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 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 ≥ 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 ± 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/μL) 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 x 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 ~22-30 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.
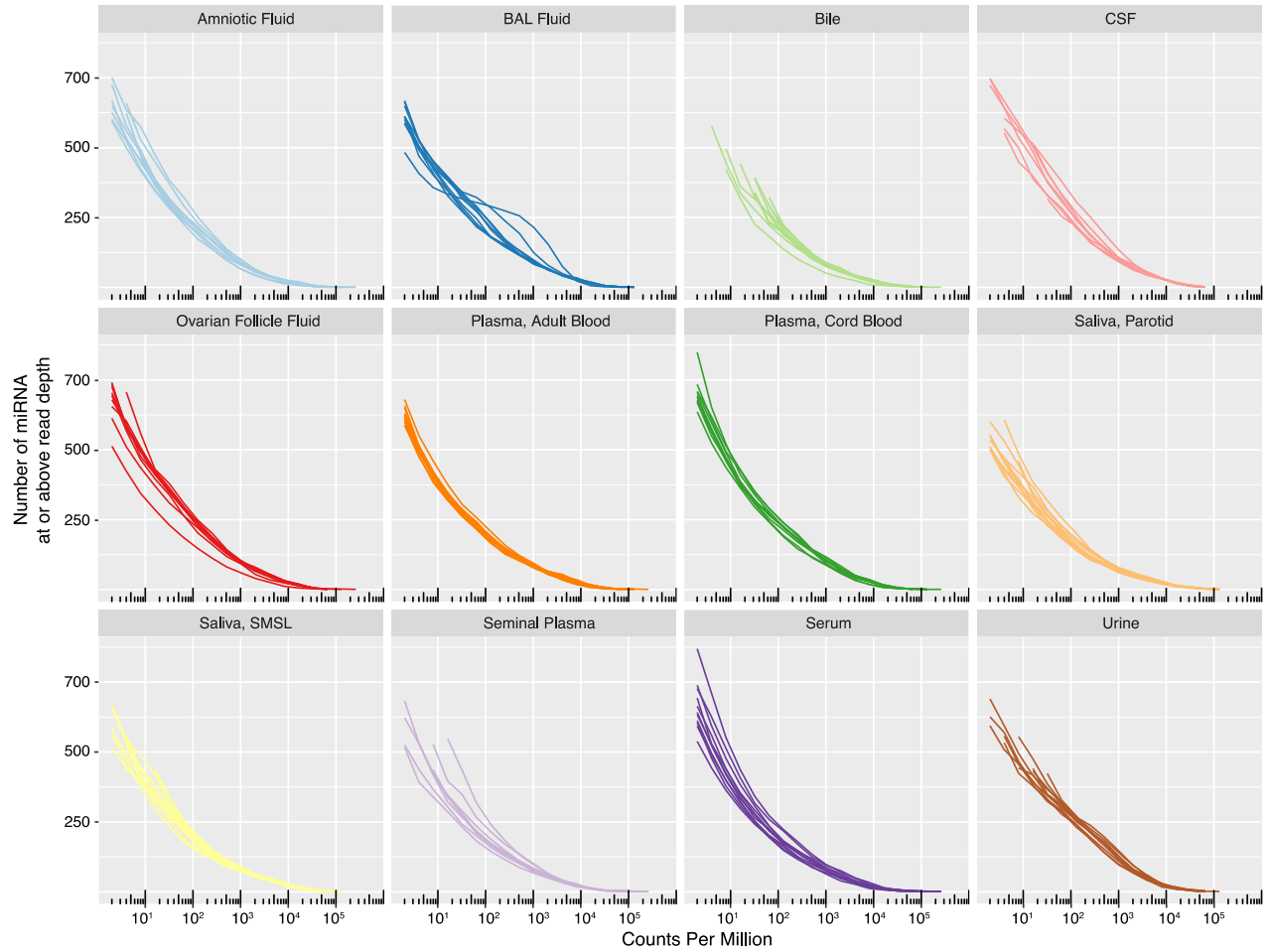
Figure S1. miRNA profiles in Individual Samples for All 12 Biofluid Types, related to Figure 2.
Plots show number of miRNAs detected as a function of read depth.
