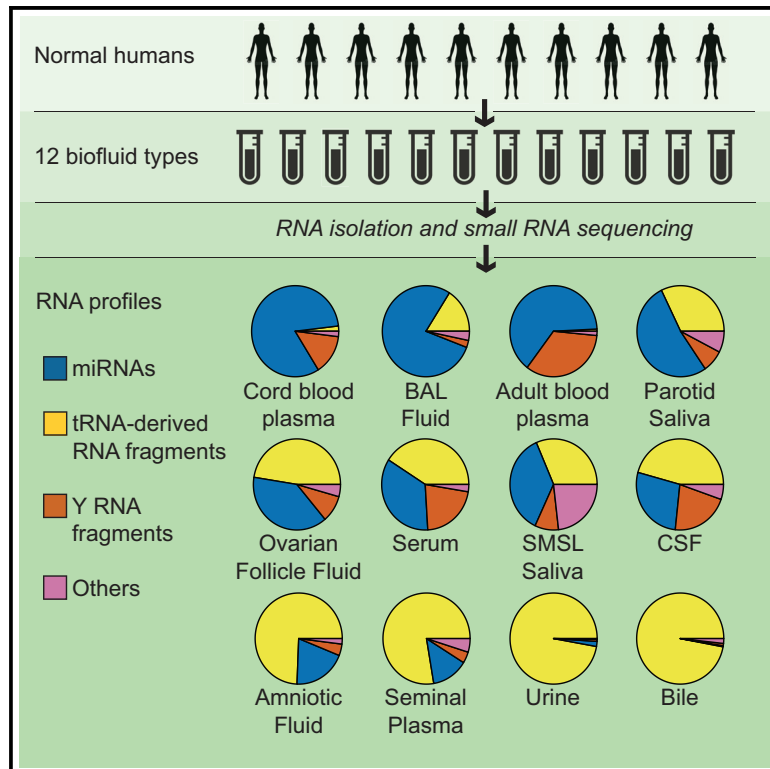


Cell Reports

Large Differences in Small RNA Composition Between Human Biofluids

Graphical Abstract



Authors

Paula M. Godoy, Nirav R. Bhakta, Andrea J. Barczak, ..., James F. Smith, Prescott G. Woodruff, David J. Erle

Correspondence

david.erle@ucsf.edu

In Brief

Using a standardized sequencing-based approach, Godoy et al. characterize small RNAs in 12 normal human biofluids. They find that each biofluid contains an extensive collection of small RNAs that belong to multiple biotypes. The relative abundance of these RNAs varies widely between biofluids.

Highlights

- RNA-seq of 12 human biofluids shows all have diverse small RNA repertoires
- Relative abundance of miRNA to tRNA-derived RNA differs markedly across biofluids
- Biofluid miRNA profiles are generally similar, but some miRNAs vary widely
- There are large differences in tRNA-derived RNAs between biofluids

Data and Software Availability

GSE112343



Large Differences in Small RNA Composition Between Human Biofluids

Paula M. Godoy,¹ Nirav R. Bhakta,² Andrea J. Barczak,¹ Hakan Cakmak,³ Susan Fisher,³ Tippi C. MacKenzie,⁴ Tushar Patel,⁵ Richard W. Price,⁶ James F. Smith,⁷ Prescott G. Woodruff,^{2,8} and David J. Erle^{1,8,9,*}

¹Lung Biology Center, University of California, San Francisco, UCSF Box 3118, San Francisco, CA 94143, USA

²Division of Pulmonary, Critical Care, Allergy and Sleep Medicine, Department of Medicine, University of California, San Francisco, UCSF Box 0130, San Francisco, CA 94143, USA

³Department of Obstetrics, Gynecology & Reproductive Sciences, University of California, San Francisco, UCSF Box 0916, San Francisco, CA 94143, USA

⁴Department of Surgery, University of California, San Francisco, UCSF Box 0570, San Francisco, CA 94143, USA

⁵Department of Transplantation, Mayo Clinic, 4500 San Pablo Road, Jacksonville, FL 32224, USA

⁶Department of Neurology, University of California, San Francisco, UCSF Box 0870, San Francisco, CA 94143, USA

⁷Department of Urology, University of California, San Francisco, UCSF Box 1695, San Francisco, CA 94143, USA

⁸Cardiovascular Research Institute, University of California, San Francisco, UCSF Box 0130, San Francisco, CA 94143, USA

⁹Lead Contact

*Correspondence: david.erle@ucsf.edu

<https://doi.org/10.1016/j.celrep.2018.10.014>

SUMMARY

Extracellular microRNAs (miRNAs) and other small RNAs are implicated in cellular communication and may be useful as disease biomarkers. We systematically compared small RNAs in 12 human biofluid types using RNA sequencing (RNA-seq). miRNAs and tRNA-derived RNAs (tDRs) accounted for the majority of mapped reads in all biofluids, but the ratio of miRNA to tDR reads varied from 72 in plasma to 0.004 in bile. miRNA levels were highly correlated across all biofluids, but levels of some miRNAs differed markedly between biofluids. tDR populations differed extensively between biofluids. Y RNA fragments were seen in all biofluids and accounted for >10% of reads in blood plasma, serum, and cerebrospinal fluid (CSF). Reads mapping exclusively to Piwi-interacting RNAs (piRNAs) were very rare, except in seminal plasma. These results demonstrate extensive differences in small RNAs between human biofluids and provide a useful resource for investigating extracellular RNA biology and developing biomarkers.

INTRODUCTION

RNAs released from cells have been detected in many biofluids (Patton et al., 2015). Although some reports suggest that large RNAs, including functional mRNAs, can be present in biofluids (Ni et al., 2002; Skog et al., 2008), most extracellular RNAs (exRNAs) are small RNAs (Hoy and Buck, 2012). Many reports have focused on microRNAs (miRNAs), which are small (typically 21–22 nt) RNAs produced by intracellular processing of larger precursor RNAs (Argyropoulos et al., 2013; Lee et al., 2010; Mitchell et al., 2008; Weber et al., 2010; Williams et al., 2013).

Several pathways have been proposed to lead to release of miRNAs from cells, and extracellular miRNAs can be found within exosomes and in complexes containing argonaute proteins (Arroyo et al., 2011) or lipoproteins (Patton et al., 2015). Extracellular miRNAs can enter cells and may target mRNAs in those cells (Patton et al., 2015). The potential of miRNAs as disease biomarkers is being explored by many investigators (Argyropoulos et al., 2013; Barger et al., 2016; Gray et al., 2017; Mitchell et al., 2008).

Small RNA biotypes other than miRNAs have also been detected in biofluids. Piwi-interacting RNAs (piRNAs) are ~23- to 30-nt RNAs involved in transcriptional and post-transcriptional silencing of transposons and other targets in germ cells (Czech and Hannon, 2016). Sequences mapping to piRNAs have been reported in human seminal fluid plasma (Hong et al., 2016) as well as in blood plasma (Freedman et al., 2016), saliva (Bahn et al., 2015), and urine (Yeri et al., 2017). Many small RNAs found in biofluids are thought to be derived from larger RNAs by specific processing events or by nonspecific degradation. tRNA-derived RNAs (tDRs) are generated by cleavage of tRNAs at specific sites (Gebetsberger and Polacek, 2013). The variety of tDRs is extensive since there are many (>600) human tRNA genes, and multiple fragment types, including 5' halves, 3' halves, 5' tRNA fragments (tRFs), 3' tRFs, and internal tRFs (i-tRFs), have been identified (Loher et al., 2017). Small RNAs derived from mRNAs, long non-coding RNAs, rRNAs (Semenov et al., 2004), and Y RNAs (Dhahbi et al., 2013) have also been detected in some biofluids. Their functions are not yet well understood.

Extracellular RNAs have been detected in at least 15 biofluids to date (Sohel, 2016). The total concentration of RNA varies widely between biofluids, with certain biofluids such as breast milk and seminal fluid being more concentrated than more dilute biofluids like cerebrospinal fluid (CSF) and urine (Weber et al., 2010). Given the difficulties inherent in absolute quantification of exRNAs, most analyses have focused on relative quantification of specific RNAs. This approach has identified some



Table 1. Study Participant Demographic Summary

Biofluid	Sex	Age (y) ^a	Race					Ethnicity						
			African American	Asian	White	Other (mixed)	Unknown	Pacific Islander	Hispanic	Non-Hispanic	Declined to State	Unknown		
Amniotic fluid	10F	Unknown	-	-	-	-	-	-	-	-	-	-	-	10
BAL fluid	4F, 6M	29.5 (28.3–33.0)	3	1	6	-	-	-	1	9	-	-	-	-
Bile	6F, 6M	57.5 (53.3–60.3)	-	-	12	-	-	-	-	11	-	-	-	-
Cord blood plasma (maternal demographics)	10F	31 (29.5–33.5)	1	-	4	2	2	2	3	5	4	-	-	2
Cord blood plasma (fetal demographics)	6F, 4M	0 (0–0)	1	2	2	3	2	2	3	4	4	-	-	3
CSF	3F, 8M	40 (37.5–43.0)	6	1	4	-	-	-	-	-	11	-	-	-
Ovarian follicle fluid	5F	28 (26–28)	-	2	3	-	-	-	-	-	5	-	-	-
Plasma	9F, 3M	26 (24.8–35.3)	-	7	4	-	-	-	1	2	10	-	-	-
Serum	8F, 3M	26 (24.5–30.0)	-	7	3	-	-	-	1	1	10	-	-	-
Urine	8F, 2M	27.5 (25.3–43.8)	-	7	2	-	-	-	-	1	9	-	-	-
Parotid saliva	8F, 5M	29 (24–30)	-	3	8	-	-	-	2	2	-	-	-	11
Submandibular and sublingual saliva	8F, 7M	29 (27–33)	-	3	10	-	-	-	2	2	-	-	-	13
Seminal fluid	10M	40.5 (33.8–40.5)	1	3	6	-	-	-	-	1	9	-	-	-

F, female; M, male. Dashes indicate no data.

^aMedian (interquartile range).

differences in small RNA content between biofluid types. For example, a PCR-based study of 714 miRNAs identified certain miRNAs that were abundant in most of the 12 biofluids studied along with other miRNAs that were enriched in specific biofluids (Weber et al., 2010). However, a more complete understanding of small RNA differences between biofluids is problematic, since published studies rely on a diverse set of methods with different biases that preclude direct comparisons. Furthermore, most studies have focused primarily or exclusively on miRNAs, and much less information is available about other RNA biotypes.

One objective of the NIH Extracellular RNA Communication Consortium (ERCC) is to identify the range of RNAs present in human biofluids. To address this goal, we compared the small RNA populations of a large and diverse collection of human biofluids using one standard RNA sequencing (RNA-seq) approach that was validated as part of a multicenter ERCC study (Giraldez et al., 2017). This approach was designed for miRNAs but can also detect other small RNAs with a 5' phosphate and a 3' hydroxyl group. Our analysis of a total of 129 samples of 12 biofluid types from human donors reveals the presence of complex RNA repertoires in all biofluids and major differences in RNA composition between biofluid types. These results are publicly available through the exRNA Atlas (<https://exrna-atlas.org>).

RESULTS

Characteristics of the Study Population

We obtained samples of 12 different biofluid types (Table 1). With the exception of bile samples, which were obtained from patients who had previously undergone liver transplantation and had intact liver function, all samples were obtained from healthy subjects. For each biofluid, we obtained 5–15 samples. In general, each sample of a given biofluid was obtained from a different participant, except that the 10 ovarian follicle fluid samples were obtained from five participants who each provided two samples. Plasma, serum, and urine samples were obtained from a separate cohort of 12 individuals who each provided samples of each of these three biofluids. For saliva, another cohort of 15 participants provided samples of both parotid saliva and submandibular and sublingual (SMSL) saliva. For amniotic fluid, bronchoalveolar lavage (BAL) fluid, bile, cord blood plasma, CSF, and seminal fluid, 10–12 participants from separate cohorts each provided a single sample of one biofluid.

Sequencing and Quality Control

RNA-seq reads were aligned to the human transcriptome using the Genboree/exceRpt pipeline (<http://www.genboree.org/index.html>). This approach resulted in a substantial proportion of read counts assigned to piRNAs in adult blood plasma (median 13% assigned to piRNA), cord blood plasma (5%), and serum (6%) samples. Large proportions of reads that mapped to piRNA in these three biofluids mapped to a single piRNA sequence (piRNAbase hsa_piR_016658, GenBank: DQ592931; 79% of piRNA reads in adult blood plasma, 48% in cord blood plasma, and 43% in serum). However, these reads also map to a Y RNA (RNY4), and each of these biofluids also contains a large

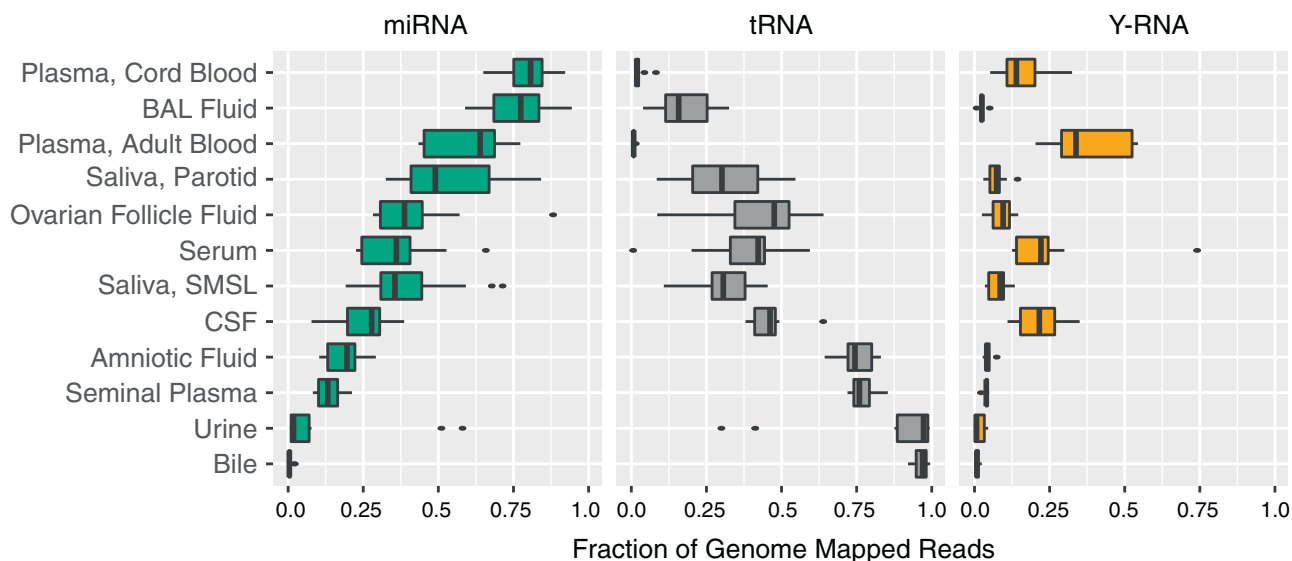


Figure 1. Distribution of RNA Biotypes Differs between Biofluids

Reads mapping to miRNAs, tRNAs, Y-RNAs, piRNAs, mRNAs, or other RNA biotypes as a fraction of total reads mapping to the human transcriptome. Boxes represent median and interquartile ranges, whiskers represent 1.5 times the interquartile range, and dots represent outliers.

number of other reads that map to other regions of RNY4. In the aggregate, >99.7% of plasma or serum reads mapping to piRNAs also mapped to RNA(s) from a different biotype. Based upon these observations, we changed from the default order of read assignment to align to Gencode RNAs (which include Y RNAs) before aligning to piRNAs. Using this approach, 0.3% of seminal plasma reads mapped to piRNAs, whereas <0.1% of reads in other biofluids mapped to piRNAs.

