



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

Introduction to KNIME

Applications KNIME Analytics Platfo... [Terminal]

15:41 knimeuser



Trash



File System



Home



KNIME application



KNIME application

Workflows
templates r...Metanodes
templates r...

KNIME Analytics Platform

File Edit View Help

Quick Access

KNIME Explorer

EXAMPLES (knime-guest@http://pt) LOCAL (Local Workspace) 3.5.3c _Metanodes templates

Workflow Coach

Recommended Nodes

File Reader

Node Repository

IO

Manipulation

Column

Binning

Auto-Binner

Auto-Binner (Apply)

Numeric Binner

Binner (Dictionary)

CAIM Binner

CAIM Applier

Convert & Replace

Filter

Welcome to KNIME Analytics Platform

Open for Innovation KNIME

Welcome to KNIME Analytics Platform!

New to KNIME? Looking for resources to get started?

- Register for emails with introductory tips [here](#).
- Explore our [Quickstart Guide](#).
- Check out [7 things to do after installing KNIME Analytics Platform](#)
- Find more hints and how-tos in the [Learning Hub](#).
- And register for our release and event emails right [here](#).

This page will be displayed upon startup but you can customize the content using the checkboxes at the bottom.

Updates for the following components are available:

- KNIME Public Server Access
- KNIME ServerSpace
- OpenMS

Outline

An outline is not available.

Console

KNIME Console

```
***** Welcome to KNIME Analytics Platform v3.7.2.v201904170949 ****
*** Copyright by KNIME AG, Zurich, Switzerland ***
*****
```

Log file is located at: /home/knimeuser/knime-workspace/.metadata/knime/knime.log

380M of 577M

Auto-Binner

This node allows to group numeric data in intervals - called bins. There are two naming options for the bins and two methods which define the number and the range of values that fall in a bin. Please use the "Numeric Binner" node if you want to define custom bins.

Dialog Options

Column Selection:
Columns in the include list are processed separately. The columns in the exclude list are omitted by the node.

Binning Method:
Use **Fixed number of bins** for bins with equal **width** over the domain range or bins that have an equal **frequency** of element

CC BY SA

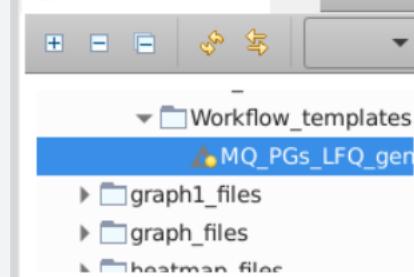
File Edit View Node Help



75%



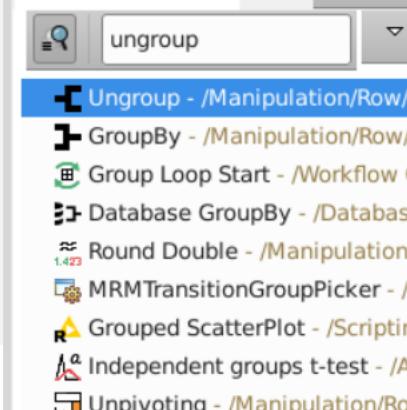
KNIME Explorer



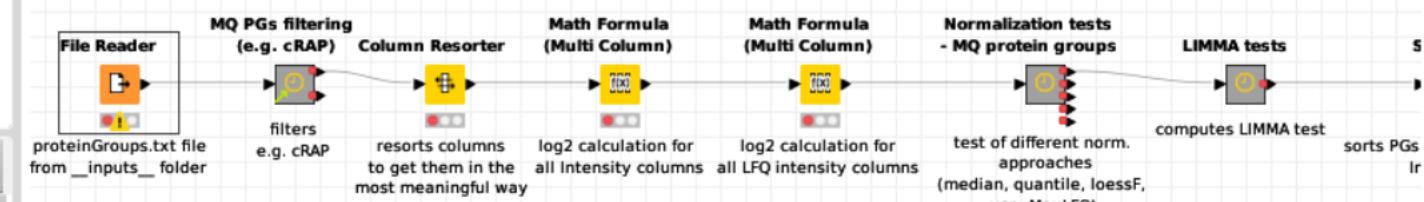
Workflow Coach

Recommended Nodes

Node Repository



*0: MQ_PGs_LFQ_general_0.6a



Example dataset

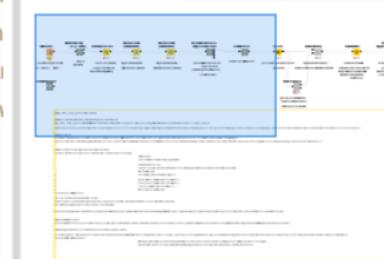


```

# MQ_PGs_LFQ_general workflow
## Description of the workflow and general remarks
MQ_PGs_LFQ_general is a KNIME workflow developed for the general processing of label-free bottom-up mass spectrometry data.
Please note, that you should understand e.g. data structure and experimental design used within the study to apply the correct processing approach! The pr
## Input data

```

Outline



Console

KNIME Console

```

WARN Column Rename (Regex) 0:844:0:0:850 Pattern replace resulted in duplicate column names; resolved c
WARN CASE Switch Data (Start) 0:844:0:0:73 Errors overwriting node settings with flow variables: Unknown
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unknown
WARN Column Rename (Regex) 0:844:239:0:850 Pattern replace resulted in duplicate column names; resolved c
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unknown
WARN Column Rename (Regex) 0:844:300:277:0:428 Pattern did not match any column name, leaving input unc
WARN Column Rename (Regex) 0:844:300:290:0:428 Pattern did not match any column name, leaving input unc

```

Applications

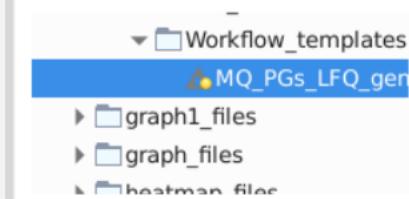
KNIME Analytics Platfor...

09:51 knimeuse

File Edit View Node Help



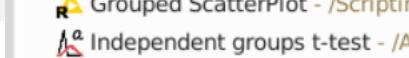
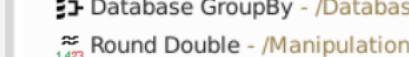
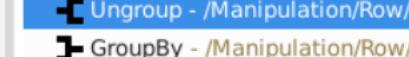
KNIME Explorer



Workflow Coach



Recommended Nodes

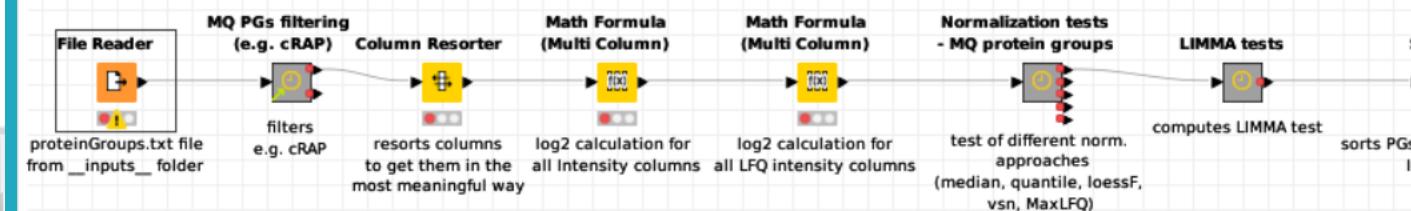


KNIME Analytics Platform

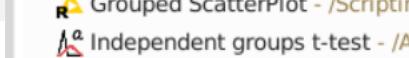
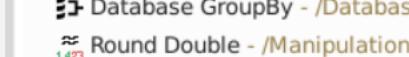
75%



*0: MQ_PGs_LFQ_general_0.6a

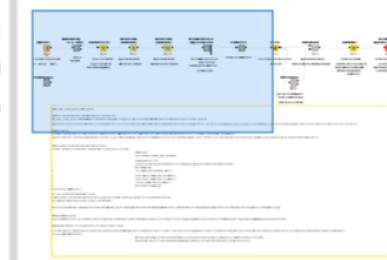


Example dataset



```
# MQ_PGs_LFQ_general workflow
## Description of the workflow and general remarks
MQ_PGs_LFQ_general is a KNIME workflow developed for the general processing of label-free bottom-up mass spectrometry data.
Please note, that you should understand e.g. data structure and experimental design used within the study to apply the correct processing approach! The pr
## Input data
```

Outline



Console

KNIME Console

```
WARN Column Rename (Regex) 0:844:0:0:850 Pattern replace resulted in duplicate column names; resolved c
WARN CASE Switch Data (Start) 0:844:0:0:73 Errors overwriting node settings with flow variables: Unkn
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unkn
WARN Column Rename (Regex) 0:844:239:0:850 Pattern replace resulted in duplicate column names; resolved c
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unkn
WARN Column Rename (Regex) 0:844:300:277:0:428 Pattern did not match any column name, leaving input unc
WARN Column Rename (Regex) 0:844:300:290:0:428 Pattern did not match any column name, leaving input unc
```

Workflow editor

File Reader

This node can be used to read data from an ASCII file or URL location. It can be configured to read various formats.

When you open the node's configuration dialog and provide a filename, it tries to guess the reader's settings by analyzing the content of the file. Check the results of these settings in the preview table. If the data shown is not correct or an error is reported, you can adjust the settings manually (see below).

The file analysis runs in the background and can be cut short by clicking the "Quick scan", which shows if the analysis takes longer. In this case the file is not analyzed completely, but only the first fifty lines are taken into account. It could happen then, that the preview appears looking fine, but the execution of the File Reader fails, when it reads the lines it didn't analyze. Thus it is recommended you check the settings, when you cut an analysis short.

KNIME Explorer 1.6a

File Reader

This node can be used to read data from an ASCII file or URL location. It can be configured to read various formats.

When you open the node's configuration dialog and provide a filename, it tries to guess the reader's settings by analyzing the content of the file. Check the results of these settings in the preview table. If the data shown is not correct or an error is reported, you can adjust the settings manually (see below).

The file analysis runs in the background and can be cut short by clicking the "Quick scan", which shows if the analysis takes longer. In this case the file is not analyzed completely, but only the first fifty lines are taken into account. It could happen then, that the preview appears looking fine, but the execution of the File Reader fails, when it reads the lines it didn't analyze. Thus it is recommended you check the settings, when you cut an analysis short.

```
# MQ_PGs_LFQ_general workflow
## Description of the workflow and general remarks
MQ_PGs_LFQ_general is a KNIME workflow developed for the general processing of label-free bottom-up mass spectrometry data.
Please note, that you should understand e.g. data structure and experimental design used within the study to apply the correct processing approach! The pr
## Input data
```

Outline

Console

KNIME Console

WARN Column Rename (Regex) 0:844:0:0:850 Pattern replace resulted in duplicate column names; resolved c
WARN CASE Switch Data (Start) 0:844:0:0:73 Errors overwriting node settings with flow variables: Unkn
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unkn
WARN Column Rename (Regex) 0:844:239:0:850 Pattern replace resulted in duplicate column names; resolved
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unkn
WARN Column Rename (Regex) 0:844:300:277:0:428 Pattern did not match any column name, leaving input unc
WARN Column Rename (Regex) 0:844:300:290:0:428 Pattern did not match any column name, leaving input unc

330M of 408M

Applications

KNIME Analytics Platfor...

09:51 knimeuse

File Edit View Node Help

KNIME Analytics Platform



75%



Quick Access

KNIME Explorer

-
- Workflow_templates
- MQ_PGs_LFQ_general_0.6a
- graph1_files
- graph_files
- hostman_files

Workflow Coach

Recommended Nodes

Node Repository

ungroup

Ungroup - /Manipulation/Row/

GroupBy - /Manipulation/Row/

Group Loop Start - /Workflow C

Database GroupBy - /Database

Round Double - /Manipulation/

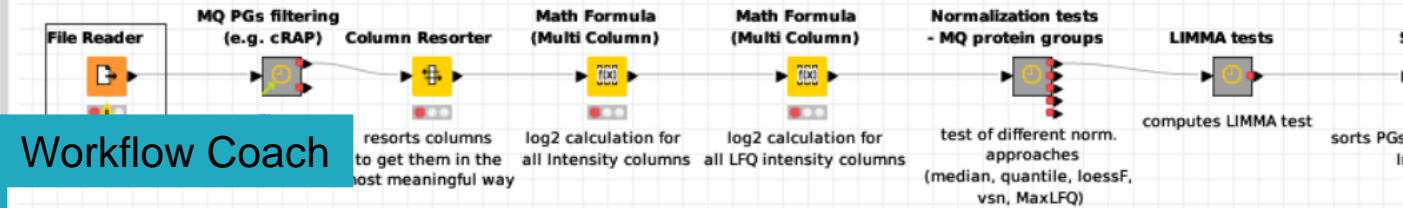
MRMTransitionGroupPicker - /

Grouped ScatterPlot - /Scriptin

Independent groups t-test - /Al

Unpivoting - /Manipulation/Ro

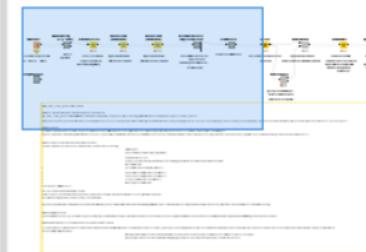
*0: MQ_PGs_LFQ_general_0.6a



Example dataset

```
# MQ_PGs_LFQ_general workflow
## Description of the workflow and general remarks
MQ_PGs_LFQ_general is a KNIME workflow developed for the general processing of label-free bottom-up mass spectrometry data.
Please note, that you should understand e.g. data structure and experimental design used within the study to apply the correct processing approach! The pr
## Input data
```

Outline



Console

KNIME Console

```
WARN Column Rename (Regex) 0:844:0:0:850 Pattern replace resulted in duplicate column names; resolved c
WARN CASE Switch Data (Start) 0:844:0:0:73 Errors overwriting node settings with flow variables: Unkn
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unkn
WARN Column Rename (Regex) 0:844:239:0:850 Pattern replace resulted in duplicate column names; resolved c
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unkn
WARN Column Rename (Regex) 0:844:300:277:0:428 Pattern did not match any column name, leaving input unc
WARN Column Rename (Regex) 0:844:300:290:0:428 Pattern did not match any column name, leaving input unc
```

File Edit View Node Help



75%



KNIME Explorer



- Workflow_templates
- MQ_PGs_LFQ_general_0.6a
- graph1_files
- graph_files
- hostman_files

Workflow Coach



Recommended Nodes

Node Repository



Ungroup - /Manipulation/Row/

GroupBy - /Manipulation/Row/

Group Loop Start - /Workflow C

Database GroupBy - /Database

Round Double - /Manipulation/

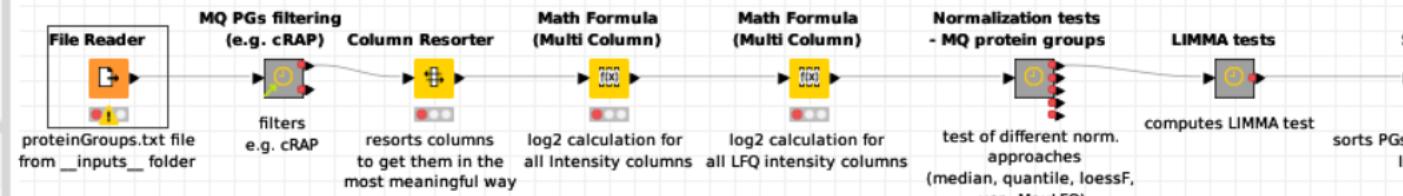
MRMTransitionGroupPicker - /

Grouped ScatterPlot - /Scriptin

Independent groups t-test - /A

Unpivoting - /Manipulation/Ro

*0: MQ_PGs_LFQ_general_0.6a

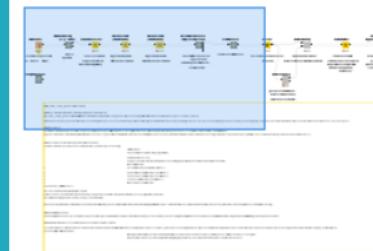


Example dataset

Node repository

```
# MQ_PGs_LFQ_general workflow
## Description of the workflow and general remarks
MQ_PGs_LFQ_general is a KNIME workflow developed for the general processing of label-free bottom-up mass spectrometry data.
Please note, that you should understand e.g. data structure and experimental design used within the study to apply the correct processing approach! The pr
## Input data
```

Outline



Console

KNIME Console

```
WARN Column Rename (Regex) 0:844:0:0:850 Pattern replace resulted in duplicate column names; resolved c
WARN CASE Switch Data (Start) 0:844:0:0:73 Errors overwriting node settings with flow variables: Unkn
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unkn
WARN Column Rename (Regex) 0:844:239:0:850 Pattern replace resulted in duplicate column names; resolved c
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unkn
WARN Column Rename (Regex) 0:844:300:277:0:428 Pattern did not match any column name, leaving input unc
WARN Column Rename (Regex) 0:844:300:290:0:428 Pattern did not match any column name, leaving input unc
```



330M of 408M





KNIME Analytics Platform

File Explorer *0: MQ_PGs_LFQ_general_0.6a Node Description

File Reader

This node can be used to read data from an ASCII file or URL location. It can be configured to read various formats. When you open the node's configuration dialog and provide a filename, it tries to guess the reader's settings by analyzing the content of the file. Check the results of these settings in the preview table. If the data shown is not correct or an error is reported, you can adjust the settings manually (see below).

The file analysis runs in the background and can be cut short by clicking the "Quick scan", which shows if the analysis takes longer. In this case the file is not analyzed completely, but only the first fifty lines are taken into account. It could happen then, that the preview appears looking fine, but the execution of the File Reader fails, when it reads the lines it didn't analyze. Thus it is recommended you check the settings, when you cut an analysis short.

Workflow Coach

Recommended Nodes

Node Repository

Ungroup - /Manipulation/Row/

GroupBy - /Manipulation/Row/

Group Loop Start - /Workflow C

Database GroupBy - /Database

Round Double - /Manipulation/

MRMTransitionGroupPicker - /

Grouped ScatterPlot - /Scriptin

Independent groups t-test - /Al

Unpivoting - /Manipulation/Ro

Outline

Console

KNIME Console

```

WARN Column Rename (Regex) 0:844:0:0:850 Pattern replace resulted in duplicate column names; resolved c
WARN CASE Switch Data (Start) 0:844:0:0:73 Errors overwriting node settings with flow variables: Unkn
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unkn
WARN Column Rename (Regex) 0:844:239:0:850 Pattern replace resulted in duplicate column names; resolved c
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unkn
WARN Column Rename (Regex) 0:844:300:277:0:428 Pattern did not match any column name, leaving input unc
WARN Column Rename (Regex) 0:844:300:290:0:428 Pattern did not match any column name, leaving input unc

```

330M of 408M

CC BY SA



KNIME Analytics Platform

File Edit View Node Help

75% Quick Access

KNIME Explorer *0: MQ_PGs_LFQ_general_0.6a Node Description

File Reader

This node can be used to read data from an ASCII file or URL location. It can be configured to read various formats. When you open the node's configuration dialog and provide a filename, it tries to guess the reader's settings by analyzing the content of the file. Check the results of these settings in the preview table. If the data shown is not correct or an error is reported, you can adjust the settings manually (see below).

The file analysis runs in the background and can be cut short by clicking the "Quick scan", which shows if the analysis takes longer. In this case the file is not analyzed completely, but only the first fifty lines are taken into account. It could happen then, that the preview appears looking fine, but the execution of the File Reader fails, when it reads the lines it didn't analyze. Thus it is recommended you check the settings, when you cut an analysis short.

Workflow Coach

Recommended Nodes

Node Repository

ungroup

Ungroup - /Manipulation/Row/

GroupBy - /Manipulation/Row/

Group Loop Start - /Workflow C

Database GroupBy - /Database

Round Double - /Manipulation/

MRMTransitionGroupPicker - /

Grouped ScatterPlot - /Scriptin

Independent groups t-test - /Al

Unpivoting - /Manipulation/Ro

MQ_PGs_LFQ_general workflow

```

# MQ_PGs_LFQ_general workflow
## Description of the workflow and general remarks
MQ_PGs_LFQ_general is a KNIME workflow developed for the general processing of label-free bottom-up mass spectrometry data.
Please note, that you should understand e.g. data structure and experimental design used within the study to apply the correct processing approach! The pr
## Input data

```

Console

KNIME Console

WARN Column Rename (Regex) 0:844:0:0:850 Pattern replace resulted in duplicate column names; resolved c
WARN CASE Switch Data (Start) 0:844:0:0:73 Errors overwriting node settings with flow variables: Unknown
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unknown
WARN Column Rename (Regex) 0:844:239:0:850 Pattern replace resulted in duplicate column names; resolved c
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unknown
WARN Column Rename (Regex) 0:844:300:277:0:428 Pattern did not match any column name, leaving input unc
WARN Column Rename (Regex) 0:844:300:290:0:428 Pattern did not match any column name, leaving input unc

330M of 408M

CC BY SA



KNIME Analytics Platform

File Edit View Node Help

Quick Access

KNIME Explorer

- Workflow_templates
 - MQ_PGs_LFQ_general_0.6a
 - graph1_files
 - graph_files
 - hostman_files

Workflow Coach

Recommended Nodes

- Node Repository
 - ungroup
- Ungroup - /Manipulation/Row/
- GroupBy - /Manipulation/Row/
- Group Loop Start - /Workflow C
- Database GroupBy - /Database
- Round Double - /Manipulation/
- MRMTransitionGroupPicker - /
- Grouped ScatterPlot - /Scriptin
- Independent groups t-test - /A
- Unpivoting - /Manipulation/Ro

***0: MQ_PGs_LFQ_general_0.6a**

Node description

This node can be used to read data from an ASCII file or URL location. It can be configured to read various formats.