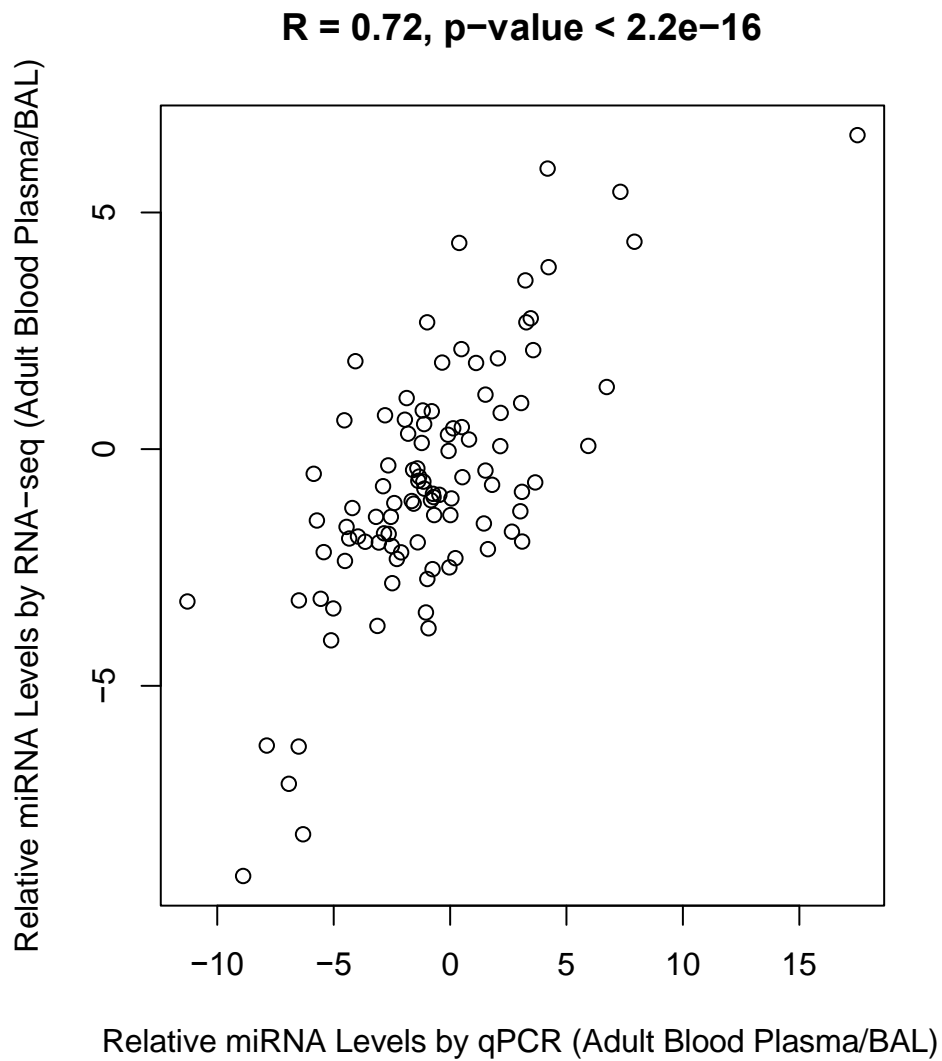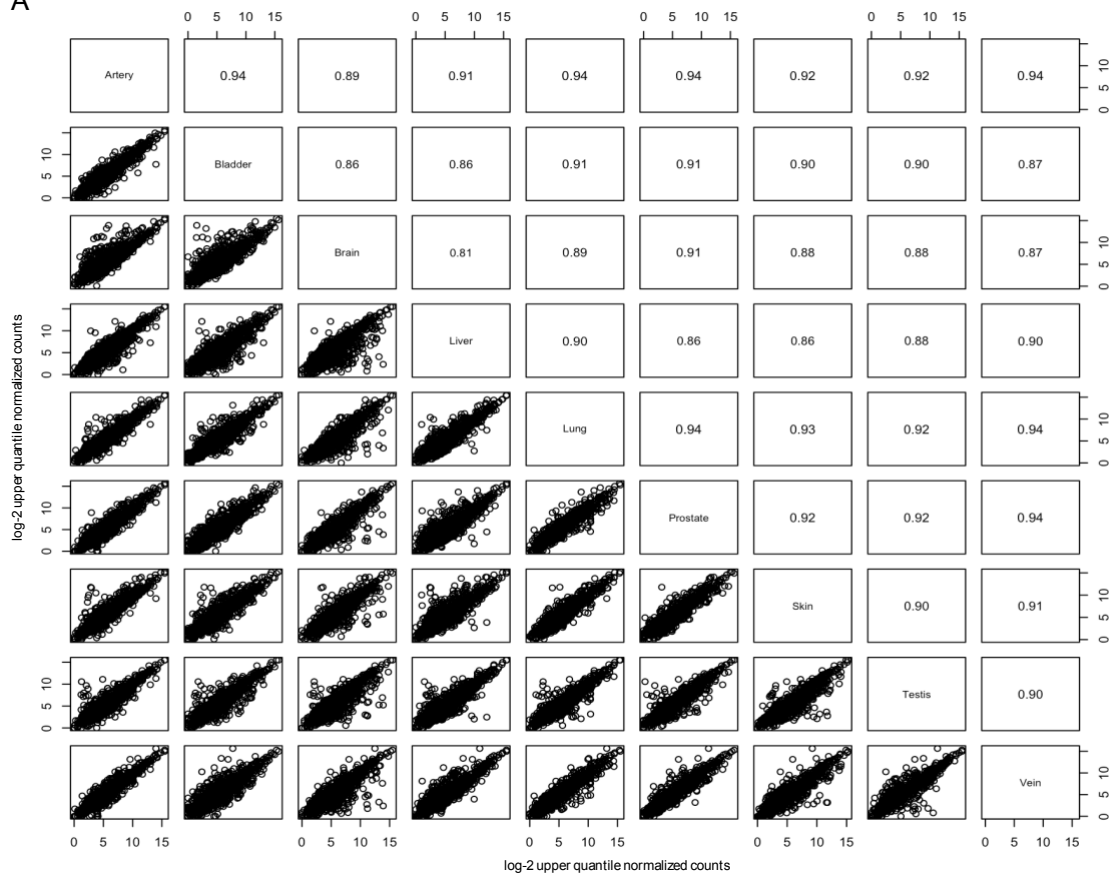
R = 0.72, p−value < 2.2e−16

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