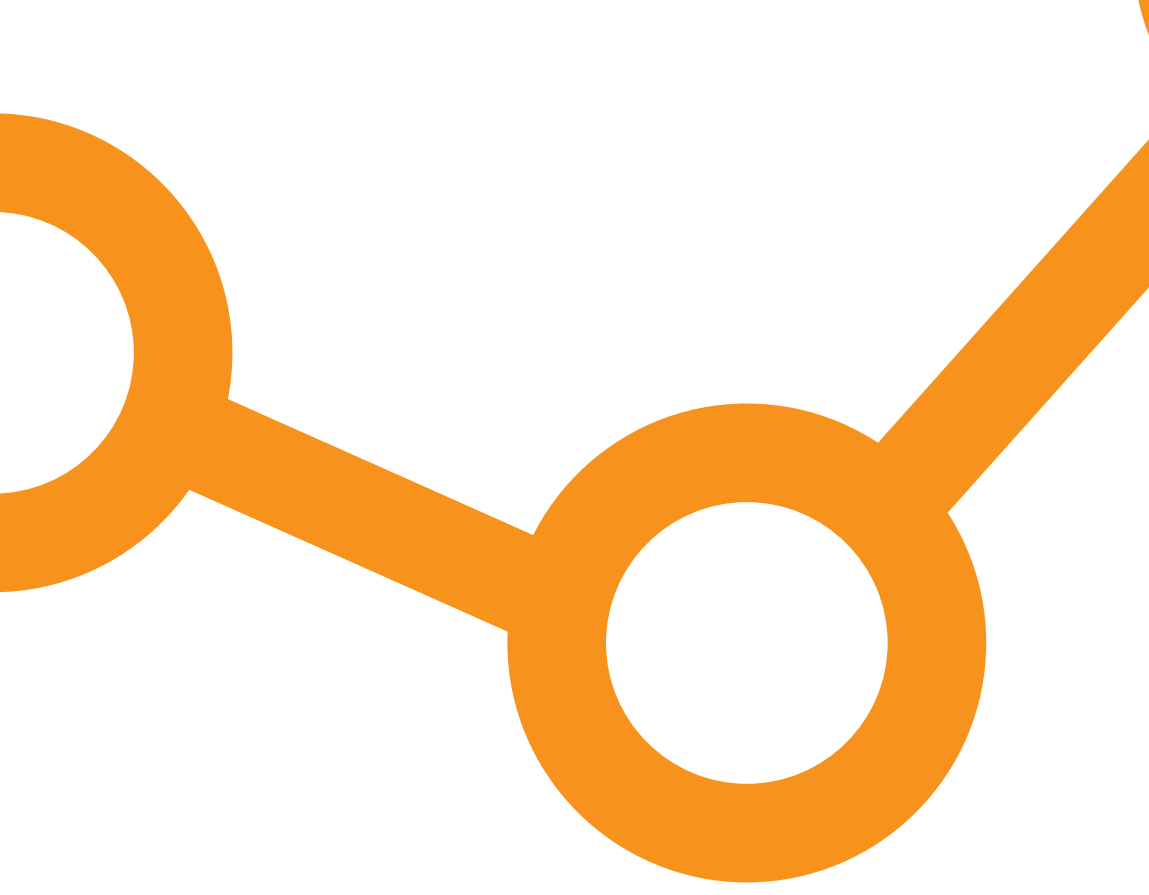


miND

miND® spike-in

Cat no: KT-041-MIND

Instruction Manual v1.0

September 2022

TAmiRNA

stability for life.

Tables

Figures

1 | Product Summary

2 | Important Pre-Analytical Considerations

3 | Lab Protocol

Further information and technical notes can be found at **www.tamirna.com/mind-spike-ins/**

