

# miRTop: An open source community project for the development of a unified format file for miRNA data

Thomas Desvignes<sup>2</sup>, Karen Ellbeck<sup>3</sup>, Ioannis S. Vlachos<sup>4</sup>, Bastian Fromm<sup>5</sup>, Marc K. Halushka<sup>6</sup>, Michael Hackenberg<sup>7</sup>, Gianvito Urgese<sup>8</sup>, Elisa Ficarra<sup>8</sup>, Shruthi Bandyadka<sup>9</sup>, Jason Sydes<sup>2</sup>, Peter Batzel<sup>2</sup>, John H. Postlethwait<sup>2</sup>, Phillipe Loher<sup>10</sup>, Eric London<sup>10</sup>, Aristeidis G. Telonis<sup>10</sup>, Isidore Rigoutsos<sup>10</sup>, [Lorena Pantano](#)<sup>1</sup>

<sup>1</sup> Harvard School TH Chan of Public Health, Boston, MA, USA. Email: [lpantano@hsph.harvard.edu](mailto:lpantano@hsph.harvard.edu)

<sup>2</sup> Institute of Neuroscience, University of Oregon, Eugene, OR, USA.

<sup>3</sup> University of Utah, Biomedical Informatics, UT, USA.

<sup>4</sup> Brigham & Women's Hospital, Broad Institute of MIT and Harvard, Harvard Medical School, Cambridge, MA, USA.

<sup>5</sup> Department of Tumor Biology. Institute for Cancer Research. Norwegian Radium Hospital. Oslo University Hospital, Norwegian.

<sup>6</sup> Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

<sup>7</sup> Bioinformatics Group. University of Granada, Spain.

<sup>8</sup> Politecnico di Torino, Italy.

<sup>9</sup> Partners Personalized Medicine, Cambridge, MA, USA.

<sup>10</sup> Computational Medicine Center, Thomas Jefferson University, Philadelphia, PA, USA

**Project Website:** <http://mirtop.github.io>

**Source Code:** <https://github.com/miRTop/mirtop>

**License:** MIT

MicroRNAs (miRNAs) are small RNA molecules (20-27 nt long) that are involved in eukaryotic gene regulation. They regulate targeted genes by RNA complementarity, generally between the miRNA seed region and the 3' UTR of messenger RNAs. They have become relevant to the study of developmental stages and diseases. With the advance of sequencing technology, the cost of detecting the miRNA transcriptome has decreased substantially, and has led to the discovery of isomiRs (with slight sequence variations relative to the annotated miRNAs). As a consequence, the amount of miRNA data has increased exponentially in the last 5 years, together with the number of pipelines to analyze it. However, there is still a lack of consensus standards in storing and sharing such data, generating diverse output files; hindering comparison between tools, data sharing, and development of downstream analyses independent of the tools used for miRNA detection and quantification. Here, we present a community based, open source project to work toward the standardization of miRNA pipelines and encourage the development of downstream tools for visualization, differential expression, sample clustering, and model prediction analyses. This project is an international collaboration, with experts from different countries that have been developing miRNA pipelines and resources. We have described a standard format for the output of miRNA detection and quantification tools using small RNA-seq data. The format is based on the GFF3 standard in order to support reference coordinates and parent/child relationships between the features. The [format](#) contains information for each sequence and its annotation to the miRNA precursor, the definition of reference or isomiR sequence, its quality, the isomiR type, and abundance in the data set. Moreover, we support a command line python tool to manage the miRNA GFF3 format ([miRTop](#)). Currently, miRTop can convert the output of commonly used small RNA-Seq pipelines, such as seqbuster, isomiR-SEA, sRNAbench, and Prost!, as well as BAM files to the miRNA GFF3 format. Importantly, the miRge pipeline has adapted the GFF3 format natively. miRTop can convert miRNA GFF3 files to a count matrix that can be easily imported to any downstream tool (i.e. for differential expression analysis). Furthermore, the tool can export the miRNA GFF file to the [isomiRs](#) package, improving the usability of this tool and rendering it independent of the upstream methods used for quantifying miRNA sequences. Together, we believe that the miRTop project will improve accessibility and uniformity of the results of miRNA pipelines, the possibility to easily run benchmarking analyses, and promote the development of miRNA tools downstream of the detection and quantification steps.