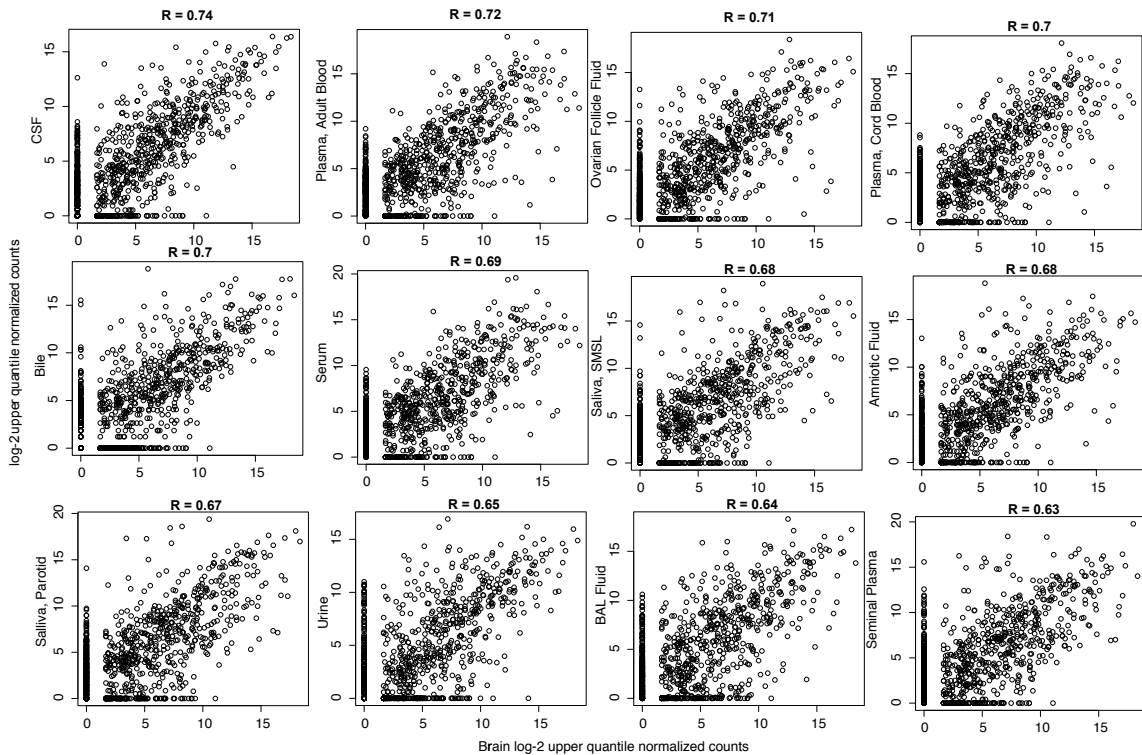
Figure S3. Comparisons of miRNA Profiles from Nine Tissues and Organs and 12 Biofluids and Brain, related to Figure 2.