Revision history

Date	Revision
09/2022	1.0

1

Product Summary

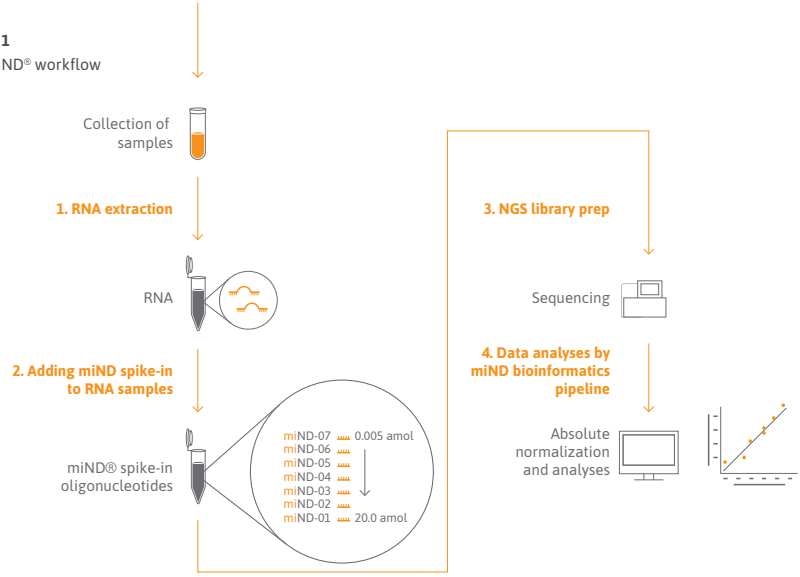
Intended-Use

Intended use

The miND® (**mi**croRNA **N**GS **D**ata Analysis) spike-in has been developed for **quality control** of experiments and **absolute quantitation** of microRNAs in any biological matrix and species (Khamina et al. 2022). The miND® spike-in is added to an RNA sample during the library preparation. The miND® spike-in consists of seven oligonucleotides that are provided in a specific ratio to cover the broad concentration range of endogenous small RNAs. A unique design of the miND® spike-in reduces sequencing bias and ensures precise quantitation of small RNA (Lutzmayer, Enugutti, and Nodine 2017).

The miND® spike-in sequences are detected in the NGS data along with the endogenous small RNAs. Read counts of the miND® spike-in and endogenous miRNAs are used to calculate absolute concentrations (amol/µL or molecules/µL). This conversion can either be achieved by using our miND® NGS data pipeline (Figure 1) (Diendorfer et al. 2022) or by the incorporation of the provided scripts in an already established NGS data analysis workflow.

Figure 1
The miND® workflow



miND® spike-in design features

Each of seven miND® spike-in consists of a unique 13-nucleotide core sequence that is flanked by four randomized nucleotides on the 5' and 3' ends (Table 1) resulting in 65,536 different RNA oligonucleotides per spike-in. The presence of random nucleotides on the 5' and 3' ends of each miND® spike-in sequence is expected to minimize the ligation bias of the core sequence. The miND® spike-in oligonucleotides contain a 5' phosphate group.

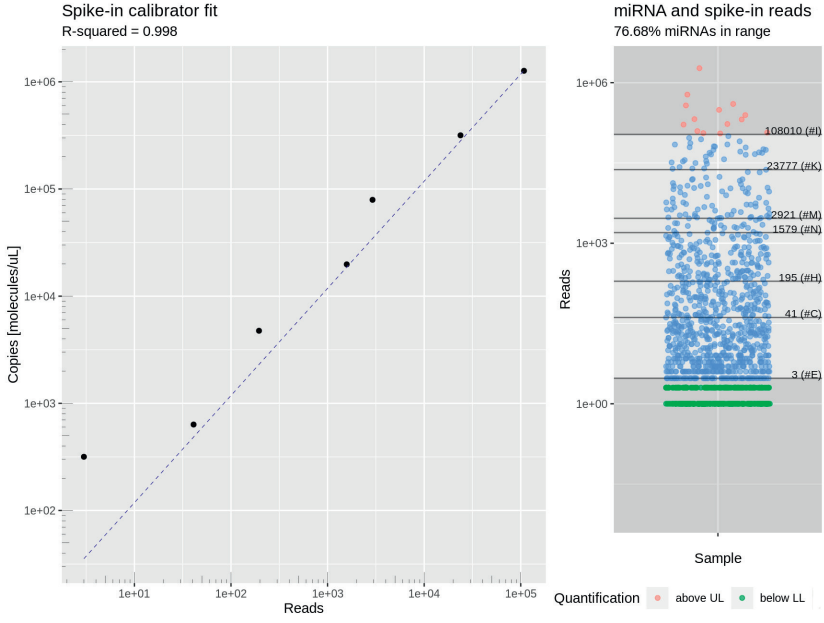
Table 1
The miND® spike-in core sequences

