APPLICATION NOTE

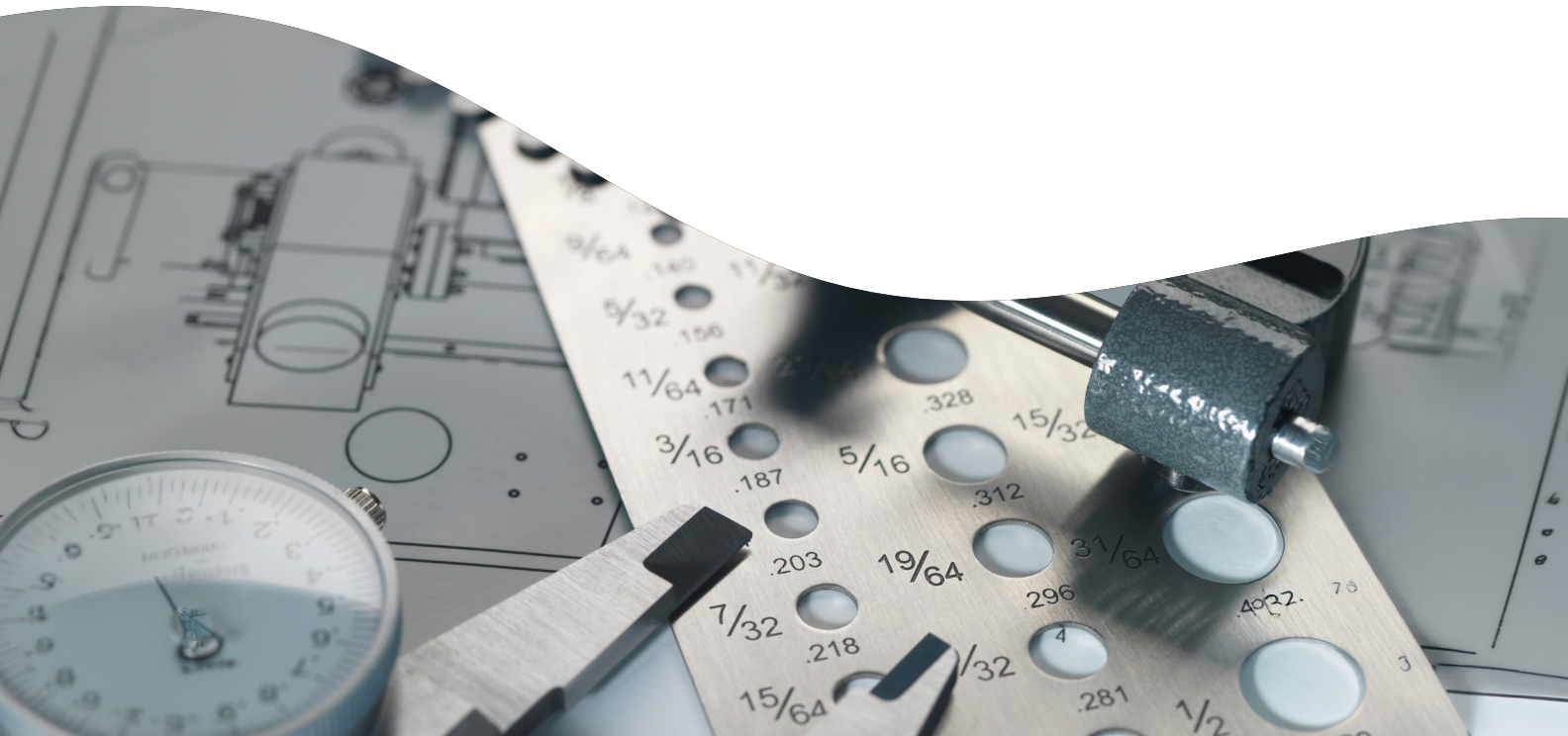
Evaluating miND Spike-In Controls for quantitative normalization of plasma miRNA sequencing.

