iScience, Volume *23*

Supplemental Information

Profiling Extracellular Long RNA Transcriptome in Human Plasma and Extracellular Vesicles for Biomarker Discovery

Rodosthenis S. Rodosthenous, Elizabeth Hutchins, Rebecca Reiman, Ashish S. Yeri, Srimeenakshi Srinivasan, Timothy G. Whitsett, Ionita Ghiran, Michael G. Silverman, Louise C. Laurent, Kendall Van Keuren-Jensen, and Saumya Das

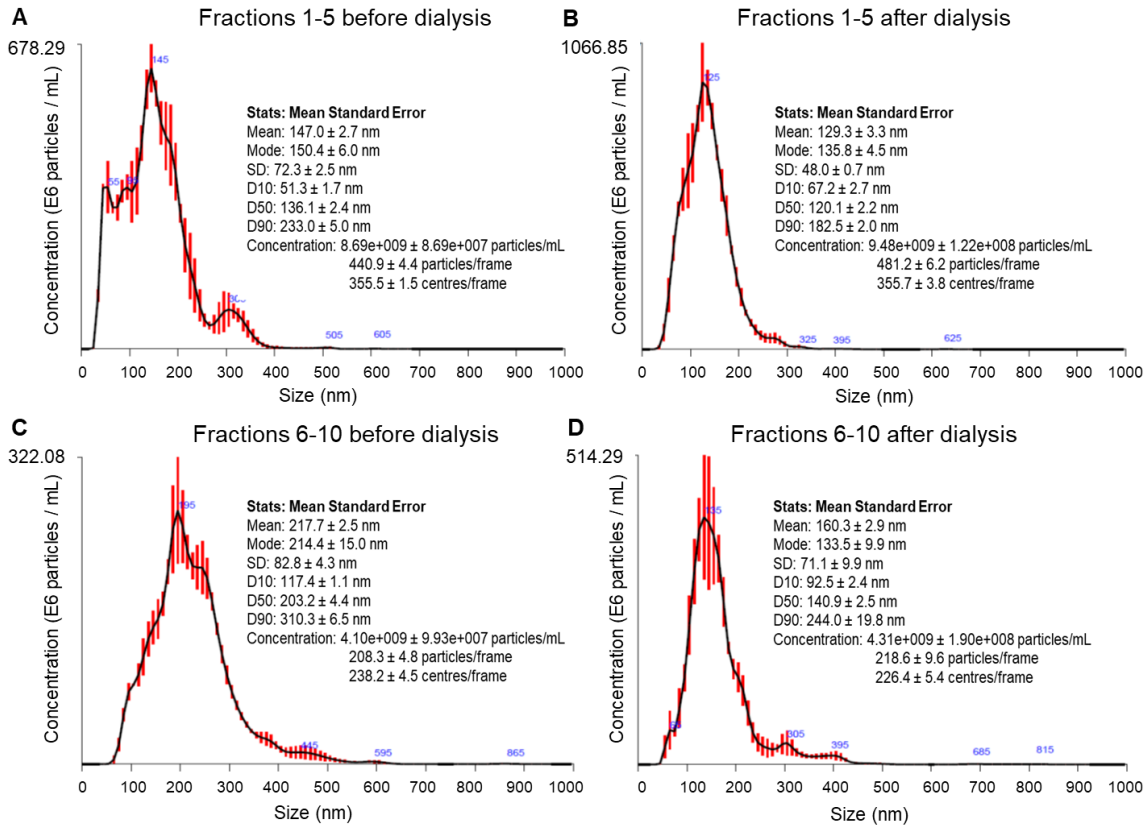

Figure S1. Related to Figure 3 and Transparent Methods. Nanoparticle Tracking Analysis of EVs before and after dialysis using the NanoSight LM10. (A) Fractions 1-5 before dialysis, (B) Fractions 1-5 after 24 hours of dialysis in 1X PBS buffer, (C) Fractions 6-10 before dialysis, (B) Fractions 6-10 after 24 hours of dialysis in 1X PBS buffer.

