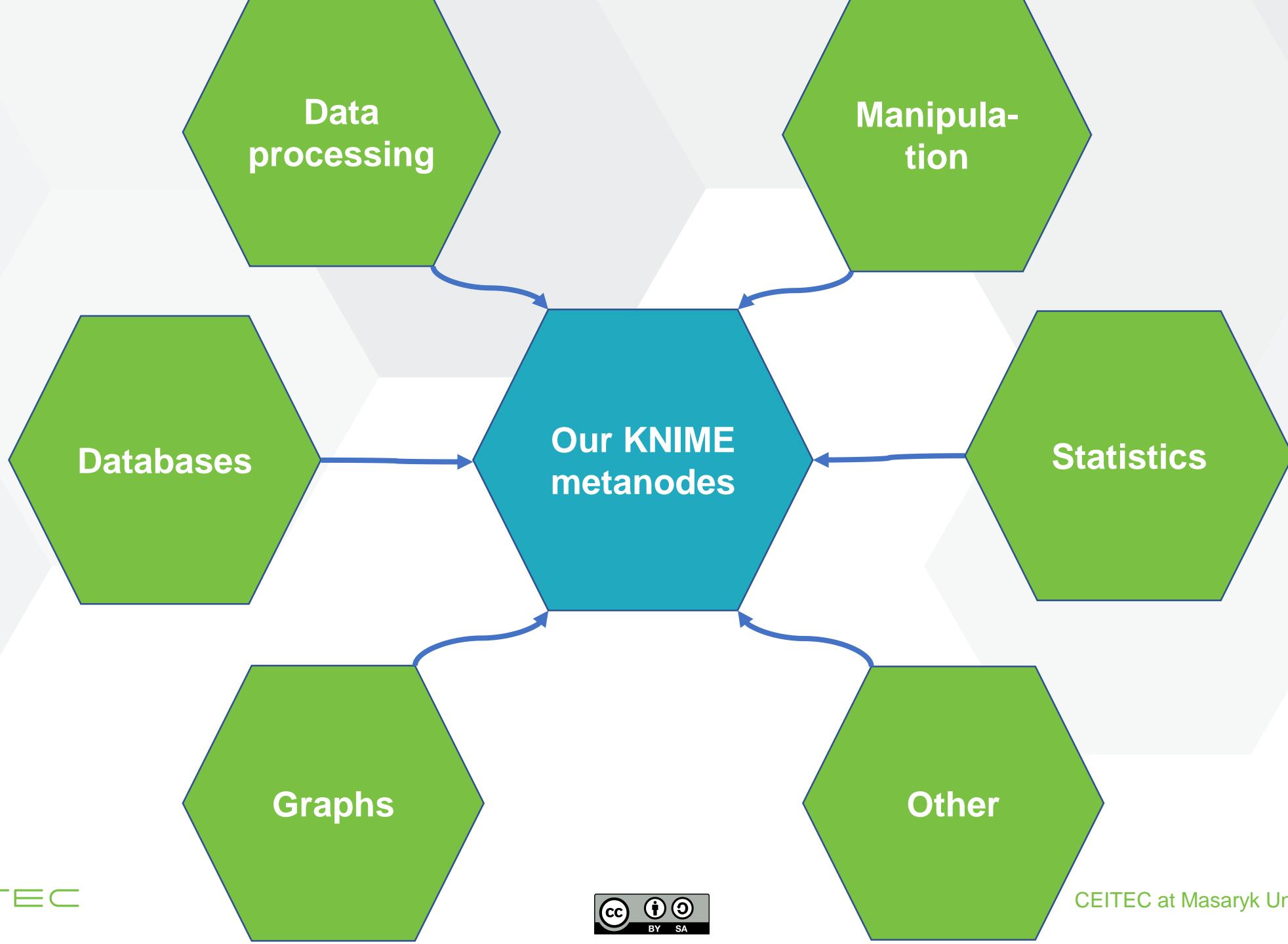
- morning session – theoretical part

- 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

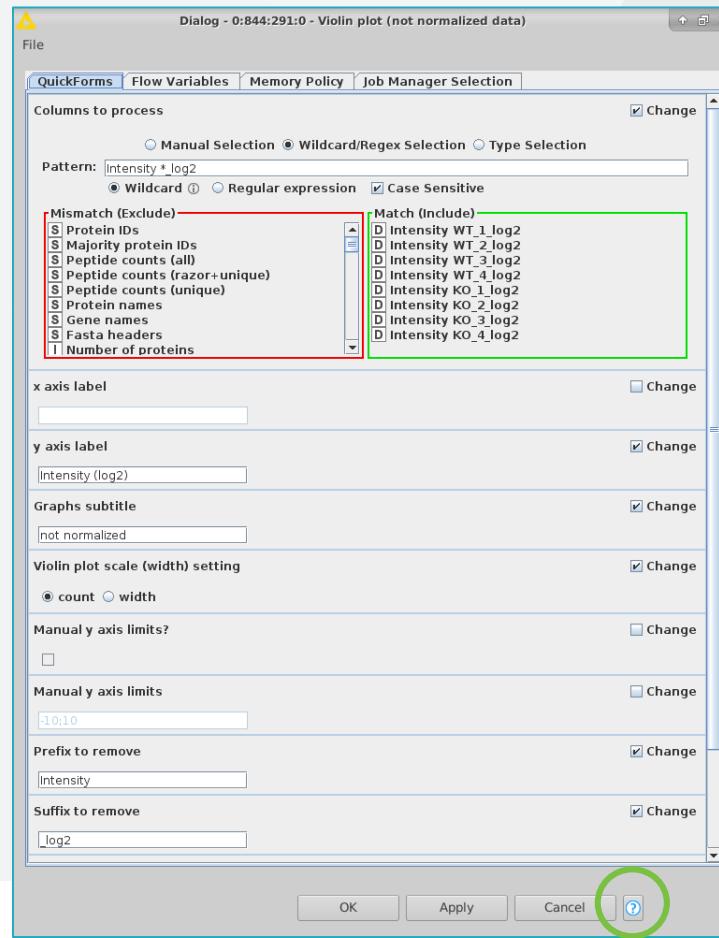
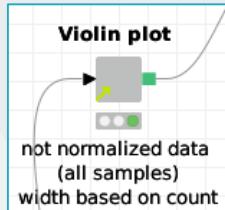




KNIME metanodes making concept

- 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

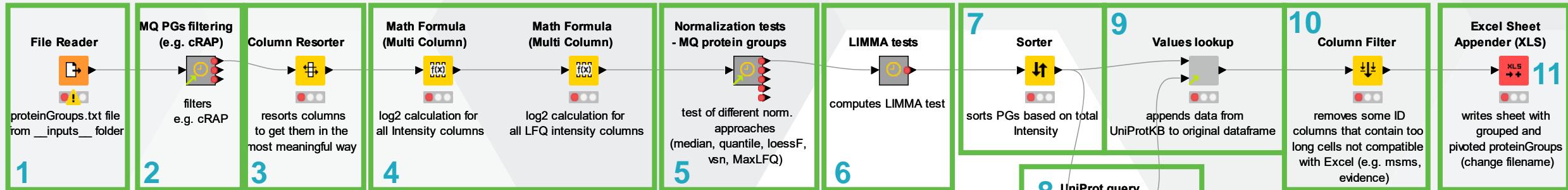
Our KNIME metanodes availability

- 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



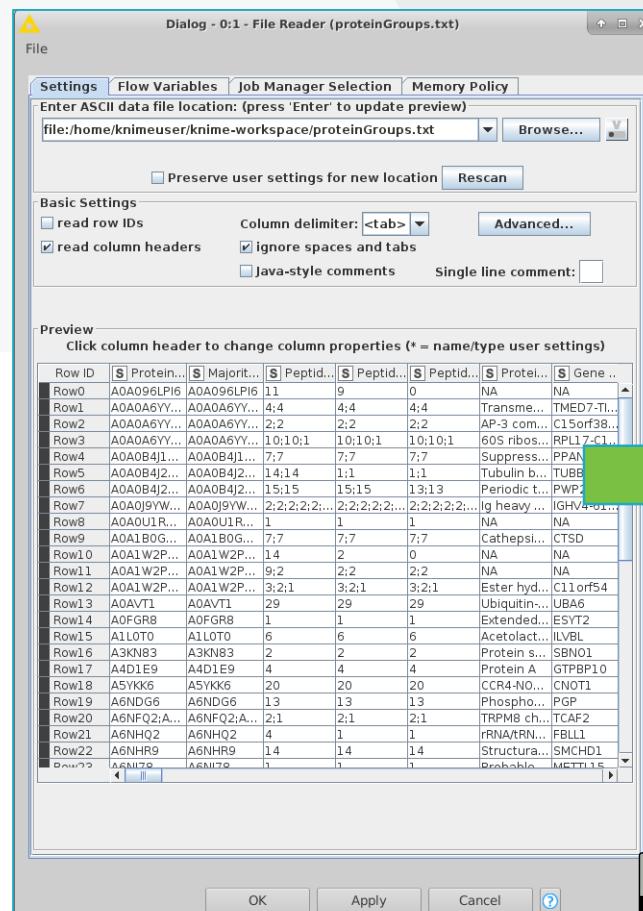
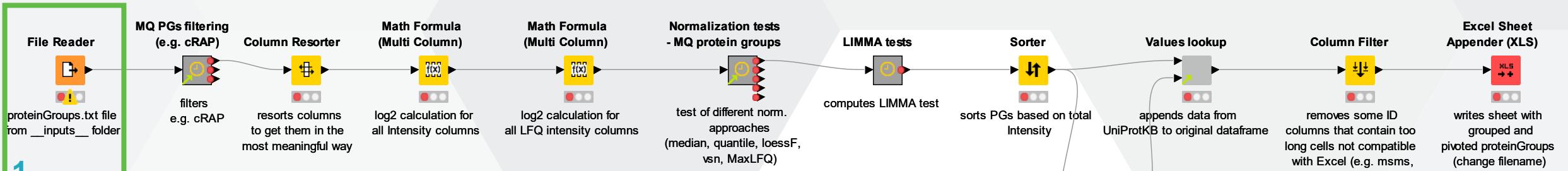
Example MaxQuant proteinGroups file processing workflow

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



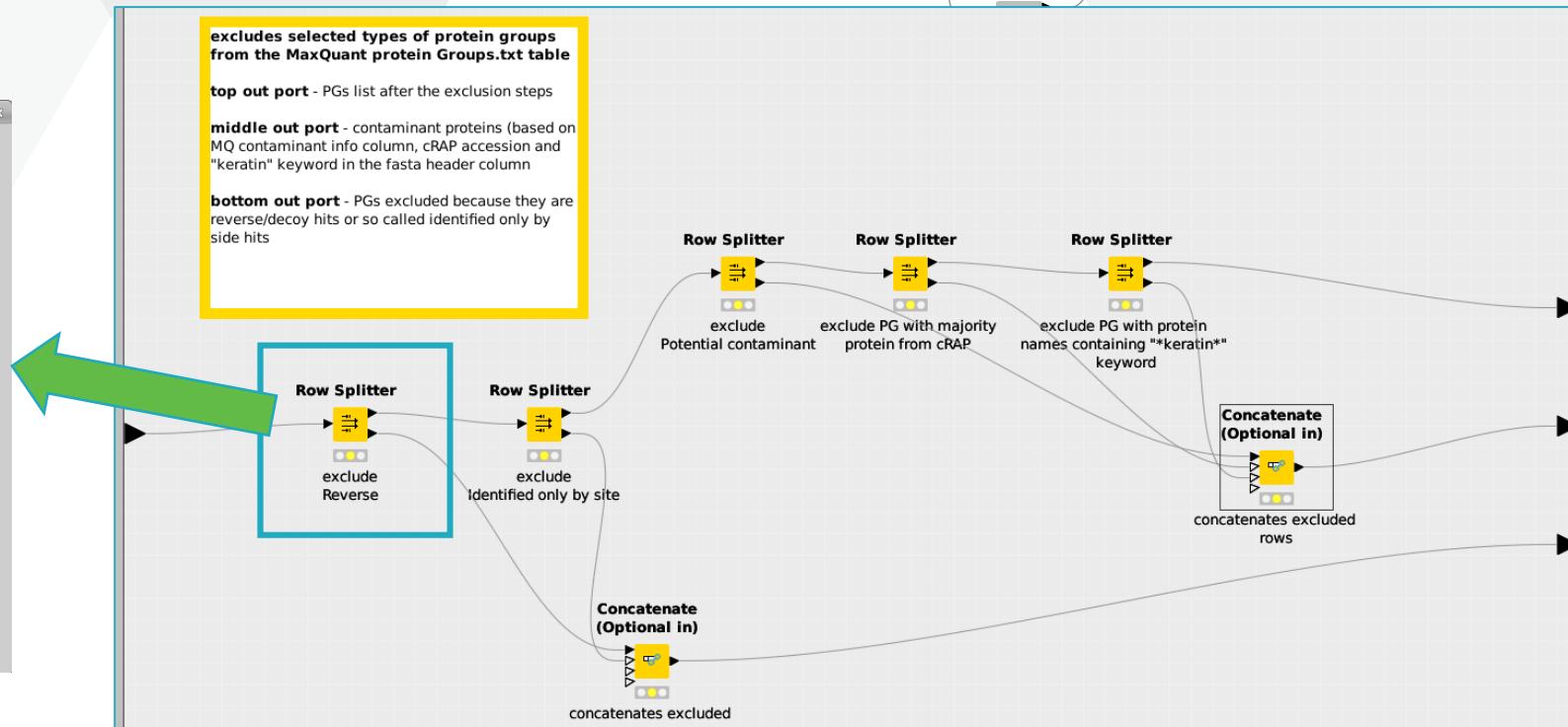
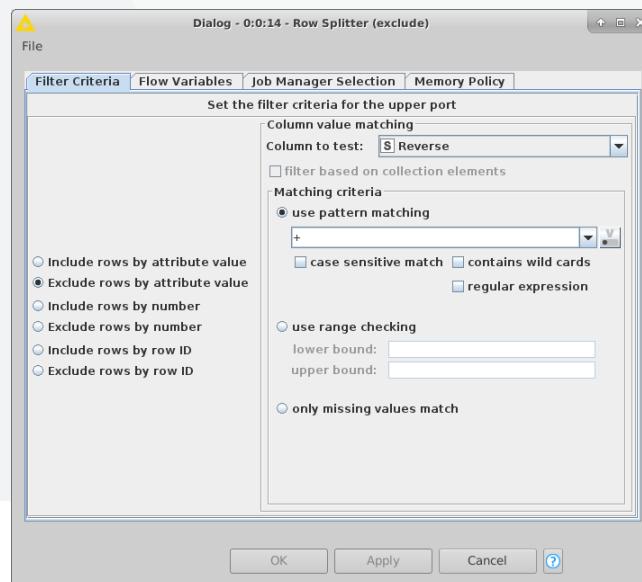
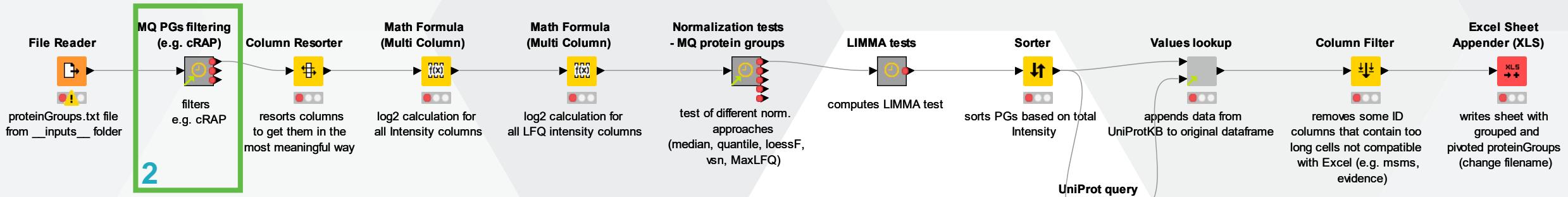
File Table - 0:1 - File Reader (proteinGroups.txt)

File Hilitc Navigation View

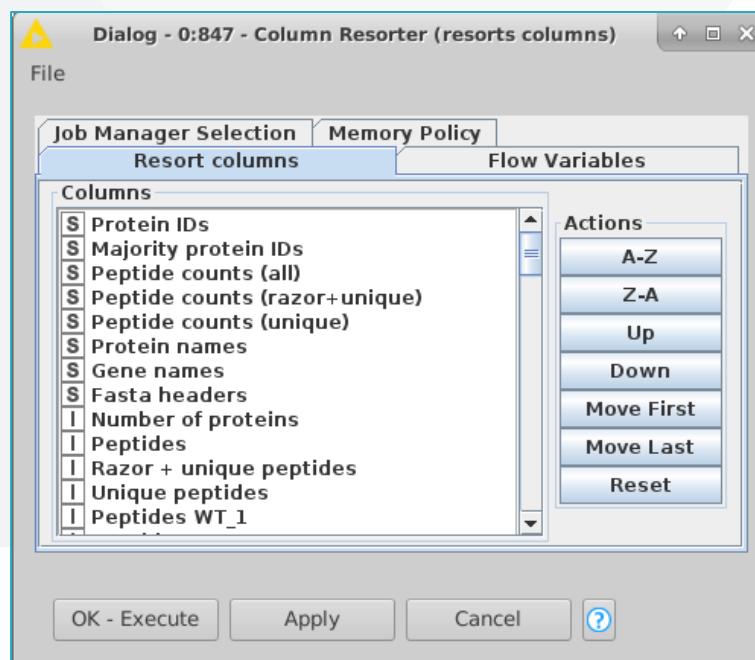
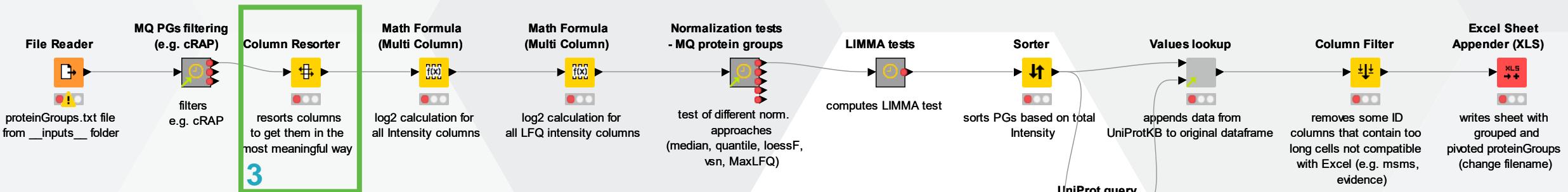
Table "proteinGroups.txt" - Rows: 3855 Spec - Columns: 81 Properties Flow Variables

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...	4;4	4;4	4;4	Transme...	TMED7-TL...	Protein TMED7-TCAM2 OS=Homo sapiens GN=TMED7
Row2	A0A0A6YY...	A0A0A6YY...	2;2	2;2	2;2	AP-3 com...	C15orf38...	Protein C15orf38-AP3S2 OS=Homo sapiens GN=C15orf38
Row3	A0A0A6YY...	A0A0A6YY...	10;10;1	10;10;1	10;10;1	60S ribos...	RPL17-C1...	Protein RPL17-C18orf32 OS=Homo sapiens GN=RPL17
Row4	A0A0B4J1...	A0A0B4J1...	7;7	7;7	7;7	Suppress...	PPAN-P2...	HCG2039996 OS=Homo sapiens GN=PPAN-P2RY1
Row5	A0A0B4J2...	A0A0B4J2...	14;14	1;1	1;1	Tubulin b...	TUBB3	Uncharacterized protein OS=Homo sapiens PE=1
Row6	A0A0B4J2...	A0A0B4J2...	15;15	15;15	13;13	Periodic t...	PWP2	Uncharacterized protein OS=Homo sapiens GN=L02
Row7	A0A0J9YW...	A0A0J9YW...	2;2;2;2;2...	2;2;2;2;2...	2;2;2;2;2...	Ig heavy...	IGHV4-61...	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	RPS10-NUDT3 readthrough OS=Homo sapiens GN
Row11	A0A1W2P...	A0A1W2P...	9;2	2;2	2;2	Ester hyd...	C11orf54	Uncharacterized protein OS=Homo sapiens PE=4
Row12	A0A1W2P...	A0A1W2P...	3;2;1	3;2;1	3;2;1	Ubiquitin...	UBA6	Ubiquitin-like modifier-activating enzyme 6 OS=Homo
Row13	A0AVT1	A0AVT1	29	29	29	Extended...	ESYT2	Extended synaptotagmin-2 OS=Homo sapiens GN=ESYT2
Row14	A0FGR8	A0FGR8	1	1	1	Acetolact...	ILVBL	Acetolactate synthase-like protein OS=Homo sapi
Row15	A1L0TO	A1L0TO	6	6	6	Protein s...	SBNO1	Protein strawberry notch homolog 1 OS=Homo sa
Row16	A3KN83	A3KN83	2	2	2	Protein A	GTPBP10	Protein A OS=Homo sapiens GN=GTPBP10 PE=1 S
Row17	A4D1E9	A4D1E9	4	4	4	CCR4-NO...	CNOT1	Protein CCR4-NOT transcription complex subunit 1 OS=Ho
Row18	A5YKK6	A5YKK6	20	20	20	Phospho...	PGP	Glycerol-3-phosphate phosphatase OS=Homo sap
Row19	A6NDG6	A6NDG6	13	13	13	TRPM8 ch...	TCAF2	TRPM8 channel-associated factor 2 OS=Homo sap
Row20	A6NFQ2;A...	A6NFQ2;A...	2;1	2;1	2;1	rRNA/tRN...	FBLL1	rRNA/rRNA 2'-O-methyltransferase fibrillarin-like pro
Row21	A6NHQ2	A6NHQ2	4	1	1	Structura...	SMCHD1	Structural maintenance of chromosomes flexible H
Row22	A6NHR9	A6NHR9	14	14	14	Probable...	METTL15	Probable methyltransferase-like protein 15 OS=Ho
Row23	A6NJ78	A6NJ78	1	1	1	RanBP2-li...	RGPD3;R...	RanBP2-like and GRIP domain-containing protein 3 OS=Homo sapiens
Row24	A6NKT7;Q...	A6NKT7;Q...	18;16	1;1	1;1	Nuclear e...	POM121C...	Nuclear envelope pore membrane protein POM121C OS=Homo sapiens
Row25	A8CG34;Q...	A8CG34;Q...	4;3	4;3	4;3	Nucleosi...	NUDT19	Nucleoside diphosphate-linked moiety X motif 19 OS=Homo sapiens
Row26	A8MXV4;R...	A8MXV4	3;1;1	3;1;1	3;1;1	MIF4G do...	MIF4G	MIF4G domain-containing protein OS=Homo sapiens
Row27	A9UHW6	A9UHW6	2	2	2	Unconve...	MYO1G	Unconventional myosin-Ig OS=Homo sapiens GN=MYO1G
Row28	B0I1T2	B0I1T2	39	39	38	Mitochon...	SLC25A10	cDNA FLJ60124, highly similar to Mitochondrial dic
Row29	B4DLN1;Q...	B4DLN1;Q...	14;7	14;7	7;7	RanBP2-like and GRIP domain-containing protein 3 OS=Homo sapiens		
Row30	P0CG08;B...	P0CG08;B...	1;1	1;1	1;1	Golgi pH ...	GPR89B;...	Golgi pH regulator B OS=Homo sapiens GN=GPR89B
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 sapiens