When you open the node's configuration dialog and provide a filename, it tries to guess the reader's settings by analyzing the content of the file. Check the results of these settings in the preview table. If the data shown is not correct or an error is reported, you can adjust the settings manually (see below).

The file analysis runs in the background and can be cut short by clicking the "Quick scan", which shows if the analysis takes longer. In this case the file is not analyzed completely, but only the first fifty lines are taken into account. It could happen then, that the preview appears looking fine, but the execution of the File Reader fails, when it reads the lines it didn't analyze. Thus it is recommended you check the settings, when you cut an analysis short.

File Reader

```
# MQ_PGs_LFQ_general workflow
## Description of the workflow and general remarks
MQ_PGs_LFQ_general is a KNIME workflow developed for the general processing of label-free bottom-up mass spectrometry data.
Please note, that you should understand e.g. data structure and experimental design used within the study to apply the correct processing approach! The pr
## Input data
```

Outline

Console

KNIME Console

WARN Column Rename (Regex) 0:844:0:0:850 Pattern replace resulted in duplicate column names; resolved c
 WARN CASE Switch Data (Start) 0:844:0:0:73 Errors overwriting node settings with flow variables: Unkn
 WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unkn
 WARN Column Rename (Regex) 0:844:239:0:850 Pattern replace resulted in duplicate column names; resolved c
 WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unkn
 WARN Column Rename (Regex) 0:844:300:277:0:428 Pattern did not match any column name, leaving input unc
 WARN Column Rename (Regex) 0:844:300:290:0:428 Pattern did not match any column name, leaving input unc

330M of 408M

CC BY SA

Applications

KNIME Analytics Platfor...

09:51 knimeuse

File Edit View Node Help



75%



Quick Access

KNIME Explorer

-
- Workflow_templates
 - MQ_PGs_LFQ_general
 - graph1_files
 - graph_files
 - hostman_files

Workflow Coach

Recommended Nodes

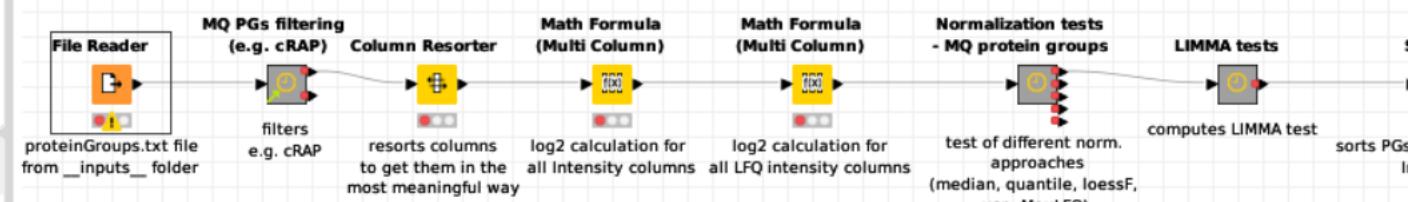
Node Repository

- ungroup
- Ungroup - /Manipulation/Row/
- GroupBy - /Manipulation/Row/
- Group Loop Start - /Workflow C
- Database GroupBy - /Database
- Round Double - /Manipulation/
- MRMTransitionGroupPicker - /
- Grouped ScatterPlot - /Scriptin
- Independent groups t-test - /A
- Unpivoting - /Manipulation/Ro

Outline

*0: MQ_PGs_LFQ_general

Node alignment and execution



Example dataset

ungroup

```
# MQ_PGs_LFQ_general workflow
## Description of the workflow and general remarks
MQ_PGs_LFQ_general is a KNIME workflow developed for the general processing of label-free bottom-up mass spectrometry data.
Please note, that you should understand e.g. data structure and experimental design used within the study to apply the correct processing approach! The pr
## Input data
```

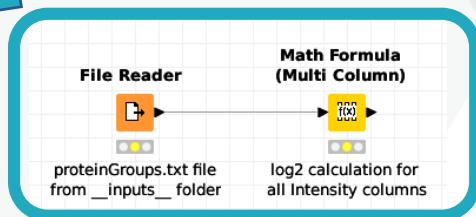
Console

KNIME Console

```
WARN Column Rename (Regex) 0:844:0:0:850 Pattern replace resulted in duplicate column names; resolved c
WARN CASE Switch Data (Start) 0:844:0:0:73 Errors overwriting node settings with flow variables: Unkn
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unkn
WARN Column Rename (Regex) 0:844:239:0:850 Pattern replace resulted in duplicate column names; resolved c
WARN CASE Switch Data (Start) 0:844:239:0:73 Errors overwriting node settings with flow variables: Unkn
WARN Column Rename (Regex) 0:844:300:277:0:428 Pattern did not match any column name, leaving input unc
WARN Column Rename (Regex) 0:844:300:290:0:428 Pattern did not match any column name, leaving input unc
```

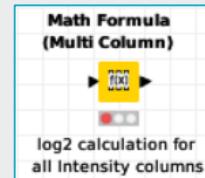
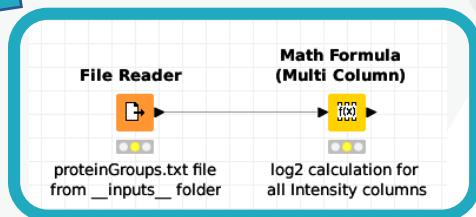
KNIME environment and metanodes

Nodes

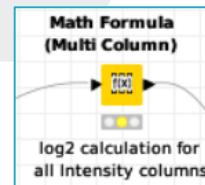


KNIME environment and metanodes

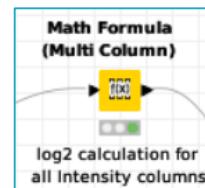
Nodes



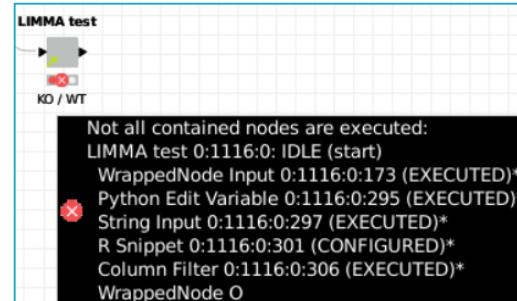
Not connected to other nodes
Not executed



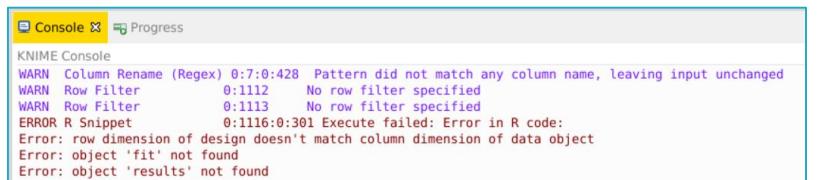
Connected to other nodes
Not executed



Connected to other nodes
Successfully executed

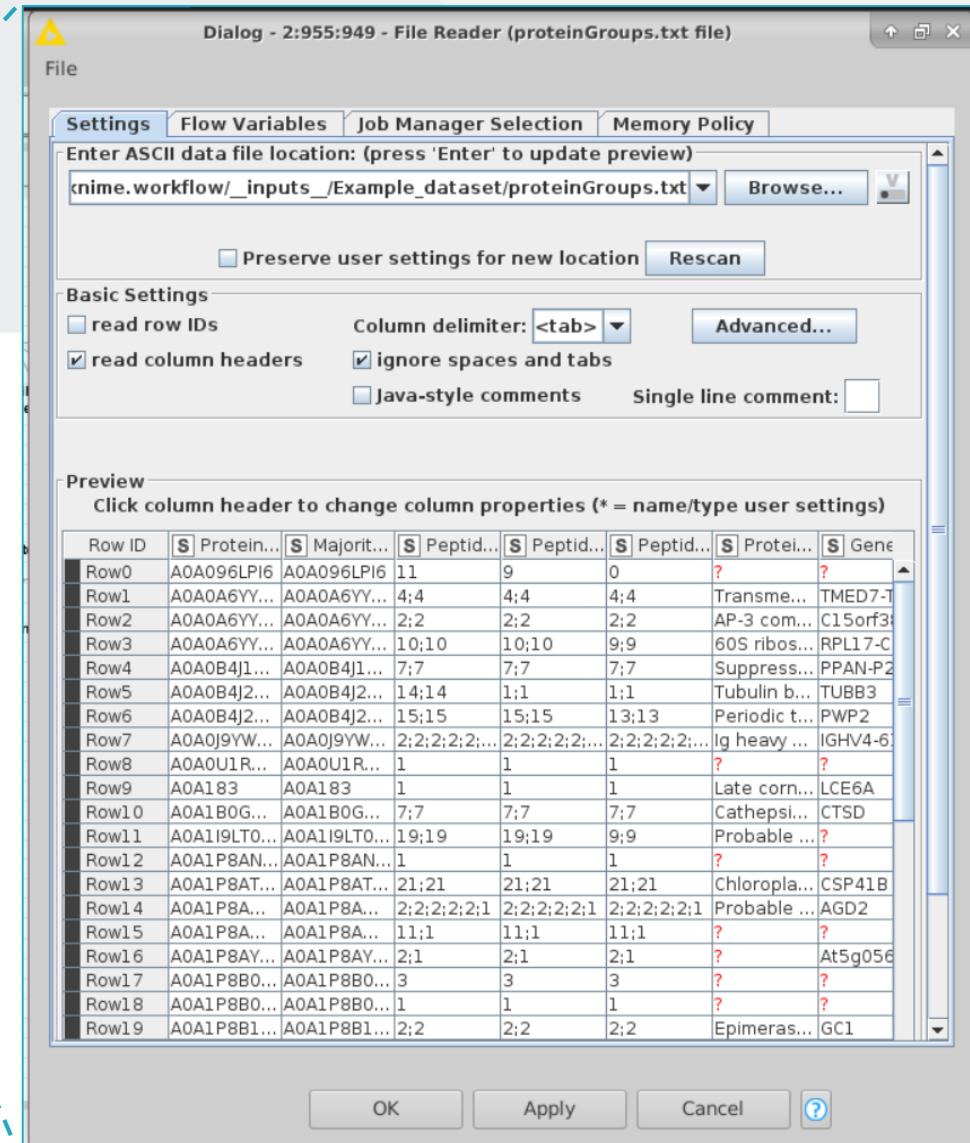
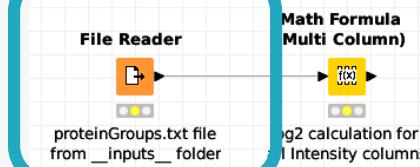


Connected to other nodes
Error while node execution

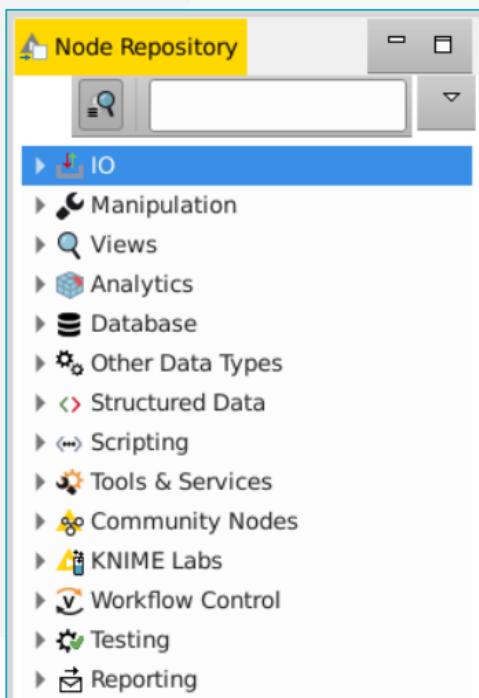


KNIME environment and metanodes

Nodes



In-built KNIME metanodes



IO

- File Reader
- Excel Reader
- CSV Reader
- Table Reader
- CSV Writer
- Table Writer
- Excel Sheet Appender
- Table Creator

Manipulation

- Column/Row Filter
- Column Rename
- Math Formula
- String Manipulation
- Column Splitter
- Column Appender
- Joiner
- Concatenate
- Column Resorter
- Missing Value
- Row Filter/Splitter
- GroupBy
- Pivoting
- Rule Engine
- Transpose
- Sorter
- Shuffle

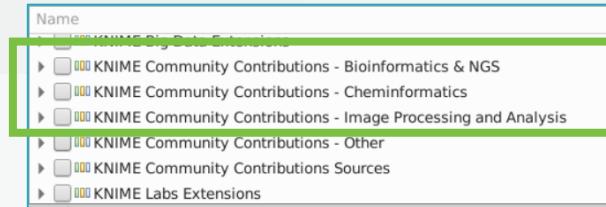
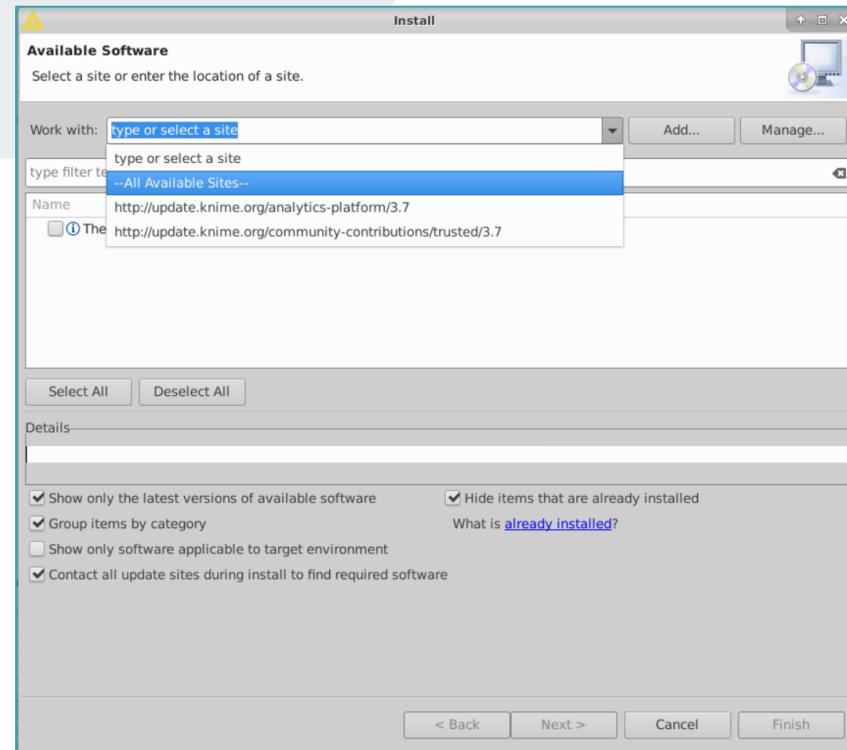
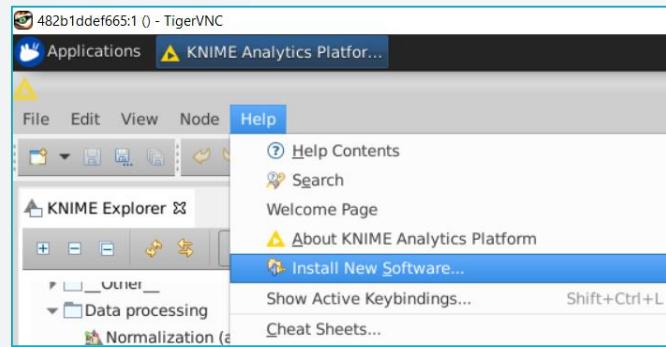
Views

- Bar Chart
- Box Plot
- Scatter Plot
- ROC Curve
- Histogram
- Line Plot

Scripting

- Java
- Python
- R

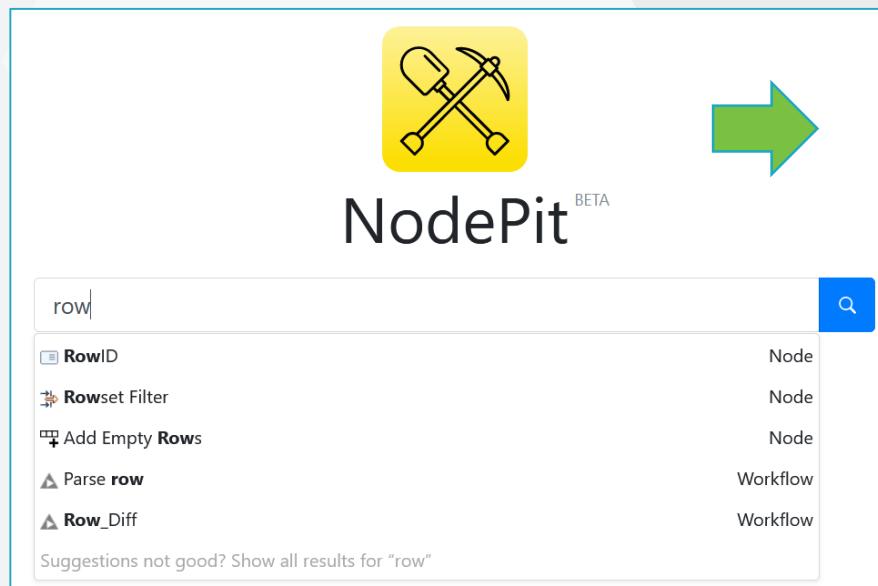
Installation of additional plugins / software



OpenMS
PIA

NodePit

Where to seek help or inspiration



The screenshot shows the NodePit documentation page for the "Row Filter" node. At the top, there are navigation links: Nodes (636), Workflows (763), Products (0), Installable Units (2), and Users (2). Below the navigation is a list of nodes:

- Row Filter Streamable**
Nodes > Manipulation > Row > Filter
Allows filtering of data rows by certain criteria, such as row ID, attribute value, and row number range.
- Row Splitter Streamable**
Nodes > Manipulation > Row > Filter
Allows splitting of the input table by certain criteria, such as row ID, attribute value, and row number range.
- Key-Collection HiLite Translator**
Nodes > Manipulation > Row
Translates hilite events from a row containing a collection cell with row keys to the original rows.
- Duplicate Row Filter**
Nodes > Manipulation > Row > Filter
Removes or marks duplicate rows.
- RowID Streamable**
Nodes > Manipulation > Row > Other
Node to replace the RowID and/or to create a column with the values of the current RowID.
- Spark Row Filter**
Nodes > Tools & Services > Apache Spark > Row
The Spark Row Filter allows rows to be excluded from the input DataFrame/RDD.

NodePit

RowID

Streamable KNIME Base Nodes version 4.0.2.v201909251340 by KNIME AG, Zurich, Switzerland

This node can be used to replace the RowID of the input data with the values of another column (by converting the values to a string) or a generated row id of the format: Row0, Row1, Row2, ... The user has additional options to ensure uniqueness and to handle missing values.

It can also be used to create a new column, which contains the RowID as a value.

If both options are selected the node appends a new column with the values of the current RowID to the table and replaces the current RowID with the values of the selected column or the generated row id.

Note: Highlighting does not work across this node if the "Enable hiliting" option is disabled.

Options

Replace RowID with selected column values: If selected the node replaces each RowID with the corresponding string value of the selected column or with a generated key.

New RowID column: The column to replace the current RowID. If none is selected a row id is generated with the format: Row0, Row1, Row2, ...

Ensure uniqueness: If checked the node ensures uniqueness by appending a counter (x) to duplicates where x is incremented for each appearance.

Handle missing values: All missing values are replaced with '?' if this check box is ticked. We recommend activating the 'Ensure uniqueness' check box to handle any duplicate missing values.

Enable hiliting: If enabled, a map is maintained joining the old with the new RowID. Depending of the number of rows, enabling this feature might consume a lot of memory.

Create new column with the RowID values: If this is selected the node creates a new column with the value of the current RowID.

New column name: The name of the new column that is being added to the table.

Input Ports

► The data table to be processed.

Output Ports

► Data with the replaced RowID and/or a new column with the current RowID values.

Best Friends (Incoming)

- Transpose (7 %)
- Row Filter (6 %) Streamable
- Column Filter (6 %) Streamable
- Joiner (5 %)
- GroupBy (4 %)

[Show all 444 recommendations](#)

Best Friends (Outgoing)

- Joiner (14 %)
- Column Filter (8 %) Streamable
- Transpose (6 %)
- Row Filter (4 %) Streamable
- String Manipulation (3 %) Streamable

[Show all 590 recommendations](#)

Workflows

- GASIC
- Factiva_parser
- PhD_Articles
- wrong-column-name-13973
- Amazon Reviews Preprocessing

[Show all 348 workflows](#)

Information about the node, options, following nodes provided

KNIME community and help

The screenshot shows the KNIME forum homepage. At the top, there's a navigation bar with links for Hub, Blog, Forum, Events, Career, Contact, Download, and a search icon. Below the navigation is a breadcrumb menu: SOFTWARE / SOLUTIONS / LEARNING / PARTNERS / COMMUNITY / ABOUT. There are also Log In and search icons.

