Blocking unwanted sequences by target specific high-affinity binding oligonucleotides during transcriptome library preparation

Celine Everaert°[1,2], Jasper Verwilt°[1,2], Kimberly Verniers[1,2], Niels Vandamme[2,3], Alvaro Marcos Rubio[1,2], Jo Vandesompele[1,2] and Pieter Mestdagh[1,2]*

[1] Department of Biomolecular Medicine, Ghent University, Ghent, Belgium

[2] Cancer Research Institute Ghent, Ghent University, Ghent, Belgium

[4] VIB Center for Inflammation Research, Vlaams Institute voor Biotechnologie, Zwijnaarde, Belgium

* To whom correspondence should be addressed. Email: Pieter.Mestdagh@UGent.be

"The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors"

ABSTRACT

RNA sequencing has become the gold standard for transcriptome analysis but comes with an inherent limitation with respect to quantification of low abundant transcripts. In contrast to microarray technology, RNA-sequencing reads are proportionally divided across transcripts. Therefore, low abundant RNAs can be out-competed by highly abundant—and sometimes non-informative—RNA species. We developed an easy-to-use strategy based on high-affinity RNA-binding oligonucleotides to block reverse transcription and PCR-amplification of specific RNA transcripts, hereby substantially reducing their abundance in the final sequencing library. To demonstrate the broad application potential of our method, we applied it to different RNA molecules and library preparation strategies, including YRNAs in small RNA-sequencing of human plasma samples, mitochondrial rRNAs in both 3' end sequencing and long-read sequencing, and MALAT1 in single-cell 3' end sequencing. We demonstrate that the blocking strategy is extremely efficient, reproducible and specific, and generally results in better transcriptome coverage and complexity. Our method does not require modifications of the library preparation procedure apart from adding blocking oligonucleotides to the RT reaction and can thus be easily integrated in virtually any RNA sequencing library preparation protocol.

INTRODUCTION

RNA sequencing has become the gold standard for transcriptome characterization. Researchers have developed numerous RNA-sequencing library preparation procedures to quantify various transcriptome subsets, including polyA+ RNA-sequencing, total RNA-sequencing, 3' end RNA-sequencing and small RNA-sequencing. Regardless of the library preparation method, RNA-sequencing reads are distributed across RNA transcripts proportionally to RNA transcript abundance.

Consequently, highly abundant RNA species, which are often deemed non-informative, can dominate the RNA-sequencing library, and hamper the detection of transcripts with lower abundance. A well-known example is ribosomal RNA (rRNA), which typically accounts for more than 80% of all RNA transcripts (1) in cellular or tissue RNA. Another well-documented example is YRNA4 fragments. YRNAs are non-coding, evolutionary conserved RNA species with a length of 80-110 nucleotides. Four different human YRNAs are known: hY1, hY3, hY4 and hY5 (2). YRNAs are readily fragmented in cells undergoing apoptosis in a caspase-dependent, Dicer-independent manner (3, 4). The resulting fragments reside in cultured cells (4), solid tumors (5) and multiple biofluids (6–8). More specifically, a 30-33 nucleotide 5'-end hY4 fragment is abundantly present in human plasma, serum and saliva (6–8), potentially serving a physiological function (9). In small RNA sequencing libraries of serum or plasma RNA, this fragment can account for more than 30% of all reads (10, 11) and even up to 70% in platelet-rich plasma. Superfluous amounts of hY4 fragments negatively impact the library complexity, requiring deeper sequencing to retrieve information about the other small RNA species in the library.

Removing sequence fragments derived from these excessively abundant transcripts from a sequencing library is instrumental in obtaining sufficient coverage of the informative fraction of the transcriptome without having to sequence libraries to extreme depth. Researchers have tackled this issue using two approaches: depleting the abundant transcripts or enriching the RNA species of interest. Strategies to deplete undesired and abundant transcripts encompass probe-based pull-down (1, 12), targeted duplex specific nuclease (DSN) digestion (13, 14) or not so random (NSR) primers (15, 16)—relying on a reverse transcription reaction that is negatively biased towards rRNA. On the other hand, molecular methods can enrich informative RNA transcripts using biotinylated probes to capture genes of interest (17). Capture probes can be designed to bind a sequence feature shared by many genes, like the polyA-tail (cfr. polyA+ RNA-sequencing), or exonic sequences in genes of interest (cfr. RNA capture-seq (18–21)). Alternatively, methods like 3' end sequencing apply polyA-priming to convert polyadenylated RNAs to cDNA selectively for further library preparation.

Depletion methods applying pull-down of the unwanted RNA tend to perform inconsistently, and their efficiency drops significantly when applied to fragmented RNA in e.g. biofluids or fixed tissues (22). Methods developed for small RNA depletion as used by Van Goethem et al. (23) are often labor-intensive and result in material loss by washing steps. A CRISPR-based strategy was recently described that cleaves unwanted sequences targeted with a single guide RNA (24). However, CRISPR-based technologies generally include a PCR step and multiple wash steps (24), making the protocol significantly longer and more prone to material loss.

Oligonucleotides containing modified nucleic acids, such as locked nucleic acids (LNAs), have been used for multiple applications, owing to their increased melting temperature and high discriminatory potential (25–27). More specifically, researchers have used LNA oligonucleotides to block the PCR amplification of specific XBP1 transcripts (28). A patent describing the use of LNA oligonucleotides to block reverse transcription and amplification of hemoglobin mRNA from whole blood during RT-qPCR (29) further exemplifies their potential and applicability.

Here, we describe an easy-to-implement method using locked LNA-modified oligonucleotides that bind unwanted RNA transcripts and block their reverse transcription and PCR amplification during RNA-sequencing library preparation. We applied our method to different abundant RNA species and RNA-sequencing library preparation strategies, including small RNA-sequencing, 3' end sequencing, long-read sequencing and single-cell 3' end sequencing. We demonstrate that the applied method, which does not require additional steps in the library prep procedure, is highly efficient and does not affect quantification of untargeted genes.

MATERIAL AND METHODS

YRNA blocking in human blood plasma samples

Samples and sample collection For the healthy donor experiments, venous blood was drawn from an elbow vein of two healthy donors in three EDTA tubes (BD Vacutainer Hemogard Closure Plastic K2-Edta Tube, 10 ml, #367525) using the BD Vacutainer Push blood collection set (21G needle). Collection of blood samples was according to the Ethical Committee of Ghent University Hospital approval EC/2017/1207, following the ICH Good Clinical Practice rules, and written informed consents of all donors were obtained. The tubes were inverted 5 times and centrifuged within 15 minutes after blood draw (400 g, 20 minutes, room temperature, without brake). Per donor, the upper plasma fraction were pipetted (leaving approximately 0.5 cm plasma above the buffy coat) and pooled in a 15 ml tube. After gently inverting, five aliquots of 220 µl platelet-rich plasma (PRP) were snap-frozen in 1.5 ml LoBind tubes (Eppendorf Protein LoBind microcentrifuge tubes Z666548 - DNA/RNA) in liquid nitrogen and stored at -80 °C. The remaining plasma was centrifuged (800 g, 10 minutes, room temperature, without brake) and transferred to a new 15 ml tube, leaving approximately 0.5 cm plasma above the separation. This plasma was centrifuged a 3rd time (2500 g, 15 minutes, room temperature, without brake), and transferred to a 15 ml tube, leaving approximately 0.5 cm above the separation. The resulting platelet-free plasma (PFP) was gently inverted, snap-frozen in five aliquots of 220 µl and stored at -80 °C. The entire plasma preparation protocol was finished in less than two hours. 200 µl PRP or PFP was used for each RNA isolation. For the spike-in RNA titration experiment, the protocol was identical, except for the fact that 4 EDTA tubes of 10 ml were drawn and that the second centrifugation step was different (1500 g, 15 minutes, room temperature, without brake).

For the cancer patient experiment, plasma samples are acquired from ProteoGenex (Inglewood, Uniteds States of America) under EC/2017/1515 from Ghent University Hospital. Blood was collected in EDTA vacutainer tubes. After inversion (10 times) the vacutainer tubes are centrifuged at 4 °C for 10 minutes at 1500 g without brakes. The plasma is then transferred into a 15 mL centrifuge tube and centrifuged for a second time for 10 minutes at 1500 g. Finally, the plasma was transferred into cryovials and stored at -80 °C until shipment. The cancer types included are colorectal cancer (CRC), lung adenocarcinoma (LUAD) and prostate cancer (PRAD).

RNA isolation and spike-in controls Total RNA was isolated from platelet-free (PFP) and platelet-rich plasma (PRP) using the miRNeasy Serum/Plasma Kit (Qiagen, Hilden, Germany, 217184). An input volume of 200 µl was used for each sample. For the cancer patient experiment, 2 µl of 1x RC PFP spikes was added to the plasma during isolation. The elution volume was 14 µl, 2µl of 1x LP PFP spikes (ThermoFisher) were added. Detailed description for the spike-in controls can be found in the exRNAQC study (30). From this total volume, 5 µl was used for the library preparation. For the healthy donor experiment, the eluate of multiple parallel extractions was pooled according to original biofluid type (PRP or PFP) and afterwards split in six aliquots of 5 µl this to minimize extraction bias. No gDNA removal step was included after RNA isolation.

YRNA LNA design The YNRA4 fragment (32 nucleotides) was tiled with 16 bp long complementary nucleotides resulting in 17 possible designs. This full set of antisense oligonucleotides was mapped against the human transcriptome (Ensembl v84) and miRbase. Only oligonucleotides with no off-targets when 3 mismatches are allowed were retained. Of the retained LNAs, the oligonucleotide with the highest Tm was chosen. The resulting fully modified LNA (ACCCACTACCATCGGA, targeting TCCGATGGTAGTGGGT) has a Tm of 89.9 °C. In addition to the fully LNA modified oligo, for the same sequence 2'-O-methyl and 2'-methoxy-ethoxy modified nucleotides and half modified (alternating modified – non-modified nucleotides) oligos were ordered at Integrated DNA Technologies. Sequences are available in Supplemental Table 1.

