Revision history

Date	Revision	Details about change
13.04.2026	1.0	Initial creation of the document

miND® spike-in

Frequently Asked Questions

Small RNA Sequencing Controls for Quality Control & Absolute Quantitation

Product overview

What are spike-in controls for small RNA sequencing?

Spike-in controls for small RNA sequencing are synthetic RNA oligonucleotides of known sequence and concentration that are added to total RNA samples prior to library preparation. They provide a stable reference that is carried through every step of the sequencing workflow. As a result, spike-ins help monitor technical variability introduced during sample processing. Differences in spike-in recovery between samples can thereby reveal inconsistencies in library preparation, sequencing depth, or other technical biases.

When spike-ins are designed as ratiometric sequences they provide an additional advantage: they allow calibration of sequencing reads against known input amounts. This enables absolute or near-absolute normalization of miRNAs.

Why do I need absolute normalization of small RNA sequencing data?

Standard small RNA sequencing data normalization methods, such as reads per million genome mapping reads (RPM), assume that the overall small RNA content is roughly constant between samples. In practice, this assumption is not always true, especially in:

- Comparisons between sample types (e.g. plasma vs. cerebrospinal fluid), where RNA composition and total RNA abundance differ substantially
- Method development and benchmarking, where spike-ins can be used to compare RNA extraction or library preparation protocols
- Disease vs. healthy comparisons, where global shifts in miRNA or other small RNA expression may occur
- Cross-batch or cross-laboratory comparisons, where differences in extraction or library preparation efficiency introduce technical variation

Relative normalization under these conditions can mask true biological signals. External spike-in controls bypass this problem because they are added at a known amount independently of the sample.

What is the difference between relative and absolute normalization in small RNA-seq?

- **Relative normalization** (e.g. RPM, TPM): Scales read counts within a sample relative to the total number of reads. This approach is fast and simple to compensate for varying sequencing depths but assumes that the overall small RNA composition is constant between samples. If one sample has a large overall increase or decrease in specific RNA classes, this scaling factor becomes distorted.

- **Absolute normalization with spike-ins:** Uses external spike-ins of known concentration to generate a calibration curve that converts sequencing read counts into absolute abundance values, such as molecules/µl. Because normalization is anchored to an external standard, absolute normalization does not rely on assumptions about total small RNA content. This enables more accurate comparisons across biologically distinct sample types, disease states, sequencing batches and laboratories.

What makes a good spike-in control for small RNA or miRNA sequencing?

An ideal spike-in for small RNA sequencing should:

- **Match the size of endogenous targets:** miRNAs are 20-22-nt long. Spike-ins must be in the size range of the target to be processed identically to endogenous miRNAs. General-purpose RNA spike-ins (e.g. ERCC, >250-nt) are not suitable.
- **Cover a wide dynamic range:** Multiple spike-ins at different concentrations spanning 3-4 orders of magnitude are needed to capture the full abundance of endogenous miRNAs.
- **Have no homology to endogenous sequences:** The spike-in sequences must be artificial to avoid cross-mapping to known miRNA or small RNAs.
- **Be resistant to ligation bias:** Adapter ligation efficiency in small RNA-seq is highly sequence-dependent and is a known source of technical bias. Randomized flanking nucleotides at the 5' and 3' ends incorporated in the spike-in sequences reduce sequence specific adapter ligation artifacts that would distort calibration.
- **Exhibit high integrity and purity:** HPLC purification removes truncated synthesis by-products that would inflate read counts.

The miND® spike-ins were designed with all of these criteria in mind. See the sections below for details.

What is the miND® spike-in control?

The miND® spike-in controls are a panel of seven synthetic 21-nucleotide long RNA oligonucleotides designed to serve as external standards for small RNA sequencing experiments. They enable two key functions (1) **quality control** of library preparation workflows and (2) **absolute quantification** of miRNA NGS data by converting sequencing read counts into absolute copy numbers (molecules/µl).

What is the design principle behind the miND® spike-ins?

Each spike-in consists of a unique 13-nucleotide core sequence flanked by 4 randomized nucleotides at each end. This unique design makes the spike-ins themselves less prone to biased detection across the calibration range. The seven spike-ins are supplied at defined molar ratios spanning approximately four orders of magnitude (0.005 to 20 amol/µl), covering the full dynamic range of endogenous circulating miRNAs.

