SUPPLEMENTAL INFORMATION

Supplemental Methods

RNA quantification and quality control

Total exRNA from selected samples was quantified using the Qubit RNA HS Assay Kit (ThermoFisher, cat# Q32852) on a Qubit 2.0 Fluorometer (ThermoFisher) per manufacturer's instructions (199 µl working solution for 1 µl sample or standard) but using 0.5 OD_{260} of a 2'-O-methylated let-7a RNA as a standard to calculate the concentration based on the following formula.

$$c_{sample}(ng/\mu l) = \frac{RFU(sample) - RFU(reagent\ blank)}{V\ (\mu l) \bullet 7/ng}$$

Where RFU are the relative fluorescence units reported by the device. The RFU of the blank samples was averaged across 2-3 readings; usual background readings were 30-35 RFUs.

Total RNA from cells and tissues was quantified on a NanoDrop UV spectrophotometer, and the RNA integrity was determined on an Agilent Bioanalyzer 2100 with Agilent RNA 6000 Pico Chips.

Small RNA-sequencing

We performed barcoded small RNA-sequencing (sRNA-seq) as published previously (1) but with additional barcoded 3'-adapters to allow multiplexing of up to 24 samples; the detailed number of samples per library and the number used in this study is indicated in Supplemental Data 1. The cDNA library preparation started with a 3' ligation of a pre-adenylated DNA oligonucleotide for each individual sample. The 3'-adapter master mix contained a cocktail of 10

equimolar 22-nt synthetic RNAs (Supplemental Data 13, cocktail 2). The samples were pooled after the 3' adapter ligation, size selected for small RNAs but the upper limit was chosen to be 45-nt instead of 24-nt used for classic miRNA-seq to allow longer mRNA fragments to be included; that is the size selection was 19- to 45-nt. RNA ligated to the 3' adapter was gel purified, followed by 5' ligation of an RNA oligonucleotide and another size selection and gel purification. The cDNA library preparation was completed by second strand synthesis using SuperScript III, and PCR amplification. The resulting amplicon was quality controlled on an Agilent TapeStation with a High Sensitivity D1000 ScreenTape and used for cluster generation and sequencing 50-bp single-read on a HiSeq 2500 sequencer in rapid run mode.

RNA-sequencing of cell and tissue samples

RNA-seq tissue and cell profiles were generated de novo or downloaded from the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra). Please see Supplemental Table for details.

De novo RNA-seq libraries were prepared by the Genomics Core Facility of The Rockefeller University using the Illumina TruSeq Stranded mRNA LT protocol according to the manufacturer's protocol but using NEB's Protoscript II reverse transcriptase for the first-strand cDNA synthesis. RNA input was 400 ng total RNA. Individual RNAseq libraries were quality controlled on an Agilent TapeStation with a High Sensitivity D1000 ScreenTape. The libraries were sequenced on an Illumina NextSeq 500 sequencer 75-bp paired-end in high-output mode in the Genomics Core Facility of The Rockefeller University.

Bioinformatics analysis

FASTQ data file generation from de novo sequenced RNA-seq libraries

Raw image data were converted to FASTQ files by the Genomics Core Facility of The Rockefeller University using Illumina's bcl2fastq conversion software version 2.18.0.12 (NextSeq 500) or version 1.8.4 (HiSeq 2500).

sRNA-seq read annotation

Short reads were annotated following a hierarchical annotation process as described(2, 3). FASTQ files were collapsed and demultiplexed using custom Perl scripts, followed by alignment with the BWA aligner version 0.7.9a (4) to a combined transcriptome reference followed by mapping of the residual unmapped reads to a combined genome reference.

Mapping parameters were set to allow for a maximum of 2 mismatches and reporting up to 999 alignments.

bwa aln -n 2 -t 16 -i 2 -R 1000 <transcriptome_reference.fa> \ <input_unique_reads>.fa> <input_unique.sai> 2>>log.txt

After the initial read alignment the BAM file was converted to a custom text (hits) file and the final annotation category was assigned in a hierarchical manner(2, 3) using custom Perl scripts (available upon request) to give RNA transcript categories with higher abundance or higher likelihood of being the true category precedence over less abundant transcripts.

bwa samse -n 1000 <transcriptome_reference.fa> <input_unique.sai> <input_unique_reads>.fa> 2>> log.txt | <custom_script.pl> <RNA_type_definition> <RNA_hierarcy> <output_directory> 2> <output.hits> 2>>log.txt

The transcriptome reference (available upon request) consisted of spike-in sequences for oligonucleotides (Supplemental Data 13), size marker oligonucleotides (5), and adapter sequences used during the cDNA library preparation (5), sequences from expression plasmids for recombinant enzymes used during cDNA library preparation (pET16b-Rnl2(1-249)K227Q), AddGene; pET28a), and curated miRNA(6), tRNA (7), and piRNA (8) references. Small nuclear (snRNAs) and small nucleolar (sno) RNAs (7SK, U1, U2, U4, U5, U6, U7; snoRNAs), small cytoplasmic (sc) and small Cajal-bodies (sca) RNAs (7SL, RNY1, RNY3, RNY4, RNY5, VTRNA1, VTRNA2), lincRNAs (MALAT, XIST, NEAT1, 11orf95, FAM157A, FAM157B, FAM91A2, FTX, MEG8, SNHG15), rRNAs (28S, 18S, 5.8S, 5S), and ciRS7 circular RNA were obtained from the UCSC database, release hg38. The diatom (accession numbers: AM236073.1, KM675692.1, KM816803.1) and hepatitis B (AB981582.1) and hepatis C (accession numbers: NC_004102, NC_009823, NC_009824, NC_009825, NC_009826, NC_009827) references were obtained from GenBank.