TruSeq small RNA library prep The TruSeq small RNA library prep sequencing kit (Illumina, San Diego, CA, USA) was used for library preparation according to manufacture instructions except for the changes listed below. After adaptor ligation and before the reverse transcription step, 2 µl LNA with a concentration of 0.25 µM (LNA1x) or 2.5 µM (LNA10x) was added to 14 µl of the adaptor ligated RNA. In the experiment with the cancer patient samples, only the 0.25µM (LNA1x) concentration was analyzed as we showed that the 10-fold higher concentration had no added value. As a negative control for LNA blocking (LNA0x), 2 µl of water was added to 14 µl of RNA. Next, 6 µl of each sample was used to start the reverse transcription and continue the library prep. Since the input amounts are low, the number of PCR cycles was set at 16 (while the manufacture recommends 11) during the final PCR step.

Pippin prep and sequencing Size selection for 125–163 bp was performed on all libraries using 3% agarose dye-free marker H cassettes on a Pippin Prep (Sage Science, Beverly, MA, USA). Next, the libraries were purified by precipitation using ethanol and resuspended with 10 mM Tris-HCl buffer (pH 8.0) with Tween 20. After dilution, the libraries were quantified using the KAPA Library Quantification Kit (Roche Diagnostics, Diegem, Belgium, KK4854). Healthy donor samples were sequenced using a NextSeq 500 using the NextSeq 500 High Output Kit v2.5 (75 cycles) (Illumina, San Diego, CA, USA). The library was loaded at a concentration of 2.0 pM with 10% PhiX, a total of 268 M reads was obtained. The cancer patient samples were loaded on one lane of a NovaSeq 6000 (Illumina, San Diego, CA, USA) instrument at a concentration of 300 pM with 10% PhiX. The NovaSeq 6000 SP Reagent Kit v1.5 (100 cycles) (Illumina, San Diego, CA, USA) was used (paired-end, 2x 50 cycli, only the first read was used for subsequent analysis) resulting in 267M reads. For the chemical

modification comparison experiment, one lane of a NovaSeq 6000 SP Reagent Kit v1.5 (100 cycles) (Illumina, San Diego, CA, USA, 20028401) (Illumina, San Diego, CA, USA) (1x100 bp) was used, loading 300 pM with 10% PhiX resulting in a total of 548M reads.

Quantification analysis For the quantification of small RNAs a dedicated in-house developed small RNA seq pipeline was used. This pipeline starts with adaptor trimming using Cutadapt (v1.8.1) (31), reads shorter than 15 nt and those in which no adaptor was found, were discarded. The reads with a low quality are discarded by using the FASTX-Toolkit (v0.0.14) (32) set at a minimum quality score of 20 in at least 80% of nucleotides. Reads belonging to our spike-in controls (both RC as LP) are counted and filtered out. The spike reads are subtracted from the fasta files and reads are counted. Subsequently, the reads were mapped with Bowtie (v1.1.2) (33) allowing one mismatch. At the end of the pipeline, the mapped reads are annotated by matching genomic coordinates of each read with genomic locations of miRNAs (obtained from miRBase, v20) and other small RNAs (obtained from UCSC GRCh37/hg19 and Ensembl v84). The original fastq-files and the count tables are submitted in EGA (EGAS00001006023). To compare within an experiment without bias, the samples are downsampled to 13M reads (concentration experiment), 6.5M reads (modification experiment) and 7M reads (cancer experiment).

Computational assessment Further data processing, including generation of visualizations and statistical testing, was done using R (v3.6.0) (34) including packages as tidyverse (v1.2.1) (35), biomaRt (v2.40.4) (36, 37), broom (v0.5.2) (38). For differential expression analysis limma-voom (v3.40.6) (39) was used on a filtered matrix with at least 10 reads per million (RPM) per miRNA over all samples.

Mitochondrial ribosomal RNA blocking in cell lysates

Cell culturing and RNA extraction We used HEK293T cells that were lysed with SingleShot lysis buffer (Bio-Rad, United States of America).

MtRNA LNA design From previous experiments, we noticed that there are three transcripts without poly(A) tail that are abundant (0.1-2% of all counts) in 3' end sequencing data of the cultured cells: MT-RNR1, MT-RNR2 and RNA45S. We inspected the RNA sequencing data using IGV_2.7.2 (40) and confirmed the presence of an adenosine-rich region flanking the abundant fragments observed in the sequencing library. For MT-RNR2, two different fragments were associated with an internal poly(A) stretch and both were contributing to the high number of gene counts. We investigated a design region of about 50 bases overlapping with the abundant fragments and used Bowtie (v1.2.3) (33) to map several 16 base-long putative LNA sequences. We retained the oligos with the lowest number of off-target hits. We then checked their binding capacities and biochemical characteristics. Sequences are available in Supplemental Table 1.

LNA treatment We combined four different LNA mixes (MT-RNR2_1, MT-RNR2_2, MT-RNR1 and RNA24S) to have a final solution containing each LNA at 25 µM (100x). We mixed 2 µl of LNA to 3 µl of RNA sample. From this solution, we used 2.5 µl as input for the library preparation.

Library preparation For the library preparation, we used the QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen, Austria). We performed the 'low input' version of the protocol.

Sequencing We sequenced the libraries using the NovaSeq 6000 SP Reagent Kit v1.5 (100 cycles) (Illumina, San Diego, CA, USA) on a NovaSeq 6000 (Illumina, San Diego, CA, USA) instrument at a concentration of 300 pM with 10% PhiX.

Quantification analysis Because we sequenced the libraries for too many cycles, thereby sequencing into the sequencing adaptor, we used Seqtk (v1.3) to remove the last 75 bases of each read. We then used BBMap (v38.26) to trim off the poly(A) tails and adapter sequences and to perform quality trimming. Next, all FASTQ files were subsampled to 2,000,000 reads with Seqtk (v1.3) and mapped to the genome using STAR (v2.6.0). We used SAMtools (v1.9) to count the reads mapping to the LNA targeted genomic regions. We used htseq-count (v0.11.0) (41) to generate the overall counts. Before initial trimming, before quality trimming and after quality trimming, we used FastQC (v0.11.9) to investigate the quality of the reads.

Computational assessment We used R (v4.1.0) (34) and tidyverse (v1.3.1) (35) and biomaRt (v2.48.3) (36, 37) to analyze and visualize the computationally generated data.

Mitochondrial ribosomal RNA blocking in direct-cDNA long-read sequencing

Cell culturing and harvesting We cultured HEK293T cells in RPMI complete medium to 80% confluence in a T75. The cells were washed with 2 ml versene and incubated with 2 ml of trypsin for 3 minutes at 37 °C. We neutralized the mixture with 8 ml fresh medium. We centrifuged for 5 minutes at 2000 rcf at 4 °C and removed the supernatants. We resuspended the cells in 1 ml of QIAzol and flash-froze the mixture in liquid nitrogen.

RNA extraction and quality control We extracted RNA using the RNeasy Micro Kit (Qiagen, Hilden, Germany, 217184) according to the manufacturer's protocol. We checked the quality of the RNA (RQN = 10) using the Fragment Analyzer RNA Kit (Agilent, United States of America).

LNA treatment We combined four different LNA mixes (MT-RNR2_1, MT-RNR2_2, MT-RNR1 and RNA24S) to have a final solution containing each LNA at 25 µM. We then made a 10-fold dilution series to obtain three different LNA solutions: LNA1x (0.25 µM), LNA10x (2.5 µM) and LNA100x (25 µM). For each library preparation, 2 µg of total RNA was mixed with 2 µl of the corresponding LNA dilution. 1 µl of RNase-free water was added to 2 µg of total RNA as a non-treated sample (LNA0x). The samples are place on ice for 5 minutes.

Library preparation We prepared direct-cDNA libraries using the SQK-DCS109 Kit (Oxford Nanopore Technologies, United Kingdom). The exact protocol was followed except for the following changes: the RNA-bead binding steps were performed for 5 minutes on a Hula Mixer and 5 minutes on the bench at room temperature; the RNA elution steps were performed for 5 minutes at 37 °C and 5 minutes on a Hula Mixer at room temperature; and 300 µl of 80% ethanol was used for the beads wash steps.

Oxford Nanopore sequencing We sequenced each library using two Flongle Flow Cells (Oxford Nanopore Technologies, United Kingdom) with a MinION Sequencer (Oxford Nanopore Technologies, United Kingdom). Sequencing was either stopped after 24 hours or when no more pores were available.

Quantification analysis We basecalled the raw fast5 files using Guppy (v3.5.2) (42) on a GPU. We grouped reads per sample and used Pychopper (v2.3.1) (43) to identify full-length transcripts containing both primer sequences. We mapped the reads with Minimap2 (v2.11) (44) and extracted reads mapping to the target fragment location using SAMtools (v1.11) (45). We then used NanoComp (v1.12.0) (46) to check the read length and quality of each sample.

Computational assessment We used R (v4.1.0) (34) and tidyverse (v1.3.1) (35) and biomaRt (v2.48.3) (36, 37) to analyze and visualize the computationally generated data.