A total of 145 samples were analyzed. 16 samples failed quality control despite repeat analysis (see [Experimental Procedures](#) for description of quality control criteria). The remaining 129 samples were used for the analyses presented here. For these samples, the mean number of reads that passed quality control was 12.8×10^6 , and mean reads exceeded 7×10^6 per sample for each biofluid type ([Table S1](#)). The proportions of reads that were shorter than 18 nt (no mapping attempted) and reads mapped to rRNA differed between biofluid types ([Table S1](#)). In three biofluid types (parotid and submandibular and sublingual saliva and CSF), the proportion of reads that mapped to the human transcriptome was relatively low. Saliva contains a relatively high proportion of reads that map to bacterial genomes ([Yeri et al., 2017](#)). The basis of the low CSF alignment rate is unclear but our miRNA mapping rate for CSF (2.5% of reads passing quality control) was similar to the 1.5% rate reported in a previous study of CSF from healthy controls and individuals with amyotrophic lateral sclerosis ([Waller et al., 2018](#)). 54.7% of CSF reads aligned to genomes of other species. These included bacterial (11.3%), fungal (7.6%), and archaeal (0.05%) sequences as well as many sequences that could not be unambiguously assigned even at the kingdom level. These non-human sequences likely represent contaminants that were relatively frequent in normal CSF libraries due to a low concentration of exRNAs in this biofluid.

Relative Abundances of RNA Biotypes Differ Widely between Biofluids

The fraction of transcriptome-aligned reads mapping to different biotypes of human RNAs varied markedly between biofluid types ([Figure 1](#); [Table S1](#)). The median rate of mapping to miRNAs was >50% for adult and cord blood plasma and BAL fluid, whereas tDRs represented >50% of reads in bile, urine, seminal plasma, and amniotic fluid. miRNAs and tDRs each accounted for >25% of reads in the remaining biofluids (parotid saliva and submandibular and sublingual saliva, ovarian follicle fluid, serum, and CSF). The relative abundance of miRNA to tRNA reads varied by $>10^3$ -fold (from 72 in plasma to 0.004 in bile). Y RNA fragments represented >10% of mapped reads in adult and cord blood plasma, serum, and CSF but <0.8% of reads in urine and bile, with intermediate levels in other biofluids. miRNAs, tRNAs, and Y RNAs accounted for >90% of all mapped reads except in submandibular and sublingual saliva, which had the largest proportion of reads mapping to portions of protein coding genes (mRNAs, 7.6%), small nuclear RNAs (snRNAs) (2.3%), retained introns (1.5%), and Gencode “processed transcripts” (1.2%). This may reflect the presence of RNA degrading enzymes, cellular debris, or microbial-derived small RNAs that map to the human genome in saliva samples. Reads that mapped to piRNAs, but not to miRNAs, tRNAs, Y RNAs, or other Gencode transcripts, represented 0.30% of reads in seminal plasma and <0.10% of reads in other biofluids. These results demonstrate that multiple classes of small RNAs are represented in each biofluid type studied, but the relative abundance of these classes varies widely between biofluids.

miRNA Profiles

In each biofluid, we detected hundreds of miRNAs, but a small number of miRNAs accounted for a large proportion of miRNA

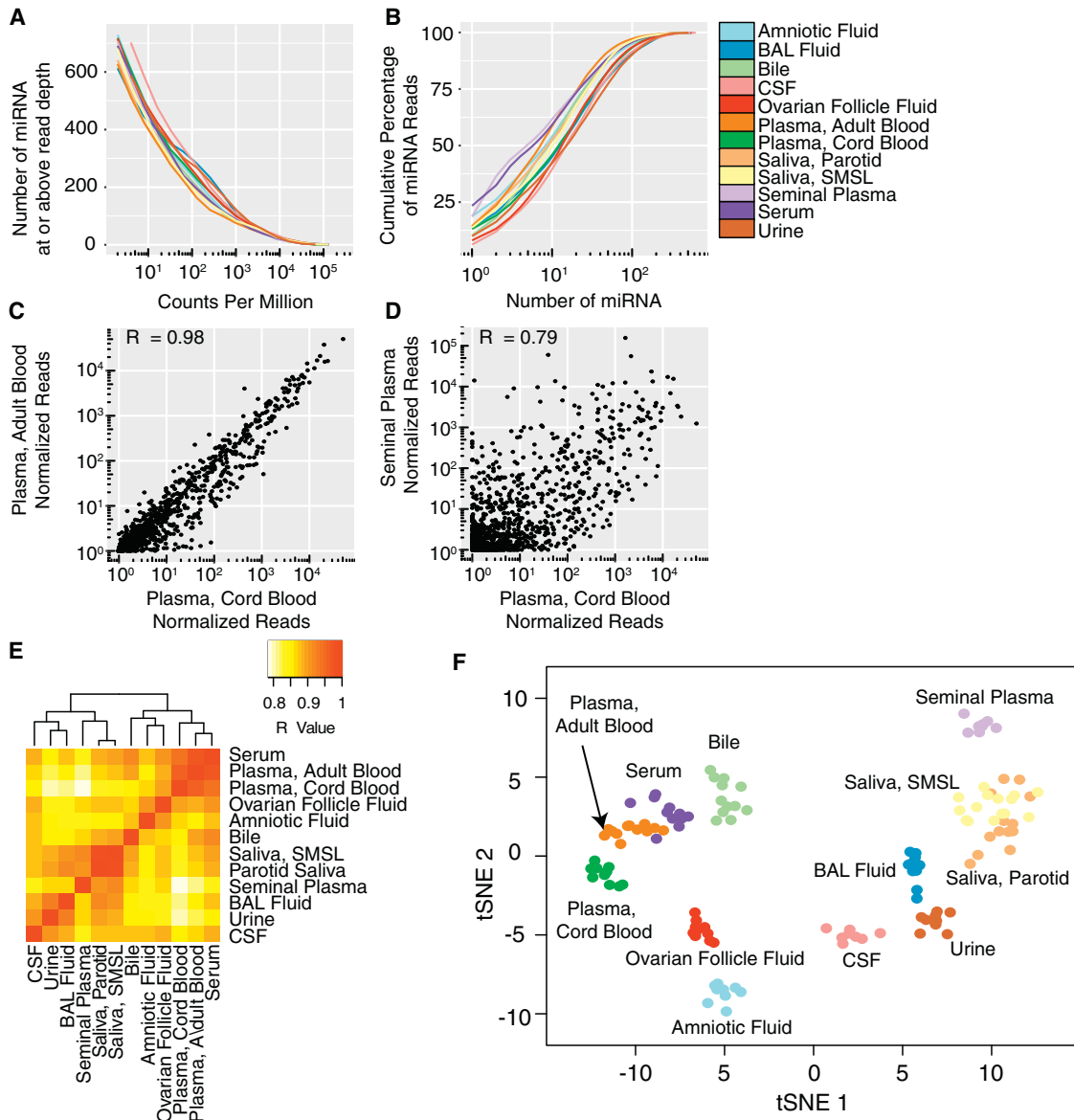


Figure 2. miRNA Profiles in 12 Biofluid Types

(A) Number of miRNAs detected as a function of read depth.

(B) Cumulative distribution of miRNA reads. See also [Figure S1](#).

(C and D) Examples of pairwise correlations between biofluids for cord blood plasma versus adult blood plasma (C) and cord blood plasma versus seminal plasma (D). Each point represents the median normalized read count for a single miRNA for the indicated biofluids. One normalized read count was added to each measurement to allow representation of log read counts for miRNAs with no reads.

(E) Correlations for all pairs of biofluids.

(F) tSNE plot produced using miRNA read counts. Each point represents a single biofluid sample.

read counts ([Figures 2A, 2B, and S1](#); [Table S2](#)). Between 395 (parotid saliva) and 541 (CSF) miRNAs had a median of ≥ 10 reads per million total miRNA reads in each biofluid. The 10 most frequent miRNAs represented between 39% (CSF) and 62% (serum) of total miRNA reads ([Table S3](#)). Despite marked differences in small RNA classes between biofluids, pairwise correlation coefficients for miRNA read counts between biofluids were generally high ([Figures 2C–2E](#)). Correlations were highest

between blood-derived biofluids (plasma, cord plasma, and serum; $R = 0.94–0.98$) and between saliva samples from the two different sites ($R = 0.98$). Seminal fluid and cord blood plasma were the least correlated ($R = 0.79$). Using all 2,153 miRNAs detected in any sample, transfer stochastic neighbor embedding (tSNE) analysis revealed that samples of most biofluid types formed distinct clusters ([Figure 2F](#)). Samples of saliva from two different sites (parotid and submandibular and

sublingual glands) formed overlapping clusters and two serum samples could not be clearly distinguished from the cluster of adult plasma samples. Therefore, although correlations in miRNA levels between biofluids were high, each biofluid (with the exception of the saliva samples from two different sites) had a distinct pattern of miRNAs.

To validate the ability of our RNA-seq method to detect differences in miRNA levels between biofluids, we used quantitative PCR (qPCR) to measure 103 miRNAs in independent samples of two biofluid types: adult blood plasma and BAL fluid (Table S4). qPCR and RNA-seq results were correlated ($R = 0.72$, $p < 2.2 \times 10^{-16}$; Figure S2; Table S4). Of these 103 miRNAs, 38 were at least 2-fold more abundant in one of these two biofluids compared with the other by RNA-seq (false discovery rate [FDR] < 0.05). For 28 of these 38 miRNAs, concordant and statistically significant (FDR < 0.05) differences were also found by qPCR.

A total of 15 miRNAs had much higher relative abundance in one biofluid than any other biofluid ($\geq 10^3$ reads/ 10^6 total miRNA reads, >10 -fold higher in one biofluid than all other biofluids, adjusted $p < 0.05$ for pairwise comparisons with all other biofluids by negative binomial Wald test; Table S5). In some cases, these levels are likely explained by high expression of the miRNAs by cell types that are in direct contact with the biofluid. Three of the four miRNAs with higher levels in amniotic fluid (miR-483-5p, miR-1247-5p, and miR-433-3p) are highly enriched in extraembryonic cells (amniotic epithelial cells, placental epithelial cells, or chorionic membranes) (de Rie et al., 2017). The three miRNAs with much higher levels in CSF are miR-9-3p, which is highly enriched in the vertebrate nervous system; miR-1911-5p, a brain-specific miRNA that was detected in CSF exosomes, but not blood plasma exosomes (Yagi et al., 2017); and miR-1298-5p, which is among the most abundant miRNAs in CSF exosomes (Yagi et al., 2017). miR-891a, which had much higher expression in seminal plasma, has been reported to be among the most abundant miRNAs in epididymis (Li et al., 2012). Combining results all blood-derived biofluids (adult and cord blood plasma and serum) into a single group and comparing with results each of the other biofluids did not identify any miRNAs that met the above criteria for much higher expression in blood-derived biofluids versus all other biofluids. Similarly, combining results from parotid and submandibular and sublingual saliva samples did not identify any miRNAs with much higher expression in saliva versus all other biofluids.

A comparison of umbilical cord blood plasma and adult blood plasma revealed that 18 miRNAs differed in relative abundance (Table S6). Three miRNAs (miR-487b-3p, miR-376c-3p, and miR-127-3p) were >5 -fold higher in cord blood plasma. A previous report detected similar differences between cord and adult blood plasma and showed relatively high levels of each of these miRNAs in placenta (Williams et al., 2013). One miRNA, let-7b-5p, was >10 -fold lower in cord blood plasma; this large difference was also seen previously (Williams et al., 2013).

We identified six groups of five or more miRNAs with similar abundance patterns across biofluids using Bayesian relevance network analysis (Ramachandran et al., 2017) (Figure 3). The largest group (group 1) contained 45 miRNAs. A subgroup (1A) containing 31 of these 45 miRNAs had levels that were highest

in amniotic fluid and cord blood plasma. All 31 subgroup 1A miRNAs are derived from the 14q32 cluster, which is the largest miRNA cluster in the human genome (54 miRNAs) (Hill et al., 2017). Subgroup 1B contained 14 miRNAs that were highest in cord blood plasma and adult blood plasma and serum but low in amniotic fluid. These include two pairs of miRNAs produced from the same pre-miRNA (miR-486-5p and miR-486-3p and miR-126-5p and miR-126-3p) and three miRNAs produced from a cluster on chromosome 17 (miR-451a, miR-144-3p, and miR-4732-3p). Group 2 (8 miRNAs) includes 5 miRNAs from the X chromosome miR-506-514 cluster; these miRNAs were highest in ovarian follicle fluid and were also relatively abundant in seminal plasma. Group 3 (23 miRNAs) was enriched for miRNAs derived from the two miR-200 family clusters, miRs-200b/a/429 and miRs-200c/141. The miR-200 family has important roles in epithelial cells (Korpál and Kang, 2008) and miRNAs in this group were most abundant in seminal plasma and saliva and least abundant in blood-derived biofluids. Group 4 comprises 7 miRNAs that were relatively abundant in bile, and group 5 comprises 5 miRNAs that were relatively abundant in CSF. Group 6 (23 miRNAs) included a subgroup of 17 miRNAs with highest relative abundance in urine and a second subgroup with 6 members of the miR-34/449 family that were most abundant in CSF, BAL, and amniotic fluid.

We were also interested in exploring the relationship between miRNA levels in biofluids and miRNA levels in organs and tissues. Analysis of data from a previous miRNA microarray study (Ludwig et al., 2016) showed that pairwise correlations for miRNA levels in nine different tissues or organs that are in contact with the biofluids we studied, like correlations between biofluids, were also high ($R = 0.73$ – 0.92 ; Figure S3). Since technical differences between small RNA-seq and microarrays precluded direct comparisons between these two datasets, we used our small RNA-seq method to analyze RNA from one human organ: the brain. Brain miRNA levels were correlated with our biofluids (R ranging from 0.63 in seminal fluid to 0.74 in CSF; Figure S3; Table S2). These data are consistent with the hypothesis that intracellular miRNA levels are one major determinant of extracellular miRNA levels. Further ERCC-supported studies of relationships between exRNA levels and levels of RNAs in different organs, tissues, and primary cell types are ongoing.

tDR Profiles

Sequences aligning to tRNAs were detected in all biofluids, although the frequencies of tDR reads varied considerably between biofluids (Figure 1). Assigning tDRs to the human genome and transcriptome is challenging, since many tDRs map to multiple tRNAs corresponding to the same anticodon, a different anticodon for the same amino acid, or even different amino acids. The Genboree pipeline aggregates tDR reads by amino acid. To analyze tDR reads in more detail, we used MINTmap (Loher et al., 2017), a tDR analysis tool that counts each unique sequence, including those that differ in length by as little as 1 nt, separately. Using this approach, between 954 (bile) and 4,997 (parotid saliva) tDRs had a median of ≥ 10 reads per million total tDR reads in each biofluid (Figures 4A and S4; Table S2). The 10 most frequent tDR reads represent between 18% (adult blood plasma) and 77% (BAL) of total tDR reads (Figure 4B). Pairwise

