Small RNA-sequencing for Analysis of Circulating miRNAs: Benchmark Study

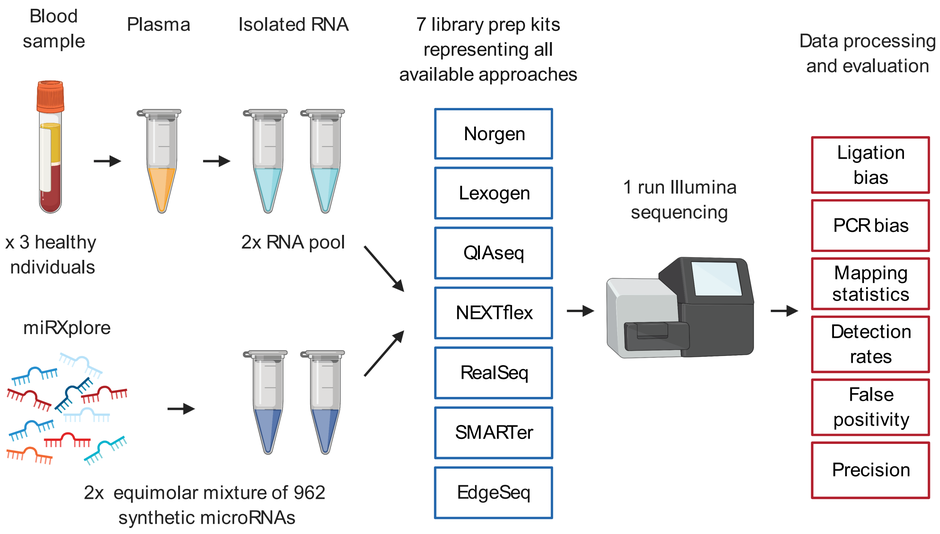
Sarka Benesova[1,2], Peter Androvic[1,3], Eva Rohlova[1,4], Mikael Kubista[1,5], and Lukas Valihrach[1]

[1]Laboratory of Gene Expression, Institute of Biotechnology CAS, BIOCEV, 252 50 Vestec, Czech Republic, [2]Department of Informatics and Chemistry, Faculty of Chemical Technology, University of Chemistry and Technology, 166 28 Prague, Czech Republic, [3]Laboratory of Growth Regulators, Faculty of Science, Palacky University, 783 71, Olomouc, Czech Republic,
[4]Department of Anthropology and Human Genetics, Faculty of Science, Charles University, 128 43 Prague, Czech Republic, [5]TATAA Biocenter AB, 411 03, Gothenburg, Sweden
Contact: sarka.benesova@ibt.cas.cz

Background

The application of **small RNA sequencing** (small RNA-seq) for profiling of **circulating microRNAs** (miRNAs), novel **promising biomarkers**, is increasing. However, the accuracy of the method is **limited by bias** introduced by ligation, PCR or polyadenylation. Therefore recently, **new approaches** were developed to prevent the biases. Here, we present **comparison of all small RNA-Seq library preparation approaches** that are commercially available for quantification of miRNAs in biofluids.