The genomic references consisted of the human genome reference build 38 (primary assembly including random and unplaced contigs but without haplotypic regions) and the E. coli K-12 genome.

The read annotation hierarchy was as follows (with descending priority):

CalibratorIllumina *(not used for this study)*
CalibratorTuschlSet1 *(cocktail 1 in sRNA-seq library preparation and Supplemental Data 13)*
CalibratorTuschlSet2 *(cocktail 2 in sRNA-seq library preparation and Supplemental Data 13)*
CalibratorTuschlSet3 *(not used for this study)*
CalibratorTuschlSet4 *(not used for this study)*
Adapter
SizeMarker
Plasmid
EColiK12
Diatom
rRNA
MTrRNA
tRNA
MTtRNA
snRNA
snoRNA
MTmRNA
scRNA
scaRNA
miRNA
mRNA
circRNA
lincRNA
rRNAPrecursor
tRNAPrecursor
snRNAPrecursor
snoRNAPrecursor
scRNAPrecursor
scaRNAPrecursor
piRNA
miRNALowConfidence
snoRNALowConfidence
MTGenome
Genome
Viral

Thus, synthetic RNAs like added calibrator oligonucleotides, barcoded adapters, and size marker RNAs are getting highest priority. Among the endogenous RNAs those with higher transcript abundance are more highly ranked than those with less abundance. This approach will assign a

read that is mapping to an rRNA transcript with 2 mismatches and at the same time to an mRNA transcript without mismatch to the rRNA transcript.

Metagene analysis of exRNA reads

The Bowtie (version 1.2.1) options used were:

bowtie -v 2 -a -f <transcriptome_reference> <input.fasta>

Where <transcriptome_reference> represented 5'UTRs, CDS, and 3'UTRs. Only UTRs longer than 15-nt were considered. The transcriptome reference was based on Ensembl release 67 with additional manual curation (manuscript in preparation). A Perl script was employed to count the number of reads matching to UTR's and CDS. If a read matched to several categories its count was split equally across the categories.

Calculating transcript coverage by ex-mRNA fragments (Figures 2A, B; Figure 5E)

Plots showing transcript coverage for selected transcripts separate for each treatment group (untreated RNA samples, PNK-treated RNA samples) and each of the four sample types (serum, EDTA plasma, ACD plasma, heparin plasma) were generated in the R language using the Bioconductor packages GenomicAlignments (1.17.3) and ggbio (1.29.5). BAM files of exRNA alignments were filtered for mRNA reads 15-nt or longer, mapping perfectly to the transcript of interest and with at most 2 mapping locations. Exon intron boundaries were as defined in

Ensembl transcripts ENST00000251595.6 for HBA2, ENST00000295897.4 for ALB, and ENST00000368733.3 for S100A8.

BAM files used for Figure 5E (S100A8 alignments) were down-sampled to 600,000 reads using samtools (version 1.2).

```
samtools view –hs 21.fraction <in.bam> –o <subsampled.bam>
```

Where *fraction* represented the to be subsampled of the original BAM file to be subsampled, e.g. to sample 600,000 reads from a file starting with 6,000,000 alignments: 21.10.

Tissue and cell RNA-seq analysis

The FASTQ files from polyA and total RNA RNA-seq reads from tissues or cells were aligned to the human genome build 38 using the STAR aligner version 2.0.4j (9). The general format of the STAR command was:

```
STAR --genomeDir <genome_index> --readFilesIn <infile.fastq.gz> \
   --clip3pAdapterSeq <adapter_seq> \
   --outFileNamePrefix <outname> --runThreadN 8 \
   --outFilterMismatchNmax 2 \
   --readFilesCommand gunzip -c --outSAMunmapped Within \
   --outSAMtype BAM Unsorted --outStd BAM_Unsorted | \
   samtools sort -m 1G -@ 8 - -o <outname.sorted.bam>
```

A gene expression matrix was generated using the featureCounts version 1.5.1 (10) program based on Ensembl release 88 counting multimapping reads fractionally. The generic featureCounts command was:

```
featureCounts -M --fraction -a Homo_sapiens.GRCh38.88.gtf
```

Supplemental References

1. Hafner, M, Renwick, N, Farazi, TA, Mihailovic, A, Pena, JT, Tuschl, T. Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. *Methods*. 2012;58(2):164–170.
2. Brown, M, Suryawanshi, H, Hafner, M, Farazi, TA, Tuschl, T. Mammalian miRNA curation through next-generation sequencing. *Front Genet*. 2013;4:145.
3. Farazi, TA et al. Bioinformatic analysis of barcoded cDNA libraries for small RNA profiling by next-generation sequencing. *Methods*. 2012;58(2):171–187.
4. Li, H, Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
5. Max, KEA et al. Human plasma and serum extracellular small RNA reference profiles and their clinical utility. *Proc Natl Acad Sci U S A*. 2018;115(23):E5334–E5343.
6. Landgraf, P et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*. 2007;129(7):1401–1414.
7. Gogakos, T, Brown, M, Garzia, A, Meyer, C, Hafner, M, Tuschl, T. Characterizing Expression and Processing of Precursor and Mature Human tRNAs by Hydro-tRNAseq and PAR-CLIP. *Cell Rep*. 2017;20(6):1463–1475.
8. Williams, Z et al. Discovery and Characterization of piRNAs in the Human Fetal Ovary. *Cell Rep*. 2015;13(4):854–863.