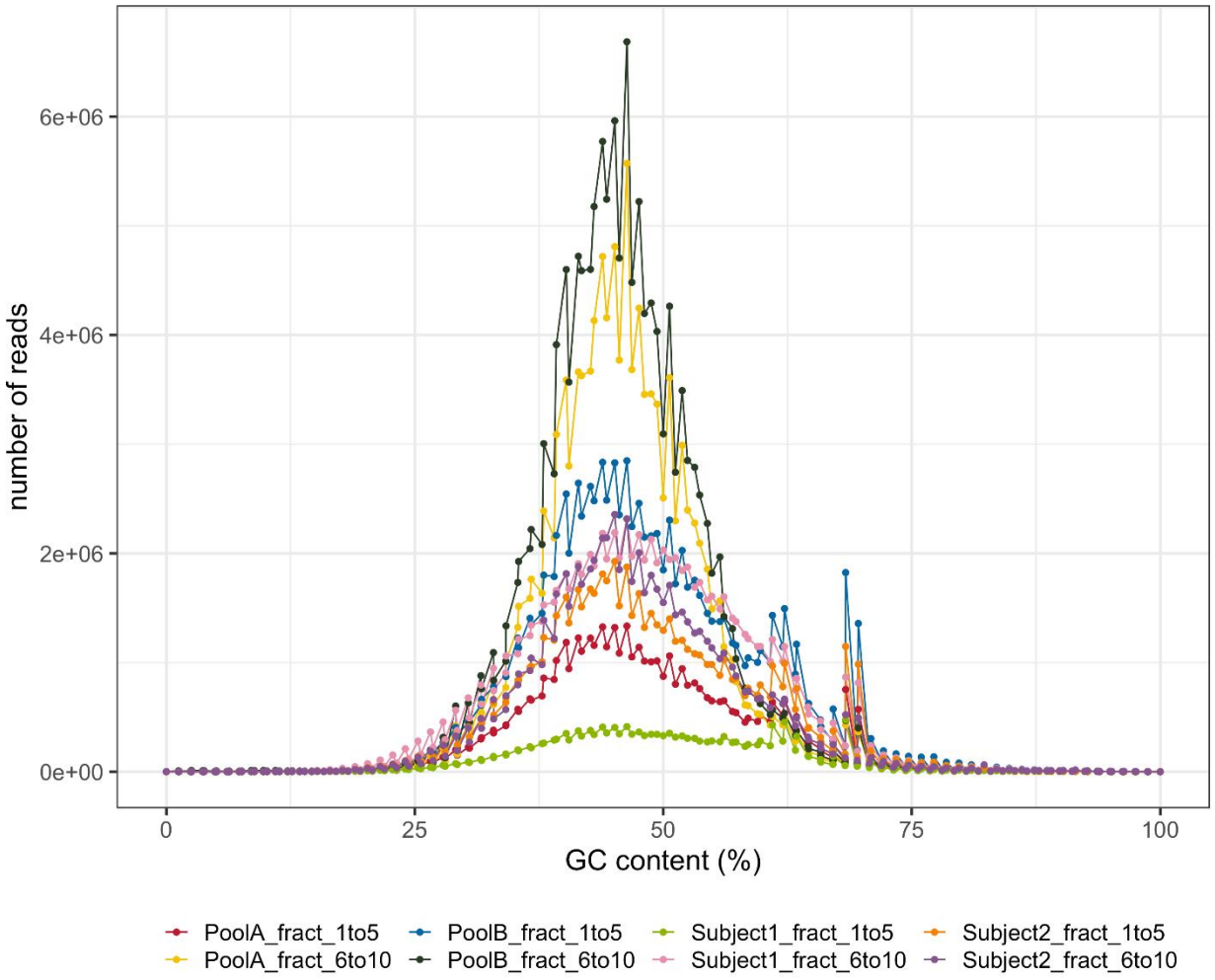

Figure S2. Related to Figure 5, Table 3 and Transparent Methods. Read GC content distribution of long RNA transcripts detected in fractions 1-5 and 6-10 for both individual and pooled samples.

Table S1. Related to Figure 2, Table 1 and Transparent Methods. Spearman's correlation from normalized counts in all samples. The mean correlation for each pool, kit, and site is shown.