MALAT1 blocking in single-cell 3' end sequencing for *peripheral blood mononuclear cells* (PBMCs)

PBMCs preparation Whole blood was collected in EDTA tubes. Blood was transferred to Leucosep filtered tubes (Greiner Bio-One) containing 15 ml of Ficoll Paque Plus (Cytiva, Washington, D.C., USA, 17144002) and diluted (1:2) with the same volume of 1X DPBS (ThermoFisher, Waltham, MA, USA, 14190144). Samples were centrifuged at room temperature for 18 minutes at 800 rcf, and the resulting buffy coat containing the PBMCs was extracted. The extracted PBMCs were centrifuged and washed twice with 1X DPBS (ThermoFisher, Waltham, MA, USA, 14190144). A counting sample was taken, and the cell viability and concentration were assessed with a Neubauer chamber, counting at least two different squares. PBMCs were then resuspended in freezing mix (complete medium (RPMI + 1% pen/strep + 10% FCS) + 10% DMSO) in cryovials with no more than 10 million cells. Vials were stored first at -80 °C inside a freezing container for 24 h and then at -150 °C. The sample was thawed just before live-death sorting.

MALAT1 LNA design Based on previous 3' end sequencing data from PBMCs, the optimal design space was identified (Supplemental Figure 8). We noticed two internal poly(A) sequences contributing to the high number of counts. Next, we designed and characterized the best LNA sequences following the similar steps as before (see 'mtRNA LNA design', but with a length of 18 nucleotides). Sequences are available in Supplemental Table 1.

LNA treatment The LNAs were diluted at 125 µM concentration of which 2 µl was used. For the pre-RT blocking, 2 µl of the oligonucleotide mix was added to the master mix (including the RT reagent, template switching oligo, reducing reagent B and RT enzyme C). The master mix is than combined with the cell suspension to a total volume of 80 µl. For the pre-cDNA amplification blocking, 2 µl of the oligonucleotide mix was added to the cDNA amplification mix (including Amp Mix and cDNA primers).

Library preparation Sorted single-cell suspensions were resuspended in PBS+0.04% BSA at an estimated final concentration of 1000 cells/µl and loaded on a Chromium GemCode Single Cell

Instrument (10x Genomics, Pleasonton, CA, USA, 1000204), Chip G (10x Genomics, Pleasonton, CA, USA, #2000177) to generate single-cell gel beads-in-emulsion (GEM). We prepared the scRNA-Seq libraries using the GemCode Single Cell 3' Gel Bead and Library kit, version NextGEM 3.1 (10x Genomics, Pleasonton, CA, USA, PN-1000121) according to the manufacturer's instructions.

Sequencing The Chromium libraries were equimolarly pooled and loaded on a NovaSeq 6000 (Illumina, San Diego, CA, USA) instrument in standard mode with a final loading concentration of 340 pM and 2% PhiX. With a SP100 cycles (Illumina, San Diego, CA, USA, 20028401) kit a total of 952M reads with q30 of 91.32% were obtained.

Quantification analysis Demultiplexing of the bcl files was performed with cellranger mkfastq (v6.0.1), after which gene counts per cell were obtained with cellranger count (v6.0.1).

Computational assessment The count matrixes were loaded into R (v4.1.0) (34) and further processed, including the integration and annotation, with Seurat (v4.0.3) (47). The cells were not filtered. The data was analyzed and visualized using tidyverse (v1.2.1) (35).

For further clarity of experimental designs, we also request a list of essential sub-heading under the Materials and Methods Section, for availability of reagents, deposited data, programs, web sites, etc. These rules apply to all manuscripts, regardless of Standard research or Methods submissions.

Reagents: This sub-heading should include enzymes, antibodies, kits, specialized commercial instruments, non-standard chemicals, peptides, recombinant proteins, et al. with reference to company name, location, and catalog #.

- 1X DPBS (ThermoFisher, Waltham, MA, USA, 14190144).
- Chromium GemCode Single Cell Instrument (10x Genomics, Pleasonton, CA, USA, 1000204)
- Direct cDNA sequencing kit (Oxford Nanopore Technologies, UK, SQK-DCS109)
- Eppendorf Protein LoBind microcentrifuge tubes (Eppendorf, Hamburg, Germany, Z666548)
- Ficoll Paque Plus (Cytiva, Washington, D.C., USA, 17144002
- Flongle Flow cell (Oxford Nanopore Technologies, UK, FLO-FLG001)
- Fragment Analyzer RNA Kit (Agilent, USA, DNF-471-0500)
- GemCode Chip G (10x Genomics, Pleasonton, CA, USA, 2000177)
- GemCode Single Cell 3' Gel Bead and Library kit, version NextGEM 3.1 (10x Genomics, Pleasonton, CA, USA, PN-1000121)
- KAPA Library Quantification Kit (Roche Diagnostics, Diegem, Belgium, KK4854)
- MinION sequencer (Oxford Nanopore Technologies, UK, MIN-101B)
- miRNeasy Serum/Plasma Kit (Qiagen, Hilden, Germany, 217184).
- NextSeq 500 High Output Kit v2.5 (75 cycles) (Illumina, San Diego, CA, USA, 20024906)
- NextSeq 500 Sequencing System (Illumina, San Diego, CA, USA, SY-415-1001)
- NovaSeq 6000 Sequencing System (Illumina, San Diego, CA, USA, 20012850)

- NovaSeq 6000 SP Reagent Kit v1.5 (100 cycles) (Illumina, San Diego, CA, USA, 20028401)
- Pippin Prep (Sage Science, Beverly, MA, USA, PIP0001).
- QuantSeq 3' mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen, Austria, 139.96)
- RNeasy Micro Kit (Qiagen, Hilden, Germany, 217184)
- SingleShot lysis buffer (Bio-Rad, United States of America, 1725080)
- TruSeq small RNA library prep sequencing kit (Illumina, San Diego, CA, USA, RS-200)
- Vacutainer Hemogard Closure Plastic K2-Edta Tube, 10 ml, (BD, Franklin Lakes, NJ, USA, 367525)
- Vacutainer Push blood collection set (BD, Franklin Lakes, NJ, USA, 368657)

Biological Resources: This sub-heading should list all cell lines, organisms and strains, plasmid vectors, et al. Each resource should include its repository with location and resource #, web link, and reference, if appropriate.

- HEK293T (ATCC, Manassas, VA, USA)

Computational Resources: Authors should ensure that all computational resources used (databases, web servers etc.) are appropriately acknowledged. URLs should be provided in the text and peer-reviewed articles describing these resources (if available) should be cited and included in the Reference list.

RESULTS

To prevent incorporation of unwanted RNA transcripts or fragments in RNA-sequencing libraries, we reasoned that LNA-modified oligonucleotides would block reverse transcription and PCR amplification when bound downstream of the priming site because of their extremely high affinity to RNA and DNA. The approach we took to design blocking LNA oligonucleotides depends on the characteristics of the unwanted RNA sequence and the library prep procedure. We selected four different library prep procedures and defined high abundant, and mostly unwanted, target RNA sequences for LNA oligonucleotide design (Figure 1). These include YRNA in human blood plasma small RNA-sequencing libraries, mitochondrial rRNA in 3' end sequencing libraries and long read sequencing libraries of HEK293T cells, and MALAT1 in single-cell 3' end sequencing libraries of PMBCs. To block RT and PCR of short fragments like YRNA in small RNA-seq libraries, we designed the 18 nucleotide LNA to be complementary to the 3' end of the 30 nucleotide YRNA fragment (Figure 1A). For longer fragments, like mitochondrial rRNA and MALAT1, the LNA oligonucleotide was designed to bind directly downstream of the poly(A) RT-priming site (Figure 1B-D). As the LNA oligonucleotides are added directly to the RT reaction (see details in Material and Methods for each of the protocols), no additional steps are required in the RNA-library prep protocol.

<u>YRNA blocking in plasma samples for TruSeq small RNA sequencing</u>

Efficient blocking of YRNA4 in PRP and PFP

We first focused on blocking RT and amplification of YRNA4 fragments in human blood plasma small RNA-seq libraries. Blocking efficiency was tested on both platelet-rich plasma (PRP) and platelet-free plasma (PFP) of healthy donors, with PRP having the highest fraction of YRNA4 fragments. Two different concentrations of a blocking LNA oligonucleotide (0.25 µM and 2.5 µM, referred to as LNA1x and LNA10x, respectively) were spiked in the RT reaction of the TruSeq small RNA library prep and results were compared to the standard workflow. Only 0.09% YRNA4 in PFP and up to 0.16% YRNA4 in PRP (Figure 2A) was observed when adding LNA1x to the RT reaction, or respectively a 477- and 468-fold reduction compared to the standard protocol. Increasing the LNA concentration 10-fold (LNA10x) provided no benefit compared to LNA1x, with a 228-fold and 262-fold reduction of YRNA4. The strong reduction in YRNA4 fragments was accompanied by a strong increase in fraction of microRNA reads, from 49.55% to 79.67% for PFP and from 17.24% to 74.61% for PRP. Since the LNA1x condition resulted in sufficient reduction in YRNA and increase in microRNA read fraction, we decided to use the 1x concentration (0.25 µM) for the subsequent experiments.

YRNA4 blocking increases microRNA coverage and preserves fold changes

We then evaluated the reproducibility of our YRNA4 blocking protocol by comparing miRNA abundance between technical library preparation replicates. Reproducibility upon YRNA4 blocking was similar to that of the standard workflow, as evidenced by similar Pearson (0.999-1) and Spearman (0.70-0.78 for PFP and 0.81-0.82 for PRP) correlation coefficients (Figure 2B). To investigate if abundance of some miRNAs is affected by YRNA4 blocking (due to off-target effects), we compared miRNA abundance (reads per million) between the YRNA4 blocking procedure and the standard workflow. Only one miRNA showed a high standardized residual (>2) in all samples and replicates. This miRNA (miR-106b-3p) showed a consistent lower abundance in the LNA1x libraries compared to the control. Of note, we did not observe any sequence similarity between the YRNA4 fragment and miR-106-3p, suggesting that non-specific binding of the LNA is unlikely. In general, miRNA expression correlations between the standard protocol and LNA1x spike protocol (Figure 2C) were comparable to these of technical replicates (Pearson correlation = 1.00, Spearman correlation = 0.67-0.72 for PFP and 0.81 for PRP). In PFP, the impact of YRNA4 blocking on the number of detected miRNAs was limited, with only nine additional miRNAs detected upon subsampling for library size correction (Figure 2D). As expected, the miRNAs that were uniquely detected in either the standard or LNA1x spike protocol were low abundant (Supplemental Figure 1A&1B). In PRP, we detected 183 additional miRNAs in the LNA1x spike protocol. All except three miRNAs detected with the standard protocol were also detected with the LNA1x spike protocol (Figure 2E, Supplemental Figure 1C). Not only does YRNA4 blocking increase the number of detected miRNAs, it also results in increased miRNA coverage (a 3 fold median RPM increase) (Supplemental Figure 1D). Taken together, YRNA4 blocking in plasma small RNA-sequencing libraries improves both miRNA library complexity and miRNA coverage.