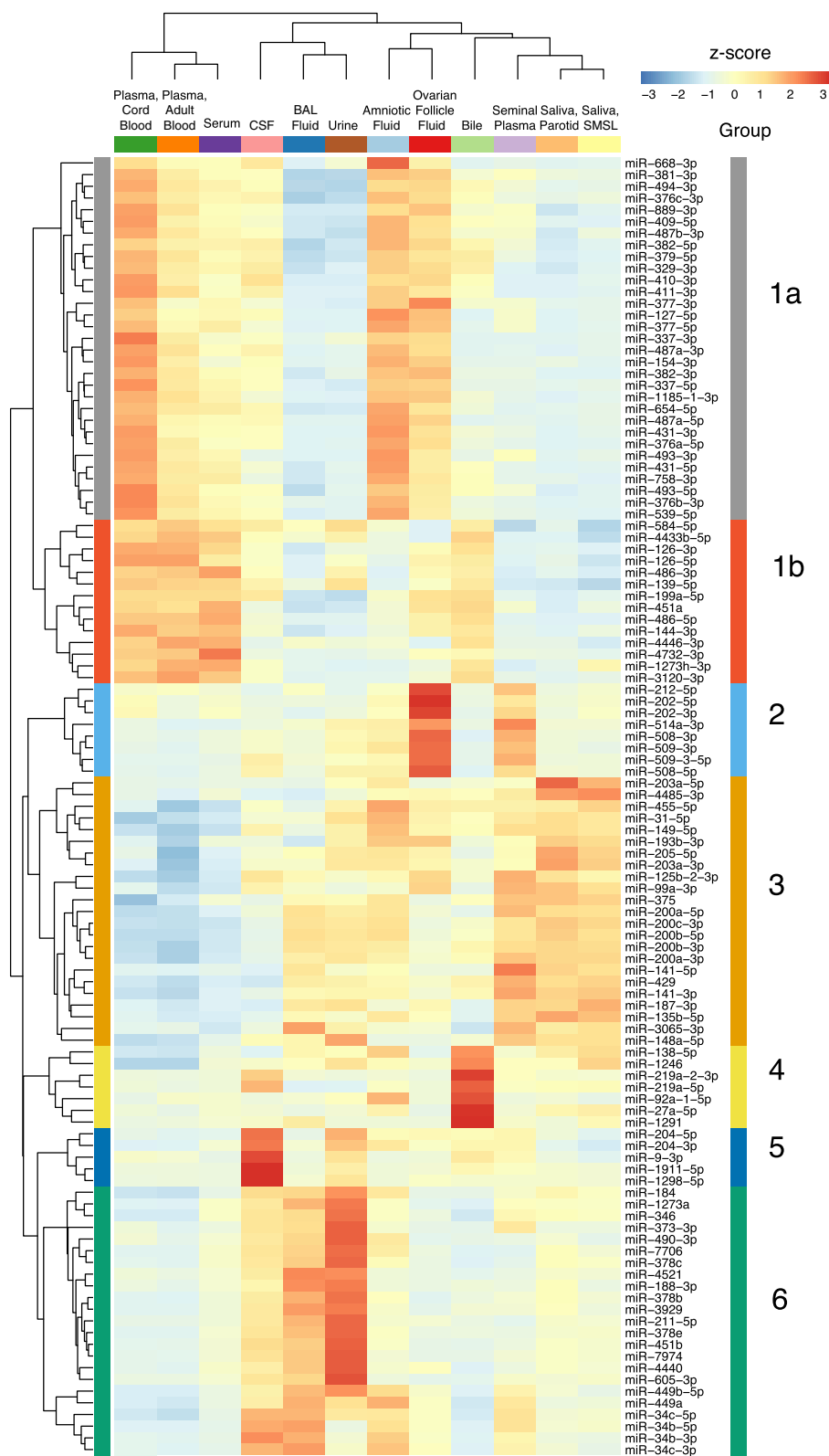


Figure 3. miRNAs with Highly Correlated Read Counts across 12 Biofluids

Hierarchical clustering heatmap depicting scaled miRNA read counts for six groups (1–6) of five or more miRNAs with similar abundance patterns across biofluids using Bayesian relevance network analysis. Z scores indicate levels of miRNA relative to levels of the same miRNA in other biofluids.

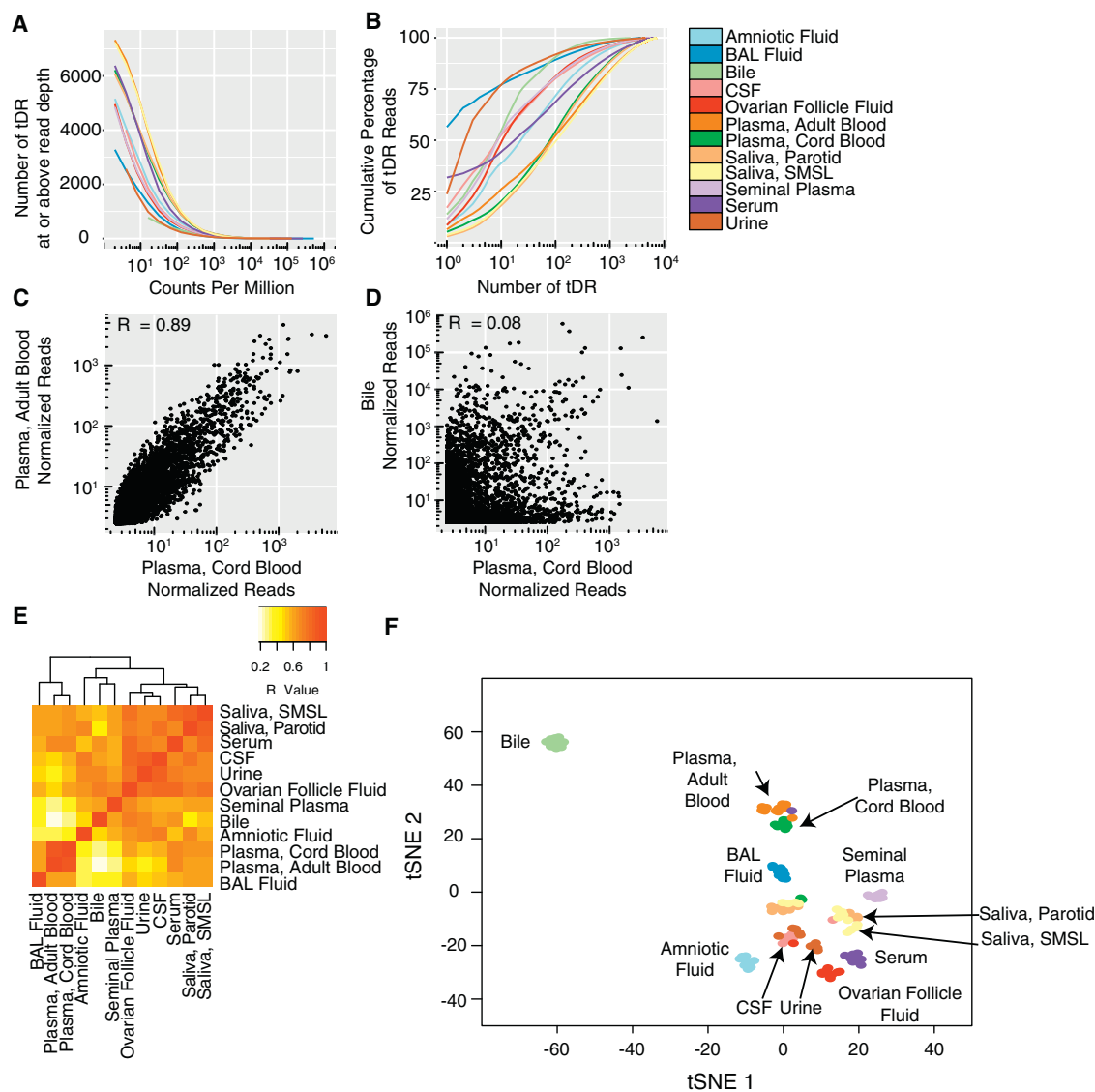


Figure 4. tDR Profiles across 12 Biofluid Types

(A) Number of tDRs detected as a function of read depth.

(B) Cumulative distribution of tDR reads. See also Figure S4.

(C and D) Examples of pairwise correlations between biofluids for cord blood plasma versus adult blood plasma (C) and cord blood plasma versus bile (D). Each point represents the median normalized read count for a single tDR for the indicated biofluids. One normalized read count was added to each measurement to allow representation of read counts for tDRs with no reads on a log scale.

(E) Correlations for all pairs of biofluids.

(F) tSNE plot produced using tDR read counts. Each point represents a single biofluid sample.

correlation coefficients for tDR read counts between biofluids ranged from 0.08 to 0.89 (Figures 4C–4E) and were typically far lower than those found for pairwise correlations using miRNA reads. Correlations were highest between adult and cord blood plasma ($R = 0.89$) and between saliva samples from the two different sites ($R = 0.84$). The proportion of mapped reads that represented tDRs was much higher for serum (37%) than for adult plasma (0.8%), and the pairwise correlation for tDR read counts between serum and adult plasma was only moderate ($R = 0.65$). Using all 8,672 tDRs detected in any sample, tSNE

analysis revealed that samples of most biofluid types formed largely distinct clusters (Figure 4E). As with miRNAs, parotid and submandibular and sublingual glands saliva samples formed overlapping clusters. Urine and CSF samples were also overlapping. Although tSNE analysis based on miRNAs produced overlapping clusters of serum and plasma samples, these biofluid types were clearly distinct when the tSNE analysis was based on tDRs.

To further analyze tDRs, we grouped together tDRs based on the amino acid that corresponds to the tRNA of origin (Figures 5A

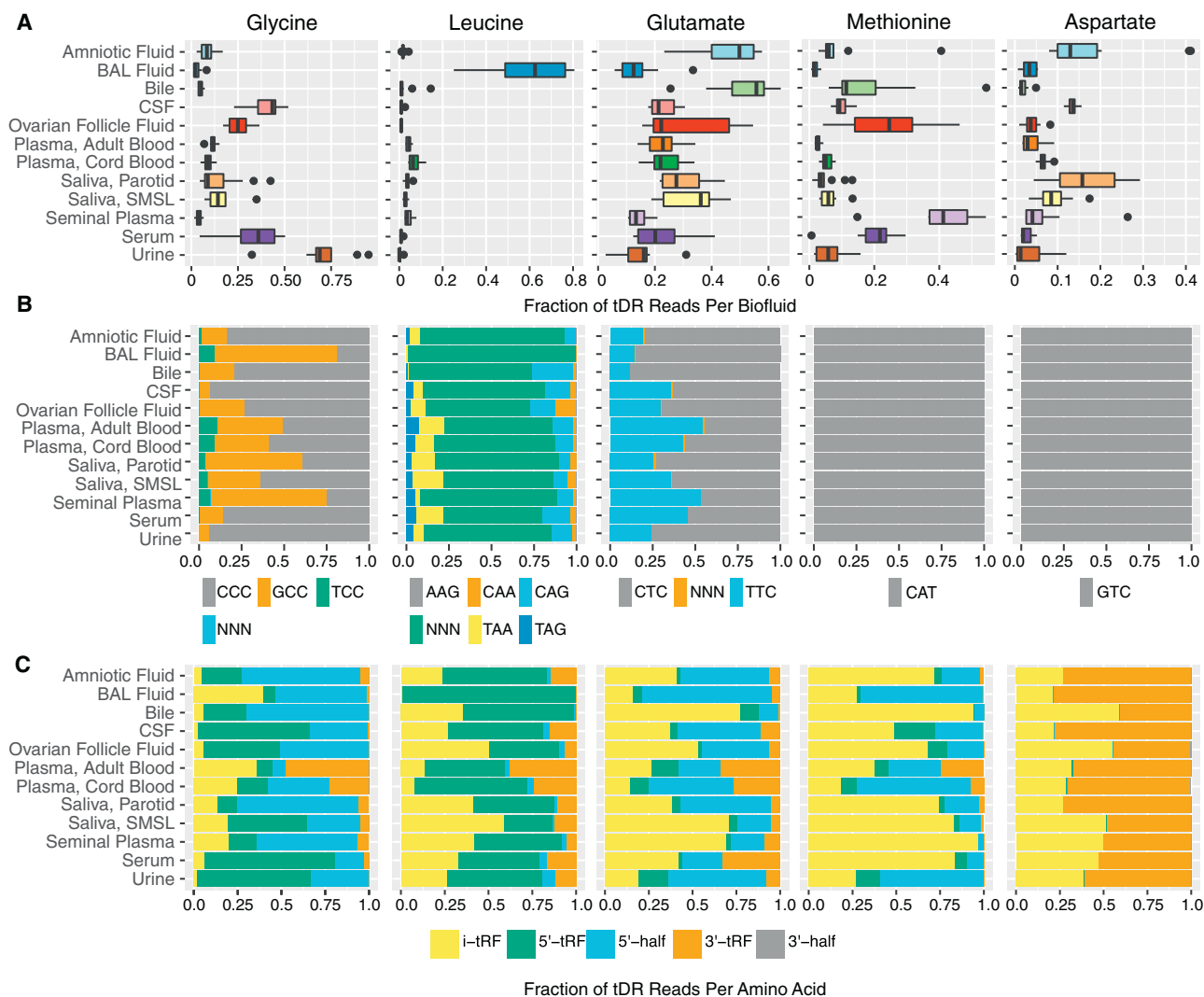


Figure 5. tDR Abundance by Amino Acid, Anticodon, and Fragment Type

(A) tDR abundance by amino acid. Boxes represent median and interquartile ranges, whiskers represent $1.5 \times$ the interquartile range. Dots represent outliers.

(B) tDR abundance by anticodon.

(C) tDR abundance by fragment type.

Data are shown for tDRs from the five most highly represented tRNAs. Data for other tDRs are shown in Figures S5–S7.

and S5). tDRs that could not be unambiguously assigned to an amino acid were excluded from this analysis. Glycine tDRs were predominant in urine, serum, and CSF. Leucine tDRs were predominant in BAL (62%) but were much less frequently found (<15%) in other biofluids. Glutamic acid tDRs were predominant in amniotic fluid, bile, plasma from cord blood and adult blood, and both types of saliva, whereas methionine tDRs were predominant in seminal plasma. Glycine, glutamate, and methionine each represented >20% of tDR reads in ovarian follicle fluid. Median tDR reads for tyrosine, asparagine, phenylalanine, and isoleucine tDRs were <1% each for all biofluid types. Normalized read counts for many tDR-associated amino acids differed markedly between biofluids. For example, tRNA reads aligning to glycine made up 68.9% of tDR reads in urine but $\leq 43.4\%$ of tDR reads in every other biofluid. A short

(16 nt) read mapping to the 5' end of leucine tRNAs accounted for the majority (59%) of all tDR reads in BAL fluid samples but was rare in all other biofluids (0.003%–1.6%).

For amino acids encoded by more than one codon, we examined the distribution of reads for each possible anticodon (Figures 5B and S6). For some amino acids (e.g., glycine), anticodon frequency varied substantially between biofluids, but for others (e.g., leucine), anticodon frequency was less variable. We also classified tDRs by fragment type (Loher et al., 2017). Fragment types differed substantially according to the tRNA of origin (Figures 5C and S7). For example, in most biofluids, glycine tDRs mapped primarily to the 5' region (5' tRFs and 3' halves), whereas methionine and histidine tDRs mapped primarily to internal regions of tRNAs (i-tRFs, fragments beginning after the first position and ending before the non-templated CCA addition). To

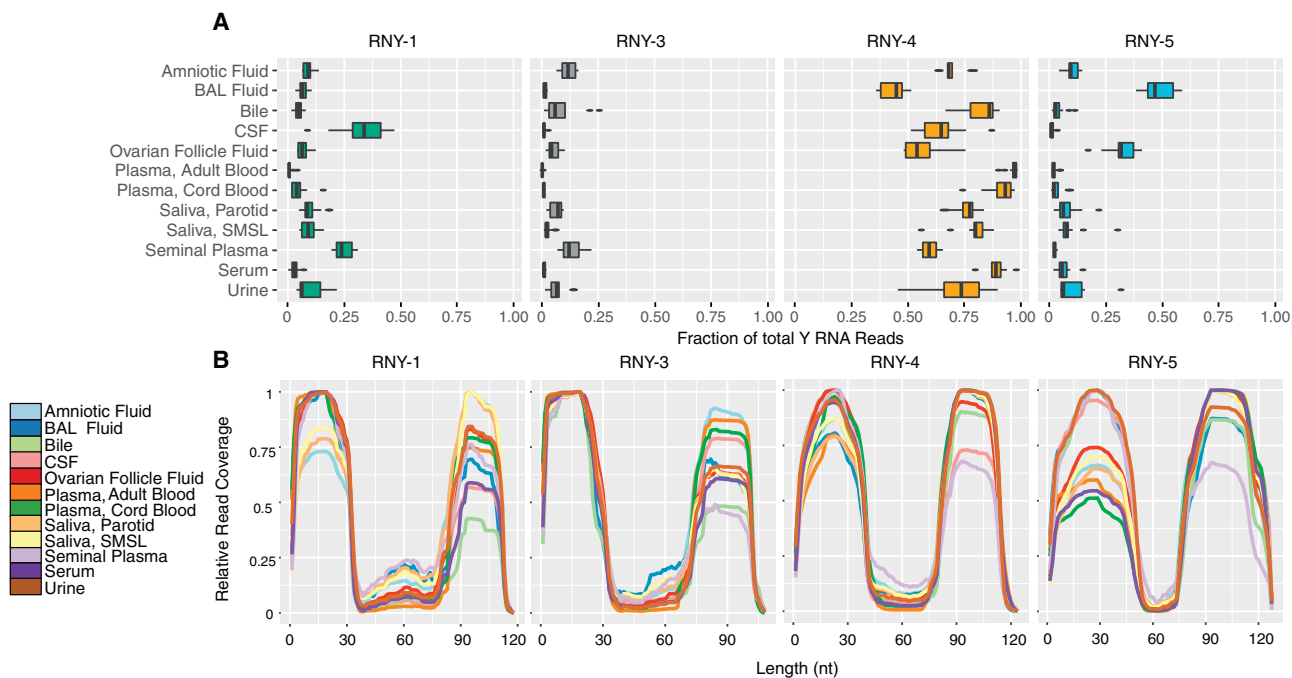


Figure 6. Y RNA Fragments in 12 Biofluid Types

(A) Distribution of Y RNA reads by Y RNA gene. Boxes represent median and interquartile ranges, whiskers represent 1.5 × the interquartile range. Dots represent outliers.