Compatibility & Application

Are the miND® spike-ins compatible with different library preparation methods?

Yes. We have tested the compatibility and performance of the miND® spike-ins with several library preparation kits, including NEXTFLEX™ Small RNA Sequencing Kit v4 (Revvity).

Other small RNA library preparation protocols should also be compatible; however, performance validation data may not be available.

Which sample types can I use with miND® spike-ins?

The spike-in concentration range has been selected to match the typical endogenous miRNA abundance in biofluid samples and have been extensively tested on:

- Biofluids: serum, plasma, cerebrospinal fluid, synovial fluid, urine, and saliva
- Extracellular vesicles and non-vesicular fractions extracted from biofluids and cell culture supernatant

For recommendations on how to use the spike-ins with cellular and tissue input, see sections below.

What is the recommend spike-in input volume?

The spike-ins have been developed and optimized for use with biofluids and other low RNA input samples. The standard use is therefore for total RNA inputs ranging from 0.5 – 50 ng at which the addition of 1 µl spike-in is recommended. For use with inputs >50 ng total RNA the recommended spike-in volume scales with the total RNA input to ensure that the spike-in signals remain proportionate to the endogenous miRNA signal.

Total RNA input	Recommended spike-in volume	Typical sample types
Standard: 0.5 – 50 ng	**1 µl**	Biofluids (serum, plasma, urine, synovial fluid), extracellular vesicles
50 – 100 ng	**1 – 2 µl**	Cells, small tissue amounts
≥ 100 ng	**≥ 2 µl**	Tissue, high-input cell experiments

Important: If you change the spike-in volume from the standard recommended 1 µl, you must update the corresponding volume in the absolute concentration calculations. See the next question for details.

I am using a non-standard spike-in volume. Do I need to adjust anything in the concentration calculations?

Yes, this is critical if you intend to work with the spike-in normalized data. The absolute concentration calculation is based on the known input amount of each spike-in. When the volume added differs from the standard 1 µl, the actual amount of spike-in introduced into the reaction changes, and the calibration curve will be shifted accordingly.

Update the spike-in input volume parameter in your analysis to match the volume used. Failing to do so, will cause a proportional error in all calculated concentrations: e.g. using 2 µl without updating the calculation will cause concentrations to appear 2x lower than the true values.

If you are using the miND pipeline (TAmiRNA), the spike-in volume is a configurable input parameter in the markdown file.

Should I dilute the spike-ins for low input samples?

Generally, no. The spike-ins are optimized for biofluid samples, which are inherently low-input matrices. The standard volume specified in the product's IFU is designed for this range and should not require dilution for typical biofluid inputs and EVs.

At what stage of the workflow should the spike-ins be added?

The spike-ins are optimized for addition directly to the extracted RNA sample immediately before small RNA library preparation.

Can the spike-ins be added prior to RNA isolation? How should the spike-in volume be adjusted?

The spike-ins are optimized for addition directly to the extracted RNA sample prior to library preparation. While pre-isolation addition is possible, it requires additional consideration and is not the default recommendation.

As a rule of thumb, use the following approach to determine the appropriate spike-in volume:

1. Identify the spike-in volume ($V_{\text{spike-in (lib prep)}}$) and RNA input volume ($V_{\text{RNA input}}$) recommended in your library preparation protocol and sample type.
2. Calculate the spike-in ratio: Divide the recommended spike-in volume by the RNA input volume.
3. Multiply this ratio by your RNA elution volume (V_{elution}).
4. Apply a correction factor based on the estimated RNA extraction efficiency (E) of your workflow to obtain the pre-isolation spike-in volume ($V_{\text{spike-in (pre-isolation)}}$).

$$V_{spike-in\ (pre-isolation)} = \frac{V_{spike-in\ (lib\ prep)}}{V_{RNA\ input}} \times V_{elution} \times \frac{1}{E}$$

Example: The protocol recommends 1 µl spike-in added to 8.5 µl RNA input as starting material for library preparation. If RNA is eluted in 50 µl and extraction efficiency is estimated at 70%, the recommended pre-isolation volume would be (1/8.5) x 50 x (1/0.7) = 8.4 µl.

Important considerations:

- **Add the spike-ins to the lysis buffer** or lysis master mix prior to sample addition. Do not add the spike-ins directly into the biological sample (e.g. plasma).