We then assessed the impact of YRNA4 blocking on differential miRNA abundance between samples. To address this, we examined the miRNA fold changes between PFP sample from patients with diverse tumor types (colorectal cancer or CRC (n=4), prostate adenocarcinoma or PRAD (n=4) and

lung adenocarcinoma or LUAD (n=4)) that were processed with the standard and LNA1x spike protocol (Supplemental Figure 2). Differences in miRNA abundance between cancer types were highly concordant between both datasets (Figure 3A). At high sequencing depth (7M reads), both methods detect an equal number of miRNAs that are differential expressed. However, when subsampling reads to lower sequencing depth, the impact on differential miRNA expression was lower in case of YRNA4 blocking (Figure 3B). At 0.7M reads, the concordance between the significantly differential genes for each replicate is 64.9% for YRNA4 blocking while its 60.0% for the original protocol. When using 7M reads, the concordance is 100% for both protocols.

LNA is most efficient modification to block library incorporation

As fully modified LNA oligonucleotides are relatively expensive, we evaluated the YRNA4 blocking potency of cheaper base modifications that are known to improve oligonucleotide binding affinity such as 2'-O-methyl (2'OME) and 2'-methoxy-ethoxy (2'MOE). We observed that an LNA modified YRNA4 oligonucleotide is equally efficient (median reduction of 6.45 fold compared to a 2'MOE modified oligonucleotide (p=0.14). While the 2'MOE oligonucleotide still reduces the YRNA4 abundance by 6.11 fold, the 2'OME is less efficient and resulted in reduction of just 1.22 fold (p=0.005). In addition, we investigated the potency of partially modified (i.e. every other nucleotide) oligonucleotides for both LNA, 2'OME and 2'MOE. We observed that the partially modified LNA YRNA4 oligonucleotide was as potent as a fully modified LNA YRNA4 oligonucleotide (YRNA4 fold change reduction of 6.90, p=0.383) and still outperforms fully modified 2'OME YRNA4 oligonucleotides (p=0.0005) (Figure 3C).

<u>rRNA blocking in 3' end sequencing data</u>

During reverse transcription, oligo(T) primers can bind internal poly(A) sequences of mitochondrial and nuclear ribosomal RNA species, which eventually get incorporated in the RNA-sequencing library (up to 2% of all reads, as found in previous sequencing data (Figure 1C-E)). We designed LNA oligonucleotides to inhibit reverse transcription of three mitochondrial rRNA fragments (MT-RNR1 and two fragments from MT-RNR2) and one fragment from nuclear rRNA RNA45S (Figure 1). All four oligonucleotides were added to the RT reaction of a 3' end library preparation on eight cell lysates, and data were compared to that of a standard 3' end library preparation workflow. Addition of the LNA oligonucleotides resulted in an average reduction of the counts per million of 16.2x, 19.2x, 8.6x and 3.2x for RNA45S, MT-RNR1, MT-RNR2 fragment 1 and MT-RNR2 fragment 2, respectively (Figure 4A). To evaluate reproducibility of the method, we compared abundance of all detected genes between biological replicates for both the standard protocol and the LNA spike protocol. The Pearson (0.982-0.997) and Spearman (0.852-0.879) correlation coefficients were high for every comparison (Supplemental Figure 3) and there was no significant difference in reproducibility between the standard and blocking protocol. To investigate whether blocking RNA45S and the MT-RNR1/2 fragments is beneficial for gene detection, we evaluated the number of detected genes for different sequencing depths (Figure 4B). For shallow sequencing depth (1-2 million reads), the number of detected genes was higher in the blocking protocol compared to the standard protocol. Finally, we investigated potential off-target effects of the blocking oligonucleotides by comparing gene expression

values between the control and blocking protocol. Out of 12,077 detected genes, we identified two genes that showed divergent gene expression values for all biological replicates: MT-ATP8 and H4C3 (Figure 4C). We did not observe significant sequence complementarity between the LNA oligonucleotides and these presumed off-targets. In conclusion, LNA oligonucleotides can efficiently and specifically block the incorporation of a variety of transcripts in 3' end RNA-seq libraries.

rRNA blocking for long-read polyA+ transcript sequencing

Additionally, we explored whether the rRNA (RNA45S, MT-RNR1, MT-RNR2) blocking strategy as described above can also be applied to Oxford Nanopore Technologies (ONT) sequencing of poly(A)-primed cDNA libraries. More specifically, we performed direct-cDNA sequencing to investigate the blocking effect on just the reverse transcription step. We added three different concentrations (0.25 µM, 2.5 µM and 25 µM, referred to as LNA1x, LNA10x and LNA100x) of the same rRNA LNA oligonucleotides as used in the 3' end sequencing data to the reverse transcription reaction of four different samples. For all targeted fragments, we observed a strong decrease in counts per million with increasing concentration of LNA oligonucleotides, except for 45S pre-ribosomal RNA in the LNA100x condition (Figure 5). Unexpectedly, we also observed a mild but consistent decrease in overall read length distribution with increasing concentration of LNA oligonucleotides (Supplemental Figure 4). The quality scores of the reads did not vary (Supplemental Figure 5). These results show the potential of LNA oligonucleotides to prevent reverse transcription (and thus sequencing) of specific RNA molecules in ONT long-read sequencing experiments.

MALAT1 blocking in single-cell 3' end sequencing of PBMCs

We finally evaluated if our method would also be applicable to single-cell RNA-sequencing. More specifically, we designed two LNA oligonucleotides to block MALAT1 in single-cell 3' end sequencing libraries of PBMCs. In PBMCs, MALAT1 can consume > 40% percent of reads through priming of internal poly(A) stretches (Supplemental Figure 6). The LNA oligonucleotides were added either before reverse transcription (pre-RT), which occurs in the gel bead-in-emulsion (GEMs), or before cDNA amplification, when the GEMs are pooled (pre-PCR). Both protocols show a decrease in MALAT1 reads (6-fold (x% reduction) for the pre-RT and 4-fold (y% reduction) for the pre-PCR blocked libraries) (Figure 6A). For some cell types, e.g. erythrocytes and regulatory T-cells, the initial MALAT1 proportions were higher, resulting in a more drastic reduction (Supplemental Figure 7). For the pre-RT sample, we observed a larger mitochondrial derived RNA fraction, which might indicate cell death (Figure 6A). We therefore focused our analysis on the pre-PCR protocol. UMAP representation of cells based on single-cell RNA-seq data from both the pre-PCR blocking and standard protocol revealed tight clustering of cell types independent of protocol (Figure 6B), implying that the MALAT1 LNA oligonucleotide in the pre-PCR protocol has minimal impact on gene expression. This was further demonstrated by a perfect correlation (Spearman and Pearson correlation = 1.00) of gene expression values between the pre-PCR blocking and standard protocol (Supplemental Figure 8). We observed a small but significant increase in the mean number of detected genes per cell in the pre-PCR protocol; 1173 genes with at least two counts in the control

sample and 1192 in the pre-PCR blocking sample (p = 5.751e-05) (Figure 6A). Of note, the increase in detected genes was cell type dependent. For CD16 monocytes, although they are not the cells with the highest original MALAT1 read percentages, the pre-PCR protocol resulted in 91 additional genes (+9%).

DISCUSSION

We demonstrate that high-affinity binding oligonucleotides can be applied to block reverse transcription and PCR amplification of various RNA transcripts in a plethora of library preparation protocols. We present a method that is both reproducible and robust, and that can drastically increase the detection and coverage of (low abundant) genes in the library. Next, we provided evidence that the LNA molecules exert their blocking activity during reverse transcription as well as during PCR, indicated by clear fragment depletion in the PCR-independent Oxford Nanopore Technologies protocol and the post-RT single-cell sample, respectively. Additionally, we show how the impact of blocking certain fragments on gene detection and coverage strongly depends on the abundance of the blocked fragments, the sample type and the sequencing depth. For example, the impact of YRNA4 blocking on gene (i.e. miRNA) detection and coverage was much larger for PRP than PFP, most likely because of higher YRNA4 abundance in PRP. At lower sequencing depth, we expect a more pronounced impact in PFP samples.

Our method has several advantages compared to existing protocols. Besides the highly convenient and easy to use protocol, no library material is lost because of enrichment or washing steps, hereby maximizing detection sensitivity.