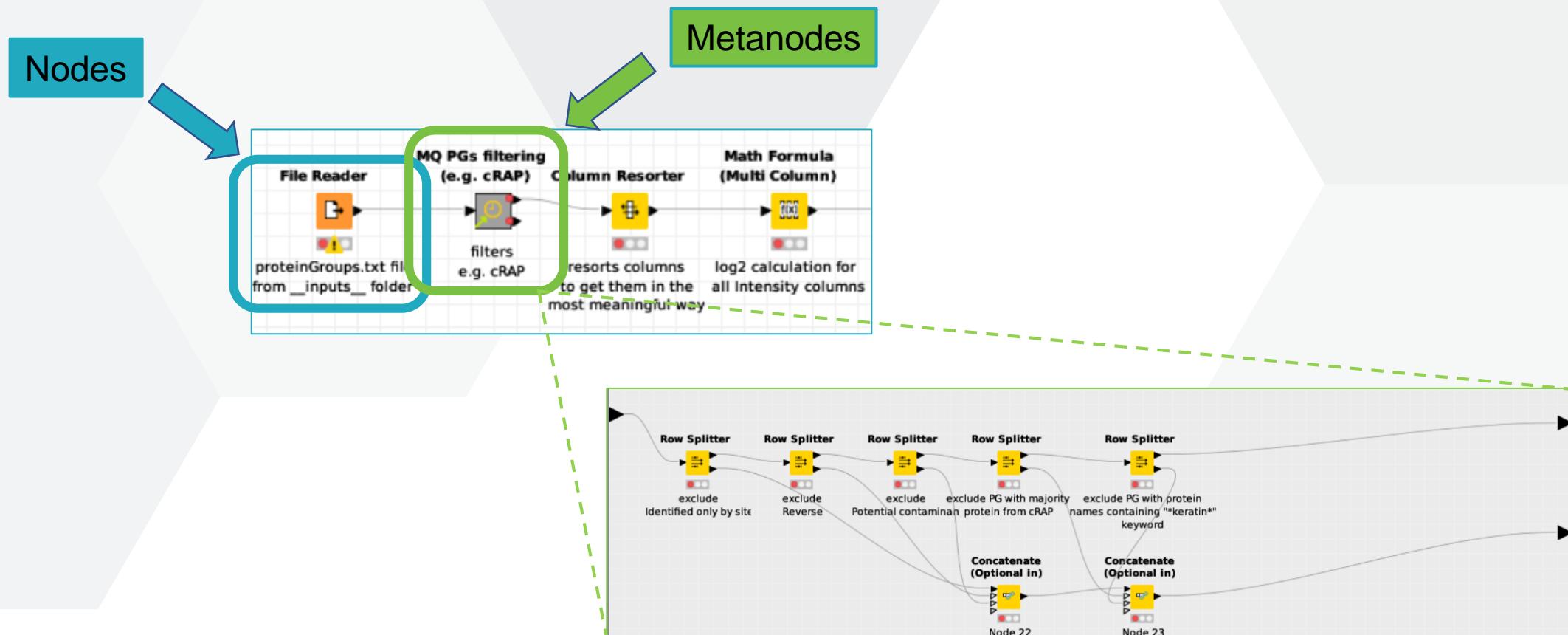
The main content area has two sections: "all categories" and "Categories". Under "Categories", there are several sections:

- KNIME Analytics Platform**: For discussions related to KNIME Analytics Platform. Topics per week: 61. Recent topic: "Don't save start / end nodes for variable connection" by user M, posted 26m ago.
- KNIME Extensions**: For discussions related to KNIME Extensions and Integrations. Topics per week: 8. Recent topic: "DB Writer writing only 25 rows per second to MySQL database" by user KNIME Analytics Platform, posted 36m ago.
- Community Extensions**: For discussions related to extensions developed by the KNIME community. Topics per week: 1. Recent topic: "workflow not executed on scheduled time...skipped few workflow and jumped to next." by user KNIME Server, posted 1h ago.
- Partner Extensions**: For discussions related to extensions developed by our KNIME partners. Topics per week: 87. Recent topic: "How to read all files in a folder with a loop" by user KNIME Analytics Platform, posted 3h ago.
- KNIME Server**: For discussions related to KNIME Server. Topics per week: 4. Recent topic: "Obtain a distance vector from a Data Frame" by user KNIME Analytics Platform, posted 5h ago.
- Special Interest Groups**: For discussions related to various special interest groups. Topics per week: 1. Recent topic: "Parsing and Pivoting JSON" by user KNIME Development, posted 5h ago.
- KNIME Development**: For discussions related to KNIME development. Topics per week: 2. Recent topic: "How to handle merged cells importing from Excel" by user KNIME Analytics Platform, posted 6h ago.

On the right side, there's a "Latest" section listing recent posts:

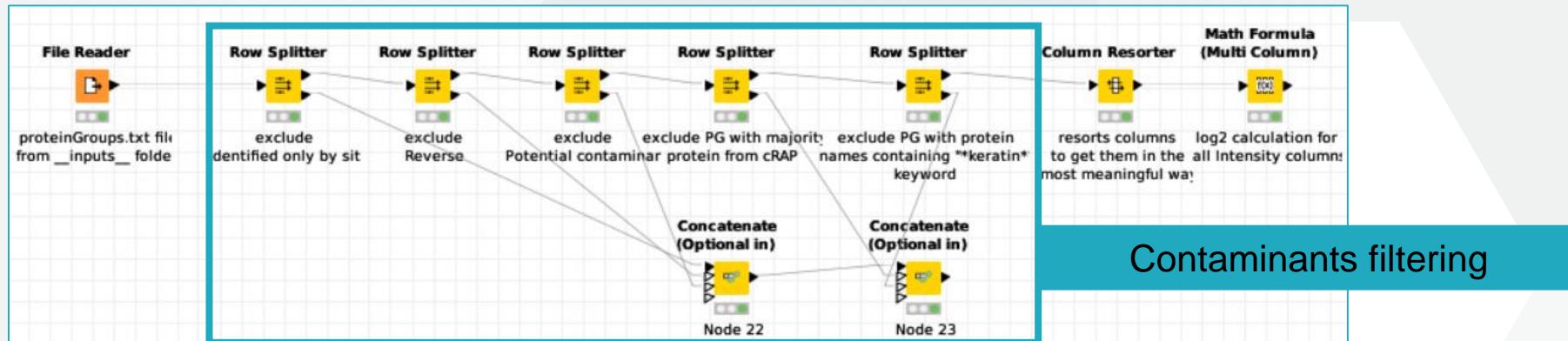
- "Don't save start / end nodes for variable connection" by user M, posted 26m ago.
- "DB Writer writing only 25 rows per second to MySQL database" by user KNIME Analytics Platform, posted 36m ago.
- "workflow not executed on scheduled time...skipped few workflow and jumped to next." by user KNIME Server, posted 1h ago.
- "How to read all files in a folder with a loop" by user KNIME Analytics Platform, posted 3h ago.
- "Obtain a distance vector from a Data Frame" by user KNIME Analytics Platform, posted 5h ago.
- "Parsing and Pivoting JSON" by user KNIME Development, posted 5h ago.
- "How to handle merged cells importing from Excel" by user KNIME Analytics Platform, posted 6h ago.
- "Get Request Node" by user KNIME Analytics Platform, posted 11h ago.

KNIME environment and metanodes



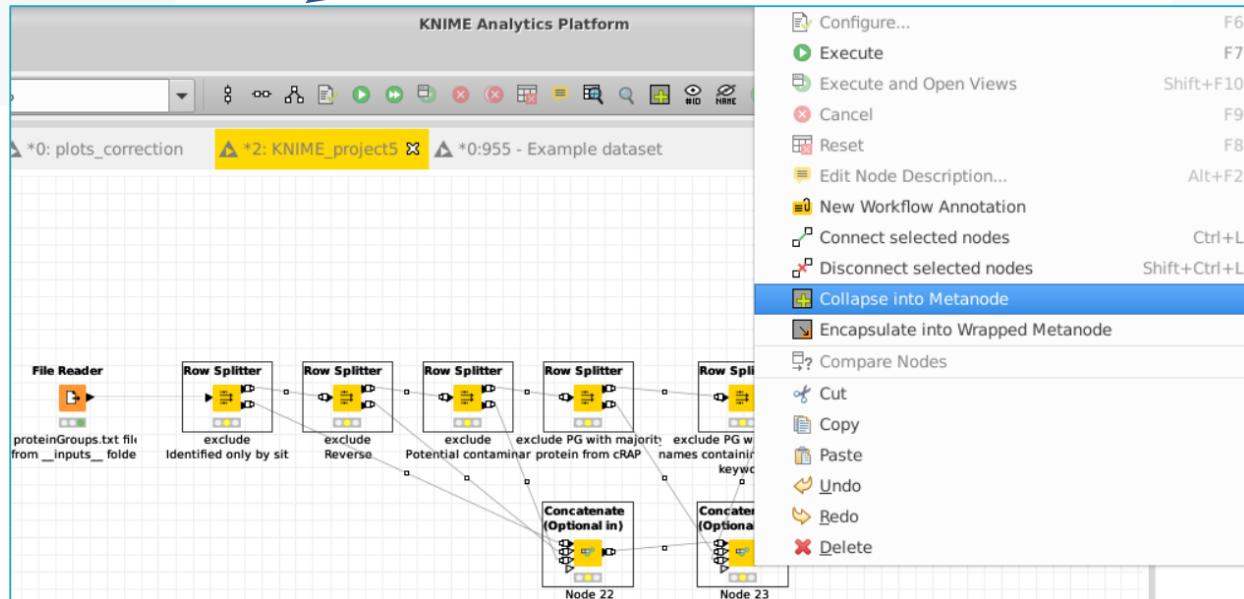
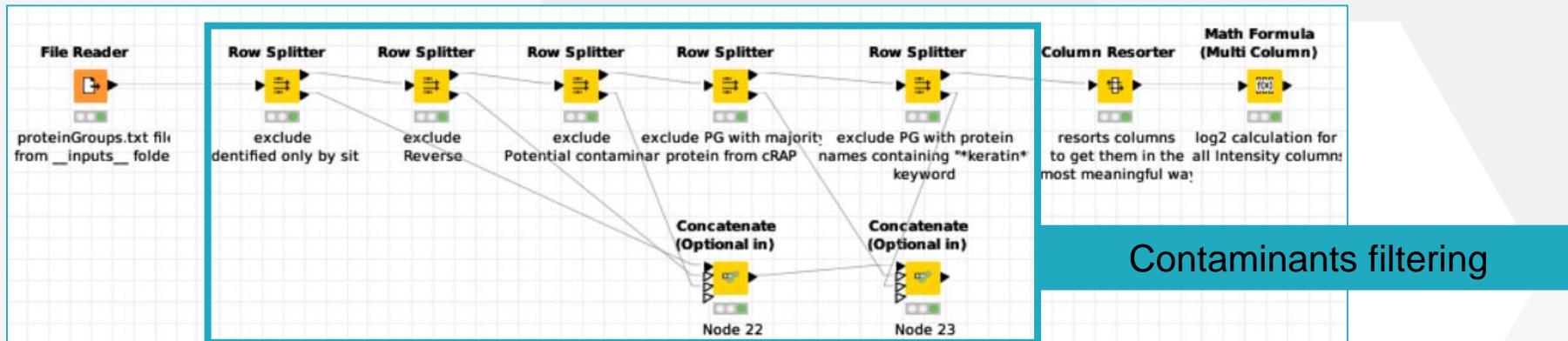
Another layer of data processing
Making workflow more structured and organized

KNIME environment and metanodes



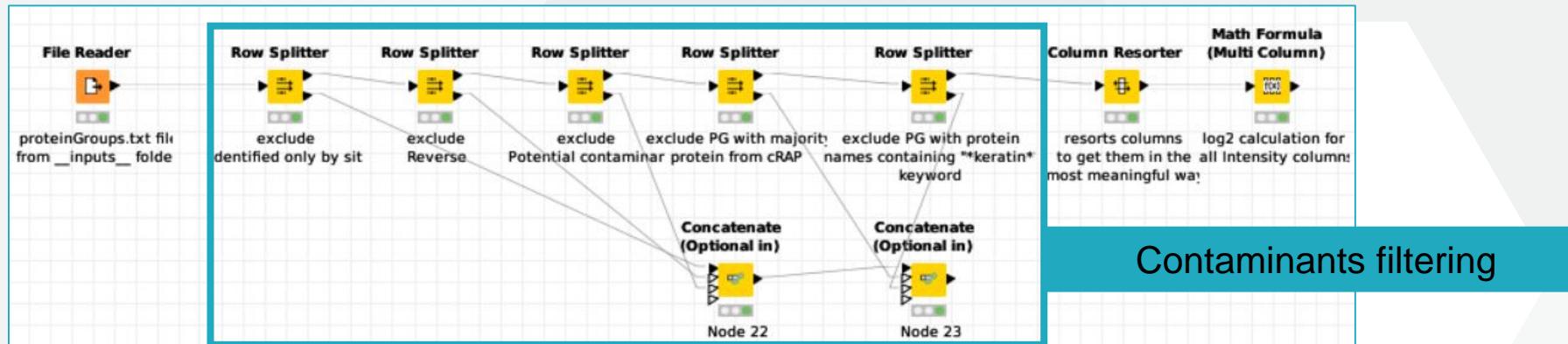
Collapsing nodes into Metanode
Structuring the workflow, making it easily readable

KNIME environment and metanodes



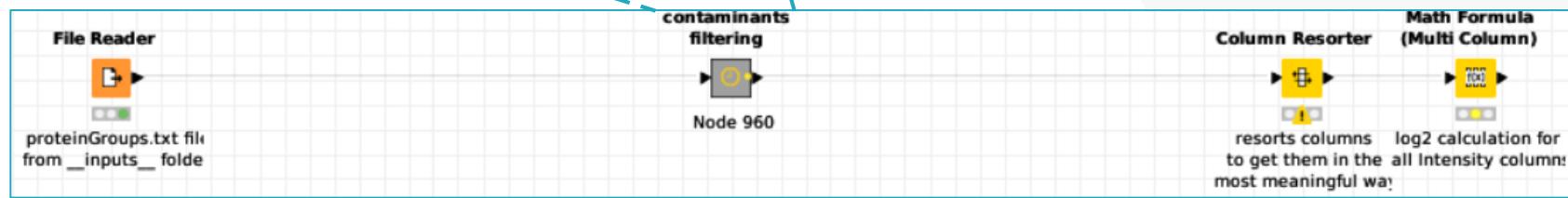
Collapsing nodes into Metanode
Structuring the workflow, making it easily readable

KNIME environment and metanodes



Contaminants filtering

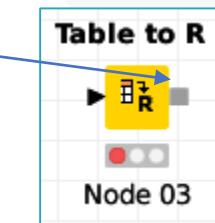
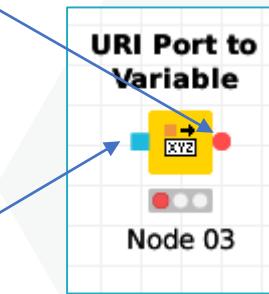
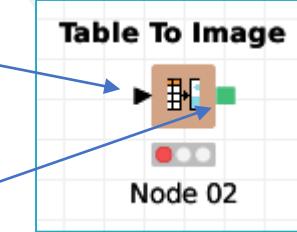
Collapsing nodes into Metanode
Structuring the workflow, making it easily readable



Different port types

tables holding data
variables used when scripting

Data
Flowvariable
PMML
Database Connection
Database Query
Distance Measure
DocumentVectorPortObject
FileStorePrefixURIPort
FileStoreReferenceURIPort
FileStoreURIPort
FilterDefinition
Gradient Boosting Model
Image
KnimeConnection
Moz API Connection
OpenNLPNERTaggerModelPortObject
Outlier
Python
R Workspace
Regression Tree
Regression Tree
Remote Connection
StanfordNERModelPortObject
Tree Ensembles
URI Object
URIPortObject managed by FileStore
VectorHashingPortObject
de.mpicbg.knime.scripting.r.port.RPortType2



images

whole R workspace

file located on the hard drive



Row ID	Protein IDs	Majority protein IDs	Peptide counts (all)	Peptide counts (Razor...)	Peptide counts (unique)	Protein names	Gene names	Fasta headers	Number of proteins	Peptides	Razor + unique peptides	Unique peptides	Peptides WT_1	Peptides WT_2
Row0	A0A096LPI6	A0A096LPI6	11		0	NA	NA	Uncharacterized protein (Fragment) OS=Homo sapiens PE=4 SV=2	1	11	9	0	7	6
Row1	A0A0A6YY...	A0A0A6YY...	4;4		4;4	Transme...	TMED7-TI...	Protein TMED7-TICAM2 OS=Homo sapiens GN=TMED7-TICAM2 PE=...	2	4	4	4	3	1
Row2	A0A0A6YY...	A0A0A6YY...	2;2		2;2	AP-3 com...	C15orf38...	Protein C15orf38-AP3S2 OS=Homo sapiens GN=C15orf38-AP3S2 ...	2	2	2	2	1	0
Row3	A0A...													
Row4	A0A...													
Row5	A0A0B4J2...	A0A0B4J2...	14;14		1;1	Tubulin b...	TUBB3	Uncharacterized protein OS=Homo sapiens PE=1 SV=1;Tubulin b...	2	14	1	1	11	11
Row6	A0A0B4J2...	A0A0B4J2...	15;15		13;13	Periodic t...	PWF...						15	13
Row7	A0A0J9YW...	A0A0J9YW...	2;2;2;2;2;...		2;2;... 2;2;2;2;2;...	Ig heavy ...	IGHV						2	2
Row8	A0A0U1R...	A0A0U1R...	1		1	NA	NA	Uncharacterized protein (Fragment) OS=Homo sapiens PE=1 SV=1	1	1	1	1	0	0
Row9	A0A1B0G...	A0A1B0G...	7;7		7;7	Cathepsi...	CTSD	Uncharacterized protein OS=Homo sapiens PE=1 SV=1;Cathepsi...	2	7	7	7	1	2
Row10	A0A1W2P...	A0A1W2P...	14		0	NA	NA	Uncharacterized protein OS=Homo sapiens PE=4 SV=1	1	14	2	0	11	11
Row11	A0A1W2P...	A0A1W2P...	9;2		2;2	NA	NA	RPS10-NUDT3 readthrough OS=Homo sapiens GN=RPS10-NUDT3 ...	2	9	2	2	7	8
Row12	A0A1W2P...	A0A1W2P...	3;2;1		3;2;1	Ester hyd...	C11orf54	Uncharacterized protein OS=Homo sapiens PE=4 SV=1;Ester hyd...	3	3	3	3	1	1
Row13	A0AVT1	A0AVT1	29		29	Ubiquitin-...	UBA6	Ubiquitin-like modifier-activating enzyme 6 OS=Homo sapiens GN=...	1	29	29	29	9	8
Row14	A0FGR8	A0FGR8	1		1	Extended...	ESYT2	Extended synaptotagmin-2 OS=Homo sapiens GN=ESYT2 PE=1 S...	1	1	1	1	0	0
Row15	A1L0T0	A1L0T0	6		6	Acetolact...	ILVBL	Acetolactate synthase-like protein OS=Homo sapiens GN=ILVBL P...	1	6	6	6	1	0
Row16	A3KN83	A3KN83	2		2	Protein s...	SBN01	Protein strawberry notch homolog 1 OS=Homo sapiens GN=SBNO...	1	2	2	2	0	0
Row17	A4D1E9	A4D1E9	4		4	Protein A	GTPBP10	Protein A OS=Homo sapiens GN=GTPBP10 PE=1 SV=1	1	4	4	4	0	1
Row18	A5YKK6	A5YKK6	20		20	CCR4-NO...	CNOT1	CCR4-NOT transcription complex subunit 1 OS=Homo sapiens GN...	1	20	20	20	6	6
Row19	A6NDG6	A6NDG6	13		13	Phospho...	PGP	Glycerol-3-phosphate phosphatase OS=Homo sapiens GN=PGP P...	1	13	13	13	7	7
Row20	A6NFQ2;A...	A6NFQ2;A...	2;1		2;1	TRPM8 ch...	TCAF2	TRPM8 channel-associated factor 2 OS=Homo sapiens GN=TCAF2...	2	2	2	2	1	1
Row21	A6NHQ2	A6NHQ2	4		1	rRNA/tRN...	FBLL1	rRNA/tRNA 2'-methyltransferase fibrillarin-like protein 1 OS=Hom...	1	4	1	1	3	4
Row22	A6NHR9	A6NHR9	14		14	Structura...	SMCHD1	Structural maintenance of chromosomes flexible hinge domain-co...	1	14	14	14	2	3
Row23	A6NJ78	A6NJ78	1		1	Probable ...	METTL15	Probable methyltransferase-like protein 15 OS=Homo sapiens GN...	1	1	1	1	0	0
Row24	A6NKT7;Q...	A6NKT7;Q...	18;16		1;1	RanBP2-li...	RGPD3;R...	RanBP2-like and GRIP domain-containing protein 3 OS=Homo sapi...	2	18	1	1	5	8
Row25	A8CG34;Q...	A8CG34;Q...	4;3		4;3	Nuclear e...	POM121C...	Nuclear envelope pore membrane protein POM 121C OS=Homo s...	2	4	4	4	0	0
Row26	A8MXV4;R...	A8MXV4	3;1;1		3;1;1	Nucleosi...	NUDT19	Nucleoside diphosphate-linked moiety X motif 19 OS=Homo sapie...	3	3	3	3	1	0
Row27	A9UHW6	A9UHW6	2		2	MIF4G do...	MIF4GD	MIF4G domain-containing protein OS=Homo sapiens GN=MIF4GD ...	1	2	2	2	0	0
Row28	B01T2	B01T2	39		38	Unconven...	MYO1G	Unconventional myosin-Ig OS=Homo sapiens GN=MYO1G PE=1 SV...	1	39	39	38	25	30
Row29	B4DLN1;Q...	B4DLN1;Q...	14;7		7;7	Mitochon...	SLC25A10	cDNA FLJ60124, highly similar to Mitochondrial dicarboxylate carrie...	2	14	14	7	10	10
Row30	P0CG08;B...	P0CG08;B...	1;1		1;1	Golgi pH ...	GPR89B;...	Golgi pH regulator B OS=Homo sapiens GN=GPR89B PE=1 SV=1;...	2	1	1	1	0	0
Row31	C9JAW5;Q...	C9JAW5;Q...	1;1		1;1	HIG1 dom...	HIGD1A	Uncharacterized protein OS=Homo sapiens PE=4 SV=1;HIG1 dom...	2	1	1	1	0	0
Row32	E7ENX8;P...	E7ENX8	11;2;1;1		1;1	2;2;1;1	NA	Uncharacterized protein (Fragment) OS=Homo sapiens PE=3 SV=1	4	11	11	2	7	6
Row33	E7EVH7;Q...	E7EVH7;Q...	11;11;3		1;1			Uncharacterized protein OS=Homo sapiens PE=4 SV=1;Kinesin lig...	3	11	11	11	4	4
Row34	E9PAV3;Q...	E9PAV3	5;1		1;1			Uncharacterized protein OS=Homo sapiens PE=4 SV=1;Myeloperoxidase-associated complex subunit alpha, muscle-s...	2	5	5	5	3	5
Row35	E9PL57;Q...	E9PL57;Q...	4;3;1		4;3;1	NEDD8	NEDD8-MDP1	readthrough (Fragment) OS=Homo sapiens GN=NEDD8-MDP1 ...	3	4	4	4	2	3
Row36	E9PLD3;E...	E9PLD3;E...	1;1		1;1	Uncharac...	C11orf98	Uncharacterized protein OS=Homo sapiens PE=4 SV=1;Uncharac...	2	1	1	1	1	1
Row37	E9PLN8;Q...	E9PLN8;Q...	1;1		1;1	Vitamin K...	VKORC1	Uncharacterized protein (Fragment) OS=Homo sapiens PE=4 SV=...	2	1	1	1	1	1
Row38	EEUEP2;P...	EEUEP2;P...	2;2		2;2	2 oxipoxy...	PCKDHA	Uncharacterized protein OS=Homo sapiens PE=4 SV=1;2 oxipoxy...	2	2	2	2	1	1

