

Be part of the solution: total RNA-seq from human biofluids and extracellular vesicles

DATE: November 22, 2019

AUTHOR: Bhavna Chaudhuri, PhD

CATEGORIES: [Research News](#) | [Cancer](#) | [NGS](#)



The unrealized potential of cell-free RNA

RNA profiling can be a powerful tool to investigate the biomarker potential of human biofluids. While circulating DNA has emerged as the most popular focus of liquid biopsy analyses, extracellular RNA present in plasma, urine, and extracellular vesicles (EVs) can also provide valuable information about disease states. However, RNA sequencing in biofluids is technically challenging due to low input amounts, large dynamic range, and variable degradation of RNA, all of which hamper straightforward quantification. Furthermore, current methods for sequencing-based total RNA profiling of cell-free samples fail to provide a comprehensive analysis because of limitations such as short fragment length, low number of quantified genes, and ribosomal RNA contamination. In a [recent study](#), researchers from the Center for Medical Genetics at the Ghent University and Biogazelle, Zwijnaarde (Belgium) assessed the performance of a strand-specific total RNA library preparation method (SMARTer Stranded Total

RNA-Seq Kit - Pico Input Mammalian; Pico v2) for RNA-seq analysis of different biofluids and derived extracellular vesicles. In an extensive assessment of the Pico v2 kit they demonstrate that the kit provides an accurate, precise, and sensitive method to quantify total RNA in human biofluids. Some of their key findings are discussed below.

Putting Pico v2 to the test

Read distribution and sample type

RNA-seq of platelet-rich plasma (ePRP) and platelet-free plasma (ePFP) from two donors collected in EDTA tubes revealed that more than 75% of the unique reads in ePRPs were derived from mitochondrial RNA (mtRNA) in contrast to ePFP, which had three times less mtRNA and most reads mapped to nuclear DNA. The authors found that the number of unmapped and multi-mapped reads was similar between plasma with and without platelets but that the distribution between exonic, intronic, and intergenic reads differed between platelet-rich and platelet-free plasma (Figure 1).

The authors also sequenced conditioned medium from breast cancer cells (CM), platelet-free plasma from a third healthy donor collected in citrate tubes (cPFP), and urine from a prostate cancer patient, in addition to EVs from these three fluids. The mapping rates were found to differ substantially among biofluids and their EVs, with the fraction of unique reads ranging from 7.69% in cPFP EVs to 90.2% in EVs isolated from conditioned medium (Figure 1, Panel A). The unique reads from almost all samples mainly mapped to nuclear DNA (Figure 1, Panel B). Platelet-free plasma contained 25.8% mtRNA, which was comparable to the percentages in the healthy donors. Most reads that mapped to nuclear DNA were found to be exonic except for those from cPFP EVs, which contained a larger fraction of intronic and intergenic reads (Figure 1, Panel C).

The kit generated strand-specific sequencing reads in all the biofluids with cDNA fragment sizes ranging from 70–400 nt, with a peak around 90 nt for the plasma samples and around 180–190 nt for the other samples (the plasma samples and derived EVs presented with the shortest fragment lengths). The authors also quantified the relative amounts of RNA present per ml of biofluid and found that ePFP and cPFP contained 17 times less RNA than ePRP, urine, or conditioned medium. EVs derived from the biofluids had significantly less RNA, ranging from 2,763 times less for conditioned medium to 7.6 times less for urine. There were marked differences between the numbers of genes identified from EVs from platelet-free citrate plasma and platelet-free EDTA plasma, with the citrate plasma containing substantially fewer genes. Taken together, these findings indicate that methods for plasma sample isolation may require careful testing and optimization.

