Research &
Pharma Solutions

Small RNA Sequencing

Tech Note

Small RNA Sequencing Service

Small RNA sequencing focuses on short RNA molecules which play an important role in silencing and post-transcriptional gene expression regulation processes. Sequencing of small RNAs allows the analysis of these molecules in more detail and therefore can be an essential part in the identification of novel biomarkers to obtain more valuable research data.

"We have selected the most suitable library preparation protocol for our customers"

For establishing our small RNA sequencing service, three small RNA library preparation protocols (X, Y, Z) were considered and assessed according to different aspects. Here we present the establishment of our favorite sequencing protocol and discuss the reasons why we think this protocol is the best choice.

A very important aspect for small RNA sequencing is the option to prepare libraries from very limited amounts of sample material. For one of the chosen library preparation kits (X) the lowest input amount is 100 ng of total RNA compared to 1 ng total RNA for the other two kits (Y, Z). An additional advantage of protocols Y and Z is the gel-free size selection during library preparation, whereas kit X requires a size selection via gel purification. Due to the lack of a low input option, library preparation kit X was excluded from further investigations.

Another crucial property of a library preparation protocol is reproducibility. Thus, we continued with the two kits (Y, Z) and prepared small RNA libraries with RNA from the same source. Library preparation was carried out according to the vendor's original protocol.

The sequencing data of samples prepared with kit Y and Z were compared. Figure 1 shows that technical replicates generated with protocol Y have a high variance, whereas samples prepared with protocol Z group according to the origin of the sample. This indicates that library preparation kit Z leads to more reproducible results than kit Y.

"The high reproducibility rate is one of the reasons why we chose protocol Z (Bioo Scientific NEXTflex Small RNA-Seq Kit v3) for our small RNA sequencing service."