(B) Y RNA read-mapping positions. We determined the number of reads covering each nucleotide of each full-length Y RNA. Values are normalized to the position of each Y RNA with the largest number of reads in each biofluid.

examine differences in tDR fragments between biofluids in more detail, we constructed read coverage maps for four tDRs with the largest numbers of mapped reads (Figure S7). Inspection of these coverage maps reveals large differences in fragment types across biofluids (e.g., increased 3' coverage in amniotic fluid for three of these four tRNAs) and illustrates major differences in fragment sizes (e.g., shorter 5' fragments for tRNA7 LeuAAG compared with the other three tRNAs).

Y RNA Profiles

We found reads mapping to each of the four human Y RNA genes in every biofluid (Figure 6A; Tables S2 and S7). In each biofluid except for BAL fluid, most reads that mapped to Y RNA mapped to RNY4. In BAL fluid, the mean mapping rate for RNY4 was 43%, and RNY5 accounted for 49%. Most reads mapped to either the 5' region or the 3' region of full-length Y RNAs, with relatively few reads mapping to the central portion of Y RNAs (Figure 6B). The proportion of 5' reads to 3' reads varied between biofluids and between different Y RNA genes. For example, RNY4 fragments more frequently mapped to the 5' end in seminal plasma and CSF but more frequently mapped to the 3' end in adult blood plasma, BAL fluid, and saliva.

piRNA Profiles

As discussed previously, some biofluids contained substantial proportions of reads that mapped to both piRNAs and other Gencode RNAs (especially RNY4). The majority of reads that mapped to piRNAs also mapped to other RNA biotypes, except

in one biofluid: seminal plasma. 57% of seminal plasma reads that mapped to piRNAs did not map to another biotype. After exclusion of ambiguously mapped reads (see Experimental Procedures), the number of piRNAs detected was higher in seminal fluid (5,588 distinct piRNAs in one or more samples) than in other fluids (94–1,068 piRNAs). The most frequent 57 piRNAs accounted for 50% of seminal plasma reads that mapped exclusively to piRNAs (Table S2).

DISCUSSION

We applied a standardized RNA-seq approach to identify small RNAs in 12 human biofluid types. This work extends previous analyses of biofluid exRNAs by applying a method that allows for relative quantification of small RNAs from multiple biotypes to a diverse collection of normal biofluids. In each biofluid tested, the majority of mapped reads could be assigned to miRNAs, tDRs, or Y RNAs. Additional reads mapping to piRNAs, mRNAs, long non-coding RNAs, snRNAs, and other RNAs were also detected but were less common. Relative levels of different RNA biotypes differed dramatically between biofluids. The relative abundance of the two major biotypes, miRNAs and tDRs, varied by $>10^3$ -fold across the set of 12 biofluids. Our results are consistent with those from another recent study that found relatively high levels of tDRs in urine and relatively high levels of Y RNA fragments in plasma (Yeri et al., 2017). Despite differences in overall small RNA composition, miRNA levels were generally highly correlated between biofluids ($R = 0.79$ – 0.98).

tDR levels were less well correlated ($R = 0.08$ – 0.89). Biofluids that had highly similar profiles for small RNAs of one biotype sometimes had very different profiles for RNAs of other biotypes. For example, adult blood plasma and serum miRNA levels were highly correlated ($R = 0.98$), but tDR levels in these biofluids were much less well correlated ($R = 0.68$). Our findings demonstrate the extensive diversity of small RNA populations within each biofluid type and highlight both similarities and differences between biofluids.

The methods we used have advantages and limitations compared with those used in previous studies. We used a uniform approach to RNA isolation, RNA-seq library preparation, sequencing library size selection, sequencing, and data analysis for all samples. The RNA-seq method we used reduces bias by employing oligonucleotides with random nucleotide sequences for adaptor ligation (Giraldez et al., 2017). Use of a permissive size selection strategy rather than one targeting miRNAs allowed us to identify RNAs of many biotypes. We used a publicly available analysis pipeline, and our data are publicly available, which allows for reanalysis of the primary data as databases and analysis tools are updated. To obtain a broad view of small RNAs in biofluids, we extracted RNA from unfractionated biofluids. Other studies have shown differences in small RNA content of specific biofluid compartments such as exosomes and argonaute-2-containing ribonucleoprotein complexes (Arroyo et al., 2011), and further work will be required to understand the compartmentalization of RNAs in the large set of biofluids we studied. A major limitation of the RNA-seq method we used is that this method is designed for sequencing of RNAs with a 5' phosphate and a 3' hydroxyl group. RNAs with various 5' end modifications (hydroxyl, triphosphate, or 5' cap) or 3' end modifications (2'-3'-cyclic phosphate, 2'-O-methyl) as well as RNAs with certain structures are more difficult or impossible to detect with this method (Raabe et al., 2014). RNAs that undergo many types of post-translational modifications, including those that are commonly found in tRNAs or after miRNA editing, may not be detected or may fail to map to the human genome. The proportion of reads that aligned to the human genome was variable, and for some biofluids, the absolute number of reads aligning to certain biotypes was low. We focused our comparative analyses on large differences in RNAs that were abundant in at least one biofluid, and our approach may not detect differences in RNAs that less abundant but nonetheless have biological importance. Differences in centrifugation speeds of biofluids could also be responsible for some of the observed differences between biofluid types. In addition, although we excluded samples with apparent blood contamination, the presence of the erythrocyte miRNA miR-451a (among the top 10 miRNAs in ovarian follicle fluid and bile as well as plasma and serum) might reflect low-level blood contamination of some samples.

Our approach was designed to compare levels of each RNA between samples. A qPCR analysis (Figure S2) validated our ability to measure relative differences in RNA levels by small RNA-seq. For both RNA-seq and qPCR, detection of RNAs depends upon both RNA abundance and technical biases. Differences in technical bias are likely the major explanation of why miRNAs with the highest RNA-seq read counts in our study differ markedly from miRNAs that were most highly detected (lowest

cycle threshold [Ct]) in a prior qPCR-based study of many biofluids (Weber et al., 2010). Therefore, detection rates should not be used as direct measures of the abundance of different RNAs within any sample.

miRNA profiles of these 12 biofluids were remarkably similar overall. Comparison of biofluid miRNA levels with miRNA levels in a single human organ (brain) suggests that cellular miRNA levels are a major determinant of biofluid miRNA abundance (Supplemental Experimental Procedures; Figure S3), but there may be other factors related to miRNA export from cells (Villarroya-Beltri et al., 2013) or extracellular miRNA stability that also affect miRNA abundance in biofluids (Taylor and Gercel-Taylor, 2008; Gilad et al., 2008). We identified large differences in levels of some miRNAs between biofluids. These may arise from differences in local production of cellular miRNAs, since we found higher levels of extraembryonic cell miRNAs in amniotic fluid, nervous system cell miRNAs in CSF, and an epididymal miRNA in seminal plasma. In many cases, miRNAs produced from the same pre-miRNA or the same miRNA cluster had highly correlated levels across the full set of biofluids, which also suggests that miRNA production is a major determinant of differences in miRNA abundance. In addition to differences in local production, differences in extracellular miRNAs could arise from preferential secretion of miRNAs from cells (Valadi et al., 2007) or differences in processing or elimination of extracellular miRNAs. Since extracellular miRNAs have been shown to be capable of entering cells may modulate gene expression (Patton et al., 2015), our results suggest that differences in miRNA composition between biofluids may be functionally important.

We found large differences in tDRs across biofluids. tRNAs can be processed into tDRs by sequence-specific enzymatic cleavage events (Lee et al., 2009), and tDRs have been previously identified in human biofluids, including human serum and urine, and in exosomes from human semen (Dhahbi, 2015; Yeri et al., 2017). We found differences in tRNA of origin and fragment type across the set of biofluids. Many factors, including tRNA levels, tRNA cleavage, tDR secretion, and biodistribution and elimination of extracellular tDRs, may contribute to determining levels of tDRs. Intracellular tDRs have been reported to have many roles, including inhibiting protein translation during cellular stress (Ivanov et al., 2011) and mediating intergenerational inheritance (Chen et al., 2016; Sharma et al., 2016). Roles of extracellular tDRs remain enigmatic, but the presence of relatively high levels of extracellular tDRs in biofluids such as bile, urine, seminal plasma, and amniotic fluid (where tDRs represented the majority of all mapped small RNA reads) suggests that extracellular tDRs may be functionally important. At least some of the functional roles of intracellular tDRs are quite specific. For example, a tDR from the Gly-GCC tRNA, but not other tDRs, was shown to regulate retroelement-driven transcription of highly expressed genes in preimplantation embryos (Sharma et al., 2016). Differences in tDR populations between biofluids may therefore have important functional implications.

Y RNA fragments represented a substantial proportion of small RNAs detected in some biofluids, including adult and cord blood plasma, serum, and CSF. A previous study also reported many Y RNA fragment reads in human serum and plasma

(Dhahbi et al., 2013). That study found that >95% of fragments mapped to the 5' end of Y RNAs. Our results, obtained using an RNA-seq method developed to reduce bias, revealed a substantially higher proportion of 3' Y RNA fragments. Intact cellular Y RNAs are required for the initiation of DNA replication, regulation of RNA stability, and cellular responses to stress (Kowalski and Krude, 2015). The role of Y RNA fragments is less well understood. It has been reported that RNY5 fragments produced in cancer cell extracellular vesicles trigger rapid cell death in primary cells, but not in cancer cells (Chakraborty et al., 2015), suggesting that further studies of the biogenesis, trafficking, and function of Y RNA fragments found in biofluids are warranted.

Blood plasma and serum contained a substantial number of small RNAs with sequences that map to piRNAs, but we urge caution about inferring that these are derived from piRNAs. The large majority of piRNA-mapped reads we obtained in our plasma and serum samples also mapped to other RNA biotypes (especially Y RNAs) that are abundant in these biofluids. In both our study and a previous study (Freedman et al., 2016), the piRNA with the largest number of mappings shares sequence with a fragment of RNY4, a Y RNA. We therefore confined our analysis of piRNAs to sequences that did not map to RNAs of other biotypes. piRNAs have prominent roles in the testis (Girard et al., 2006), and even using this more conservative approach, we found that a substantial number and variety of piRNA sequences were detected in seminal plasma. A previous report also identified piRNAs in seminal plasma and identified a set of piRNAs that were associated with infertility (Hong et al., 2016). Although previously suggested to be germline specific, piRNAs also play roles in non-gonadal cells (Ishizu et al., 2012). Non-gonadal cells might therefore be a source of piRNAs in biofluids outside the reproductive tract. However, we found relatively few reads that mapped only to piRNAs in biofluids other than seminal plasma. The functional significance of extracellular piRNAs is not yet known.

By using a uniform RNA-seq-based approach, we produced an extensive catalog of small RNAs that are present in a large set of human biofluids. The RNA composition of biofluids differed widely, with major differences in the distribution of biotypes and of specific RNAs within each biotype. This work provides a resource for investigators seeking to understand the production, distribution, and function of exRNAs. Biofluid small RNAs are being explored as disease biomarkers, and our work also helps to identify RNAs that are present in multiple biofluids and may have potential as novel biomarkers.

EXPERIMENTAL PROCEDURES

Human Subjects

With the exception of bile samples, which were obtained at the Mayo Clinic following liver transplantation, all samples were obtained from healthy subjects enrolled in a variety of studies conducted at the University of California, San Francisco. Protocols were approved by institutional review boards at the Mayo Clinic (bile) and the University of California, San Francisco (all other specimen types). Except for cord blood plasma, which was obtained at birth, all other samples were obtained from studies enrolling subjects aged 18 years or older. Details of sample collection and storage for each sample type are provided in [Supplemental Experimental Procedures](#).

RNA Isolation, Library Preparation, and Sequencing

Frozen biofluid samples used for small RNA-seq experiments were thawed and centrifuged at $2,000 \times g$ for 5 min at 4°C. RNA was isolated from 200 μ L of biofluid using the QIAGEN miRNEasy Micro Kit according to the manufacturer's protocol, except that 1 mL Qiazol and 180 μ L chloroform were used. Small RNA-seq libraries were prepared using 4N protocol D as previously described (Giraldez et al., 2017; [Supplemental Experimental Procedures](#)).

Sequence Alignment

All FASTQ files were processed using the exceRpt small RNA-seq pipeline version 4.6.2 available on the Genboree Workbench (<http://www.genboree.org/index.html>). Sequence reads were clipped 4 nt from the invariant portions of both 5' and 3' adapters to remove the four randomized nucleotide in the adapters. Clipped reads were aligned with a one-mismatch allowance. To exclude piRNAs that mapped to other biotypes, we changed the order of mapping assignment to count Gencode alignments in preference to piRNA alignments (alignment order: miRNAs, tRNAs, all annotations from Gencode, piRNAs, and circular RNA). We also eliminated piRNA mappings if the large majority (>98%) of reads mapping to a given piRNA also mapped to an RNA of a different biotype. To determine whether the high proportion of CSF reads that did not map to the human genome mapped to other genomes, we used the Genboree exogenous alignment option to map CSF samples to exogenous miRNAs in miRbase and all sequenced genomes in Ensembl and NCBI. No mismatches were permitted for exogenous alignments.

RNA-Seq Data Inclusion Criteria

RNA-seq results were included if (1) $\geq 100,000$ reads mapped to the transcriptome, (2) the number of reads mapping to annotated transcripts represented at least 50% of the number of reads mapping to the human genome, and (3) $R \geq 0.85$ for correlation of miRNA reads with most other samples of the same biofluid. For samples that failed to meet inclusion criteria on initial analysis, duplicate biofluid samples were used for a second round of RNA isolation and library preparation. If the repeat analysis met inclusion criteria, then these results were used for all analyses. If the repeat analysis did not meet the inclusion criteria, then data from that sample were not used for subsequent analyses. There was no case in which replicate analysis of a sample gave consistent results ($R \geq 0.85$ for miRNAs), but these results were poorly correlated ($R < 0.85$) with most other samples of the same biofluid type.

RNA Biotype Distribution Analysis

We used the biotype counts file generated by the small RNA-seq pipeline on the Genboree Workbench. Since Y RNA reads represented a large proportion of Gencode transcript reads in some biofluids, we used the Gencode read counts file to quantify Y RNA reads separately from other Gencode reads.

miRNA Analysis

miRNA diversity analyses were performed using the mean normalized miRNA read counts for all samples of a given biofluid type. Read counts for each miRNA in each sample were normalized by dividing by the total number of miRNA read counts in that sample. For analyses of the numbers of miRNAs detected as a function of read depth, we excluded all miRNAs with mean read counts $< 1/10^6$ total miRNA reads (lower limit of detection). For biofluids where miRNAs represented a small proportion of total RNA reads, the total number of miRNA reads for all samples was $< 10^6$. For those biofluids, the lower limit of detection was defined as 10^6 divided by the total number of miRNA reads for that biofluid. For analyses of cumulative distributions, we ranked miRNAs in descending order of normalized mean read counts for a given biofluid and calculated the cumulative sum until we included enough miRNAs to account for 99.9% of miRNA reads.

