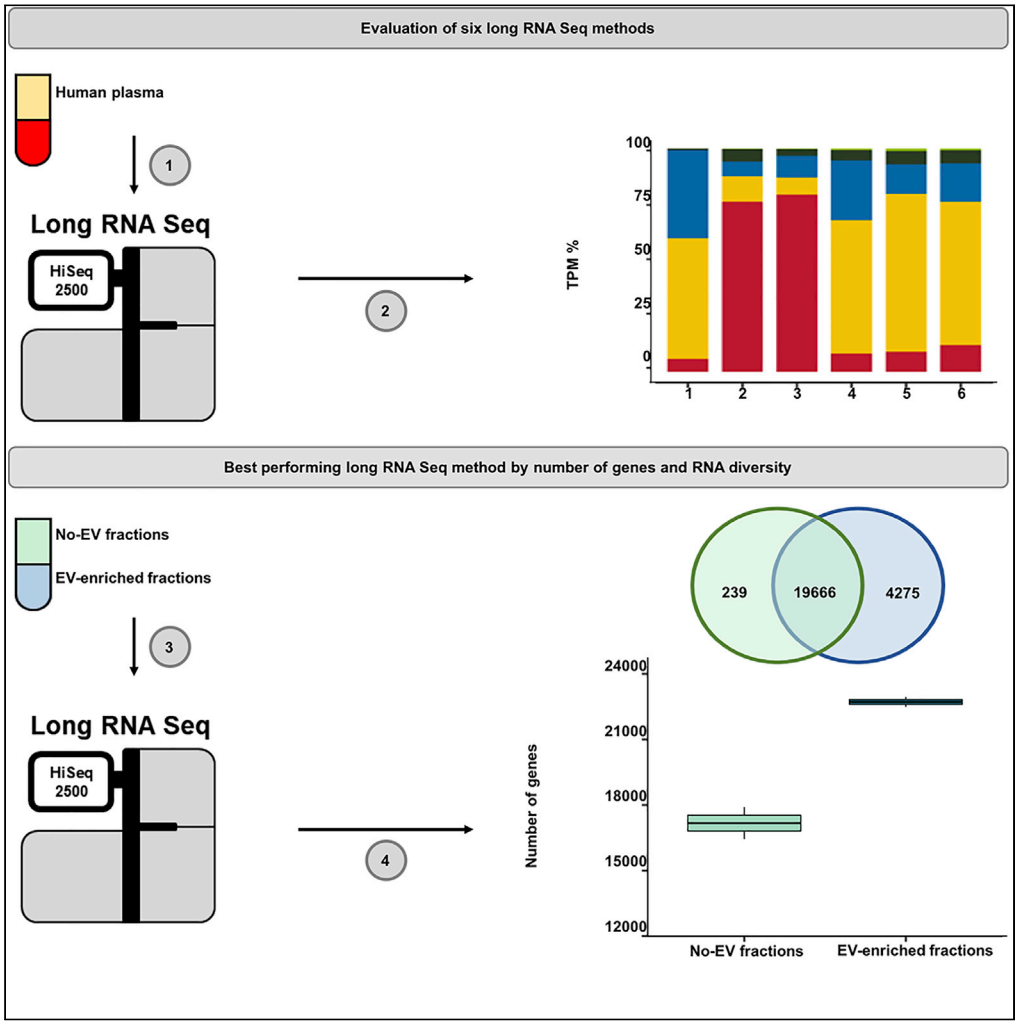
iScience

CellPress
OPEN ACCESS

Article

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



Rodosthenis S. Rodosthenous, Elizabeth Hutchins, Rebecca Reiman, ..., Louise C. Laurent, Kendall Van Keuren-Jensen, Saumya Das

kjensen@tgen.com (K.V.K.-J.)
sdas@mgh.harvard.edu (S.D.)

HIGHLIGHTS

Low-input RNA-sequencing kits were evaluated using plasma extracellular RNA

Thousands of unique and diverse long RNA transcripts were detected at >80% coverage

Extracellular vesicle-enriched fractions had a unique set of long RNA transcripts

Rodosthenous et al., iScience *23*, 101182
June 26, 2020
https://doi.org/10.1016/j.isci.2020.101182

iScience

CellPress
OPEN ACCESS

Article

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

Rodosthenis S. Rodosthenous,[1,5] Elizabeth Hutchins,[2,5] Rebecca Reiman,[2] Ashish S. Yeri,[1] Srimeenakshi Srinivasan,[4] Timothy G. Whitsett,[2] Ionita Ghiran,[3] Michael G. Silverman,[1] Louise C. Laurent,[4] Kendall Van Keuren-Jensen,[2,6,*] and Saumya Das[1,6,7,*]

SUMMARY

The recent discovery of extracellular RNAs in blood, including RNAs in extracellular vesicles (EVs), combined with low-input RNA-sequencing advances have enabled scientists to investigate their role in human disease. To date, most studies have been focusing on small RNAs, and methodologies to optimize long RNAs measurement are lacking. We used plasma RNA to assess the performance of six long RNA sequencing methods, at two different sites, and we report their differences in reads (%) mapped to the genome/transcriptome, number of genes detected, long RNA transcript diversity, and reproducibility. Using the best performing method, we further compare the profile of long RNAs in the EV- and no-EV-enriched RNA plasma compartments. These results provide insights on the performance and reproducibility of commercially available kits in assessing the landscape of long RNAs in human plasma and different extracellular RNA carriers that may be exploited for biomarker discovery.

INTRODUCTION

RNA is an essential biomolecule that plays an important role in diverse cellular functions. Due to this central role, RNA expression has been extensively studied in the context of diagnosis, prognosis, and treatment of complex diseases. Technological advances, especially on the development of RNA sequencing (RNA-seq), provide new opportunities for discoveries related not only to gene expression but also to differing transcript isoforms, splice variants, and gene fusions in an unbiased way (Byron et al., 2016). In fact, RNA-seq-based tests have already made it into clinical applications, such as the FoundationOne Heme test (Foundation Medicine) that employs RNA-seq toward gene fusion detection in blood cancers (Intlekofer et al., 2018), the GEM ExTra test (Ashion Analytics) that integrates exome sequencing and RNA-seq for clinical use (Borad et al., 2014; Nasser et al., 2015), and the ExoDx Prostate test (Exosome Diagnostics) that utilizes RNA-Seq data from extracellular vesicles (i.e., exosomes) isolated from urine (McKiernan et al., 2016). Therefore, the clinical potential for RNA-seq demands further methodological testing toward protocols that maximize efficiency, RNA species output, and can be performed on small sample volumes, and/or low inputs of RNA.

Blood remains the most commonly collected biofluid in the clinic and in most diseases, it represents an ideal source of accessible biological information from diverse tissues. Blood contains a range of biomarkers including proteins, metabolites, DNA, and RNA that can be measured and analyzed for the development of blood-based biomarkers in diverse disease types such as cancer, cardiovascular disease, and neurodegenerative diseases. Recently, multiple efforts have been focusing on the development of diagnostic and therapeutic applications that are based on extracellular RNA (exRNA), primarily RNA encapsulated in extracellular vesicles (EVs) or carried in other carrier subtypes (Srinivasan et al., 2019). EVs (i.e, exosomes and microvesicles) are typically 20–1000 nm vesicles that are released from cells into the blood circulation (or other biofluids) and contain proteins, lipids, DNA, and RNA molecules (reviewed in (Raposo and Stoorvogel, 2013; Yanez-Mo et al., 2015)). The discovery of RNA molecules in blood EVs and the proof that EVs provide protection to RNA molecules from being degraded by RNAses (Wang et al., 2010) lead to an increased interest in the profiling of RNAome in blood EVs under different conditions. In fact, EVs contain

[1]Cardiovascular Research Center, Massachusetts General Hospital, Harvard Medical School, 185 Cambridge Street, Simches 3rd Floor, Boston, MA 02114, USA

[2]Division of Neurogenomics, The Translational Genomics Research Institute, Phoenix, AZ 85004, USA

[3]Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA

[4]Department of Obstetrics, Gynecology, and Reproductive Sciences, Sanford Consortium for Regenerative Medicine, University of California, San Diego, La Jolla, CA 92093, USA