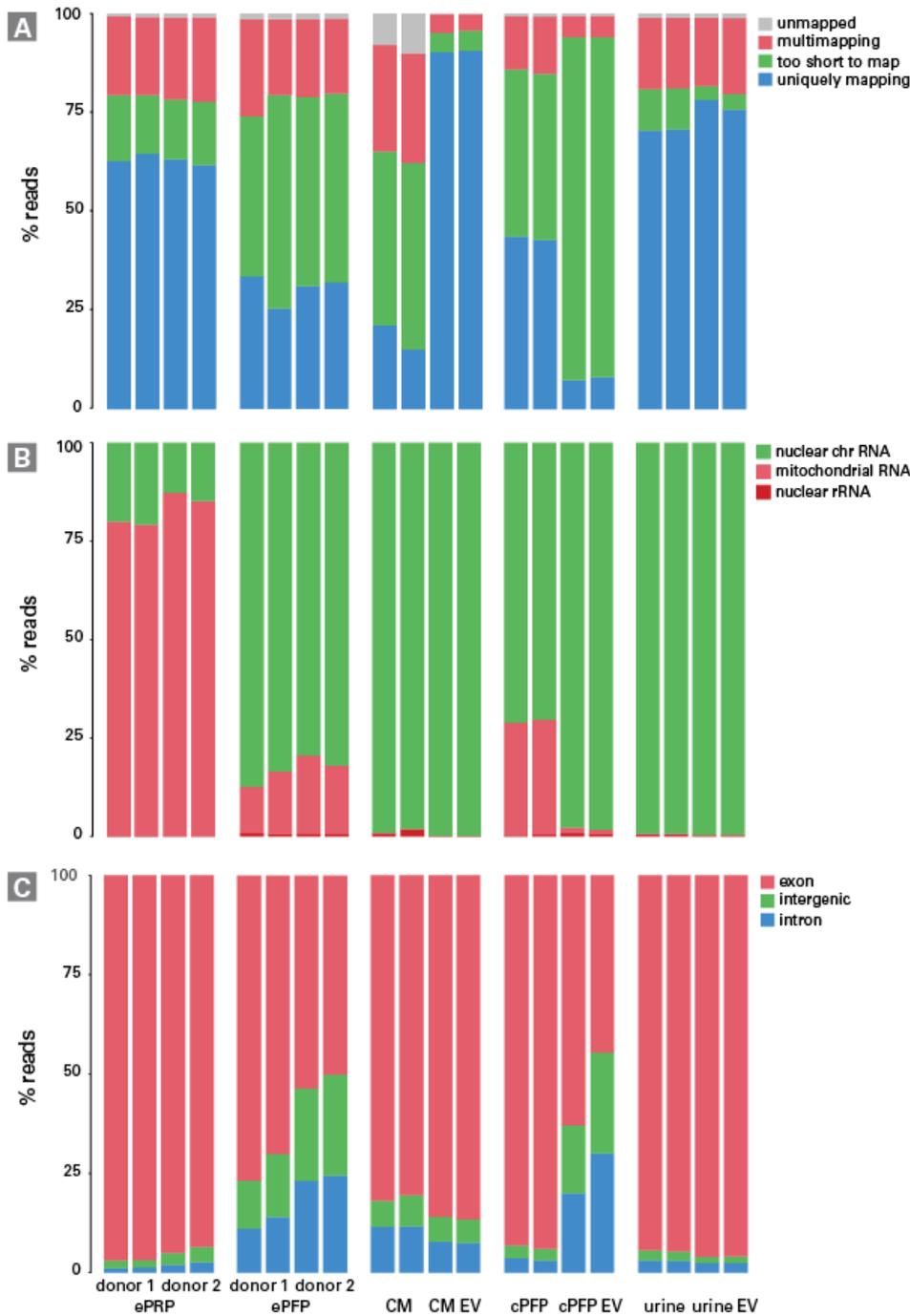


Figure 1. Read distribution of all libraries differs among samples. **Panel A.** Percentage of reads assigned as too short to map, uniquely mapping, multi-mapping, or unmapped. **Panel B.** Percentage of reads derived from nuclear RNA, mitochondrial RNA, and ribosomal RNA per sample. **Panel C.** Percentage of the reads originating from nuclear chromosomes derived from exonic, intronic, and intergenic regions per sample. All data was quantified with STAR. Figure used from Everaert et al. 2019 under [Creative Commons License 4.0](https://creativecommons.org/licenses/by/4.0/).

The SMARTer Stranded Total RNA-Seq method was able to generate full-length RNA-seq libraries for different human biofluids and their respective EVs. The method generated reproducible read distribution results for technical replicates, both at the RNA isolation and library preparation level. The results were specific to biofluid sample type.

Reproducibility

Scatter plots of the read counts from technical replicates of the ePRP and ePFP samples showed that gene counts were reproducible between independent RNA extractions of the same plasma sample (Figure 2). Cumulative distribution plots showed that reproducibility was even better when replication was introduced at the time of library preparation compared to RNA isolation.