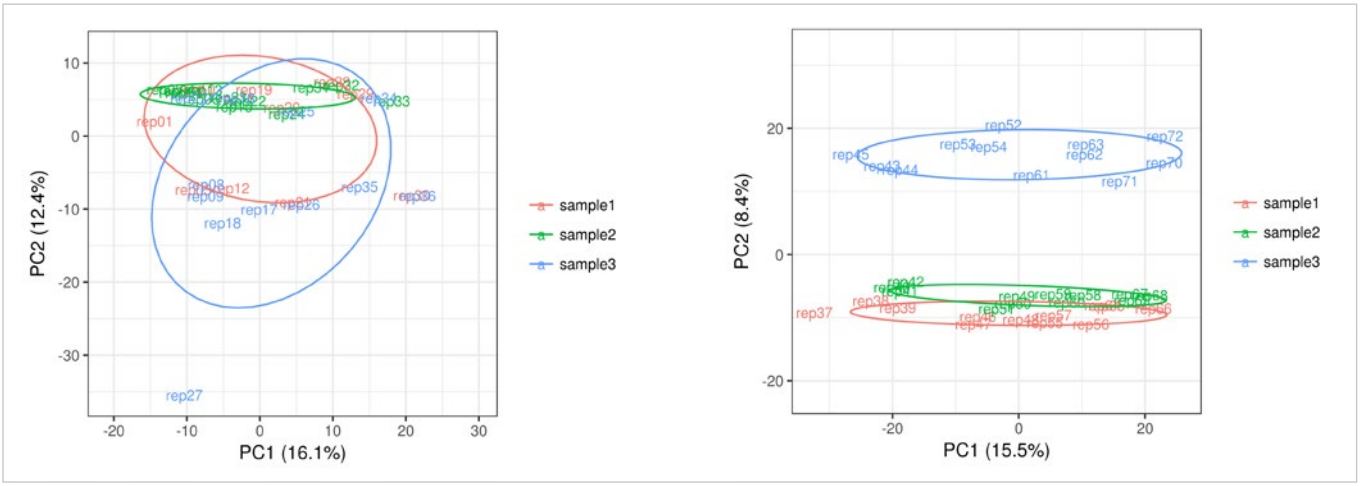


***Figure 1**: Principal component analysis (PCA) plot of samples prepared with library preparation kit Y (left) and Z (right). Libraries of three samples were prepared with different RNA input amounts (6 ng, 60 ng, 200 ng and 1000 ng total RNA) in triplicates, each. Next generation-sequencing (single end and 50 nt read length) was performed by using an Illumina sequencing platform. Replicates are colored according to sample origin. Samples with similar expression patterns cluster together. While Library kit Y shows no current pattern, replicates prepared with library kit Z cluster nicely according to the original sample they were derived from.*

Our low input option for limited sample material

Besides our standard library preparation service with 200 ng input material we also provide a solution for very limited sample amounts. With the Bioo Scientific NEXTflex Small RNA-Seq Kit v3 we can generate libraries and reproducible sequencing results starting from 1 ng of total RNA. Figure 2 (left side) demonstrates that these little amounts of sample material are enough to allow a reproducible detection of miRNA transcripts. However, we also noticed that the RNA input amount influences the results: Experiments performed with 200 ng RNA input material yielded more reads that could be mapped to miRNAs (90.87% on average) than those performed with 1 ng RNA (66.64% on average). They also contained a higher number of different miRNAs (median 837.5) than samples with 1 ng RNA input (median 461) (Figure 2, right side). We suggest using the same amount of RNA for all samples within one experiment to ensure comparability.