[5]These authors contributed equally

[6]These authors contributed equally

[7]Lead Contact

*Correspondence: kjensen@tgen.com (K.V.K.-J.), sdas@mgh.harvard.edu (S.D.)

https://doi.org/10.1016/j.isci.2020.101182

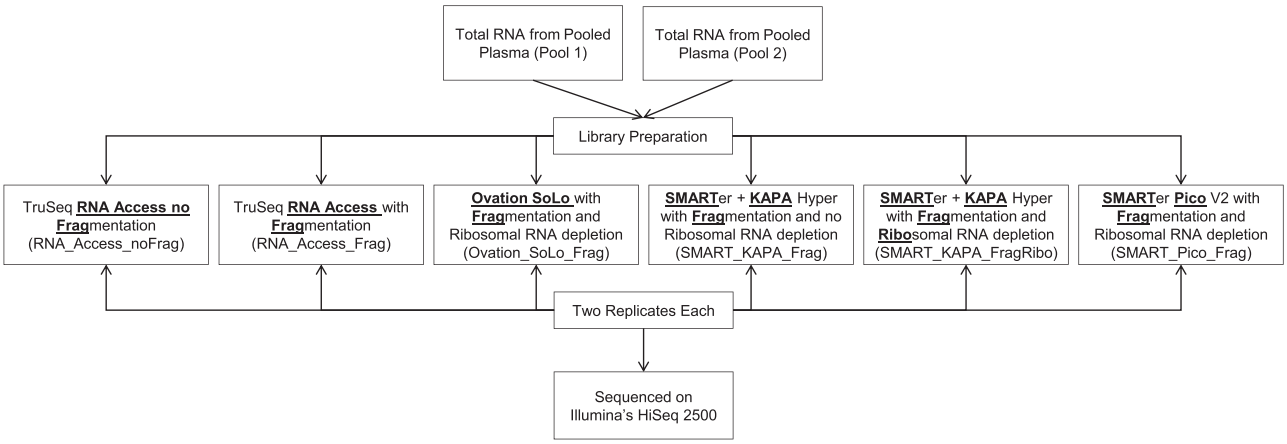


Figure 1. Total RNA from Pooled Plasma from Healthy Volunteers was Used to Prepare RNA-Sequencing Libraries to Evaluate Six Different Kits/Conditions. The total RNA was equally divided, and the libraries were constructed in duplicates. All libraries were sequenced on Illumina's HiSeq 2500 platform under the same conditions.

all the RNA species that are found in the cell as well (i.e., mRNAs, tRNAs, lncRNAs, rRNAs, miRNAs, etc.) (Murillo et al., 2019). However, electrophoretic analysis of EV RNAs reveal that apart from intact small RNAs, they also contain fragments of longer RNAs (Lasser et al., 2011; Skog et al., 2008), which can make standard sequencing more difficult. Numerous studies have shown that RNAs detected in blood circulating EVs are associated with disease prognosis, diagnosis, and progression and can be therefore used for the development of clinical tests (Ingenito et al., 2019; Liu et al., 2019; Quinn et al., 2015). However, successful development of such biomarkers requires standardized and reproducible RNA-Seq protocols that can be used to measure EV-containing RNAs from small volumes of blood and therefore low yields of RNA input.

Our group has previously focused on the development and optimization of RNA-Seq methods to study small RNAs, such as miRNAs, tRNAs, and piRNAs in biofluids (Murillo et al., 2019; Shah et al., 2017; Yeri et al., 2018). However, equivalent methodology to profile the longer RNAs and their fragments have not been well reported. In this study, we focus on methodology to profile fragments of protein-coding and long non-coding RNA transcripts (e.g., mRNAs, lncRNAs, and other long non-coding RNAs) in biofluids and extracellular RNA carriers. In brief, we took plasma from healthy volunteers and divided the plasma into two independent, uniform pools, extracted the RNA, and compared the RNA profiles obtained across six different RNA-sequencing library preparation kits and two different laboratory sites to determine optimal performance as measured by the number of reads mapping to the genome/transcriptome and RNA species diversity. Using the best-performing RNA-seq kit based on these two metrics, we examined the profile of RNAs in EV-enriched and no-EV plasma compartments isolated from both pooled plasma or plasma from individual human subjects.

RESULTS

To allow for the systematic comparison of library construction kits/conditions, we employed total RNA from two independent pools of plasma samples. We divided the total RNA from both pools equally among the six different RNA sequencing kits/conditions and constructed libraries in duplicate at two independent sites. Following sample preparation, we performed long RNA sequencing using Illumina's HiSeq 2500 platform to assess genome and transcriptome mapping percentage (Figure 1). In order to standardize the number of input reads for downstream analysis and comparison across kits, FASTQs were randomly and uniformly down-sampled to 50 million read pairs prior to genome and transcriptome mapping.

We found that all six tested kits/conditions tended to have similar percentages of reads mapped to the genome across pools, although Ovation SoLo (OS) showed the lowest percentage uniquely mapping reads in both pools (Figure 2A). The percentage of reads mapped to the transcriptome was higher in the TruSeq RNA Access kit (now called RNA Exome) with or without fragmentation, whereas the other kits had ~50% or fewer of reads mapped to the transcriptome (Figure 2B). To measure the RNA biotypes captured across

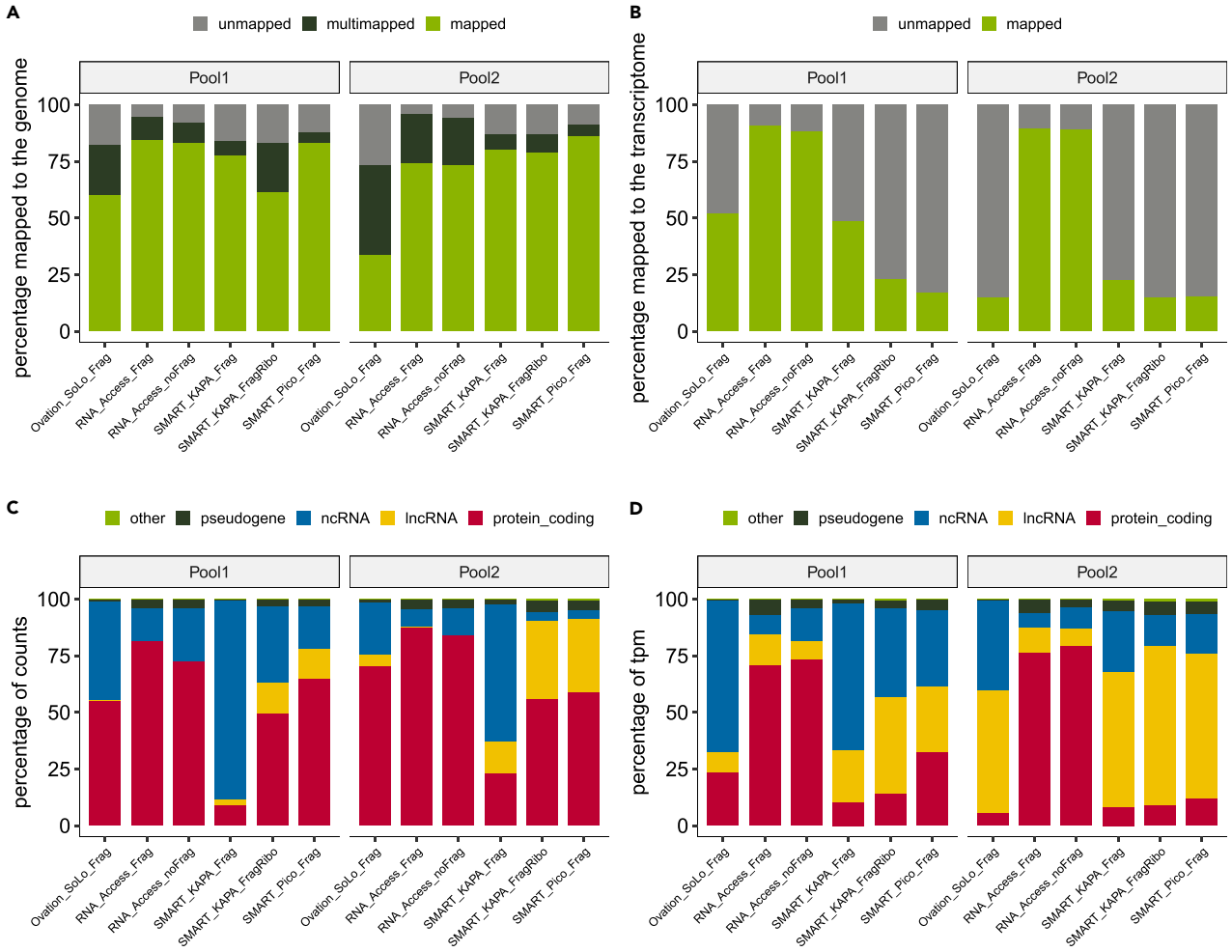


Figure 2. Assessment of Genome and Transcriptome (hg38) Mapping Percentage by RNA-Sequencing Library Preparation Kit/Condition and by RNA Species Using Total RNA from Pooled Plasma.