2. filters out reverse hits and contaminants

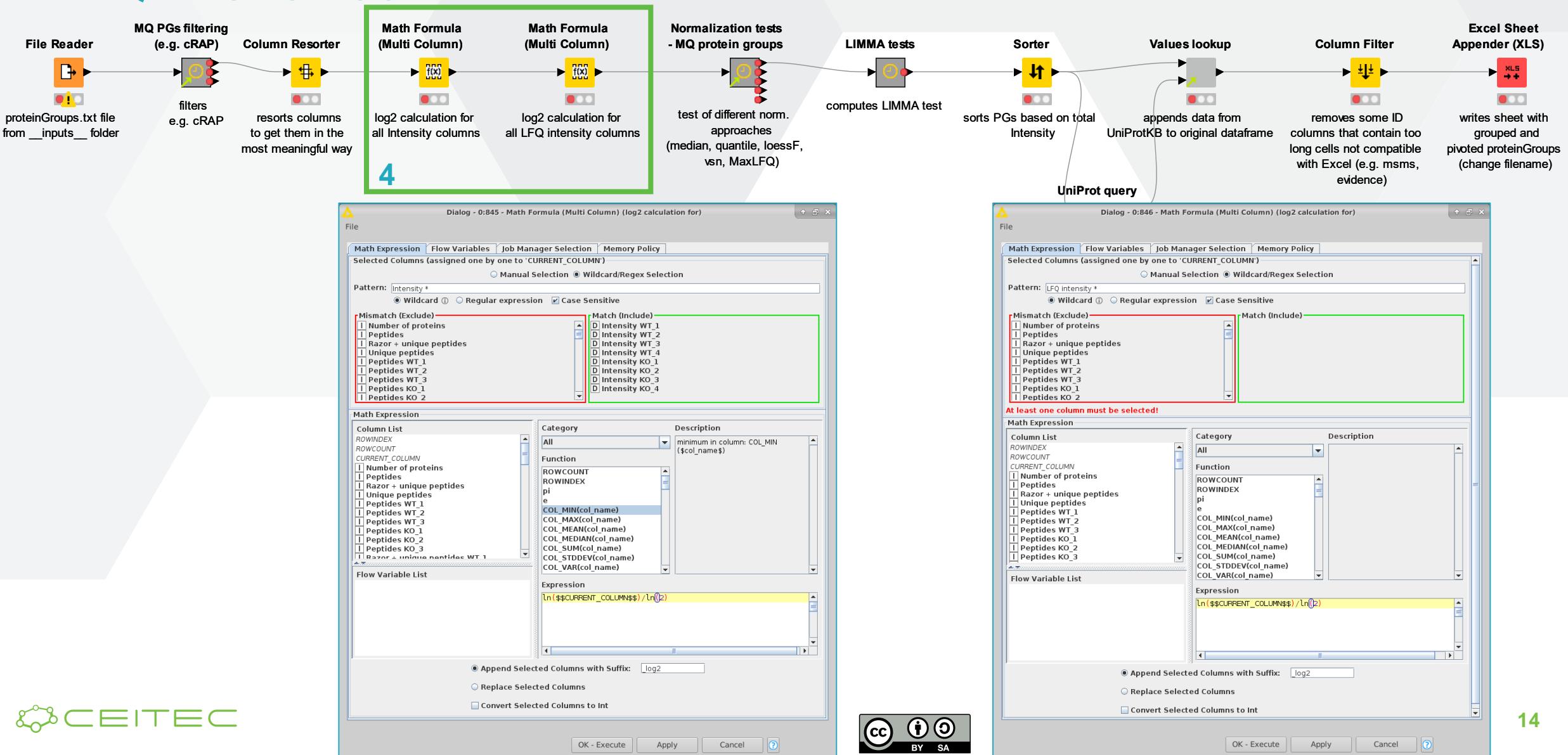


3. resorts the samples/replicates ordering (figures)

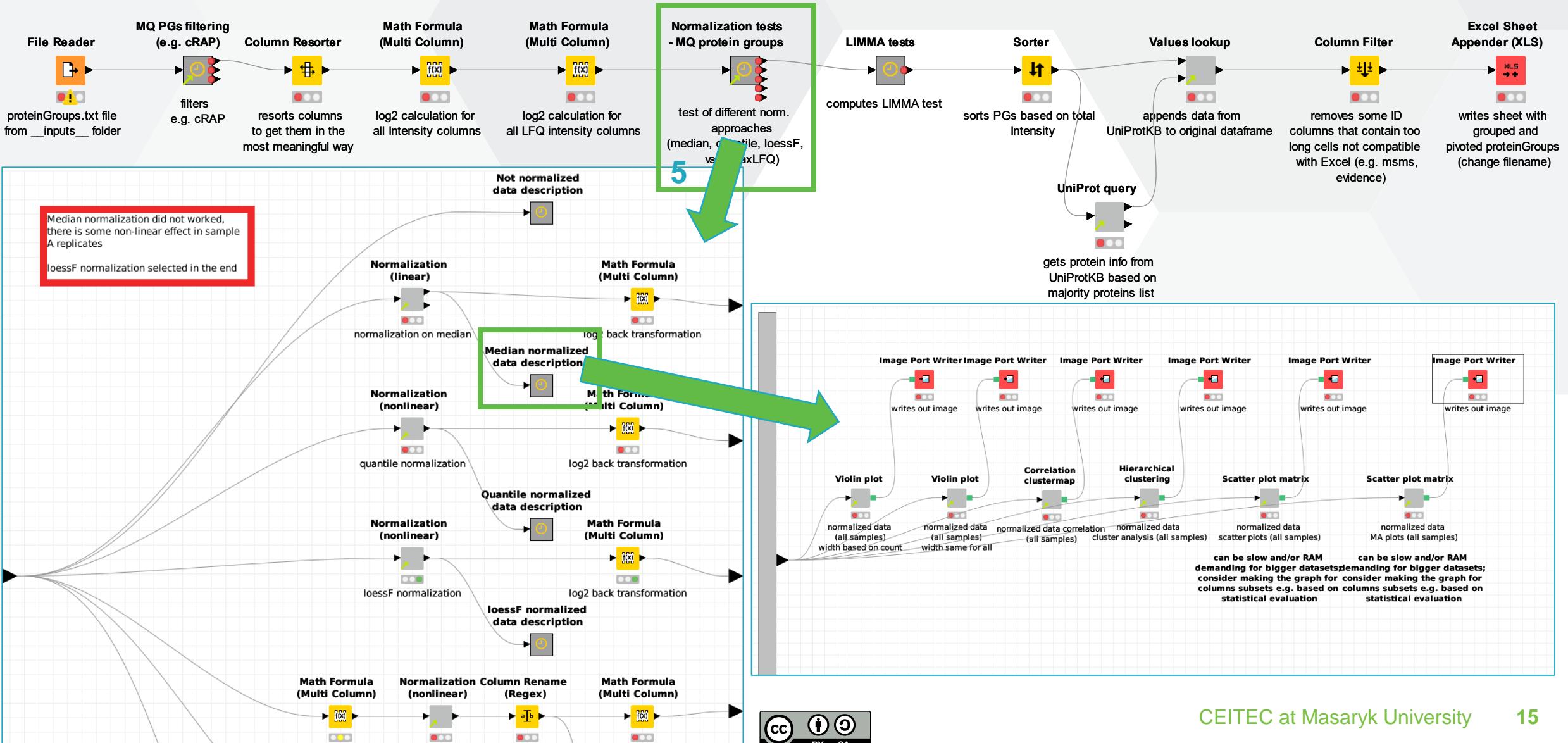


- order of data columns important for e.g. their visualization – order is kept the same as in the table
- the node enables you to resort e.g. intensity columns in the most meaningful way for viewing

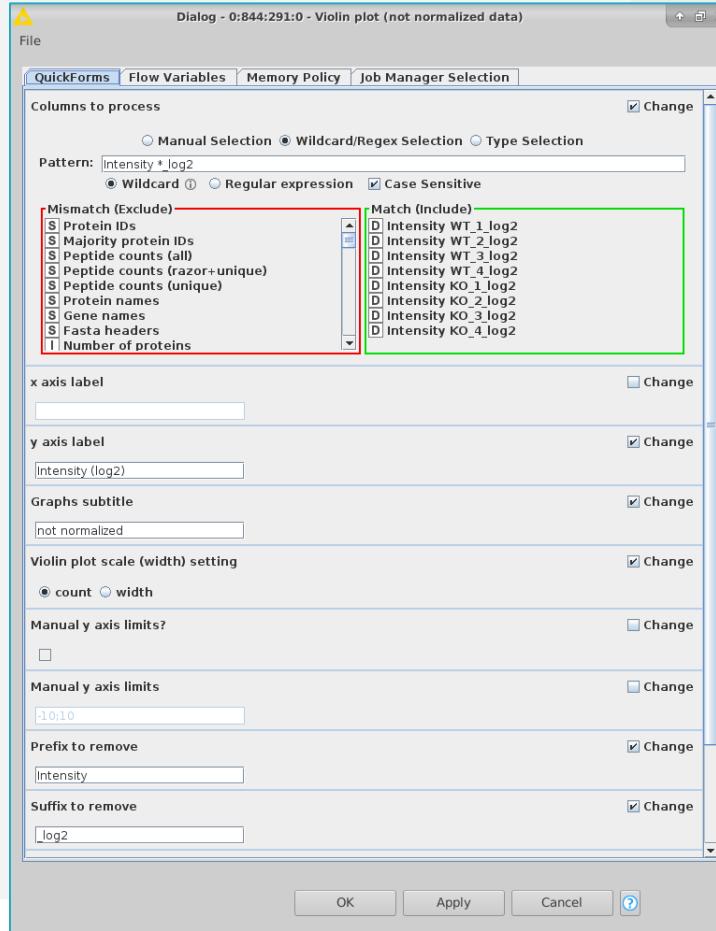
4. data transformation (log2) – protein groups Intensities and LFQ intensities



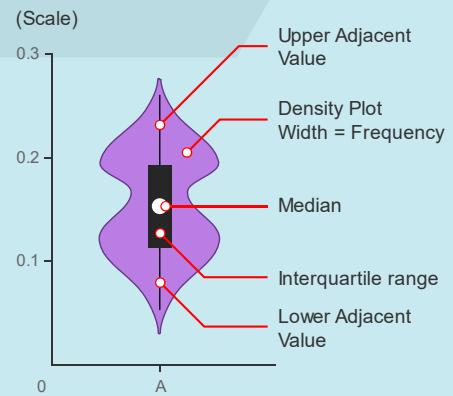
5. data normalization approach selection and data check



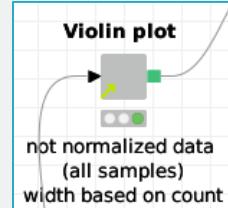
Metanodes examples – Violin plot



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

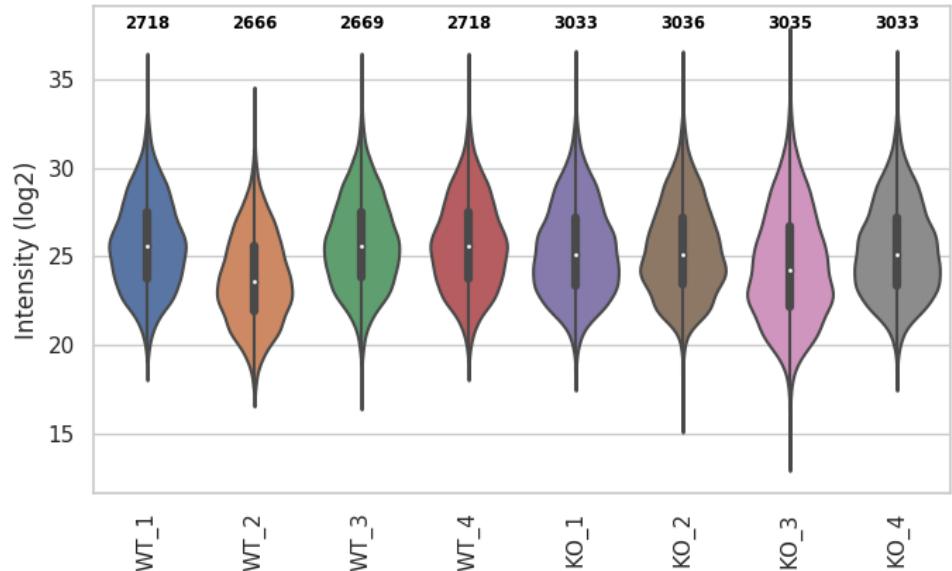


- Comparison of samples:** number of cases, distribution, basic statistics (min, max, median)

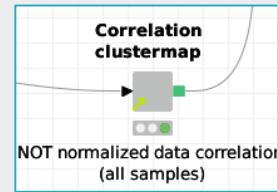
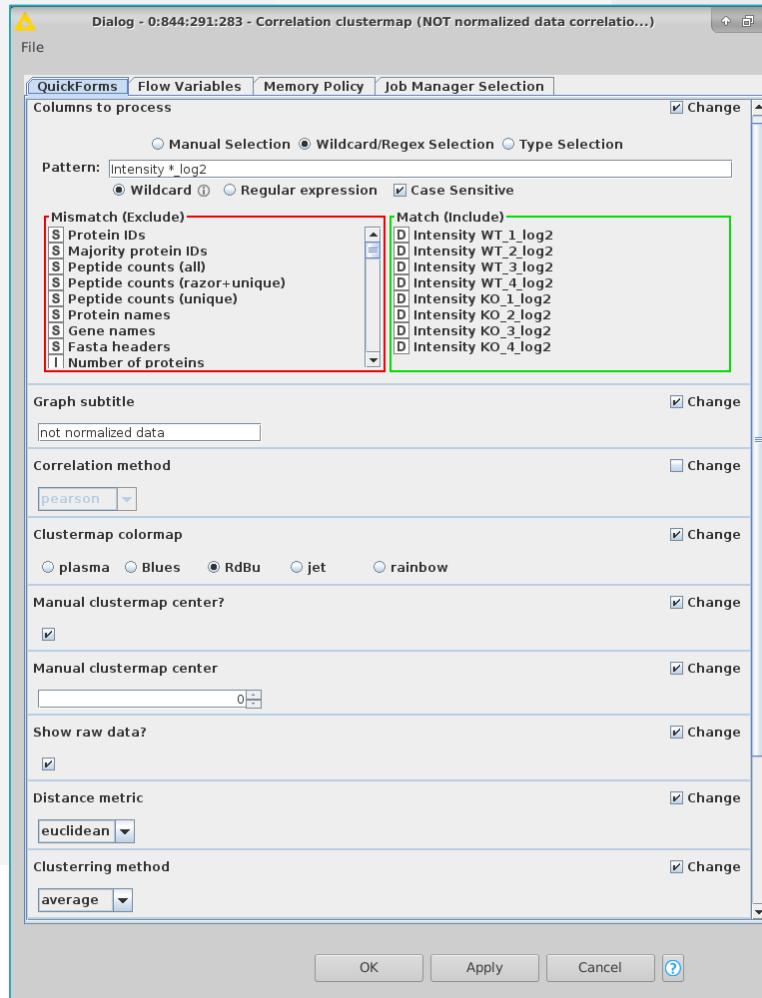
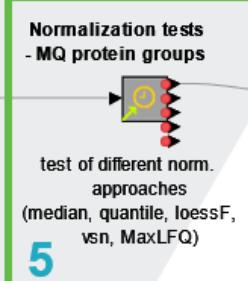


Normalization tests - MQ protein groups
test of different norm. approaches
(median, quantile, loessF, vsn, MaxLFQ)
5

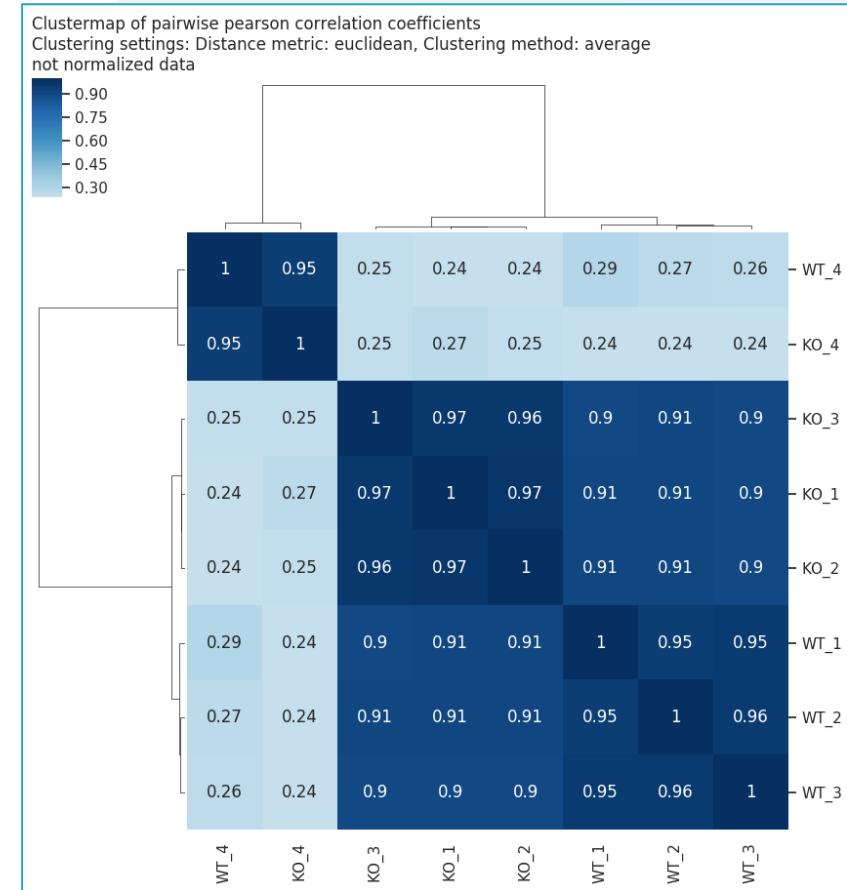
Violin graphs and number of observed values: violins width corresponds to the number of values not normalized



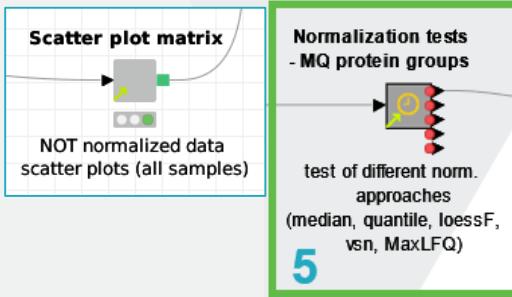
Metanodes examples – Correlation clustermap



- 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? Change

Type of regression: lowess Change

Show unity line? Change

Impute missing values? Change

Value to impute if requested

Construct MA plot like graph? Change

manual x axis limits: None Change

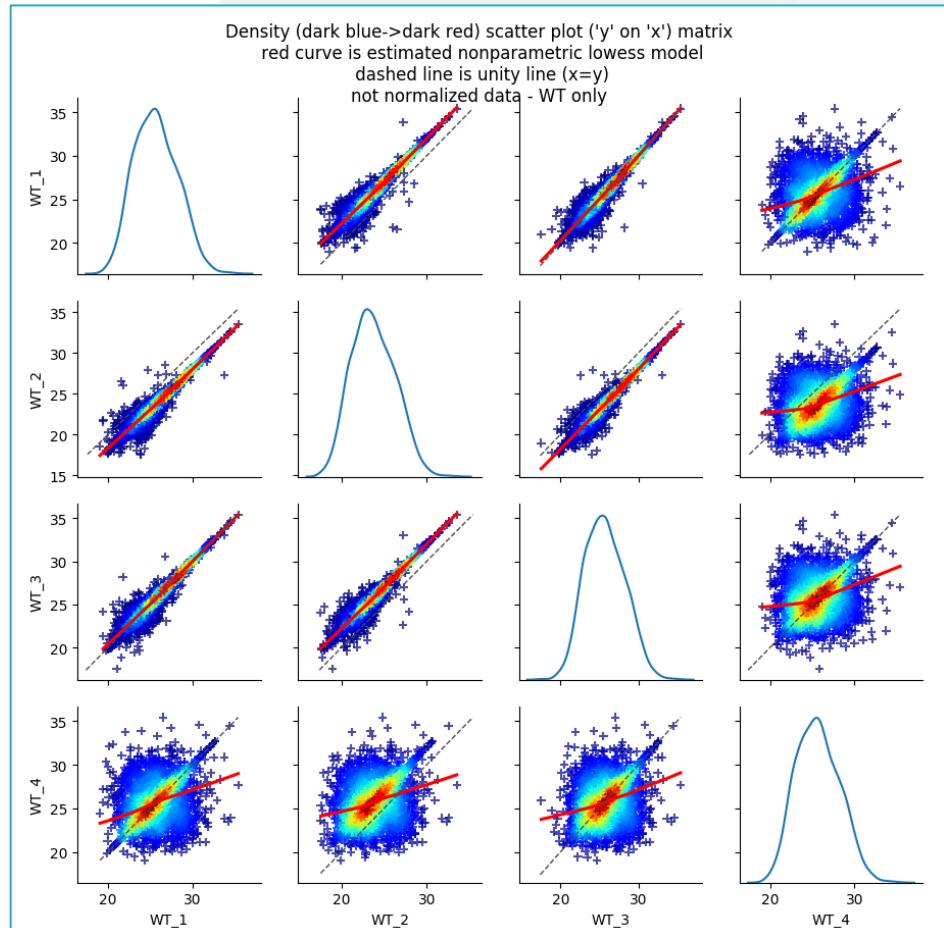
manual y axis limits: None Change

Prefix to remove: Intensity Change

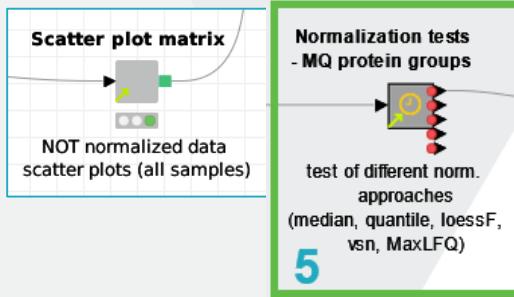
Suffix to remove: log2 Change

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



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)

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

Match (Include)

- Intensity WT_1.log2
- Intensity WT_2.log2
- Intensity WT_3.log2
- Intensity WT_4.log2

Graphs subtitle: not normalized data Change

Fit regression curve?