Experimental design



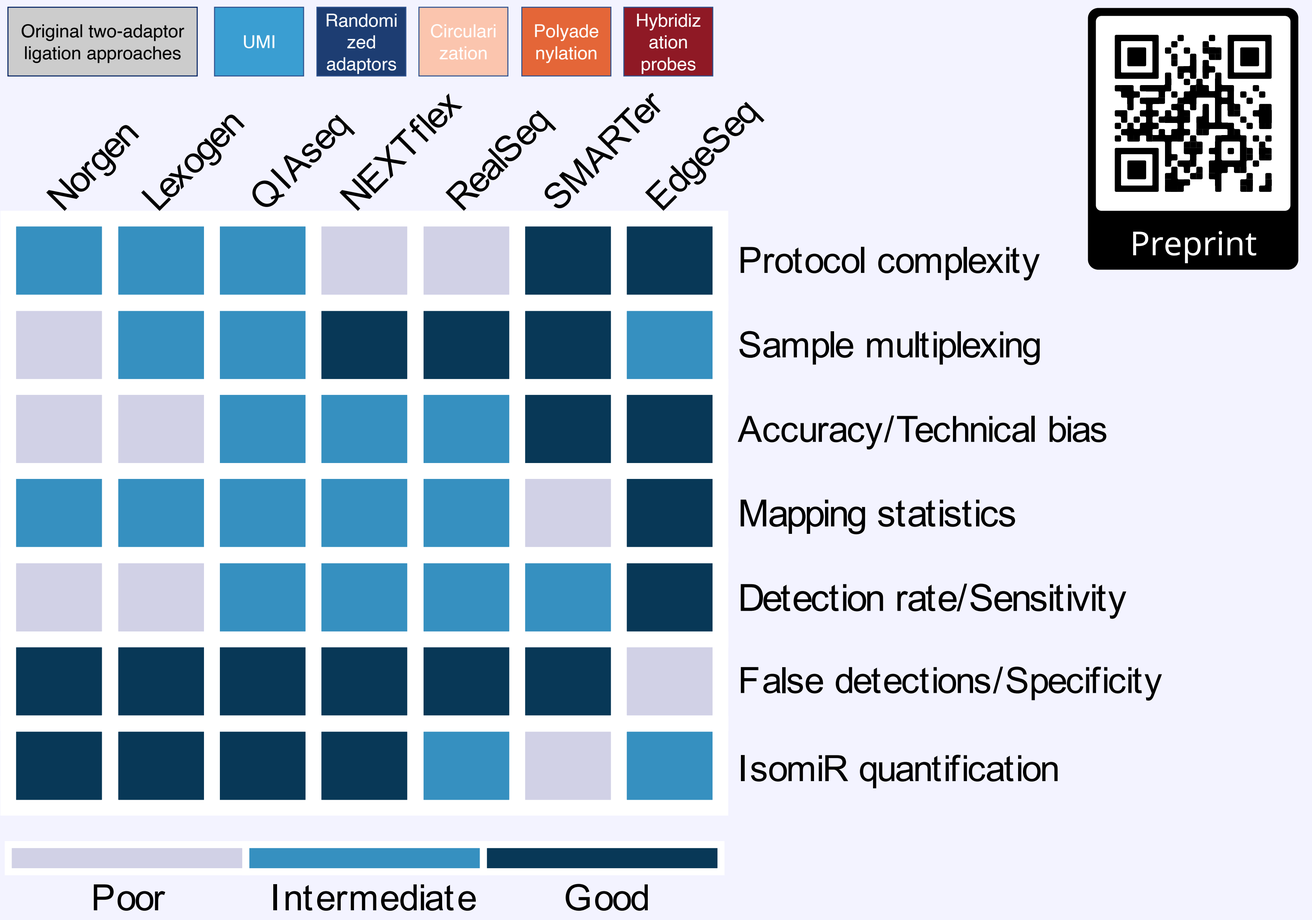
Conclusions

- There is **no single protocol outperforming others** across all metrics.
- **Protocol-specific biases** give rise to limited overlap of measured miRNAs profiles between methods and cause their low correlation.
- The correlation can be improved by **normalization** approach utilizing bias ratios **learnt from miRXplore** sample.
- The application of UMIs has only marginal effect on overall quantification bias.
- Traditional two-steps ligation methods introduce large bias.
- Protocols performing well in majority of metrics use capture probes (EdgeSeq) or randomized adaptors (NEXTflex).

Funding

This study was supported by Czech Science Foundation P303/18/21942S, Czech Health Research Council NU21-08-00286 and institutional support RVO 86652036 and is a part of specific university research project A1-FCHT-2021-003. We thank vendors and local distributors of small RNA-Seq preparation protocols for discounts or free sample kits that were used in this study.

Selection of suitable protocol for small RNA-seq



Preprint

Normalization of data with factors learnt from equimolar pool leads to higher between protocol correlations

Between-protocol reproducibility for plasma samples before and after data correction using bias ratios learnt on miRXplore samples. P-value from two-tailed paired t-test.