(A) Percentage of input reads for each RNA pool and kit/condition that were uniquely mapped (green), multimapped (black), or unmapped (gray) to the human genome using STAR.

(B) Percentage of input reads for each RNA pool and kit/condition that were quasimapped (green) or unmapped (gray) to the human transcriptome using salmon.

(C) Percentage of uniquely mapped read counts (light green bar in panel A) assigned to the human transcriptome using featureCounts for each RNA pool and kit/condition and represented by the following RNA species: protein coding (red), lncRNAs (yellow), non-coding RNAs (blue), pseudogenes (black), and other RNAs (including RNAs to be experimentally confirmed, immunoglobulin genes, and T cell receptor genes; green).

(D) Percentage of transcripts per kilobase million (TPM) quasimapped to the human transcriptome (light green bar in panel B) for each RNA pool and kit/condition and represented by the following RNA species: protein coding (red), lncRNAs (yellow), non-coding RNAs (blue), pseudogenes (black), and other RNAs (green).

kits, we took the reads mapping to the transcriptome and displayed the percentage of reads assigned to each of four RNA biotypes defined by Ensembl/Havana/Vega (protein-coding, lncRNAs, ncRNAs, and pseudogenes) and calculated the percentage of counts (Figure 2C) and transcripts per kilobase million (TPM) (Figure 2D) represented by each biotype category. From this analysis, we found that the SMARTer Pico v2 and SMARTer/KapaHyper (Frag and FragRibo) kits had a higher proportion of lncRNA counts and TPMs across the pools. As expected, due to its capture probe design, the RNA Access kit showed the highest percentage of protein coding RNA when analyzed both by counts and TPMs (Figures 2C and 2D). However, when we looked at diversity of RNA species performance, the SMARTer Pico V2 with Fragmentation and Ribosomal RNA depletion (SMART_Pico_Frag) showed the highest number (154,942) of unique transcripts detected as compared with all other kits (Table 1). To assess reproducibility of each kit, we calculated the Spearman's correlation from DESeq2-normalized counts for all pools and

Kit	Mean	SD
Ovation_SoLo_Frag	53,695	13,109
RNA_Access_Frag	43,215	3,568
RNA_Access_noFrag	45,961	7,756
SMART_KAPA_Frag	138,658	9,927
SMART_KAPA_FragRibo	113,303	34,667
SMART_Pico_Frag	154,942	14,618

Table 1. Number of Long RNA Transcripts Detected in Human Plasma across Evaluated Kits.

replicates. The mean correlation for each pool, kit, and site combination listed is shown in Table S1. We found that all kits had a comparable reproducibility within and between site(s), except for the Ovation_SoLo_Frag and SMART_KAPA_FragRibo kits, which showed lower reproducibility.

Our main objective was to assess the long RNA profile of EV-enriched plasma fractions as per the traditional markers (i.e, CD9, CD63, and Alix) and whether is different than other fractions. EVs were isolated using iodixanol cushioned-density gradient centrifugation (C-DGUC), which enriches EVs based on their density (Witwer et al., 2013). Any isolation method will target slightly different EV populations and could give slightly different results regarding the associated exRNAs. However, the C-DGUC method is considered very stringent in separating EVs and lipoproteins (19) and is recommended in the MISEV guidelines to achieve better separation of EVs (20). All samples were dialyzed to remove iodixanol prior to downstream analyses (Figure 3A). As depicted in Figure S1, dialysis did not have any impact on the numbers of EVs;

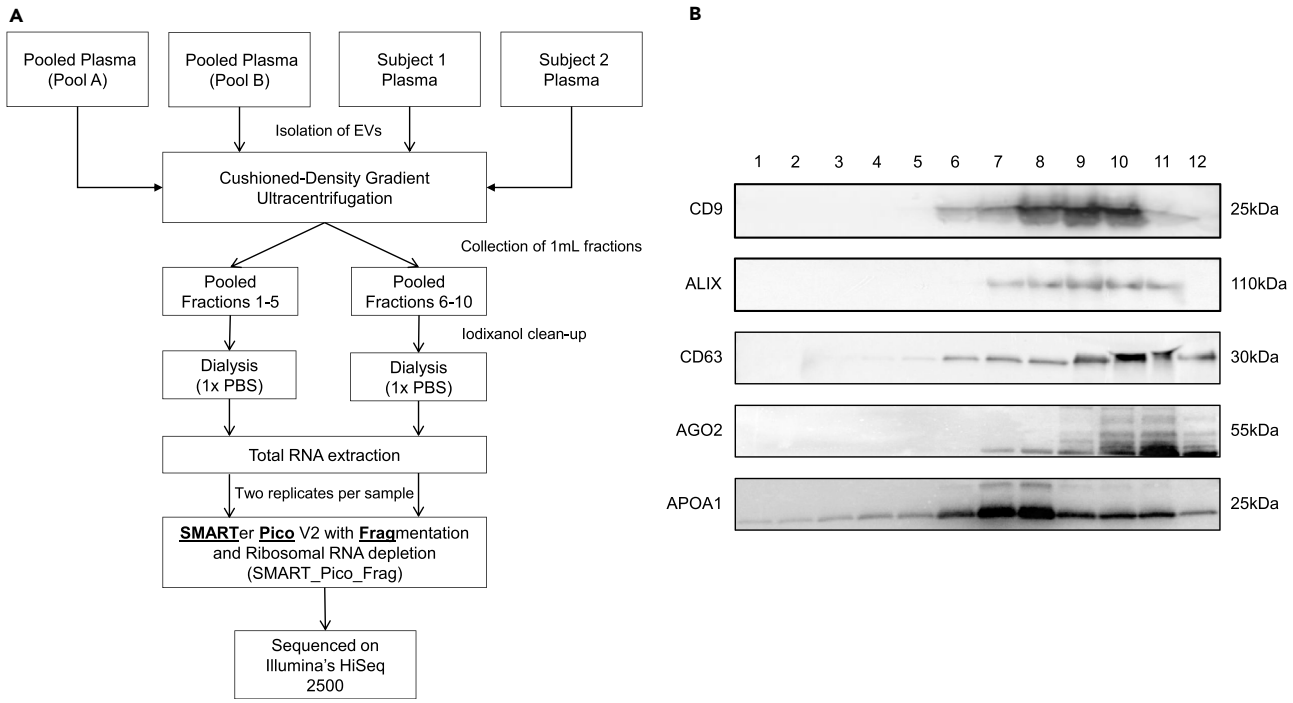


Figure 3. Isolation of Total RNA from EV- and No-EV-Enriched Plasma Fractions for RNA Sequencing

(A) Total RNA from extracellular vesicles (EVs) isolated from either pooled or individual plasma samples was used to compare gene expression between no-EV (fractions 1–5) and EV-enriched (fractions 6–10) plasma compartments. The different sizes of EVs were isolated by cushioned-density gradient ultracentrifugation using iodixanol and purified with dialysis prior to RNA extraction. RNA-sequencing libraries were constructed in duplicates using the SMARTer Pico v2 kit and subsequently sequenced on Illumina's HiSeq 2500 platform.