For pairwise correlations and tSNE analyses, we normalized read counts by DESeq2 (Love et al., 2014) and determined mean read counts for all samples of each biofluid. We calculated the Pearson correlation coefficient of each biofluid pair using R. We generate tSNE plots using the Rtsne function of the Rtsne R package with perplexity 20 and a maximum iteration of 5,000. We plotted the results using the xyplot function from the lattice R package. We used Bayesian relevance networks (Ramachandran et al., 2017) to generate a list of

co-expressing miRNAs. We normalized miRNA read counts with DESeq2, transformed the counts data using the varianceStabilizingTransformation function, identified the top 25% most frequent miRNAs (by mean), and then selected the top 13% most variable as input to the Bayesian relevance networks algorithm. We selected all co-expressing miRNAs with a Bayesian correlation ≥ 0.80 , which had an estimated FDR of 0.012. We used the pheatmap function from the pheatmap R package to generate a hierarchical clustering diagram for all miRNAs belonging to networks with 3 or more miRNAs.

Methods used for qPCR miRNA validation experiments are described in [Supplemental Experimental Procedures](#).

tDR Analysis

We used MINTmap ([Loher et al., 2017](#)) to analyze tDRs. We removed adapters, trimmed 4 nt from each end of the read and removed low-quality reads (using standard Genboree parameters) from all fastq files with the fastx-toolkit. Fastq files were then processed individually by MINTmap. We only counted alignments that mapped exclusively to annotated tRNA regions to reduce ambiguity. For analyses of amino acid or anticodon read counts, reads that mapped ambiguously (to multiple amino acids or anticodons) were deemed “undetermined.” All reads were aggregated by sum and normalized by the total number of tDR reads within each sample or biofluid type. To analyze tRNA coverage, we took the normalized mean of each sample and aggregated read counts by biofluid and the start and end position.

Y-RNA Analysis

Using the bam files generated by the Genboree pipeline using the “Upload Full Results” option, we searched for any alignment to the 4 human Y RNAs (RNY1-201, RNY3-201, RNY4-201, and RNY5-201) and considered each unique fragment a distinct read. To plot read coverage, we followed the same strategy as outlined above for tDRs.

DATA AND SOFTWARE AVAILABILITY

The accession number for the raw sequence data for bile reported in this paper is GEO: GSE112343. The accession numbers for raw data for seminal plasma, amniotic fluid and cord blood plasma, CSF, and ovarian follicle fluid reported in this paper are dbGAP: phs001692.v1.p1, phs001693.v1.p1, phs001694.v1.p1, and phs001695.v1.p1, respectively. The accession numbers for the read counts for all samples reported in this paper are exRNA Atlas (<https://exrna-atlas.org/>): EXR-DERLE1PHASE1PRIOT-AN (parotid and submandibular and sublingual saliva) and EXR-DERLE1PHASE1OPEN-AN (all other biofluids). Due to issues related to patient confidentiality, raw data are not currently available for adult blood plasma, serum, urine, BAL fluid, and parotid and submandibular and sublingual saliva samples. If consent and approval are granted in the future, the raw data will be made available at that time.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and seven tables and can be found with this article online at <https://doi.org/10.1016/j.celrep.2018.10.014>.

ACKNOWLEDGMENTS

This publication is part of the NIH ERCC paper package and was supported by the NIH Common Fund’s exRNA Communication Program. We thank the individuals who participated in the study by providing samples; Christine P. Nguyen and Bobby Antalek for assistance with collection and annotation of the BAL fluid, adult blood plasma, serum, and urine samples; Serena Spudich for access to CSF samples collected in studies that she directed; Michael Li and Suresh Garudadri for performing qPCR experiments; and William Thistlethwaite (Baylor College of Medicine) for help with depositing data. This work was supported by the NIH Extracellular RNA Communication Program

(grant U01HL126493 to P.G.W. and D.J.E. and grants UH2TR000884 and UH3TR000884 to T.P.).

AUTHOR CONTRIBUTIONS

P.M.G., N.R.B., A.J.B., P.G.W., and D.J.E. designed the study. N.R.B., H.C., S.F., T.C.M., T.P., R.W.P., J.F.S., and P.G.W. recruited study participants and collected biofluid samples. P.M.G. performed the RNA-seq experiments and analyzed the data. N.R.B. designed, performed, and analyzed the qPCR experiments. P.M.G. and D.J.E. wrote the paper. All authors contributed to manuscript editing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 25, 2018

Revised: May 26, 2018

Accepted: October 2, 2018

Published: October 30, 2018

REFERENCES

- Argyropoulos, C., Wang, K., McClarty, S., Huang, D., Bernardo, J., Ellis, D., Orchard, T., Galas, D., and Johnson, J. (2013). Urinary microRNA profiling in the nephropathy of type 1 diabetes. *PLoS ONE* 8, e54662.
- Arroyo, J.D., Chevillet, J.R., Kroh, E.M., Ruf, I.K., Pritchard, C.C., Gibson, D.F., Mitchell, P.S., Bennett, C.F., Pogosova-Agadjanyan, E.L., Stirewalt, D.L., et al. (2011). Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc. Natl. Acad. Sci. USA* 108, 5003–5008.
- Bahn, J.H., Zhang, Q., Li, F., Chan, T.M., Lin, X., Kim, Y., Wong, D.T., and Xiao, X. (2015). The landscape of microRNA, Piwi-interacting RNA, and circular RNA in human saliva. *Clin. Chem.* 61, 221–230.
- Barger, J.F., Rahman, M.A., Jackson, D., Acunzo, M., and Nana-Sinkam, S.P. (2016). Extracellular miRNAs as biomarkers in cancer. *Food Chem. Toxicol.* 98 (Pt A), 66–72.
- Chakraborty, S.K., Prakash, A., Nechooshtan, G., Hearn, S., and Gingeras, T.R. (2015). Extracellular vesicle-mediated transfer of processed and functional RNY5 RNA. *RNA* 21, 1966–1979.
- Chen, Q., Yan, M., Cao, Z., Li, X., Zhang, Y., Shi, J., Feng, G.H., Peng, H., Zhang, X., Zhang, Y., et al. (2016). Sperm tsRNAs contribute to intergenerational inheritance of an acquired metabolic disorder. *Science* 351, 397–400.
- Czech, B., and Hannon, G.J. (2016). One loop to rule them all: the ping-pong cycle and piRNA-guided silencing. *Trends Biochem. Sci.* 41, 324–337.
- de Rie, D., Abugessaisa, I., Alam, T., Arner, E., Arner, P., Ashoor, H., Åström, G., Babina, M., Bertin, N., Burroughs, A.M., et al.; FANTOM Consortium (2017). An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat. Biotechnol.* 35, 872–878.
- Dhabhi, J.M. (2015). 5’ tRNA halves: the next generation of immune signaling molecules. *Front. Immunol.* 6, 74.
- Dhabhi, J.M., Spindler, S.R., Atamna, H., Boffelli, D., Mote, P., and Martin, D.I. (2013). 5’-YRNA fragments derived by processing of transcripts from specific YRNA genes and pseudogenes are abundant in human serum and plasma. *Physiol. Genomics* 45, 990–998.
- Freedman, J.E., Gerstein, M., Mick, E., Rozowsky, J., Levy, D., Kitchen, R., Das, S., Shah, R., Danielson, K., Beaulieu, L., et al. (2016). Diverse human extracellular RNAs are widely detected in human plasma. *Nat. Commun.* 7, 11106.
- Gebetsberger, J., and Polacek, N. (2013). Slicing tRNAs to boost functional ncRNA diversity. *RNA Biol.* 10, 1798–1806.
- Gilad, S., Meiri, E., Yogeve, Y., Benjamin, S., Lebanony, D., Yerushalmi, N., Benjamin, H., Kushnir, M., Cholak, H., Melamed, N., et al. (2008). Serum microRNAs are promising novel biomarkers. *PLoS ONE* 3, e3148.

- Giraldez, M.D., Spengler, R.M., Etheridge, A., Godoy, P.M., Barczak, A.J., Srinivasan, S., De Hoff, P.L., Tanriverdi, K., Courtright, A., Lu, S., et al. (2017). Accuracy, reproducibility and bias of next generation sequencing for quantitative small rna profiling: a multiple protocol study across multiple laboratories. *bioRxiv*. <https://doi.org/10.1101/113050>.
- Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202.
- Gray, C., McCowan, L.M., Patel, R., Taylor, R.S., and Vickers, M.H. (2017). Maternal plasma miRNAs as biomarkers during mid-pregnancy to predict later spontaneous preterm birth: a pilot study. *Sci. Rep.* **7**, 815.
- Hill, K.E., Kelly, A.D., Kuijjer, M.L., Barry, W., Rattani, A., Garbutt, C.C., Kissick, H., Janeway, K., Perez-Atayde, A., Goldsmith, J., et al. (2017). An imprinted non-coding genomic cluster at 14q32 defines clinically relevant molecular subtypes in osteosarcoma across multiple independent datasets. *J. Hematol. Oncol.* **10**, 107.
- Hong, Y., Wang, C., Fu, Z., Liang, H., Zhang, S., Lu, M., Sun, W., Ye, C., Zhang, C.-Y., Zen, K., et al. (2016). Systematic characterization of seminal plasma piRNAs as molecular biomarkers for male infertility. *Sci. Rep.* **6**, 24229.
- Hoy, A.M., and Buck, A.H. (2012). Extracellular small RNAs: what, where, why? *Biochem. Soc. Trans.* **40**, 886–890.
- Ishizu, H., Siomi, H., and Siomi, M.C. (2012). Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. *Genes Dev.* **26**, 2361–2373.
- Ivanov, P., Emara, M.M., Villen, J., Gygi, S.P., and Anderson, P. (2011). Angiogenin-induced tRNA fragments inhibit translation initiation. *Mol. Cell* **43**, 613–623.
- Korpala, M., and Kang, Y. (2008). The emerging role of miR-200 family of microRNAs in epithelial-mesenchymal transition and cancer metastasis. *RNA Biol.* **5**, 115–119.
- Kowalski, M.P., and Krude, T. (2015). Functional roles of non-coding Y RNAs. *Int. J. Biochem. Cell Biol.* **66**, 20–29.
- Lee, Y.S., Shibata, Y., Malhotra, A., and Dutta, A. (2009). A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* **23**, 2639–2649.
- Lee, L.W., Zhang, S., Etheridge, A., Ma, L., Martin, D., Galas, D., and Wang, K. (2010). Complexity of the microRNA repertoire revealed by next-generation sequencing. *RNA* **16**, 2170–2180.
- Li, Y., Wang, H.-Y., Wan, F.-C., Liu, F.-J., Liu, J., Zhang, N., Jin, S.-H., and Li, J.-Y. (2012). Deep sequencing analysis of small non-coding RNAs reveals the diversity of microRNAs and piRNAs in the human epididymis. *Gene* **497**, 330–335.
- Loher, P., Telonis, A.G., and Rigoutsos, I. (2017). MINTmap: fast and exhaustive profiling of nuclear and mitochondrial tRNA fragments from short RNA-seq data. *Sci. Rep.* **7**, 41184.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
- Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., Rheinheimer, S., Meder, B., Stähler, C., Meese, E., and Keller, A. (2016). Distribution of miRNA expression across human tissues. *Nucleic Acids Res.* **44**, 3865–3877.
- Mitchell, P.S., Parkin, R.K., Kroh, E.M., Fritz, B.R., Wyman, S.K., Pogosova-Agadjanyan, E.L., Peterson, A., Noteboom, J., O'Briant, K.C., Allen, A., et al. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl. Acad. Sci. USA* **105**, 10513–10518.
- Ni, H., Capodici, J., Cannon, G., Communi, D., Boeynaems, J.M., Karikó, K., and Weissman, D. (2002). Extracellular mRNA induces dendritic cell activation by stimulating tumor necrosis factor-alpha secretion and signaling through a nucleotide receptor. *J. Biol. Chem.* **277**, 12689–12696.
- Patton, J.G., Franklin, J.L., Weaver, A.M., Vickers, K., Zhang, B., Coffey, R.J., Ansel, K.M., Btleloch, R., Goga, A., Huang, B., et al. (2015). Biogenesis, delivery, and function of extracellular RNA. *J. Extracell. Vesicles* **4**, 27494.
- Raabe, C.A., Tang, T.-H., Brosius, J., and Rozhdestvensky, T.S. (2014). Biases in small RNA deep sequencing data. *Nucleic Acids Res.* **42**, 1414–1426.
- Ramachandran, P., Sánchez-Taltavull, D., and Perkins, T.J. (2017). Uncovering robust patterns of microRNA co-expression across cancers using Bayesian Relevance Networks. *PLoS ONE* **12**, e0183103.
- Semenov, D.V., Kuligina, E.V., Shevryrina, O.N., Richter, V.A., and Vlassov, V.V. (2004). Extracellular ribonucleic acids of human milk. *Ann. N.Y. Acad. Sci.* **1022**, 190–194.
- Sharma, U., Conine, C.C., Shea, J.M., Boskovic, A., Derr, A.G., Bing, X.Y., Belleanne, C., Kucukural, A., Serra, R.W., Sun, F., et al. (2016). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science* **351**, 391–396.
- Skog, J., Würdinger, 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.
- Sohel, M.H. (2016). Extracellular/circulating microRNAs: release mechanisms, functions and challenges. *Achiev. Life Sci.* **10**, 175–186.
- Taylor, D.D., and Gercel-Taylor, C. (2008). MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecol. Oncol.* **110**, 13–21.
- Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J.J., and Lötval, J.O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.* **9**, 654–659.
- Villarroya-Beltri, C., Gutiérrez-Vázquez, C., Sánchez-Cabo, F., Pérez-Hernández, D., Vázquez, J., Martín-Cofreces, N., Martínez-Herrera, D.J., Pascual-Montano, A., Mittelbrunn, M., and Sánchez-Madrid, F. (2013). Sumoylated hnRNP2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs. *Nat. Commun.* **4**, 2980.
- Waller, R., Wyles, M., Heath, P.R., Kazoka, M., Wolff, H., Shaw, P.J., and Kirby, J. (2018). Small RNA sequencing of sporadic amyotrophic lateral sclerosis cerebrospinal fluid reveals differentially expressed miRNAs related to neural and glial activity. *Front. Neurosci.* **11**, 731.
- Weber, J.A., Baxter, D.H., Zhang, S., Huang, D.Y., Huang, K.H., Lee, M.J., Galas, D.J., and Wang, K. (2010). The microRNA spectrum in 12 body fluids. *Clin. Chem.* **56**, 1733–1741.
- Williams, Z., Ben-Dov, I.Z., Elias, R., Mihailovic, A., Brown, M., Rosenwaks, Z., and Tuschl, T. (2013). Comprehensive profiling of circulating microRNA via small RNA sequencing of cDNA libraries reveals biomarker potential and limitations. *Proc. Natl. Acad. Sci. USA* **110**, 4255–4260.
- Yagi, Y., Ohkubo, T., Kawaji, H., Machida, A., Miyata, H., Goda, S., Roy, S., Hayashizaki, Y., Suzuki, H., and Yokota, T. (2017). Next-generation sequencing-based small RNA profiling of cerebrospinal fluid exosomes. *Neurosci. Lett.* **636**, 48–57.
- Yeri, A., Courtright, A., Reiman, R., Carlson, E., Beecroft, T., Janss, A., Siniard, A., Richholt, R., Balak, C., Rozowsky, J., et al. (2017). Total extracellular small RNA profiles from plasma, saliva, and urine of healthy subjects. *Sci. Rep.* **7**, 44061.