  Note: Spike-ins added directly to a biological sample are exposed to the sample matrix before protective lysis conditions are established. This leads to degradation by RNases present in the sample.

- If extraction efficiency is unknown, an estimate of 50-70% (E = 0.5-0.7) is a reasonable starting point.
- Optimization and pilot testing are strongly recommended before applying pre-isolation spike-in addition to a full experiment, as recovery can vary between extraction kits, sample matrices and library preparation methods.
- Absolute quantification is not recommended when spike-ins are added prior to RNA isolation as the fraction of spike-ins recovered is not precisely known.

Quality

How is the quality of miND spike-in batches ensured?

Each new miND® spike-in LOT undergoes a two-stage quality control process before release:

1. **Concentration verification by RT-qPCR**

   The concentration of the highest-abundance spike-in is first verified by RT-qPCR. This step confirms that the lot was produced at the correct concentrations.

2. **NGS validation**

   Lots that pass the RT-qPCR step are validated by sequencing. Plasma RNA extracted with two different RNA extraction kits, each prepared in three technical replicates, is used to run the full workflow. During this NGS-QC step, the following parameters are confirmed:

   - Presence and detection of all 7 spike-ins
   - Correct abundance
   - Expected vs. observed concentrations via linear regression
   - Coverage of endogenous plasma miRNAs
   - Presence of the expected full-length 21-nt sequences (no truncations)

Should I expect LOT-to-LOT variability between miND® spike-in LOTs?

As the spike-ins are present at very low molar concentrations, slight variations between synthesis LOTs are inherent and expected. TAmiRNA produces LOTs in bulk to support comparability over multiple years.

To ensure comparability for larger-scale or long-term studies, we recommend purchasing sufficient stock from a single LOT to cover the entire study. The lyophilized miND® spike-in kit (KT-041-MIND-96) provides long-term stability, making it well suited for extended studies where consistent performance over time is critical.

How are the miND® spike-ins purified? Could truncated species interfere with calibration?

The miND® spike-ins are HPLC purified to minimize the presence of truncated synthesis by-products. In addition, we performed a comprehensive assessment of spike-in sequence lengths across >5,000 samples. This analysis consistently demonstrates that the vast majority of aligned spike-in reads correspond to the expected full-length 21-nt sequences.

Data Analysis

The following questions cover the bioinformatics workflow for processing miND® spike-in data. They complement the wet-lab guidance above and address the most frequent questions we receive about the open-source scripts and Docker tools published at https://github.com/tamirna.

Which tools does TAmiRNA provide for analyzing miND® spike-in data, and do I have to use the full miND® pipeline?

You do not have to run the full miND® pipeline. We provide three layers of tooling on our public GitHub (https://github.com/tamirna), and you can enter the workflow at whichever level fits your existing setup:

- **miND (full pipeline)**: Snakemake pipeline that runs the entire workflow from raw FASTQ to a final HTML report with QC, miRNA counts, spike-in calibration, and differential expression. Recommended when you do not already have a small RNA-seq pipeline in place. The publication describing the pipeline is Diendorfer et al., F1000Research 2022.
- **mind-spikein-docker (quantification only)**: a dockerized Snakemake workflow that takes trimmed, quality-filtered FASTQ files as input and produces miRNA counts (via miRDeep2) and spike-in counts (via bbduk). Use this if you already have your own adapter trimming and QC and just want consistent miRNA and spike-in quantification.
- **mind-spike-in-concentrations (calibration only)**: an R script (miND-spikein.R) that takes the miRNA and spike-in count files produced by the docker tool and fits the calibration model, calculates absolute concentrations in molecules/µL, and writes per-sample QC plots and a spikein_stats.csv summary. Use this if you already have count tables from your own aligner and want only the calibration step.

If you are integrating spike-in quantification into an existing pipeline, the simplest path is to add the spike-in FASTA as an additional reference in your alignment step and then run miND-spikein.R on the resulting counts.

I have my own pipeline. What input file format does miND-spikein.R expect?

The script reads two files per sample from input_data_path/<sample_id>/:

- <sample_id>.mirnas.csv: tab-separated table of miRNA read counts. The first column header must be #miRNA and the read-count column must be named read_count. This matches the output of miRDeep2 quantifier.pl. If a miRNA appears multiple times (e.g. mature and star), the script keeps the maximum count per ID.
- <sample_id>.spikeins.txt: standard bbduk stats output produced by mapping reads against the spike-in FASTA. The script reads the library size from line 2 and the per-spike-in #Name/Reads table starting at line 4. Spike-in IDs must be of the form #miND-NN, where NN runs from 01 to 07 (the seven core sequences).