Introduction

Small RNA sequencing is widely used to profile microRNAs (miRNAs) in biofluids such as plasma. Circulating miRNAs are attractive biomarkers because they are stable in extracellular environments and reflect physiological and pathological processes. As a result, plasma miRNA profiling is increasingly applied in oncology, cardiovascular disease, and biomarker discovery.

Despite its broad adoption, quantitative interpretation of miRNA sequencing data remains challenging, as read counts reflect not only true miRNA abundance but also technical variability introduced during the workflow. RNA extraction efficiency, adapter ligation bias, reverse transcription efficiency, PCR amplification, and sequencing depth can all influence measured miRNA levels.

These challenges are particularly pronounced in biofluid samples, where total RNA input is extremely low and varies substantially between samples. Plasma RNA preparations often contain picogram- to low-nanogram RNA quantities. In addition, hemolysis and sample handling can alter the apparent abundance of specific miRNAs. Kirschner et al. demonstrated that even mild, visually undetectable hemolysis can significantly elevate miR-16 and miR-451 in plasma, reinforcing the need for robust normalization controls[1].



Normalization is therefore essential to distinguish biological differences from technical variability. Most studies rely on relative approaches such as reads per million (RPM) or library size scaling. While these correct for sequencing depth, the comparable RNA composition between samples, which often does not hold in biofluid studies. For example, a small number of highly expressed miRNAs can dominate plasma libraries, leading to underestimation of lower-abundance transcripts when normalized using total read counts[2].

An additional complication is the lack of universally stable endogenous reference RNAs in plasma or serum. Unlike cellular transcriptomics experiments, biofluids lack consistently expressed reference transcripts across individuals and disease states. Several studies have shown that commonly used endogenous miRNA controls exhibit variability across disease states, samples and processing protocols, limiting their use for normalization [3,4].

Synthetic spike-in controls provide an alternative strategy for absolute or semi-absolute normalization. When added at known concentrations after extraction but prior to library preparation, they act as internal calibration standards to monitor technical performance. Because their input is known, spike-ins enable estimation of the relationship between sequencing reads and molecular abundance, supporting conversion of read counts into approximate RNA copy numbers and improving comparability across samples.

The miND® Spike-In Controls consist of a defined mixture of seven synthetic small RNAs spanning multiple concentration levels[5], providing a calibration framework for small RNA sequencing. In this application note, we evaluate the performance of miND spike-ins in libraries generated from human plasma RNA using the NEXTFLEX™ Small RNA-Seq Kit v4. The objective of this work is to assess spike-in detection, quantitative linearity, and compatibility with a small RNA sequencing workflow.

Methods

Human plasma RNA was isolated from 200 µL of plasma using the miRNeasy® Serum/Plasma Kit (QIAGEN). The miND® Spike-In Controls were added to the sample prior to library preparation according to the manufacturer's recommendations. Small RNA libraries were then prepared using the NEXTFLEX Small RNA-Seq Kit v4. Three technical replicate libraries were generated.

Library quality control was performed on the Fragment Analyzer 5200 system (Agilent). Final libraries were pooled equimolarly, quantified using a Qubit® fluorometer (Thermo Fisher Scientific), and loaded at 650 pM with 10% PhiX spike-in on an Illumina® NextSeq2000™ platform with 72 single-read cycles aiming to obtain approximately 7 million reads per sample. A target depth of ~7 million reads per library was selected as a practical balance between sensitivity for biologically relevant miRNAs, and maintaining sufficient read capacity for reliable detection of the low abundance spike-ins added to the sample. Additional sequencing may be required when maximizing detection of low-abundance endogenous miRNAs is the primary objective[6].

Small RNA sequencing data analysis was performed using the miND® analysis pipeline[7].

Results

Sequencing depth and spike-in representation

Sequencing yielded 7.1–7.4 million total reads per library, demonstrating consistent sequencing depth across replicates. Spike-in reads accounted for ~0.19 % of total reads, indicating that the incorporation of spike-ins occupies only a small fraction of capacity (Figure 1A and B).