9. Dobin, A et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
10. Liao, Y, Smyth, GK, Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–930.
11. Garzia, A et al. The E3 ubiquitin ligase and RNA-binding protein ZNF598 orchestrates ribosome quality control of premature polyadenylated mRNAs. *Nat Commun*. 2017

Supplemental Figure Legends

Supplemental Figure 1. exRNA sample characteristics for libraries 1-4. (A) Quantification of total exRNA from 450 µl sample from healthy control donors used for libraries 1-4. Samples were quantified using the Qubit RNA HS Assay as described in the supplemental methods. Only samples with residual sample volume left after sRNA-seq cDNA library preparation were measured. Differences were tested with a linear-fixed-effects model and Tukey's post-hoc test in the R language. It should be noted that the T4 PNK treatment started with an aliquot of the untreated RNA. Box plots show the median and first and third quartiles (lower and upper hinges). Whiskers extend at most 1.5 x interquartile range from the hinges, any data outside this is shown as individual outlier points. (B) Read length distribution of uniquely mapping and multi-mapping mRNAs reads across different blood sample types. Included were only reads without mismatch to the reference transcriptome.

Supplemental Figure 2. Read coverage of ex-tRNAs. (A) Read coverage of the Gly-tRNA isoacceptor from the untreated heparin sample of donor "Control 1" (sample "Control1_heparin_untreated" in Supplemental Data 1) together with the tRNA secondary

structure (box). The anticodon is highlighted in yellow and bases protected by the RNA-binding protein ZNF598 based on PAR-CLIP(11) are colored blue. The pileup was scaled to 11 bins using the following formula for each base: (((count/maximum count) ** 1/4) + 0.5)/0.1. (B) Showing the 10 most frequently sequenced sequences (corresponding to the 5'-end) with corresponding read count.

Supplemental Figure 3. Differences in the ex- miRNA and ex-mRNA profile and read coverage by blood sample type. (A) Differences in abundance of cell-free, circulating miRNAs (top row) and mRNAs (bottom row); see Supplemental Data 14 for hypothesis tests. (B) t-SNE plot based on the exRNA profiles of all RNA categories of the four different sample types from six healthy donors (perplexity 4). (C) Read coverage for the small nuclear RNAs (snRNAs) U1 (left), and U2 (right) in the different sample types. Rectangle in the left sub-figure indicates coverage loss of U1 region in EDTA and ACD plasma. Bar (core) on the left (U1) and right (U2) indicates core region resistant to nuclease digestion.

Supplemental Figure 4. Capture of top expressed transcripts from selected solid tissues in different sample types. The 1,000 most abundant nuclear mRNA transcripts from the selected cell types and tissues from bulk RNA-seq that collected 5 unique or 10 total ex-mRNA reads in at least 3 of the 6 donors per sample type were considered captured. The x out of 1,000 captured transcripts (x axis) were ordered in descending order by the tissue specificity score (TSS, y axis). Transcripts with a TSS greater than 3 were highlighted in red and listed space permitting. Note that bulk sequencing data of heterogenous and well perfused tissues like lung will contain hematopoietic or erythropoietic/hematopoietic-specific transcripts within the top 1,000 genes of

the sample. It should also be noted that most reads for e.g. the MYBPC3 (left lower panel) and all for MIOX (kidney panels) were shorter than 17-nt with a high likelihood of misassignments.

Supplemental Figure 5. Treatment of total extracellular RNA with T4 polynucleotide kinase (T4 PNK) followed by small RNA-sequencing (sRNA-seq) in a pilot cohort of chest pain patients (ACS) and matched controls. This figure is analogous to Figure 1B, C and Supplemental Figure 1 showing basic read and annotation characteristics (A, B) as well as the effect of PNK end-treatment on mRNA capture (C) for libraries 5 and 6, and the library preparation strategy for libraries 5 and 6. Matching untreated and treated RNA samples from each donor were ligated to the same barcoded 3'-adapter. See Supplemental Data 1 for details on all sub-samples and libraries.. (A) Differences in read annotation for endogenous RNA classes for untreated RNA and PNK-treated RNA using initial annotation settings (up to 2 mismatches, unlimited multi-mapping). (A) Read length distribution for reads annotated as mRNAs (nuclear and mitochondrial) with 0, 1, and 2 mismatches to the reference transcriptome; multi-mapping reads and reads mapping to only one transcript (uniquely mapping) are shown in different colors as indicated. Note that for the final mRNA downstream analysis only mRNA reads without mismatch (i.e. 0 mismatch), maximum 2 mapping locations, and 15-nt or longer were considered. (C) Differences in nuclear-encoded mRNA capture between untreated and PNK-treated RNA using strict (final) annotation criteria (no mismatch and up to two mapping locations).

Supplemental Figure 6. Unsupervised hierarchical clustering of RNA spike-in calibrators. Shown are spike-in calibrator RNAs profiles of samples from library 5 and 6. The calibrator

RNA (cocktail 2) was added at the 3'-adapter ligation step as described. Columns and rows were clustered using complete linkage and Euclidean distance (row dendrogram was removed. Clustering was done in the R language using the aheatmap package.