Oligo	Sequence (5' - 3')	Molar amount (amol) in 1 µL of miND spike-in
miND-01	(N)(N)(N)(N)ACGAUCGGCUCUA(N)(N)(N)(N)	20
miND-02	(N)(N)(N)(N)UGAACGUCCGUAC(N)(N)(N)(N)	5
miND-03	(N)(N)(N)(N)UCUCGCGCGCGUU(N)(N)(N)(N)	1.25
miND-04	(N)(N)(N)(N)CGAGUAAUGAACG(N)(N)(N)(N)	0.3125
miND-05	(N)(N)(N)(N)GCUACACACGUCG(N)(N)(N)(N)	0.075
miND-06	(N)(N)(N)(N)UAUUCGCGGUGAC(N)(N)(N)(N)	0.01
miND-07	(N)(N)(N)(N)ACCUCCGUUUACG(N)(N)(N)(N)	0.005

The miND® spike-in simultaneously enables quality control and absolute quantitation of miRNAs across different sample types:

- miND® spike-in serves as a quality control for small RNA-sequencing experiments to confirm the dynamic range and sensitivity of the assay
- miND® spike-in are used to generate a linear regression model to calculate absolute concentrations of endogenous microRNAs (Figure 2)

Figure 2
The miND® spike-in calibrator fit and miRNA spike-in reads distribution



The miND® spike-in was developed based on the principles described in the following publication:

- Lutzmayer et al. Novel small RNA spike-in oligonucleotides enable absolute normalization of small RNA-Seq data. 2017 Sci Rep https://doi.org/10.1038/s41598-017-06174-3

Product Components

The miND® spike-in contains the following components:

One 1.5 mL tube with lyophilized miND® spike-in

The entire workflow consists of three main steps:

1. Dissolve the dry miND® spike-in in nuclease-free water (NFW) (Note: NFW is not provided with the kit.). Aliquot spike-ins to enhance stability after reconstitution.
2. Add the miND® spike-in to an RNA sample before the NGS library preparation
3. Immediately proceed with the NGS library preparation according to the manufacturer's protocol

Storage and Stability

The dry miND® spike-in are shipped at ambient temperature and **must be stored at -20°C upon arrival**. Shelf life of the dry miND® spike-in is at least 18 months. Avoid repeated freezing and thawing as this may lead to degradation. Consider preparing aliquots: each tube of the miND® spike-in is sufficient for 250 NGS library preparations.

2

Important Pre-Analytical Considerations

Choice of Sample Type and Protocols for RNA isolation and small RNA-sequencing library preparation

The miND® spike-in has been extensively tested and optimised for plasma samples processed according the following conditions:

- RNA extraction from 200 µL of plasma with either miRNeasy Mini kit (Cat. 217004, Qiagen) or Maxwell RSC miRNA Tissue kit (Promega, Madison, WI, USA, AS1460) following the manufacturer's protocols.
- NGS workflow according to the RealSeq®-Biofluids Plasma/Serum miRNA Library Kit for Illumina® sequencing (Cat. 600-00012, 600-00024, 600-00048) and the RealSeq®-Dual Biofluids Plasma/ Serum miRNA Library Kit for Illumina® sequencing (Cat. 700-00024, 700-00048)
- Illumina sequencing using the following parameters: minimal read length 50 bp, minimal required sequencing depth is 7.5 million reads per sample.

For other sample types as well as RNA isolation and small RNA-sequencing protocols, it is important to take into consideration the following factors:

- The yield and efficiency of small RNA isolation varies between different isolation and purification kits. Therefore, the choice of the RNA isolation or purification kit can impact the results.
- The selection of the small RNA-sequencing library preparation protocols might impact miND® spike-in performance. Therefore, if other than the recommended library kits are used, users need to perform a pilot experiment with a limited number of samples in order to ensure that the miND® spike-in covers the concentration range of the endogenous small RNAs in the target samples.
- The miND® spike-in core sequences were mapped against the following genomes: *Homo sapiens*, *Mus musculus*, *Bos taurus*, *Rattus norvegicus*, *Sus scrofa*, *Caenorhabditis elegans*, *Drosophila melanogaster* and *Danio rerio* demonstrating minimal overlap with the target genomes. If miND® spike-in is planned to be used with samples from other organisms, it is recommended to reach out TAmiRNA team (e-mail: support@tamirna.com) in order to discuss additional bioinformatic analysis.
- We suggest to use one lot of the miND® spike-in within a project in order to reduce any potential impact of the lot-to-lot variability on the generated results.