Type of regression: lowess Change

Show unity line? Change

Impute missing values? Change

Value to impute if requested: 15 Change

Construct MA plot like graph?

Manual x axis limits: None Change

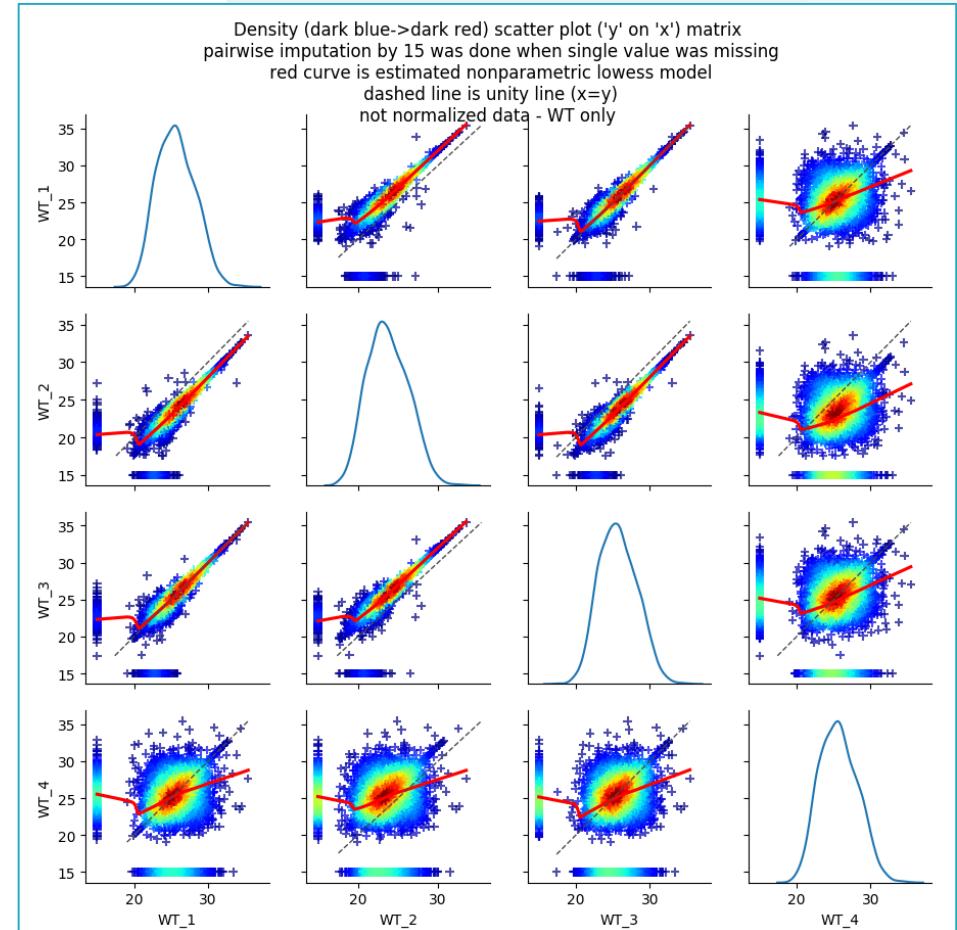
Manual y axis limits: None Change

Prefix to remove: Intensity Change

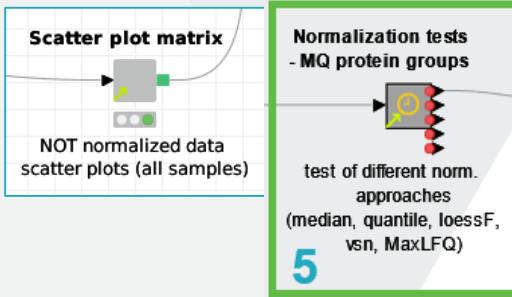
Suffix to remove: log2 Change

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



5

QuickForms Flow Variables Memory Policy Job Manager Selection

Columns to process

- Manual Selection Wildcard/Regex Selection Type Selection

Pattern: Intensity WT*_log2

Mismatch (Exclude)

- 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

Match (Include)

- D Intensity WT_1.log2
- D Intensity WT_2.log2
- D Intensity WT_3.log2
- D Intensity WT_4.log2

Graphs subtitle

not normalized data

Fit regression curve?

Type of regression

Show unity line?

Impute missing values?

Value to impute if requested

Construct MA plot like graph?

manual x axis limits

manual y axis limits

Prefix to remove

Intensity

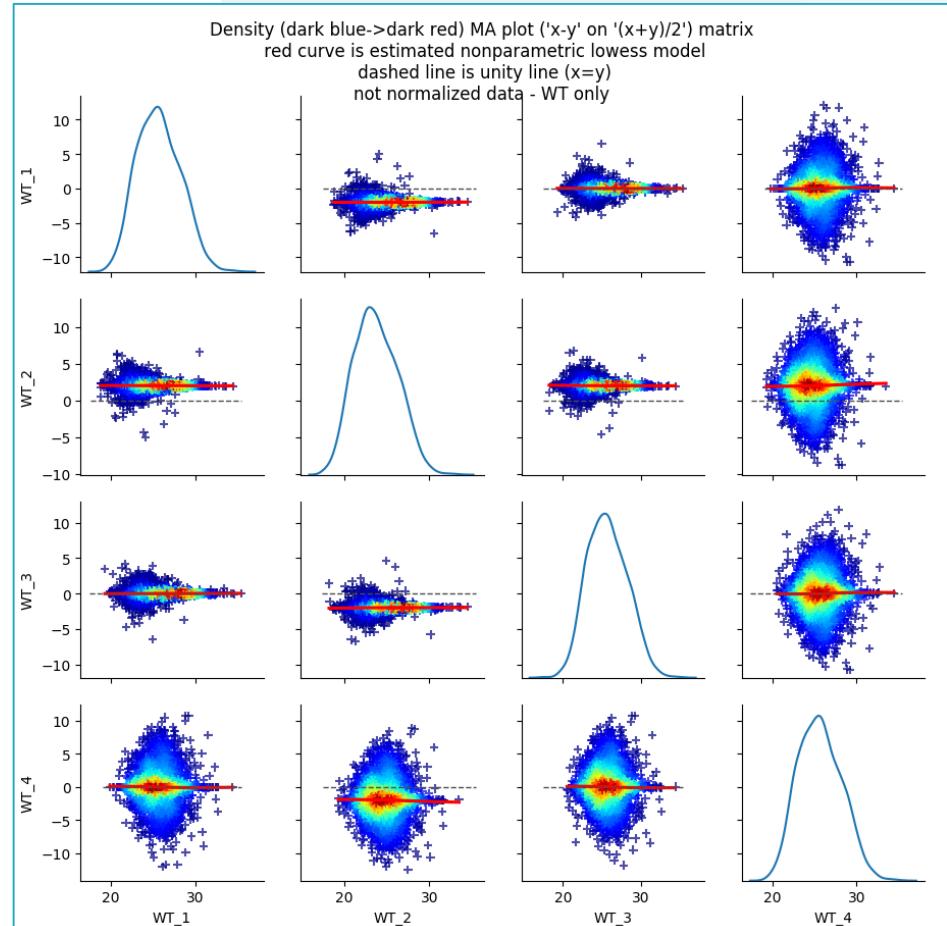
Suffix to remove

log2

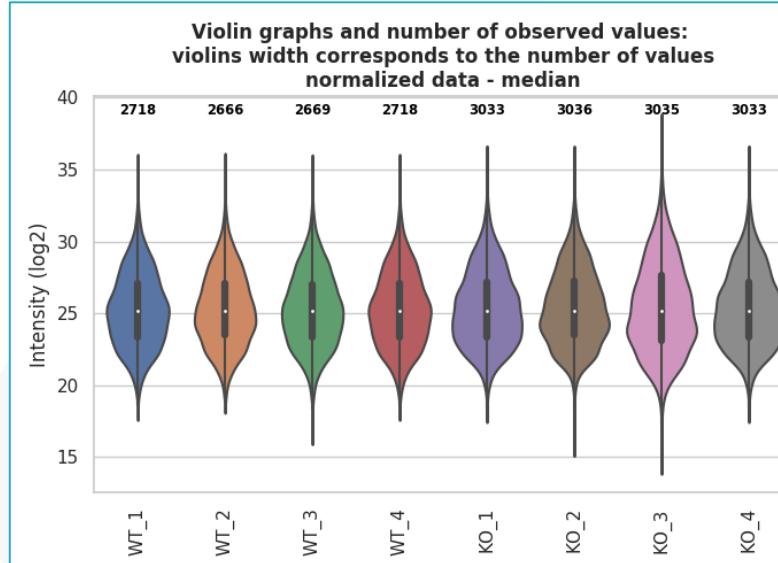
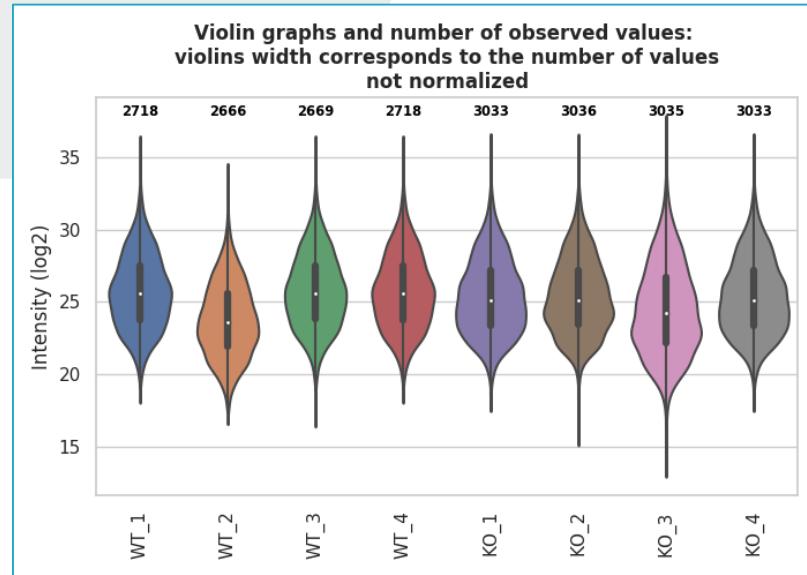
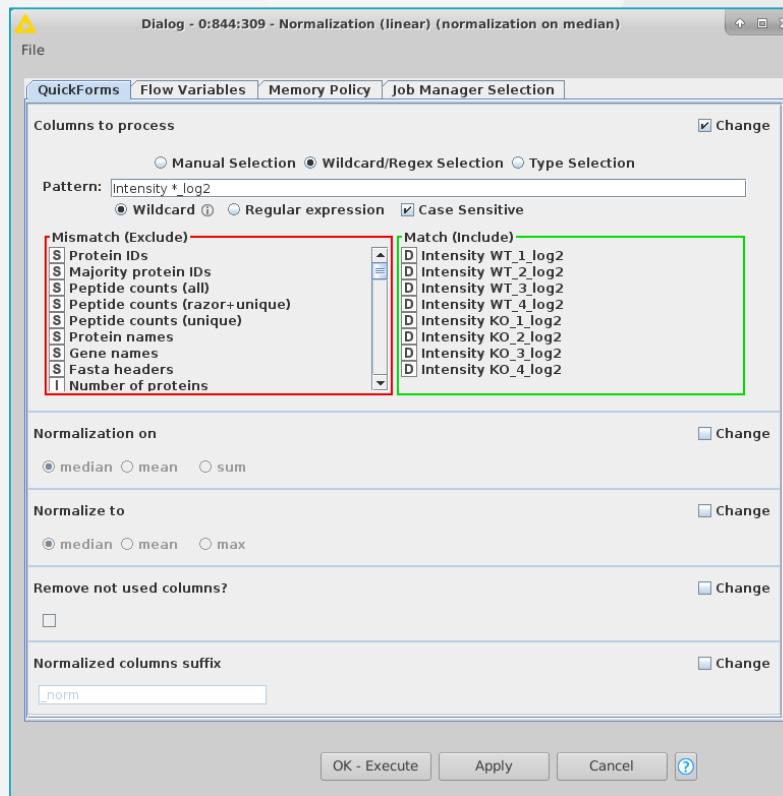
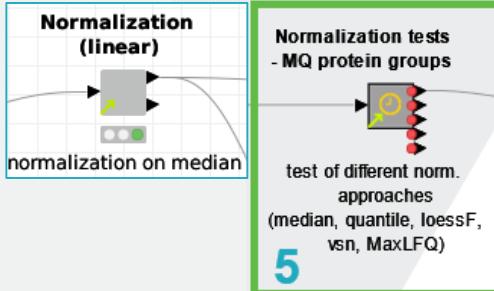
Size

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”

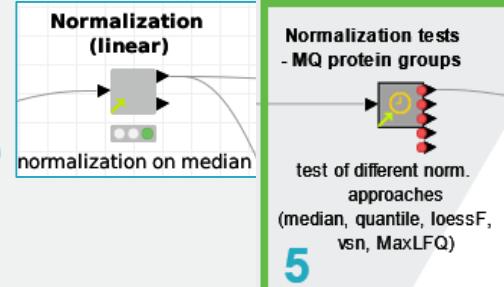


Metanodes examples – Normalization (linear)

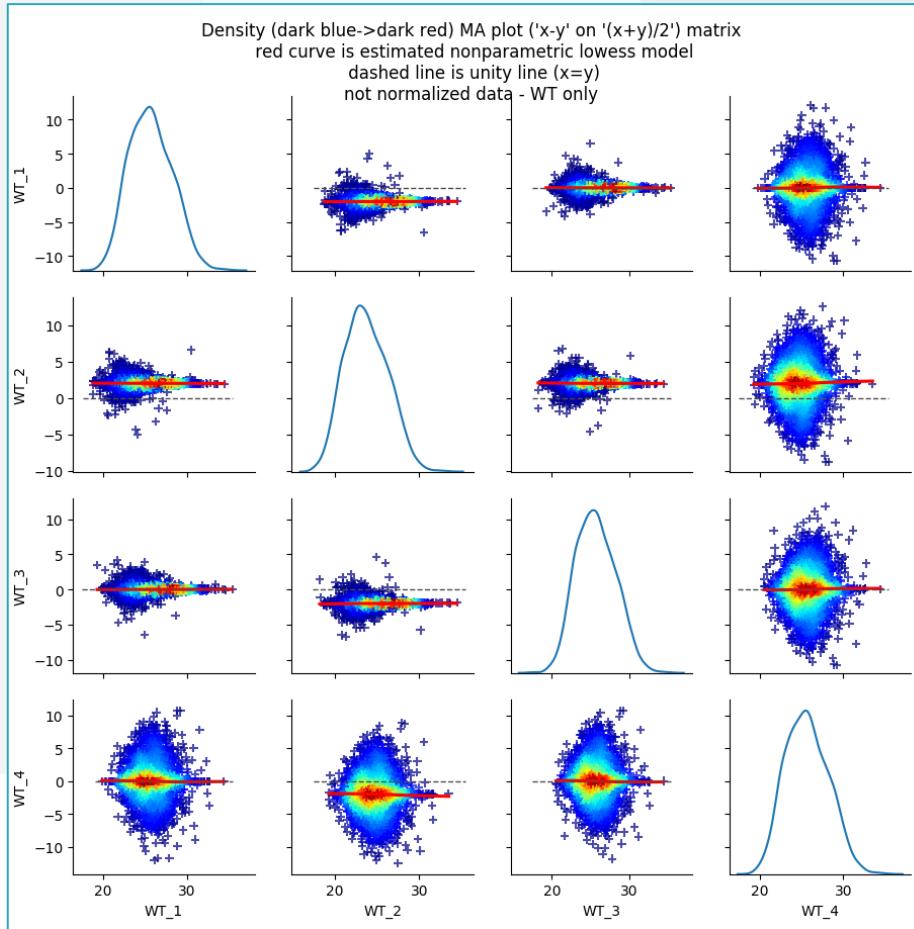


- Removal of **linear** technical processing artefacts from data
- Violin plots are just moved up or down to have the median value equal along all column selected for normalization
- subtracts e.g. column median (“Normalize on” option) from all columns and adds e.g. median (“Normalize to” option) of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

Metanodes examples – Normalization (linear)

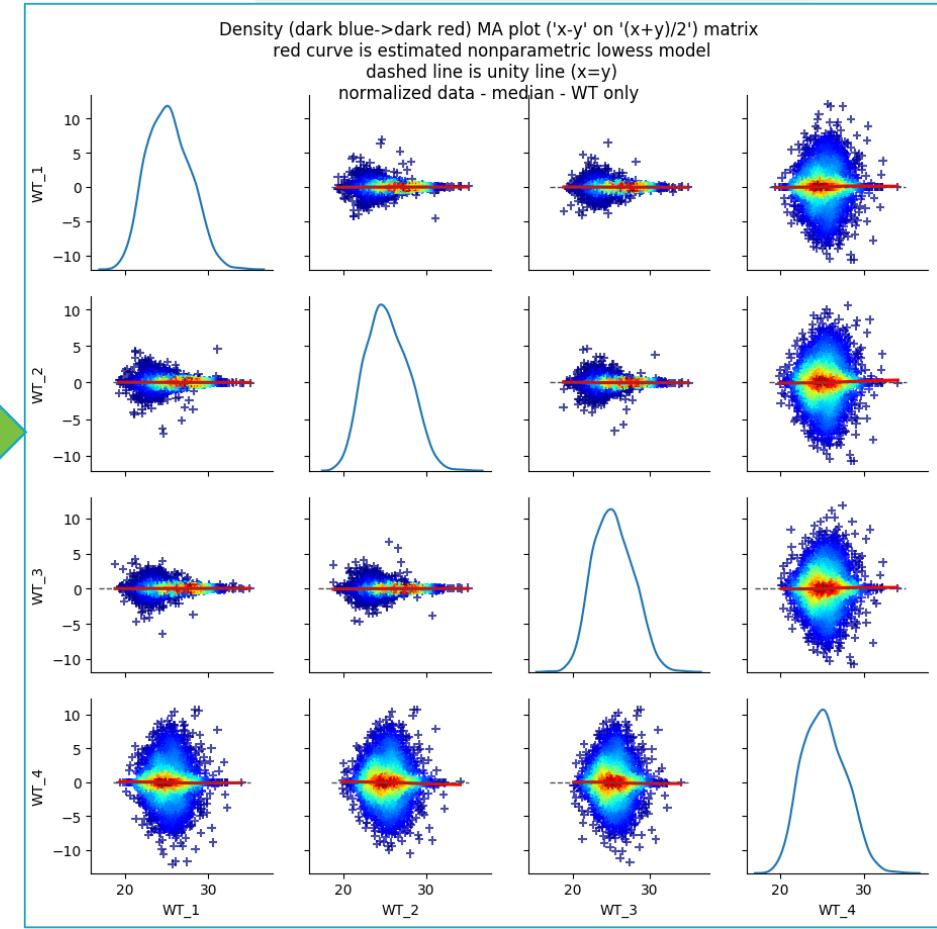


WT replicates only – prior normalization

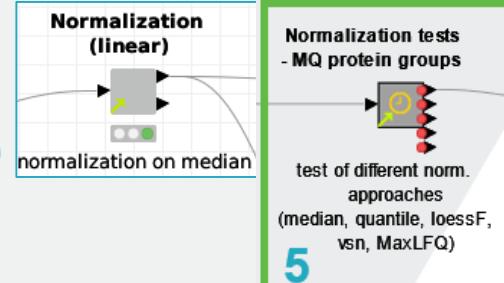


median
normalization

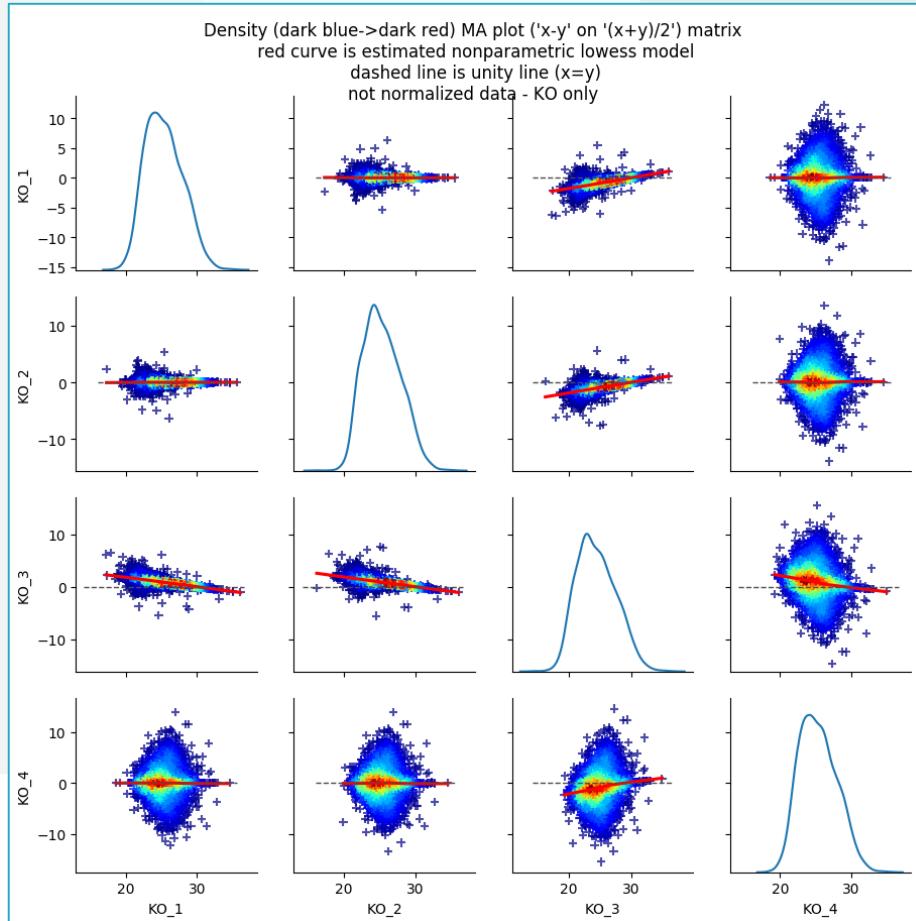
WT replicates only – after normalization