Within site comparisons

Kit	Pool1_Site1	Pool1_Site2	Pool2_Site1	Pool2_Site2
OvationSoLo_Frag	0.89	0.84	0.81	0.81
RNA_Access_Frag	0.91	0.91	0.9	0.94
RNA_Access_noFrag	0.91	0.92	0.92	0.92
SMART_KAPPA_Frag	0.9	0.93	0.94	0.91
SMART_KAPPA_FragRibo	0.76	0.88	0.93	0.86
SMART_Pico_FragRibo	0.95	0.84	0.96	0.9

Between sites comparison

Kit	Pool1	Pool2
OvationSoLo_Frag	0.74	0.7
RNA_Access_Frag	0.86	0.87
RNA_Access_noFrag	0.85	0.88
SMART_KAPPA_Frag	0.87	0.88
SMART_KAPPA_FragRibo	0.73	0.84
SMART_Pico_FragRibo	0.82	0.88

Table S2. Related to Figure 4 and Transparent Methods. RNA concentrations from fractions 1-5 and 6-10 isolated using differential gradient centrifugation.

Sample Name	Plasma Volume into C-DGUC (mL)	Fractions	Total RNA (ng)
Subject 1	1	1-5	1.96
Subject 1	1	6-10	3.36
Subject 2	1	1-5	1.79
Subject 2	1	6-10	4.72
Pool A	3	1-5	3.01
Pool A	3	6-10	10.1
Pool B	3	1-5	3.22
Pool B	3	6-10	10.81

Transparent Methods

Samples for plasma RNA analysis

Human plasma from 10 healthy male and 10 healthy female donors 21-45 years of age were collected, processed, and combined to create a male pool and a non-pregnant female pool by the laboratory of Dr. Ionita Ghiran at Beth Israel Deaconess Medical Center (BIDMC). The BIDMC IRB approved the protocol (#2001P000591) to consent participants and collect samples according to the Declaration of Helsinki principles (20). Blood was collected from a peripheral vein using a 19g butterfly needle with K_2EDTA as the anticoagulant at room temperature and centrifuged at 500 *x* g for 10 min (20). The supernatant was removed and re-centrifuged at 2,500 x g for 10 minutes. The plasma was divided into 1 mL aliquots and stored at -80^{o}C until exRNA isolation was performed.

***Plasma RNA isolation*.**

Total RNA from the pool of healthy human male plasma was isolated using the miRNeasy Serum/Plasma Kit (Qiagen, Cat. No. 217184) as previously described (20). In brief, 6 mL of QIAzol Lysis Reagent was added to 1.2 mL of plasma. After vortexing and incubating for 5 minutes at room temperature, 1.2 mL of chloroform was added, followed by vigorous shaking for 15 seconds. Samples were incubated for 3 minutes at room temperature and centrifuged for 15 minutes at 12,000 x g at 4°C. The upper aqueous phase was transferred to a new tube where 1.5 volumes of 100% ethanol was added. 700 µL of the mixture was added to an assembled RNeasy MinElute spin column and centrifuged for 15 seconds at 1,000 x g at room temperature. This step was repeated until the rest of the sample had been loaded. The spin column was washed and centrifuged three times: the first wash was with 700 µL Buffer RWT and centrifuged for 15 seconds at 8,000 x g at room temperature, second with 500 µL Buffer RPE and centrifuged for 15 seconds at 8,000 x g at room temperature, and third wash was with 500 µL of fresh 80% ethanol and

centrifuged for 2 minutes at ≥8,000 x g at room temperature. The lid of the spin column was opened and spun at full speed spin for 5 minutes at room temperature to remove residual ethanol. RNA was extracted from the column by applying 30 µL of RNase-free water directly to the column and centrifuging for 1 minute at 100 x g and for another minute at full speed. The eluted volume was equally divided in 5 µL aliquots and frozen at -80°C.

Library Preparation and RNAseq Conditions. As this study aimed to compare six different library preparation kits/conditions, the differing kit protocols were used as follows:

SMARTer Universal Low Input RNA Kit for Sequencing + KAPA Hyper Prep Kit with (SMART_KAPA_FragRibo) and without (SMART_KAPA_Frag) Ribosomal Depletion. For each RNA sample, 10 ng total RNA was used for the double-stranded cDNA synthesis. For the library preparation with ribosomal depletion (SKF_FragRibo), the ribosomal RNA was depleted using Illumina's Ribo-Zero Gold rRNA Removal Kit (Illumina, Cat. No. MRZG12324). The ribosomal RNA-depleted RNA was further purified with the NucleoSpin RNA XS columns (Macherey Nagel, Cat. No. 740902.10) according to Takara Bio's, but no longer supported, "Protocol for Removal of rRNA from Small Amounts of Total RNA." The double-stranded cDNA was synthesized from both the the ribo-depleted (SMART_KAPA_FragRibo) or not (SMART_KAPPA_Frag) RNA samples using the SMARTer Universal Low Input RNA Kit for Sequencing (Takara Bio, Cat. No. 634940) with a 16-cycle PCR. The concentration of cDNA was measured with the Qubit dsDNA HS Reagent (ThermoFisher Scientific, Cat. No. Q32854). In the SMART_KAPA_FragRibo group, 6 ng of double-stranded cDNA for three of the four samples (Pool 1 Replicate 2, Pool 2 Replicate 1, and Pool 2 Replicate 2) and all of it for the fourth sample (Pool 1 Replicate 1) that had an undetectable amount of cDNA was further fragmented with the Covaris E220 sonicator (Peak Incident Power = 140W, Duty Factor = 10%, Cycles/Burst = 200, Treatment Time = 80 sec). In the SMART_KAPA_Frag group, 10ng of

double-stranded cDNA from all four samples was fragmented using the same Covaris parameters. The fragmented cDNA from both groups was then prepared into libraries using the KAPA Hyper Prep Kit (KAPA Biosystems, Cat. No. KK8504). This library preparation included a combination end repair and A-tailing reaction, followed by a 4°C overnight ligation of uniquely barcoded adapters to each sample at a 50:1 adapter to insert molar ratio, and then a 12-cycle enrichment PCR. The size of each final library was determined by TapeStation High Sensitivity D1000 ScreenTape (Agilent Technologies, Cat. No. 5067-5584 & Cat. No. 5067-5603), and the concentration was measured with KAPA SYBR FAST Universal qPCR Kit (KAPA Biosystems, Cat. No. KK4824). Libraries were then combined into an equimolar pool, which was also measured for size and concentration. The pool was clustered onto a paired-end flowcell (Illumina, Cat. No. PE-401-3001) with a 20% v/v PhiX v3 spike-in (Illumina, Cat. No. FC-110-3001) and sequenced on Illumina's HiSeq 2500 with TruSeq SBS Kit v3-HS chemistry (Illumina, Cat. No. FC-401-3002) to 50 million read pairs per library. The first and second reads were each 83 bases.

TruSeq RNA Access with (RNA_Access_Frag) and without (RNA_Access_noFrag) Fragmentation. For each RNA sample, 10ng of total RNA was prepared into Illumina-compatible, pre-capture libraries using the TruSeq RNA Access kit (Illumina, Cat. No. RS-301-2001), which is now called TruSeq RNA Exome (Illumina, Cat. No. 20020189). For the library preparation with fragmentation (RNA_Access_Frag), the only difference was an extra step of chemical and thermal RNA fragmentation (94 °C for 8 minutes) which was done prior to the following common steps of double-strand cDNA synthesis, A-tailing, end repair, uniquely barcoded adapter ligation, and a 15-cycle enrichment PCR. Pre-capture libraries were measured for size using Agilent's High Sensitivity D1000 ScreenTape (Agilent Technologies, Cat. No. 5067-5584 & Cat. No. 5067-5603) and concentration by the Qubit dsDNA HS Assay kit (ThermoFisher Scientific, Cat. No. Q32851). For the capture step, 200 ng of each of the pre-

capture cDNA libraries was pooled and each pool contained four libraries. The capture step included two overnight biotinylated probe hybridizations and streptavidin bead selections followed by a 10-cycle enrichment PCR. The coding-region enriched library pool was measured for size using Agilent's High Sensitivity D1000 ScreenTape (Agilent Technologies, Cat. No. 5067-5584 & Cat. No. 5067-5603) and concentration via the KAPA SYBR FAST Universal qPCR Kit (KAPA Biosystems, Cat. No. KK4824). Libraries were then combined into an equimolar pool which was also measured for size and concentration. The pool was clustered onto a paired-end flowcell (Illumina, Cat. No. PE-401-3001) with a 1% v/v PhiX v3 spike-in (Illumina, Cat. No. FC-110-3001) and sequenced on Illumina's HiSeq 2500 with TruSeq SBS Kit v3-HS chemistry (Illumina, Cat. No. FC-401-3002) to 50 million read pairs per library. The first and second reads were each 83 bases. Since these libraries were prepared, Illumina has repackaged and rebranded their TruSeq RNA Access kit as TruSeq RNA Exome, wherein they sell the pre-capture portion of the kit as TruSeq RNA Library Prep for Enrichment (Illumina, Cat. No. 20020189), the capture portion as TruSeq RNA Enrichment (Illumina, Cat. No. 20020490), and the actual capture probes as Illumina Exome Panel-Enrichment Oligos (Illumina, Cat. No. 20020183) separately. No changes to the chemistry portions of the protocol were made.