(B) Western blot analysis of CD9, CD63, and Alix markers for extracellular vesicles, AGO2 for Argonaute 2 proteins, and APOA1 for high-density lipoproteins in all fractions (1–12) retrieved from iodixanol cushioned-density gradient ultracentrifugation and dialysis of plasma samples.

however, we observed a reduction in the average size of particles in both fraction pools as measured by NanoSight LM10. This may be due to the removal of iodixanol, which may cause the formation of larger aggregates. We next aimed to determine any differences in genome/transcriptome mapping percentage, representation of RNA species, and gene expression between fractions enriched in EVs and those that do not contain EVs. For this, we chose the SMARTer Pico V2 with Fragmentation and Ribosomal RNA depletion (SMART_Pico_Frag) protocol due to its overall performance and ability to capture the highest diversity of long RNA transcripts. As depicted in Figure 3B, fractions 6–10 were predominantly positive for the traditional EV markers CD9, CD63, and Alix (compared with fractions 1–5) and therefore we decided to proceed with two pools of fractions (1–5 and 6–10) for the RNA-seq. It is worth noting that fractions 6–10 were positive for other RNA carriers such as the Argonaute 2 (AGO2) protein and high-density lipoproteins (HDL), which was the only marker present in all fractions (Figure 3B). HDL generally segregates with denser fractions (as for small RNA carrier HDL, 21). ApoA1, used as a marker for lipoproteins, is not specific for HDL and may be exchanged between different classes of lipoproteins, including chylomicrons and VLDLs. We did not have control of the fasting state of the individuals whose plasma comprised the pool plasma used for analysis and the western blot. For the individual samples analyzed (subject 1 and 2), the samples were collected from fasting individuals. We recommend, for future studies, that researchers collect samples in the fasting state to eliminate this potential confounder.

RNA was isolated from fractions 1–5 and 6–10 using ExoRNeasy, and the RNA was quantified for each sample in triplicate using Quant-iT Ribogreen RNA Assay according to ThermoFisher's low-range Ribogreen protocol. Fractions 6–10 consistently had more RNA than fractions 1–5 (Table S2). Analysis of the RNA-seq data showed that fractions 6–10 tended to have higher percentages of reads mapped to the genome and transcriptome across both the pooled and individual plasma samples compared with fractions 1–5 (Figures 4A and 4B). As for the RNA species between the two fraction pools, we found that the percentage of counts and TPMs were similar across the pooled and individual plasma samples (Figures 4C and 4D). In addition, fractions 6–10 showed an increase in transcript diversity, with 67,297 and 74,716 transcripts detected in the pools and subjects, respectively, as compared with 45,057 and 34,664 transcripts detected in fractions 1–5 (Table 2).

Lastly, we sought to examine any differences in the number of genes detected in fractions 1–5 and 6–10 in both the pooled and individual plasma samples. The total number of genes detected in pooled plasma was 24,180, which was very comparable to the number of genes (i.e., 24,613) detected in individual plasma samples. However, when comparing the two fraction pools (1–5 and 6–10), we found that the total number of genes detected in fractions 6–10 in both the pooled (23,941) and individual (24,513) samples was higher than the number of genes detected in the pooled (19,905) and individual (19,558) plasma samples in fractions 1–5 (Figure 5A). Although a comparable number of genes were commonly detected in both fraction pools in the pooled and individual plasma samples (19,666 and 19,458 respectively), a much higher number of genes was uniquely detected in fractions 6–10 as compared with fractions 1–5 in both the pooled (4,275 versus 239 genes) and individual samples (5,055 versus 100 genes) (Figures 5B and 5C). In pooled plasma samples, the uniquely expressed genes in fractions 6–10 represented 18% of the total number of genes, whereas in individual plasma samples, this number represented 21% of the total number of genes. Table 3 shows the number of genes and transcripts detected in fractions 1–5 and 6–10 for the pooled samples and the subject samples. GC content analysis revealed that the distribution pattern was similar for all samples. However, we observed a consistently higher number of reads in fractions 6–10 compared with fractions 1–5 in both the pooled and individual samples that was more apparent around the 50% mean GC content peak (Figure S2).

One interesting question is whether or not we can detect full gene coverage or we detect only fragments. Figure 6A shows the transcript length distribution of RNA biotypes by abundance for each of the sample fractions, and Figure 6B shows the corresponding mean transcript length, in basepairs, of the transcript length distribution shown in Figure 6A. A large number of protein coding RNA and lncRNAs are detected across a range of RNA transcript lengths. Table 4 describes the number of genes detected at > 80% coverage for transcripts of different lengths. These data help describe and highlight the decreased genes counts and increased TPMs associated with lncRNAs in the gene and transcript plots for Figures 4C and 4D. The low number of gene counts reflects the low abundance of lncRNAs compared with mRNAs in these samples. However, the increased TPMs dedicated to lncRNAs in 4D is due to the length estimate that is included for Salmon outputs; RNA fragments are detected and divided by their length, which is slightly smaller for the detected lncRNAs than the mRNA lengths observed in the distribution of Figure 6.

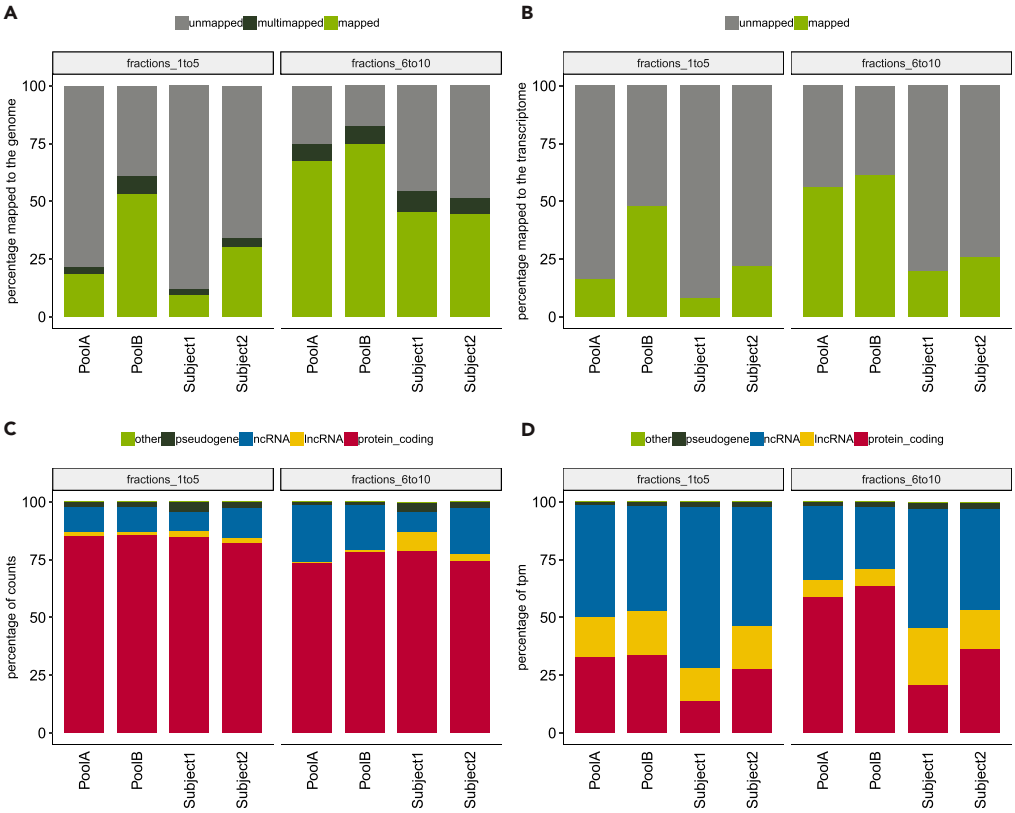


Figure 4. Assessment of Genome and Transcriptome (hg38) Mapping Percentage by RNA-Sequencing Library Preparation Kit/Condition and by RNA Species Using Total RNA from between no-EV (Fractions 1–5) and EV-Enriched (Fractions 6–10) Plasma Compartments.

The different fractions were isolated from either pooled plasma (i.e., Pool A and B) or plasma from individual subjects (i.e., Subject 1 and 2).

(A) Percentage of input reads for each RNA pool/subject and kit/condition by fractions that were uniquely mapped (green), multimapped (black), or unmapped (gray) to the human genome using STAR.

(B) Percentage of input reads for each RNA pool/subject and kit/condition by plasma compartment that were quasimapped (green) or unmapped (gray) to the human transcriptome using salmon.