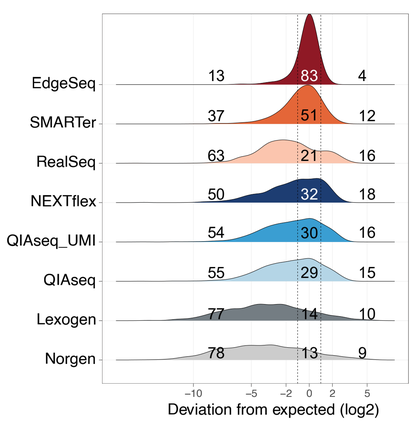
Ligation is the main contributor of the overall bias in the small RNA-seq

Percentage of variance in QIAseq data (miRXplore sample) explained by ligation bias, PCR bias and replicates.

First two bases of miRNA affects ligation of adaptor in RealSeq

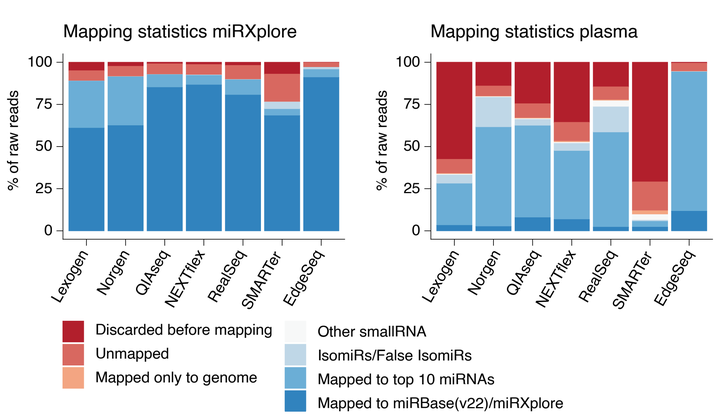
Percentage variance in QIAseq data (miRXplore sample) explained by miRNA sequence characteristics.

Accuracy/Technical bias



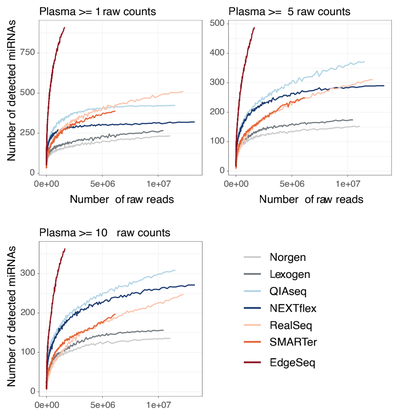
Density plots show distribution of log_2-fold change between measured and expected value in miRXplore. Dashed lines show two-fold deviation from expected value; numbers indicate percentage of miRNAs within and outside of the two-fold range.

Mapping statistics



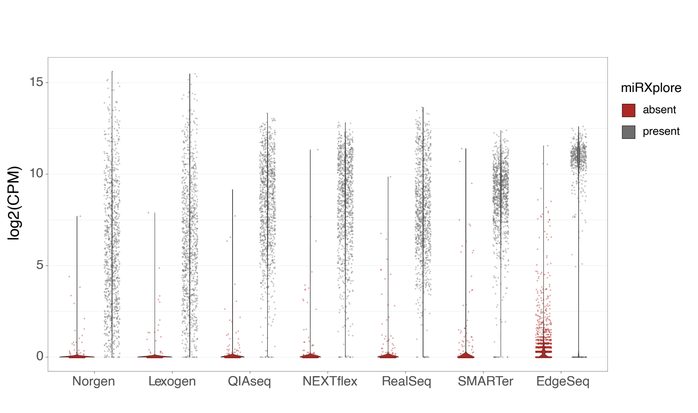
Mapping statistics for miRXplore and plasma samples. Top 10 = 10 most expressed miRNAs; discarded before mapping = sequence length out of 16-29 bp or only adaptors.

Detection rate/Sensitivity



Dependency of number of detected miRNAs in plasma on sequencing depth and various detection thresholds (1, 5, 10 raw reads).

False detections/Specificity



Violin plots showing measured level of true and false positive miRNAs measured in miRXplore samples.