A SampleContrastSheet.xlsx is not the input to this script. The contrast sheet is the entry point for the full miND® pipeline; it is not consumed by the spike-in calibration script directly. Only the per-sample CSV and TXT count files are needed.

If you produce counts with a different aligner (e.g. STAR, bowtie, salmon), build the two files in the format above before running the script, or reuse the calibration logic in your own code (see the next question for the formula).

How exactly are read counts converted to molecules/µL? Can you show the formula?

The script fits an ordinary least-squares linear regression of the known molar concentration of each spike-in on its observed read count, forced through the origin:

$$\text{concentration}_i = \text{slope} \times \text{read\_count}_i \quad \text{(intercept fixed at 0)}$$

The slope and the model R-squared are derived from the seven calibrator points using R's lm(concentration ~ 0 + rc, data = sample_spikeins). The known molar concentration of each spike-in in the final sample is calculated from its stock concentration as:

concentration_in_sample [molecules/µL] = stock_concentration [amol/µL] × spikein_volume [µL] × 6.02214e5 / final_volume [µL]

Defaults in the published script: spikein_volume = 1 µL and final_volume = 9.5 µL (1 µL spike-in plus 8.5 µL RNA input, matching the RealSeq-Biofluids working volume). Stock concentrations of the seven spike-ins miND-01 to miND-07 are 20, 5, 1.25, 0.3125, 0.075, 0.01 and 0.005 amol/µL respectively.

If you use a different spike-in volume or a different library prep with a different RNA input volume, update both parameters in the script. Otherwise all calculated concentrations will be off by the corresponding factor (e.g. using 2 µL spike-in without changing the parameter will make endogenous miRNAs look 2x less abundant than they actually are).

What do the QC fields in spikein_stats.csv mean, and when is a sample safe to use for absolute quantification?

For every sample the script writes one row to spikein_stats.csv and a <sample>_spikein_qc.pdf plot. The fields are:

- spikeins_detected: number of the seven core sequences with at least one read.
- intercept, slope, rsq: parameters of the calibration line. Intercept is fixed at zero by design.
- spikein_lower_limit, spikein_upper_limit: lowest and highest spike-in read counts in this sample. miRNAs with counts inside this range are quantified by interpolation; outside the range they are extrapolated and the range column in the per-sample concentrations file is marked too low or too high.
- mirnas_in_range: number of detected miRNAs that fall inside the calibrator range. The QC PDF reports the percentage of miRNAs in range.
- qc: free-text combination of: OK; warning (not all spikeins detected) when 5 or 6 of 7 are detected and calibration still runs; warning (less than 50% of miRNAs in spikeins range) when the dynamic range of the library is not well covered; FAILED (R squared less than 0.95) when the calibration line does not fit the calibrators and absolute concentrations should not be used; FAILED (2 or more spike ins missing) when the model cannot be fitted at all.

Use absolute concentrations only for samples flagged OK or warning. Samples flagged FAILED should be excluded from spike-in normalized analyses; you may still use their relative miRNA counts.

I want to run differential expression analysis on top of the spike-in calibration. Where do the spike-in counts belong?

Spike-in reads must be excluded from the count matrix that goes into the DEA tool (edgeR, DESeq2, limma-voom). They are calibrators, not biological features, and leaving them in will inflate library sizes and distort dispersion estimates. Drop all rows where the feature ID starts with #miND- before constructing the DGEList or DESeqDataSet.

There are two common ways to use the spike-in information in a DEA:

- Spike-ins as size factors / offsets: compute a per-sample scaling factor from the calibration slope (samples where one read corresponds to fewer molecules/µL have higher library efficiency and need to be scaled down accordingly) and pass it to DGEList$samples$norm.factors (edgeR) or sizeFactors() (DESeq2) in place of the default TMM or median-of-ratios factors. This is the most direct way to remove global composition bias between groups.
- Filter to miRNAs in calibrator range, then DEA on absolute concentrations: use the range == "in range" column from each sample's _concentrations.csv to keep only miRNAs with reliable absolute values, then run your statistical test on the calibrated molecules/µL values rather than on raw counts.