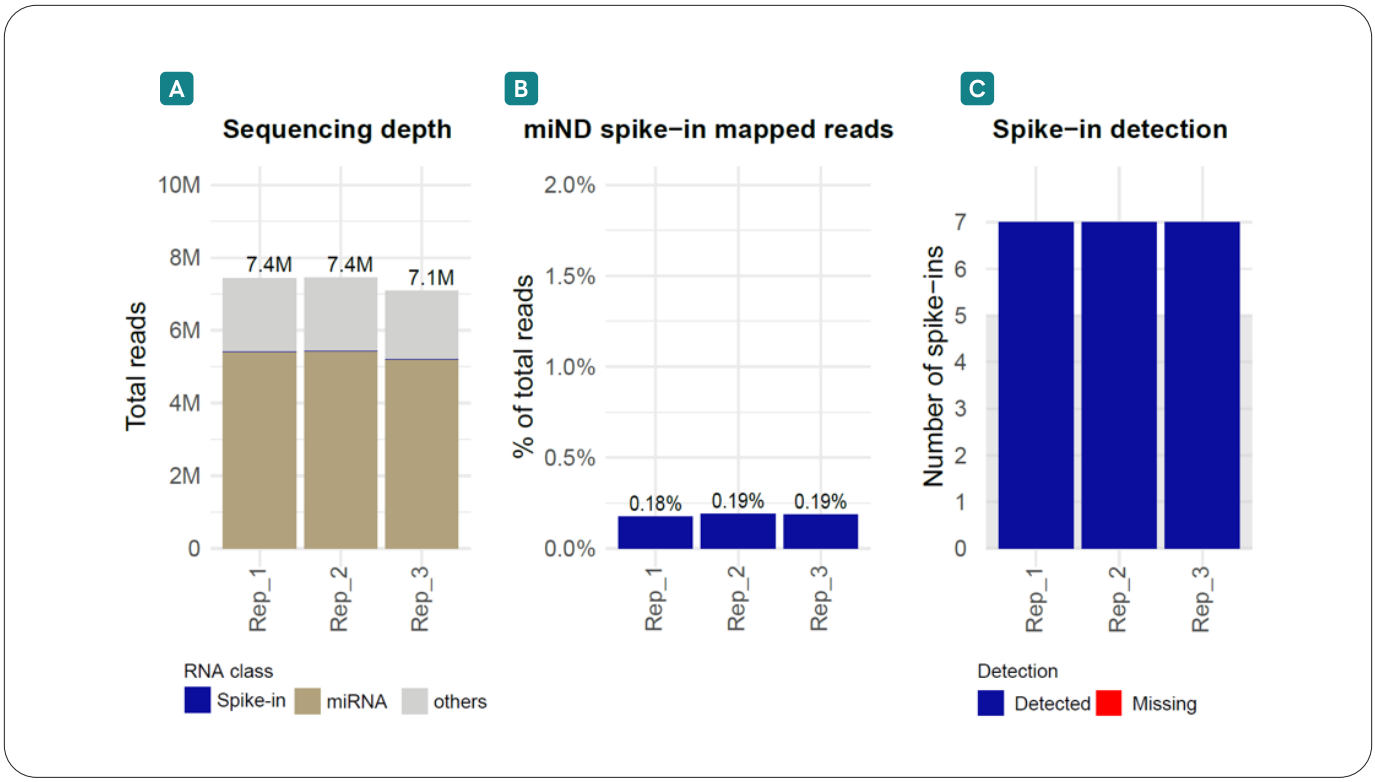


Figure 1: **Robust detection of miND® spike-in controls in NEXTFLEX** small RNA libraries. **A**. Sequencing depth across three technical replicates, showing consistent distribution of reads mapping to miRNAs, spike-ins, and other RNA species. **B**. miND® spike-ins represent ~0.18–0.19 % of total reads, indicating reliable detection with minimal impact on sequencing capacity. **C**. All spike-in molecules were detected across replicates, demonstrating robust incorporation and reproducible performance of the spike-in controls.

