Analysis of miRNA expression profiles in tissue samples by next generation sequencing

Kui Qian, Eeva Auvinen, Dario Greco, and Petri Auvinen
University of Helsinki, Finland
petri.auvinen@helsinki.fi

MicroRNAs (miRNA) are short single-stranded RNA molecules that have been demonstrated to play an important role in regulation of gene expression in many organisms. With the advantages of next generation sequencing, new opportunities have arisen in identifying and quantifying miRNAs, and investigating their functions. We wanted to characterize the expression profiles of miRNAs in human tissue samples by utilizing a deep sequencing approach based on SOLiD 4 System. In this study, we outlined a typical miRNA-seq data processing workflow including reads mapping, counts preprocessing and differentially expressed miRNA testing.

Information about the known human miRNAs were obtained from miRBase database (Version 16). Sequence reads from SOLiD4 platform were first mapped to the 1368 annotated known miRNAs and star-miRNAs. Thereafter, the observed counts of reads for each of the miRNAs were collected for statistical testing to identify differentially expressed miRNAs. We compared the results using different parameters in preprocessing and several widely used statistical packages in Bioconductor.

The miRNA expression levels displayed a very large range as reflected by the number of sequence reads, which varied from single counts for rare miRNAs to several millions of reads for the most abundant miRNAs. We provided an overview of strategies used for miRNA sequencing, and a set of useful methods and available tools for data analysis.