Storage and Stability of the dissolved miND® spike-in and RNA

- Store the dry miND® spike-in upon arrival at -20°C.
- Aliquot and store the dissolved miND® spike-in at -80°C. Always keep the dissolved miND® spike-in on ice when handling and return immediately after use to -80°C. Avoid repeated freeze-thawing cycles.
- All samples and miND® spike-in aliquots should be stored in nuclease-free plastic tubes with minimized absorption rates for nucleic-acids ("low-binding").
- RNA samples must be stored at -80°C for long-term storage and kept on ice when handling and return immediately after use to -80°C.

Working with RNA

- Ensure that you work in RNase-free environment and that you use molecular grade NFW only.
- Always work with fresh, disposable plastic consumables and wear gloves.
- Handle carefully to avoid contamination.
- Spin down all reaction and sample tubes before opening.
- Clean all surfaces with commercially available RNase decontamination solutions.
- Use filter barrier pipette tips to avoid aerosol-mediated contamination.
- Designated solutions, tips, pipets and other materials and equipment should be assigned for RNA work only.

3

Lab Protocol

Essential components

miND® spike-in kit

Consumables and Instruments Not Supplied by TAmiRNA

Molecular grade nuclease-free water
Nuclease-free, low nucleic acid binding tubes (1.5 mL)
Nuclease-free, filter pipette tips
Vortexer
Calibrated pipettes
Centrifuge for <2 mL tubes
small RNA NGS library preparation kit

Wet-Lab Protocol

Notes

1. Dissolve the dry miND® spike-in in sterile nuclease-free water

Step 1	Briefly centrifuge the tube containing the dried miND® spike-in to collect the oligo pellet at the bottom of the tube.
Step 2	Add 250 µL of the nuclease-free water.
Step 3	Briefly vortex or mix by pipetting 8-10 times up and down. Leave the tube on ice for 20 minutes.
Step 4	Prepare aliquots that can be used for a single experiment. The volume of the aliquots can be determined based on the experimental plan (it is recommended to add 1 µL of the dissolved miND® spike-in to an RNA sample before the NGS library preparation).
Step 5	Store the aliquots of the dissolved miND® spike-in at -80°C.

2. Add the miND® spike-in to an RNA sample before the small RNA NGS library preparation.

		Notes
Step 1	Thaw an aliquot of the dissolved miND® spike-in on ice.	
Step 2	Mix the miND® spike-in by gently tapping the tube and briefly centrifuge to collect the liquid at the bottom of the tube.	
Step 3	Add 1 µL of the miND® spike-in to each RNA sample right before starting the NGS library preparation experiment.	
Step 4	Immediately proceed with the NGS library preparation according to the manufacturer's protocol	

Data Analysis

The miND® NGS data analysis pipeline processes NGS raw data and performs the conversion of read counts for endogenous small RNAs to the absolute concentrations (amol or molecules/µL). Furthermore, the miND® NGS data analysis pipeline compiles all results in a simple but comprehensive report providing an overview of overall QC data, unsupervised clustering analysis, normalized microRNA count matrices, and differential expression analysis based on raw NGS data (.fasta/.fastq or .bam files). Alternatively, users can incorporate the scripts for annotation and interpretation of miND® spike-in into their own data analysis pipelines.

The absolute quantification of miRNAs using miND® spike-in requires the following steps:

- mapping and quantification of microRNAs and other small RNAs.
- mapping and quantification of miND® spike-in
- calculation of absolute concentrations using a linear regression

We understand that these steps can be performed with many different tools. In order to give a starting point, we provide an exemplary workflow how these steps can be performed using commonly used open-source software. We have published our in-house pipeline miND® (Diendorfer et al. 2022) that can be used for the whole process of miRNA mapping from raw NGS data.

The following steps assume that NGS data is already preprocessed and available in **quality checked and adapter trimmed fastq or fasta files**.