Spike-in detection

All seven spike-in molecules were consistently detected across replicates, demonstrating robust incorporation of the spike-in controls and reproducible detection across libraries (Figure 1C).

Quantitative calibration

The spike-in panel spans four orders of magnitude in abundance, ranging from approximately ~1 copy/µL to ~10,000 copies/µL. This broad dynamic range enables construction of a calibration model covering low- and high-abundance miRNAs. Calibration was performed by comparing the known spike-in concentrations with the observed sequencing read counts, enabling estimation of endogenous miRNA abundance per µL of RNA input.

A strong log-log linear relationship was observed across the concentration range, with a coefficient of determination (R^2 = 0.988), indicating that sequencing counts closely track the expected spike-in input levels (Figure 2A).

Distribution of endogenous miRNA signals

Using the spike-in calibration range, 80.6 % of detected miRNAs fell within the quantitative window defined by the spike-in controls, indicating that the panel effectively brackets the expression levels of most miRNAs detected in the dataset. ~20 % of miRNAs were detected outside this range (Figure 2B).

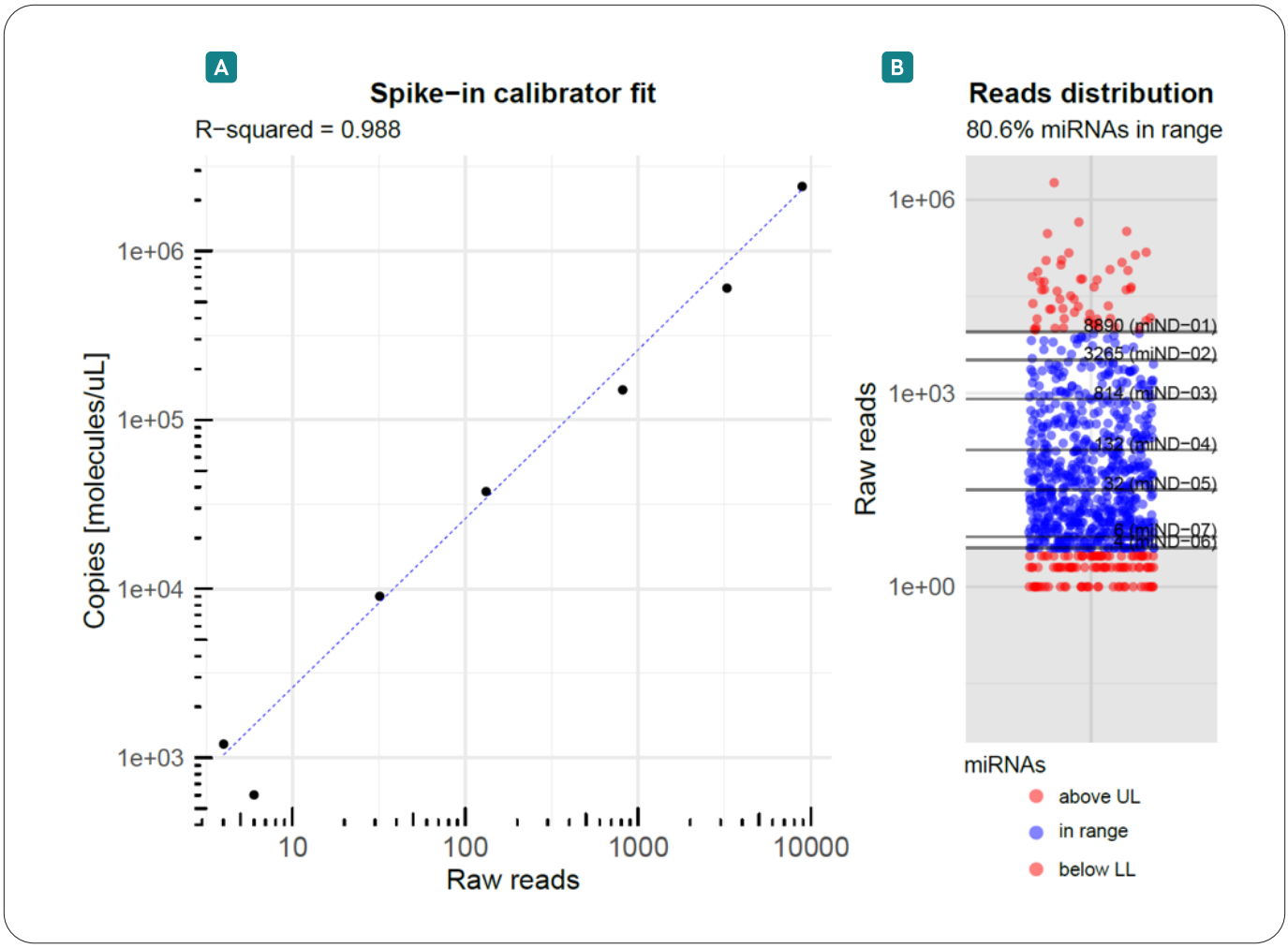


Figure 2: **Quantitative calibration using miND® spike-in controls.** A. Relationship between log-transformed sequencing reads and spike-in concentration, showing a strong linear fit (R^2 = 0.988) across the 7 spike-in molecules. B) Distribution of endogenous miRNA read counts relative to the spike-in-defined quantitative range, with 80.6 % of detected miRNAs falling within the calibration window.

Conclusion

These results demonstrate that miND® Spike-In Controls provide a reliable quantitative framework for plasma small RNA sequencing when used with the NEXTFLEX Small RNA-Seq Kit v4. Spike-ins added prior to library preparation were consistently detected across replicates while representing only a small fraction of sequencing reads.