Cell Reports, Volume 25

Supplemental Information

Large Differences in Small RNA

Composition Between Human Biofluids

Paula M. Godoy, Nirav R. Bhakta, Andrea J. Barczak, Hakan Cakmak, Susan Fisher, Tippi C. MacKenzie, Tushar Patel, Richard W. Price, James F. Smith, Prescott G. Woodruff, and David J. Erle

Cell Reports, Volume 25

Supplemental Information

Large Differences in Small RNA

Composition Between Human Biofluids

Paula M. Godoy, Nirav R. Bhakta, Andrea J. Barczak, Hakan Cakmak, Susan Fisher, Tippi C. MacKenzie, Tushar Patel, Richard W. Price, James F. Smith, Prescott G. Woodruff, and David J. Erle

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Sample Collection and Storage for Samples Analyzed by Small RNA-seq

Amniotic Fluid

Amniotic fluid samples were collected prospectively from healthy pregnant women undergoing diagnostic amniocentesis or at term delivery. Samples were centrifuged at 1,320 x *g* for 10 min at room temperature and then frozen at -80°C.

Cord Blood Plasma

Blood was collected from umbilical cords in purple top tubes at the time of term deliveries. Tubes were centrifuged at 1,320 x *g* for 10 min at room temperature and supernatants were frozen at -80°C.

BAL Fluid

BAL fluid was collected from healthy control subjects during bronchoscopy in a completed clinical study (NCT01484691). All participants were between 18 and 70 years of age. Exclusion criteria included any history of asthma or allergic rhinitis, airway hyper-responsiveness as defined by a $\geq 20\%$ decrease in forced expiratory volume in 1 s in response to inhalation of 8 mg/mL methacholine, any cigarette smoking within one year, or >10 total pack-years smoking history. Each of two segments of either the right middle lobe or lingula (chosen per participant by randomization) was lavaged serially with 50 mL aliquots of normal saline (200-300 mL total per participant) and wall suction. Recovered volumes were 110 ± 22 ml (mean \pm SD). Specimen traps were kept on ice. BAL fluid was filtered through 2-ply gauze to remove large clumps of mucus, centrifuged for 5 min at 300 x *g* at 4°C, and stored at -80°C.

Bile

Bile was collected from an externally draining tube placed within the common bile duct at the time of liver transplantation. Samples were collected 2-3 weeks following transplantation in a Therapak Single-Specimen Collection 90 mL container, transferred to a 15mL polypropylene tube, and then centrifuged for 10 min at 3,000 x *g* at 4°C. Aliquots of bile were stored at -78°C.

CSF

CSF samples were collected by lumbar puncture from HIV-seronegative healthy control volunteers in San Francisco, California in the context of studies of HIV-1 infection and the

central nervous system. Lumbar punctures were performed for research purposes only. CSF was collected in polypropylene tubes and placed immediately on wet ice for transport to the UCSF Core Virology Laboratory where it was processed within 1 h. After centrifugation at 600 x *g* for 10 min to remove cells, supernatants were aliquoted and snap frozen for storage at -80°C.

Ovarian Follicle Fluid

Healthy egg donors underwent ovarian follicular stimulation with gonadotropins per UCSF clinic protocols. Ultrasounds were performed to assess and confirm uterine lining and ovarian follicle maturation. Approximately 36 h after human chorionic gonadotropin injection when a sufficient number of mature-sized follicles had developed, the follicles (≥ 16 mm diameter) were aspirated for egg retrieval using negative pressure into 10 mL tubes prior to transvaginal ultrasound-guided oocyte retrieval. Two independent samples of 3.5-5.0 mL of follicular fluid were collected from each study participant, one from each ovary. Visual inspection and Combur-9-test urine sticks (sensitivity 5 erythrocytes/ μ L) were used to exclude samples with blood contamination. Each sample was centrifuged at 400 x *g* for 10 min at room temperature. Supernatants were aliquoted and immediately frozen at -80 °C.

Adult Blood Plasma, Serum, and Urine

Blood and urine were collected from healthy adults ages 18-70 with no evidence of coronary heart disease, peripheral vascular disease, congestive heart failure, cerebrovascular disease, cancer, chronic lung disease, endocrinopathy, or renal disease by subject report or by history and physical examination. Subjects fasted for a minimum of 8 h prior to blood and urine collection. Blood was collected using a 21-gauge needle and the first 5 ml was discarded. Plasma was prepared from EDTA-containing vacutainer tubes via centrifugation for 10 min at 500 x *g* at 23 °C, followed by transfer of the top layer to fresh tubes (avoiding buffy coat), followed by centrifugation for 10 min at 2000 x *g* at room temperature and transfer of the top 80% of the plasma to fresh tubes for storage at -80°C. Serum was collected from red top vacutainer tubes that were kept at room temperature for 30-45 min to allow for coagulation, followed by centrifugation for 10 min at 2000 x *g* at room temperature. Urine was collected by clean catch in a sterile collection container, followed by centrifugation for 10 min at 2000 x *g* at room temperature. The top 80% was transferred to fresh tubes followed by storage at -80°C.

Saliva

SMSL and parotid saliva were collected as described (Albertolle et al., 2015). Briefly, subjects thoroughly rinsed their mouths with water prior to sample collection. Salivary flow was stimulated by the application of citric acid to the tongue. Parotid secretions were obtained by using a Lashley cup. SMSL saliva was obtained by using a Block and Brotman collector that fit around the gland openings. The secretions were collected on ice. Immediately thereafter, a protease inhibitor cocktail (Pierce) was added and samples were briefly vortexed, divided into 1 mL aliquots, and frozen at -80°C .

Seminal Plasma

Semen samples were collected through masturbation into a sterile cup and centrifuged to remove sperm. Semen analysis was performed according to WHO 2010 guidelines in the UCSF Center for Reproductive Health Andrology Laboratory. All study participants had normal semen parameters. Samples were centrifuged for 5 min at $428 \times g$ and supernatants were stored at -80°C .

RNA Isolation, Library Preparation, and Sequencing

The small RNA-sequencing method used to generate libraries relies on a version of the TruSeq Small RNA Library Preparation Kit modified by using randomized adapters, adding PEG to the adapter-RNA solution, and including steps to enzymatically remove excess adapter after 3' ligation. Multiple libraries with unique indexes were pooled, purified using the Qiagen MinElute PCR Purification Kit per the manufacturer's recommendations. Libraries were size-selected using the PippinPrep (Sage Science) with a 3% agarose gel. In pilot experiments, we adjusted the size selection parameters to maintain a low proportion of adapter dimers (132 bp) and maximize the proportion of library with inserts of $\sim 22\text{-}30$ bp. We selected the PippinPrep broad range option to deplete sequences <137 bp and >166 bp but maintain a range of insert sizes within this range. Size-selected DNA was sequenced on an Illumina HiSeq 4000 (single end 50 base mode).

qPCR validation

Samples collected for qPCR experiments were obtained from healthy subjects enrolled in a study conducted at the University of California, San Francisco. Samples were collected from 11 males and 8 females. Of these 19 subjects, 1 is Native America, 3 are Hispanic, 4 are Asian, and 11 are Caucasian. The median age was 32 years (interquartile range, 27.5 - 42.5). For BAL

fluid collection, segmental lavage with two 50 mL aliquots of warmed sterile normal saline in each of two segments for a total instilled volume of 200 mL with typical return ranging from 80-120 mL. Lavage fluid was filtered through 2-ply gauze. BAL fluid was collected into 15 mL polypropylene conical tubes and centrifuged at 300 x g for 5 minutes at 4°C. 1 mL of BAL supernatant was aliquoted into 1.5 mL Eppendorf O-ring tubes and stored at -80°C. For collection of adult blood plasma, blood was collected in K2 EDTA purple top tubes using a 21-gauge needle. Tubes were centrifuged at 1000 x g for 10 minutes, aliquoted into 250 µL volumes, and stored at -80°C. 400 µL of BAL fluid was extracted with the Qiagen miRNEasy Mini kit as per the manufacturer's protocol and eluted in 50 µL water. 200 µL of adult blood plasma was extracted with the mirVana PARIS kit as per the manufacturer's protocol and eluted into 100 µL water. 2 µL of input RNA was used for reverse transcription.

We performed quantitative PCR (qPCR) for selected miRNAs using methods described previously (Moltzahn et al., 2011 and Seumois et al., 2012). In brief, we applied a stem-loop-based multiplex quantitative reverse transcription PCR method and pre-amplifications (14 cycles), followed by purification of multiplex PCR products and subsequent uniplex analysis on a microfluidics chip (Fluidigm, South San Francisco, CA). An amplification curve quality score of ≥ 0.5 was required; otherwise, the reaction was considered to have failed amplification.

Technical duplicates were run and Ct's averaged as the initial step. Samples then underwent global mean normalization (Mestdagn et al., 2009) using all miRNAs that yielded a Ct in any given sample. As all of the above steps were done on Ct data, the normalized expression values were considered to be on a log base 2 scale. The ratios plotted in Figure S2 were calculated by subtracting BAL fluid normalized values from adult blood plasma values within each platform. To generate p-values, we performed Student's t-test for each miRNA. P-values were adjusted by FDR to correct for multiple testing.

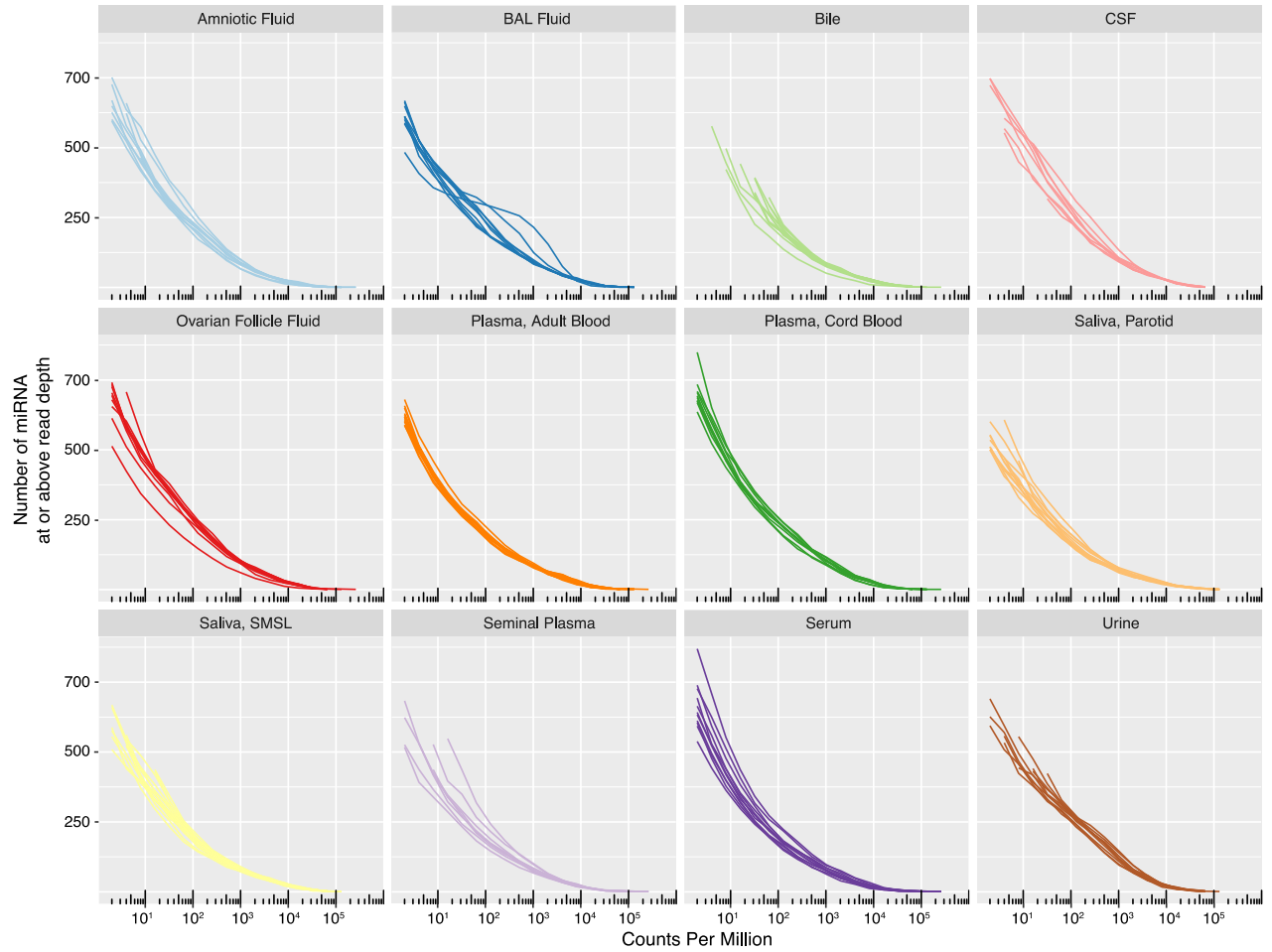


Figure S1. miRNA profiles in Individual Samples for All 12 Biofluid Types, related to Figure 2.

Plots show number of miRNAs detected as a function of read depth.

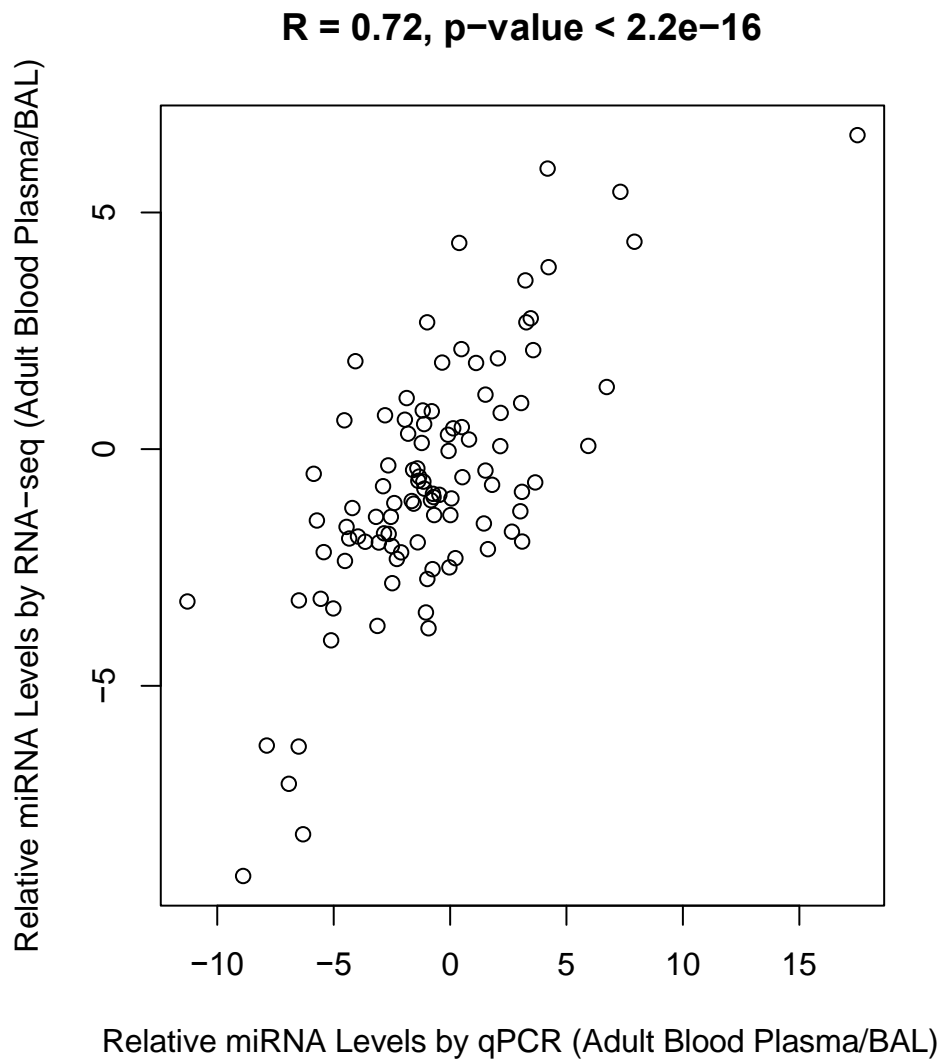
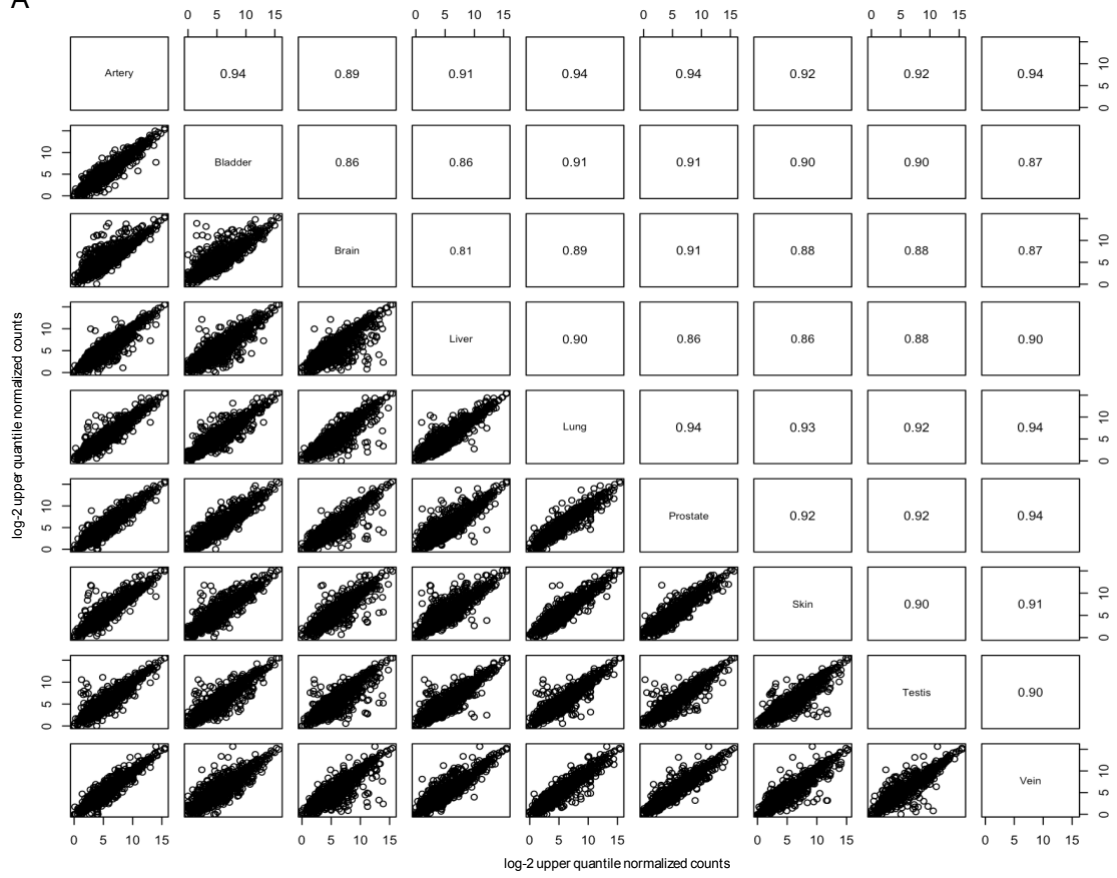