(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
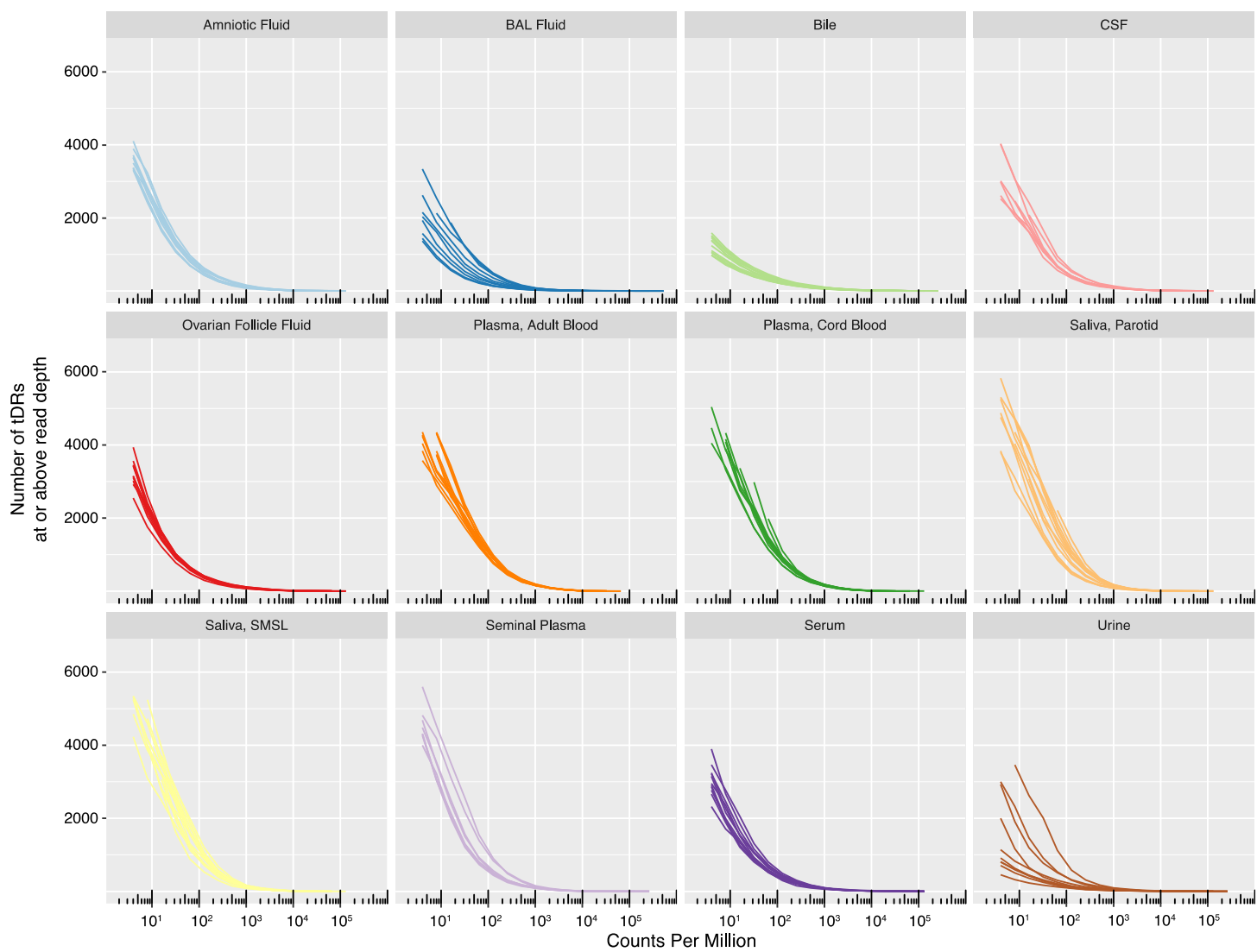
Figure S4. tDR profiles in Individual Samples for All 12 Biofluid Types, related to Figure 4.