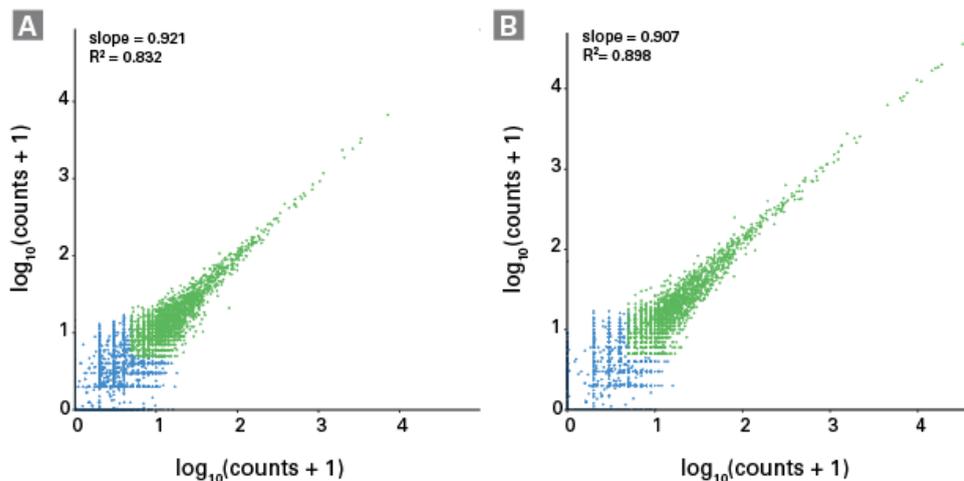


Figure 2. RNA isolation replicates of ePRP and ePFP show high repeatability. ePRP (Panel A) and ePFP (Panel B) replicate correlation with filtered (counts <4, blue) and retained genes (counts \geq 4, green) resulted in high Pearson correlation of 0.912 and 0.948, respectively. Figure used from Everaert et al. 2019 under [Creative Commons License 4.0](#).

The percentage of counts assigned to five gene types (protein-coding genes, lncRNA genes, miscellaneous RNA genes, pseudogenes, and other genes) differed between the biofluids. For example, ePRP contained a high number of pseudogene reads, resulting from mitochondrial genes, whereas ePFP reads mainly mapped to protein-coding genes. The authors also calculated the absolute numbers per gene biotype and concluded that the method was able to pick up many different classes of RNA molecules from many types of samples, including:

- Approximately 3,000 **lncRNA** genes could be distinguished in EVs isolated from urine. cPFP contained around 1,500 lncRNAs, while they could detect almost no lncRNAs in EVs isolated from this plasma. As expected, ePFP contained more lncRNAs than ePRP.
- cPFP and urinary EVs showed substantially more **circular RNAs** compared to other samples. Circular RNAs are presumed to be more stable and less degraded compared to linear forms, and are therefore considered ideal candidates for cancer biomarker discovery studies.

Biological differences in RNA content among biofluids

The comparison of RNA abundance profiles revealed the number of genes shared between the EVs and their biofluids of origin (Table I). Urine and urinary EVs showed great similarity in abundance of genes while plasma and plasma EVs differed most from each other. Similar studies including many more samples in each biofluid group could lead to new insights into specific RNA molecules that do and do not get transported via extracellular vesicles.

Table I. Comparison of RNA abundance profiles of genes shared between the EVs and their biofluids of origin.

Sample	Shared genes	Unique genes	Unique to EVs
Conditioned medium	4,891	755	1,853
Urine	>10,000	521	900
cPFP	1,528	7,211	70

Nucleic acids present in all sorts of biofluids and their EVs are promising biomarkers for diagnosis, prognosis, therapy response, and disease monitoring. Total RNA sequencing of biofluid samples allows the study of posttranscriptional regulation and provides a much more comprehensive view of the transcriptome. It enables the detection of various RNA biotypes including circular RNAs, histone RNAs, and long noncoding RNAs in addition to predefined targets or polyadenylated RNAs. [SMARTer Stranded Total RNA-Seq Kit v2—Pico Input Mammalian](#) provides a robust new method for RNA seq studies from low-input samples.

“ The advantage of the SMARTer Stranded Total RNA-Seq method is its potential to process low amounts of input material. Indeed, collecting samples is often the bottleneck of fundamental, (pre)clinical and translational research projects and being able to disseminate large amounts of information from only 200 μ L (or less) can substantially impact research progress.

References

Everaert, C. *et al.* [Performance assessment of total RNA sequencing of human biofluids and extracellular vesicles](#). *Sci. Reps.* **9**, 17574 (2019).



Takara Bio USA, Inc.

United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999
FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES. © 2023 Takara Bio Inc. All Rights Reserved. All trademarks are the property of Takara Bio Inc. or its affiliate(s) in the U.S. and/or other countries or their respective owners. Certain trademarks may not be registered in all jurisdictions. Additional product, intellectual property, and restricted use information is available at takarabio.com.