Metanodes examples – Normalization (linear)

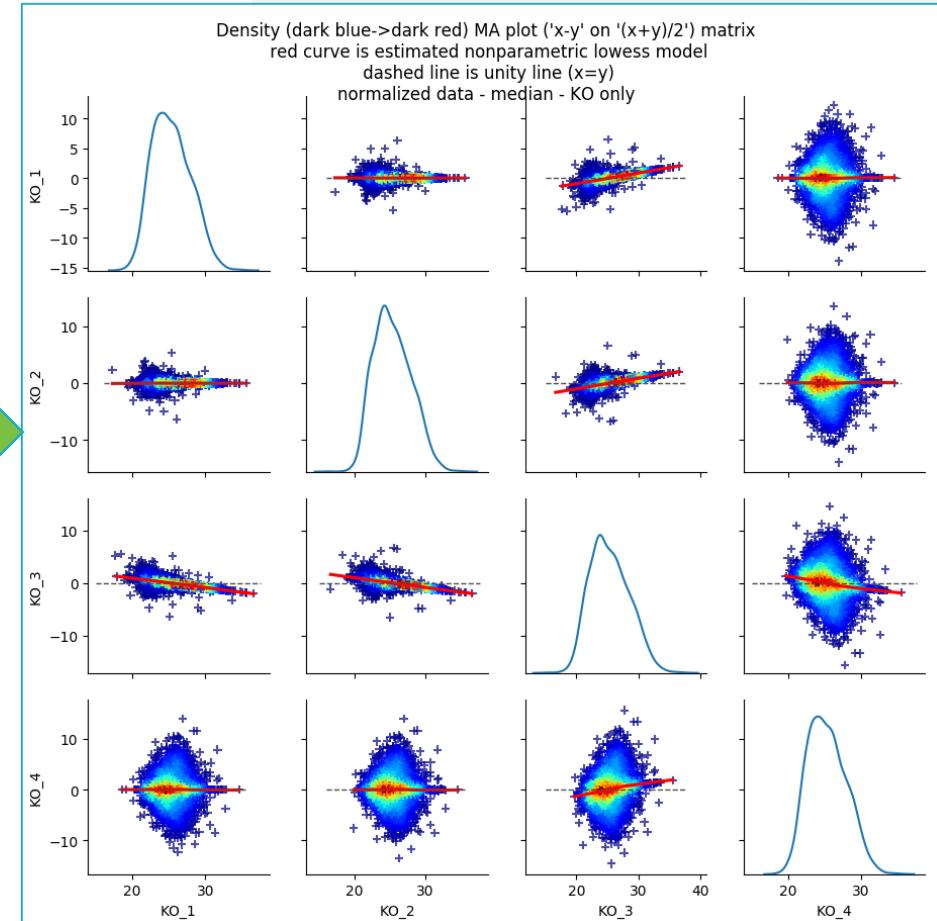


KO replicates only – prior normalization

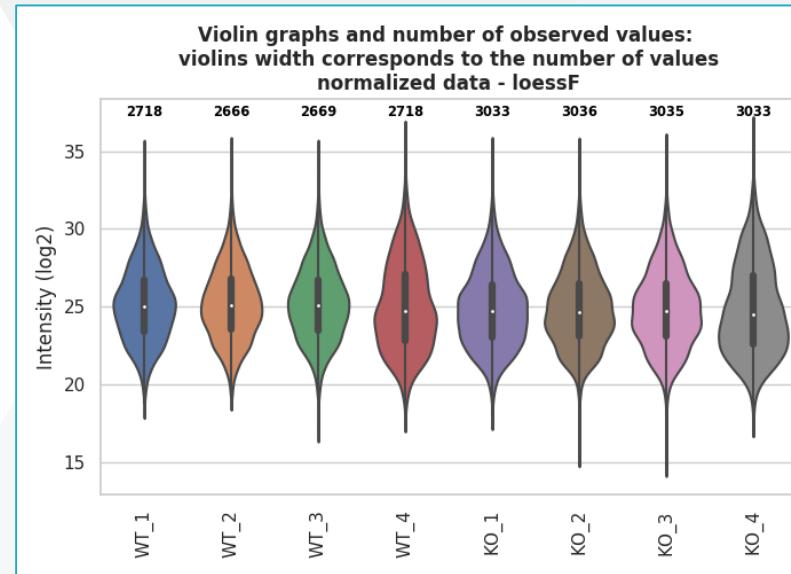
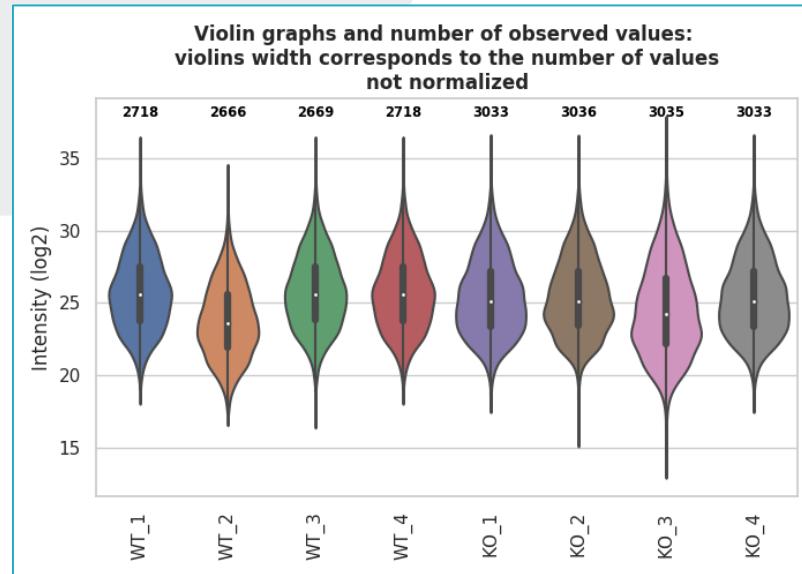
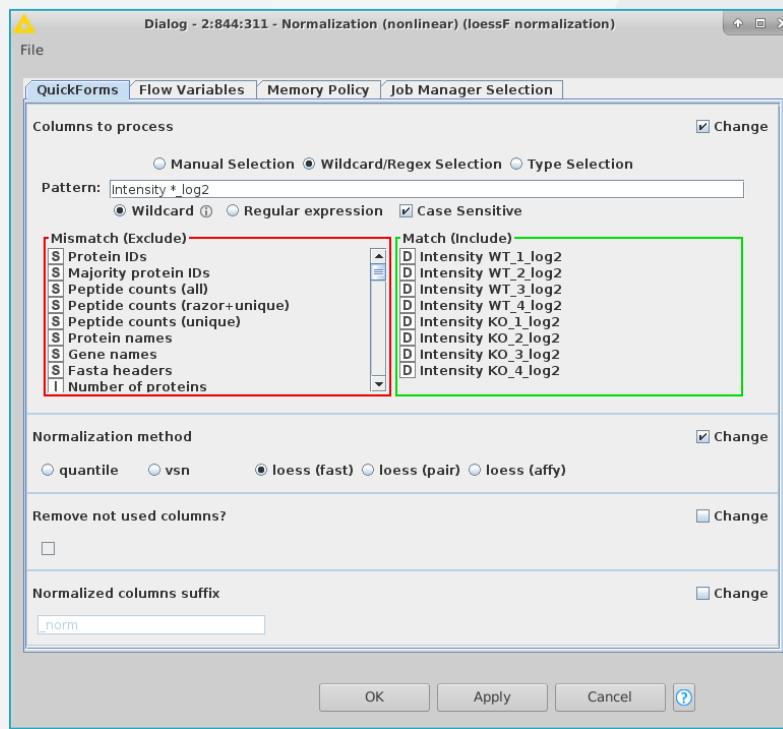
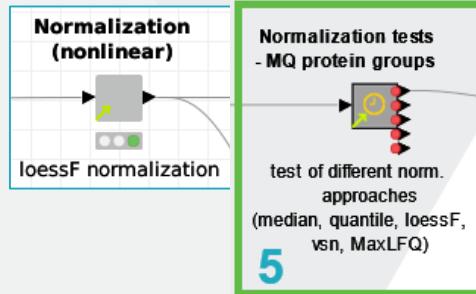


median
normalization

KO replicates only – after normalization

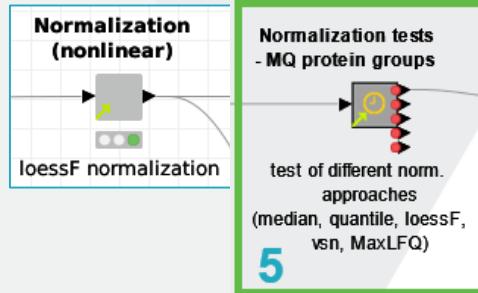


Metanodes examples – Norm. (nonlinear)

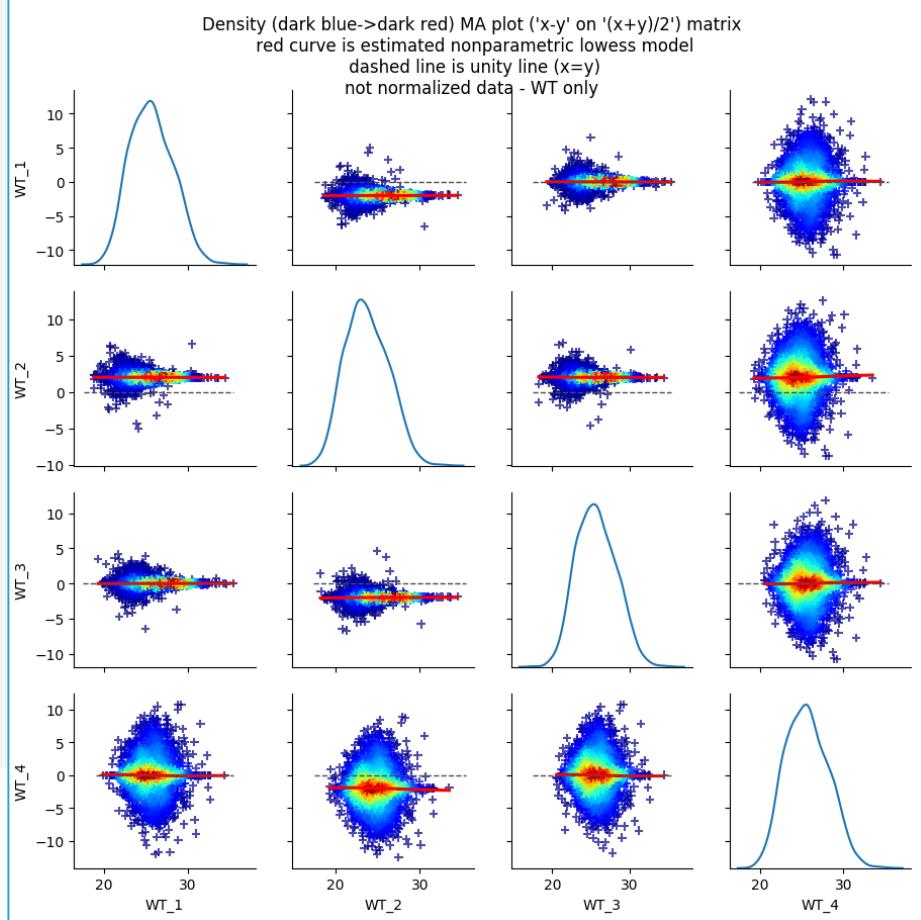


- Removal of **nonlinear** technical processing artefacts from data
- concrete data processing way depends on the selected normalization technique
- check the metanode description for implementation details and selected R packages for publications and algorithm details

Metanodes examples – Norm. (nonlinear)

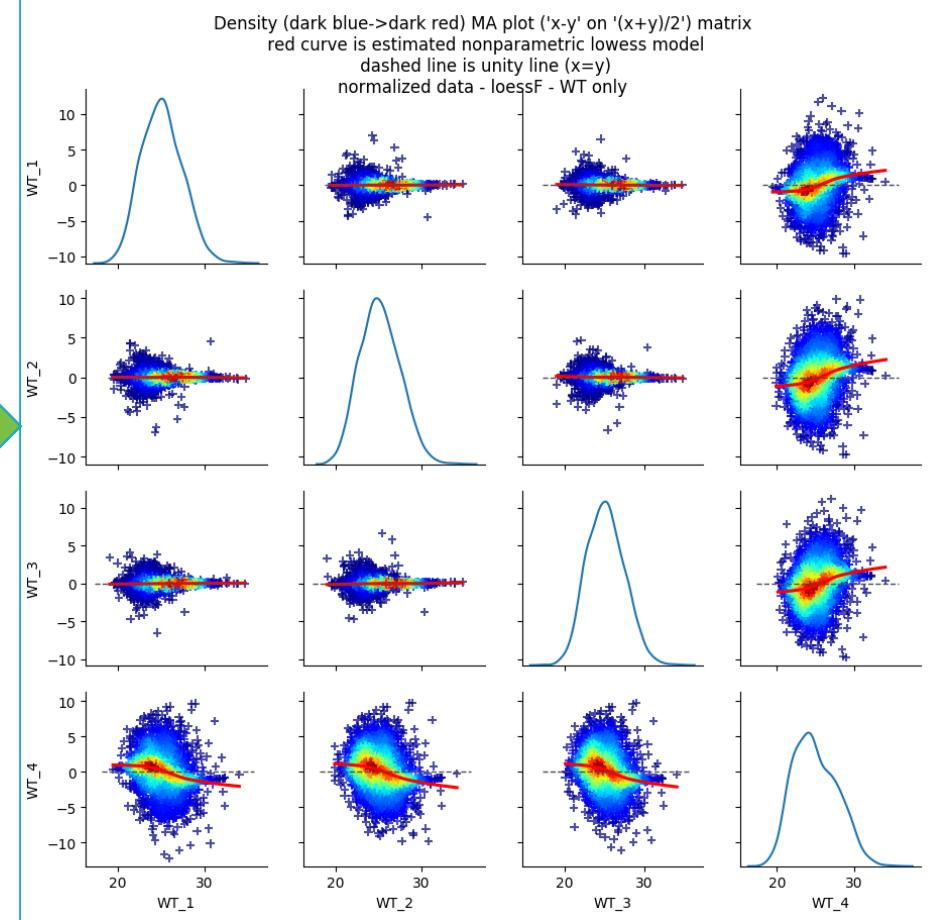


WT replicates only – prior normalization

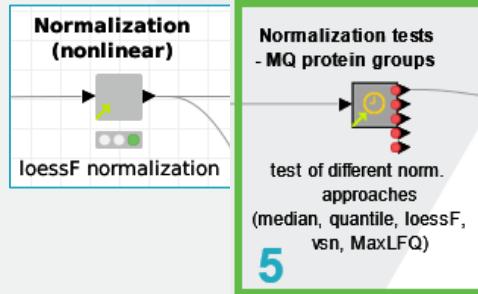


loess (fast)
normalization

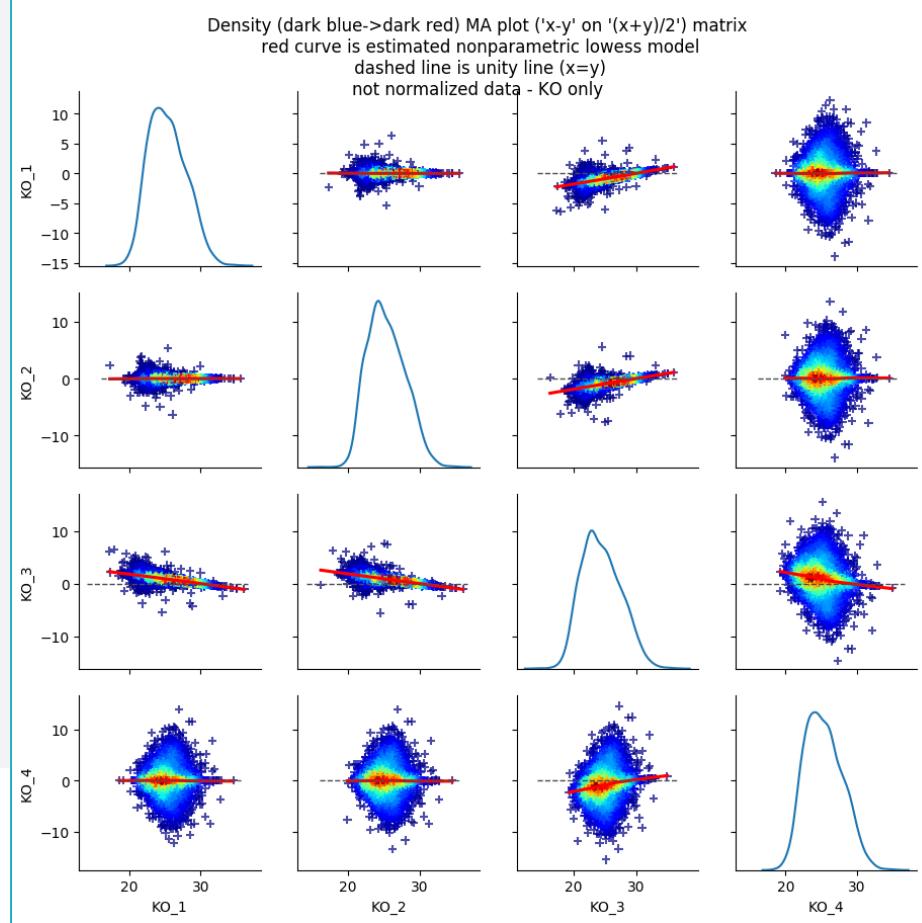
WT replicates only – after normalization



Metanodes examples – Norm. (nonlinear)