Rows: individual protein groups

Columns: characteristics of particular e.g. protein groups



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

Row ID	Protein IDs	Majority protein IDs	Peptide counts (all)	Peptide counts (Razor...)	Peptide counts (unique)	Protein names	Gene names	Fasta headers	Number of proteins	Peptides	Razor + unique peptides	Unique peptides	Peptides WT_1	Peptides WT_2
Row0	A0A096LPI6	A0A096LPI6	11		0	NA	NA	Uncharacterized protein (Fragment) OS=Homo sapiens PE=4 SV=2	1	11	9	0	7	6
Row1	A0A0A6YY...	A0A0A6YY...	4;4		4;4	Transme...	TMED7-TI...	Protein TMED7-TICAM2 OS=Homo sapiens GN=TMED7-TICAM2 PE=...	2	4	4	4	3	1
Row2	A0A0A6YY...	A0A0A6YY...	2;2		2;2	AP-3 com...	C15orf38...	Protein C15orf38-AP3S2 OS=Homo sapiens GN=C15orf38-AP3S2 ...	2	2	2	2	1	0
Row3	A0A...													
Row4	A0A...													
Row5	A0A0B4J2...	A0A0B4J2...	14;14		1;1	Tubulin b...	TUBB3	Uncharacterized protein OS=Homo sapiens PE=1 SV=1;Tubulin b...	2	14	1	1	11	11
Row6	A0A0B4J2...	A0A0B4J2...	15;15		13;13	Periodic t...	PWF...						15	13
Row7	A0A0J9YW...	A0A0J9YW...	2;2;2;2;2;...		2;2;... 2;2;2;2;2;...	Ig heavy ...	IGHV						2	2
Row8	A0A0U1R...	A0A0U1R...	1		1	NA	NA	Uncharacterized protein (Fragment) OS=Homo sapiens PE=1 SV=1	1	1	1	1	0	0
Row9	A0A1B0G...	A0A1B0G...	7;7		7;7	Cathepsi...	CTSD	Uncharacterized protein OS=Homo sapiens PE=1 SV=1;Cathepsi...	2	7	7	7	1	2
Row10	A0A1W2P...	A0A1W2P...	14		0	NA	NA	Uncharacterized protein OS=Homo sapiens PE=4 SV=1	1	14	2	0	11	11
Row11	A0A1W2P...	A0A1W2P...	9;2		2;2	NA	NA	RPS10-NUDT3 readthrough OS=Homo sapiens GN=RPS10-NUDT3 ...	2	9	2	2	7	8
Row12	A0A1W2P...	A0A1W2P...	3;2;1		3;2;1	Ester hyd...	C11orf54	Uncharacterized protein OS=Homo sapiens PE=4 SV=1;Ester hyd...	3	3	3	3	1	1
Row13	A0AVT1	A0AVT1	29		29	Ubiquitin-...	UBA6	Ubiquitin-like modifier-activating enzyme 6 OS=Homo sapiens GN=...	1	29	29	29	9	8
Row14	A0FGR8	A0FGR8	1		1	Extended...	ESYT2	Extended synaptotagmin-2 OS=Homo sapiens GN=ESYT2 PE=1 S...	1	1	1	1	0	0
Row15	A1L0T0	A1L0T0	6		6	Acetolact...	ILVBL	Acetolactate synthase-like protein OS=Homo sapiens GN=ILVBL ...	1	6	6	6	1	0
Row16	A3KN83	A3KN83	2		2	Protein s...	SBN01	Protein strawberry notch homolog						
Row17	A4D1E9	A4D1E9	4		4	Protein A	GTPBP10	Protein A OS=Homo sapiens GN=GTPBP10						
Row18	A5YKK6	A5YKK6	20		20	CCR4-NO...	CNOT1	CCR4-NOT transcription complex su						
Row19	A6NDG6	A6NDG6	13		13	Phospho...	PGP	Glycerol-3-phosphate phosphatase OS=Homo sapiens GN=PGP P...	1	13	13	13	7	7
Row20	A6NFQ2;A...	A6NFQ2;A...	2;1		2;1	TRPM8 ch...	TCAF2	TRPM8 channel-associated factor 2 OS=Homo sapiens GN=TCAF2...	2	2	2	2	1	1
Row21	A6NHQ2	A6NHQ2	4		1	rRNA/tRN...	FBLL1	rRNA/tRNA 2'-methyltransferase fibrillarin-like protein 1 OS=Hom...	1	4	1	1	3	4
Row22	A6NHR9	A6NHR9	14		14	Structura...	SMCHD1	Structural maintenance of chromosomes flexible hinge domain-co...	1	14	14	14	2	3
Row23	A6NJ78	A6NJ78	1		1	Probable ...	METTL15	Probable methyltransferase-like protein 15 OS=Homo sapiens GN...	1	1	1	1	1	0
Row24	A6NKT7;Q...	A6NKT7;Q...	18;16		1;1	RanBP2-li...	RGPD3;R...	RanBP2-like and GRIP domain-containing protein 3 OS=Homo sapi...	2	18	1	1	5	8
Row25	A8CG34;Q...	A8CG34;Q...	4;3		4;3	Nuclear e...	POM121C...	Nuclear envelope pore membrane protein POM 121C OS=Homo s...	2	4	4	4	0	0
Row26	A8MXV4;R...	A8MXV4	3;1;1		3;1;1	Nucleosi...	NUDT19	Nucleoside diphosphate-linked moiety X motif 19 OS=Homo sapie...	3	3	3	3	1	0
Row27	A9UHW6	A9UHW6	2		2	MIF4G do...	MIF4GD	MIF4G domain-containing protein OS=Homo sapiens GN=MIF4GD ...	1	2	2	2	0	0
Row28	B01T2	B01T2	39		38	Unconven...	MYO1G	Unconventional myosin-Ig OS=Homo sapiens GN=MYO1G PE=1 SV...	1	39	39	38	25	30
Row29	B4DLN1;Q...	B4DLN1;Q...	14;7		7;7	Mitochon...	SLC25A10	cDNA FLJ60124, highly similar to Mitochondrial dicarboxylate carrie...	2	14	14	7	10	10
Row30	P0CG08;B...	P0CG08;B...	1;1		1;1	Golgi pH ...	GPR89B;...	Golgi pH regulator B OS=Homo sapiens GN=GPR89B PE=1 SV=1;...	2	1	1	1	0	0
Row31	C9JAW5;Q...	C9JAW5;Q...	1;1		1;1	HIG1 dom...	HIGD1A	Uncharacterized protein OS=Homo sapiens PE=4 SV=1;HIG1 dom...	2	1	1	1	0	0
Row32	E7ENX8;P...	E7ENX8	11;2;1;1		1;1	2;2;1;1	NA	Uncharacterized protein (Fragment) OS=Homo sapiens PE=3 SV=1	4	11	11	2	7	6
Row33	E7EVH7;Q...	E7EVH7;Q...	11;11;3		1;1			Uncharacterized protein OS=Homo sapiens PE=4 SV=1;Kinesin lig...	3	11	11	11	4	4
Row34	E9PAV3;Q...	E9PAV3	5;1		1;1			Uncharacterized protein OS=Homo sapiens PE=4 SV=1;Typepeptide-associ...	2	5	5	5	3	5
Row35	E9PL57;Q...	E9PL57;Q...	4;3;1		4;3;1	NEDD8	NEDD8-MDP1	readthrough (Fragment) OS=Homo sapiens GN=NEDD8-MDP1 ...	3	4	4	4	2	3
Row36	E9PLD3;E...	E9PLD3;E...	1;1		1;1	Uncharac...	C11orf98	Uncharacterized protein OS=Homo sapiens PE=4 SV=1;Uncharac...	2	1	1	1	1	1
Row37	E9PLN8;Q...	E9PLN8;Q...	1;1		1;1	Vitamin K...	VKORC1	Uncharacterized protein (Fragment) OS=Homo sapiens PE=4 SV=...	2	1	1	1	1	1
Row38	EEUEP2;P...	EEUEP2;P...	2;2		2;2	2 oxipo...	PCKDHA	Uncharacterized protein OS=Homo sapiens PE=4 SV=1;2 oxipo...	2	2	2	2	1	1

Columns: characteristics of particular e.g. protein groups

Data processing -> “column-wise”
Adding new columns, not new rows (in comparison to Excel)

Rows: individual protein groups

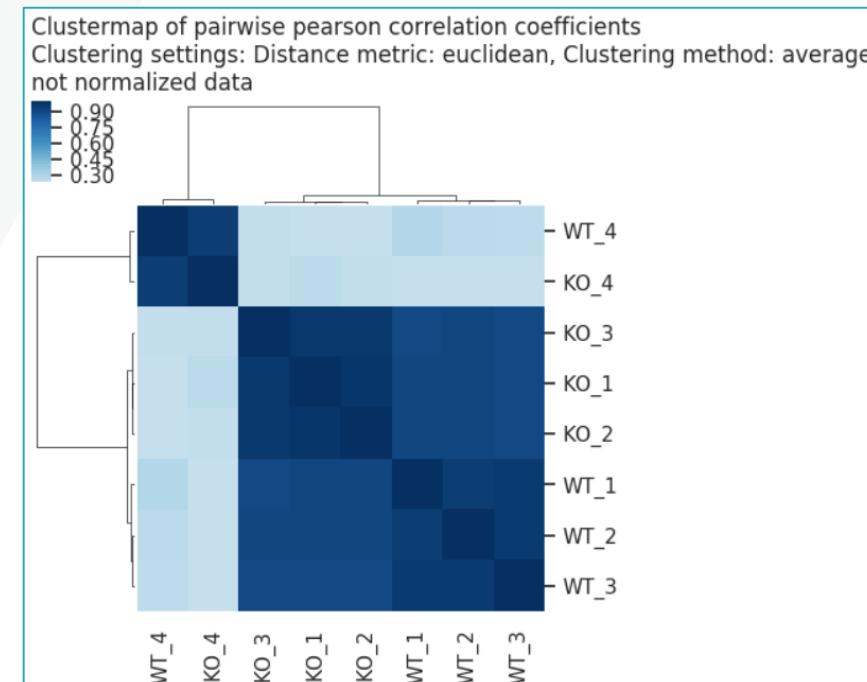
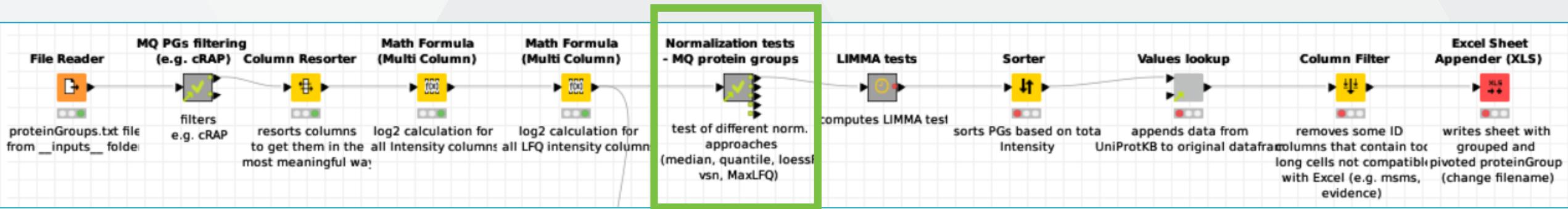


Row ID	Protein IDs	Majority protein IDs	Peptide counts (all)	Peptide counts (razor...)	Peptide counts (uniq...)	Protein names	Gene names	Fasta headers	Number of proteins	Peptides	Razor + unique peptides	Unique peptides	Peptides WT_1	Peptides WT_2
Row0	A0A096LPI6	A0A096LPI6	11	9	0	NA	NA	Uncharacterized protein (Fragment) OS=Homo sapiens PE=4 SV=2	1	11	9	0	7	6
Row1	A0A0A6YY...	A0A0A6YY...	4;4	4;4	4;4	Transmembrane domain-containing protein TMED7_TL	TMED7_TL	Protein TMED7_TL; Uncharacterized protein TMED7_TL OS=Homo sapiens GN=TMED7_TL PE=2 SV=2	2	4	4	4	3	1
Row2	A0A0A6YY...	A0A0A6YY...	2;2	2;2	2;2	AP-2-associated protein kinase	AP-2APK	Uncharacterized protein AP-2-associated protein kinase OS=Homo sapiens PE=2 SV=2	2	2	2	2	1	0
Row3	A0A0A6YY...	A0A0A6YY...	10;10;1	10;10;1	10;10;1	60S ribosomal protein L10	L10	Uncharacterized protein (Fragment) OS=Homo sapiens PE=4 SV=2	1	10	10	10	9	9
Row4	A0A0B4J1...	A0A0B4J1...	7;7	7;7	7;7	Superoxide dismutase	SOD1	Superoxide dismutase OS=Homo sapiens PE=3 SV=1	2	7	7	7	2	3
Row5	A0A0B4J2...	A0A0B4J2...	14;14	1;1	1;1	Tubulin beta-1 chain	TUBB1	Tubulin beta-1 chain OS=Homo sapiens PE=1 SV=1	2	14	1	1	11	11
Row6	A0A0B4J2...	A0A0B4J2...	15;15	15;15	13;13	Peroxiredoxin 2	PRDX2	Peroxiredoxin 2 OS=Homo sapiens PE=2 SV=2	2	15	15	13	6	4
Row7	A0A0J9YW...	A0A0J9YW...	2;2;2;2;2...	2;2;2;2;2...	2;2;2;2;2;...	Ig heavy chain	IGH	Ig heavy chain OS=Homo sapiens PE=1 SV=1	14	2	2	2	1	1
Row8	A0A0U1R...	A0A0U1R...	1	1	1	NA	NA	Uncharacterized protein OS=Homo sapiens PE=1 SV=1	1	1	1	1	0	0
Row9	A0A1B0G...	A0A1B0G...	7;7	7;7	7;7	Cathepsin D	CATD	Cathepsin D OS=Homo sapiens PE=1 SV=1	2	7	7	7	1	2
Row10	A0A1W2P...	A0A1W2P...	14	2	0	NA	NA	Uncharacterized protein OS=Homo sapiens PE=4 SV=1	1	14	2	0	11	11
Row11	A0A1W2P...	A0A1W2P...	9;2	2;2	2;2	NA	NA	RPS10-NUDT3 readthrough OS=Homo sapiens GN=RPS10-NUDT3	2	9	2	2	7	8
Row12	A0A1W2P...	A0A1W2P...	3;2;1	3;2;1	3;2;1	Ester hydrolase	C11orf54	Uncharacterized protein OS=Homo sapiens PE=4 SV=1; Ester hydrolase OS=Homo sapiens PE=1 SV=1	3	3	3	3	1	1
Row13	A0AVT1	A0AVT1	29	29	29	Ubiquitin-conjugating enzyme UBA6	UBA6	Ubiquitin-like modifier-activating enzyme 6 OS=Homo sapiens GN=UBA6 PE=1 SV=1	1	29	29	29	9	8
Row14	A0FGR8	A0FGR8	1	1	1	Extended synaptotagmin-2	SYT2	Extended synaptotagmin-2 OS=Homo sapiens GN=SYT2 PE=1 SV=1	1	1	1	1	0	0
Row15	A1L0T0	A1L0T0	6	6	6	Acetolactate synthase	ILVBL	Acetolactate synthase-like protein OS=Homo sapiens GN=ILVBL PE=1 SV=1	1	6	6	6	1	0
Row16	A3KN83	A3KN83	2	2	2	Protein S100-A1	SBNO1	Protein strawberry notch homolog 1 OS=Homo sapiens GN=SBNO1 PE=1 SV=1	1	2	2	2	0	0
Row17	A4D1E9	A4D1E9	4	4	4	Protein A	GTPBP10	Protein A OS=Homo sapiens GN=GTPBP10 PE=1 SV=1	1	4	4	4	0	1
Row18	A5YKK6	A5YKK6	20	20	20	CCR4-NOT transcription complex subunit 1	CNOT1	CCR4-NOT transcription complex subunit 1 OS=Homo sapiens GN=CNOT1 PE=1 SV=1	1	20	20	20	6	6
Row19	A6NDG6	A6NDG6	13	13	13	Phosphoglycerate kinase	PGP	Glycerol-3-phosphate phosphatase OS=Homo sapiens GN=PGP PE=1 SV=1	1	13	13	13	7	7
Row20	A6NFQ2;A...	A6NFQ2;A...	2;1	2;1	2;1	TRPM8 channel-associated factor 2	TCAF2	TRPM8 channel-associated factor 2 OS=Homo sapiens GN=TCAF2 PE=1 SV=1	2	2	2	2	1	1
Row21	A6NHQ2	A6NHQ2	4	1	1	rRNA/tRNA	FBLL1	rRNA/tRNA 2'-O-methyltransferase fibrillarin-like protein 1 OS=Homo sapiens GN=FBLL1 PE=1 SV=1	1	4	1	1	3	4
Row22	A6NHR9	A6NHR9	14	14	14	Structural protein of chromatin	SMCHD1	Structural maintenance of chromosomes flexible hinge domain-containing protein 1 OS=Homo sapiens GN=SMCHD1 PE=1 SV=1	1	14	14	14	2	3
Row23	A6NJ78	A6NJ78	1	1	1	Probable methyltransferase	METTL15	Probable methyltransferase-like protein 15 OS=Homo sapiens GN=METTL15 PE=1 SV=1	1	1	1	1	1	0
Row24	A6NKT7;Q...	A6NKT7;Q...	18;16	1;1	1;1	RanBP2-like domain-containing protein 3	RANBP2L3	RanBP2-like and GRIP domain-containing protein 3 OS=Homo sapiens GN=RANBP2L3 PE=1 SV=1	2	18	1	1	5	8
Row25	A8CG34;Q...	A8CG34;Q...	4;3	4;3	4;3	Nuclear envelope pore membrane protein POM121C	POM121C	Nuclear envelope pore membrane protein POM121C OS=Homo sapiens GN=POM121C PE=1 SV=1	2	4	4	4	0	0
Row26	A8MXV4;R...	A8MXV4	3;1;1	3;1;1	3;1;1	Nucleoside diphosphate-linked moiety X motif 19	NUDT19	Nucleoside diphosphate-linked moiety X motif 19 OS=Homo sapiens GN=NUDT19 PE=1 SV=1	3	3	3	3	1	0
Row27	A9UHW6	A9UHW6	2	2	2	MIF4G domain	MIF4GD	MIF4G domain-containing protein OS=Homo sapiens GN=MIF4GD PE=1 SV=1	1	2	2	2	0	0
Row28	B011T2	B011T2	39	39	38	Unconventional myosin-Ig	MYO1G	Unconventional myosin-Ig OS=Homo sapiens GN=MYO1G PE=1 SV=1	1	39	39	38	25	30
Row29	B4DLN1;Q...	B4DLN1;Q...	14;7	14;7	7;7	Mitochondrial membrane protein	SLC25A10	cDNA FLJ60124, highly similar to Mitochondrial dicarboxylate carrier OS=Homo sapiens GN=SLC25A10 PE=1 SV=1	2	14	14	7	10	10
Row30	P0CG08;B...	P0CG08;B...	1;1	1;1	1;1	Golgi pH regulator B	GPR89B	Golgi pH regulator B OS=Homo sapiens GN=GPR89B PE=1 SV=1	2	1	1	1	0	0
Row31	C9JAW5;Q...	C9JAW5;Q...	1;1	1;1	1;1	HIG1 domain-containing protein	HIGD1A	Uncharacterized protein OS=Homo sapiens PE=4 SV=1; HIG1 domain-containing protein OS=Homo sapiens PE=1 SV=1	2	1	1	1	0	0
Row32	E7ENX8;P...	E7ENX8	11;2;1;1	11;2;1;1	2;2;1;1	NA	NA	Uncharacterized protein (Fragment) OS=Homo sapiens PE=3 SV=1	4	11	11	2	7	6
Row33	E7EVH7;Q...	E7EVH7;Q...	11;11;3	11;11;3	11;11;3	Kinesin light chain	KLC1	Uncharacterized protein OS=Homo sapiens PE=4 SV=1; Kinesin light chain OS=Homo sapiens PE=1 SV=1	3	11	11	11	4	4
Row34	E9PAV3;Q...	E9PAV3	5;1	5;1	5;1	Nascent polypeptide-associated complex subunit alpha	NACA	Nascent polypeptide-associated complex subunit alpha, muscle-specific OS=Homo sapiens GN=NACA PE=1 SV=1	2	5	5	5	3	5
Row35	E9PL57;Q...	E9PL57;Q...	4;3;1	4;3;1	4;3;1	NEDD8	NEDD8	NEDD8-MDP1 readthrough (Fragment) OS=Homo sapiens GN=NEDD8-MDP1 PE=1 SV=1	3	4	4	4	2	3
Row36	E9PLD3;E...	E9PLD3;E...	1;1	1;1	1;1	Uncharacterized protein	C11orf98	Uncharacterized protein OS=Homo sapiens PE=4 SV=1; Uncharacterized protein OS=Homo sapiens PE=1 SV=1	2	1	1	1	1	1
Row37	E9PLN8;Q...	E9PLN8;Q...	1;1	1;1	1;1	Vitamin K-dependent protein X	VKORC1	Uncharacterized protein (Fragment) OS=Homo sapiens PE=4 SV=1	2	1	1	1	1	1
Row38	EEUEP2;P...	EEUEP2;P...	2;2	2;2	2;2	2-oxoacid oxidoreductase	RCDDH4	Uncharacterized protein OS=Homo sapiens PE=4 SV=1; Uncharacterized protein OS=Homo sapiens PE=1 SV=1	2	2	2	2	1	1