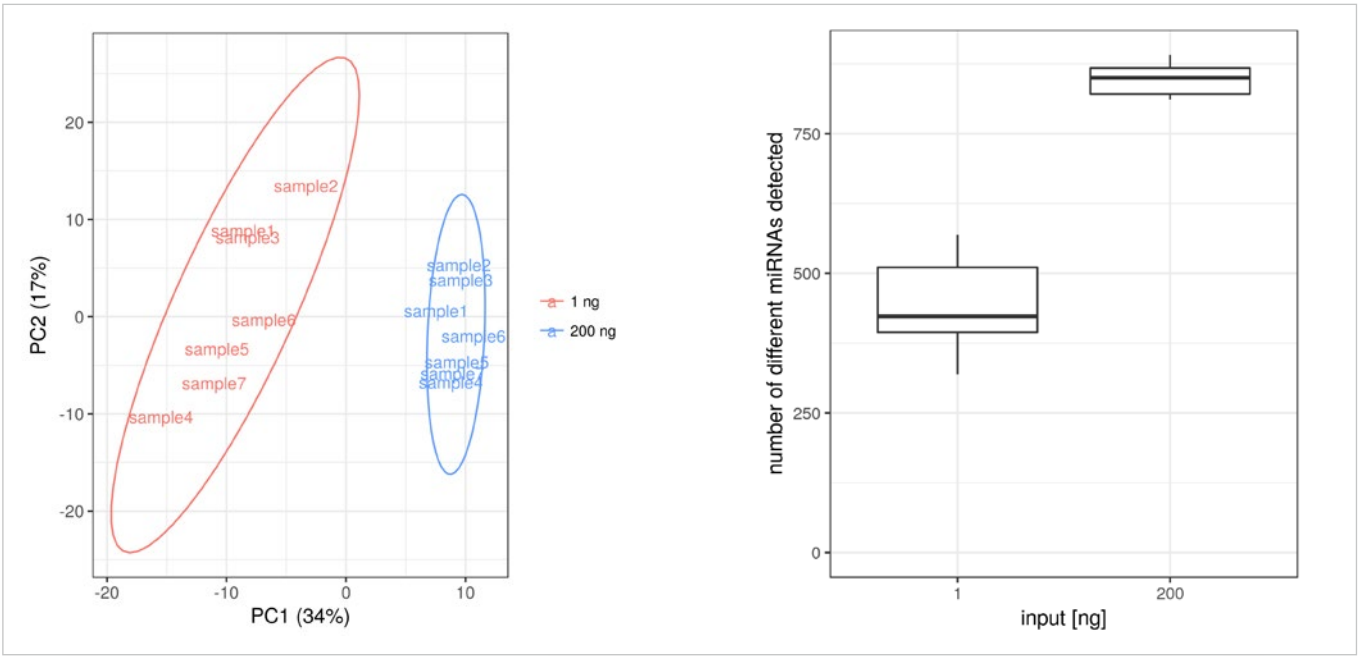


***Figure 2**: PCA plot of sequenced libraries prepared with 200 ng (blue) and 1 ng (red) RNA (left): reproducible results can be generated with as little as 1 ng RNA input. Corresponding boxplot of detected miRNAs with 200 ng and 1 ng RNA input (right).*

Samples are sequenced with the latest Illumina sequencing technology

To offer our customers the best possible technical solution for small RNA sequencing, we are working with the latest Illumina sequencing technologies.

For your individual projects we highly recommend sequencing on SP flow cells (theoretical output 65 gigabases) NovaSeq 6000. Of course, other flow cell options are available as well (S1, S2, S4) for small RNA sequencing on NovaSeq 6000.

In summary, library preparation with NEXTflex Small RNA-Seq kit v3 (Bioo Scientific) leads to reproducible results even with very limited amounts of sample. Depending on the batch size, there are different flow cells available for small RNA sequencing on NovaSeq 6000.

Small RNA NGS Service at a Glance

Your sequencing project can be planned individually:

RNA: RNA isolation of different sample materials, e.g., cells, PAX-tubes, tissue or you can send isolated total RNA

Library preparation: with our standard protocol (200 ng total RNA) or the low input option (please send at least 5 ng total RNA)

Sequencing: with NovaSeq6000 Sequencing System, single end (1x 50 nt) with at least 10 mio reads per sample. Depth and read lenght can be adapted.

Data delivery: download from our secure server or delivery on HDD

Bioinformatics: raw data delivery or bioinformatically processed data (alignment, counting of miRNAs, normalization with miRge, differential gene expression, annotation with miRBase)

Turnaround time: 2-4 weeks

We recommend at least three replicates per group for analysis of differential expression.

Depending on the sample size of a project there are different flow cells available for small RNA sequencing on NovaSeq 6000.

Overview of the sample processing procedure using the NEXTflex Small RNA-Seq Library Preparation kit