While we generally observe potent blocking of targeted transcripts, we also observed minor unwanted effects. With ONT sequencing, the read lengths decreased with increasing concentrations of LNA oligonucleotides. One possible explanation is that the LNA oligonucleotide might indistinctly inhibit the reverse transcription reaction. We investigated only full-length (adaptor-to-adaptor) reads and did not find any off-target effects related to sequence complementarity. Second, adding LNA oligonucleotides to the GEMs during 3’ end sequencing resulted in more mtRNA reads. Since the cells are only lysed after 18 min, the LNA oligonucleotides may enter the living cells and induce cell death. The larger the fraction of mtRNA genes for specific cell types, the fewer genes that are detected. Third, the optimal concentration of LNA oligonucleotide may be application and target dependent. A dedicated optimization step is warranted for optimal performance. This is reflected in the scRNA-sequencing experiment, where the benefit (in means of number of genes) depends on the cell type. Factors to consider are original fraction of the targeted read(s) and input RNA concentration of the library preparation protocol. Finally, we observed a limited number of off-target effects upon addition of certain LNA oligonucleotides (for instance MT-AT8 and H4C3 in the 3’ end sequencing experiment). We did not observe significant sequence complementarity between the LNA oligonucleotides and the presumed off-targets. Nevertheless, off-target effects are not entirely unexpected given the relatively short length of the LNAs, their high RNA-binding capacity and the small design space. To improve specificity, increasing oligonucleotide length or reducing the number of LNA nucleotides to lower

binding affinity may prove beneficial. A final limitation of the method is that it may only be applicable to small RNA-sequencing, or library prep methods employing an oligo(T) reverse transcription primer or a gene-specific targeted RT primer. When the priming is random, it is not possible to design a single LNA oligonucleotide to block reverse transcription of the entire fragment. One option would be to design multiple LNA oligonucleotides spanning the entire transcript, but this would become rather expensive. While the cost of an LNA modified oligonucleotide is relatively high, the amount required for efficient blocking is small. Even at low synthesis scale, several hundreds of reactions can be performed, resulting in a limited per-sample cost. Moreover, blocking unwanted transcripts may help reduce the sequencing cost.

We believe the method presented here is versatile, and can be used for other applications not investigated here, including hemoglobin mRNA blocking in whole blood samples (up to 70% of all mRNA in whole blood (48)) or trypsin mRNA in pancreatic RNA samples. While we only investigated mixtures of up to four different LNA oligonucleotides, it would be possible to combine many more, and block a variety of different fragments in one sample. Such mixtures can be designed specifically for unique and challenging sample types, which can contain several highly expressed, uninformative fragments (49).

In conclusion, we present a novel and broadly applicable tool to specifically block unwanted RNA species during RNA sequencing library preparations by simply adding a target-specific LNA oligonucleotide to the RT or PCR reaction.

AVAILABILITY

The generated sequencing data is available through EGA with accession ID EGAS00001006023

ACCESSION NUMBERS

EGAS00001006023

SUPPLEMENTARY DATA

Supplementary Data statement:

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENT

Celine Everaert: Conceptualization, Methodology, Software, Formal Analysis, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization, Funding Acquisition. **Jasper Verwilt:** Conceptualization, Methodology, Software, Formal Analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization. **Kimberly Verniers:** Investigation, Writing – Review & Editing. **Niels Vandamme:** Investigation,

Resources, Writing – Review & Editing, Funding Acquisition. **Alvaro Marcos Rubio:** Investigation, Resources, Writing – Review & Editing. **Jo Vandesompele:** Conceptualization, Resources, Writing – Review & Editing, Supervision, Funding Acquisition. **Pieter Mestdagh:** Conceptualization, Resources, Writing – Review & Editing, Supervision, Funding Acquisition.

FUNDING

This work was funded by 'Fonds Wetenschappelijk Onderzoek' Flanders; Ghent University; Kom op tegen Kanker (Stand up to Cancer), the Flemish cancer society and Stichting Tegen Kanker

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. O'Neil,D., Glowatz,H. and Schlumpberger,M. (2013) Ribosomal RNA Depletion for Efficient Use of RNA-Seq Capacity. In *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc., Hoboken, NJ, USA, Vol. 103, pp. 4.19.1-4.19.8.

2. Hendrick,J.P., Wolin,S.L., Rinke,J., Lerner,M.R. and Steitz,J.A. (1981) Ro small cytoplasmic ribonucleoproteins are a subclass of La ribonucleoproteins: further characterization of the Ro and La small ribonucleoproteins from uninfected mammalian cells. *Mol. Cell. Biol.*, **1**, 1138–1149.

3. Rutjes,S.A., Van Der Heijden,A., Utz,P.J., Van Venrooij,W.J. and Pruijn,G.J.M. (1999) Rapid nucleolytic degradation of the small cytoplasmic Y RNAs during apoptosis. *J. Biol. Chem.*, **274**, 24799–24807.

4. Nicolas,F.E., Hall,A.E., Csorba,T., Turnbull,C. and Dalmay,T. (2012) Biogenesis of Y RNA-derived small RNAs is independent of the microRNA pathway. *FEBS Lett.*, **586**, 1226–1230.

5. Meiri,E., Levy,A., Benjamin,H., Ben-David,M., Cohen,L., Dov,A., Dromi,N., Elyakim,E., Yerushalmi,N., Zion,O., *et al.* (2010) Discovery of microRNAs and other small RNAs in solid tumors. *Nucleic Acids Res.*, **38**, 6234–6246.

6. Ishikawa,T., Haino,A., Seki,M., Terada,H. and Nashimoto,M. (2017) The Y4-RNA fragment, a potential diagnostic marker, exists in saliva. *Non-coding RNA Res.*, **2**, 122–128.

7. Dhahbi,J.M., Spindler,S.R., Atamna,H., Boffelli,D., Mote,P. and Martin,D.I.K. (2013) 5′-YRNA fragments derived by processing of transcripts from specific YRNA genes and pseudogenes are abundant in human serum and plasma. *Physiol. Genomics*, **45**, 990–998.

8. Ninomiya,S., Kawano,M., Abe,T., Ishikawa,T., Takahashi,M., Tamura,M., Takahashi,Y. and Nashimoto,M. (2015) Potential Small Guide RNAs for tRNase ZL from Human Plasma,

Peripheral Blood Mononuclear Cells, and Cultured Cell Lines. *PLoS One*, **10**, e0118631.

9. Ninomiya,S., Ishikawa,T., Takahashi,M., Seki,M. and Nashimoto,M. (2019) Potential physiological roles of the 31/32-nucleotide Y4-RNA fragment in human plasma. *Non-coding RNA Res.*, **4**, 135–140.

10. Dhahbi,J.M., Spindler,S.R., Atamna,H., Boffelli,D. and Martin,D.I.K. (2014) Deep Sequencing of Serum Small RNAs Identifies Patterns of 5′ tRNA Half and YRNA Fragment Expression Associated with Breast Cancer. *Biomark. Cancer*, **6**, BIC.S20764.

11. Yan,Y., Wang,X., Venø,M.T., Bakholdt,V., Sørensen,J.A., Krogdahl,A., Sun,Z., Gao,S. and Kjems,J. (2017) Circulating miRNAs as biomarkers for oral squamous cell carcinoma recurrence in operated patients. *Oncotarget*, **8**, 8206–8214.

12. Benes,V., Blake,J. and Doyle,K. (2011) Ribo-Zero Gold Kit: improved RNA-seq results after removal of cytoplasmic and mitochondrial ribosomal RNA. *Nat. Methods*, **8**, iii–iv.

13. Archer,S.K., Shirokikh,N.E. and Preiss,T. (2014) Selective and flexible depletion of problematic sequences from RNA-seq libraries at the cDNA stage. *BMC Genomics*, **15**, 401.

14. Yi,H., Cho,Y.-J., Won,S., Lee,J.-E., Yu,H.J., Kim,S., Schroth,G.P., Luo,S. and Chun,J. (2011) Duplex-specific nuclease efficiently removes rRNA for prokaryotic RNA-seq. *Nucleic Acids Res.*, **39**, e140.

15. Armour,C.D., Castle,J.C., Chen,R., Babak,T., Loerch,P., Jackson,S., Shah,J.K., Dey,J., Rohl,C.A., Johnson,J.M., *et al.* (2009) Digital transcriptome profiling using selective hexamer priming for cDNA synthesis. *Nat. Methods*, **6**, 647–649.

16. Arnaud,O., Kato,S., Poulain,S. and Plessy,C. (2016) Targeted reduction of highly abundant transcripts using pseudo-random primers. *Biotechniques*, **60**, 169–174.

17. Levin,J.Z., Berger,M.F., Adiconis,X., Rogov,P., Melnikov,A., Fennell,T., Nusbaum,C., Garraway,L.A. and Gnirke,A. (2009) Targeted next-generation sequencing of a cancer transcriptome enhances detection of sequence variants and novel fusion transcripts. *Genome Biol.*, **10**, R115.

18. Mercer,T.R., Gerhardt,D.J., Dinger,M.E., Crawford,J., Trapnell,C., Jeddeloh,J.A., Mattick,J.S. and Rinn,J.L. (2012) Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.*, **30**, 99–104.

19. Mercer,T.R., Clark,M.B., Crawford,J., Brunck,M.E., Gerhardt,D.J., Taft,R.J., Nielsen,L.K., Dinger,M.E. and Mattick,J.S. (2014) Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. *Nat. Protoc.*, **9**, 989–1009.

20. Clark,M.B., Mercer,T.R., Bussotti,G., Leonardi,T., Haynes,K.R., Crawford,J., Brunck,M.E.,

Cao,K.A.L., Thomas,G.P., Chen,W.Y., *et al.* (2015) Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing. *Nat. Methods*, **12**, 339–342.

21. Morlion,A., Everaert,C., Nuytens,J., Hulstaert,E., Vandesompele,J. and Mestdagh,P. (2021) Custom long non-coding RNA capture enhances detection sensitivity in different human sample types. *RNA Biol.*, **18**, 215–222.

22. Herbert,Z.T., Kershner,J.P., Butty,V.L., Thimmapuram,J., Choudhari,S., Alekseyev,Y.O., Fan,J., Podnar,J.W., Wilcox,E., Gipson,J., *et al.* (2018) Cross-site comparison of ribosomal depletion kits for Illumina RNAseq library construction. *BMC Genomics*, **19**, 199.