The full miND® pipeline performs the spike-in filtering automatically in its built-in DEA step.

Troubleshooting

I am seeing insufficient spike-in reads/not all spike-ins are detected. What should I check?

If you do not detect all seven spike-in sequences work through the following checklist:

- **Sequencing depth:** A minimum of 7.5 million reads per sample is recommended for reliable detection of all seven spike-ins. Very shallow sequencing will cause the lowest-abundance spike-ins to fall below the detection threshold.
- **Reconstitution:** Confirm that the correct spike-in reconstitution volume has been used as outlined in the latest version of the manual and that the spike-ins are mixed thoroughly before use.
- **Spike-in addition step:** Confirm that the spike-ins were added to the RNA right before library preparation.
- **Spike-in volume:** Verify that an appropriate spike-in volume has been used. For high input/complexity samples, the spike-in volume may be increased. See the recommendations above.
- **Library QC:** Check if the library QC (TapeStation/Fragment Analyzer) profile looks like expected.
- **miRNA mappings:** Check if the miRNA mappings look as expected. A low recovery of low abundance miRNAs may indicate insufficient sequencing depth.
- **Freeze-thaw cycles:** Repeated freeze-thaw cycles can degrade the spike-in oligos. Aliquot the working stock to minimize freeze-thaw cycles and store resuspended aliquots at -80°C.
- **Clean workspace:** To minimize the potential of degradation of the spike-ins, make sure that you clean your workspace with an appropriate RNase decontamination solution before starting.

If the issue persists, contact ngs@revivity.com with your sequencing QC data and sample processing information for further assistance.

I get zero spike-in counts in my bioinformatics analysis even though spike-ins were added. What should I check?

When spike-ins were confirmed to have been added and the library preparation QC (e.g. TapeStation/Fragment Analyzer profiles) looks as expected, the typical causes are:

- **Bioinformatics**: Ensure that the spike-in sequences are included in the reference set used for read alignment. The miND pipeline (TAmiRNA) includes spike-in sequences by default, for other pipelines, add the spike-in FASTA sequences for mapping.
- **RNA input amounts >100 ng**: The spike-ins have been validated for total RNA input amounts up to 100 ng. Using higher RNA inputs may reduce or completely obscure spike-in representation relative to endogenous small RNAs, potentially resulting in very low or undetectable spike-in counts.

What if I detect fewer than all seven spike-in sequences? Can I still use them for normalization?

This depends on how many spike-ins are detected and which ones are missing:

- **≥ 5 spike-ins detected:** Normalization is supported, the calibration curve remains robust.
- **≤ 4 spike-ins detected:** Normalization is not recommended. This typically indicates a technical issue. Investigate the root cause before interpreting results.

Note: If only the lowest-concentration spike-ins (e.g. miND-06, miND-07) are missing, this may simply reflect the assay's detection limit rather than a failed experiment. In that situation, review the total sequencing depth and overall library complexity.

Spike-in reads make up a disproportionately large fraction of my total mapped reads. What does this indicate?

If the spike-in reads represent an unusually high proportion of your total mapped reads, this may be a signal that the endogenous RNA input was lower than expected or that an excessive amount of spike-ins was added to the sample.

- **Review the RNA isolation protocol**: Losses during RNA extraction (e.g. inefficient elution, wrong protocol for the sample type) can dramatically reduce endogenous RNA recovery.
- **Spike-in reconstitution volume**: Verify that the spike-ins were reconstituted using the correct volume specified in the latest version of the IFU.

The spike-in detection is highly variable between my replicates. What could cause this?

High between sample variability in spike-in countstypically points to pipetting imprecision as the primary cause.

To improve reproducibility, check the following:

- **Use a calibrated pipette:** Verify that the pipette used for spike-in addition is properly calibrated and performs accurately in the 1 µl range. For very small volumes (≤1 µl), reverse pipetting can improve precision and consistency.
- **Mix thoroughly before use:** Ensure that the spike-in working stock is properly resuspended, completely thawed, and mixed thoroughly before addition to samples.
- **Prepare a master mix when possible:** Some library preparation workflows allow spike-ins to be added directly to an adapter ligation or reaction master mix. This reduces replicate-to-replicate variability introduced by repeated individual pipetting steps.

How do I contact for technical support?

Contact us via email (ngs@revvity.com) or through the online form on our website Contact Us - Technical Support | Revvity.