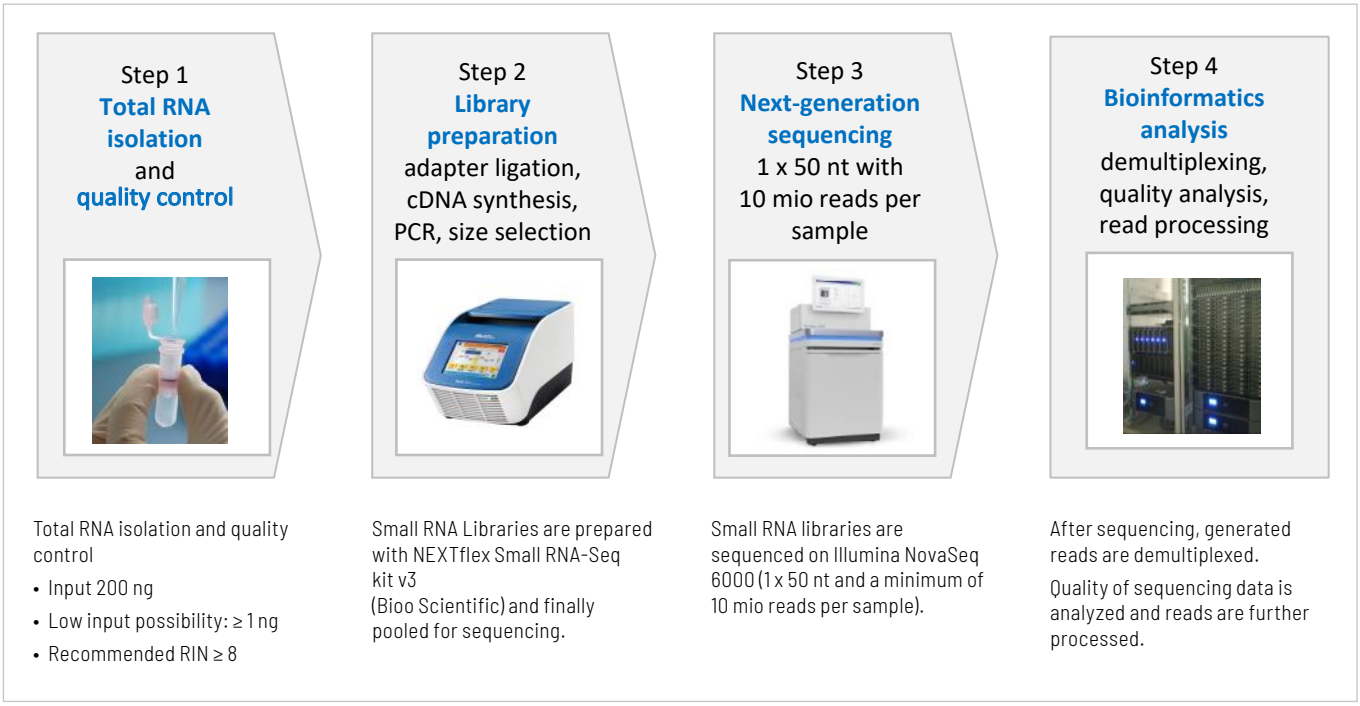


***Figure 3**: Experimental workflow for small RNA sequencing: From sample preparation to final data analysis.*

Examples for ready-to-publish data provided by CeGaT

The data of your sequencing project can be delivered as raw data or can be bioinformatically processed according to your requirements, including a final report with ready-to-publish data:

- Alignment
- Counting of miRNAs
- Normalization (miRge)
- Differential gene expression
- Annotation (miRBase)

For analysis of differential gene expression, reads can be mapped to known miRNAs and pre-miRNAs (hairpins). Via group comparison, significantly expressed miRNAs can be found (Figure 4A).

Furthermore, your data can be visualized, which allows you to identify the main results at a glance. Different methods can be used to visualize the relation between samples or sample groups: hierarchical clustering according to their similarity (Figure 4B) and principle component analysis (PCA) (Figure 5A). A PCA reduces the dimension of the data room to those two components which explain most of the differences between the samples. The expression pattern can also be visualized in detail via heatmap (Figure 5B).

A

Group 1	Group 2	Number of significant miRNAs (padj<0.05)
A	B	335
A	C	318
A	D	257
B	C	295
B	D	268
C	D	316

B

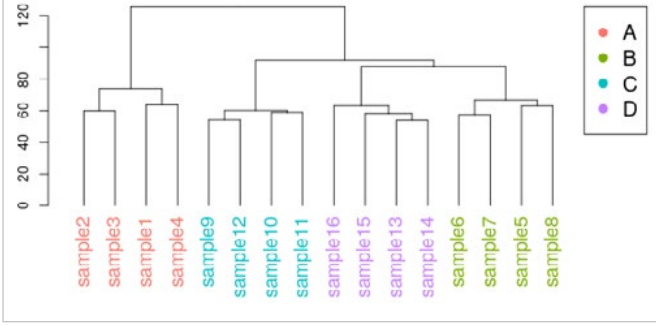
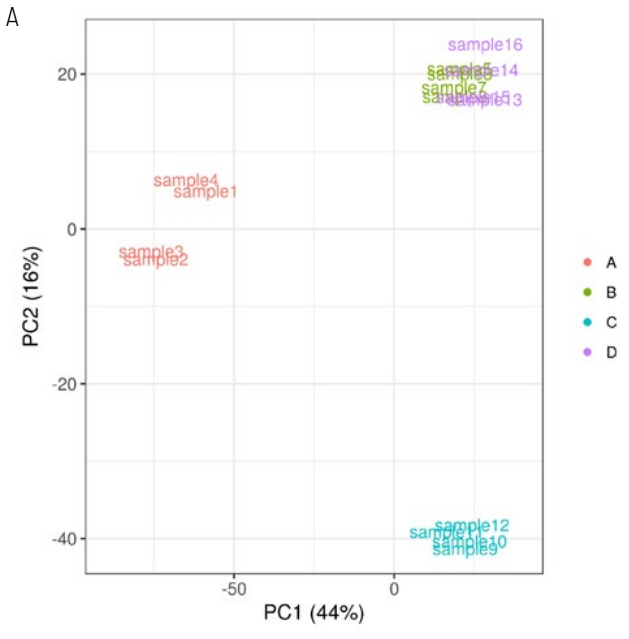


