Application of Small RNA-Seq in the Development of Biomarkers for Cognitive Diseases

Introduction
MicrRNAs, regulatory non-coding RNAs of 18-22 nucleotides, are promising biomarkers for different health outcomes because they are tissue specific, stable in extracellular space, can be reliably detected in tissues and diverse body fluids (e.g. blood, cerebrospinal fluid (CSF)) and expression profiles can be linked to health status.\textsuperscript{1-5} In order to investigate potential biomarkers for cognitive diseases, the NEXTFLEX\textsuperscript{®} small RNA-seq kit v3 with UDI was used to sequence microRNAs from human post-mortem tissue, serum and CSF from patients. Additionally, mRNAs extracted from post-mortem human tissue were sequenced using the NEXTFLEX\textsuperscript{®} Rapid Directional RNA-Seq kit 2.0, to predict biological effects of microRNA expression changes.

Methods

MicroRNA Sequencing of Tissue, Serum and CSF Samples
MicrRNAs extracted from post-mortem brain tissue slices, serum or CSF were sequenced using the NEXTFLEX\textsuperscript{®} small RNA-seq kit v3 with UDI on an Illumina\textsuperscript{®} NovaSeq\textsuperscript{®} 50 cycles SP flowcell in single-end mode. The obtained data was pre-processed using cutadapt and miRge2. Statistical analysis was done according to the R-ODAF best practice sequencing pipeline (https://github.com/MCTverheijen/R-ODAF) to filter out low read counts and spurious spikes.
RNA Sequencing of Tissue Samples

Ribo-depleted RNA from post-mortem brain tissue slices were sequenced using the NEXTFLEX® Rapid Directional RNA-Seq kit 2.0 with Unique Dual Index barcodes on an Illumina® NovaSeq® 200 cycle S1 flowcell in paired-end. The R-ODAF best practice sequencing pipeline (https://github.com/MCTverheijen/R-ODAF) was used for pre-processing and statistical analysis (including removal of low read counts and spurious spikes).

Quality Control

Quality control of the sequencing data was performed using fastQC combined with MultiQC to merge the individual fastQC outputs.

Results

Sequence Quality

All reads were sequenced with high quality. All bases of all reads depicted Q-score above 30 and most bases even above 35. The NEXTFLEX® small RNA-seq kit v3 with UDI predominantly captured microRNAs ranging from 23 to 28 nt (Fig 1A) in the analyzed samples and the NEXTFLEX® Rapid Directional RNA-Seq kit 2.0 sequenced RNAs with lengths above 87 nt (Fig 1B).

Sequence Depth and Mapping

Tissue, serum and CSF microRNAs: The sequencing depth of the human microRNAs varied between 1.8 M and 4.7 M. This variation was due to sample type (Fig 2A). For the microRNAs extracted from brain tissue, 60–76.8% of the sequenced reads could be mapped to the genome (STAR+RSEM), while this was only 4.4–30.1% for serum and 0.1–1.6% for CSF. This lower percentage of mapped reads is typical of circulating microRNA sequencing.6

On average, 784 microRNAs were detected in brain samples (sd:72), 360 in serum samples (sd:71) and 103 CSF samples (sd:39). These can be further investigated for their potential application as biomarkers for cognitive diseases.

Tissue RNAs: RNAs sequenced from post-mortem brain slices were sequenced with depths varying from 5.3 M to 10.8 M reads (Fig 2C). On average, 38.7% of these reads could be mapped (STAR+RSEM) to the genome (sd: 4.9%). These sequenced RNAs could be further used to investigate biological processes involved in cognitive diseases. Furthermore, interactions between these RNAs and microRNAs can be investigated to identify processes deregulated by microRNAs in cognitive diseases.

Figure 1. FastQC-MultiQC quality control.
References


Figure 2. Sequenced reads and mapped reads.

Conclusion

We have sequenced RNA and microRNA from postmortem human tissue, as well as circulating microRNAs in serum and CSF using the NEXTFLEX® Rapid Directional RNA-Seq kit 2.0 and the NEXTFLEX® small RNA-seq kit v3 with UDI, on an Illumina® NovaSeq® 6000 sequencer. The sequencing metrics obtained are consistent with the expectations for these samples. Additionally, several differentially expressed entities have been found and will be investigated further. We therefore conclude that these library kits are suitable for the search of biomarkers associated with neurological disorders.

Department of Toxicogenomics (TGX)
Universiteitssingel 40, 6229 ER Maastricht Postbus 616 6200 MD Maastricht
https://www.maastrichtuniversity.nl/research/toxicogenomics