Supplemental Figure 7. Changes of the top 9 (by FDR) ex-mRNAs higher in ACS patients (n = 6) than controls (n = 10; library 6, PNK-treated). Box plots show the median and first and third quartiles (lower and upper hinges). Whiskers extend at most 1.5 x interquartile range from the hinges, any data outside this is shown as individual outlier points.

Supplemental Data

Supplemental Data 1. Master table in Excel spreadsheet format containing individual level sample information and read annotation as per initial annotation settings (reads 12-nt and longer, up to two mismatches, multi-mapping allowed) shown as read counts and percentages.

Supplemental Data 2. Effect of mismatches and multi-mapping on the percentage of reads 12-nt and longer annotated as mRNAs using initial mapping/annotation settings (up to two mismatches to reference, multi-mapping allowed).

Supplemental Data 3. Excel spreadsheet containing the results of the differential analysis of ex-miRNAs comparing the 4 different sample types used in libraries 1 to 4. Shown are raw and Benjamini-Hochberg corrected (FDR) P values for the overall/ANOVA-like analysis and Benjamini-Hochberg corrected (FDR) P values of the pairwise comparisons.

Supplemental Data 4. Excel spreadsheet containing the results of the differential analysis of ex-mRNAs comparing the 4 different sample types used in libraries 1 to 4. Only ex-mRNA fragments without mismatch, 15-nt or longer, and at maximum mapping to two transcripts of the transcriptome reference were allowed for this analysis. Shown are raw and Benjamini-Hochberg corrected (FDR) P values for the overall/ANOVA-like analysis and Benjamini-Hochberg corrected (FDR) P values of the pairwise comparisons.

Supplemental Data 5. Excel spreadsheet containing the results of the gene set analysis of mRNAs comparing the 4 different sample types used in libraries 1 to 4. C2 gene sets (from the MSigDB Collections of the Broad Institute) containing the terms "ribosome", "translation", or "inflammation" were used as input. A self-contained test as implemented in the mroast function of the Bioconductor package edgeR was used. The mRNA input count matrix and the design model were identical to the differential analysis (Dataset 3).

Supplemental Data 6. Excel spreadsheet containing the tissue and cell RNA-seq samples used to calculate the tissue-specific score (TSS). Details about the tissue and cells samples are given in sheet "README", and the gene expression matrix (transcripts per million, TPM) in sheet "Supplemental Data 6".

Supplemental Data 7. Excel spreadsheet containing the 169 mRNA genes (first column) with a tissue-specificity score > 3 detected in circulation in either serum or the plasma samples EDTA, ACD, or heparin. The other columns show the corresponding expression values in TPM (transcripts per million) of these genes in the selected tissues.

Supplemental Data 8. Percentage of the top 200, 500, 1,000 cellular transcripts by gene expression in the displayed cell and tissues that were captured as ex-mRNAs in each sample type if the mRNA transcript reference collected 5 or more unique reads (15-nt and longer, no mismatch, up to 2 mapping locations) in 3 or more of the 6 donors per sample type.

Supplemental Data 9. ACS and control group demographics. Basic demographics and blood parameters of 6 patients with acute coronary syndrome (ACS) and 10 matched controls used in libraries 5 (untreated RNA) and 6 (PNK-treated RNA). All 16 donors were male. P values were calculated using the Wilcoxon rank sum test.

Supplemental Data 10. Excel spreadsheet containing the individual level clinical chemistry and laboratory data for the six ACS patients and the 10 matched controls whose RNA was processed in sRNA-seq libraries 5 (untreated) and 6 (PNK-treated).

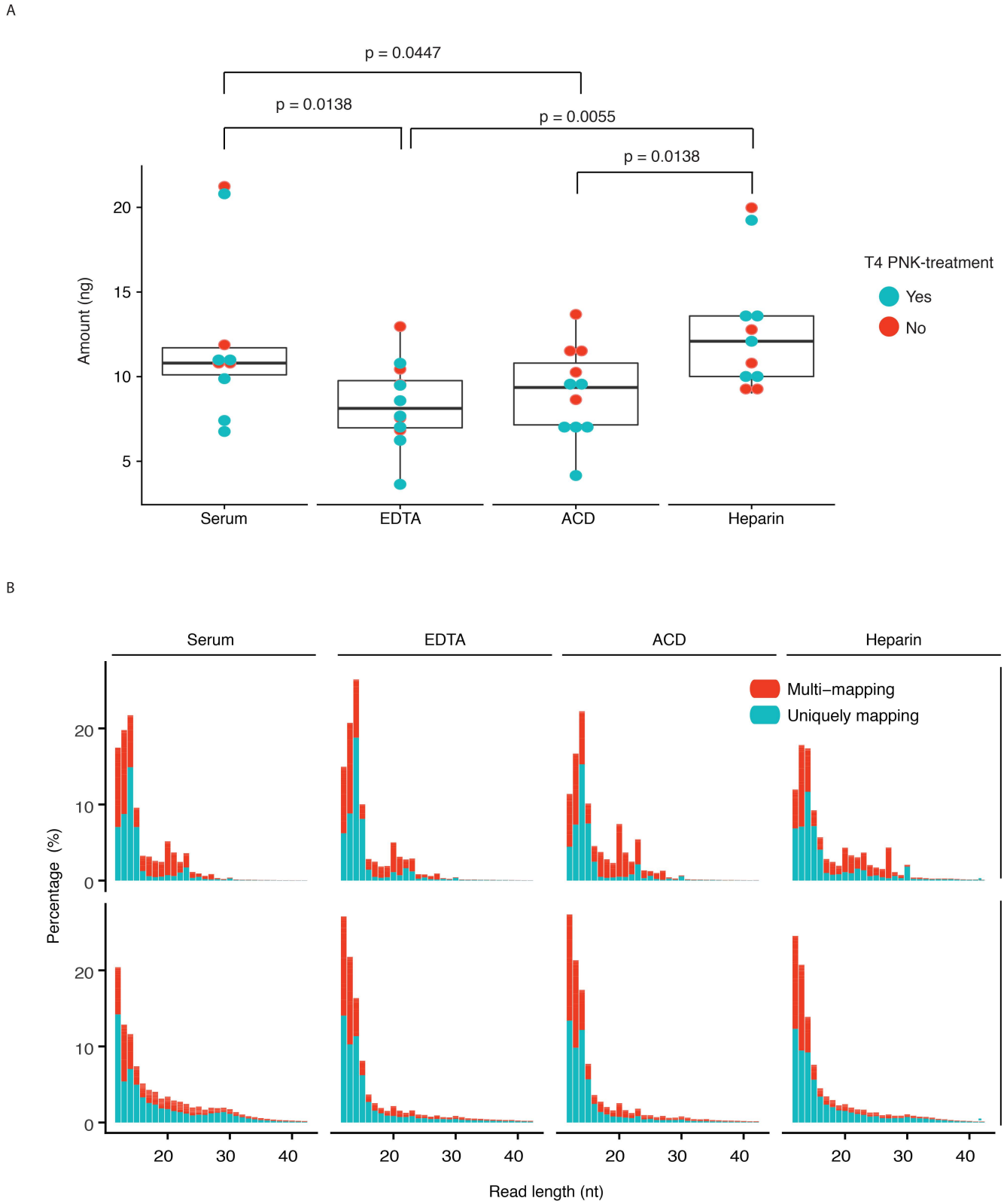
Supplemental Data 11. Excel spreadsheet containing the results of the differential analysis of ex-miRNAs (library 5 only) comparing the 6 ACS patients with age-matched controls.