The strong relationship between spike-in concentration and sequencing reads establishes a calibration range spanning multiple orders of magnitude that encompassed approximately 80.6 % of the endogenous miRNAs detected in the libraries.

Together, these results support the use of miND spike-ins with the NEXTFLEX workflow to enable quantitative normalization of plasma miRNA sequencing data.

References

1. Kirschner, M.B. et al. (2011). Haemolysis during sample preparation alters microRNA content of plasma. *PLoS ONE*. 6:e24145. doi: 10.1371/journal.pone.0024145.

2. Mestdagh, P. et al. (2014). Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nature Methods*. 11:809–815. doi: 10.1038/nmeth.3014.

3. Moldovan, L. et al. (2014). Methodological challenges in utilizing miRNAs as circulating biomarkers. *Journal of Cellular and Molecular Medicine*. 18:371–390. https://doi.org/10.1111/jcmm.12236.

4. Schwarzenbach, H. et al. (2015). Data normalization strategies for microRNA quantification. *Clinical Chemistry*. 61:1333–1342. doi:10.1373/clinchem.2015.239459.

5. Khamina, K., et al. (2022). A MicroRNA Next-Generation-Sequencing Discovery Assay (miND) for Genome-Scale Analysis and Absolute Quantitation of Circulating MicroRNA Biomarkers. *International Journal of Molecular Sciences*. 23(3):1226. doi:10.3390/ijms23031226.

6. Hong, L., et al. (2021). Systematic evaluation of multiple qPCR platforms, NanoString and miRNA-Seq for microRNA biomarker discovery in human biofluids. *Scientific Reports*. 11:4435. doi:10.1038/s41598-021-83365-z.

7. Diendorfer, A., et al. (2022). miND (miRNA NGS Discovery Pipeline): A Small RNA-Seq Analysis Pipeline and Report Generator for microRNA Biomarker Discovery Studies. *F1000Research*. 11(233). doi:10.12688/f1000research.94159.1.