Ovation SoLo RNA-Seq System (Ovation_SoLo_Frag). For each RNA sample, 10 ng of total RNA was prepared into Illumina-compatible libraries using Ovation SoLo RNA-Seq System, Human (Nugen, Cat. No. 0500-32). Library preparation included double-stranded cDNA synthesis, fragmentation, end repair, uniquely barcoded adapter ligation, qPCR PCR 1 optimization, PCR 1, ribodepletion, and PCR2 following the manufacturer's protocol. After ligation, the optimal number of library amplification 1 cycles for each library was measured by qPCR with EvaGreen Dye (Biotium, Cat. No. 31000) according to Nugen's Ovation SoLo protocol instructions (i.e., 14 cycles for Pool 1 and 16 cycles for Pool 2 samples). After amplification 1 step, library concentration was measured with Qubit dsDNA HS Assay kit

(ThermoFisher Scientific, Cat. No. Q32851), and 10 ng of each library was then loaded into InDA-C ribodepletion reaction and subsequent library amplification 2 (2 cycles + 6 cycles for all samples). Libraries were measured for size via TapeStation (Agilent High Sensitivity D1000 ScreenTape & Sample Buffer, Cat. No. 5067-5584 & Cat. No. 5067-5603) and concentration via qPCR (KAPA Biosystems, Cat. No. KK4824), before combined into an equimolar pool. The pool was clustered onto a paired-end flowcell (Illumina, Cat. No. PE-401-3001) with no PhiX v3 spike-in and sequenced on Illumina's HiSeq 2500 with TruSeq SBS Kit v3-HS chemistry (Illumina, Cat. No. FC-401-3002) to 50 million read pairs per library. Ovation SoLo Custom R1 Primer was used to prime for read 1 sequencing, while standard Illumina sequencing primers were used to prime for all other sequencing reads. The first and second reads were each 83 bases.

SMARTer Stranded Total RNA-Seq Kit v2 – Pico Input Mammalian (SMART_Pico_Frag).

For each RNA sample, indexed, Illumina-compatible, double-stranded cDNA libraries were synthesized from 10ng total RNA with Takara Bio's SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian kit (Takara Bio, Cat. No. 634411). Library preparation included chemical and thermal RNA fragmentation (94 °C for 2 min), cDNA synthesis, a 5-cycle indexing PCR, ribosomal cDNA depletion, and a 14-cycle enrichment PCR. Each library was measured for size with Agilent's High Sensitivity D1000 ScreenTape and reagents (Agilent, Cat. No. 5067-5584 & 5067-5603) and concentration with KAPA SYBR FAST Universal qPCR Kit (Kapa Biosystems, Cat. No. KK4824). Libraries were combined into an equimolar pool which was subsequently measured for size and concentration. The pool was hybridized onto a paired-end flowcell (Illumina, Cat. No. PE-402-4002) with a 1% v/v PhiX Control v3 spike-in (Illumina, Cat. No. FC-110-3001) using Illumina's HiSeq Rapid Duo cBot Sample Loading Kit (Illumina, Cat. No. CT-403-2001) on a cBot. Each template-hybridized flowcell was then clustered and sequenced on

Illumina's HiSeq 2500 with HiSeq Rapid v2 chemistry (Illumina, Cat. No. FC-402-4022) to 50 million read pairs per library. The first and second reads were each 83 bases.