Plots show number of tDRs detected as a function of read depth.
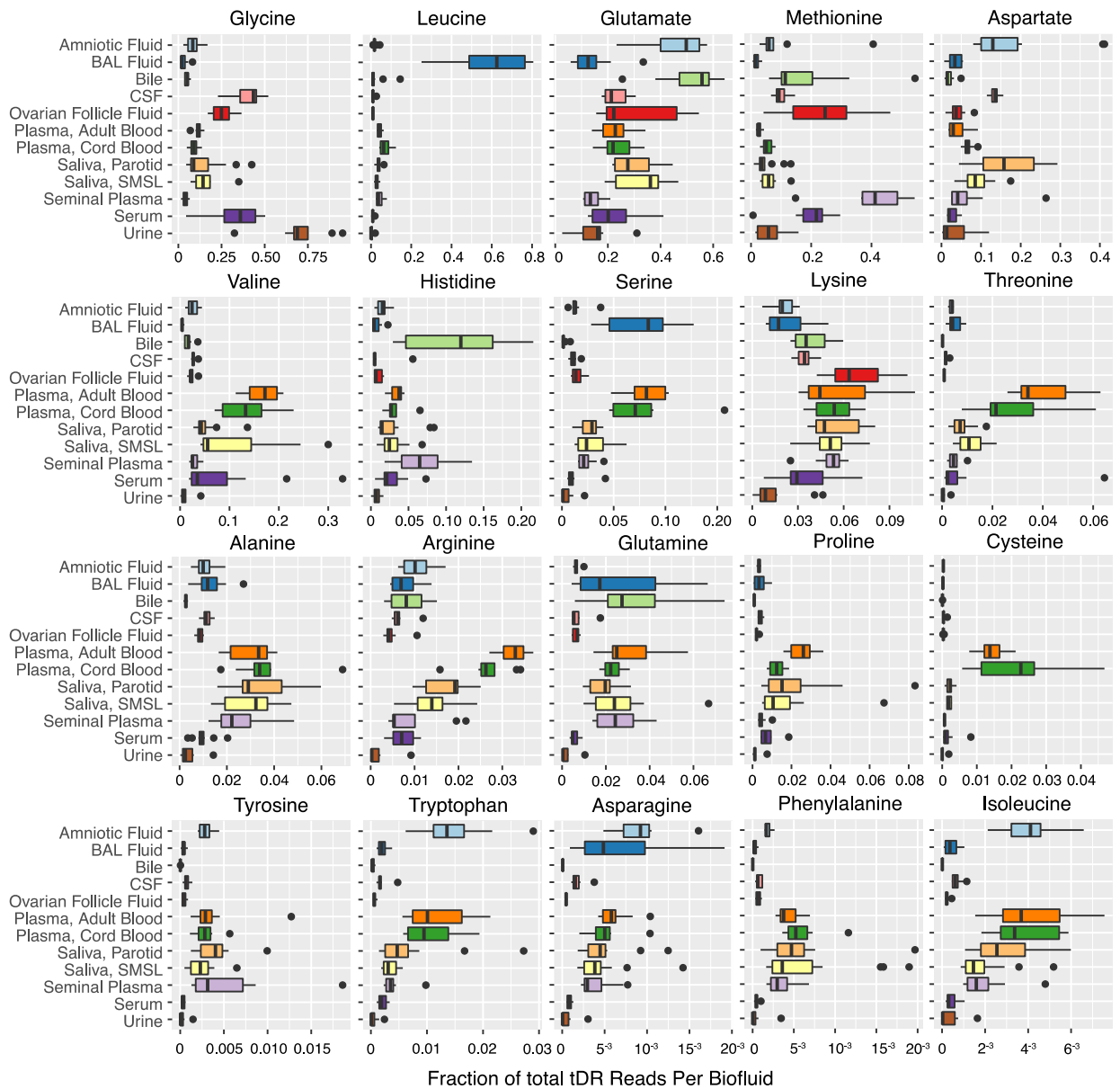
Figure S5. tDR Abundance by Amino Acid, related to Figure 5.