Columns can be of different data types:

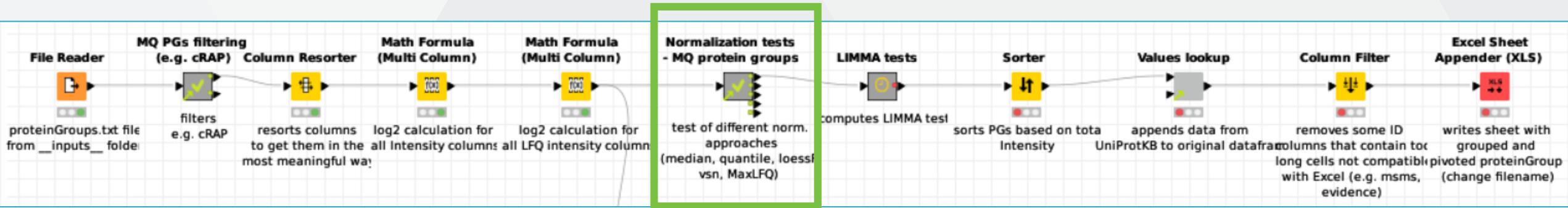
- S (string):** any text
- I (integer):** “whole” numbers
- D (double):** numbers with decimal places
- (date and time)**

KNIME environment and metanodes

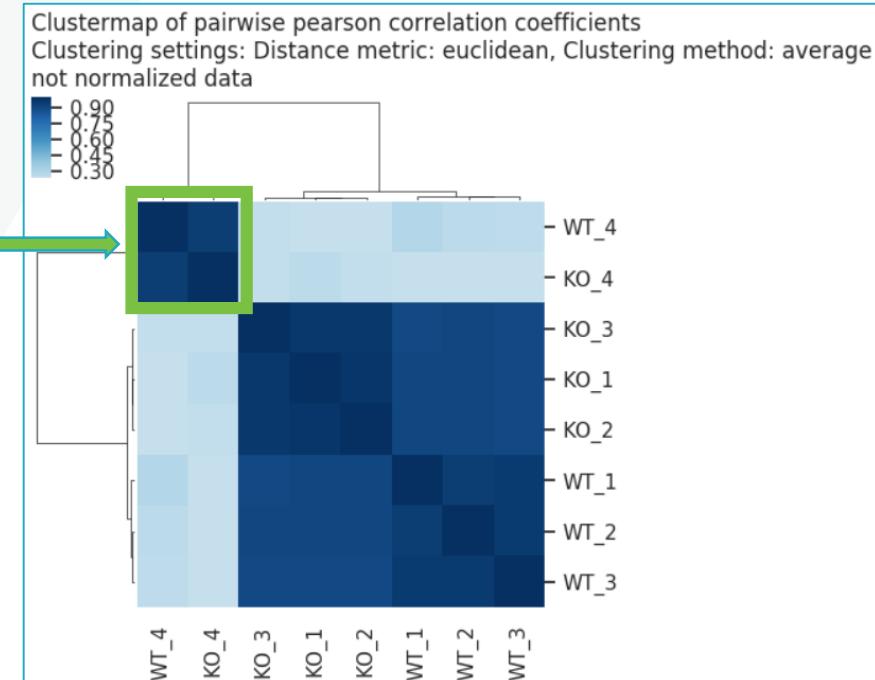


Alternative ways of data processing

KNIME environment and metanodes

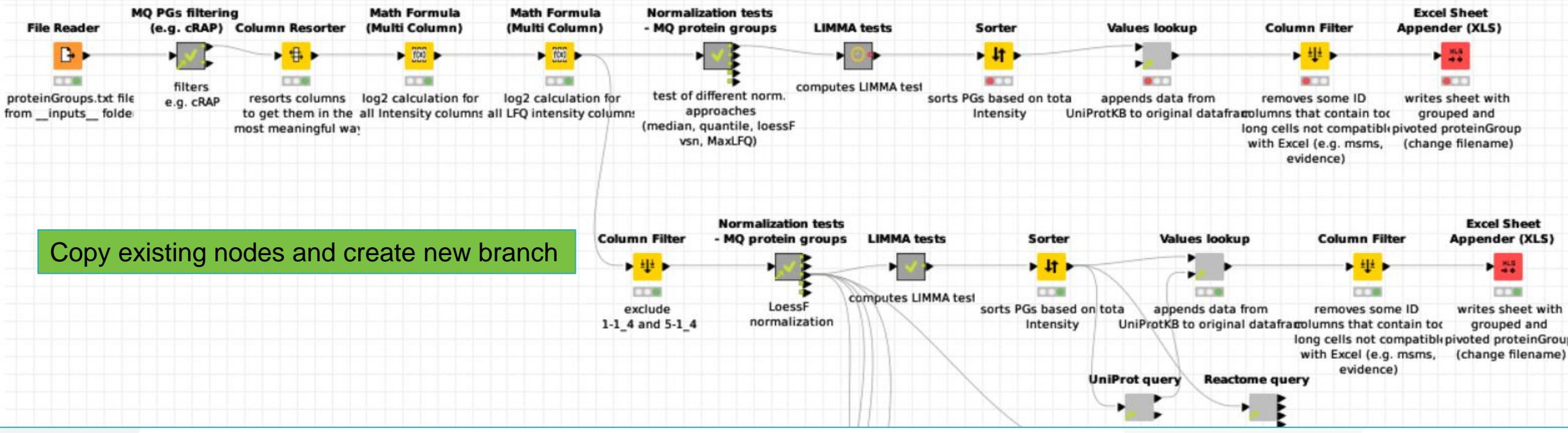


How will the clustering change
when outliers are excluded?

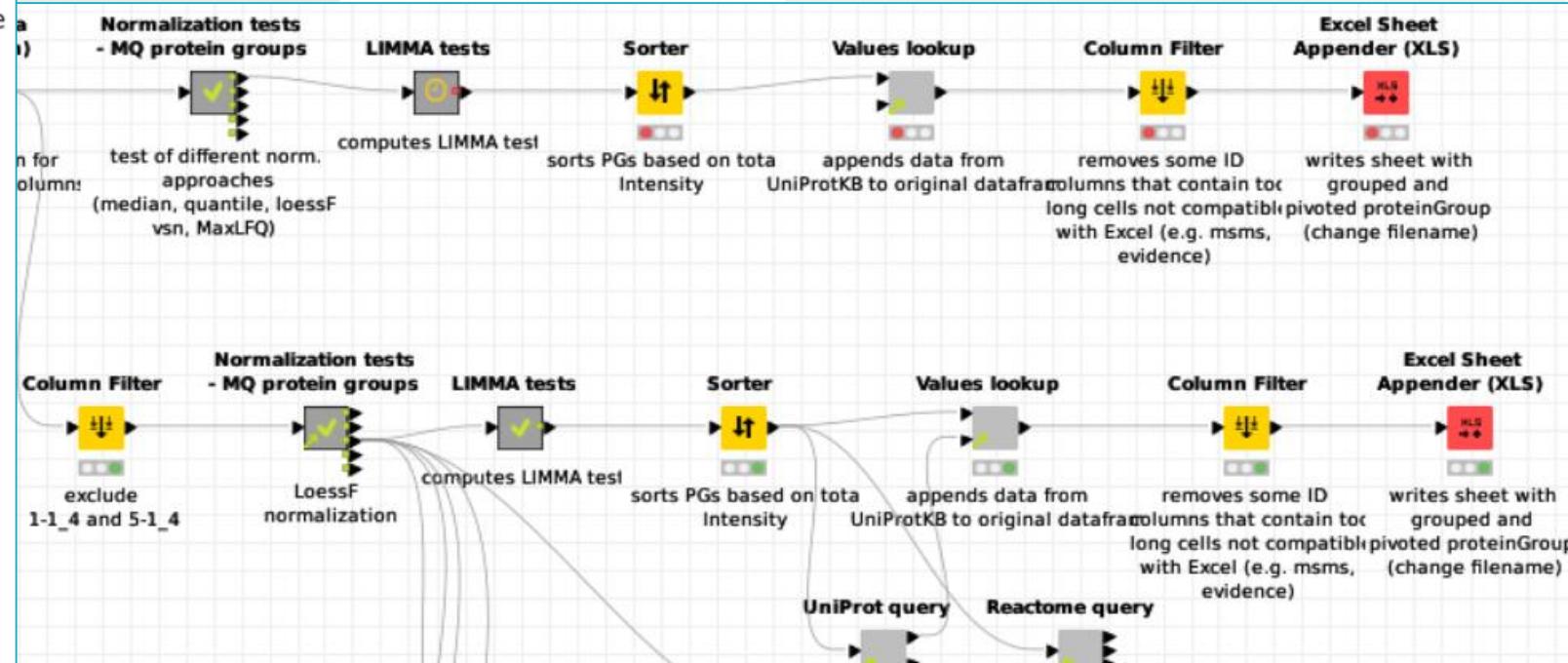
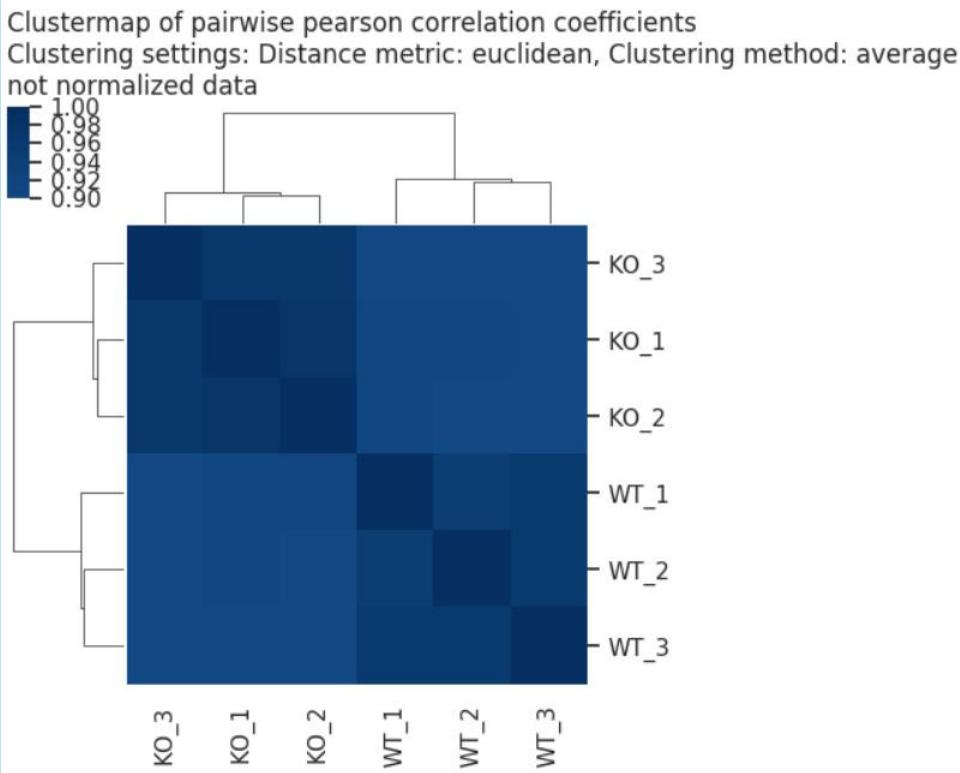


Alternative ways of data processing

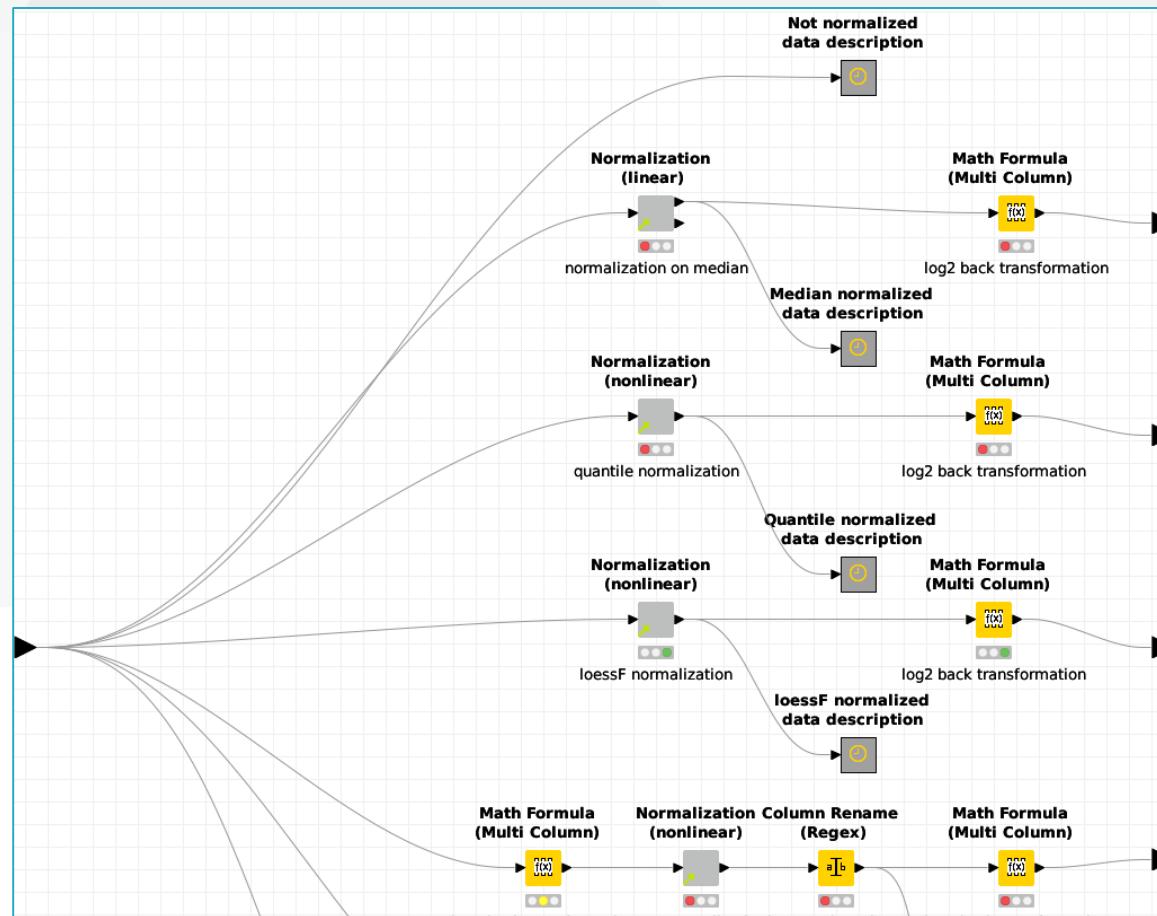
KNIME environment and metanodes



KNIME environment and metanodes

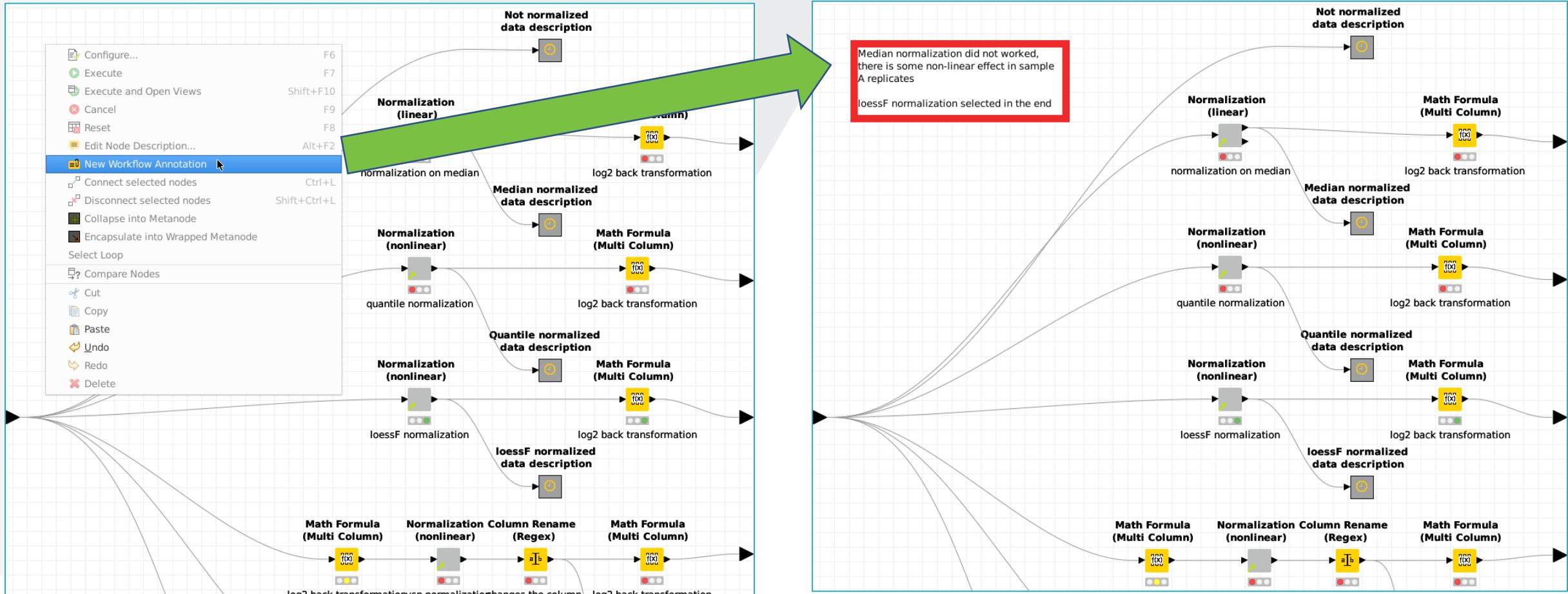


KNIME environment and metanodes



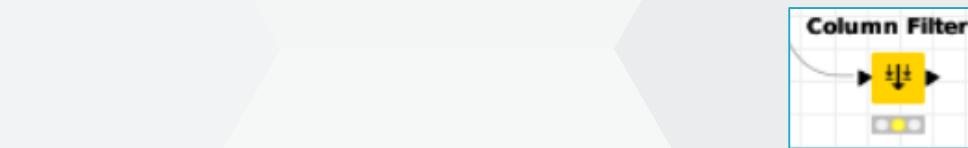
Alternative ways of data processing
Normalization metanode

