

# Application container and KNIME metanodes for reproducible processing of omics data

Potěšil D.<sup>1</sup>, Gömöryová K.<sup>2</sup>, Bednařík M.<sup>1</sup>, Cupák M.<sup>1</sup>, Bryja V.<sup>2</sup>, Zdráhal Z.<sup>1</sup>

1 – Central European Institute of Technology, Masaryk University, Brno; 2 – Department of Experimental Biology, Faculty of Science, Masaryk University, Brno

## INTRODUCTION

Omics data processing pipelines are becoming increasingly complex and harder to reproduce. Software containers like docker containers can be implemented to make the processing environment reproducible.

Here we had two main objectives:

- 1) to build an environment enabling reproducible Omics data processing in visual, workflow-like way in KNIME Analytics Platform using open-source tools;
- 2) to develop KNIME metanodes extending the environment usability for selected applications.

## MATERIALS & METHODS

Docker container was built on Ubuntu 18.04 LTS with desktop environment and additional system components like Python, R, Java and with KNIME Analytics Platform. KNIME metanodes were developed using inbuilt nodes where possible and Python/R scripts were used for selected data processing steps or advanced statistics.

## RESULTS

### Figure 1: Software container and Metanodes templates design and utilization overview

**Software container** was built on Ubuntu LTS with xfce used as desktop environment and other system libraries. Access to the container is provided by using **VNC** (visual) and **shared folder** (files). **Python** and **R** are installed with minimal list of selected packages needed for Metanodes templates. The most recent version of **KNIME Analytics Platform** is installed including **additional plug-ins** needed for Metanodes templates and general processing steps in KNIME. **Additional scripts** for selected applications like Metanodes templates git repository reset are provided.

**Software container** is built based on the **docker file** and associated scripts available in the GitHub repository. Generated docker image is available on **docker hub**.

**Helper scripts** are available in GitHub repository next to docker file and other scripts to assist in usage of the docker image (starting docker container).

**KNIME Metanodes Templates** utilize **KNIME Form nodes** to get input information from the user and inbuilt KNIME processing nodes for basic operations. **R** and/or **python** scripts are used for advanced statistics and/or processing/visualization. **Each metanode documentation** (usage, utilized tools, python or R scripts) should make further metanode or scripts modifications possible/easy.

**KNIME Metanodes Templates** can be used **within the software container** but also in **standalone KNIME installation** – it may be required to install additional KNIME plugins or R/python packages depending on the Metanode template.

### Table 1: Metanodes templates overview

The overview of current list of metanodes.

Metanode Category	Metanode template name	Comments
Data processing	Normalization (linear)	e.g. normalization on median, average value
	Normalization (nonlinear)	e.g. loessF, vsn
	Ratio calculator	Including possible missing value imputation
	Signal-to-interference (S2I) correction	e.g. for TMT data correction
	Rows normalization (linear)	Rows based normalization
Databases	UniProt query	Gets e.g. GO terms based on protein UniProt ID
Graphs	Correlation clustermap	Cluster analysis based on corr. coefficients
	Hierarchical clustering	Cluster analysis based on raw data (intensities)
	Scatter plot matrix	Pairwise scatter/MA plots
	UpSet plot (interactive)	UpSet plot application wrapper
	Violin plot	
Manipulation	Volcano plot	Interactive version in preparation
	FASTA to table	
	Substring extraction	Multiple items per cell supported
	Table to FASTA	
Statistics	Values lookup	Multiple items per cell supported
	LIMMA test	Basic design only, paired and others in preparation

## Acknowledgement

Application container and KNIME metanodes templates build on open-source programs, tools and packages, please see <https://github.com/OmicsWorkflows> for their list.

The work was supported from European Regional Development Fund-Project "CIISB4HEALTH" (No. CZ.02.1.01/0.0/0.0/16\_013/0001776).

## CONCLUSIONS AND HIGHLIGHTS

- ✓ The application container has been prepared as standardized environment for data processing using KNIME Analytics Platform and is available under the GPL3 license
- ✓ KNIME Metanodes templates were developed for selected application in proteomic data processing and are available under the GPL3 license; other templates are in preparation
- ✓ The application container and KNIME metanodes templates are regularly maintained and updated
- ✓ Community contributions are welcomed!

CEITEC – Central European  
Institute of Technology  
Žerotínovo nám. 9, 601 77 Brno, Czech Republic  
+420 549 494 366, info@ceitec.cz  
[www.ceitec.eu](http://www.ceitec.eu)

Contact person  
David Potěšil  
e-mail [david.potesil@ceitec.muni.cz](mailto:david.potesil@ceitec.muni.cz)  
phone +420 549 498 258



MUNI  
SCI

CEITEC  
Central European Institute of Technology  
BRNO | CZECH REPUBLIC



EUROPEAN UNION  
European Structural and Investment Funds  
Operational Programme Research,  
Development and Education