(C) Percentage of uniquely mapped read counts (light green bar in panel A) assigned to the human transcriptome using featureCounts for each RNA pool/subject and kit/condition by plasma compartment and represented by the following RNA species: protein coding (red), lncRNAs (yellow), non-coding RNAs (blue), pseudogenes (black), and other RNAs (including RNAs to be experimentally confirmed, immunoglobulin genes, and T cell receptor genes; green).

(D) Percentage of transcripts per kilobase million (TPM) quasimapped to the human transcriptome (light green bar in panel B) for each RNA pool/subject and kit/condition by plasma compartment and represented by the following RNA species: protein coding (red), lncRNAs (yellow), non-coding RNAs (blue), pseudogenes (black), and other RNAs (green).

Significant pathways from IPA pathway analysis of genes unique to fractions 6–10 include Glucocorticoid Biosynthesis, the Intrinsic Prothrombin Activation Pathway, multiple pathways related to thyroid hormone metabolism, and eNOS Signaling (Table 5). The complete list of pathways from IPA pathway analysis is shown in Table S3. Last, the full list of transcripts uniquely detected in fractions 1–5 or 6–10 is provided in Tables S4 and S5, respectively.

DISCUSSION

RNA sequencing from blood offers the opportunity to develop biomarkers of health and disease using plasma or sub-compartments of plasma (e.g., extracellular vesicles), which is an easily available biofluid that can be obtained non-invasively. The lack of method standardization and reproducibility has hampered the growth of this emerging technology, especially in view of the small sample volumes typically available, different compartments in which the extracellular RNA is carried, and low quantities of RNA present in most

Sample Type	Fraction	Number of Transcripts (Mean)[a]
Pooled plasma	1 to 5	45,057
Pooled plasma	6 to 10	67,297
Individual plasma	1 to 5	34,664
Individual plasma	6 to 10	74,717

Table 2. Number of Long RNA Transcripts Detected in No-EV (1–5) and EV-Enriched (6–10) Fractions from Pooled and Individual Plasma Samples.

[a]Mean number of genes come from the duplicate runs using the SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian kit.

biofluids. Previous studies, including by our group, had focused on small RNAs known to be most abundant in the extracellular compartment. However, recent studies suggest that circulating mRNA and other long RNAs may be specific disease markers. In this study, we used plasma exRNA to rigorously compare six RNA-seq library preparation kits tailored to longer (>200 nt) RNA sequences, and we present their differences in genome and transcriptome mapping percentage as well as long RNA species diversity. In

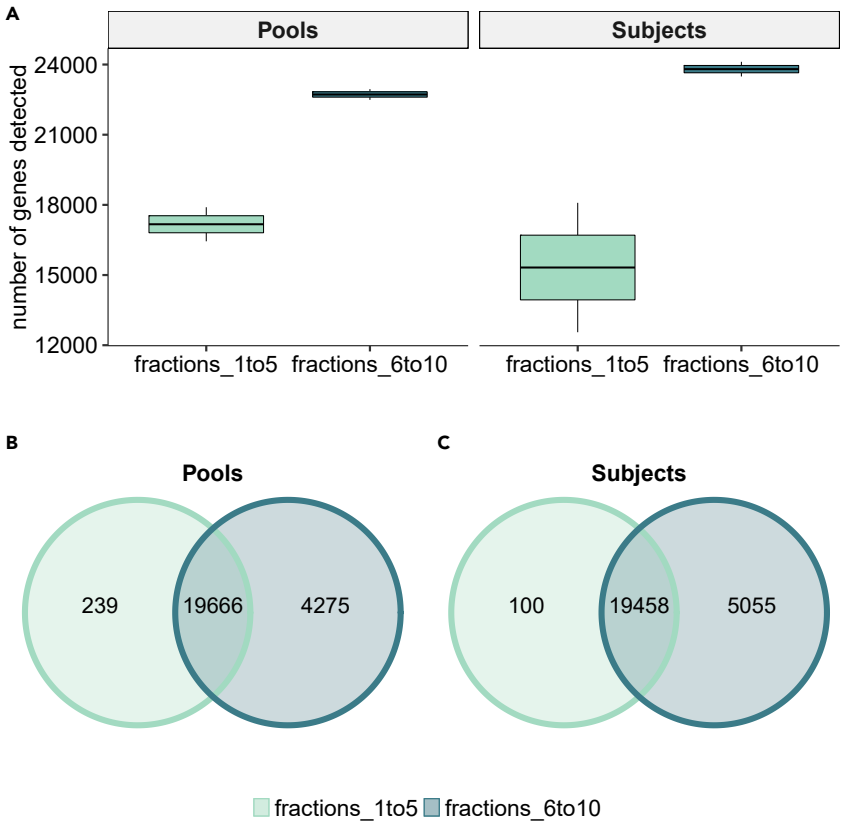


Figure 5. Evaluation of Detected Genes in Extracellular Vesicles (EVs) from Pooled or Individual Plasma Samples Using the SMARTer Pico v2 Kit.

Genes were filtered to those with an average of 10 counts across all samples and then normalized using DESeq2.

(A) Box plots showing the total number of genes detected in EV-enriched and no-EV plasma compartments by sample type (pooled versus individual plasma).

(B) Venn diagram showing the number of commonly or uniquely detected genes in no-EV (fractions 1–5) and EV-enriched (fractions 6–10) compartments from pooled plasma samples.

(C) Venn diagram showing the number of commonly or uniquely detected genes in no-EV (fractions 1–5) and EV-enriched (fractions 6–10) compartments from individual plasma samples.

	Protein Coding	lncRNA	ncRNA	Pseudogene	Other
Genes Detected Counts >10					
Pool A 1-5	10,964	2,391	87	813	63
Pool A 6-10	13,443	3,477	68	1,023	83
Pool B 1-5	12,455	3,481	134	1,187	94
Pool B 6-10	13,699	3,651	71	1,049	98
Subject 1 1-5	9,160	1,528	57	592	33
Subject 1 6-10	16,938	13,941	356	3,118	350
Subject 2 1-5	12,133	4,698	137	1,212	106
Subject 2 6-10	16,007	10,752	223	2,451	261
Transcripts Detected TPM >1					
Pool A 1-5	10,130	4,760	209	237	58
Pool A 6-10	5,485	2,084	234	150	37
Pool B 1-5	9,745	4,315	256	193	42
Pool B 6-10	5,582	2,196	270	168	38
Subject 1 1-5	8,442	4,164	125	189	42
Subject 1 6-10	12,205	12,175	494	936	191
Subject 2 1-5	9,904	5,983	260	343	69
Subject 2 6-10	9,465	8,110	376	585	121

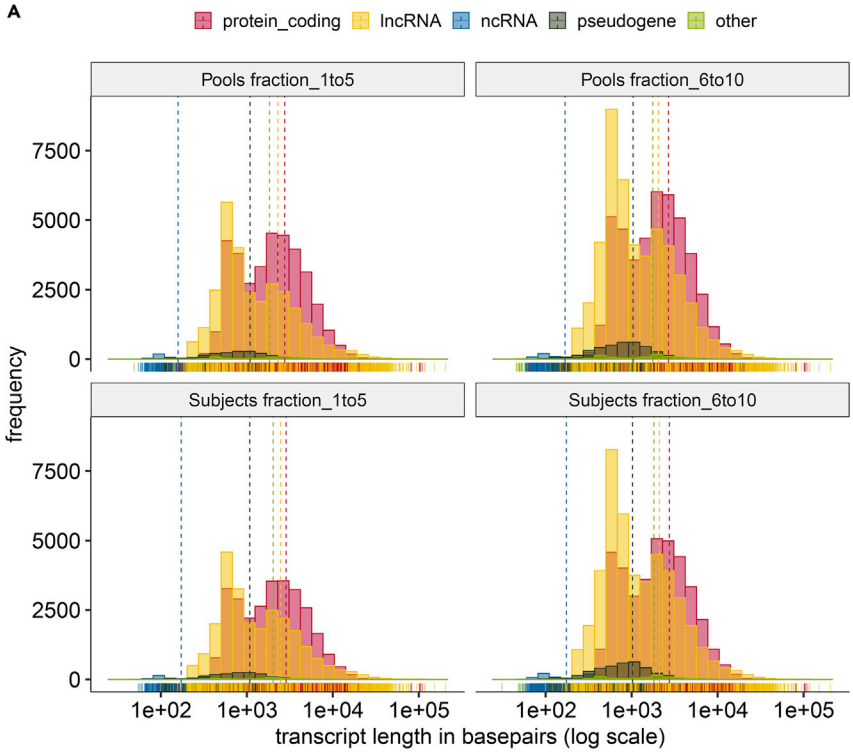
Table 3. The Number of Genes (Detected at >10 Counts) and Transcripts (Detected at >1 TPM) for Each Pooled and Subject Fraction (1–5 and 6–10).
Total genes and transcripts are broken out by RNA biotype as in Figure 4.

addition, by using the kit with the greatest demonstrated gene diversity, SMARTer Pico V2, we showed that the EV-enriched fractions (i.e, fractions 6–10) yield different genome/transcriptome mapping percentages and have a distinct gene profile than the no-EV fractions (i.e., fractions 1–5). As the kit required may differ based on the aim of any given experiment, we hope this dataset provides a reference for genome mapping percentage and long RNA species diversity in a clinically relevant biofluid, plasma.