Supplemental Data 12. Excel spreadsheet containing the results of the differential analysis of ex-mRNAs (library 6 only) comparing the 6 ACS patients with age-matched controls. Only mRNA fragments without mismatch, 15-nt or longer, and at maximum mapping to two transcripts of the transcriptome reference were allowed for this analysis. Shown are raw and Benjamini-Hochberg corrected (FDR) P values for the overall/ANOVA-like analysis (any difference among the samples?) and Benjamini-Hochberg corrected (FDR) P values of the pairwise comparisons.

Supplemental Data 13. Sequences of the synthetic RNA cocktails used as spike-ins for the RNA isolation (cocktail 1), and the cDNA library generation (cocktail 2).

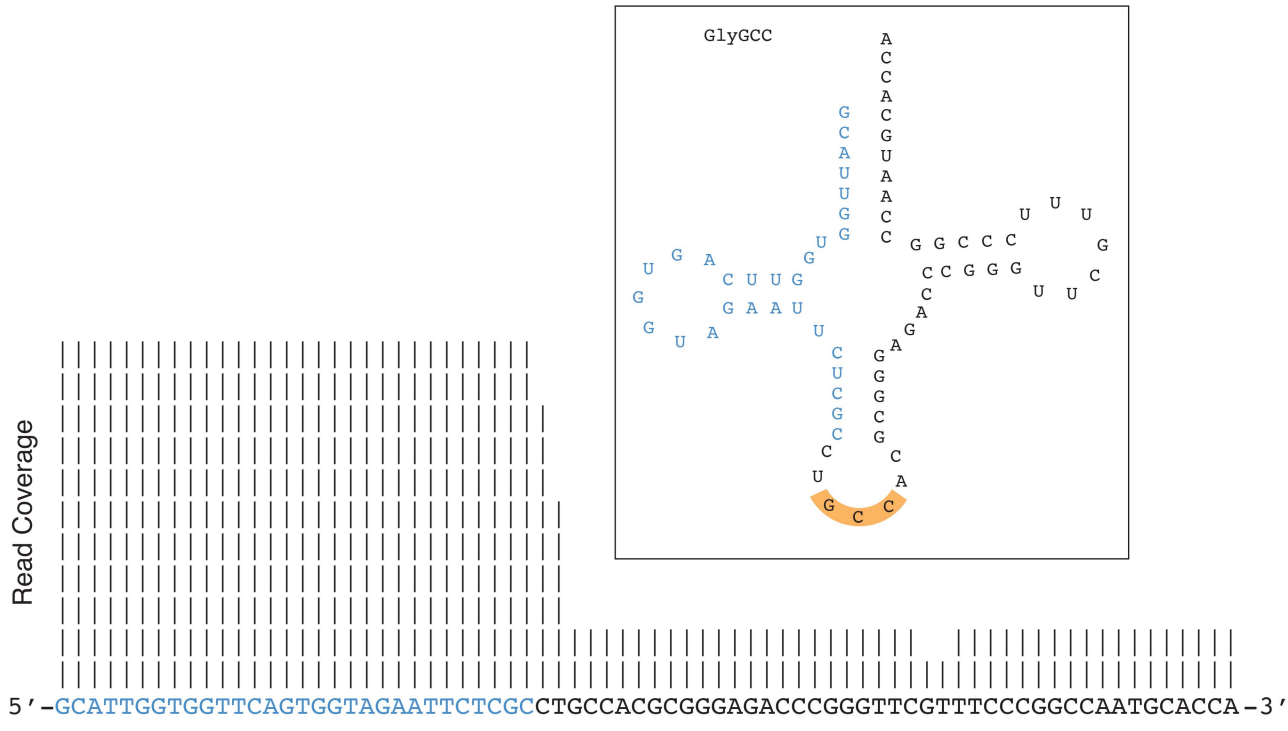
Supplemental Data 14. Benjamini-Hochberg corrected P values for pairwise comparisons as shown in Supplemental Figure 3A.