23. Van Goethem,A., Yigit,N., Everaert,C., Moreno-Smith,M., Mus,L.M., Barbieri,E., Speleman,F., Mestdagh,P., Shohet,J., Van Maerken,T., *et al.* (2016) Depletion of tRNA-halves enables effective small RNA sequencing of low-input murine serum samples. *Sci. Rep.*, **6**.

24. Hardigan,A.A., Roberts,B.S., Moore,D.E., Ramaker,R.C., Jones,A.L. and Myers,R.M. (2019) CRISPR/Cas9-targeted removal of unwanted sequences from small-RNA sequencing libraries. *Nucleic Acids Res.*, 10.1093/nar/gkz425.

25. Breitenbuecher,F., Hoffarth,S., Worm,K., Cortes-Incio,D., Gauler,T.C., Köhler,J., Herold,T., Schmid,K.W., Freitag,L., Kasper,S., *et al.* (2014) Development of a Highly Sensitive and Specific Method for Detection of Circulating Tumor Cells Harboring Somatic Mutations in Non-Small-Cell Lung Cancer Patients. *PLoS One*, **9**, e85350.

26. Singh,S.K., Nielsen,P., Koshkin,A.A. and Wengel,J. (1998) LNA (locked nucleic acids): Synthesis and high-affinity nucleic acid recognition. *Chem. Commun.*, 10.1039/a708608c.

27. Zhang,Y., Roccaro,A.M., Rombaoa,C., Flores,L., Obad,S., Fernandes,S.M., Sacco,A., Liu,Y., Ngo,H., Quang,P., *et al.* (2012) LNA-mediated anti-miR-155 silencing in low-grade B-cell lymphomas. *Blood*, **120**, 1678–1686.

28. Hummelshoj,L., Ryder,L.P., Madsen,H.O. and Poulsen,L.K. (2005) Locked nucleic acid inhibits amplification of contaminating DNA in real-time PCR. *Biotechniques*, **38**, 605–610.

29. Russell,C., Kerkof,K. and Timour,M. (2006) US20060234277A1 - Method for selectively blocking hemoglobin RNA amplification - Google Patents.

30. Consortium, exRNAQC, Anckaert,J., Cobos,F.A., Decock,A., Deleu,J., Wever,O. De, Wilde,J. De, Dhondt,B., D'huyvetter,T., Everaert,C., *et al.* (2021) Performance of RNA purification kits and blood collection tubes in the Extracellular RNA Quality Control (exRNAQC) study. *bioRxiv*, 10.1101/2021.05.11.442610.

31. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10.

32. FASTX-Toolkit.

33. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol. 2009 103*, **10**, 1–10.

34. R Core Team (2021) R: A language and environment for statistical computing. *R Found. Stat. Comput. Vienna, Austria*.

35. Wickham,H., Averick,M., Bryan,J., Chang,W., D',L., Mcgowan,A., François,R., Grolemund,G., Hayes,A., Henry,L., *et al.* (2019) Welcome to the Tidyverse. *J. Open Source Softw.*, **4**, 1686.

36. Durinck,S., Moreau,Y., Kasprzyk,A., Davis,S., De Moor,B., Brazma,A. and Huber,W. (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.

37. Durinck,S., Spellman,P.T., Birney,E. and Huber,W. (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc. 2009 48*, **4**, 1184–1191.

38. Robinson,D. (2014) broom: An R Package for Converting Statistical Analysis Objects Into Tidy Data Frames.

39. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47–e47.

40. Robinson,J.T., Thorvaldsdóttir,H., Winckler,W., Guttman,M., Lander,E.S., Getz,G. and Mesirov,J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol. 2011 291*, **29**, 24–26.

41. Anders,S., Pyl,P.T. and Huber,W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166.

42. Wick,R.R., Judd,L.M. and Holt,K.E. (2019) Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.*, **20**.

43. nanoporetech/pychopper: A tool to identify, orient, trim and rescue full length cDNA reads.

44. Li,H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100.

45. Danecek,P., Bonfield,J.K., Liddle,J., Marshall,J., Ohan,V., Pollard,M.O., Whitwham,A., Keane,T., McCarthy,S.A., Davies,R.M., *et al.* (2021) Twelve years of SAMtools and BCFtools. *Gigascience*, **10**, 1–4.

46. De Coster,W., D'Hert,S., Schultz,D.T., Cruts,M. and Van Broeckhoven,C. (2018) NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*, **34**, 2666–2669.

47. Hao,Y., Hao,S., Andersen-Nissen,E., Mauck,W.M., Zheng,S., Butler,A., Lee,M.J., Wilk,A.J., Darby,C., Zager,M., *et al.* (2021) Integrated analysis of multimodal single-cell data. *Cell*, **184**, 3573-3587.e29.

48. Field,L.A., Jordan,R.M., Hadix,J.A., Dunn,M.A., Shriver,C.D., Ellsworth,R.E. and Ellsworth,D.L. (2007) Functional identity of genes detectable in expression profiling assays following globin mRNA reduction of peripheral blood samples. *Clin. Biochem.*, **40**, 499–502.

49. Hulstaert,E., Morlion,A., Avila Cobos,F., Verniers,K., Nuytens,J., Vanden Eynde,E., Yigit,N., Anckaert,J., Geerts,A., Hindryckx,P., *et al.* (2020) Charting Extracellular Transcriptomes in The Human Biofluid RNA Atlas. *Cell Rep.*, **33**, 108552.

TABLE AND FIGURES LEGENDS

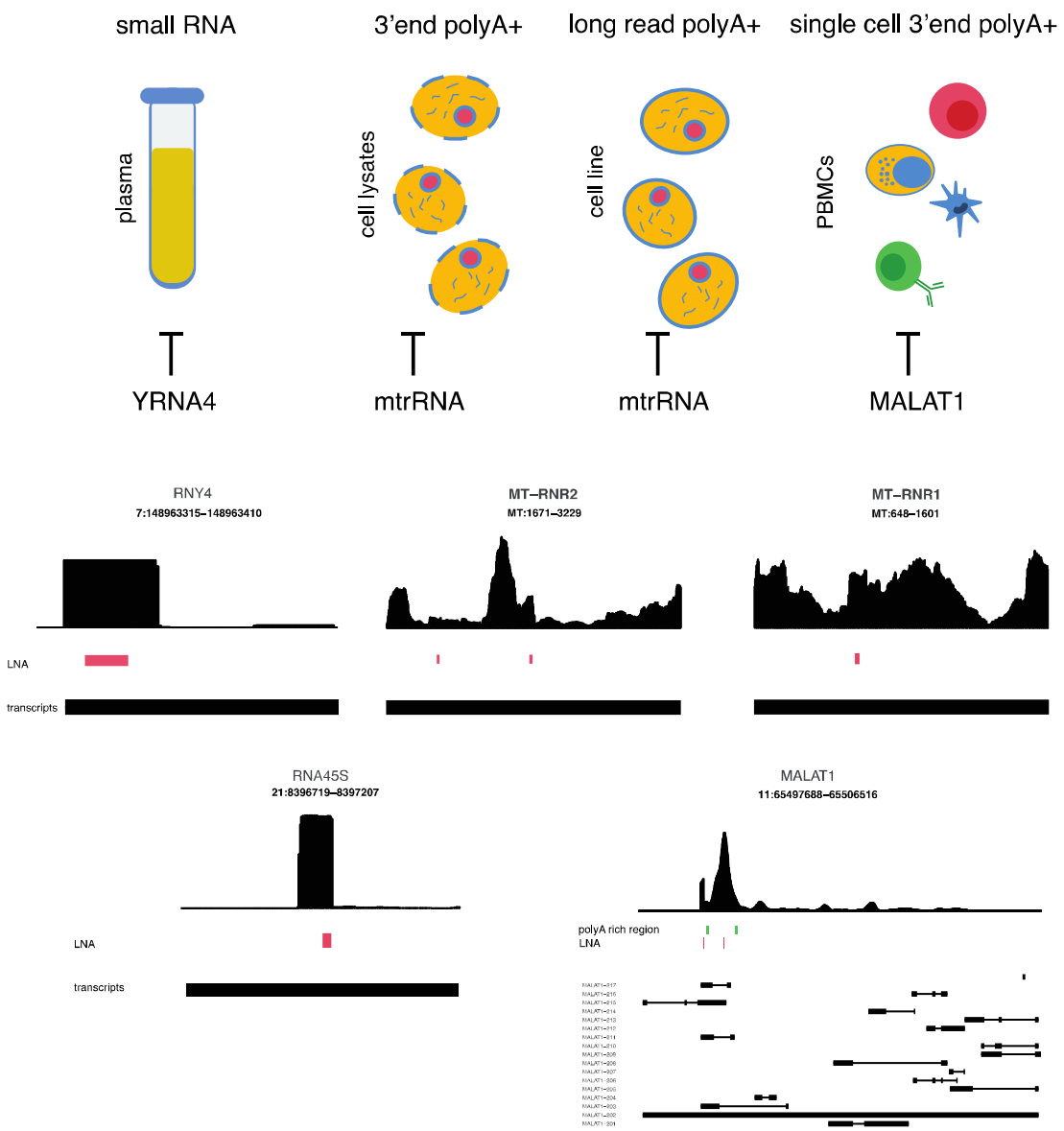

Figure 1. **Overview of targeted transcripts and investigated library preparation methods.** A. Visual representation of each of the performed RNA-sequencing library preparation methods with the associated blocked transcripts. B-F. Coverage plots of non-depleted sequencing data for the LNA targeted region. The height of the bars represents the number of reads mapping to that position in the data. The LNA and its binding location is shown in red/green under the coverage plot. The transcripts of the targeted genes are indicated at the bottom of each plot.