Figure S2. qPCR Validation of RNA-seq Measurements of miRNAs, related to Figure 2. Each point represents one of the 103 miRNAs detected by qPCR (global-mean-normalized Ct > -10) and RNA-seq (normalized reads > 4).

A



B

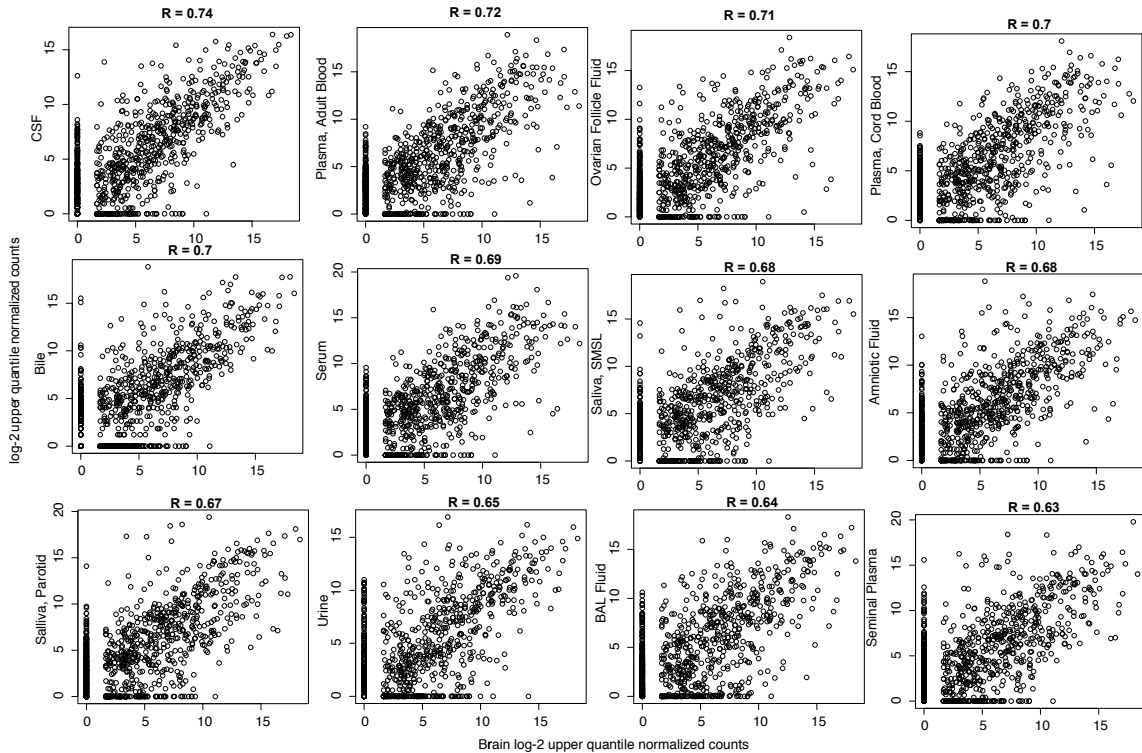


Figure S3. Comparisons of miRNA Profiles from Nine Tissues and Organs and 12 Biofluids and Brain, related to Figure 2.

(A) Data were downloaded from the Human miRNA Tissue Atlas (Ludwig et al., 2016 and <https://ccb-web.cs.uni-saarland.de/tissueatlas>). Each dot represents log-transformed quantile normalized microarray signal intensity for a single miRNA. Values above the diagonal represent Pearson correlation coefficients for each pairwise comparison.

(B) Each dot represents log-transformed upper quantile normalized small RNA-seq read counts. Values above the diagonal represent Pearson correlation coefficients for each pairwise comparison

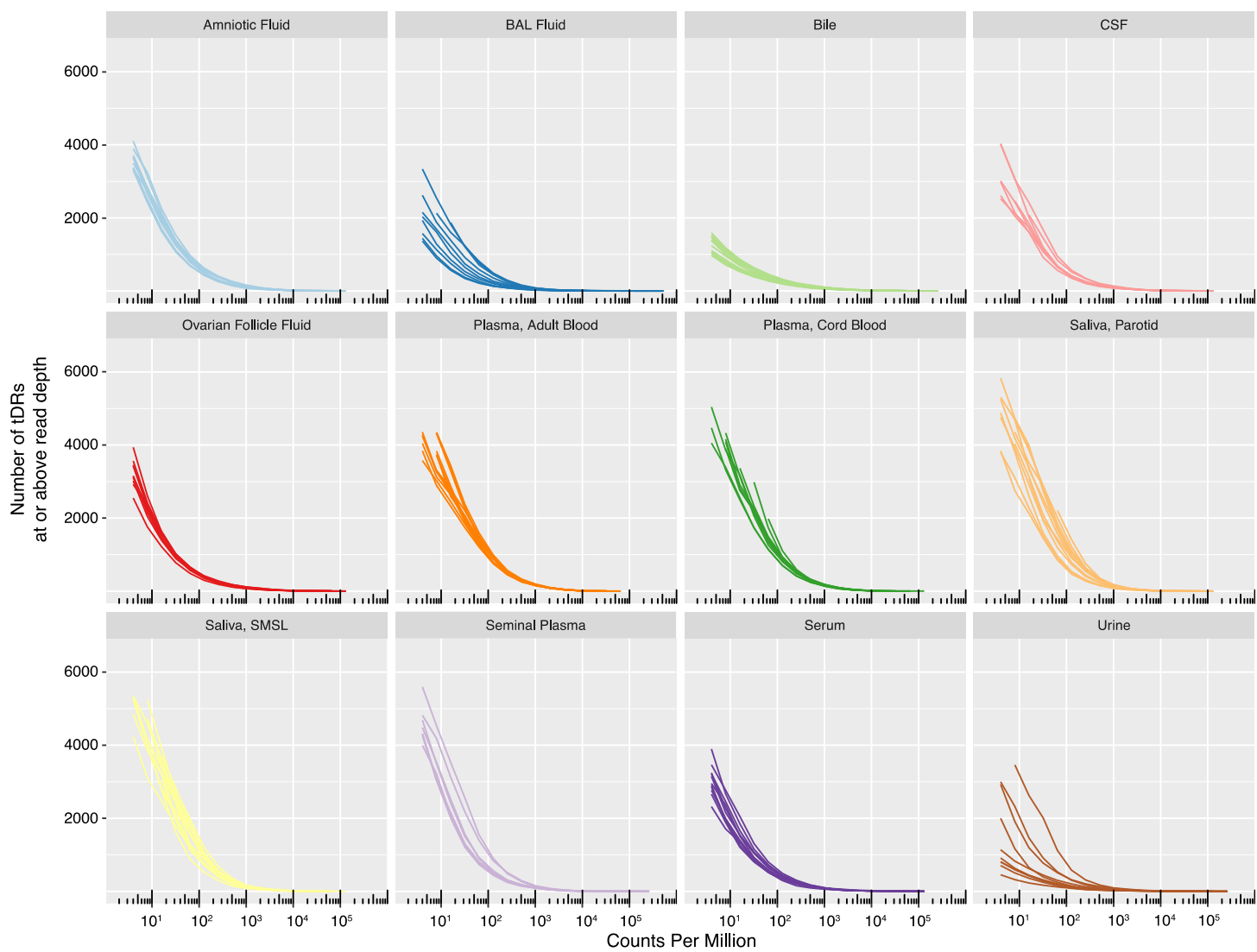


Figure S4. tDR profiles in Individual Samples for All 12 Biofluid Types, related to Figure 4. Plots show number of tDRs detected as a function of read depth.

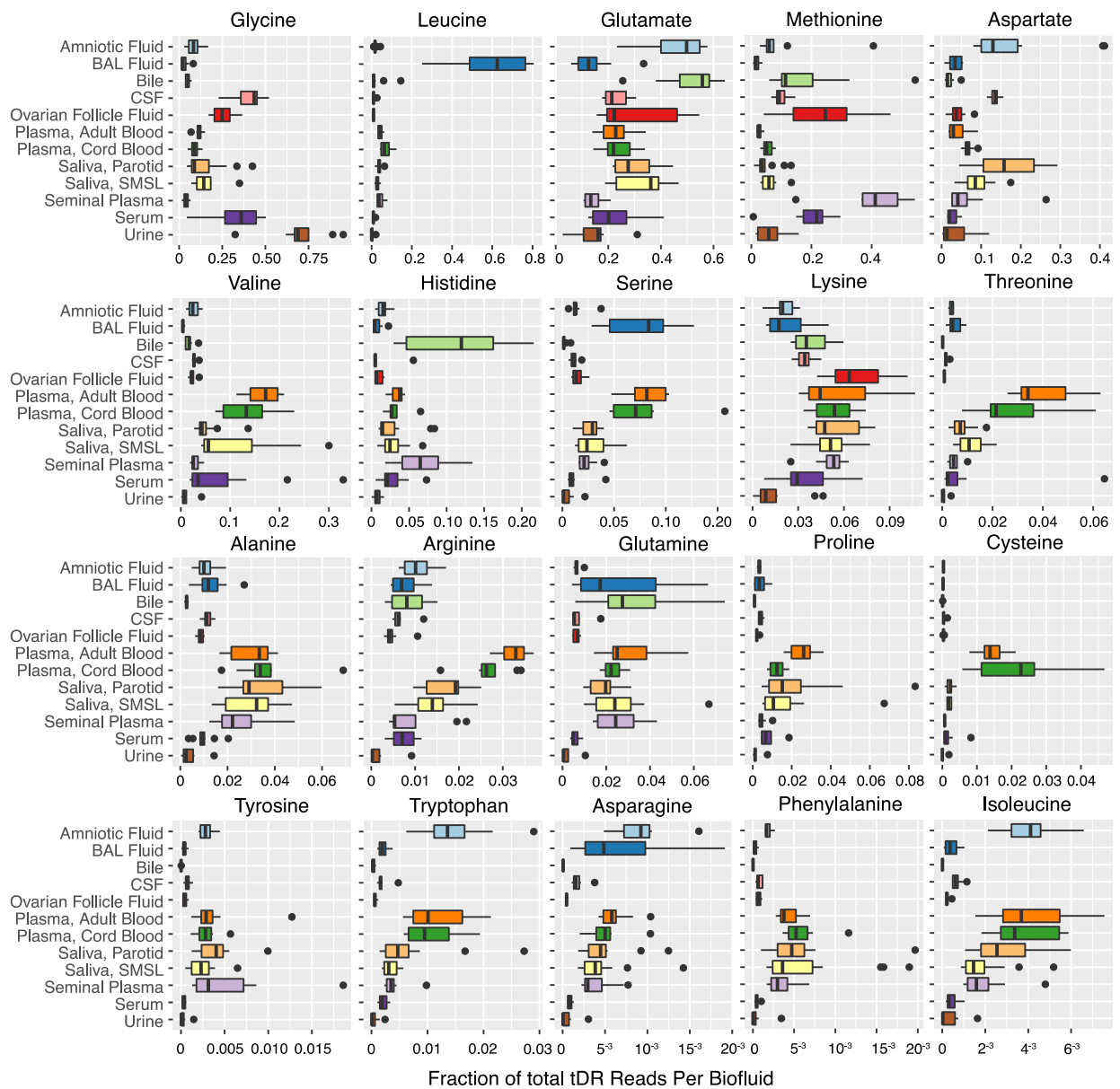


Figure S5. tDR Abundance by Amino Acid, related to Figure 5.

Fractions represent the ratio of alignments to tRNA corresponding to each amino acid divided by all tRNA alignments for that sample.