KNIME environment and metanodes

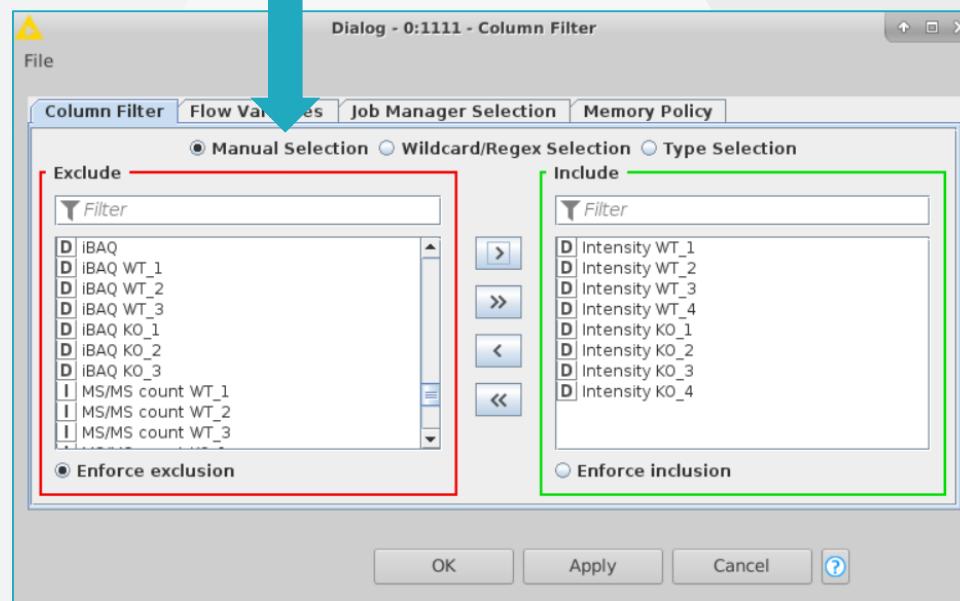


Add workflow annotation to describe your partial results or your reasons for the next processing steps

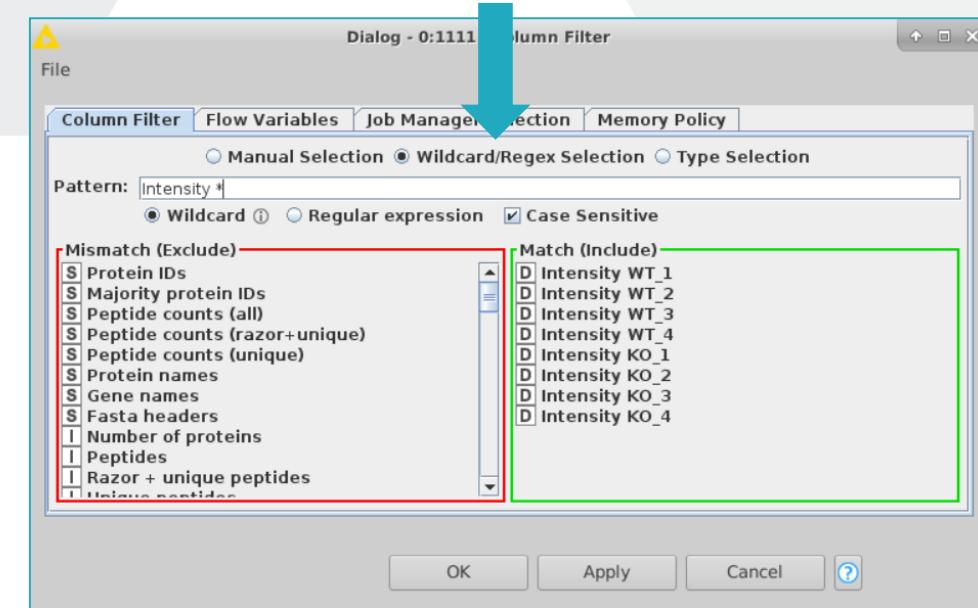
KNIME environment and metanodes



Selecting particular columns



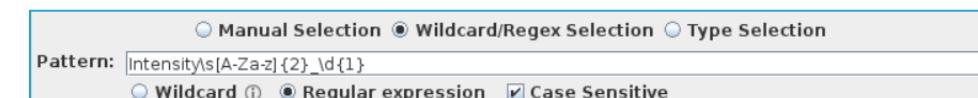
Manual Selection



Wildcard/Regex Selection

* any number of characters

? single character



KNIME environment and metanodes

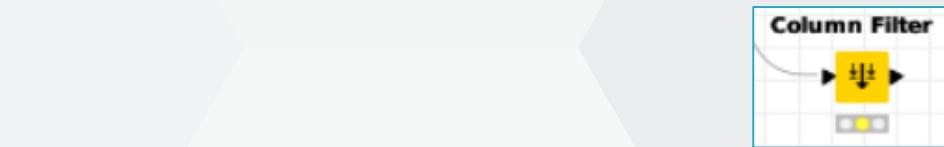


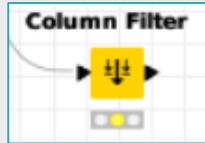
Table "proteinGroups.txt"								- Rows: 3855	- Spec - Columns: 81	Properties	Flow Variables
Row ID	Protein IDs	Majority protein IDs	Peptide counts (all)	Peptide counts (razor...)	Peptide counts (uniq...)	Protein names	Gene names				
Row0	A0A096LP16	A0A096LP16	11	9	0	NA	NA				
Row1	A0A0A6YY...	A0A0A6YY...	4;4	4;4	4;4	Transme...	TMED7-TI...				
Row2	A0A0A6YY...	A0A0A6YY...	2;2	2;2	2;2	AP-3 com...	C15orf38...				
Row3	A0A0A6YY...	A0A0A6YY...	10;10;1	10;10;1	10;10;1	60S ribos...	RPL17-C1...				
Row4	A0A0B4J1...	A0A0B4J1...	7;7	7;7	7;7	Suppress...	PPAN-P2...				
Row5	A0A0B4J2...	A0A0B4J2...	14;14	1;1	1;1	Tubulin b...	TUBB3				
Row6	A0A0B4J2...	A0A0B4J2...	15;15	15;15	13;13	Periodic t...	PWP2				
Row7	A0A0J9YW...	A0A0J9YW...	2;2;2;2;2...	2;2;2;2;2...	Ig heavy ...	IGHV4-61...					
Row8	A0A0U1R...	A0A0U1R...	1	1	1	NA	NA				
Row9	A0A1B0G...	A0A1B0G...	7;7	7;7	7;7	Cathepsi...	CTSD				
Row10	A0A1W2P...	A0A1W2P...	14	2	0	NA	NA				
Row11	A0A1W2P...	A0A1W2P...	9;2	2;2	2;2	NA	NA				
Row12	A0A1W2P...	A0A1W2P...	3;2;1	3;2;1	3;2;1	Ester hyd...	C11orf54				
Row13	A0AVT1	A0AVT1	29	29	29	Ubiquitin...	UBA6				
Row14	A0FGR8	A0FGR8	1	1	1	Extended...	ESYT2				
Row15	A1L0TO	A1L0TO	6	6	6	Acetolact...	ILVBL				
Row16	A3KN83	A3KN83	2	2	2	Protein s...	SBN01				
Row17	A4D1E9	A4D1E9	4	4	4	Protein A	GTPBP10				
Row18	A5YKK6	A5YKK6	20	20	20	CCR4-NO...	CNOT1				
Row19	A6NDG6	A6NDG6	13	13	13	Phospho...	PGP				
Row20	A6NFQ2;A...	A6NFQ2;A...	2;1	2;1	2;1	TRPM8 ch...	TCAF2				
Row21	A6NHQ2	A6NHQ2	4	1	1	rRNA/tRN...	FBL1				
Row22	A6NHR9	A6NHR9	14	14	14	Structura...	SMCHD1				
Row23	A6NJ78	A6NJ78	1	1	1	Probable...	METTL15				
Row24	A6NKT7;Q...	A6NKT7;Q...	18;16	1;1	1;1	RanBP2-li...	RGPD3;R...				
Row25	A8CG34;Q...	A8CG34;Q...	4;3	4;3	4;3	Nuclear e...	POM121C...				
Row26	A8MXV4;R...	A8MXV4	3;1;1	3;1;1	3;1;1	Nucleosi...	NUDT19				
Row27	A9UHW6	A9UHW6	2	2	2	MIF4G do...	MIF4GD				
Row28	B01T2	B01T2	39	39	38	Unconven...	MYO1G				
Row29	B4DLN1;Q...	B4DLN1;Q...	14;7	14;7	7;7	Mitochon...	SLC25A10				
Row30	P0CG08;B...	P0CG08;B...	1;1	1;1	1;1	Golgi pH ...	GPR89B;...				
Row31	C9JAW5;Q...	C9JAW5;Q...	1;1	1;1	1;1	HIG1 dom...	HIGD1A				
Row32	E7ENX8;P...	E7ENX8	11;2;1;1	11;2;1;1	2;2;1;1	NA	NA				
Row33	E7EVH7;Q...	E7EVH7;Q...	11;11;3	11;11;3	11;11;3	Kinesin li...	KLC1				
Row34	E9PAV3;Q...	E9PAV3	5;1	5;1	5;1	Nascent ...	NACA				
Row35	E9PL57;Q...	E9PL57;Q...	4;3;1	4;3;1	4;3;1	NEDD8	NEDD8-M...				
Row36	E9PLD3;E...	E9PLD3;E...	1;1	1;1	1;1	Uncharac...	C11orf98				
Row37	E9PLN8;Q...	E9PLN8;Q...	1;1	1;1	1;1	Vitamin K...	VKORC1				
Row38	EEUEP2;P...	EEUEP2;P...	2;2	2;2	2;2	2;2;2;2;2...	EEUEP2;P...				
Row39	EEUEP2;P...	EEUEP2;P...	2;2	2;2	2;2	2;2;2;2;2...	EEUEP2;P...				

Selecting particular columns

Table "default"										- Rows: 3855	- Spec - Columns: 8	Properties	Flow Variables
Row ID	Intensity WT_1	Intensity WT_2	Intensity WT_3	Intensity WT_4	Intensity KO_1	Intensity KO_2	Intensity KO_3	Intensity KO_4					
Row0	138,005,142.7...	12,334,974.4...	71,774,050.5...	15,572,388.0...	46,310,787.8...	65,411,765.7...	27,839,936.0...	28,964,138...					
Row1	?	?	?	?	13,352,870.9...	50,866,275.16	39,509,763.1...	20,286,929.2...	48,905,794...				
Row2	4,160,334,638...	?	?	?	3,590,791.711	3,540,430.878	1,125,920.393	?					
Row3	1,028,354,513...	314,222,598...	1,280,566,98...	15,563,821.8...	894,136,536...	1,134,082,69...	1,635,044,01...	17,004,683...					
Row4	6,383,802,317	4,694,509,839	34,252,056.6...	?	15,392,710.3...	29,521,666.2...	20,631,702.7...	26,533,305...					
Row5	?	?	?	?	?	22,386,282.3...	5,698,109,921	?					
Row6	45,039,556,163	6,157,646,112	83,816,383.7...	9,455,654,763	55,917,428.8...	44,384,862.8...	29,949,435.3...	6,222,054.0...					
Row7	46,969,859.13	14,405,226.6...	28,810,180.6...	?	45,626,826.2...	66,037,583.4...	25,067,767.0...	?					
Row8	?	?	?	?	?	?	?	?					
Row9	12,895,069,479	3,612,318,917	32,890,061.6...	391,393,647...	15,571,478.06	13,145,711.7...	4,972,123,133	430,975,71...					
Row10	?	?	?	?	17,645,879.3...	4,431,229,258	1,242,353,534	20,573,911...					
Row11	?	1,454,052,925	?	3,559,576,992	?	?	?	?					
Row12	7,278,034,617	2,113,210,345	6,789,013,352	213,736,654...	9,675,544,988	4,720,476,968	2,074,351,321	203,337,43...					
Row13	54,171,556,037	14,595,970.3...	68,329,068.9...	54,171,556.0...	58,593,352.1...	84,210,793.1...	38,816,331.2...	58,593,352...					
Row14	?	?	?	137,976,535...	?	3,716,122,177	?	?					
Row15	5,320,155,427	?	?	?	?	17,536,605.8...	13,779,263.62	10,612,251.5...	4,291,075.3...				
Row16	?	?	?	?	?	?	?	?					
Row17	?	2,046,101,616	10,655,091.6...	?	?	?	?	?					
Row18	55,347,664,163	13,088,101.8...	41,248,787.4...	359,827,171...	39,173,530.8...	32,866,689.33	15,099,435.2...	296,568,00...					
Row19	71,992,970,049	19,154,319.73	88,855,174.1	200,024,787...	85,563,561.2...	94,532,283.7	44,716,978.1...	246,184,02...					
Row20	890,579,777	?	?	1,366,094,69...	972,197,046	1,597,896,889	268,900,452	1,290,778.4...					
Row21	?	13,058,093.19	66,441,285.2...	5,275,453.284	50,187,124.4...	?	26,998,669.1...	11,795,860...					
Row22	59,882,689,882	6,212,733.741	56,689,054.7...	?	6,227,628,293	23,367,589.1...	5,757,035,496	7,194,820.0...					
Row23	98,544,334,268	20,479,323.18	84,373,772.8...	6,439,973,522	2,158,196.13	?	1,016,594,482	11,247,742...					
Row24	6,754,534,287	?	?	377,031,701...	8,765,007,713	8,480,396,246	2,499,980,107	404,221,96...					
Row25	?	?	?	2,766,398,83...	4,019,159,526	7,448,896,21	2,245,272,389	2,648,083,2...					
Row26	8,186,897.83	?	?	54,040,722.7...	5,461,522,912	5,859,424,259	4,621,569,045	43,169,566...					
Row27	101,043,892.8...	20,432,110.4...	89,429,608	1,206,242,33...	2,484,254,773	2,595,930,984	?	1,256,513.7...					
Row28	494,047,248.7...	11,268,416...	480,779,680...	494,047,248...	491,505,055...	522,048,431...	495,165,972...	491,505,05...					
Row29	204,876,470.3...	46,750,630.2...	172,037,542...	67,321,686.8...	130,567,297...	171,180,153...	109,584,291...	68,693,558...					
Row30	?	?	?	83,022,355.4...	?	?	?	?					
Row31	?	?	?	10,928,040.4...	5,932,887,497	?	?	?					
Row32	59,777,456.03	13,411,808.9...	73,261,192.3...	?	59,858,974.6...	66,362,341.8...	38,006,238.8...	50,187,124...					
Row33	14,671,669,374	4,404,999,811	34,091,238.5...	39,142,061.8...	33,545,136.1...	34,370,554,78	16,886,550.2...	34,878,282...					
Row34	138,916,388,388	20,677,775.3...	165,100,708.3...	1,198,523,51...	117,652,941...	134,983,192...	88,565,189.3...	1,244,078,1...					
Row35	77,916,388,388	21,069,492.6...	80,497,529.8...	?	72,284,195.8...	68,722,025.2...	38,381,032.2...	?					
Row36	31,876,367,938	5,416,680,068	27,488,776.4...	143,689,405...	23,858,141.79	20,890,186.3...	8,719,334,296	131,808,08...					
Row37	18,562,401,605	3,906,078,279	12,935,077.1...	?	18,680,678.3...	16,908,523.8...	7,826,302,112	4,950,995.7...					
Row38	2,568,251,553	645,107,373	1,699,465,973	?	2,626,481,715	2,596,831,245	617,488,994	?					
Row39	21,242,586,846	3,108,789,822	23,171,114.1...	?	20,206,887.1...	24,492,980.7...	7,895,203,846	2,732,346.1...					
Row40	?	?	?	?	?	?	?	?					

CEITEC at Masaryk University

KNIME environment and metanodes



Selecting particular columns

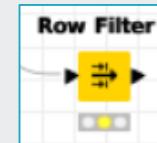
Table "proteinGroups.txt"							
	- Rows: 3855		Spec - Columns: 81		Properties		Flow Variables
Row ID	Protein IDs	Majority protein IDs	Peptide counts (all)	Peptide counts (razor...)	Peptide counts (uniq...)	Protein names	Gene names
Row0	A0A096LP16	A0A096LP16	11	9	0	NA	NA
Row1	A0A0A6YY...	A0A0A6YY...	4;4	4;4	4;4	Transme...	TMED7-TI...
Row2	A0A0A6YY...	A0A0A6YY...	2;2	2;2	2;2	AP-3 com...	C15orf38...
Row3	A0A0A6YY...	A0A0A6YY...	10;10;1	10;10;1	10;10;1	60S ribos...	RPL17-C1...
Row4	A0A0B4J1...	A0A0B4J1...	7;7	7;7	7;7	Suppress...	PPAN-P2...
Row5	A0A0B4J2...	A0A0B4J2...	14;14	1;1	1;1	Tubulin b...	TUBB3
Row6	A0A0B4J2...	A0A0B4J2...	15;15	15;15	13;13	Periodic t...	PWP2
Row7	A0A0J9YW...	A0A0J9YW...	2;2;2;2;2...	2;2;2;2;2...	Ig heavy ...	IGHV4-61...	
Row8	A0A0U1R...	A0A0U1R...	1	1	1	NA	NA
Row9	A0A1B0G...	A0A1B0G...	7;7	7;7	7;7	Cathepsi...	CTSD
Row10	A0A1W2P...	A0A1W2P...	14	2	0	NA	NA
Row11	A0A1W2P...	A0A1W2P...	9;2	2;2	2;2	NA	NA
Row12	A0A1W2P...	A0A1W2P...	3;2;1	3;2;1	3;2;1	Ester hyd...	C11orf54
Row13	A0AVT1	A0AVT1	29	29	29	Ubiquitin...	UBA6
Row14	A0FGR8	A0FGR8	1	1	1	Extended...	ESYT2
Row15	A1L0TO	A1L0TO	6	6	6	Acetolact...	ILVBL
Row16	A3KN83	A3KN83	2	2	2	Protein s...	SBNO1
Row17	A4D1E9	A4D1E9	4	4	4	Protein A	GTPBP10
Row18	A5YKK6	A5YKK6	20	20	20	CCR4-NO...	CNOT1
Row19	A6NDG6	A6NDG6	13	13	13	Phospho...	PGP
Row20	A6NFQ2;A...	A6NFQ2;A...	2;1	2;1	2;1	TRPM8 ch...	TCAF2
Row21	A6NHQ2	A6NHQ2	4	1	1	rRNA/tRN...	FBL1
Row22	A6NRH9	A6NRH9	14	14	14	Structura...	SMCHD1
Row23	A6NJ78	A6NJ78	1	1	1	Probable...	METTL15
Row24	A6NKT7;Q...	A6NKT7;Q...	18;16	1;1	1;1	RanBP2-li...	RGPD3;R...
Row25	A8CG34;Q...	A8CG34;Q...	4;3	4;3	4;3	Nuclear e...	POM121C...
Row26	A8MXV4;R...	A8MXV4	3;1;1	3;1;1	3;1;1	Nucleosi...	NUDT19
Row27	A9UHW6	A9UHW6	2	2	2	MIF4G do...	MIF4GD
Row28	B01T2	B01T2	39	39	38	Unconven...	MYO1G
Row29	B4DLN1;Q...	B4DLN1;Q...	14;7	14;7	7;7	Mitochon...	SLC25A10
Row30	P0CG08;B...	P0CG08;B...	1;1	1;1	1;1	Golgi pH ...	GPR89B;...
Row31	C9JAW5;Q...	C9JAW5;Q...	1;1	1;1	1;1	HIG1 dom...	HIGD1A
Row32	E7ENX8;P...	E7ENX8	11;2;1;1	11;2;1;1	2;2;1;1	NA	NA
Row33	E7EVH7;Q...	E7EVH7;Q...	11;11;3	11;11;3	11;11;3	Kinesin li...	KLC1
Row34	E9PAV3;Q...	E9PAV3	5;1	5;1	5;1	Nascent ...	NACA
Row35	E9PL57;Q...	E9PL57;Q...	4;3;1	4;3;1	4;3;1	NEDD8	NEDD8-M...
Row36	E9PLD3;E...	E9PLD3;E...	1;1	1;1	1;1	Uncharac...	C11orf98
Row37	E9PLN8;Q...	E9PLN8;Q...	1;1	1;1	1;1	Vitamin K...	VKORC1
Row38	EEUEP2;P...	EEUEP2;P...	2;2	2;2	2;2	2;2;2;2;2...	EEUEP2;P...