Important: This guide does not aim to provide a start-to-finish pipeline for processing of the miND® spike-in, but give a basic understanding of the steps needed to incorporate them into an existing workflow. If some steps are already included in the user's own bioinformatics pipeline, they can be omitted from this guide and the following steps adjusted to the user's needs.

Preparation of tools and data needed

Conda installation of required tools

We recommend using a package manager like conda (https://docs.conda.io/projects/conda/en/latest/) to install the following tools in an environment, but any other way to obtain and install the software tools can be used. If conda is used, install the needed tools with the following command:

```
# create a new conda environment
conda create -y --name miND

# activate the newly created environment
conda activate miND

# install mamba for quicker dependency resolution upon installation
conda install -y mamba -c conda-forge

# install needed tools
mamba install -c bioconda -c conda-forge -y python=3.9.13 bowtie=1.3.1 \
mirdeep2=2.0.1.2 bbmap=38.90 seqtk=1.3 gawk=5.0.0 wget=1.21.3
```

This will install the following tools:

- bowtie1 (http://bowtie-bio.sourceforge.net/index.shtml)
- miRDeep2 (https://github.com/rajewsky-lab/mirdeep2)
- bbmap (https://jgi.doe.gov/data-and-tools/software-tools/bbtools/)
- awk
- seqtk
- wget

Download and preperation of required data

Precursor and mature sequences of miRNAs are needed to quantify miRNAs using miRDeep2. The following bash code line will download the latest sequences from miRBase.org, extracts miRNAs of one species only, and convert the RNA format to DNA (required by some tools).

```
# download latest mature and hairpin sequences
wget -O - https://www.mirbase.org/ftp/CURRENT/mature.fa.gz | gunzip -c |
seqtk seq -l0 - > mature.fa

wget -O - https://www.mirbase.org/ftp/CURRENT/hairpin.fa.gz | gunzip -c |
seqtk seq -l0 - > hairpin.fa

grep -A 1 --no-group-separator '^>hsa' mature.fa > mature.filtered.fa
grep -A 1 --no-group-separator '^>hsa' hairpin.fa > hairpin.filtered.fa

awk 'BEGIN{RS=">";FS="\n"}NR>1{printf ">%s\n",$1; for (i=2;i<=NF;i++) \
{gsub(/U/,"T",$i); printf "%s\n",$i}}' mature.filtered.fa |
seqtk seq -l0 > mature.filtered.dna.fa

awk 'BEGIN{RS=">";FS="\n"}NR>1{printf ">%s\n",$1;for (i=2;i<=NF;i++) \
{gsub(/U/,"T",$i); printf "%s\n",$i}}' hairpin.filtered.fa |
seqtk seq -l0 > hairpin.filtered.dna.fa
```

Mapping and quantification of miRNAs

Due to the short length and special characteristics of miRNAs, it is recommended to use dedicated tools for the mapping and quantification. A popular choice is miRDeep2 (https://github.com/rajewsky-lab/mirdeep2), which can quantify the miRNAs in a sample using the following steps:
Data needed:

- miRNA precursor sequences of the species of interest (provided by https://mirbase.org)
- miRNAs mature sequences of the species of interest (provided by https://mirbase.org)
- FASTA/Q files with NGS reads (named sample_1.fq in this example)
- fasta files of miND® spike-in core sequences (provided below)
- miRNA mapping and quantification
- First step is to prepare the input fastq file to following mapping steps. The mapper.pl script from miRDeep2 will perform the following tasks:
- convert to fasta format (-e option)
- remove entries with non-canonical letters (-j option)
- discard reads shorter than 17 nucleotides (-l option)
- collapse identical reads and store the read count in the fasta header (-m option)

In addtion, the number of threads used for bowtie1 mapping can be changed with the -o option (set to 2 in this example).

```
# create a folder for each sample
mkdir sample_1 && cd sample_1

# run the mapper
mapper.pl 'sample_1.fq' -e -h -j -l 17 -o 2 -m -s 'sample_1.collapsed.fa'
```

The output file “sample_1.collapsed.fa” will be used in the following step to find and quantify miRNA reads.

Important: Additional mapping and filtering steps can be required, depending on the task on hand (e.g. prior mapping of reads to the host genome to filter unmapped reads). If this is the case, we recommend these mappings now with the collapsed fasta file and proceed with the so filtered fasta file in the following step.