***Figure 4**: Results of differential gene expression analysis.*

A: Statistics of significantly differentially expressed miRNAs in comparison between groups.

B: Hierarchical clustering of samples according to their similarity of expression data based on all genes that received at least one read. The expression data were rlog-transformed and the Euclidean distance was calculated. Samples are colored according to group.



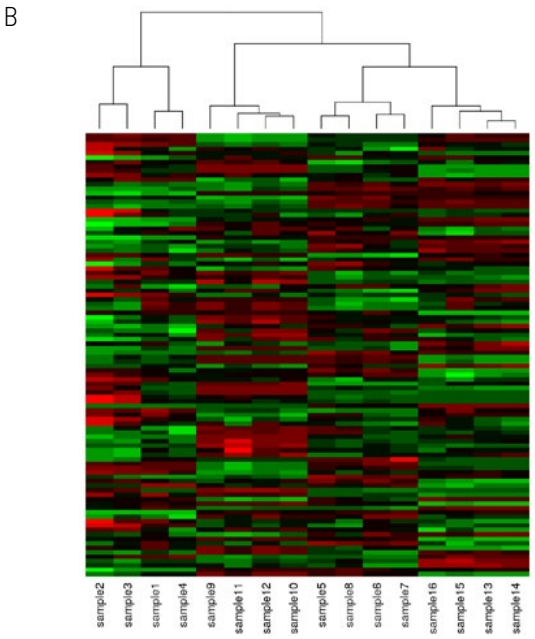


***Figure 5**: Results of differential gene expression analysis. A Principal component analyses of rlog-transformed expression data of all miRNAs that received at least one read. The percentage values on the axes describe how much of the variance between samples is captured in this principal component. Samples are colored according to group. B Heatmap of rlog-transformed expression data of all miRNAs that received at least one read.*

CeGaT

About Us

CeGaT was founded in 2009 in Tübingen, Germany. Our scientists are specialized in next-generation sequencing (NGS) for genetic diagnostics, and we also provide a variety of sequencing services for research purposes and pharma solutions. The portfolio is complemented by non-sequencing-based methods such as immunomonitoring.

Our dedicated project management team of scientists and bioinformaticians works closely with you to develop the best strategy for the realization of your project. Depending on its scope we select the most suitable library preparation and sequencing conditions on our Illumina platforms.

We would be pleased to provide you with our service. Contact us today to start planning your next project.

CeGaT GmbH
Research & Pharma Solutions
Paul-Ehrlich-Str. 23
72076 Tübingen
Germany

Phone: +49 7071 56544-333
Fax: +49 7071 56544-56
Email: rps@cegat.com
Web: www.cegat.com

DAkkS Deutsche Akkreditierungsstelle D-ML-13206-01-00

Accredited by DAkkS according to DIN EN ISO 15189:2014

CAP ACCREDITED COLLEGE of AMERICAN PATHOLOGISTS

CLIA CERTIFIED ID: 99D2130225