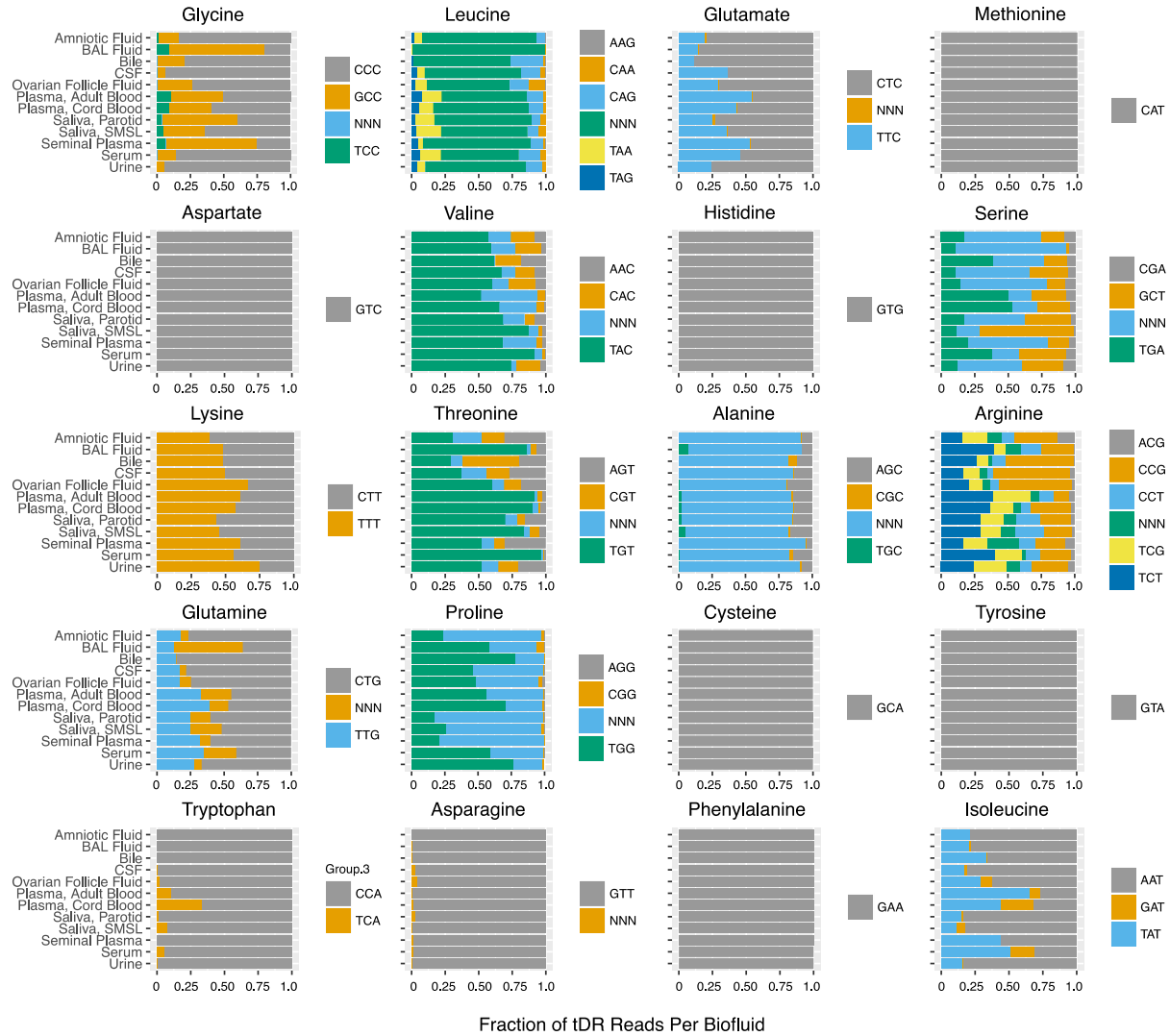


Figure S6. tDR Abundance by Anticodon, related to Figure 5.

Each fraction represents the number of alignments to a particular anticodon divided by the number of reads aligning to each amino acid. Values represent means for all samples of each biofluid type.

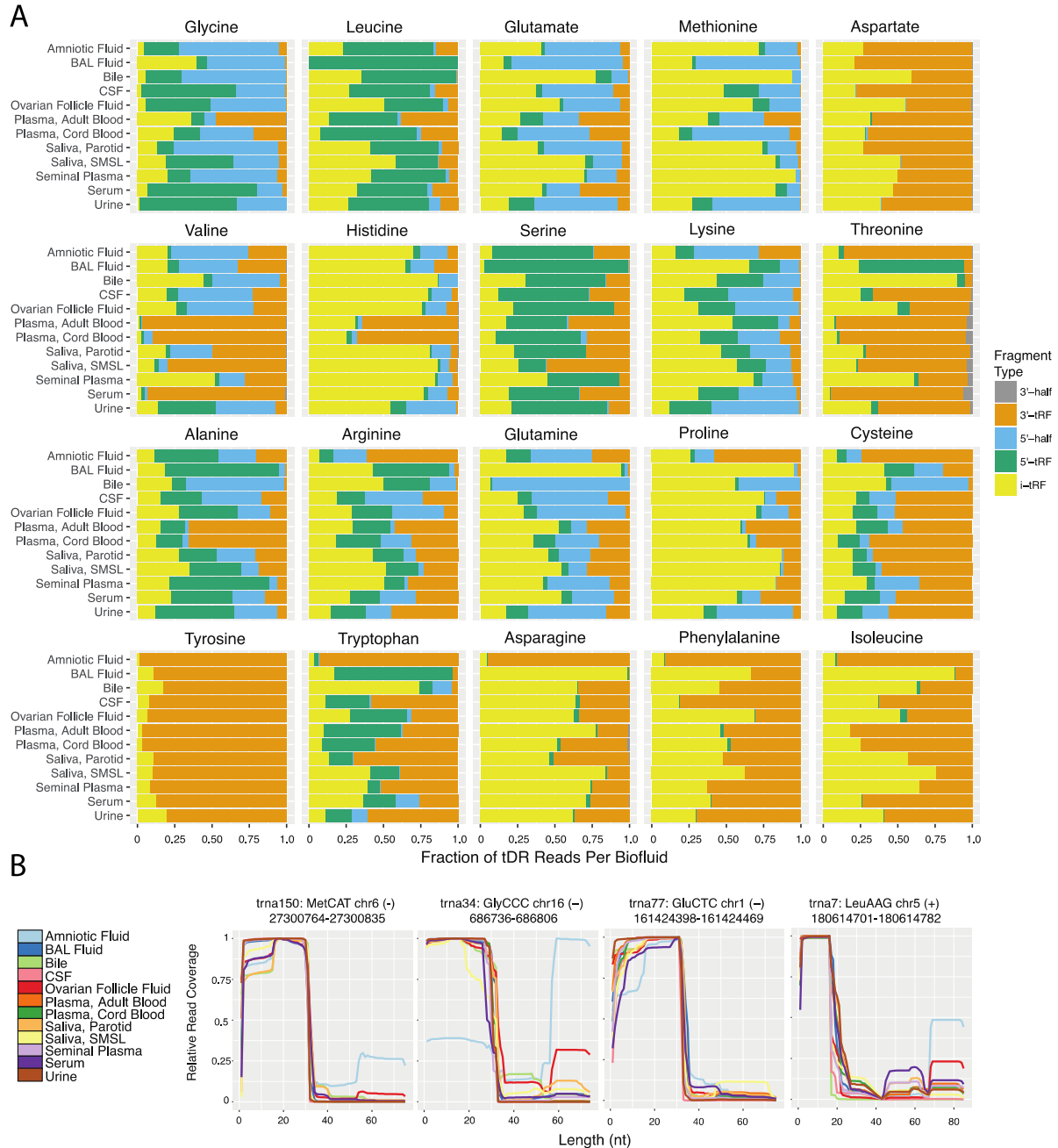


Figure S7. tDR Abundance by Fragment Type and tDR Coverage Maps, related to Figure 5.

(A) Each fraction represents the number of alignments to a particular fragment type divided by the number of reads aligning to each amino acid. Values represent means for all samples of each biofluid type.

(B) Coverage represents the fraction of reads that map to a particular tRNA gene that include a given position in the full-length tRNA. Values are normalized relative to the position with the highest coverage in each biofluid.

Table S5. miRNAs with Much Higher Expression in One Biofluid^a. Related to Figures 2 and 3.

	Amniotic Fluid	BAL	Bile	CSF	Ovarian Follicle Fluid	Plasma, Adult Blood	Plasma, Cord Blood	Serum	Seminal Plasma	Saliva, Parotid	Saliva, SMSL	Urine
<i>Highest in amniotic fluid</i>												
miR-483-5p	83650.5	50.4	4116.2	915.9	5350.7	40.9	858.1	329.2	87.8	13.5	7.2	184.9
miR-194-5p	35054.7	52.1	738.3	398.6	142.6	268.0	331.8	220.8	186.0	168.1	152.8	1599.0
miR-1247-5p	10058.4	2.2	47.2	51.9	113.4	2.6	3.6	18.9	5.6	3.0	1.0	1.0
miR-433-3p	2324.9	59.8	88.0	170.6	99.4	39.8	202.7	34.3	1.3	7.2	2.7	167.4
<i>Highest in BAL</i>												
miR-146b-5p	812.2	62270.1	1295.5	866.4	1093.7	1461.3	1236.6	1087.1	2234.0	434.5	483.8	1736.2
<i>Highest in bile</i>												
miR-27a-5p	15.0	1.4	4775.9	1.0	1.3	1.0	2.4	5.9	3.4	18.9	23.2	3.1
miR-219a-2-3p	1.0	1.0	3707.2	44.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
<i>Highest in CSF</i>												
miR-9-3p	7.8	7.6	89.3	4707.5	8.1	10.5	14.9	5.0	37.3	7.1	3.4	113.5
miR-1298-5p	1.0	1.0	1.0	3099.7	1.3	1.0	1.0	1.0	1.0	1.0	1.0	25.8
miR-1911-5p	1.0	1.0	1.2	1283.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	10.7
<i>Highest in ovarian follicle fluid</i>												
miR-132-3p	156.4	185.6	243.7	433.2	11088.5	70.3	89.2	139.4	309.5	240.2	234.1	673.4
miR-503-5p	112.2	8.0	3.4	35.8	3057.7	55.2	136.5	25.7	14.9	20.7	22.5	27.5
miR-202-3p	3.9	1.0	1.0	1.0	2387.9	1.0	7.9	4.3	53.8	1.0	3.5	1.5
<i>Highest in seminal plasma</i>												
miR-891a-5p	53.8	3.3	196.6	4.4	8.4	1.0	1.0	9.0	10613.2	2.8	78.7	139.9
<i>Highest in urine</i>												
miR-7706	99.6	280.1	15.0	85.9	15.1	9.6	8.5	33.4	8.3	66.8	22.5	3255.6

^amiRNAs with ≥ 1000 reads/ 10^6 total miRNA reads, >10 -fold higher in one biofluid than all other biofluids, and adjusted $p < 0.05$ for pairwise comparisons with all other biofluids by negative binomial Wald Test. Values represent median reads per million total miRNA reads for each biofluid type.

Table S6. miRNAs with Different Expression in Umbilical Cord Versus Adult Blood Plasma^a. Related to Figure 2.

miRNA	Cord blood plasma (reads/million total miRNA reads)	Adult blood plasma (reads/million total miRNA reads)	Fold difference cord/adult blood plasma
miR-487b-3p	1113.4	131.9	8.44
miR-376c-3p	1240.2	170.0	7.29
miR-127-3p	2929.4	442.5	6.62
miR-224-5p	1026.9	245.5	4.18
miR-409-3p	1376.2	350.9	3.92
miR-145-5p	1488.6	397.9	3.74
miR-143-3p	3596.3	1726.1	2.08
miR-25-3p	9187.4	5168.0	1.78
miR-484	4874.8	2750.3	1.77
miR-186-5p	2239.6	1350.8	1.66
miR-363-3p	1351.1	833.7	1.62
miR-652-3p	1827.7	1138.2	1.61
miR-103a-3p	3224.8	2163.2	1.49
let-7f-5p	3754.5	5685.1	0.66
let-7d-3p	724.9	1318.8	0.55
miR-151a-5p	850.7	1960.7	0.43
miR-92a-3p	18762.1	48018.4	0.39
let-7b-5p	350.1	4023.0	0.09

^aAll miRNAs that were significantly different between cord and adult blood plasma (adjusted $p < 0.05$ by the negative binomial Wald test as implemented in DESeq2).

Table S7. Y-RNA Fragments in 12 Biofluid. Related to Figure 6.

Amniotic Fluid	Mean	Median	1st Quartile	3rd Quartile	BAL Fluid	Mean	Median	1st Quartile	3rd Quartile	Bile	Mean	Median	1st Quartile	3rd Quartile
RNY1-201	9.0%	9.0%	6.9%	9.9%	RNY1-201	6.8%	6.4%	5.5%	8.0%	RNY1-201	4.9%	4.8%	3.6%	5.9%
RNY3-201	11.8%	11.7%	9.1%	15.0%	RNY3-201	1.5%	1.2%	0.9%	2.1%	RNY3-201	8.6%	6.0%	3.3%	10.4%
RNY4-201	69.4%	67.9%	67.5%	69.4%	RNY4-201	43.1%	44.7%	37.8%	47.2%	RNY4-201	82.4%	85.9%	77.6%	87.4%
RNY5-201	9.8%	9.6%	8.8%	12.7%	RNY5-201	48.6%	46.8%	43.7%	54.9%	RNY5-201	4.1%	2.8%	2.1%	4.4%
CSF	Mean	Median	1st Quartile	3rd Quartile	Ovarian Follicle Fluid	Mean	Median	1st Quartile	3rd Quartile	Plasma, Adult Blood	Mean	Median	1st Quartile	3rd Quartile
RNY1-201	32.5%	33.8%	28.7%	41.1%	RNY1-201	6.8%	6.3%	4.6%	8.1%	RNY1-201	1.0%	0.5%	0.4%	0.9%
RNY3-201	1.2%	1.0%	0.8%	1.3%	RNY3-201	5.3%	4.4%	3.4%	7.3%	RNY3-201	0.3%	0.3%	0.2%	0.3%
RNY4-201	64.9%	64.6%	57.3%	67.7%	RNY4-201	55.8%	53.9%	48.9%	59.5%	RNY4-201	96.4%	97.4%	96.4%	97.9%
RNY5-201	1.4%	0.8%	0.7%	1.7%	RNY5-201	32.1%	31.8%	30.7%	37.3%	RNY5-201	2.3%	1.7%	1.5%	2.5%
Plasma, Cord Blood	Mean	Median	1st Quartile	3rd Quartile	Saliva, Parotid	Mean	Median	1st Quartile	3rd Quartile	Saliva, SMSL	Mean	Median	1st Quartile	3rd Quartile
RNY1-201	5.0%	3.9%	1.8%	5.5%	RNY1-201	10.1%	8.6%	7.7%	10.8%	RNY1-201	9.1%	9.0%	6.2%	11.5%
RNY3-201	1.0%	0.9%	0.7%	1.4%	RNY3-201	6.4%	7.2%	3.8%	8.7%	RNY3-201	2.7%	2.1%	1.7%	3.0%
RNY4-201	90.6%	93.0%	89.3%	95.4%	RNY4-201	75.6%	77.2%	74.2%	78.5%	RNY4-201	78.9%	79.8%	79.3%	82.9%
RNY5-201	3.5%	2.1%	1.6%	3.9%	RNY5-201	8.0%	6.2%	4.8%	9.2%	RNY5-201	9.2%	7.6%	6.3%	8.3%
Seminal Plasma	Mean	Median	1st Quartile	3rd Quartile	Serum	Mean	Median	1st Quartile	3rd Quartile	Urine	Mean	Median	1st Quartile	3rd Quartile
RNY1-201	24.7%	23.8%	21.6%	28.4%	RNY1-201	3.2%	3.1%	1.9%	3.9%	RNY1-201	9.7%	6.2%	5.6%	14.4%
RNY3-201	13.4%	12.1%	9.7%	16.5%	RNY3-201	1.2%	1.0%	0.9%	1.6%	RNY3-201	7.0%	6.7%	4.3%	7.6%
RNY4-201	59.4%	59.2%	56.5%	62.4%	RNY4-201	89.1%	88.8%	87.0%	90.9%	RNY4-201	72.4%	73.5%	65.9%	81.4%
RNY5-201	2.5%	2.3%	1.8%	2.8%	RNY5-201	6.5%	5.7%	4.6%	7.7%	RNY5-201	10.9%	6.3%	5.3%	14.4%

SUPPLEMENTAL REFERENCES

Albertolle, M.E., Hassis, M.E., Ng, C.J., Cuisson, S., Williams, K., Prakobphol, A., Dykstra, A.B., Hall, S.C., Niles, R.K., Ewa Witkowska, H., *et al.* (2015). Mass spectrometry-based analyses showing the effects of secretor and blood group status on salivary N-glycosylation. *Clin Proteomics* 12, 29.

Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., *et al.* (2016). Distribution of miRNA expression across human tissues. *Nucl Acids Res*, 44, 3865–3877.

Mestdagh, P., Van Vlierberghe, P., De Weer, A., Muth, D., Westermann, F., Speleman, F., and Vandesompele, J. (2009). A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol* 10, R64.

Moltzahn, F., Olshen, A. B., Baehner, L., Peek, A., Fong, L., Stöppler, H., *et al.* (2011). Microfluidic based multiplex qRT-PCR identifies diagnostic and prognostic microRNA signatures in sera of prostate cancer patients. *Cancer Res* 71, 550–560.

Seumois, G., Vijayanand, P., Easley, C. J., Omran, N., Kalinke, L., North, M., *et al.* (2012). An integrated nano-scale approach to profile miRNAs in limited clinical samples. *Am J Clin Exp Immunol* 1, 70–89.