Filtering
Intensity * columns

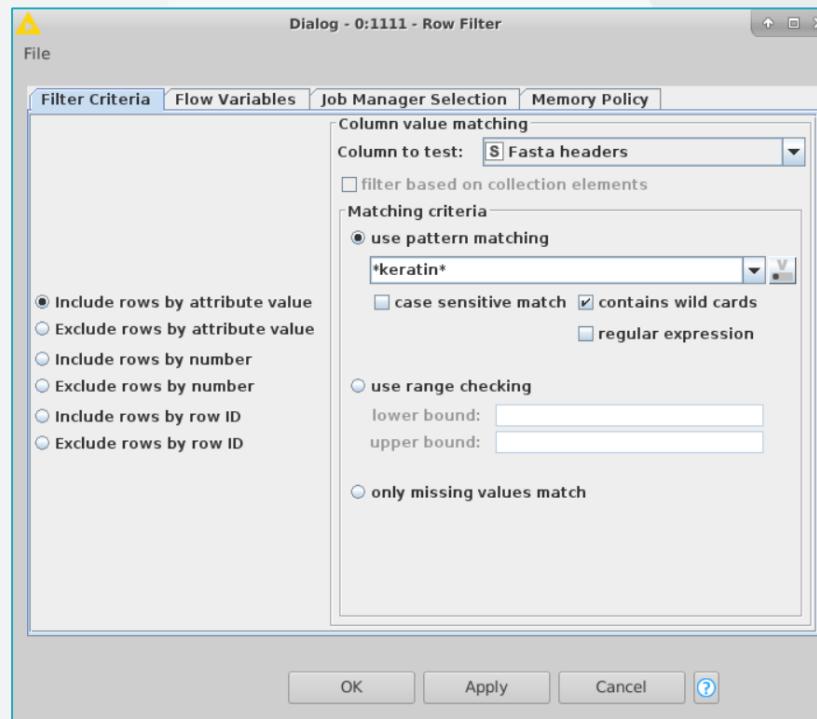
Table "default"							
	- Rows: 3855	Spec - Columns: 8	Properties	Flow Variables			
Row ID	Intensity WT_1	Intensity WT_2	Intensity WT_3	Intensity WT_4	Intensity KO_1	Intensity KO_2	Intensity KO_3
Row0	13,805,142.7...	12,334,974.4...	71,774,050.5...	15,572,388.0...	46,310,787.8...	65,411,765.7...	27,839,936.0...
Row1	?	?	?	?	13,352,870.9...	39,509,763.1...	20,286,929.2...
Row2	4,160,334,638	?	?	?	3,590,791.711	3,540,430,878	1,125,920,393
Row3	1,028,354,513...	314,222,598...	1,280,566,98...	15,563,821.8...	894,136,536...	1,134,082,69...	17,004,683...
Row4	6,383,802,317	4,694,509,839	34,252,056.6...	?	15,392,710.3...	29,521,666.2...	20,631,702.7...
Row5	?	?	?	?	?	22,386,282.3...	5,698,109,921
Row6	45,039,556,163	6,157,646,112	83,816,383.7...	9,455,654,763	55,917,428.8...	44,384,862.8...	29,949,435.3...
Row7	46,969,859.13	14,405,226.6...	28,810,180.6...	?	45,626,826.2...	66,037,583.4...	25,067,767.0...
Row8	?	?	?	?	?	?	?
Row9	12,895,069,479	3,612,318,917	32,890,061.6...	391,393,647...	15,571,478.06	13,145,711.7...	4,972,123,133
Row10	?	?	?	?	17,645,879.3...	4,431,229,258	1,242,353,534
Row11	?	1,454,052,925	?	3,559,576,992	?	?	?
Row12	7,278,034,617	2,113,210,345	6,789,013,352	213,736,654...	9,675,544,988	4,720,476,968	2,074,351,321
Row13	54,171,556,03	4,475,5,970.3...	68,1...	793.1...	38,816,331.2...	58,593,352...	
Row14	?	?	?	?	22,177	?	?
Row15	5,320,155,427	?	?	?	26,632	10,612,251.5...	4,291,075.3...
Row16	?	?	?	?	?	?	?
Row17	?	101.616	10,655,091.6...	?	?	?	?
Row18	55,347,664,163	13,088,101.8...	41,248,787.4...	359,827,171...	39,173,530.8...	32,866,689.33	15,099,435.2...
Row19	71,992,970,049	19,154,319.73	88,855,174.1	200,024,787...	85,563,561.2...	94,532,283.7	44,716,978.1...
Row20	890,579,777	?	?	1,366,094,69...	972,197,046	1,597,896,889	268,900,452
Row21	?	13,058,093.19	66,441,285.2...	5,275,453,284	50,187,124.4...	?	26,998,669.1...
Row22	59,882,689,882	6,212,733,741	56,689,054.7...	?	6,227,628,293	23,367,589.1...	5,757,035,496
Row23	98,544,334,268	20,479,323,18	84,373,772.8...	6,439,973,522	2,158,196.13	?	1,016,594,482
Row24	6,754,534,287	?	?	377,031,701...	8,765,007,713	8,480,396,246	2,499,980,107
Row25	?	?	?	2,766,398,83...	4,019,159,526	7,448,896,21	2,245,272,389
Row26	8,186,897.83	?	?	54,040,722.7...	5,461,522,912	5,859,424,259	4,621,569,045
Row27	101,043,892.8...	20,432,110.4...	89,429,608	1,206,242,33...	2,484,254,773	2,595,930,984	1,256,513,7...
Row28	494,047,248.7...	11,268,416...	480,779,680...	494,047,248...	491,505,055...	522,048,431...	495,165,972...
Row29	204,876,470.3...	46,750,630.2...	172,037,542...	67,321,686.8...	130,567,297...	171,180,153...	109,584,291...
Row30	?	?	?	83,022,355.4...	?	?	?
Row31	?	?	?	10,928,040.4...	5,932,887,497	?	?
Row32	59,777,456.03	13,411,808.9...	73,261,192.3...	?	59,858,974.6...	66,362,341.8...	38,006,238.8...
Row33	14,671,669,374	4,404,999,811	34,091,238.5...	39,142,061.8...	33,545,136.1...	34,370,554,78	16,886,550.2...
Row34	138,538,281.43	40,677,775.3...	165,100,708.3...	1,198,523,51...	117,652,941...	134,983,192...	88,565,189.3...
Row35	77,916,388,388	21,069,492.6...	80,497,529.8...	?	72,284,195.8...	68,722,025.2...	38,381,032.2...
Row36	31,876,367,938	5,416,680,068	27,488,776.4...	143,689,405...	23,858,141.79	20,890,186.3...	8,719,334,296
Row37	18,562,401,605	3,906,078,279	12,935,077.1...	?	18,680,678.3...	16,908,523.8...	7,826,320,112
Row38	2,568,251,553	645,107,373	1,699,465,973	?	2,626,481,715	2,596,831,245	617,488,994
Row39	21,242,586,846	3,108,789,822	23,171,114.1...	?	20,206,887.1...	24,492,980.7...	7,895,203,846
Row40	?	?	?	31,694,294.10	?	?	50,988,242

? : Missing values

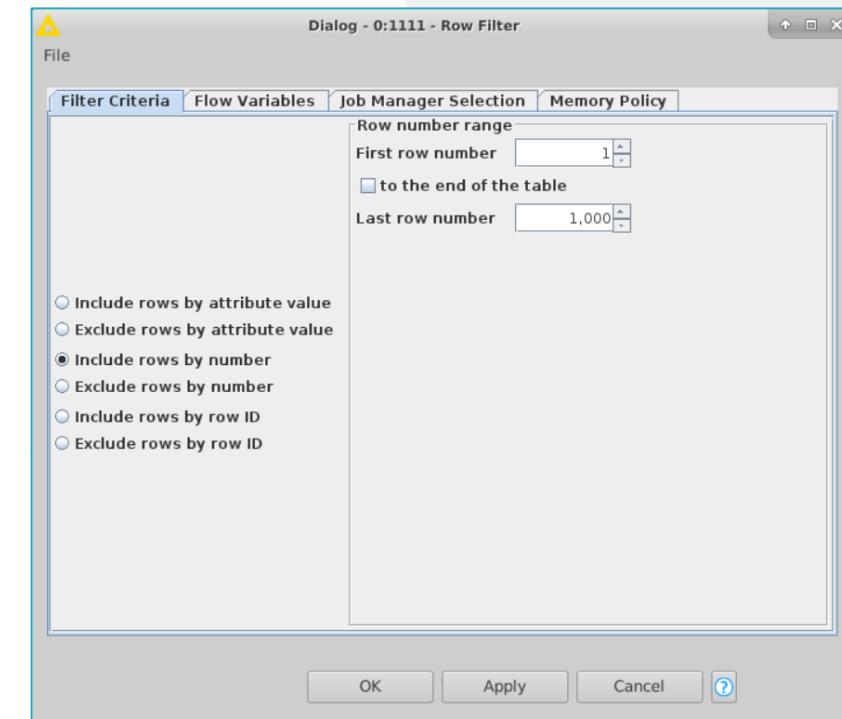
KNIME environment and metanodes



Selecting particular rows



Filtering using matching criteria

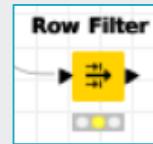


Filtering using rows range

KNIME environment and metanodes

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

Row ID	Protein IDs	Majority protein IDs	Peptide counts (all)	Peptide counts (razor...)	Peptide counts (uniq...)	Protein names	Gene names
Row0	A0A096LP16	A0A096LP16	11	9	0	NA	NA
Row1	A0A0A6YY...	A0A0A6YY...	4;4	4;4	4;4	Transme...	TMED7-Tl...
Row2	A0A0A6YY...	A0A0A6YY...	2;2	2;2	2;2	AP-3 com...	C15orf38...
Row3	A0A0A6YY...	A0A0A6YY...	10;10;1	10;10;1	10;10;1	60S ribos...	RPL17-C1...
Row4	A0A0B4J1...	A0A0B4J1...	7;7	7;7	7;7	Suppress...	PPAN-P2...
Row5	A0A0B4J2...	A0A0B4J2...	14;14	1;1	1;1	Tubulin b...	TUBB3
Row6	A0A0B4J2...	A0A0B4J2...	15;15	15;15	13;13	Periodic t...	PWP2
Row7	A0A0J9YW...	A0A0J9YW...	2;2;2;2;2;...	2;2;2;2;2;...	Ig heavy...	IGHV4-61...	
Row8	A0A0U1R...	A0A0U1R...	1	1	1	NA	NA
Row9	A0A1B0G...	A0A1B0G...	7;7	7;7	7;7	Cathepsi...	CTSD
Row10	A0A1W2P...	A0A1W2P...	14	2	0	NA	NA
Row11	A0A1W2P...	A0A1W2P...	9;2	2;2	2;2	NA	NA
Row12	A0A1W2P...	A0A1W2P...	3;2;1	3;2;1	3;2;1	Ester hyd...	C11orf54
Row13	A0AVT1	A0AVT1	29	29	29	Ubiquitin...	UBA6
Row14	A0FGR8	A0FGR8	1	1	1	Extended...	ESYT2
Row15	A1L0T0	A1L0T0	6	6	6	Acetolact...	ILVBL
Row16	A3KN83	A3KN83	2	2	2	Protein s...	SBNO1
Row17	A4D1E9	A4D1E9	4	4	4	Protein A	GTPBP10
Row18	A5YKK6	A5YKK6	20	20	20	CCR4-NO...	CNOT1
Row19	A6NDG6	A6NDG6	13	13	13	Phospho...	PGP
Row20	A6NFQ2;A...	A6NFQ2;A...	2;1	2;1	2;1	TRPM8 ch...	TCAF2
Row21	A6NHQ2	A6NHQ2	4	1	1	rRNA/tRN...	FBL1
Row22	A6NHR9	A6NHR9	14	14	14	Structura...	SMCHD1
Row23	A6NJ78	A6NJ78	1	1	1	Probable ...	METTL15
Row24	A6NKT7;Q...	A6NKT7;Q...	18;16	1;1	1;1	RanBP2-li...	RGPD3;R...
Row25	A8CG34;Q...	A8CG34;Q...	4;3	4;3	4;3	Nuclear e...	POM121C...
Row26	A8MXV4;R...	A8MXV4	3;1;1	3;1;1	3;1;1	Nucleosi...	NUDT19
Row27	A9UHW6	A9UHW6	2	2	2	MIF4G do...	MIF4GD
Row28	B0I1T2	B0I1T2	39	39	38	Unconven...	MYO1G
Row29	B4DLN1;Q...	B4DLN1;Q...	14;7	7;7	7;7	Mitochon...	SLC25A10
Row30	P0CG08;B...	P0CG08;B...	1;1	1;1	1;1	Golgi pH ...	GPR89B;...
Row31	C9JAW5;Q...	C9JAW5;Q...	1;1	1;1	1;1	HIG1 dom...	HIGD1A
Row32	E7ENX8;P...	E7ENX8	11;2;1;1	11;2;1;1	2;2;1;1	NA	NA
Row33	E7EVH7;Q...	E7EVH7;Q...	11;11;3	11;11;3	11;11;3	Kinesin li...	KLC1
Row34	E9PAV3;Q...	E9PAV3	5;1	5;1	5;1	Nascent ...	NACA
Row35	E9PL57;Q...	E9PL57;Q...	4;3;1	4;3;1	4;3;1	NEDD8	NEDD8-M...
Row36	E9PLD3;E...	E9PLD3;E...	1;1	1;1	1;1	Uncharac...	C11orf98
Row37	E9PLN8;Q...	E9PLN8;Q...	1;1	1;1	1;1	Vitamin K...	VKORC1
Row38	EEUEPDR...	EEUEPDR...	2;2	2;2	2;2	2;2;2;2	EEUEPDR...



Selecting particular rows



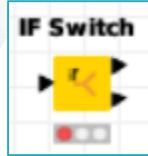
Table "proteinGroups.txt" - Rows: 2 Spec - Columns: 81 Properties Flow Variables

Row ID	Protein IDs	Majority protein IDs	Peptide counts (all)	Peptide counts (razor...)	Peptide counts (uniq...)	Protein names	Gene names	Fasta headers
Row3639	Q8N1A0	Q8N1A0	1	1	1	Keratin-like protein KRT222	KRT222	Keratin-like protein KRT222 OS=Homo sapiens GN=KRT222 PE=2 SV=1
Row3852	P04264-c...	P04264-c...	5;5;1;1;1;...	4;4;0;1;1;...	4;4;0;1;1;...	Keratin, type II cytoskelet...	KRT1	Keratin, type II cytoskeletal 1 (cRAP) OS=Homo sapiens GN=KRT1 PE...

Filtering rows containing *keratin* keyword in Fasta headers column

When you just need or want to script...

Visual programming
(in-built KNIME nodes)

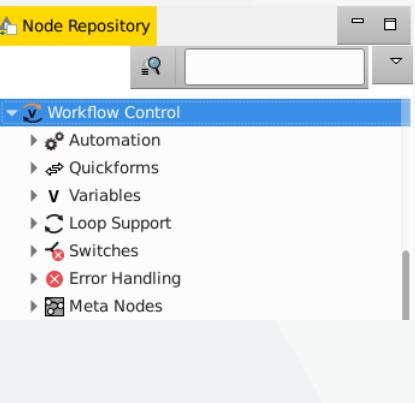


Scripting using snippets

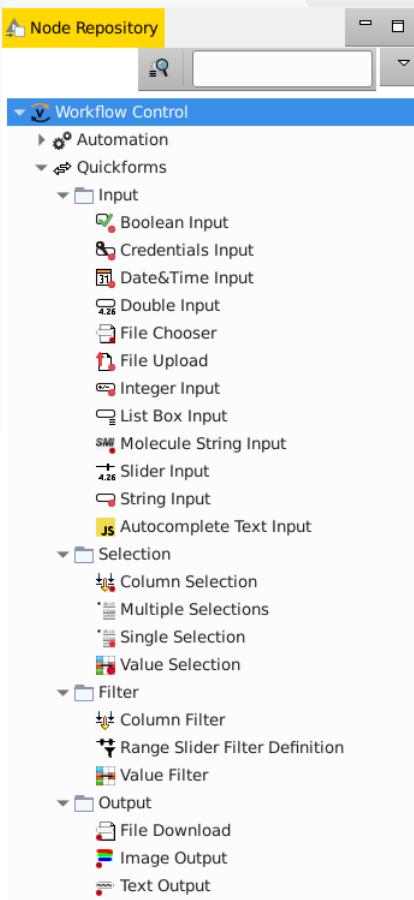


Visual programming concepts in KNIME

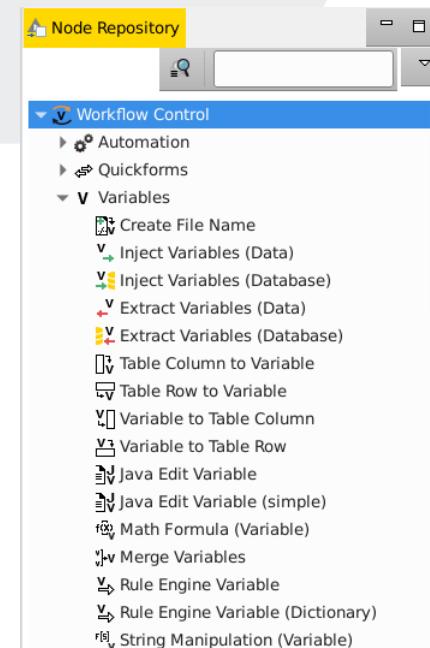
Workflow control nodes



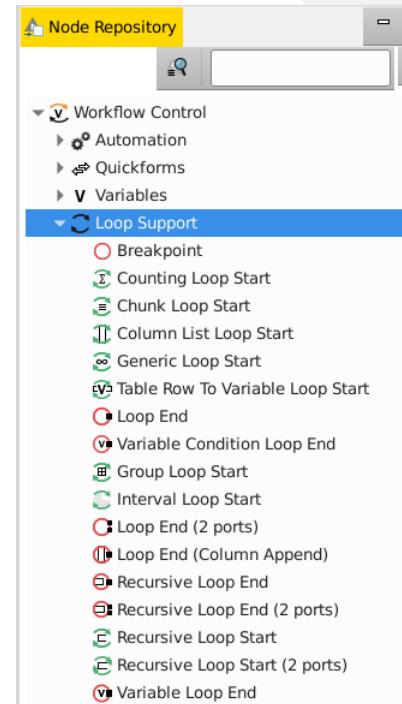
inputs support



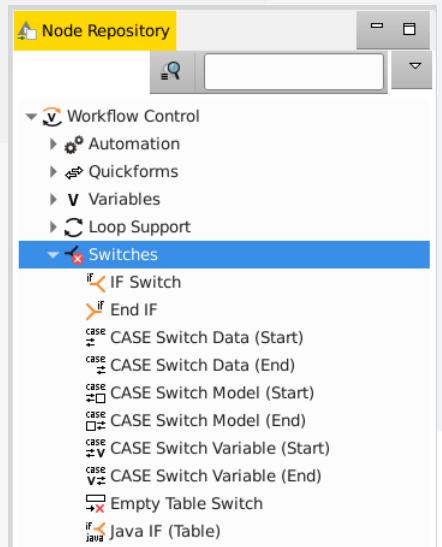
variables support



looping support



switches support



Scripting using snippets

Dialog - 0:964:0:269 - Python Script (1→2) (normalization)

File

Script Options Templates Flow Variables Job Manager Selection Memory Policy

Columns

- D Intensity 100-1_1_log2
- D Intensity 100-1_2_log2
- D Intensity 100-1_3_log2

Flow variables

- s if_switch
- s columns_suffix
- i remove_not_used
- s normalize_to
- i normalize_to (index)
- s normalization_method
- i normalization_method
- s columns_to_process
- s knime_workspace

```

1 # performs input dataframe normalization
2 import pandas
3
4 ## data import for KNIME version
5 data = input_table
6
7 # gets columns statistics based on the value on which the data is
8 columns_statistics = getattr(data, flow_variables['normalization'])
9
10 # creates empty dataframe to be filled by norm. factors
11 normalization_factors = pandas.DataFrame()
12
13 # gets normalization factors
14 normalization_factors['normalization factors'] = getattr(columns,
15
16 # normalize the data using the calculated normalization factors
17 data_normalized = data * normalization_factors['normalization fa
18
19 ## outputs normalized data
20 output_table_1 = data_normalized
21
22 ## outputs normalization factors in the orientation as input dat
23 output_table_2 = normalization_factors.T
24