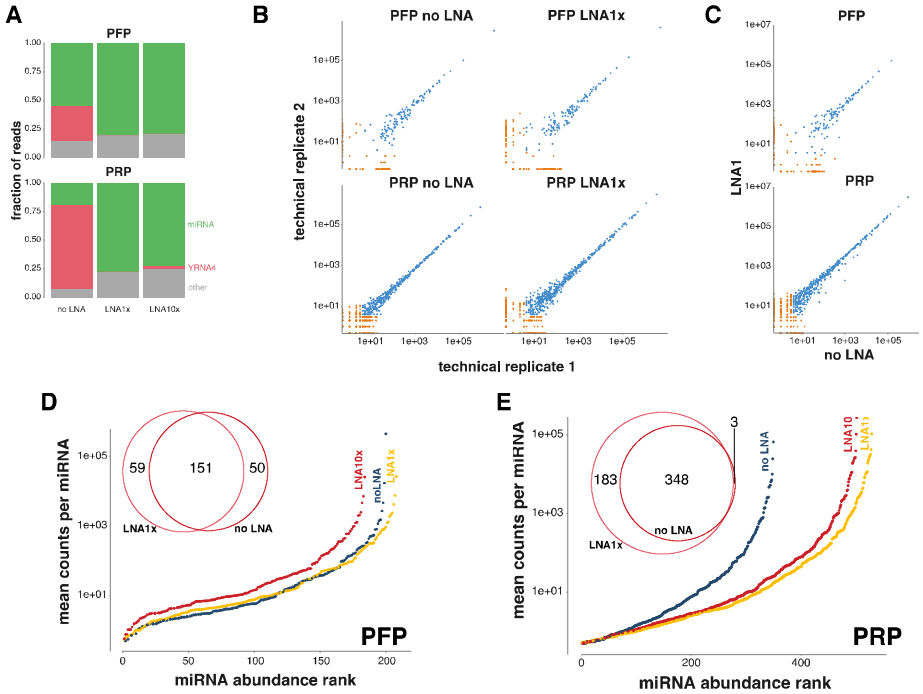

Figure 2. **LNA blocking during TruSeq Small RNA sequencing of plasma samples is efficient, increases the informative read percentages and results in more detected miRNAs and a higher coverage**. A. Read percentages for the YRNA and miRNA biotype without the LNA (noLNA) blocking and by adding the LNA in two different concentrations (LNA1x, LNA10x) to both platelet rich (PRP) and platelet free (PFP) plasma. B. Correlations of the technical replicates are equally high as C. the correlations between blocking and not blocking. D & E. The mean CPM between the technical replicates for all detected (CPM > 0.5) miRNAs and overlap between the detected miRNAs in the LNA blocking (LNA1x) and no blocking condition (noLNA) for both PFP and PRP.

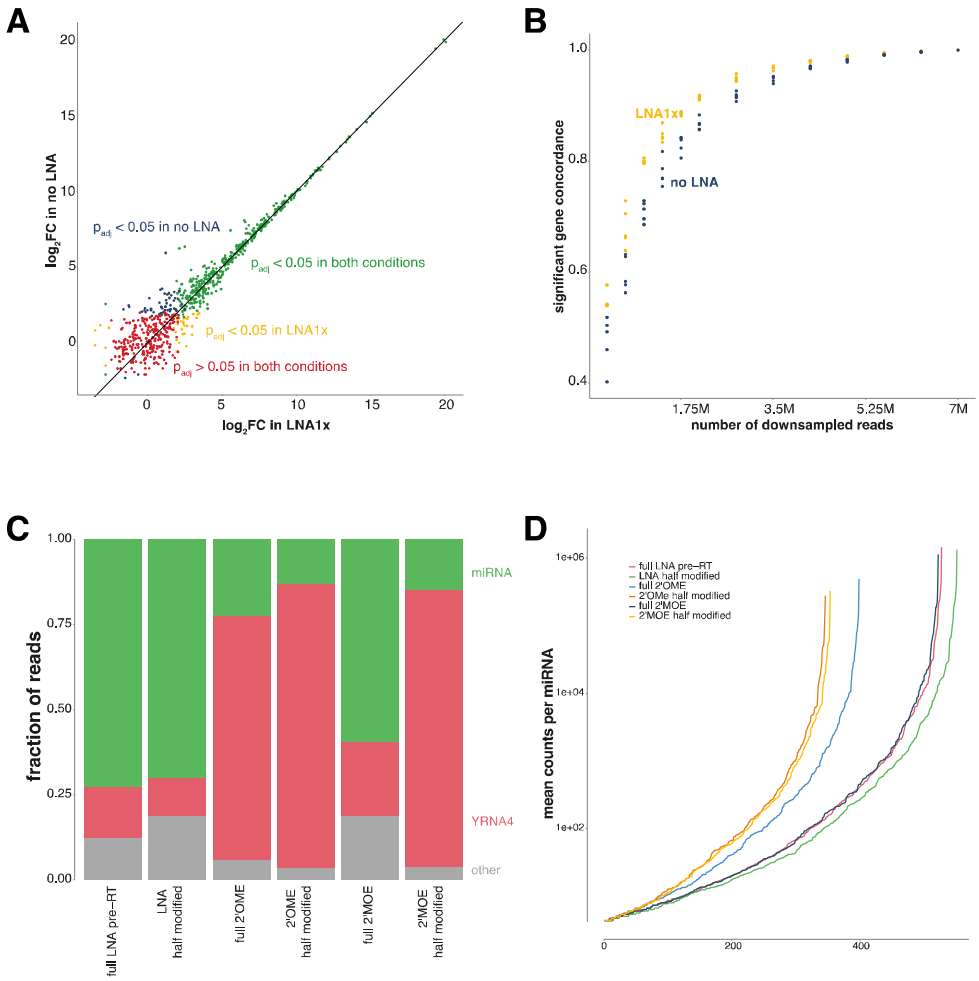

Figure 3**. At a low read depth, the detection of significant differential genes benefits from the blocking and LNA modifications are more efficient compared to 2'MOE and 2'OMe modifications.** A. Fold changes between conditions are preserved when blocking the RNY4 fragment. B. The significant gene concordance between subsamples is higher for the LNA RNY4 blocking compared to the original protocol, especially at a lower sequencing depth. C. Read percentages for the YRNA and miRNA biotype for multiple modifications. D. The number of discovered miRNAs and their coverage for the various modifications.

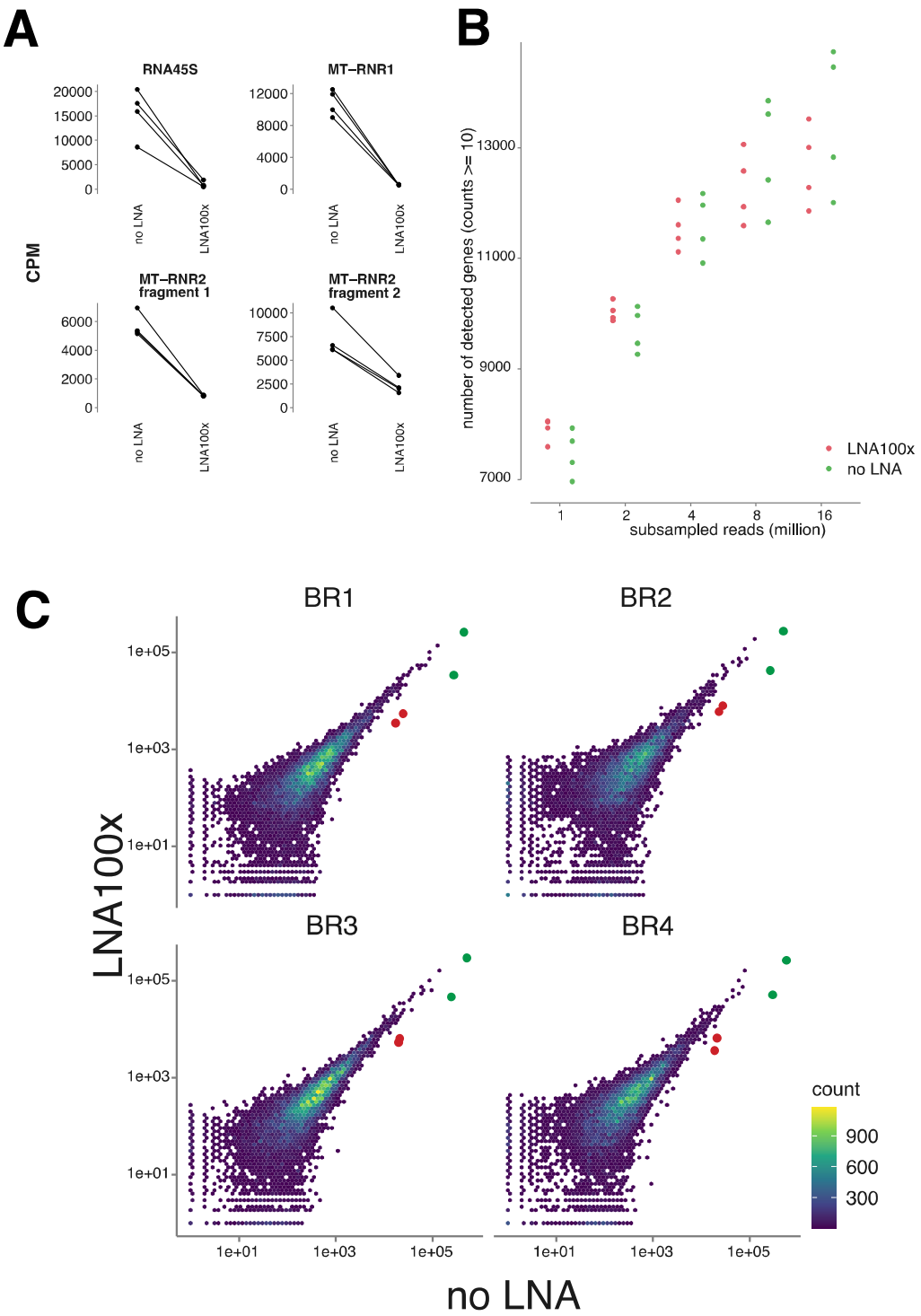

Figure 4. **LNA blocking of mtrRNA fragments (mitochondrially encoded 16S rRNA (two fragments), 45S pre−ribosomal RNA and mitochondrially encoded 12S rRNA) in QuantSeq 3'**

end sequencing. A. Counts per million (CPM) of the targeted transcripts for untreated ('noLNA') and treated ('LNA100x') samples after QuantSeq 3' end sequencing. Each dot is a sample and is colored by the sgRNA used for knock-down. Lines are drawn between samples originating from the same lysate. B. Average number of detected genes for subsampled treated and untreated samples. A gene is 'detected' once it has at least 10 counts. Each point is a biological replicate and is colored by treatment. C. Scatter plot between treated and untreated samples for each biological replicate. Each dot corresponds to a gene. The green dots are MT-RNR1 and MT-RNR2, while the red dots indicate the genes likely affected by off-target binding: H4C3 and MT-ATP8.