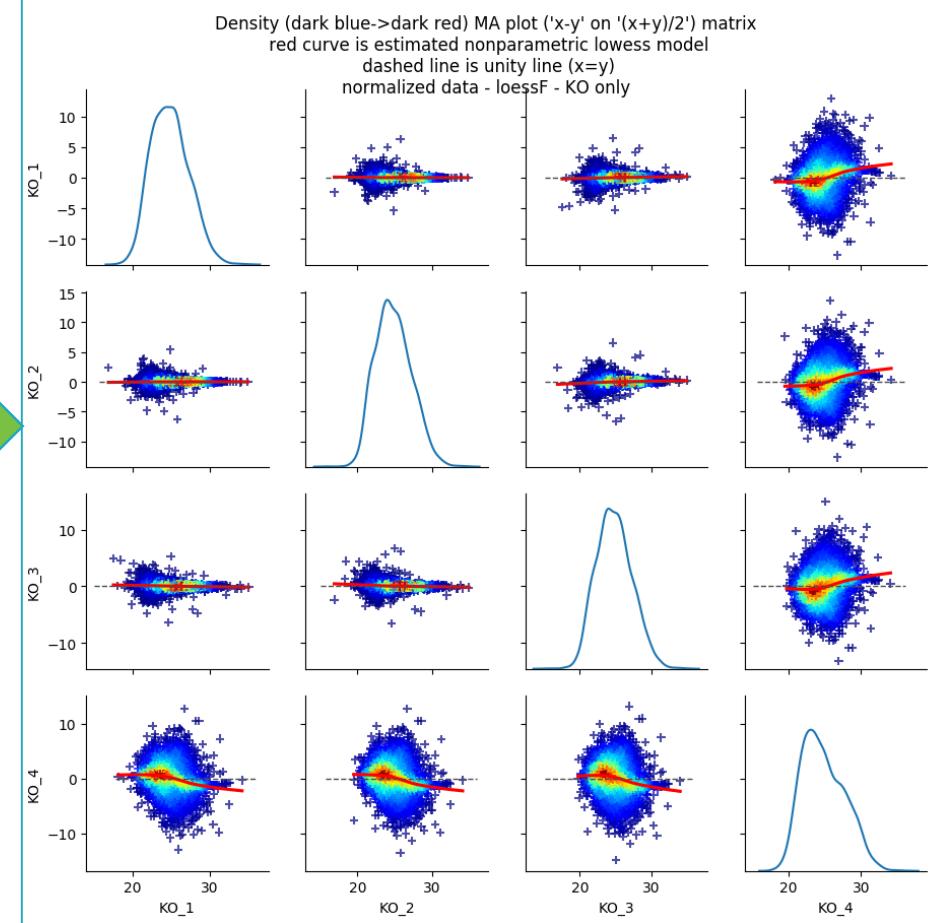


KO replicates only – prior normalization



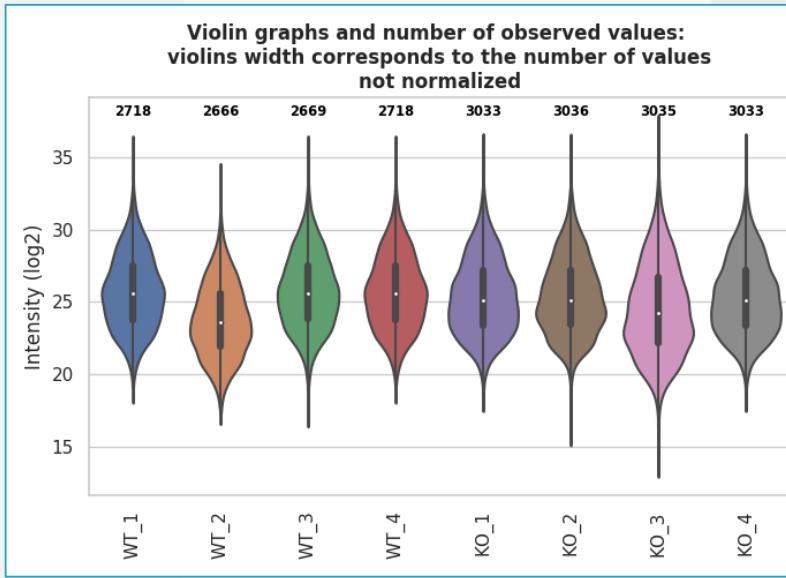
loess (fast)
normalization

KO replicates only – after normalization

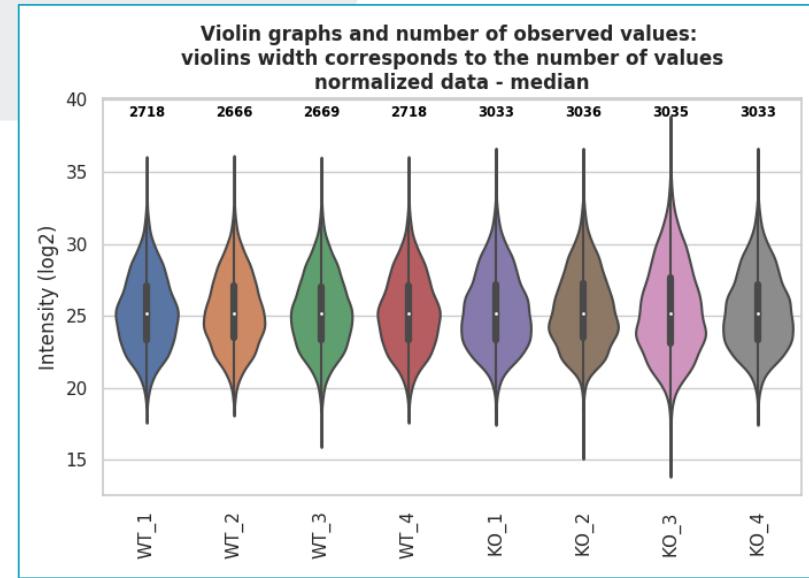


Normalization comparison – violin plots

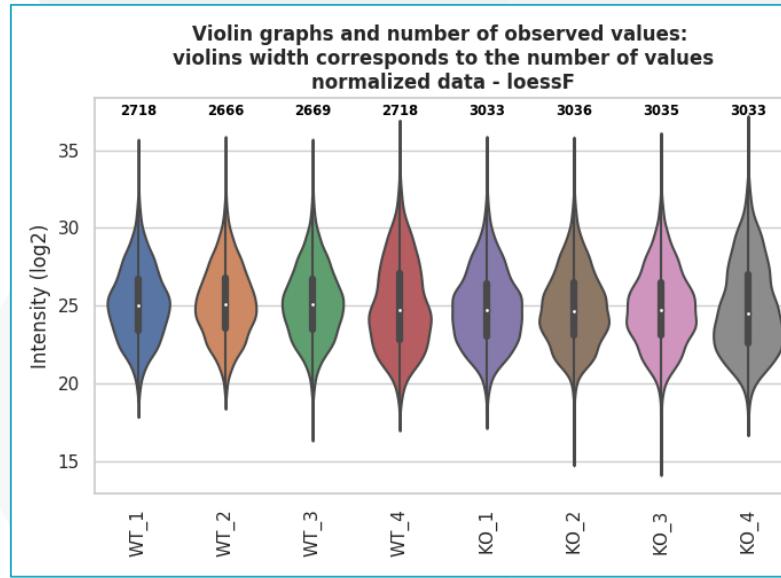
prior normalization



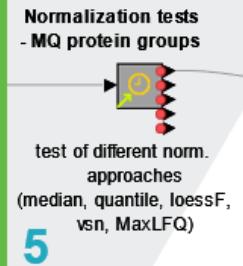
median normalization



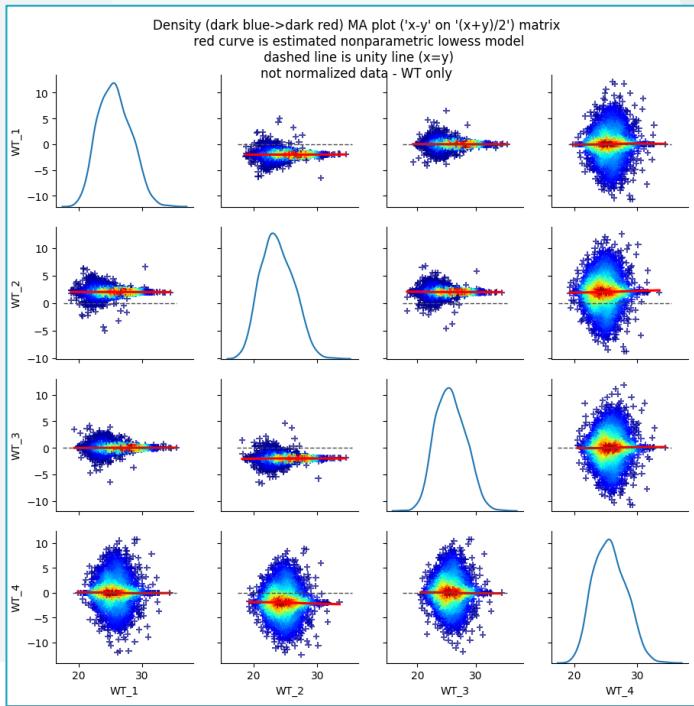
loess (fast) normalization



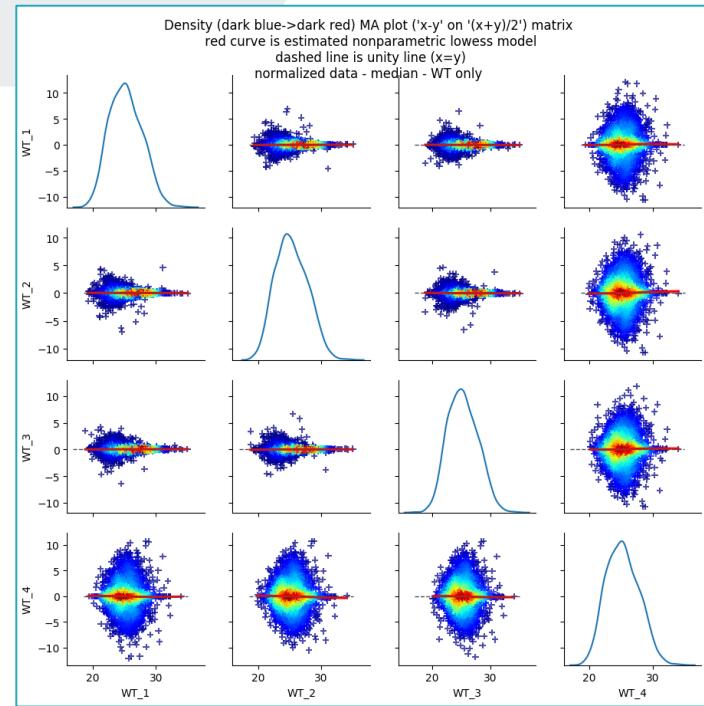
Normalization comparison – MA plots



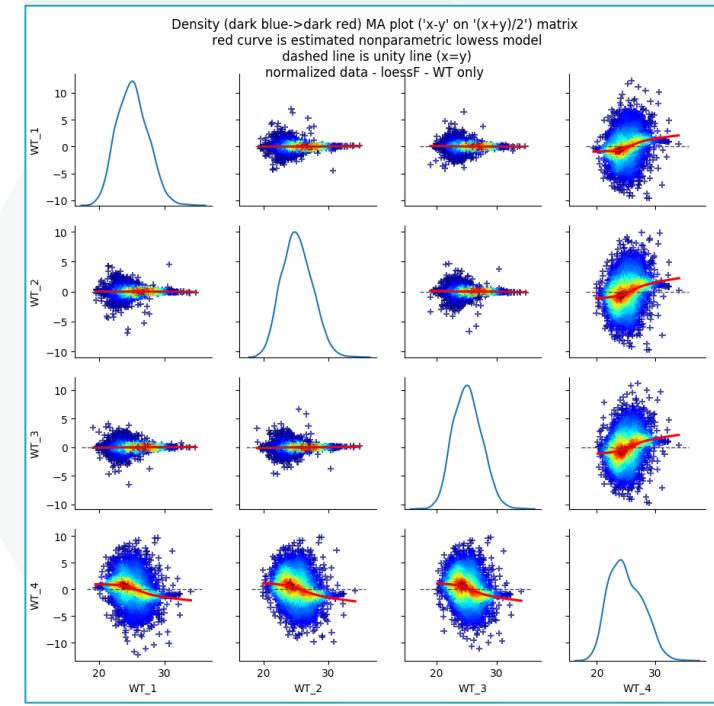
WT – prior normalization



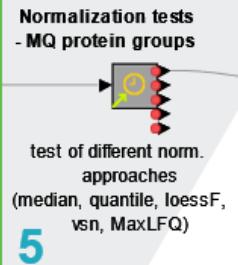
WT – median normalization



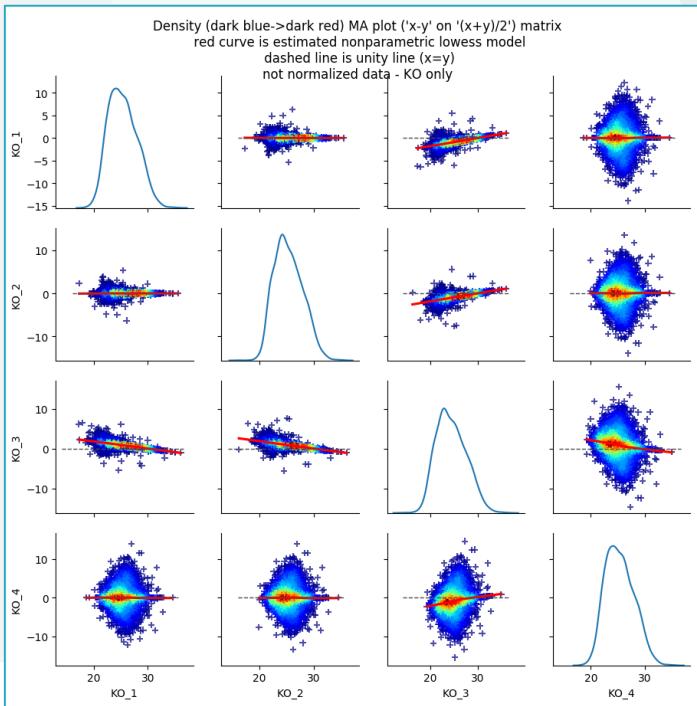
WT – loess (fast) normalization



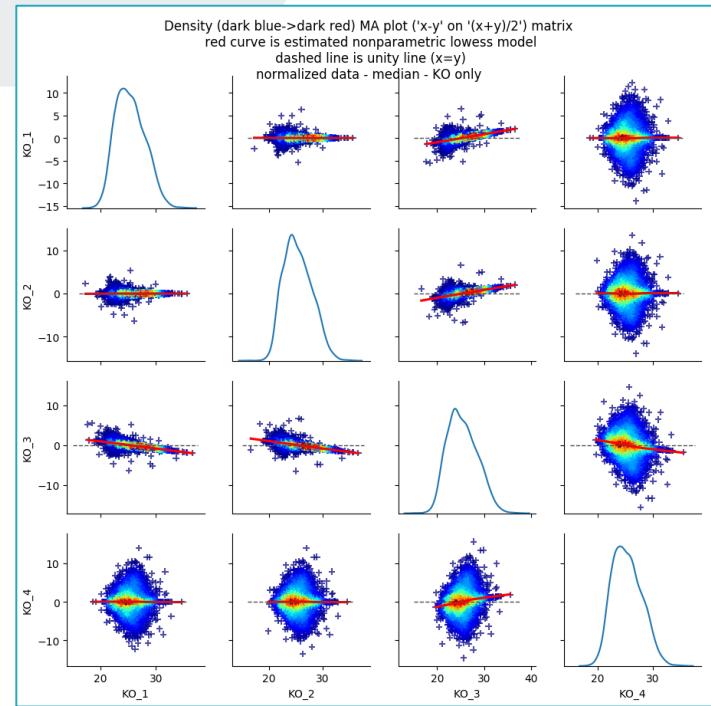
Normalization comparison – MA plots



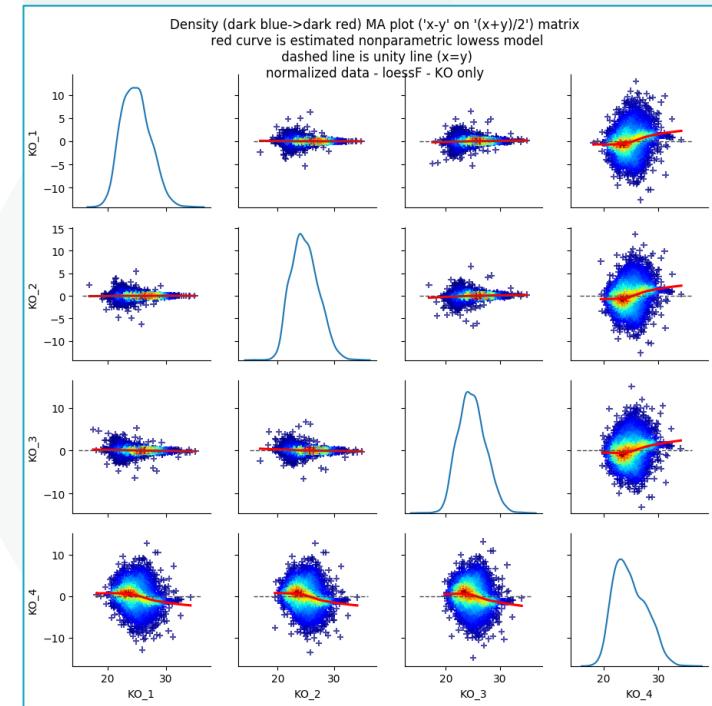
KO – prior normalization



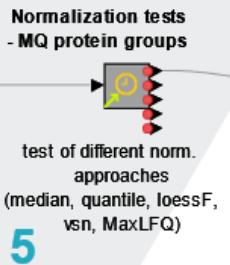
KO – median normalization



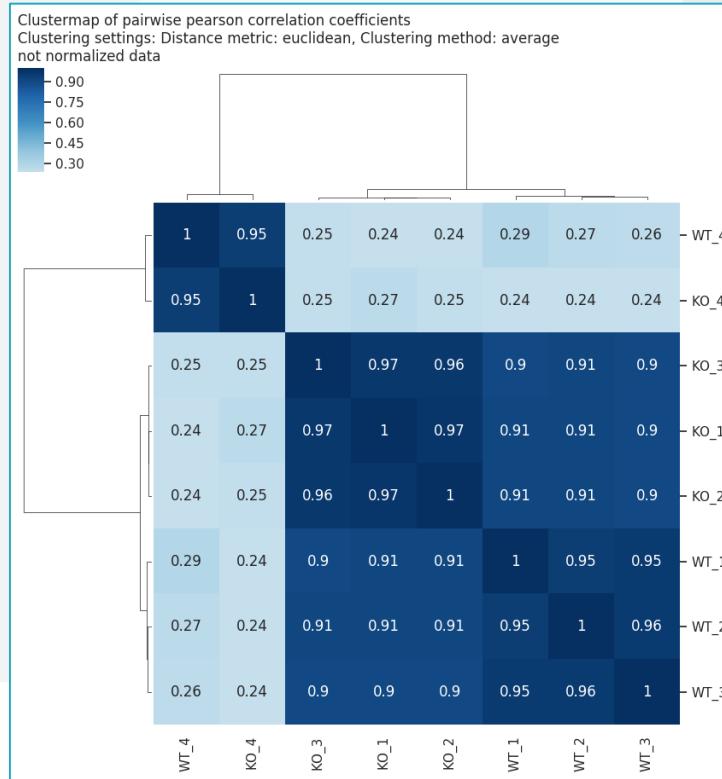
KO – loess (fast) normalization



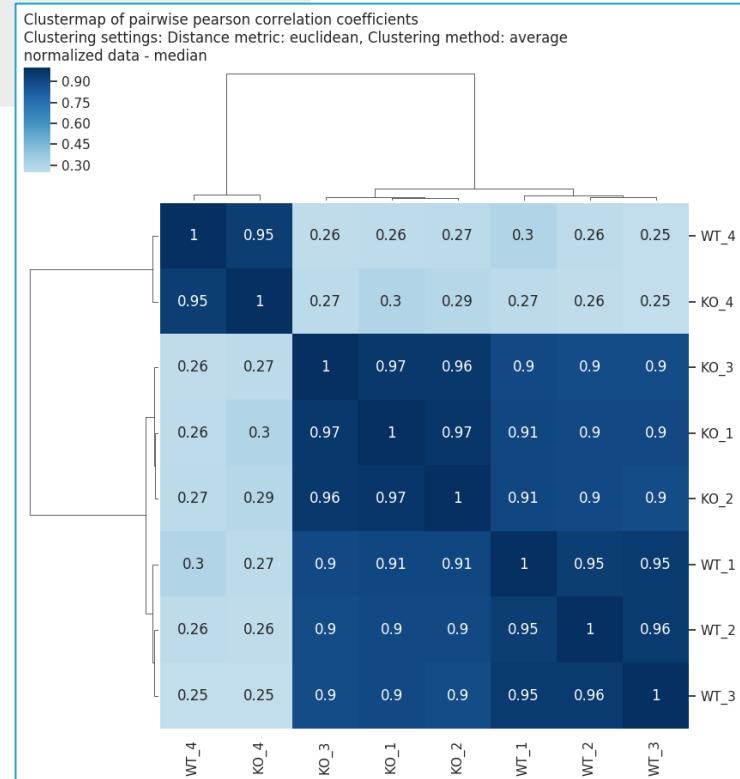
Normalization comparison – correlation clustermap



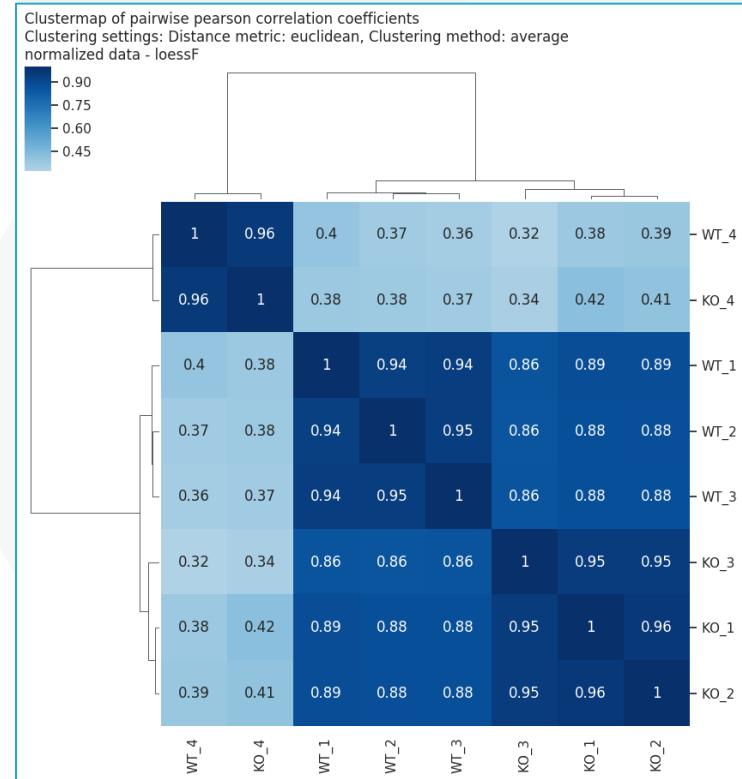
prior normalization



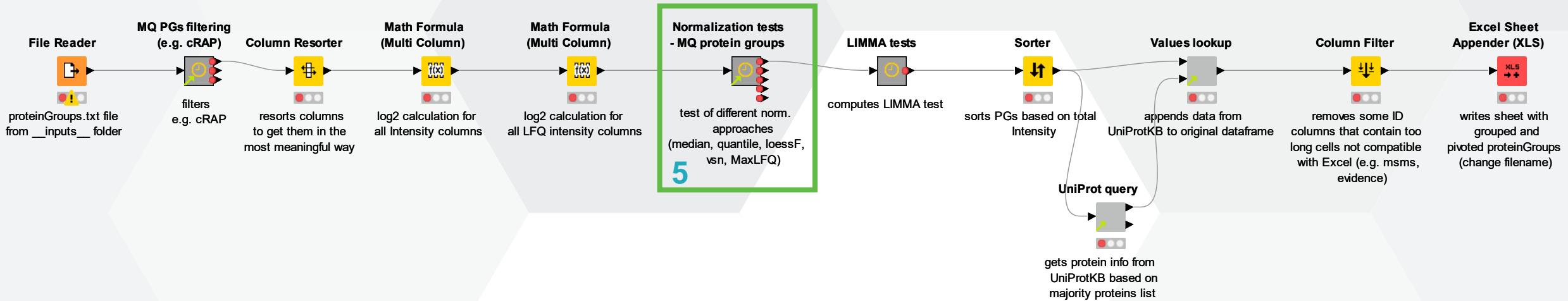
median normalization



loess (fast) normalization



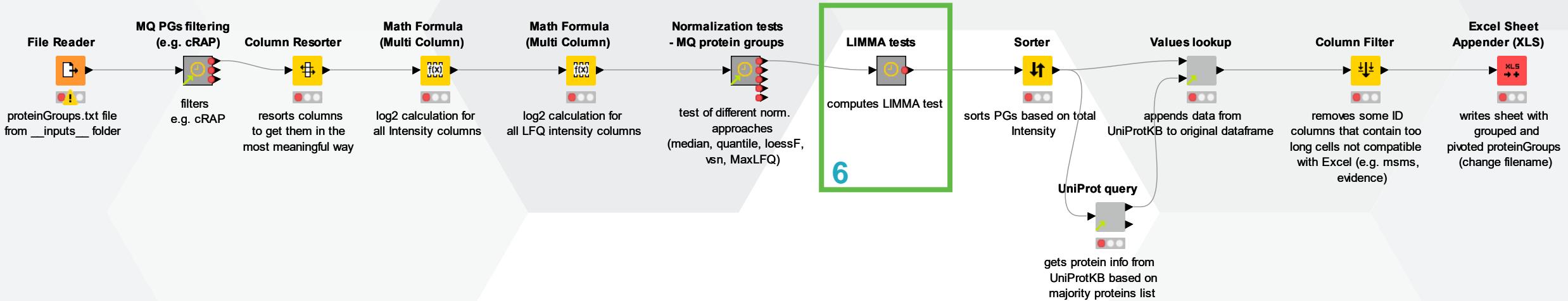
5. data normalization approach selection and data check