For the next step, miRDeep2 will be used to map the reads against the miRNA reference data set (precursor and mature sequence are needed - see “Download and preperation of required data” above on how to generate these files).

```
# this should be run IN the sample folder (e.g. sample_1)
#
# map and quantify miRNAs using miRDeep2's quantifier.pl script
# adapt the -T option to run more than two threads
quantifier.pl -p '../hairpin.filtered.dna.fa' -m '../mature.filtered.dna.fa' -y \
'default' -T 2 -d -r "sample_1.collapsed.fa"
mv miRNAs_expressed_all_samples_default.csv sample_1.mirnas.csv
```

miRDeep2 generates output in the current folder (thus the creation of a sample_1 folder and running it from inside this folder) with all the relevant mappings and quantifications. The last step will move the generically named output file in the sample folder with a sample specific name.

Mapping and quantification of miND® spike-in

The miND® spike-in requires a different approach for quantification, as it consists of a fixed core sequence and variable 3` and 5` ends. Only the core sequence will be used to quantify each spike-in. bbmap from the bbtools is well suited for this task.

miND® spike-in core sequences

The miND® spike-in core sequences are provide in fasta format to the bbmap tool. Content of the mind-spike-ins.fa:

```
#SpikeIn.miND.01
ACGATCGGCTCTA
#SpikeIn.miND.02
TGAACGTCCGTAC
#SpikeIn.miND.03
TCTCGCGCGCGTT
#SpikeIn.miND.04
CGAGTAATGAACG
#SpikeIn.miND.05
GCTACACACGTCG
#SpikeIn.miND.06
TATTCGCGGTGAC
#SpikeIn.miND.07
ACCTCCGTTTACG
```

miND® spike-ins mapping and quantification

bbduk is used to find sequences that contain the provided miND® spike-in core sequences.

```
# run this in the sample directory
bbduk.sh threads=2 -Xmx12g in='sample_1.collapsed.fa' outm=stdout.fq \
ref='../ mind-spike-ins.fa' stats='sample_1.spikeins.txt' statscolumns=5 \
k=13 maskmiddle=f rcomp=f hdist=0 edist=0 rename=t
```

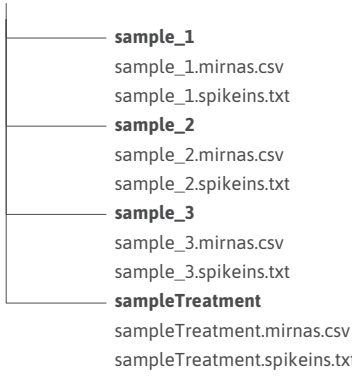
This will generate a file names “sample_1.spikeins.txt” containing counts for each of the spike-ins.

Calculation of absolute concentrations

The last step of the absolute quantification is the merging of the miRNA and the miND® spike-in mappings, followed by modelling of the concentrations and generation of figures and tables for the assessment of the model fit.
This will be done using a R script that reads in the two mapping files (miRNAs and miND® spike-in) for multiple samples and then generates a combined CSV file containing absolute quantifications based on the miND® spike-in in addition to QC information.
For each sample there must be two files: sample_1.mirnas.csv and sample_1.spikeins.txt; The script will look for pairs in the folder defined in the input_data_path variable (see top of file). Each pair is identified by the identical sample name (sample_1 in this case). If a different RNA to miND® spike-in ratio was used, the calculation of concentrations can be adapted in the setup section of the R script.

The R script contains a few comments and instructions on how to run it. It can process multiple samples at once and expects them in a directory structure like this (for 4 examplory samples and their names):

input_data_path



R script download

The R script is hosted in the most current version in our public github repository: https://github.com/tamirna

miND® spike-in quality control

The provided R script will do multiple checks to ensure sufficient detection of the miND® spike-in. Results of these checks are saved in a separate file (spikein_stats.csv) for all samples together. It will give details about the amount of detected miND® spike-ins, the number of miRNAs in range of the miND® spike-in (miRNAs with read counts higher than the lowest spike-in read count and lower than the highest spike-in), the parameters of the linear model and fit of the model (r-squared). The last column gives details about the QC results.