```

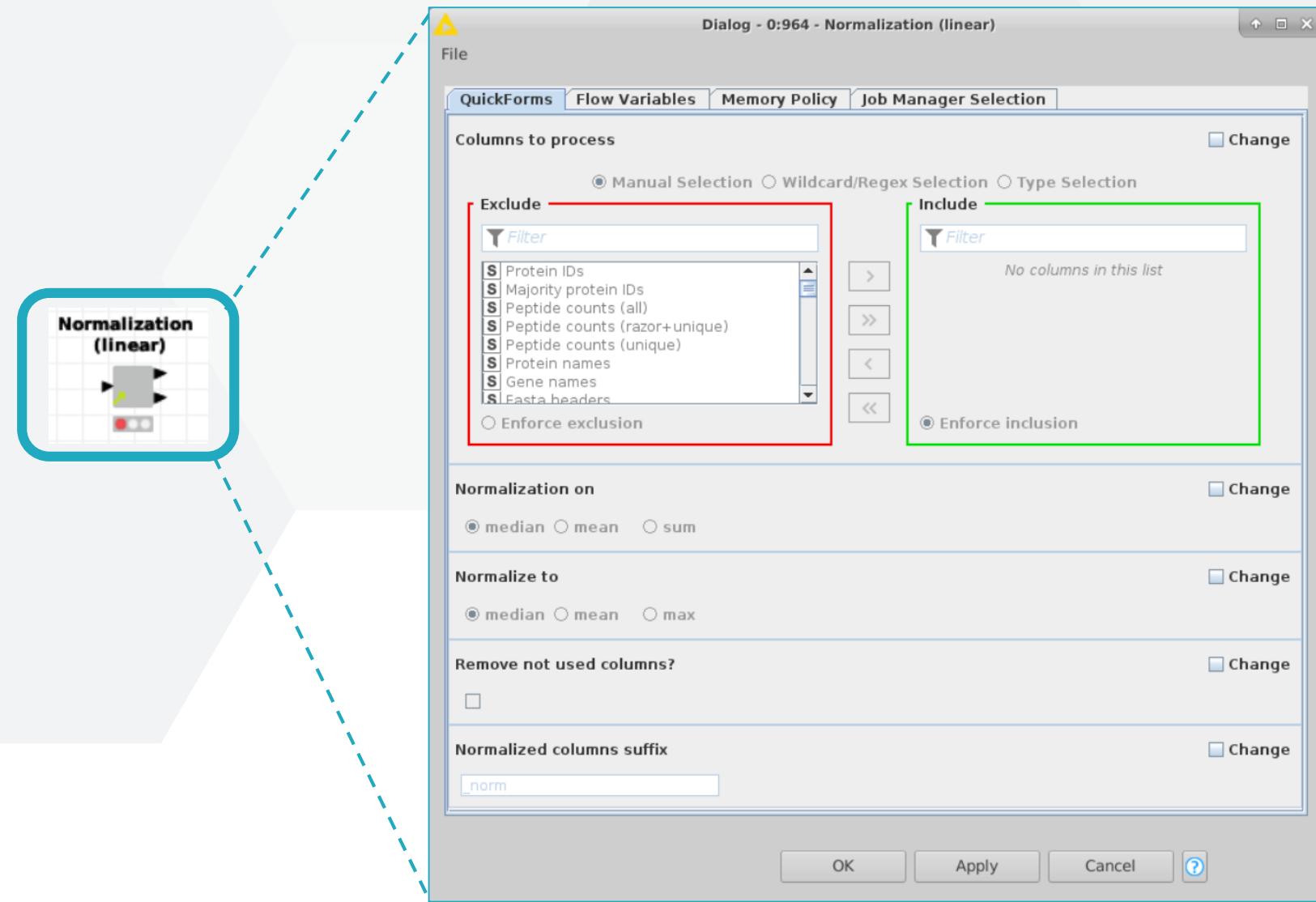
Execute script Execute selected lines Reset workspace

Successfully loaded input data into python

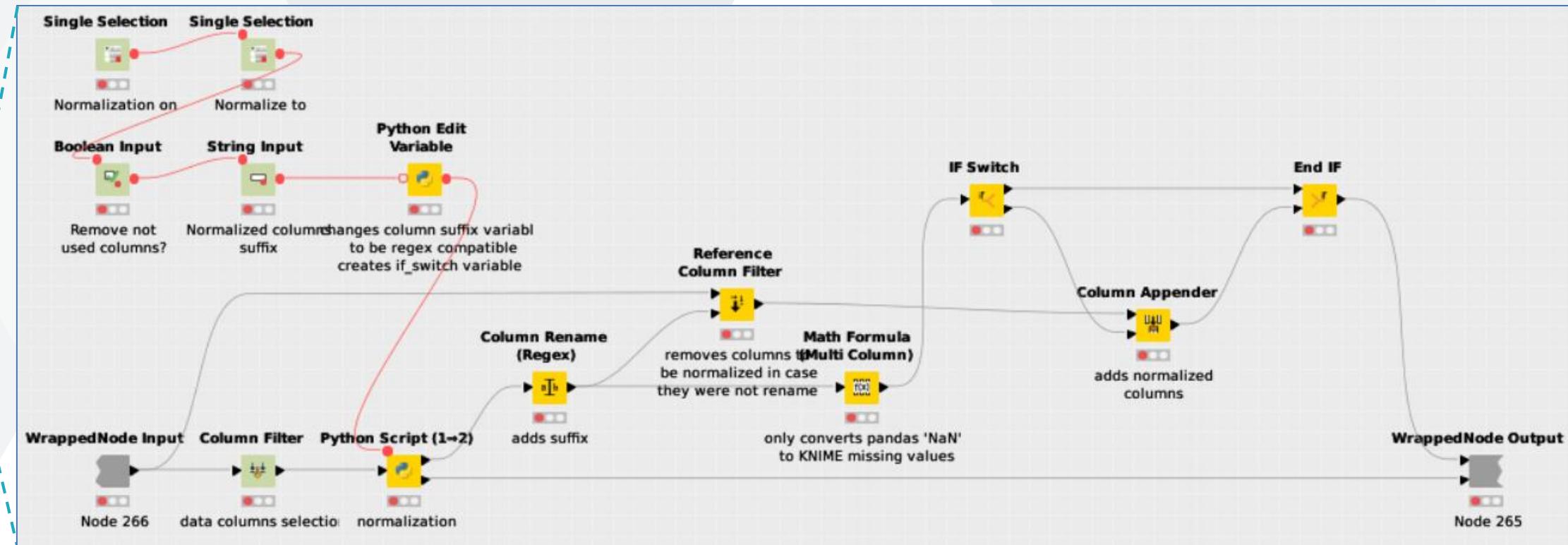
OK Apply Cancel ?

- R
- Python
- Java

Visual programming concepts in KNIME

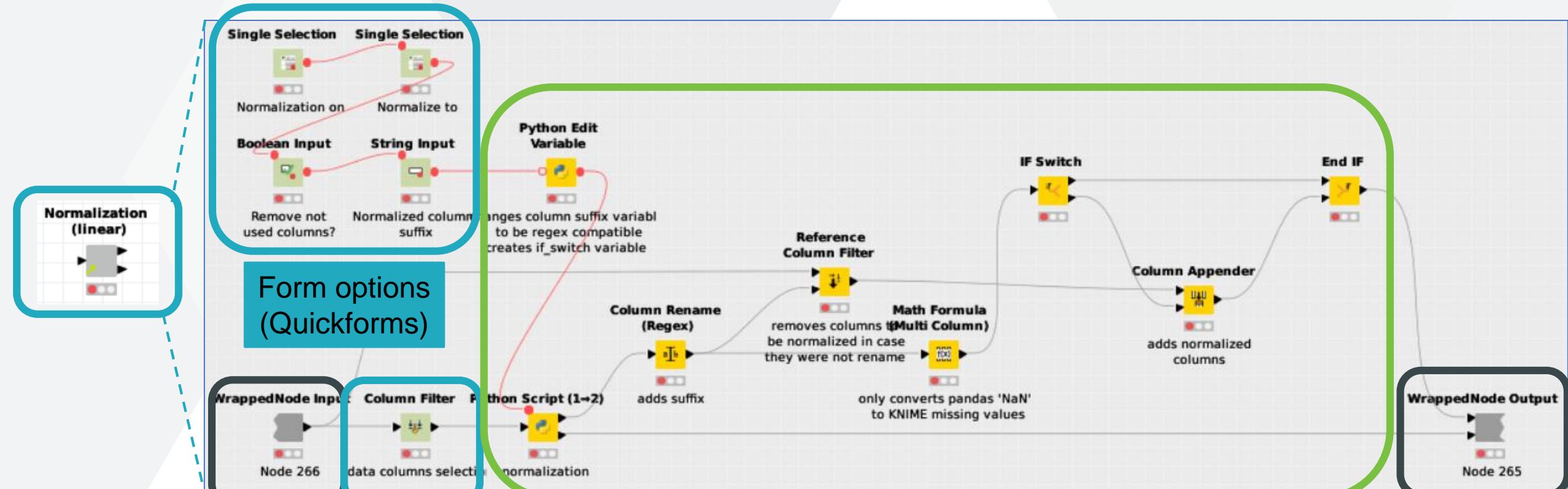


Visual programming concepts in KNIME



Inside the node... is another workflow

Visual programming concepts in KNIME



Visual programming concepts in KNIME

The diagram illustrates the visual programming concepts in KNIME, specifically focusing on data normalization and configuration.

Left Panel (Workflow Diagram):

- Normalization (linear) Node:** A node highlighted with a blue rounded rectangle. It takes a wrapped node input and produces a wrapped node output.
- Single Selection Nodes:** Two nodes highlighted with a green rounded rectangle. One is labeled "Normalization on" and the other is "Normalize to". They receive inputs from the "Normalization (linear)" node and send outputs to the "Python Edit Variable" node.
- Boolean Input and String Input Nodes:** These nodes receive inputs from the "Single Selection" nodes and send outputs to the "Python Edit Variable" node.
- Python Edit Variable Node:** This node receives inputs from the "String Input" and "Boolean Input" nodes. It has annotations: "changes column suffix variable to be regex compatible" and "creates if_switch variable". Its output goes to the "Column Rename (Regex)" node.
- Column Rename (Regex) Node:** This node receives input from the "Python Edit Variable" node and produces a wrapped node output.
- WrappedNode Input Node (Node 266):** This node receives a wrapped node input and sends it to the "Column Filter" node.
- Column Filter Node:** This node receives input from the "WrappedNode Input" node and sends it to the "Python Script (1→2)" node.
- Python Script (1→2) Node:** This node receives input from the "Column Filter" node and has annotations: "data columns selection" and "normalization". Its output goes to the "Column Rename (Regex)" node.
- Annotations:** A green dashed arrow points from the "Single Selection" nodes to the "Python Edit Variable" node. Another green dashed arrow points from the "Python Script (1→2)" node to the "Column Rename (Regex)" node.

Right Panel (Dialog Box):

Dialog - 0:964:0:258 - Single Selection (Normalization on)

Job Manager Selection Tab:

- Label:** Normalization on
- Description:** on what value within each data column the normalization should be done

Control Tab:

- Variable Name:** normalization_method
- Parameter Name:** single-selection
- Selection Type:** Radio buttons (horizontal)
- Possible Choices:** median, mean, sum
- Default Value:** median
- Limit number of visible options:**
- Number of visible options:** 5

Flow Variables Tab:

Specification of settings for particular form options

Visual programming concepts in KNIME

The diagram illustrates a KNIME workflow for data normalization. A 'Normalization (linear)' node (Node 266) is highlighted with a blue box. Its output feeds into a 'Single Selection' node, which has two parallel paths: one for 'Normalization on' and one for 'Normalize to'. These paths lead to 'Boolean Input' and 'String Input' nodes respectively. The 'Boolean Input' path also includes a 'Remove not used columns?' node. The 'String Input' path includes a 'Python Edit Variable' node. The 'Boolean Input' and 'String Input' nodes both feed into a 'Column Filter' node. The 'Column Filter' node outputs to a 'Python Script (1→2)' node, which is labeled 'normalization'. The 'Python Script' node outputs to a 'Column Rename (Regex)' node, which adds a suffix. The 'Column Rename' node outputs to a 'WrappedNode Output' node (Node 265).

A green dashed line highlights the 'Single Selection' node and its associated 'Normalization on' path.

Dialog - 0:964:0:258 - Single Selection (Normalization on)

Job Manager Selection (selected), **Memory Policy**, **Control**, **Flow Variables**

Label: Normalization on
Description: on what value within each data column the normalization should be done

Variable Name: normalization_method
Parameter Name: single-selection
Selection Type: Radio buttons (horizontal)
Possible Choices: median, mean, sum
Default Value: median
Limit number of visible options:
Number of visible options: 5

Specification of settings for particular form options
Variables can be further used for scripting

When you just need or want to script...

Python Script (1→2)

The screenshot shows the KNIME Python Script dialog (Dialog - 0:964:0:269 - Python Script (1→2) (normalization)). The dialog has tabs for Script, Options, Templates, Flow Variables, Job Manager Selection, and Memory Policy. The Script tab is active.

Script Content:

```

1 # performs input dataframe normalization
2 import pandas
3
4 ## data import for KNIME version
5 data = input_table
6
7 # gets columns statistics based on the value on which the data is
8 columns_statistics = getattr(data, flow_variables['normalization'])
9
10 # creates empty dataframe to be filled by norm. factors
11 normalization_factors = pandas.DataFrame()
12
13 # gets normalization factors
14 normalization_factors['normalization factors'] = getattr(columns,
15
16 # normalize the data using the calculated normalization factors
17 data_normalized = data * normalization_factors['normalization fa
18
19 ## outputs normalized data
20 output_table_1 = data_normalized
21
22 ## outputs normalization factors in the orientation as input dat
23 output_table_2 = normalization_factors.T
24

```

Flow Variables:

- if_switch
- columns_suffix
- remove_not_used
- normalize_to
- normalize_to (index)
- normalization_method
- normalization_method
- columns_to_process
- knime_workspace

Workspace Variables:

Name	Type	Value
knime_jupyter	module	
INT_SENTINEL	int	-2147483648
LONG_SENTINEL	int	-922337203...
flow_varia...	OrderedDict	OrderedDic...
input_table	DataFrame	I...
python_mes...	str	2
workspace	PythonKernel	<python3.P...

Buttons:

- Execute script
- Execute selected lines
- Reset workspace

Status Bar:

Successfully loaded input data into python

Dialog Buttons:

OK Apply Cancel ?

When you just need or want to script...

Python Script (1→2)

normalization

Input data

Flow variables

Code itself using input data and flow variables

Flow variables possible to be defined in the form or generated from the data

Dialog - 0:964:0:269 - Python Script (1→2) (normalization)

```

1 performs input dataframe normalization
2 import pandas
3
4 ## data import for KNIME version
5 data = input_table
6
7 # gets columns statistics based on the value on which the data is
8 columns_statistics = getattr(data, flow_variables['normalization'])
9
10 # creates empty dataframe to be filled by norm. factors
11 normalization_factors = pandas.DataFrame()
12
13 # gets normalization factors
14 normalization_factors['normalization factors'] = getattr(columns_
15
16 # normalize the data using the calculated normalization factors
17 data_normalized = data * normalization_factors['normalization fa
18
19 ## outputs normalized data
20 output_table_1 = data_normalized
21
22 ## outputs normalization factors in the orientation as input dat
23 output_table_2 = normalization_factors.T
24

```

Execute script Execute selected lines Reset workspace

Successfully loaded input data into python

OK Apply Cancel ?

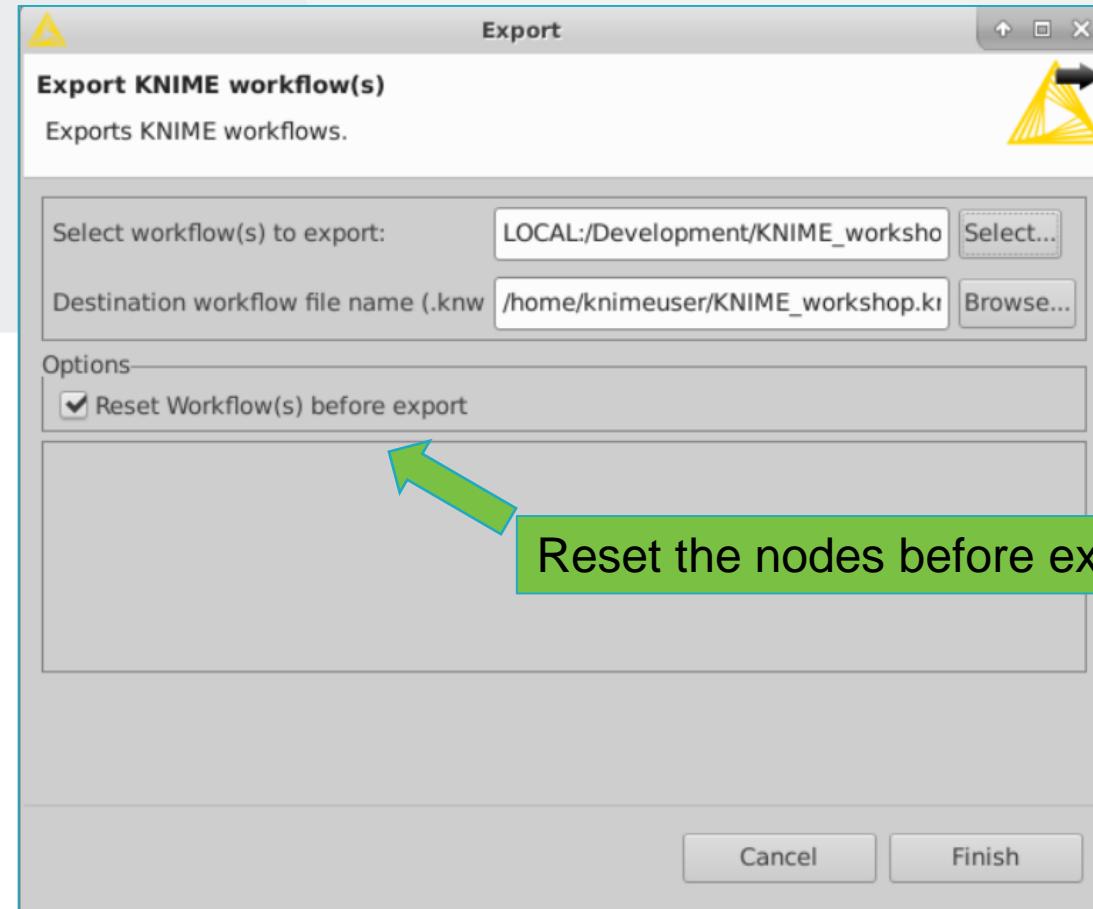
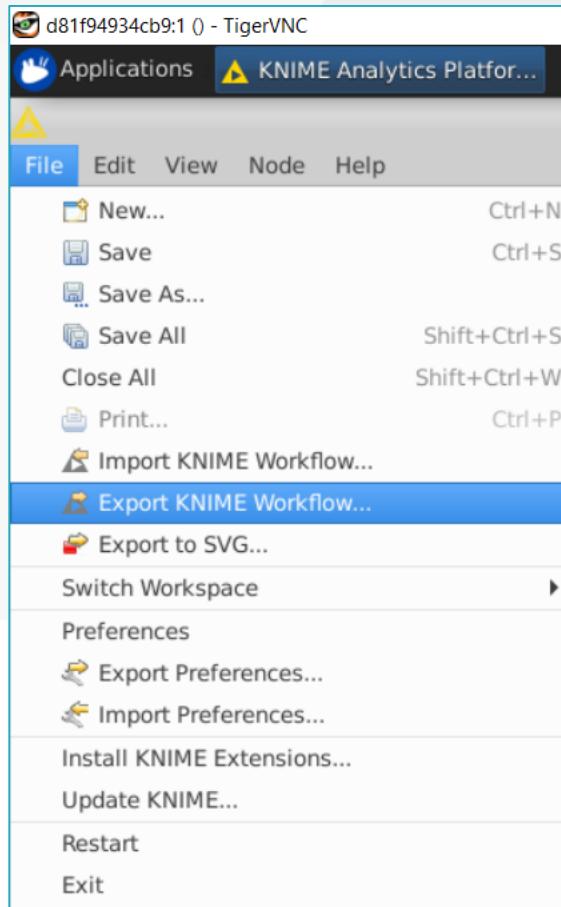
Name	Type	Value
knime_jupyter	module	
INT_SENTINEL	int	-2147483648
LONG_SENTINEL	int	-922337203...
flow_varia...	OrderedDict	OrderedDic...
input_table	DataFrame	I...
python_mes...	str	2
workspace	PythonKernel	<python3.P...

Metanodes sharing



Saving metanode as a template for sharing it with your colleagues or re-using in your workflows

Workflows sharing



Exporting workflows for easy sharing with your colleagues

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