Supplemental Figure 1



Supplemental Figure 2

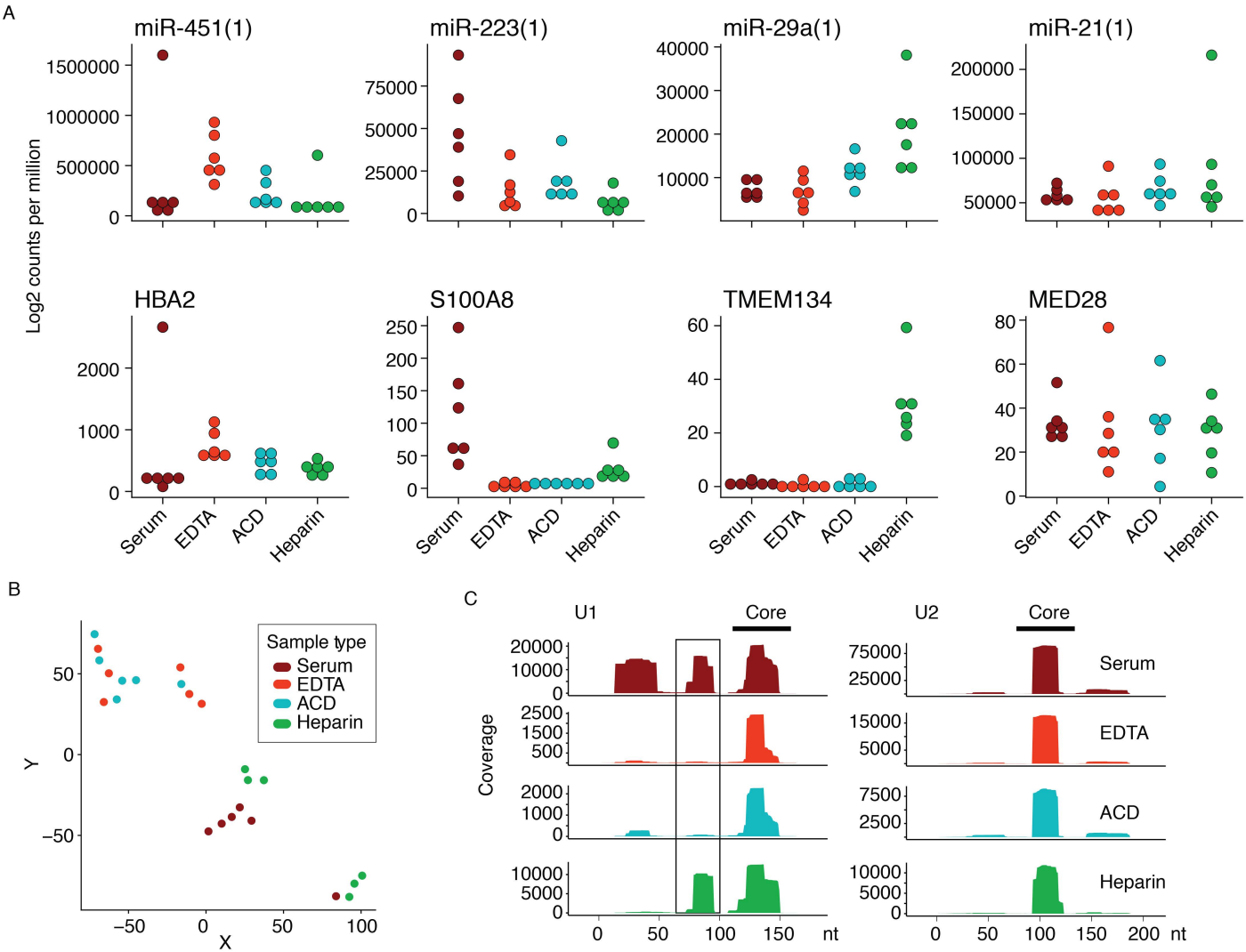
A



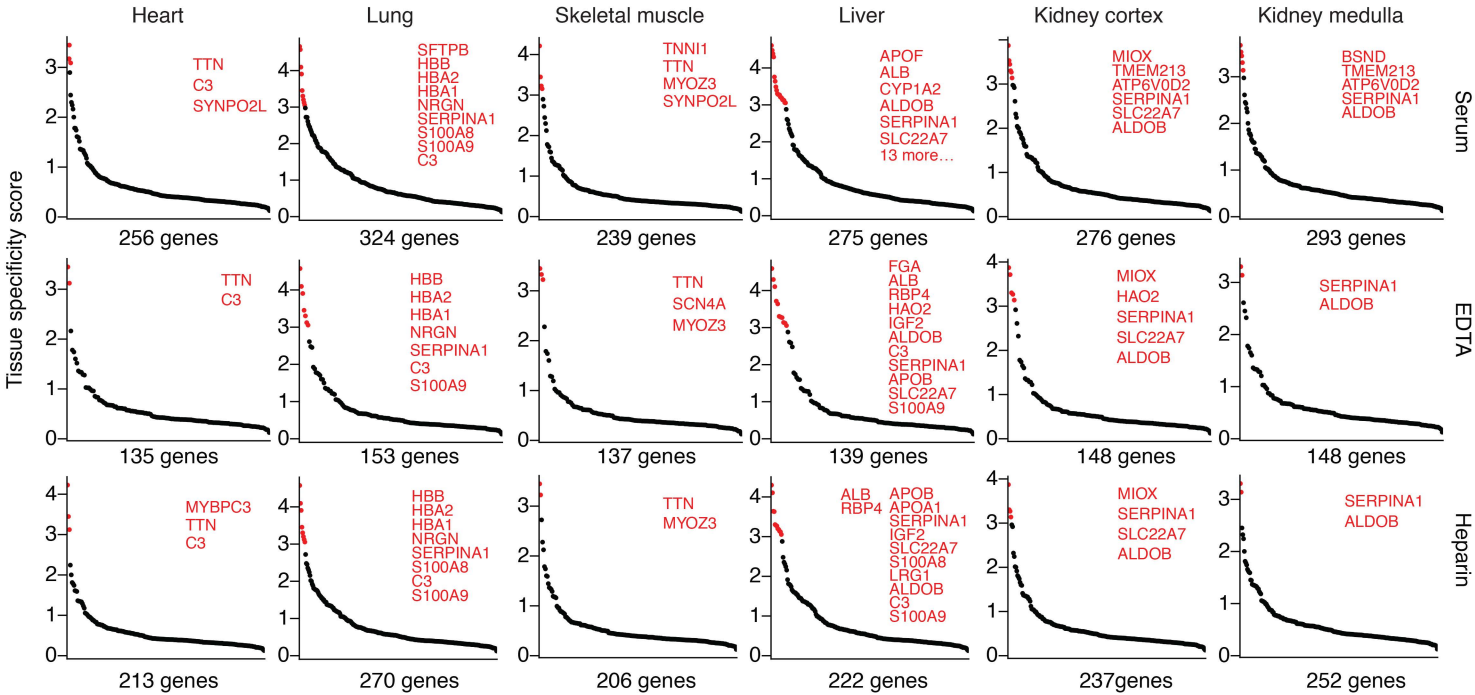
B

Sequence	Count
GCATTGGTGGTTCAGTGGTAGAATTCTCGC	914725
GCATTGGTGGTTCAGTGGTAGAATTCTCGCC	610267
GCATTGGTGGTTCAGTGGTAGAATTCTCGCCT	157870
GCATTGGTGGTTCAGTGGTAGAATTCTC	83456
GCATTGGTGGTTCAGTGGTAGAATTCTCG	75785
tCATTGGTGGTTCAGTGGTAGAATTCTCGC	26911
tCATTGGTGGTTCAGTGGTAGAATTCTCGCC	17778
TTCAGTGGTAGAATTCTCGC	12201
GGTTCAGTGGTAGAATTCTCGC	8976
GCATTGGTGGTTCAGTGGTAGAATTCT	8415