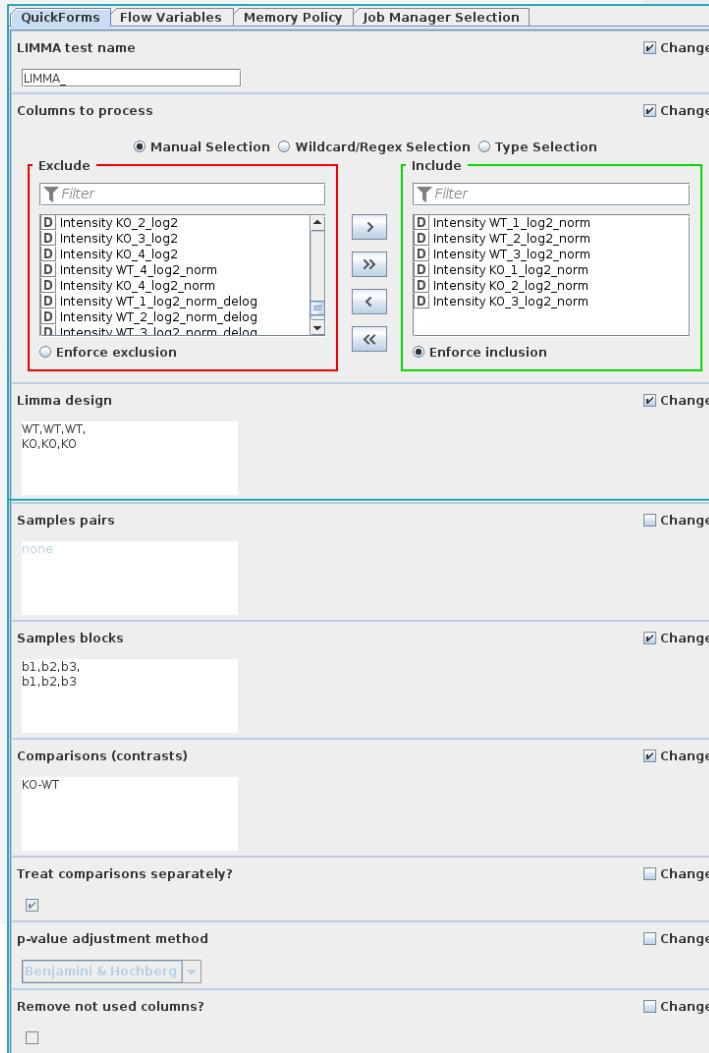
- selected outcomes of the normalization tests example

- 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

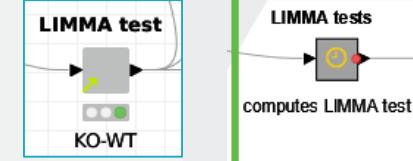
6. statistical evaluation



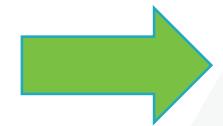
Metanodes examples – LIMMA test



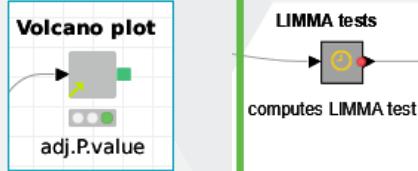
- initial statistical evaluation
 - proper design selection
 - paired design?
 - batches?
 - does not handle missing values on its own, preferable to deal with them beforehand
 - needs at least 1 value in all sample types
 - does not filter tested data based on e.g. number of numerical values in on sample 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.lo gFC	D	LIMMA_K_O-WT.Av eExpr	D	LIMMA_K_O-WT.t	D	LIMMA_K_O-WT.P. Value	D	▲ LIMM A_KO- WT....	D	LIM O-V
-9.733	26.548	-10.374	0	0	0	0	0	0	44.932		
9.14	27.739	9.742	0	0	0	0	0	0	38.907		
-9.336	26.101	-9.382	0	0	0	0	0	0	35.431		
-8.114	25.797	-8.649	0	0	0	0	0	0	29.38		
7.225	28.22	7.701	0	0	0	0	0	0	22.038		
6.991	27.612	7.452	0	0	0	0	0	0	20.246		
6.846	27.693	7.297	0	0	0	0	0	0	19.164		
-6.376	25.326	-6.797	0	0	0	0	0	0	15.82		
6.168	24.812	6.574	0	0	0	0	0	0	14.411		
9.211	24.801	6.366	0	0	0	0	0	0	12.244		
7.326	26.304	6.376	0	0	0	0	0	0	12.893		
-5.904	24.338	-6.293	0	0	0	0	0	0	12.699		
-5.893	28.173	-6.281	0	0	0	0	0	0	12.624		
-5.767	27.108	-6.147	0	0	0	0	0	0	11.835		
6.1	24.891	6.13	0	0	0	0	0	0	11.675		
5.539	26.877	5.904	0	0	0	0	0	0	10.448		
-5.535	23.023	-5.899	0	0	0	0	0	0	10.422		
-5.506	26.142	-5.869	0	0	0	0	0	0	10.255		
-5.761	25.997	-5.79	0	0	0	0	0	0	9.768		
5.396	26.941	5.751	0	0	0	0	0	0	9.605		
-5.4	25.554	-5.756	0	0	0	0	0	0	9.628		
9.915	26.444	5.73	0	0	0	0	0	0	8.31		
5.549	26.064	5.577	0	0	0	0	0	0	8.627		
-5.045	24.425	-5.377	0	0	0	0	0	0	7.632		
-4.973	24.28	-5.301	0	0	0	0	0	0	7.247		
-4.93	25.424	-5.255	0	0	0	0	0	0	7.014		
4.826	28.285	5.144	0	0	0	0	0	0	6.466		
-4.804	25.288	-5.121	0	0	0	0	0	0	6.356		
4.74	26.655	5.052	0	0	0	0	0	0	6.025		
5.615	24.714	4.887	0	0	0	0	0	0	5.149		
4.562	27.187	4.863	0	0	0	0	0	0	5.134		
-4.498	25.271	-4.794	0	0	0	0	0	0	4.82		
4.762	24.863	4.786	0	0	0	0	0	0	4.767		
5.146	24.374	4.478	0	0.001	0	0.001	0	0	3.381		
4.128	26.09	4.4	0	0.001	0	0.001	0	0	3.105		
4.049	25.65	4.316	0	0.001	0	0.001	0	0	2.755		
-6.219	26.166	-4.298	0	0.001	0	0.001	0	0	2.495		
7.255	24.117	4.193	0	0.002	0	0.002	0	0	1.885		
3.933	28.547	4.193	0	0.002	0	0.002	0	0	2.259		
-3.901	24.147	-4.158	0	0.002	0	0.002	0	0	2.121		
3.85	23.063	4.104	0	0.003	0	0.003	0	0	1.909		
6.996	24.445	4.043	0	0.004	0	0.004	0	0	1.366		



Metanodes examples – Volcano plot



QuickForms Flow Variables

Log fold change column: LIMMA_KO-WT.logFC

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

Apply -log10 on p-value column? (checked)

Zero p value replacement: 0.001

x axis (logFC) threshold: 1

y axis (p value) threshold: 0.05

Graphs subtitle:

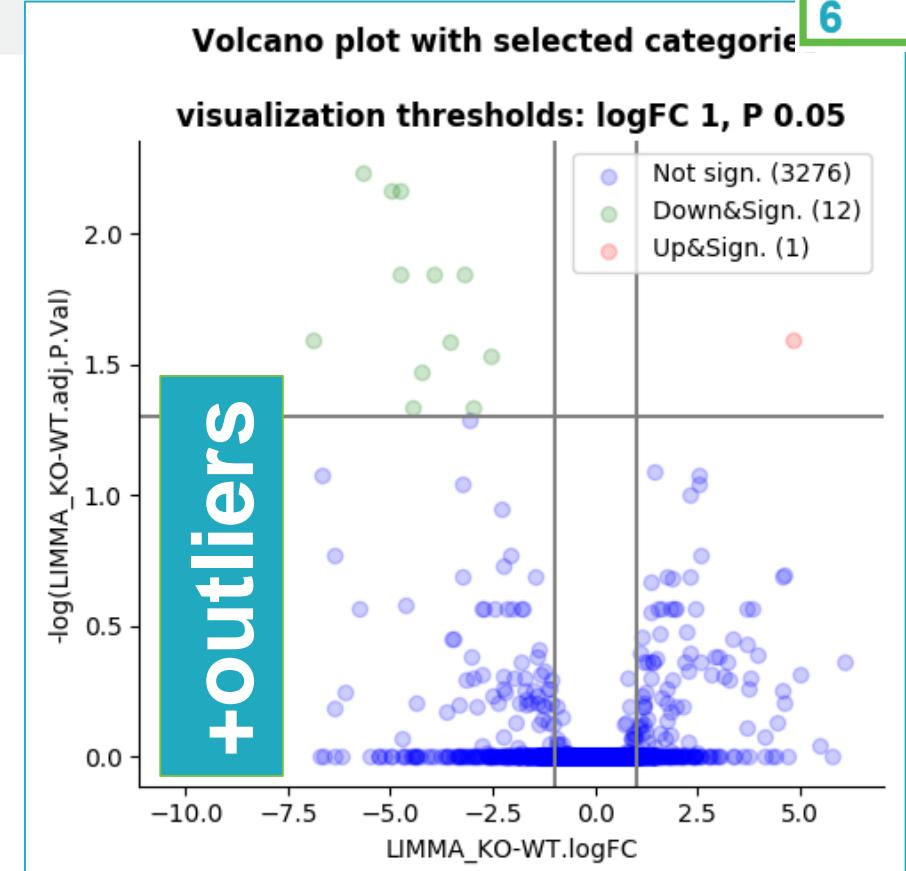
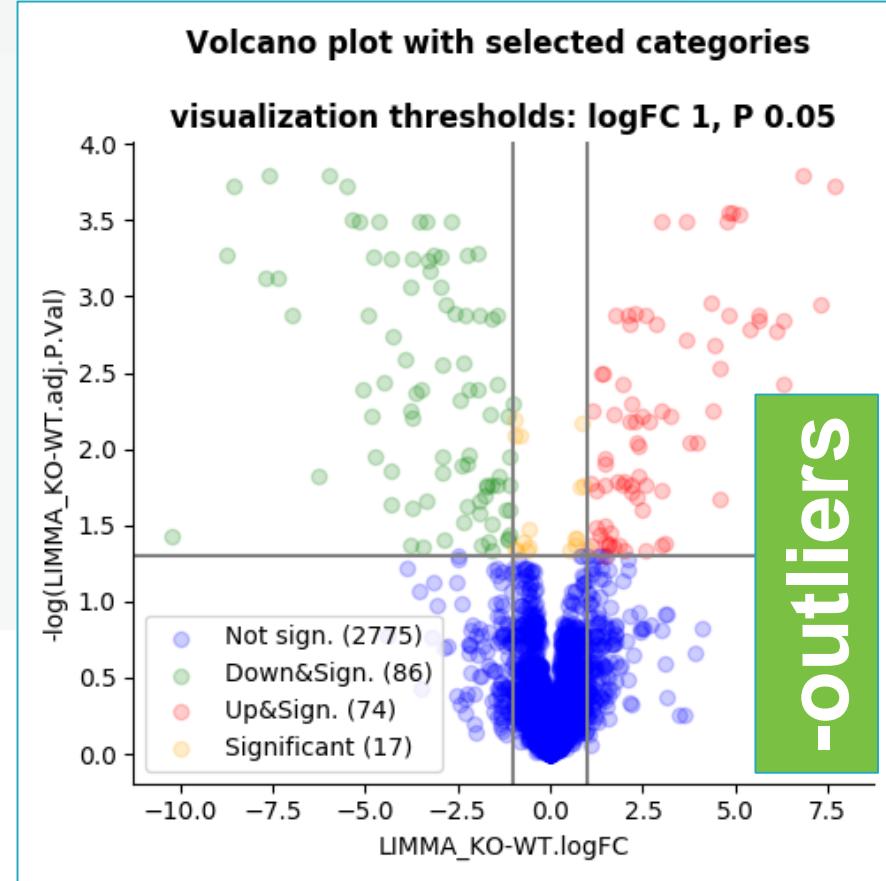
x axis label:

y axis label:

Manual x axis limits? (unchecked)

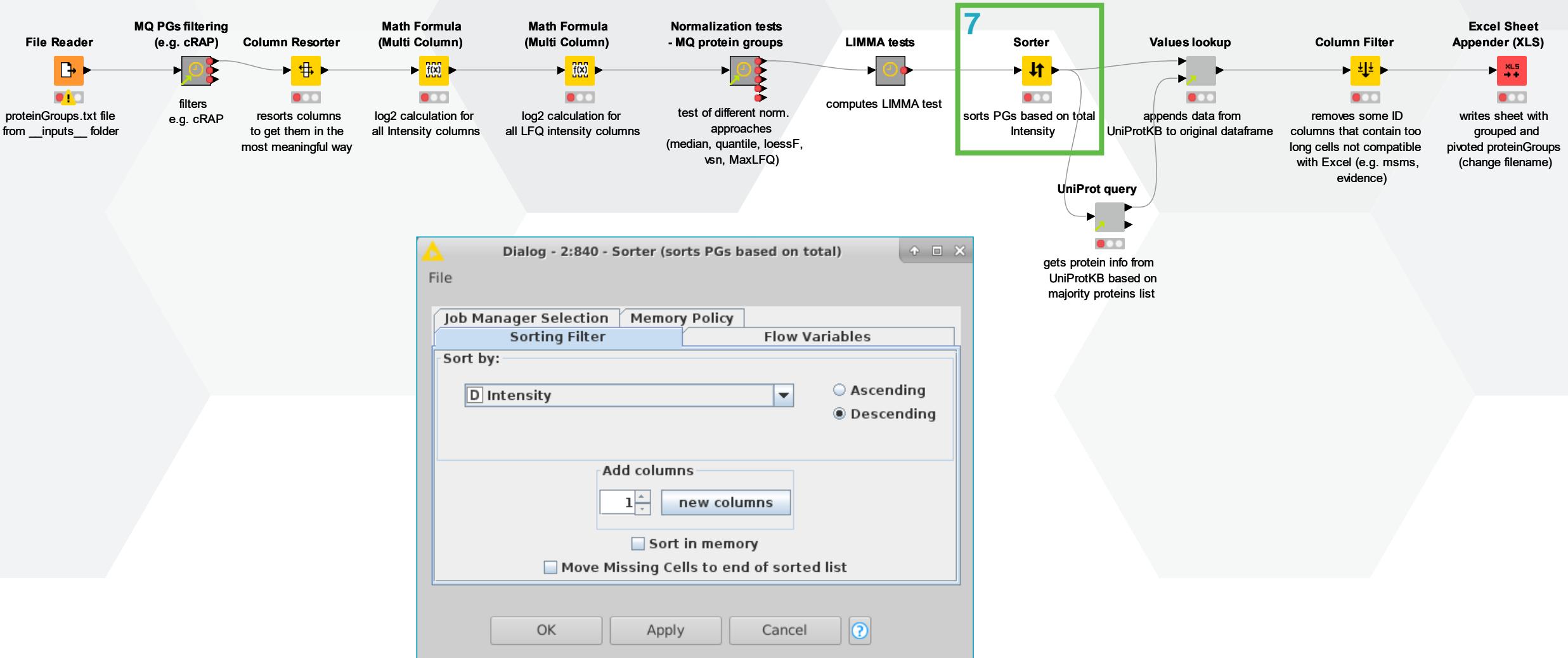
Manual y axis limits? (unchecked)

Manual y axis limits: -10:10

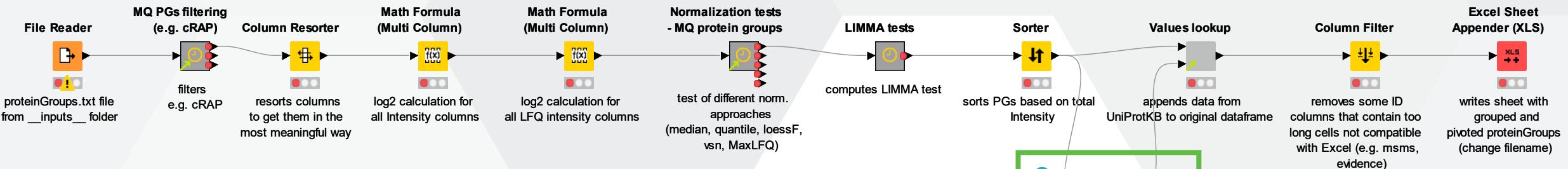


- visualize results of the statistical tests (logFC and P/adj.P values)
- categorize the results based on the specified threshold

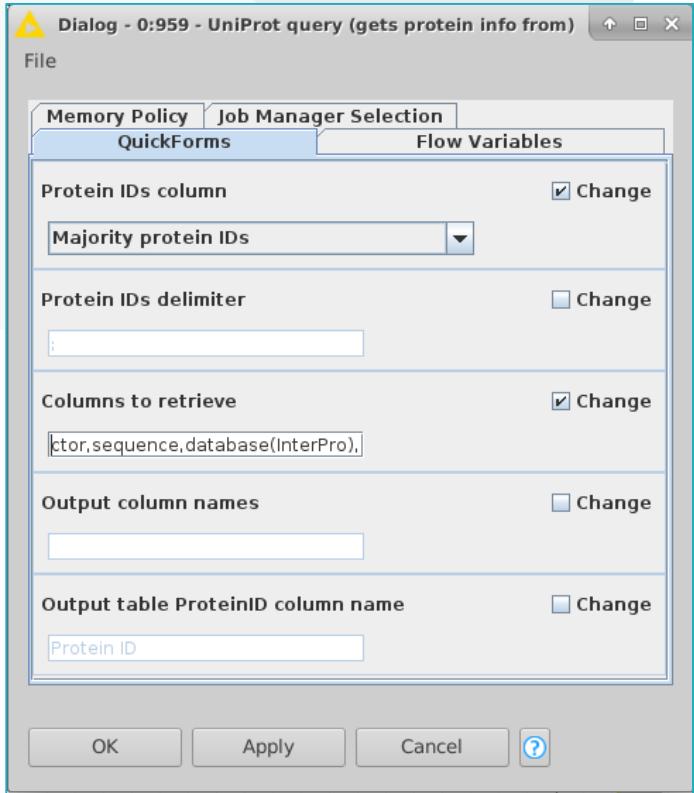
7. resorts the table (for reporting purposes only)



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



8 UniProt query
gets protein info from UniProtKB based on majority proteins list

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]			GO-0009451	PPR family; PCMP-E subfamily
Row1	THIUMP03	Homo sapiens (Human)	cytosol [GO:0005829]; RNA methylation [GO:0030488]	RNA binding...; cytosol [GO:0003724]			GO-0003724	Methyltransferase superfamily
Row2	PADI_PRC6A_A13p51260 F24M...	Arabidopsis thaliana (Mouse-ear cress)	cytoplasm [GO:0005737]; cyto...; proteasomal protein catabolic...	proteasomal protein catabolic...; endopept...; cytoplas...; cytosol [GO:000346...			GO-0005050	Peptidase T1A family
Row3	GPT1_GFAT_GFPT	Homo sapiens (Human)	cytosol [GO:0005829]; extra...; chromosome [GO:0005694]; f...	chromosome [GO:0005694]; f...; double-strand break repair vi...; FHA dom...; chromos...; membrane [GO:0016202]; nu...			GO-000372...	
Row4	MDC1_KIAA0170_NFBDI	Homo sapiens (Human)	cytosol [GO:0005829]	RNA metabolic process [GO:0...			GO-000563...	
Row5	DDX24	Homo sapiens (Human)	cytoplasm [GO:0005737]; cyto...; extracellular exosome [GO:00...	cell surface receptor signalin...; enzym...; cytoplas...; cytosol [GO:0003724]; cell wall organization [GO:00...			GO-000372...	
Row6	PRMT1_HMT2_HMT1L2_I1R1B4	Homo sapiens (Human)	cytoplasm [GO:0005737]; cyto...; extracellular exosome [GO:00...	amino acid transport [GO:00...			GO-000372...	
Row7	SLC1AS1_ASCT2_M7V1_RDR_RDRC	Homo sapiens (Human)	cytoplasm [GO:0005737]; cyto...; catalytic step 2 spliceosome [...	amino acid transport [GO:00...			GO-000372...	
Row8	AXS2_4t1g08200 T23G18.6	Arabidopsis thaliana (Mouse-ear cress)	apoplast [GO:0048046]; cytop...; ficolin-1-rich granule membra...	apoplast [GO:0048046]; cytop...; ficolin-1-rich granule membra...			GO-000382...	NAD(P)-dependent epimer...
Row9	POLD3_KIAA0039	Homo sapiens (Human)	mitochondrial organization [...	mitochondrial organization [...			GO-000552...	
Row10	SLC3A2_MDU1	Homo sapiens (Human)	amino acid transport complex...	amino acid transport [GO:00...			GO-000372...	AAATPase family
Row11	CLPP5_NCLP1_NCLP5_A1tg02...	Arabidopsis thaliana (Mouse-ear cress)	chloroplast [GO:0009507]; ch...	protein quality control for misf...; ATPase [...			GO-000372...	SLC3A transporter family
Row12	EIF4EBP2	Homo sapiens (Human)	cytoplasm [GO:0005737]; pos...	CAMP-mediated signalling [GO:0...			GO-000573...	EIF4E-binding protein family
Row13	AT5g45160_K18C1.4	Arabidopsis thaliana (Mouse-ear cress)	cytoplasm [GO:0005737]; endo...	endoplasmic reticulum [GO:00...			GO-000392...	TRAFAC class dynamin-like...
Row14	HNRNPU_C1orf199_HNRNU_SAF...	Homo sapiens (Human)	catalytic step 2 spliceosome [...	adaptive thermogenesis [GO:0...			GO-000102...	
Row15	ESD	Homo sapiens (Human)	cytoplasmic vesicle [GO:0003...	actin bin...; catalytic...; cytoplas...; cytosol [GO:0003724];...			GO-000578...	Esterase D family
Row16	ZNF638_NP220_ZFME	Homo sapiens (Human)	cytoplasm [GO:0005737]; intr...	formaldehyde catabolic proce...; cytoplas...; cytoplas...; cytosol [GO:0003724];...			GO-000369...	
Row17	POLD3_KIAA0039	Homo sapiens (Human)	cytoplasm [GO:0005737]; del...	DNA damage response, detec...; cytoplas...; cytosol [GO:0003724];...			GO-000072...	
Row18	NTR2_NTRA_A1tg17420 F5j6.1B	Arabidopsis thaliana (Mouse-ear cress)	mitochondrial matrix [GO:00...	removal of superoxide radical...; thioredox...; mitochon...			GO-000479...	Class-II pyridine nucleotide...
Row19	ACCI_EM822_QK_PAS3_A1tg3...	Arabidopsis thaliana (Mouse-ear cress)	cytosol [GO:0005829]; nucleo...	embryo development ending ...; acetyl co...; cytosol [GO:0003724];...			GO-000398...	FPP/GPP synthase family
Row20	FS1_P51g47770_MCA23.9	Arabidopsis thaliana (Mouse-ear cress)	cytoplasm [GO:0005737]; cyto...	cholesterol biosynthetic proc...; dimethyla...; cytosol [GO:0003724];...			GO-000416...	Mg-chelatase subunits Dif...
Row21	CHLD_ALB1_PDE166_A1tg052...	Arabidopsis thaliana (Mouse-ear cress)	chloroplast [GO:0009507]; cyto...	chlorophyll biosynthetic proc...; chloroph...; chloroph...; cytosol [GO:0003724];...			GO-000552...	
Row22	NC_2014768	Homo sapiens (Human)	cell body [GO:0044297]; chag...	binding of sperm to zona pell...; ATP bindi...; cell body [GO:0044297];...			GO-000219...	TCP-1 chaperonin family
Row23	LCT3_CCTG_TRIC5	Homo sapiens (Human)	contractile ring [GO:0070938];...	cell division [GO:0051301]; mi...; identical ...; contractile...; actin bin...; cytosol [GO:0003724];...			GO-00002...	MAP65/ASE1 family
Row24	PROK2AP1	Homo sapiens (Human)	extracellular patch [GO:00...	actin filament capping protein...; actin fil...; actin bind...; cytosol [GO:0003724];...			GO-000271...	Protein kinase C epsilon isoform
Row25	CAP21	Homo sapiens (Human)	clathrin-coated vesicle [GO:00...	neutrophil degranulation [GO:0...			GO-000013...	SCAMP family
Row26	SAMP1_SCAMP	Homo sapiens (Human)	cytosol [GO:0005829]; cytoplas...	cytoplasmic translation [GO:0...			GO-000218...	Eukaryotic ribosomal protein L1 family
Row27	PPIC1_A1tg47700_MCA23.2	Arabidopsis thaliana (Mouse-ear cress)	cytosol [GO:0005829]; beta-al...	cellulose catabolic process [GO:0...			GO-000582...	Glycosyl hydrolase 1 family
Row28	BGLU42_A1tg36990_MFL18.1	Arabidopsis thaliana (Mouse-ear cress)	cytosol [GO:0005829]; eukary...	formation of cytoplasmic tran...			GO-000173...	ElF-3 subunit L family
Row29	At5g25754_At5g25757	Arabidopsis thaliana (Mouse-ear cress)	cytosol [GO:0005829]; eukary...	mRNA bin...; mRNA bin...; actin bin...; cytosol [GO:0003724];...			GO-000013...	
Row30	NUM1_NMP22_NUMA	Homo sapiens (Human)	cell cortex [GO:000938]; cell ...	anastral spindle assembly [GO:0...			GO-000013...	
Row31	UBE2D3_UBC52_UBCH5C	Homo sapiens (Human)	cytosol [GO:0005829]; endope...	disorder...; cell conte...; cytosol [GO:0003724];...			GO-000012...	
Row32	EEF2_EF2	Homo sapiens (Human)	cytosol [GO:0005829]; nucleo...	apoptotic process [GO:00069...]; ATP bindi...; nucleus [GO:0005634];...			GO-000012...	Ubiquitin-conjugating enzyme E2
Row33	UAP1_SPAG2	Homo sapiens (Human)	aggresome [GO:0016235]; cyt...	aging [GO:0007568]; cellular r...; aggresome [GO:0016235];...			GO-000203...	TRAFAC class translation factor
Row34	PSMC4_MP224_TB7P	Homo sapiens (Human)	cytosol [GO:0005829]; nucleo...	UDP-N-acetylglucosamine bios...; carboxy...; cytosol [GO:0003724];...			GO-000397...	UDPG type 1 family
Row35	MARCKS_C1AK1_EM2K	Homo sapiens (Human)	cytosol [GO:0005829]; cytosol [...	anaphase-promoting complex...; ATPase a...; cytosol [GO:0003724];...			GO-000016...	AAA ATPase family
Row36	SRM2_KIAA0324_SRL300_SR...	Homo sapiens (Human)	cytoplasm [GO:0005737]; cyto...	intracellular signal transducti...; ATP bindi...; cytosol [GO:0003724];...			GO-000016...	Protein kinase superfamily...
Row37	SRM2_KIAA0324_SRL300_SR...	Homo sapiens (Human)	Cajal body [GO:0015030]; cat...	mRNA splicing, via spliceosom...; C2H2 zinc...; Cajal bod...; cytosol [GO:0003724];...			GO-000039...	CWC21 family
Row38	A1tg2820_E0K301.2	Arabidopsis thaliana (Mouse-ear cress)	cytosol [GO:0005829]; nucleo...	carboxyl...; cytosol [GO:0003724];...			GO-000561...	