For a sample to **pass the QC checks**, the following must be true:

- 5 or more miND® spike-in core sequences detected
- linear model parameters calculated
- Pearson correlation coefficient (r-squared) above 0.95

Potential Problems

Observation: No miND® spike-in sequences were detected in the generated NGS data
Potential cause:

- A tube with the miND® spike-in was not centrifuged before adding NFW and the oligo pellet was lost
- The miND® spike-in was reconstituted in a wrong volume of nuclease-free water
- Improper storage of the reconstituted miND® spike-in led to degradation
- Contamination of the miND® spike-in with RNases led to degradation
- The miND® data analysis pipeline was not used

Observation: The generated NGS data did not pass the quality control of the miND® spike-in
Potential cause:

- Less than 5 miND® spike-in core sequences detected
- Insufficient sequencing depth
- Degradation of miND® spike-ins due to the use of contaminated reagents
- Low ligation efficiency of spike-ins with the selected small RNA NGS library preparation kit

Related Services

TAmiRNA offers a broad range of high-quality RNA services performed by experts according to GLP standards, including EV/exosome characterization, and RNA isolation from any biological matrix, small RNA and mRNA NGS, qPCR, and customized bioinformatic analyses.

The miND® spike-in is compatible with a broad range of sample types and species:

- **Species compatibility:** our bioinformatic pipeline has been tested with human, mouse, rat, pig, cow, and horse samples. Any species with known microRNAs can be analyzed.
- **Sample types:** besides cells and tissues we have tested conditioned media, plasma (various anti-coagulants), serum, urine, CSF, brain microdialysate, and synovial fluid. This includes enrichment of EV/exosomes from all biofluids.
- **Laser microdissection:** the miND® service can be used to analyze dissected tissue compartments for increased precision. Learn more here: https://www.tamirna.com/space-resolved-rna-profiling-in-complex-tissues/
- **Other RNAs:** PNK-treatment of total exRNA increases coverage of mRNA and long non-coding RNAs in your data.
- **RNA-seq:** we offer mRNA-seq (polyA and total RNA) alongside our small RNA-seq workflow to generate high quality microRNA/mRNA datasets.

Read more about these services at www.tamirna.com/small-rna-sequencing-services/

Related Products

TAmiRNA offers research-use kits for analysis of microRNA biomarkers.
Read more about the products at www.tamirna.com

Further Reading

Below you find a list of publications describing the development and application of the miND® spike-in:

1. Khamina, K. et al. A MicroRNA Next-Generation-Sequencing Discovery Assay (miND) for Genome-Scale Analysis and Absolute Quantitation of Circulating MicroRNA Biomarkers. Int. J. Mol. Sci. 2022,23,1226. https://doi.org/ 10.3390/ijms23031226
2. Diendorfer, A. et al. miND (miRNA NGS Discovery pipeline): a small RNA-seq analysis pipeline and report generator for microRNA biomarker discovery studies. F1000Research 2022,11:233,1226. https://doi.org/10.12688/f1000research.94159.1
3. Gutmann, C. et al. Association of cardiometabolic microRNAs with COVID-19 severity and mortality. Cardiovasc Res. 2022, 118(2):461-474. https://doi.org/10.1093/cvr/cvab338
4. Lutzmayer, Stefan, Balaji Enugutti, and Michael D. Nodine. 2017. 'Novel Small RNA Spike-in Oligonucleotides Enable Absolute Normalization of Small RNA-Seq Data'. Scientific Reports 7 (1): 5913. https://doi.org/10.1038/s41598-017-06174-3.

Notes

Notes

Notes

Notes

Notice to purchaser

A patent was filed for the invention related to novel spike-in oligonucleotides for absolute quantitation of nucleotide sequence data under WO2018138334A1.

The purchase of this product conveys to the buyer the non-transferable right to use the purchased product and components of the product only in research conducted by the buyer (whether the buyer is an academic or for-profit entity).

Imprint

Copyright: 2022 / TAmiRNA GmbH
Text: TAmiRNA GmbH
Concept & Design: www.fuergestaltung.at
Illustrations: www.birgitbenda.at

TAmiRNA GmbH
Leberstrasse 20
1110 Vienna, Austria
+43 1 391 33 22 90
www.tamirna.com