Extracellular Vesicle Collection and RNA isolation. Whole blood was collected from three healthy donors and two patients with supraventricular tachycardia in the MGH clinic (IRB# 2017P002010). Within two hours from collection, all blood samples were centrifuged at 800 x g for 15 minutes and 1,800 x g for 10 minutes at room temperature, and plasma was immediately stored at -80°C until further analysis. For the isolation of EV-enriched fractions, we pooled the plasma from healthy donors and used 3mL for each of the two pools (i.e., Pool A and B) used for downstream analysis. The plasma from patients was not pooled and 1mL for each sample (i.e., Subject 1 and 2) was used for downstream analysis. All samples were analyzed following the iodixanol Cushioned-Density Gradient Ultracentrifugation (C-DGUC) method as described by Li et al. (24). In brief, pooled plasma wad diluted with 1X PBS (ThermoFisher Scientific, cat. No. 10010023) to a final volume of 8mL and carefully overlaid on a 2mL cushion of 60% iodixanol (Sigma-Aldrich, cat. No. D1556) in a 13.2mL centrifuge tube (Beckman Coulter, cat. No. 331372). The samples were centrifuged at 100,000 x g for 2 hours at 4 °C using a SW 41 Ti rotor (k-factor 124) and the top 7mL (out of 10mL) were discarded. The bottom 3mL volume (containing the EVs) was mixed well and three layers of 20%, 10%, and 5% of iodixanol (3mL each) were carefully overlaid on top of it, respectively, totaling a volume of 12mL. The samples were centrifuged again at 100,000 x g for 18 hours at 4 °C using a SW 41 Ti rotor (k-factor 124). Once the centrifugation step was done, 12 fractions of 1mL were collected starting from the top of the tube. To remove iodixanol, we performed an extra step of dialysis using a Float-A-Lyzer G2 device per manufacturer's instructions (Spectrum Labs, cat. No. G235059,). For this step, fractions 1-5 and 6-10 were pooled, which were subsequently used as fraction pools (5mL each) for downstream analyses. RNA from each fraction pool (fractions 1-5 and 6-10) and sample (Subject 1, Subject 2, Pool A, and Pool B) was extracted using the exoRNeasy Serum/Plasma Maxi Kit (Qiagen, Cat. No. 77064), and then quantitated with Quant-iT Ribogreen RNA Assay

(ThermoFisher Scientific, Cat. No. R11490) according to ThermoFisher's low-range Ribogreen protocol.

Library Preparation and RNAseq Conditions for extracellular RNA. For each RNA sample (fractions 1-5 and fractions 6-10 of Subject 1 and 2, and Pool A and B), indexed, Illumina-compatible, double-stranded cDNA libraries were synthesized from total extracellular RNA with Takara Bio's SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian kit (Takara Bio, Cat. No. 634411). Library preparation included chemical and thermal RNA fragmentation (94 °C for 2 min), cDNA synthesis, a 5-cycle indexing PCR, ribosomal cDNA depletion, and a 14-16 cycle enrichment PCR. Total RNA input into library preparation per sample was normalized to the lowest amount of RNA within C-DGUC isolated fractions (1-5 or 6-10) and within plasma volume input into C-DGUC (3 mL for Pool A and B, 1 mL for Subject 1 and 2), and the number of enrichment PCR cycles was dictated by the total RNA input per sample. Total input RNA for Subject 1 and 2 was 1.8 ng for fractions 1-5 and 3.4 ng for fractions 6-10. Total input RNA for Pool A and B was 3.0 ng for fractions 1-5 and 10 ng for fractions 6-10. Samples with 10 ng of total RNA input underwent 14 cycles of PCR enrichment, whereas samples with less than 10 ng of total RNA input underwent 16 cycles of PCR enrichment. Each library was measured for size with Agilent's High Sensitivity D1000 ScreenTape and reagents (Agilent, Cat. No. 5067-5584 & 5067-5603) and concentration with KAPA SYBR FAST Universal qPCR Kit (Kapa Biosystems, Cat. No. KK4824). Libraries were combined into an equimolar pool which was measured for size and concentration. The pool was clustered onto a paired-end flowcell (Illumina, Cat. No. PE-401-3001) with a 1% v/v PhiX v3 spike-in (Illumina, Cat. No. FC-110-3001) and sequenced on Illumina's HiSeq 2500 with TruSeq SBS Kit v3-HS chemistry (Illumina, Cat. No. FC-401-3002) to 70 million read pairs per library. The first and second reads were each 82 bases.