7. get

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

(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)

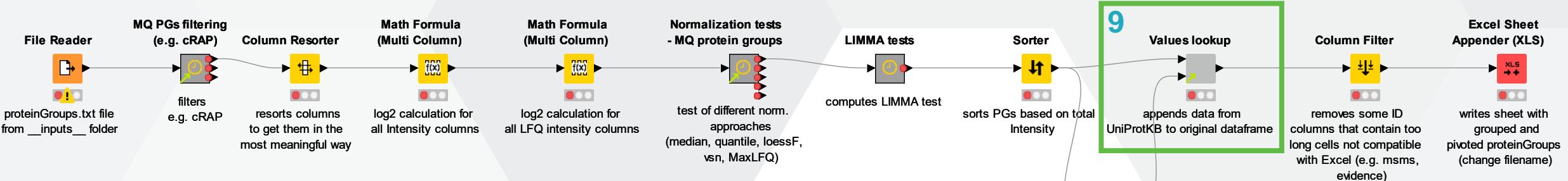
Excel Sheet Appender (XLS)


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

9. adds annotation info into the original table



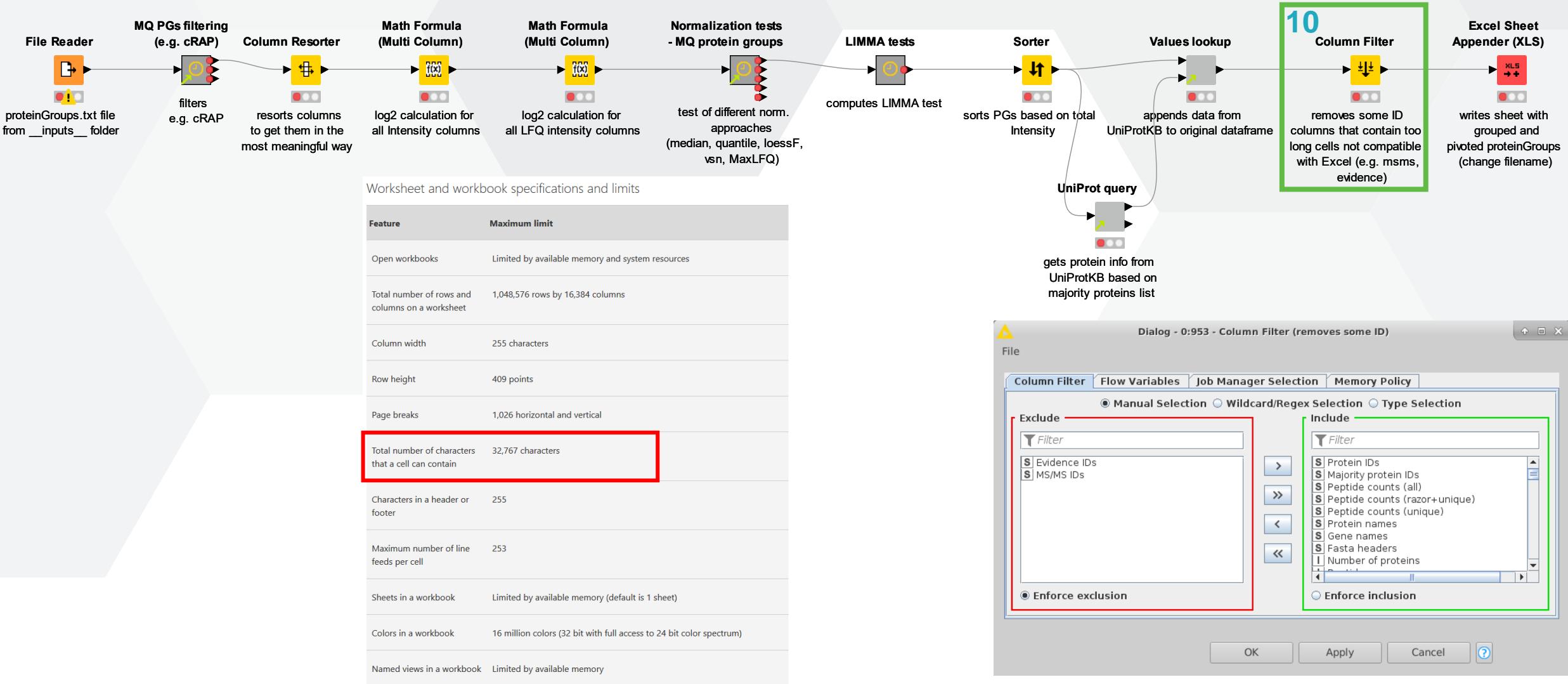
Values lookup metanode utilization

- 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)

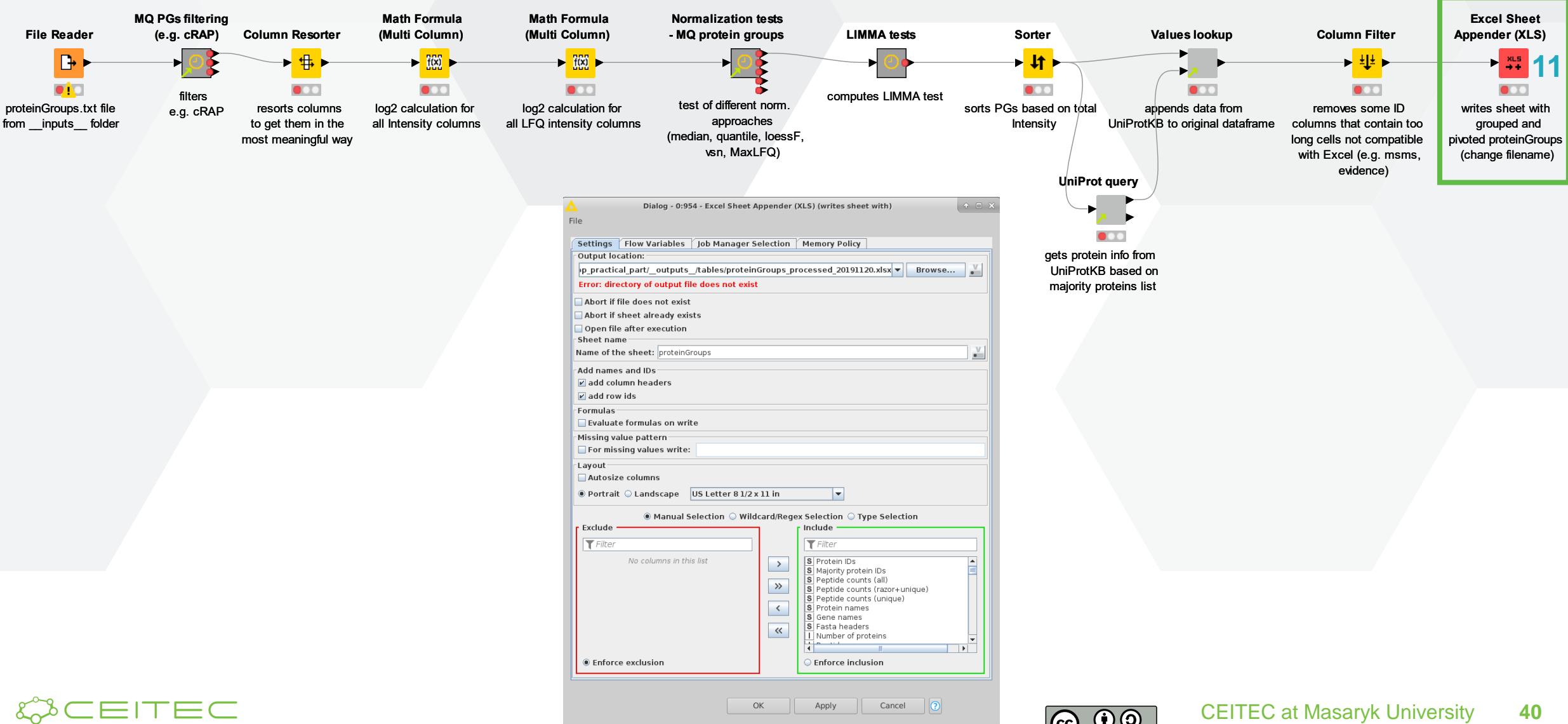
Majority protein IDs
Majority protein IDs
Top Input table IDs column
Top Input table IDs delimiter
Bottom Input table ID column
Protein ID
Columns to add from bottom input table
Manual Selection Wildcard/Regex Selection Type Selection
Exclude
Include
Added columns prefix
UniProt_
Added columns suffix
Looked up values delimiter
Remove empty strings?
Remove duplicates?
Remove not used columns?

S Majority protein IDs	S UniProt_Protein_families	S UniProt_Interactions_with	S UniProt_Cross-references	S UniProt_Cross-references_inAct
A0A096LP16	XEVFTSSLPELVKEV/LEF...	IPR029062; IP...	PF04003; &...	A0A096LP16;
A0A06YAY00;P09Y3B3	MPPRFSGAORWAAVAGR...	IPR009038; IP...	PF01105; & ...	& 09Y3B3;
A0A06YVHL;P59780	MSRIVHDGALRNKAV/QSV...	IPR022775; IP...	PF01217; IP...	& P59780;
A0A06YL6;P18621	MVRYSLDPENPTKSCKSR...	IPR001063; IP...	PF00237; & ...	A0A06YL6...
A0A084LV8;P9N055	MGOS5GRSRHOKRAAOA...	IPR007109; IP...	PF00001; PE...	A0A084LV8...
A0A084J269;P13506	IPR002453; IP...	PF00001; PE...	PF00001; PE...	& 084J269;
A0A084E25;P15269	MKFAVYRFNSNLGTVYRRG...	IPR020472; IP...	PF04003; PE...	& 015269;
A0A01BG0U03;P07339	MKHLWFVFLLVAAPRWL...	IPR007110; IP...	PF07686; & ...	& & & ...
A0A01J1QV1	MFLVLFQEMDKPCGCPGV...	IPR001461; IP...	PF07966; PE...	& & & ...
A0A01BG0U03;P07339	MOPPSLPLAULCLAAVAS...	IPR001461; IP...	PF07966; PE...	& P07339;
A0A01W2PN4	MSLHJQELPITPCATHWNE...	IPR030140; IP...	PF00400; & ...	& 09W2PN4;
A0A01W2PQ55	IPR015797; IP...	PF00293; PE...	PF00293; PE...	& 015269;
A0A01W2PB88;P9H0W9	MLMPKKRNRAIVPDLRQW...	IPR0015021; IP...	PF08925; PE...	& 09H0W9;
A0A01V8T1	MIEGSEPAVVAH;P0GEAES...	IPR032420; IP...	PF16191; IP...	A0AVT1;
A0FGR8	Q9BSJ5; A...	IPR000008; IP...	PF0068; PE...	& 0AFGR8;
A1L0T0	MTANRDAALSLSHRHPG...	IPR029055; IP...	PF02775; A...	& 1L0T0;
A3KN93	MPTEAAAQFAPGSILPFSIL...	IPR027417; IP...	PF1372; PE...	A3KN93;
A4D1E9	IPR008379; Q5P59...	IPR031167; IP...	PF01018; PE...	A4D1E9;
A5VKK6	MHCVSCA4PRKYNIGKEV...	IPR03642; IP...	PF13344; A...	A5VKK6;
A6NDG6	MLDLSLSSLSLQ5VLDN...	IPR007126; IP...	PF16415; PE...	A6NDG6;
A6HFQ2;A0A1B0GVM2	MAMAEAAFGDAR;P0...	IPR03642; IP...	PF13344; A...	A6HFQ2; &
A6NHQ2	MATIAAEAAFEALMDQ;TC...	IPR000692; IP...	PF01269; A...	A6NHQ2;
A6HNP9	MKSAASSRGGGRG...	IPR000692; IP...	PF01269; A...	A6HNP9;
A6NJT8	MAAADGGPGEGASV;GTE...	IPR03690; IP...	PF06470; A...	A6HNP9;
A6NK77;Q7Z3J3	MRLRVYFCRMVKECLCSW...	IPR02903; IP...	PF01793; A...	A6NK77; & Q...
A9C934;Q9H6A1	MSSLAQEMDQ;P0...	IPR032023; IP...	PF01465; PE...	A9C934; & Q...
A9hMv4	MOEPSEREYKQ;P0...	IPR015624; IP...	PF02854; A...	A9hMv4;
A9UHW6	IPR03673; IP...	IPR03673; IP...	PF01624; IP...	A9UHW6;
B0I1T2	MOEPSEPEYKQ;P0...	IPR036961; IP...	PF00638; PE...	B0I1T2;
B4D1N1;Q9UBK3	MSSLAQEMDQ;P0...	IPR036961; IP...	PF00638; PE...	B4D1N1; & Q...
P0C6G0;B7ZAQ6	MPLAAPRP1WGP;CLGLRA...	IPR02030; IP...	PF00535; PE...	B4D1N1; & Q...
C9JAW5;P09Y241	MSTDTGVSILPSEYEDQGS...	IPR007667; & ...	PF04588; & ...	C9JAW5; & Q...
E7ENK8	MPLSPDLRWPWTSSLVDAAL...	IPR035959; IP...	PF00664; PE...	E7ENK8;
E7EVH7;Q07866	MLPCAAGARGRGAMV...	IPR018796; IP...	PF0231; PE...	& Q07866;
E9PVAV3	MLPEGETVPAV...	IPR016641; IP...	PF01849; PE...	E9PVAV3;
E9PL57;Q15843	MLUKXTLTKIEIDIEPT...	IPR03642; IP...	PF12889; PE...	E9PL57; & Q...
E9PLD3;E9PRG8	MCAFPGGKINPRTTELKK...	IPR037691; & ...	PF17719; & ...	E9PLD3; & E...
E9PLNB8;Q9B0B6	MGSTWGPSGWVVRALC...	IPR009003; IP...	PF00089; PE...	E9PLNB8; & ...
F5H5P2;P12694	MDSPLSRELQ;P0...	IPR034616; IP...	PF00076; & ...	& P12694;
F8W031;Q9Y2B0	MKGWGWLLALLGALLTA...	IPR016142; IP...	PF11938; & ...	& Q9Y2B0;
H0YAA0	XYDYYGDYHNVRYGYE...	IPR019375; IP...	PF10231; PE...	H0YAA0;
H0YC42	MDCREMDLYEDQSPFD...	IPR00630; IP...	PF10246; PE...	H0YC42;
H3BN98	MAEGDNURSTNLAAETAS...	IPR00630; IP...	PF00410; A...	H3BN98;
I3LE03;Q9Y2R5	MEEEEMNGS0MVRMSQ...	IPR01909; IP...	PF01352; PE...	I3LE03; & Q...
K7EQ62	SYSFDPYSDMGPRLFSH...	IPR034635; IP...	PF01352; PE...	K7EQ62;
K7ERE1;Q8N2U0	MAGFAA4FRLGALGAA...	IPR006696; & ...	PF04241; & ...	& Q8N2U0;
L7N2F9;Q15836;P63027	MSATAATAPPAPAAGEG...	IPR001388; IP...	PF00957; & ...	& Q15836; ...
MOQYTO	XVMVGLDQSKMNCDR...	IPR029045; IP...	PF00378; MOQYTO;	
MQQZDB	MFAPAVGLGLVLWPLMA...	IPR015819; IP...	PF09172; PF...	