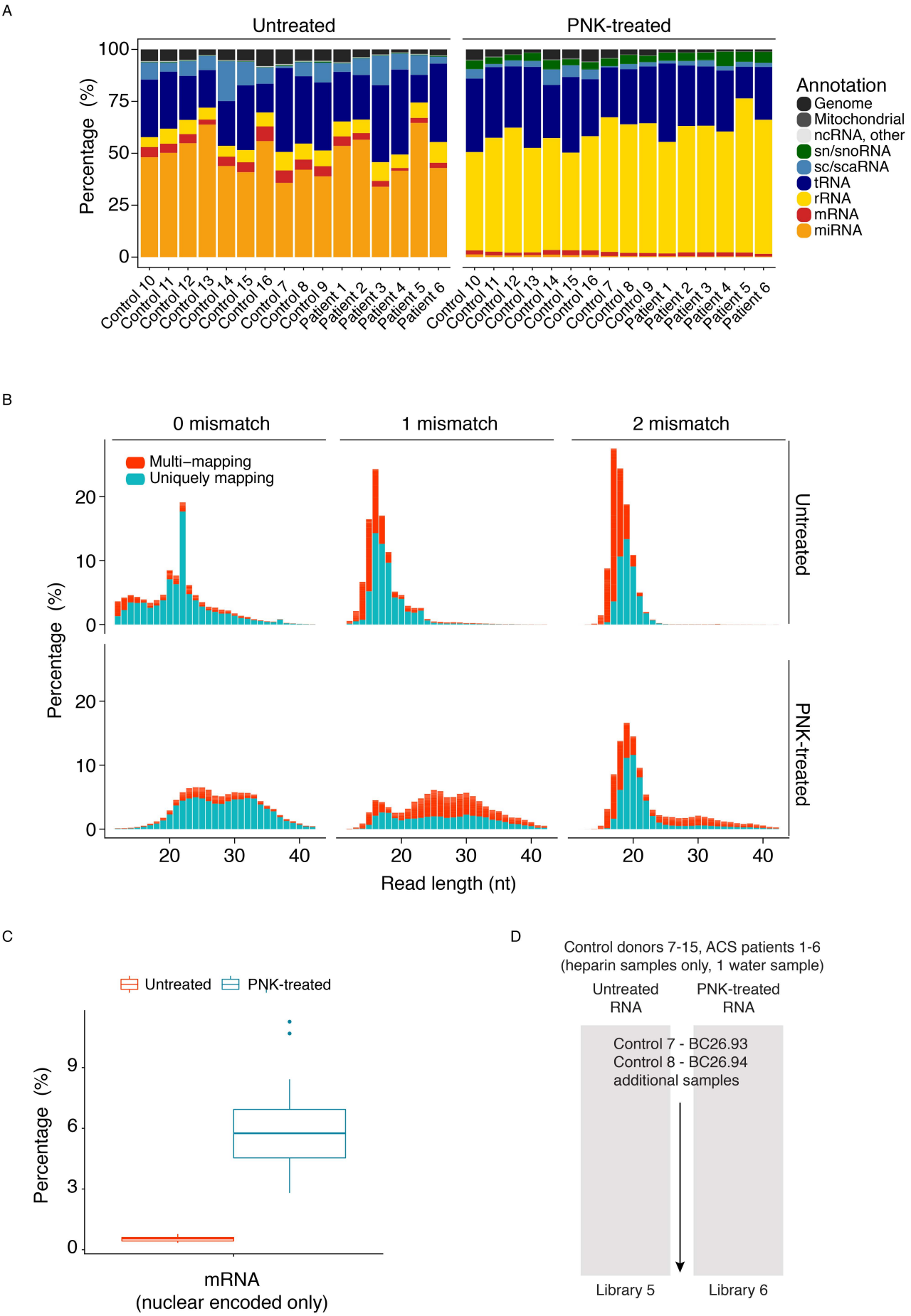
Supplemental Figure 3



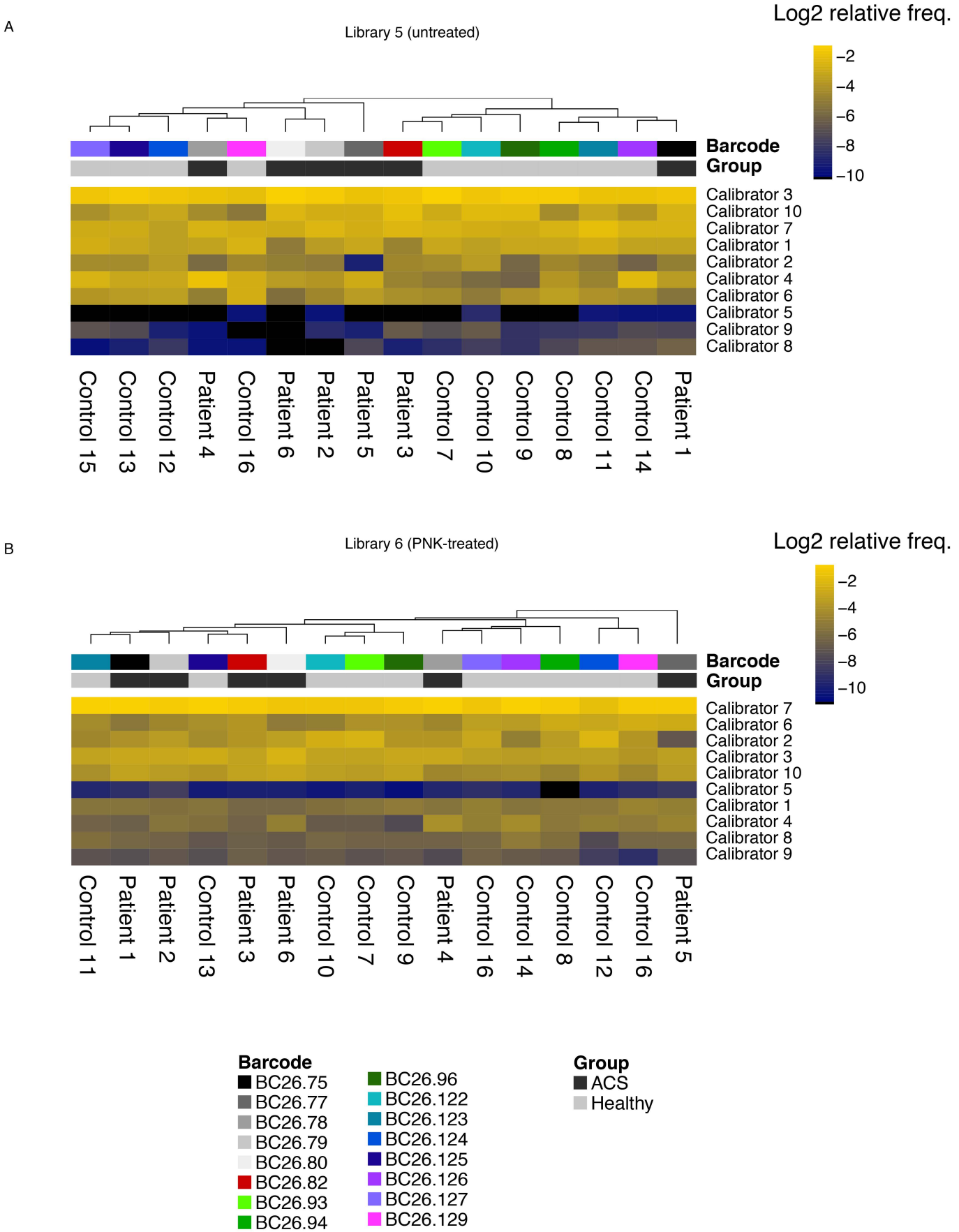
Supplementary Figure 4



Supplementary Figure 5



Supplemental Figure 6



Supplemental Figure 7