We firstly compared the genome/transcriptome mapping percentage and long RNA species diversity of six different library preparation kits/methods by using exRNA that was extracted from the same pool of plasma. Although the percentage of mapped reads varied modestly across kits, we found that the TruSeq RNA Access kit had a significantly higher percentage of reads mapped to the transcriptome, which was expected based on the design of the kit to enrich for mapping percentage of coding RNA sequences. We also found that fragmentation did not have any impact on the performance of this kit. Although the TruSeq RNA Access kit had higher percentages of counts and TPMs in protein coding genes, the SMARTer Pico v2 and SMARTer/KapaHyper kits tended to have more representation of non-coding RNAs and lncRNAs. Thus, experiments focused on protein-coding RNA versus those focused on non-coding RNA or increased RNA species diversity might choose differing library preparation kits.

The use of exRNA is an emerging research area toward the development of diagnostics and therapeutics in health tracking and disease states (Ingenito et al., 2019; Liu et al., 2019; Murillo et al., 2019; Quinn et al., 2015). Therefore, having shown that the library preparation kits vary in terms of transcriptome mapping percentage and RNA species diversity employing cell-free exRNA from pooled plasma samples, we next focused on the evaluation of exRNA from EV-enriched or no-EV fractions. For this part, we used exRNA from pooled and individual plasma samples isolated using C-DGUC. We observed that the concentration of RNA in the EV-enriched fractions (6–10) was higher in every case: 3.4–4.7 ng in the individuals samples

A



B

biotype	Pools fraction_1to5	Pools fraction_6to10	Subjects fraction_1to5	Subjects fraction_6to10
protein_coding	2758	2681	2862	2745
lncRNA	2299	2046	2481	2104
ncRNA	159	168	173	174
pseudogene	1092	1038	1083	1025
other	1838	1763	2026	1818

Figure 6. Distribution of Transcript Lengths by RNA Biotype for Each of the Sample Fractions.

(A) Histogram of the transcript lengths, by biotype, of transcripts detected in each sample type. The mean length for each biotype is displayed as a vertical dashed line.

(B) Mean transcript length, in basepairs, of transcripts detected in each sample type broken out by biotype.

and 10.1–10.8 ng in the pooled samples compared with 1.8–2.0 ng for the patient samples and 3.0–3.2 ng for the pooled samples (in fractions 1–5). EV-enriched fractions had increased percentage of mapping to the genome and transcriptome as compared with no-EV fractions that were negative for EV markers. From these data, we might surmise that the greatest amount and diversity of long RNA species are associated with the EV fractions compared with the EV-depleted (lipoprotein-rich) fractions.

Having demonstrated that the plasma exRNA in different compartments could impact the number of reads mapping to the genome/transcriptome, we next determined the impact on the number of genes detected. As most current RNA-based biomarkers rely on specific genes, profiling of the different blood exRNA compartments is of great importance. Overall, as described above, we detected more genes in EV-enriched fractions from both pooled and individual plasma samples compared with no-EV fractions. Although 19,666 and 19,458 genes were commonly detected between the two fraction pools (1–5 vs 6–10) from pooled and individual plasma samples respectively, there were ~18x more uniquely detected genes in the EV-enriched fractions compared with the no-EV fractions from pooled samples (4,279 versus 239 genes) and ~50x more in individual samples (5,055 versus 100 genes) (Figures 5B and 5C). Thus, not only a higher percentage of reads maps to the genome/transcriptome in EV-enriched fractions but a higher number of genes can be detected as well, which means higher RNA diversity. In a recent study, Wei et al. showed that

	Fraction	200–500	501–1,000	1,001–5,000	5,001–10,000	>10,000
Pool A	1 to 5	9,877	21,906	15,252	1,252	160
Pool B	1 to 5	13,067	30,275	22,624	2,140	327
Pool A	6 to 10	15,321	37,015	28,729	2,790	348
Pool B	6 to 10	14,853	34,982	26,306	2,429	309
Subject 1	1 to 5	4,219	9,263	5,363	323	49
Subject 2	1 to 5	8,039	17,035	11,668	880	148
Subject 1	6 to 10	14,736	32,723	25,662	2,299	315
Subject 2	6 to 10	12,924	27,737	20,481	1,679	231

Table 4. Number of Transcripts Detected with >80% Coverage at Varying Transcript Lengths (bp).

different exRNA fractions isolated using ultrafiltration from human glioma stem cells had a distinctly different profile of both small and long RNAs (Wei et al., 2017). Although this study focused on exRNA from glioma stem cells and looked at different exRNA fractions than our study, it agrees with our findings that different fractions do exert different exRNA profiles and highlights the importance of the establishment of an exRNA roadmap based on the different fractions in biofluids.

RNA-seq will continue to play an integral role in the development of blood-based biomarkers. The meticulous, direct comparison of sequencing methods should provide justification based on desired RNA species detection (e.g. an unbiased RNA view versus a coding gene-only approach). This work demonstrates

Ingenuity Canonical Pathways	Ratio	Molecules	p Value
Glucocorticoid biosynthesis	0.50	CYP11B1,CYP11B2,CYP17A1,CYP21A2	0.0001
Intrinsic prothrombin activation pathway	0.17	F12,FGA,FGB,KLK11,KLK13,KLK4,KLK5	0.0005
Mineralocorticoid biosynthesis	0.43	CYP11B1,CYP11B2,CYP21A2	0.0013
Phototransduction pathway	0.14	ARR3,GNAT1,GUCA1A,GUCY2D, GUCY2F,OPN4,RHO	0.0018
Thyronamine and iodothyronamine metabolism	0.67	DIO1,DIO3	0.0035
Thyroid hormone metabolism I (via deiodination)	0.67	DIO1,DIO3	0.0035
SPINK1 pancreatic cancer pathway	0.11	CELA2A,KLK11,KLK13,KLK4,KLK5,PRSS3	0.0117
Extrinsic prothrombin activation pathway	0.19	F12,FGA,FGB	0.0166
TR/RXR activation	0.08	DIO1,DIO3,FGA,G6PC,SYT12,THRSP,TRH	0.0269
Maturity onset diabetes of Young (MODY) signalling	0.15	GCK,PDX1,SLC2A2	0.0309
Coagulation system	0.11	F12,FGA,FGB,SERPIND1	0.0324
Basal cell carcinoma signalling	0.09	BMP15,FZD10,FZD7,FZD9,KIF7,WNT4	0.0347
eNOS signaling	0.07	AQP12A/AQP12B,AQP2,AQP5,AQP8, CCNA1,CHRM1,CHRM3, CHRNA3,CHRNB3,CNGA4	0.0398
Glycine betaine degradation	0.20	BHMT2,SARDH	0.0447

Table 5. Ingenuity Pathway Analysis of Genes Uniquely Detected in EV-Enriched Fractions 6 to 10.

feasibility using different library prep kits to sequence RNA from plasma and shows that the aim of the study should dictate the choice of kit. Lastly, it demonstrates that different plasma exRNA compartments comprise of a unique RNA profile, which directly impacts detection of certain RNAs from blood circulation.