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	Rep prots	Qual Data	PG Score, Q	Peptides	Data	Normalized PG Intensities - log2 Trans	BAO Data	BAO values	Ratio Cat.E	thresholds	Annotation	Protein families	
total	3522	3751	2	2	PGs present	thresholds	Low	Middle	High	KO-WT	KO-WT	KO-WT	Y	
shown	132	136	Protein	Identifications	Peptides	#replicates	13.31	24.73	31.31	KO-WT	KO-WT	KO-WT	Y	
PG I	Access.	Description	Summa	Y1	KO	WT	WT	WT	KO	KO	KO	adj.P.	Y	
679	348	P09326 CD46 antigen OS=Homo sapiens GN=CD46 PE=1 SV=2	YYYY YYYY	3	3		24.33	24.86	24.70	27.08	26.34	27.37	0.02	N
699	3045	Q86U42 Polyadenylate-binding protein 2 OS=Homo sapiens GN=	YYYY YYYY	3	3		23.23	23.61	-	26.81	26.88	26.92	0.03	N
721	1708	P43702 Proteasome subunit beta-type-3 OS=Homo sapiens GN=	YYYY YYYY	3	3		25.08	24.63	24.35	27.13	26.86	27.06	1.01	Peptidase T1B family
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		TRAFAC class OBG-HIX-like
754	1311	P26062 Proteasome subunit beta-type-6 OS=Homo sapiens GN=	YYYY YYYY	3	3		-	-	-	20.92	21.43	26.66	0.000	Peptidase T1B family
759	2597	Transferrin OS=Homo sapiens GN=PGM1P	YYYY YYYY	3	3		23.41	23.56	23.67	26.40	26.58	26.63		Translin family
817	1871	P55145 Mesencephalic astrocyte-derived neurotrophic factor C OS=Homo sapiens GN=MNDP	YYYY YYYY	3	3		21.62	-	22.68	26.44	26.44	26.76		ARMET family
819	1475	Phosphoglucomutase-1 OS=Homo sapiens GN=PGM1P	YYYY YYYY	3	3		-	21.70	22.48	26.45	26.61	26.47		Phosphoketose mutase family
847	598	P75400 Pre-mRNA-processing factor 10 homolog A OS=Homo sapiens	YYYY YYYY	3	3		-	-	-	22.24	21.32	-		PRPF40 family
915	2193	P98111 Rho GTPase-activating protein 4 OS=Homo sapiens GN=	NNNN YYYY	0	2		24.46	23.56	24.50	26.08	26.22	26.06		N
934	3827	Translocation protein SEC62 OS=Homo sapiens GN=S1	YYYY YYYY	3	3		-	-	-	27.01	25.70	25.43		SEC62 family
939	3877	Q3BPX5 Actin-related protein 2/3 complex subunit 5-like protein	YYYY YYYY	3	3		24.34	25.01	24.74	26.15	26.03	25.37		APRCS family
985	2234	Q86257 Quinone oxidoreductase OS=Homo sapiens GN=CRY2	YYYY YYYY	3	3		24.67	24.32	24.60	26.12	26.05	26.22		Zinc-containing alcohol dehyd
1022	2031	P63165 Small ubiquitin-related modifier 1 OS=Homo sapiens GN=	YYYY YYYY	3	3		23.26	23.90	24.05	26.11	25.76	26.24		Ubiquitin family, SUMO subfa
1026	5319	Q9Y636 Chloride intracellular channel protein 4 OS=Homo sapiens	YYYY YYYY	3	3		23.32	22.91	23.18	25.71	25.87	25.86		Chloride channel CLIC family
1065	2526	Condensin complex subunit 2 OS=Homo sapiens GN=SEC63	YYYY YYYY	3	3		23.64	22.78	23.54	25.54	24.59	26.04		CND2 (condensin subunit 2) f
1088	1076	P14735 Insulin-degrading enzyme OS=Homo sapiens GN=IDE PI	YYYY YYYY	3	3		24.21	24.52	24.55	25.81	25.83	25.89		Peptidase M16 family
1092	1111	P16402 Histone H1.3 OS=Homo sapiens GN=HIST1H1D PE=1 SV=2	YYYY YYYY	3	3		23.17	-	-	25.53	24.46	25.69		Histone H1/H5 family
1111	5323	Q9Y6A9 Signal peptidase complex subunit 1 OS=Homo sapiens GN=	YYYY YYYY	3	3		-	-	-	25.36	25.94	25.41		SPC51 family
1120	2325	Q27675 Secernin-1 OS=Homo sapiens GN=SCRN1 PE=1 SV=2	YYYY YYYY	3	3		-	-	-	18.32	20.02	20.55		Peptidase C63 family, Secerni
1158	653	Q9T964 ATP synthase subunit g, mitochondrial OS=Homo sapiens	YYYY YYYY	3	3		20.89	20.53	20.51	24.37	24.50	25.52		ATPase g subunit family
1200	3292	Q8NI36 WD repeat-containing protein 36 OS=Homo sapiens GN=	YYYY YYYY	3	3		22.31	21.78	21.90	25.36	25.19	25.73		N
1205	632	Q75717 WD repeat and HMG-box DNA-binding protein 1 OS=Homo sapiens	YYYY YYYY	3	3		22.31	21.78	21.90	25.36	25.19	25.73		BRX1 family
1268	3345	Q8DTN6 Ribosome biogenesis protein BRX1 homolog OS=Hom	YYYY YYYY	3	3		22.11	23.33	22.46	25.14	25.21	24.63		Protein kinase superfamily, ST
1291	714	P04804 Serine/threonine-protein kinase 10 OS=Homo sapiens G	YYYY YYYY	3	3		22.11	22.34	22.46	24.70	24.85	24.85		N
1319	2328	Q12788 Transducin beta-like protein 3 OS=Homo sapiens GN=NET	YYYY YYYY	3	3		22.01	22.87	23.08	25.18	25.11	25.41		Pyridoxine kinase family
1344	233	P00764 Pyridoxal kinase OS=Homo sapiens GN=PDXK PE=1 SV=2	YYYY YYYY	3	3		22.01	22.34	22.46	25.14	25.21	24.63		Pyridoxamine 5'-phosphate oxi
1392	3315	Q9BT23 LIM domain-containing protein 2 OS=Homo sapiens GN=	YYYY YYYY	3	3		-	13.68	13.83	24.33	24.86	24.85		PSMG2 family
1418	3640	Q96U07 Proteasome assembly chaperone 2 OS=Homo sapiens C	YYYY YYYY	3	3		-	-	-	24.78	24.38	24.32		Small GTPase superfamily, Art
1437	2160	P84085 ADP-ribosylation factor 5 OS=Homo sapiens GN=ARF5	YYYY YYYY	3	3		21.44	22.08	21.70	25.20	25.81	25.27		EMP24/GP25L family & EMP;
1450	1	DAA067YY Protein TMED7-1C/TAM2 OS=Homo sapiens GN=TMED	YYYY YYYY	2	3		-	-	-	25.01	24.64	24.58		Pyridoxamine 5'-phosphate oxi
1466	4666	Q9WV53 Pyridoxine 5-phosphate oxidase OS=Homo sapiens GN=	YYYY YYYY	3	3		-	-	-	24.36	25.00	24.64		N
1472	3711	Q96KRI Zinc finger RNA-binding protein OS=Homo sapiens GN=	YYYY YYYY	3	3		-	-	-	24.24	24.75	24.80		TOR1AI family
1509	2828	Q5TJY8 Torsin-1A-interacting protein OS=Homo sapiens GN=1	YYYY YYYY	3	3		23.02	22.96	-	24.58	24.39	24.30		YY transcription factor family
1597	1257	P25430 Transcriptional repressor pro <i>cyclin Y11</i> OS=Homo sapiens	YYYY YYYY	3	3		20.77	-	20.78	24.39	24.47	24.49		Peptidase M48A family
1614	642	Q5T844 CAAZ prenyl protease 1 homolog OS=Homo sapiens G	YYYY YYYY	3	3		23.72	22.39	22.77	24.83	24.57	24.57		Mitochondrion-specific ribos
1640	3712	Q96EY7 Pentatricopeptide repeat domain-containing protein 3,	YYYY YYYY	3	3		22.47	22.68	23.43	24.84	24.23	24.74		SnRNP Sm protein family
1660	2044	P62310 U6 snRNA-associated Sm-like protein LSM3 OS=Homo	YNNN YYYY	2	3		-	-	-	24.46	24.21	24.21		TMEM372/zigma-2 receptor f:
1724	2815	Q5BFJ2 Sigma intracellular receptor 2 OS=Homo sapiens GN=TI	YYYY YYYY	2	3		21.76	22.34	22.57	23.32	24.13	24.58		Cytidine and deoxyxycytidine
1748	4216	Q9GZXT Single-stranded DNA cytosine deaminase OS=Homo sapi	YYYY YYYY	2	3		22.71	22.87	23.08	24.36	24.65	24.08		Activator 1 small subunit fam
1772	1517	P40337 Replication factor C subunit 5 OS=Homo sapiens GN=F	YYYY YYYY	3	3		-	-	21.85	23.33	24.05	24.17		PRPB8 family
1811	3250	Q8NAV1 Pre-mRNA-splicing factor 38 OS=Homo sapiens GN=	NNNN YYYY	0	3		22.57	22.02	-	24.12	24.11	24.02		RNA polymerase betta chain f:
1821	4316	Q3H9T6 DNA-directed RNA polymerase I subunit RPA2 OS=Ho	YNNN YYYY	1	2		-	-	-	25.17	24.74	-		N
1822	2834	Q5R3T2 Rab GTPase-activating protein 1-like OS=Homo sapien	NNNN YYYY	0	3		21.33	-	21.89	24.33	23.31	23.89		Annexin family
1834	355	P09525 Annexin A4 OS=Homo sapiens GN=ANXA4 PE=1 SV=4	YYYY YYYY	1	3		19.44	-	20.38	23.63	24.30	23.64		Synaptophysin/synaptobrevin
1852	2630	Q16563 Synaptophysin-like protein 1 OS=Homo sapiens GN=SY	NNNN YYYY	0	3		20.35	21.28	22.02	23.26	24.06	24.32		Complex I LYR family
1869	2845	Q5TF6F2 Ubiquitin-associated protein 2 OS=Homo sapiens GN=	YYYY YYYY	3	3		21.74	-	21.61	23.83	23.78	23.78		Complex I NDUF10 subunit I
1880	2857	Q5USX50 Complex III assembly factor L1 RMT1 OS=Homo sapiens	YYYY YYYY	2	3		22.30	22.63	22.32	23.85	24.02	23.82		CEITEC at Masaryk University
1905	888	P06312 Immunoglobulin kappa variable 4-1 OS=Homo sapien	YYYY YYYY	3	3		20.33	19.12	18.22	23.80	24.02	24.14		41
1919	756	Q95293 NADH dehydrogenase [ubiquinone] alpha subcomplex	NNNN YYYY	1	3		-	21.62	21.78	23.65	24.04	23.96	0.005	

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=1	?	?	?	?	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...	UBQLN2	Ubiquitin-modifier-activating enzyme 6 OS=Homo sapiens GN=...	25.195	25.264	25.486	25.264	25.785	25.474
Row15	Acetyl-CoA...	LCB1	Acetyl-CoA acyltransferase synthase-like protein OS=Homo sapiens GN=LCB1 P... 21.699	?	?	?	23.402	23.072	23.906
Row7	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	Glycerol-3-phosphate phosphatase OS=Homo sapiens GN=PGP P...	25.593	25.677	25.894	25.938	25.979	25.637
Row20	TRPM8 ch...	TCAF2	TRPM8-channel-associated factor 2 OS=Homo sapiens GN=TCAF2...	19.119	?	?	19.369	20.047	19.483
Row21	rNARNL...	FBL11	rRNA/rRNA 2'-O-methyltransferase fibrillarin-like protein 1 OS=Hom...	?	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 ...	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-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	Nucleosid...	NUDT19	Nucleoside diphosphate-linked moiety X motif 19 OS=Homo sapi...	22.331	?	?	21.739	21.86	22.895
Row27	MIF4G do...	MIF4GD	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=2 SV=1;Kinesin lig...	23.118	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-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;Uncharac...	24.292	23.748	24.074	23.846	23.663	23.492
Row37	Vitamin K...	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=1	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=Or40n499 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

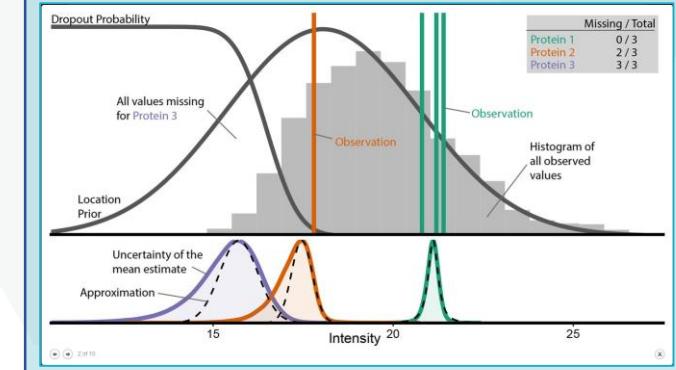
Protein mass spectrometry with label-free quantification (LFQ) is widely used for quantitative proteomics studies. Nevertheless, well-principled statistical inference procedures are still lacking, and most practitioners adopt methods from transcriptomics. These, however, cannot properly treat the principal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

The screenshot shows the 'Columns to process' configuration panel of the proDA software. It includes sections for 'Columns to process' (with options for 'Manual Selection', 'Wildcard/Regex Selection', and 'Type Selection'), 'Pattern' (set to 'norm'), 'Mismatch (Exclude)' (listing items like 'Protein IDs', 'Majority protein IDs', etc., with 'Protein IDs' highlighted in red), 'Match (Include)' (listing items like 'Intensity 1-1_1.log2_norm', 'Intensity 1-1_2.log2_norm', etc., with 'Intensity 1-1_1.log2_norm' highlighted in green), and other settings for 'proDA design', 'replicates', 'reference sample', 'contrast', and 'p-value adjustment method' (set to 'Benjamini & Hochberg').

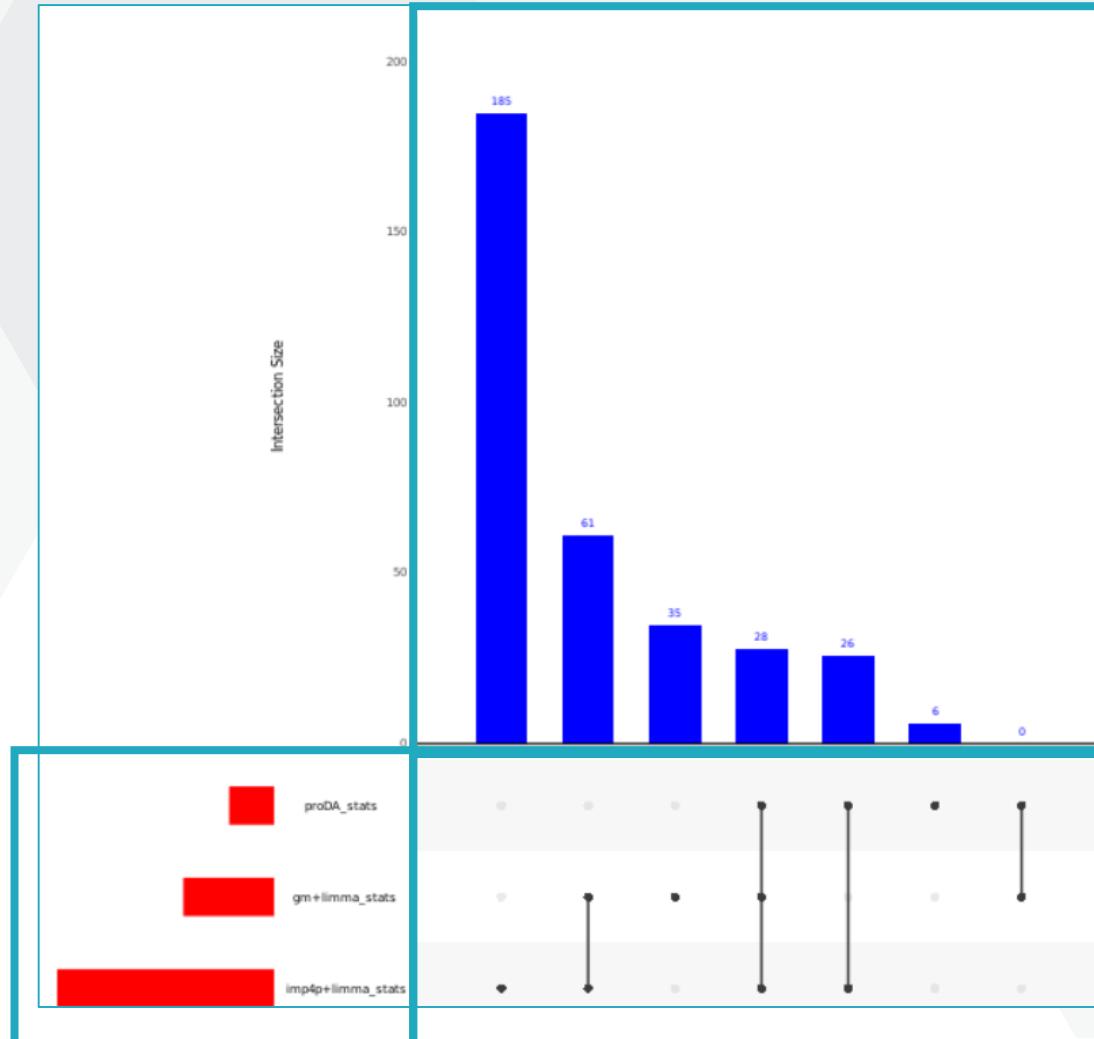
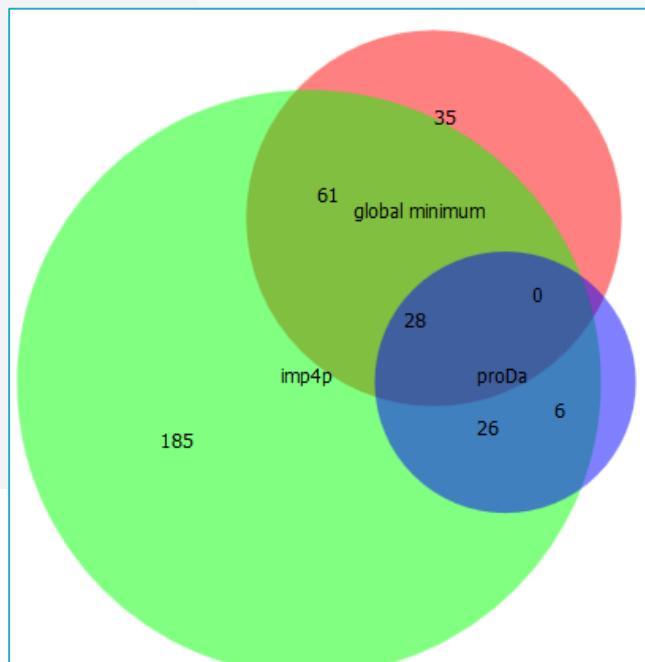
- statistical analysis with missing values still present
- similar results to global minimum imputation + LIMMA strategy
- tends to be more conservative compared to imp4p + LIMMA



(“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)



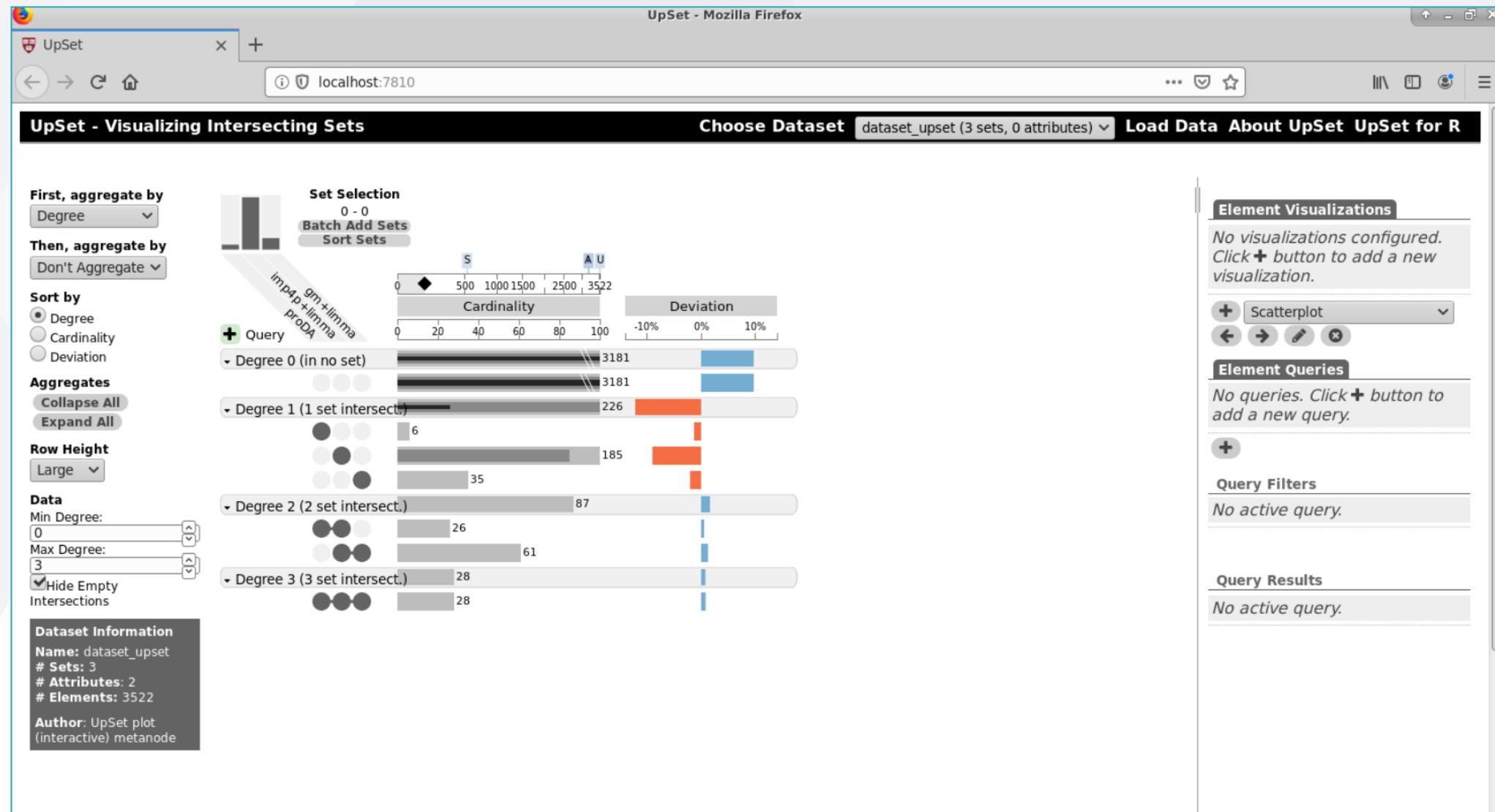
Size of particular sample

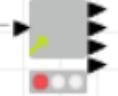


Intersections
(in which samples is particular protein present)

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)





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

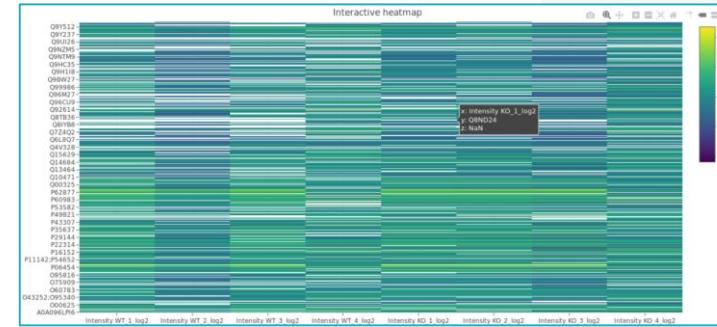
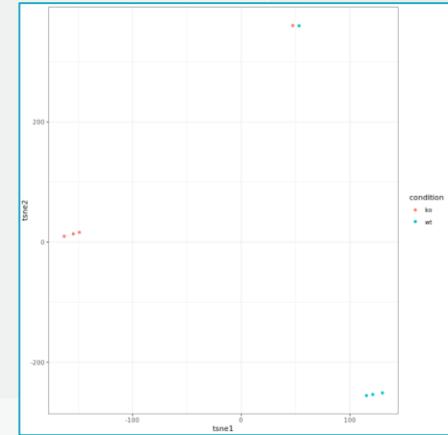
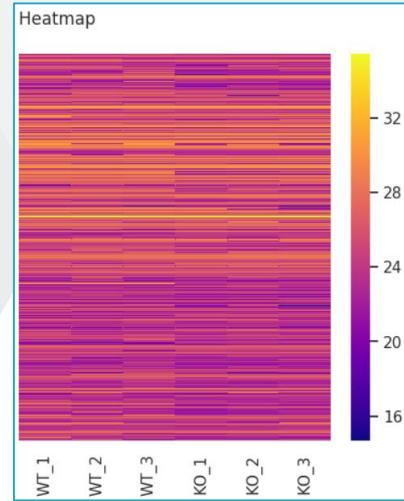
More information on the used API can be found here
<https://reactome.org/AnalysisService/#/identifiers> and here <https://reactome.org/dev/analysis>.

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;P8C6G24;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**