# Bioinformatic pipeline for comprehensive analysis of various small RNAs through RNA sequencing

Trachtova Karolina1,2, Bystry Vojtech1, Demko Martin1, Slaby Ondrej1,3

*1 Central European Institute of Technology, Masaryk University, 60177 Brno, Czech Republic*
*2 Christian Doppler Laboratory for Applied Metabolomics (CDL-AM), Medical University of Vienna, 1090 Vienna, Austria*

3 *Department of Biology, Faculty of Medicine, Masaryk University, 60177 Brno, Czech Republic*

Next-generation sequencing (NGS) is a revolutionary method that allows massive parallel sequencing of millions of DNA or RNA fragments. Although NGS is considered a state-of-the-art method, there is still a need for more comprehensive bioinformatical approaches, especially in the research of various small RNAs. In the sequencing of small RNAs, the crucial problem is accurate identification and quantification of a full spectrum of small RNA pool. Most of the available workflows are however targeted mostly at microRNA and ignore other RNA types such as snoRNA, snRNA, piRNA, or isomiRs.

We propose here a bioinformatic pipeline for accurate quantification of all known small RNAs classes. Our pipeline is divided into stand-alone modules, each focusing on one part of the sequencing data analysis (first quality control, pre-processing, RNA quantification and differential expression analysis). The most significant is the RNA quantification module, where a subsequent number of mapping rounds, utilizing reference sequences collected from several resources, ensure quantification of all different small non-coding RNAs. Custom Python tool was created to count reads assigned to different RNAs that also address an issue of multi-loci RNAs (such as piRNA) and problem of overlapping RNA annotations.

Each module provides a PDF/HTML report summarizing results, including tables, plots and their explanation and so guiding the user when exploring different small RNA expression levels. To smooth utilization of report plots for publication, we also offer an interactive application implemented in Shiny for a real-time visualization of differential expression results where the content and appearance of popular plots such as heatmap, PCA or volcano plot can be easily altered.

*Core Facility Bioinformatics of CEITEC Masaryk University is gratefully acknowledged for the obtaining of the scientific data presented here.*