Limitations of the Study

Although we only used a small number of human samples, we were able to uniformly test each of the kits using the same starting material. We also then employed the best performing kit based on mapping percentage to genome/transcriptome and ability to detect the greatest diversity of RNA species on EV-enriched and no-EV plasma compartments to further demonstrate its performance. However, many more samples will need to be used to arrive at what should be expected from normal healthy subjects and how it varies in disease. Therefore, we are not able to recommend with high confidence a one-kit-fits-all for RNA-seq experiments in plasma. Based on our findings, each library preparation kit produces a varying genome/transcriptome mapping percentage and RNA species detection, which precludes us from doing so. We, however, recommend that different kits should be chosen depending on the goals and focus of each experiment using plasma. Furthermore, we were not able to sequence RNA from highly purified RNA carriers (e.g, CD9+ EVs versus AGO2 versus HDL) due to technical limitations and we, therefore, cannot comment on the RNA-seq performance and RNA cargo of each of the abovementioned RNA carriers. We do, however, encourage future studies to focus on the rigorous isolation of each RNA carrier and perform a systematic evaluation of RNA-seq protocols on plasma exRNA compartments for the creation of a comprehensive exRNA atlas by RNA carrier in blood.

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Saumya Das MD, PhD (sdas@mgh.harvard.edu).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The source code that generates the combined GENCODE and LNCipedia gene annotation can be accessed here: https://github.com/tgen/gencode-plus-lncipedia. The RNA sequencing data have been deposited in Dryad: https://doi.org/10.5061/dryad.kh1893236.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101182.

ACKNOWLEDGMENTS

We acknowledge funding support from the NIH Extracellular RNA Communication Consortium Common Fund grants UH3TR000901 [SD], UH3TR000891 [KVKJ], UH3TR000906 and U01HL126494 [LCL], HL126497 [IG]. We also acknowledge funding support from the American Heart Association grant 16SFRN31280008 [SD] and NIH grants R01HL122547 [SD] and R01CA218500 [IG].

AUTHOR CONTRIBUTIONS

Conceptualization, K.V.K.J., S.D.; Methodology, K.V.K.J., S.D., R.S.R., R.R.; Formal Analysis, E.H., A.Y.; Investigation, R.S.R., R.R., S.S.; Resources, K.V.K.J., S.D., L.C.L., I.G., M.G.S.; Data Curation, E.H., A.Y.; Writing—Original Draft, R.S.R., E.H., R.R.; Writing—Review & Editing, R.S.R., E.H., T.G.W., K.V.K.J., S.D., I.G., M.G.S., L.C.L.; Visualization, R.S.R., E.H., K.V.K.J., S.D.; Supervision, K.V.K.J., S.D.; Funding Acquisition, I.G., L.C.L., K.V.K.J., S.D.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 28, 2020
Revised: April 20, 2020
Accepted: May 15, 2020
Published: June 26, 2020

REFERENCES

Borad, M.J., Champion, M.D., Egan, J.B., Liang, W.S., Fonseca, R., Bryce, A.H., McCullough, A.E., Barrett, M.T., Hunt, K., Patel, M.D., et al. (2014). Integrated genomic characterization reveals novel, therapeutically relevant drug targets in FGFR and EGFR pathways in sporadic intrahepatic cholangiocarcinoma. PLoS Genet. *10*, e1004135.

Byron, S.A., Van Keuren-Jensen, K.R., Engelthaler, D.M., Carpten, J.D., and Craig, D.W. (2016). Translating RNA sequencing into clinical diagnostics: opportunities and challenges. Nat. Rev. Genet. *17*, 257–271.

Ingenito, F., Roscigno, G., Affinito, A., Nuzzo, S., Scognamiglio, I., Quintavalle, C., and Condorelli, G. (2019). The role of exo-miRNAs in cancer: a focus on therapeutic and diagnostic applications. Int. J. Mol. Sci. *20*, 4687.

Intlekofer, A.M., Joffe, E., Batlevi, C.L., Hilden, P., He, J., Seshan, V.E., Zelenetz, A.D., Palomba, M.L., Moskowitz, C.H., Portlock, C., et al. (2018). Integrated DNA/RNA targeted genomic profiling of diffuse large B-cell lymphoma using a clinical assay. Blood Cancer J. *8*, 60.

Lasser, C., Alikhani, V.S., Ekstrom, K., Eldh, M., Paredes, P.T., Bossios, A., Sjostrand, M., Gabrielsson, S., Lotvall, J., and Valadi, H. (2011). Human saliva, plasma and breast milk exosomes contain RNA: uptake by macrophages. J. Transl. Med. *9*, 9.

Liu, W., Bai, X., Zhang, A., Huang, J., Xu, S., and Zhang, J. (2019). Role of exosomes in central nervous system diseases. Front. Mol. Neurosci. *12*, 240.

McKiernan, J., Donovan, M.J., O'Neill, V., Bentink, S., Noerholm, M., Belzer, S., Skog, J., Kattan, M.W., Partin, A., Andriole, G., et al. (2016). A novel urine exosome gene expression assay to predict high-grade prostate cancer at initial biopsy. JAMA Oncol. *2*, 882–889.

Murillo, O.D., Thistlethwaite, W., Rozowsky, J., Subramanian, S.L., Lucero, R., Shah, N., Jackson, A.R., Srinivasan, S., Chung, A., Laurent, C.D., et al. (2019). exRNA atlas analysis reveals distinct extracellular RNA cargo types and their carriers present across human biofluids. Cell *177*, 463–477.e15.

Nasser, S., Kurdolgu, A.A., Izatt, T., Aldrich, J., Russell, M.L., Christoforides, A., Tembe, W., Keifer, J.A., Corneveaux, J.J., Byron, S.A., et al. (2015). An integrated framework for reporting clinically relevant biomarkers from paired tumor/normal genomic and transcriptomic sequencing data in support of clinical trials in personalized medicine. Pac. Symp. Biocomput. 56–67. https://pubmed.ncbi.nlm.nih.gov/25592568/.

Quinn, J.F., Patel, T., Wong, D., Das, S., Freedman, J.E., Laurent, L.C., Carter, B.S., Hochberg, F., Van Keuren-Jensen, K., Huentelman, M., et al. (2015). Extracellular RNAs: development as biomarkers of human disease. J. Extracell. Vesicles *4*, 27495.

Raposo, G., and Stoorvogel, W. (2013). Extracellular vesicles: exosomes, microvesicles, and friends. J. Cell Biol. *200*, 373–383.

Shah, R., Yeri, A., Das, A., Courtright-Lim, A., Ziegler, O., Gervino, E., Ocel, J., Quintero-Pinzon, P., Wooster, L., Bailey, C.S., et al. (2017). Small RNA-seq during acute maximal exercise reveal RNAs involved in vascular inflammation and cardiometabolic health: brief report. Am. J. Physiol. Heart Circ. Physiol. *313*, H1162–H1167.

Skog, J., Wurdinger, T., van Rijn, S., Meijer, D.H., Gainche, L., Sena-Esteves, M., Curry, W.T., Jr., Carter, B.S., Krichevsky, A.M., and Breakefield, X.O. (2008). Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. Nat. Cell Biol. *10*, 1470–1476.

Srinivasan, S., Yeri, A., Cheah, P.S., Chung, A., Danielson, K., De Hoff, P., Filant, J., Laurent, C.D., Laurent, L.D., Magee, R., et al. (2019). Small RNA sequencing across diverse biofluids identifies optimal methods for exRNA isolation. Cell *177*, 446–462.e16.

Wang, K., Zhang, S., Weber, J., Baxter, D., and Galas, D.J. (2010). Export of microRNAs and microRNA-protective protein by mammalian cells. Nucleic Acids Res. *38*, 7248–7259.

Wei, Z., Batagov, A.O., Schinelli, S., Wang, J., Wang, Y., El Fatimy, R., Rabinovsky, R., Balaj, L., Chen, C.C., Hochberg, F., et al. (2017). Coding and noncoding landscape of extracellular RNA released by human glioma stem cells. Nat. Commun. *8*, 1145.

Witwer, K.W., Buzas, E.I., Bemis, L.T., Bora, A., Lasser, C., Lotvall, J., Nolte-'t Hoen, E.N., Piper, M.G., Sivaraman, S., Skog, J., et al. (2013). Standardization of sample collection, isolation and analysis methods in extracellular vesicle research. J. Extracell. Vesicles *2*. https://pubmed.ncbi.nlm.nih.gov/24009894/.