Nanoparticle Tracking Analysis. We measured the concentration and size distribution of isolated EVs from the different fractions using the NanoSight LM10 device (Malvern

Instruments, Westborough, MA). The device was washed with double-filtered (0.2um) 1X PBS (ThermoFisher Scientific, cat. No. 10010023) prior to each sample measurement. The settings for the camera were adjusted according to the manufacturer's instructions and Three videos of 30 seconds per sample were recorded and analyzed using the Nanoparticle Tracking Analysis software 2.3. Results are shown as mean ± standard deviation of these recordings.

Immunoblot Analysis. Dialyzed EVs (1mL) were lysed using 100uL of 10X lysis buffer (Cell Signaling, cat. No. 9803S), 10uL of protease inhibitor, 10uL of phosphatase inhibitor, and 5uL of phenylmethylsulfonyl fluoride (PMSF). Next, the lysate was loaded on an Amicon Ultra-15 Centrifugal Filter Unit (Millipore, cat. No. UFC900308) and centrifuged for 60 min at 5,000 x g to concentrate the sample down to ~200uL. For the western blot, we loaded 30 µL (1 µg/µL) of each sample on 4-20% polyacrylamide gels (Bio-Rad, cat. No. 567-1094) following the standard western blot steps. Mouse monoclonal anti-human CD9 (BioLegend, cat. No. 312102), CD63 (BD Biosciences, cat. No. 556019), Alix (BioLegend, cat. No. 634501), APOA1 (Cell Signaling Technology, cat. No. 3350S) and rabbit anti-human AGO2 (Cell Signaling Technology, cat. No. 2897S) were used at 1:500 dilutions prior to incubation with the samples.

RNA Seq data analysis. Fastq files were generated from the raw sequence files using bcl2fastq v2.19.1.403 (Illumina) using default parameters. In order to standardize input read amount across kits and sites, we randomly down-sampled the raw fastq files using seqtk v1.2-r101-dirtyn (https://github.com/lh3/seqtk) to 50 million reads and 10 million reads for plasma, and 70 million reads for EV samples. Reads were trimmed with cutadapt v1.17 (25) according to kit recommendations: -u 7 -U 7 for SMARTer/KAPA Hyper, -u 5 for Ovation SoLo, and -U 3 for SMARTerPicov2. Samples prepared with the RNA Access kit were not trimmed. Trimmed fastq files were then aligned to the GRCh38 genome with STAR v2.6.1d (26) with the following parameters: --runMode alignReads --outSAMtype BAM Unsorted --outSAMmode Full --

outSAMstrandField intronMotif --outFilterType BySJout --outSAMunmapped Within --outSAMmapqUnique 255 --outFilterMultimapNmax 20 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.1 --alignMatesGapMax 1000000 --seedSearchStartLmax 50 --alignIntronMin 20 --alignIntronMax 1000000 --alignSJoverhangMin 18 --alignSJDBoverhangMin 18 --chimSegmentMin 18 --chimJunctionOverhangMin 18 --outSJfilterOverhangMin 18 18 18 18 --alignTranscriptsPerReadNmax 50000. Following genome alignment, reads were counted with featureCounts v1.6.3, (part of the subread package) (27) using a non-redundant genome annotation combined from GENCODE 29 and LNCipedia5.2 and the following parameters: -p -t exon -g gene_id. Additionally, the strandedness parameter was passed to featureCounts according to kit as following: -s 1 for Ovation SoLo, -s 2 for RNA Access, and -s 2 for SMARTerPico v2. SMARTer/KapaHyper is unstranded, so no strandedness parameter was set. Normalized counts were generated with DESeq2 v1.22.2 (28) after first filtering out genes that had 10 or fewer reads on average. Trimmed fastq files were also quasi-mapped to the same combined GENCODE 29 and LNCipedia5.2 annotation (29) using salmon quant v0.11.3 (30) to estimate transcripts per million (TPMs) with the following parameters: --libType A --numBootstraps 100 --seqBias --gcBias -dumpEq. Transcript coverage was calculated by creating a bedgraph with bedtools coverage v2.27.1 using the aligned BAM file from STAR and the non-redundant genome annoation.

Ingenuity Pathway Analysis. A list of genes uniquely detected in fractions 6-10, defined as having a mean on greater than 10 counts and not expressed an any fraction 1 to 5 sample, was uploaded to Ingenuity (Qiagen) for pathway analysis. A core analysis was then performed using human data on this unique list of genes. The resulting pathways are listed in Supplementary Figure 2.