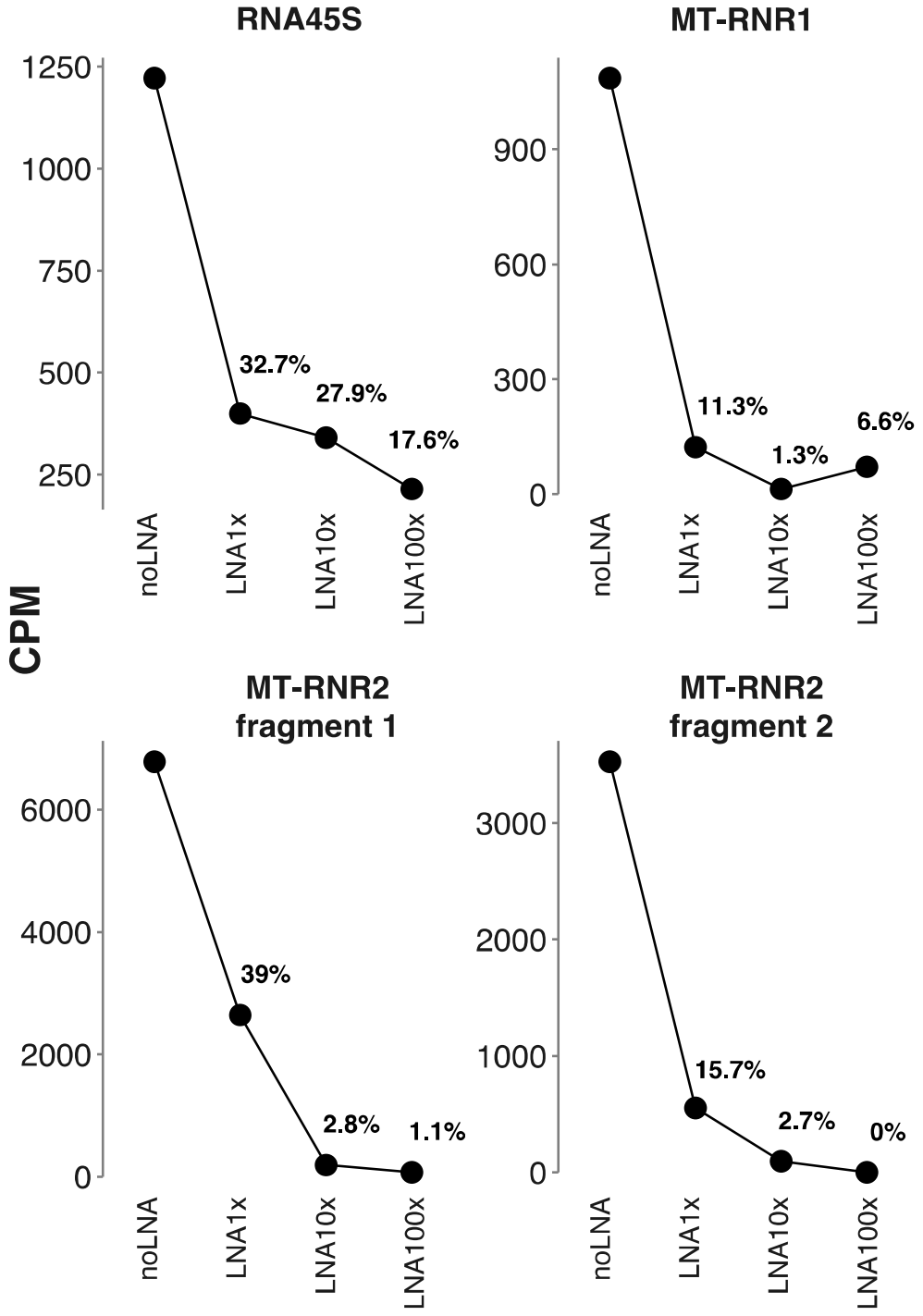

Figure 5. **LNA blocking of mtrRNA fragments (mitochondrially encoded 16S rRNA (two fragments), 45S pre−ribosomal RNA and mitochondrially encoded 12S rRNA) in Oxford Nanopore direct-cDNA sequencing**. Counts per million (CPM) of the targeted transcripts for

untreated ('noLNA') and treated ('LNA1x', 'LNA10x' and 'LNA100x') samples after Oxford Nanopore direct-cDNA sequencing. Fractions of CPM for treated samples relative to untreated samples are printed as percentages.

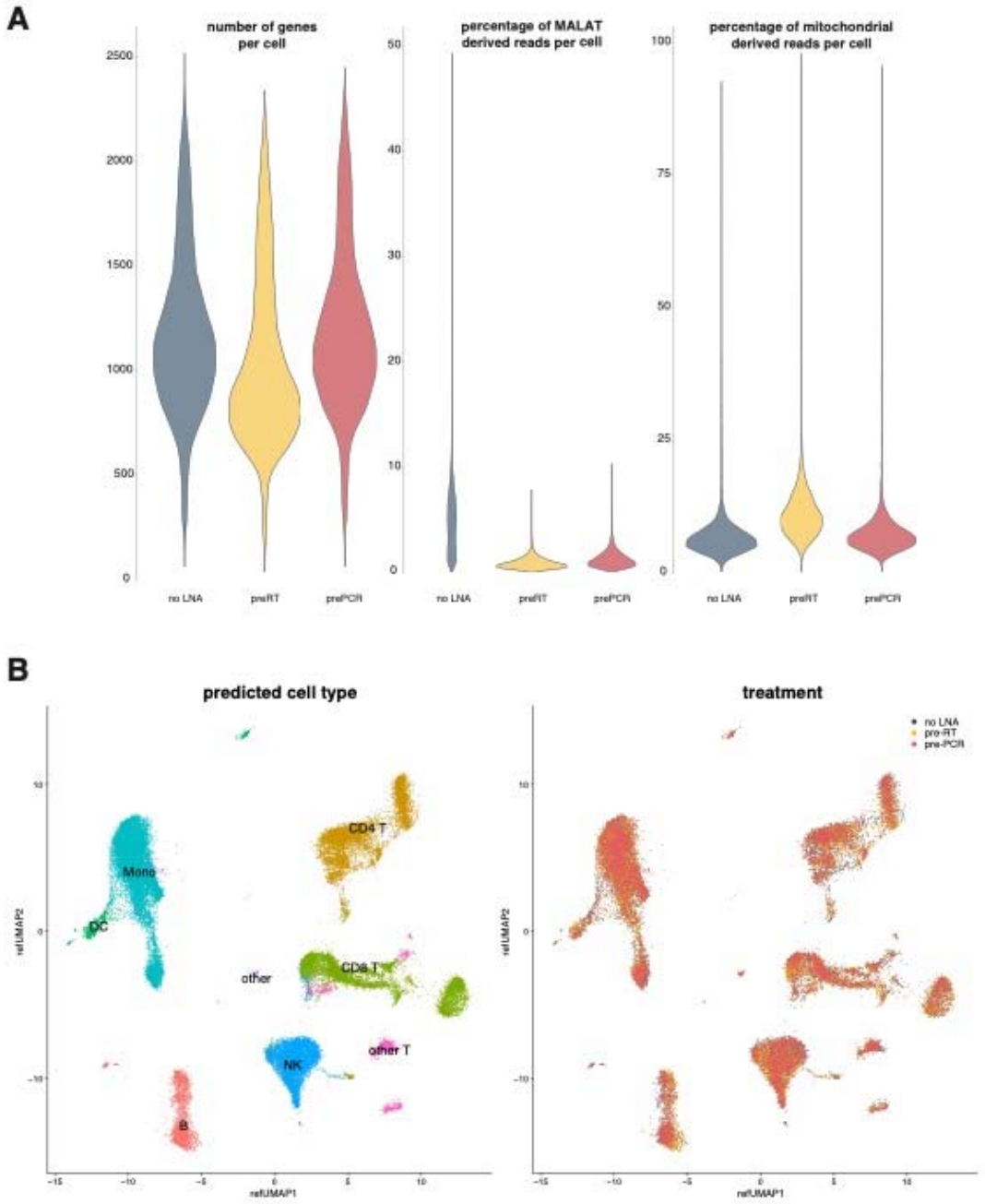

Figure 6. **MALAT blocking prePCR genes in 10X Chromium 3' end sequencing experiments is efficient and does not impact other genes.** A. Quality metric distributions for the standard protocol, preRT and prePCR blocking, indicating a slightly higher number of detected genes in prePCR

blocking, lower MALAT percentages in both preRT and prePCR blocking and higher mitochondrial RNA (a cell death proxy) in preRT blocking. B. Highly similar UMAP representations after integration of the samples are obtained for all protocols.