Yanez-Mo, M., Siljander, P.R., Andreu, Z., Zavec, A.B., Borras, F.E., Buzas, E.I., Buzas, K., Casal, E., Cappello, F., Carvalho, J., et al. (2015). Biological properties of extracellular vesicles and their physiological functions. J. Extracell. Vesicles *4*, 27066.

Yeri, A., Courtright, A., Danielson, K., Hutchins, E., Alsop, E., Carlson, E., Hsieh, M., Ziegler, O., Das, A., Shah, R.V., et al. (2018). Evaluation of commercially available small RNASeq library preparation kits using low input RNA. BMC Genomics *19*, 331.

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

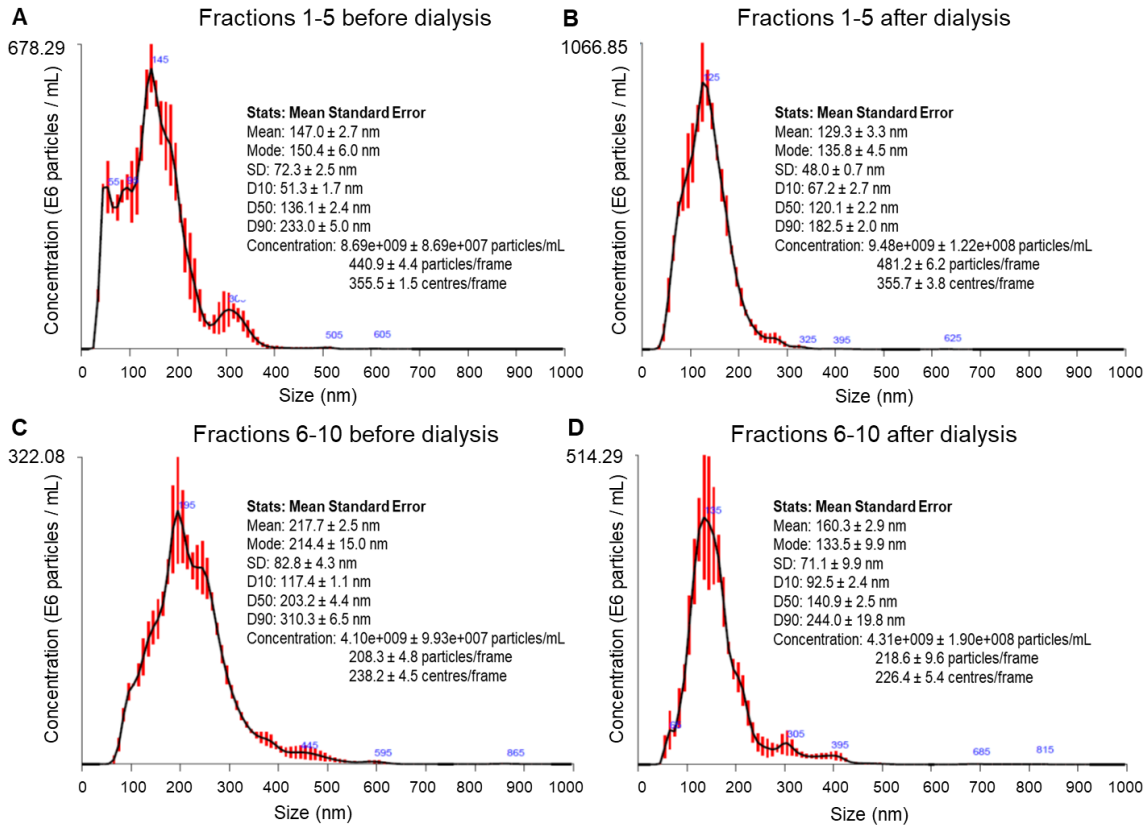


Figure S1. Related to Figure 3 and Transparent Methods. Nanoparticle Tracking Analysis of EVs before and after dialysis using the NanoSight LM10. (A) Fractions 1-5 before dialysis, (B) Fractions 1-5 after 24 hours of dialysis in 1X PBS buffer, (C) Fractions 6-10 before dialysis, (B) Fractions 6-10 after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

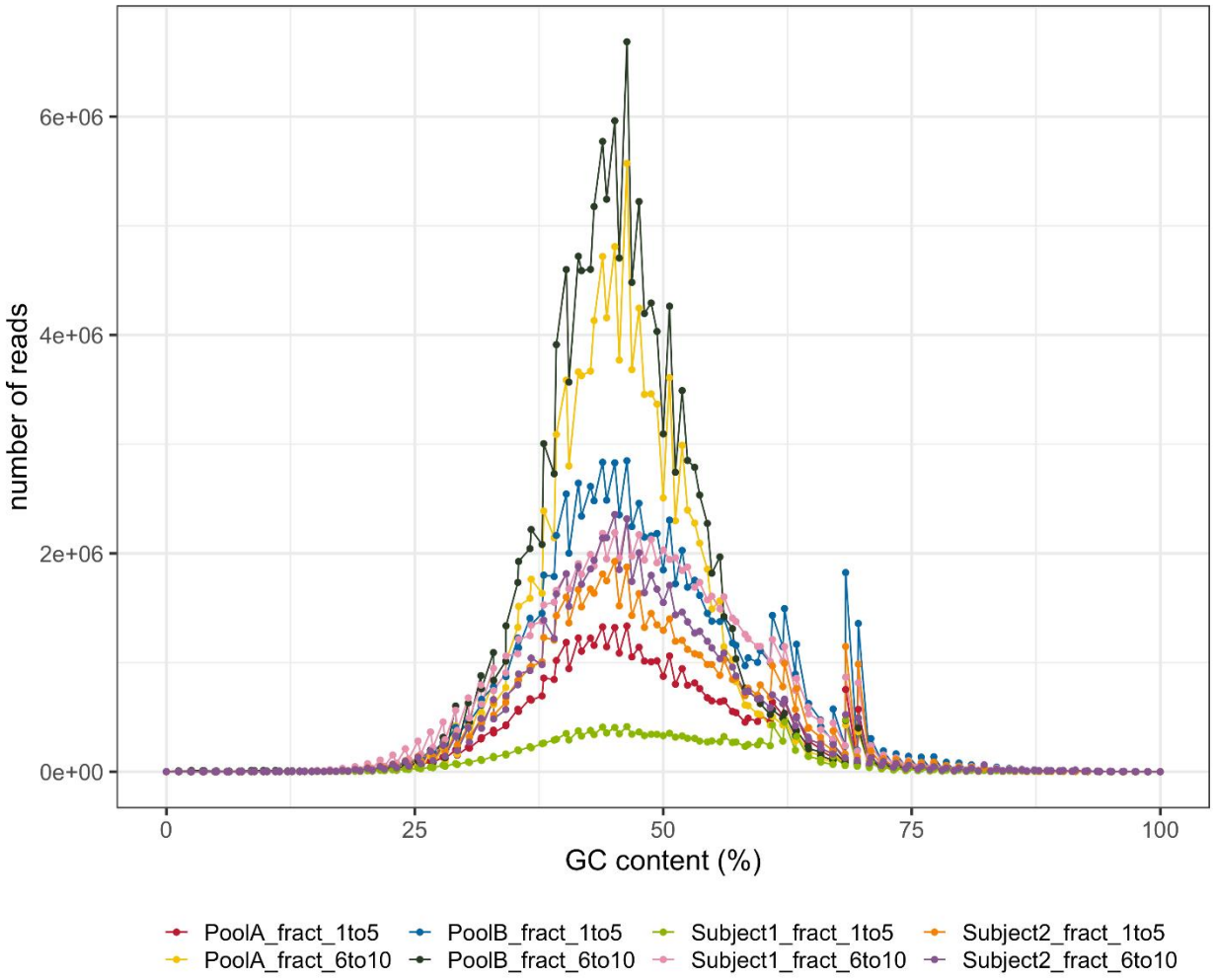


Figure S2. Related to Figure 5, Table 3 and Transparent Methods. Read GC content distribution of long RNA transcripts detected in fractions 1-5 and 6-10 for both individual and pooled samples.

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.