Fractions represent the ratio of alignments to tRNA corresponding to each amino acid divided by all tRNA alignments for that sample.
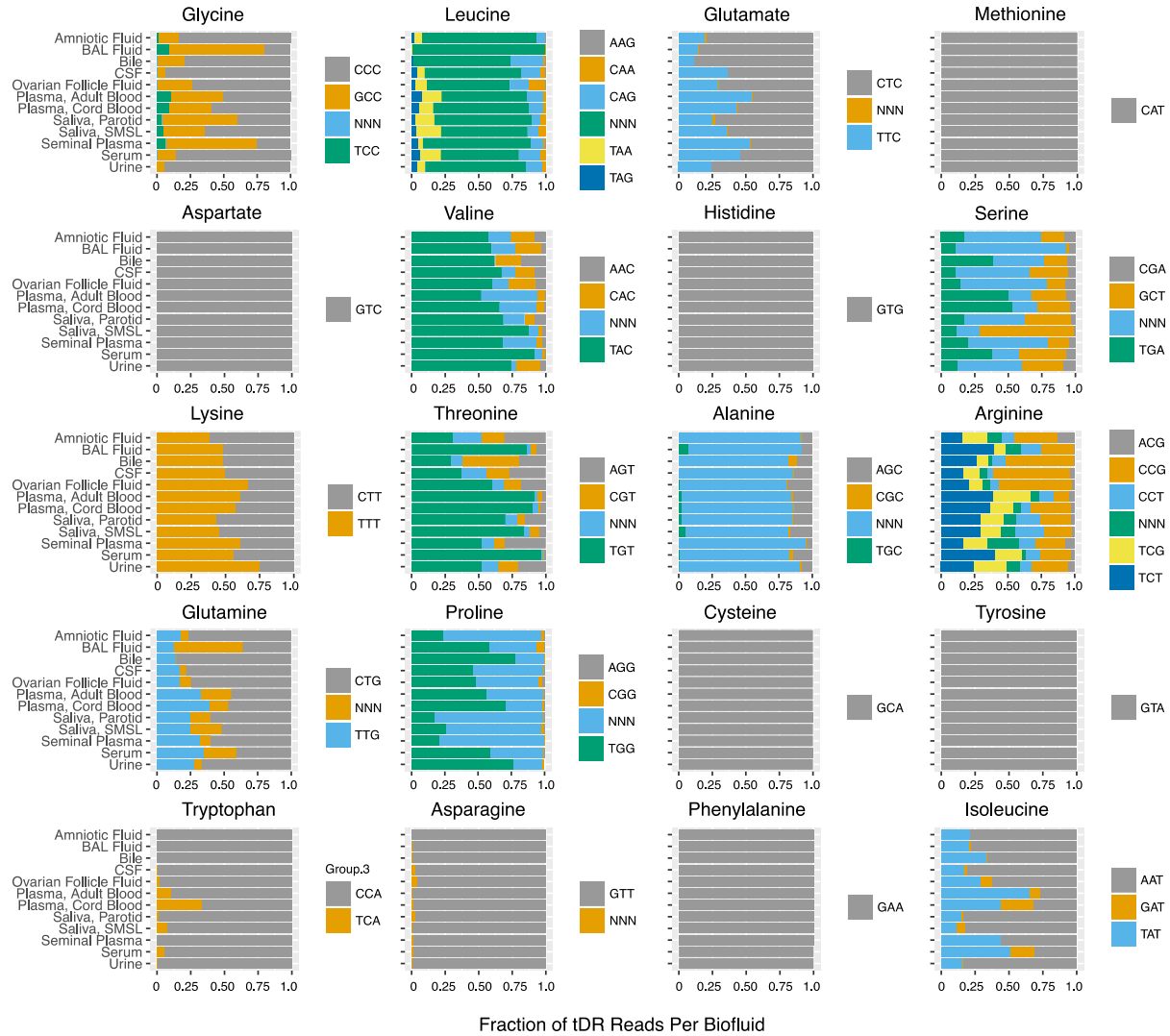
Figure S6. tDR Abundance by Anticodon, related to Figure 5.
Each fraction represents the number of alignments to a particular anticodon divided by the number of reads aligning to each amino acid. Values represent means for all samples of each biofluid type.
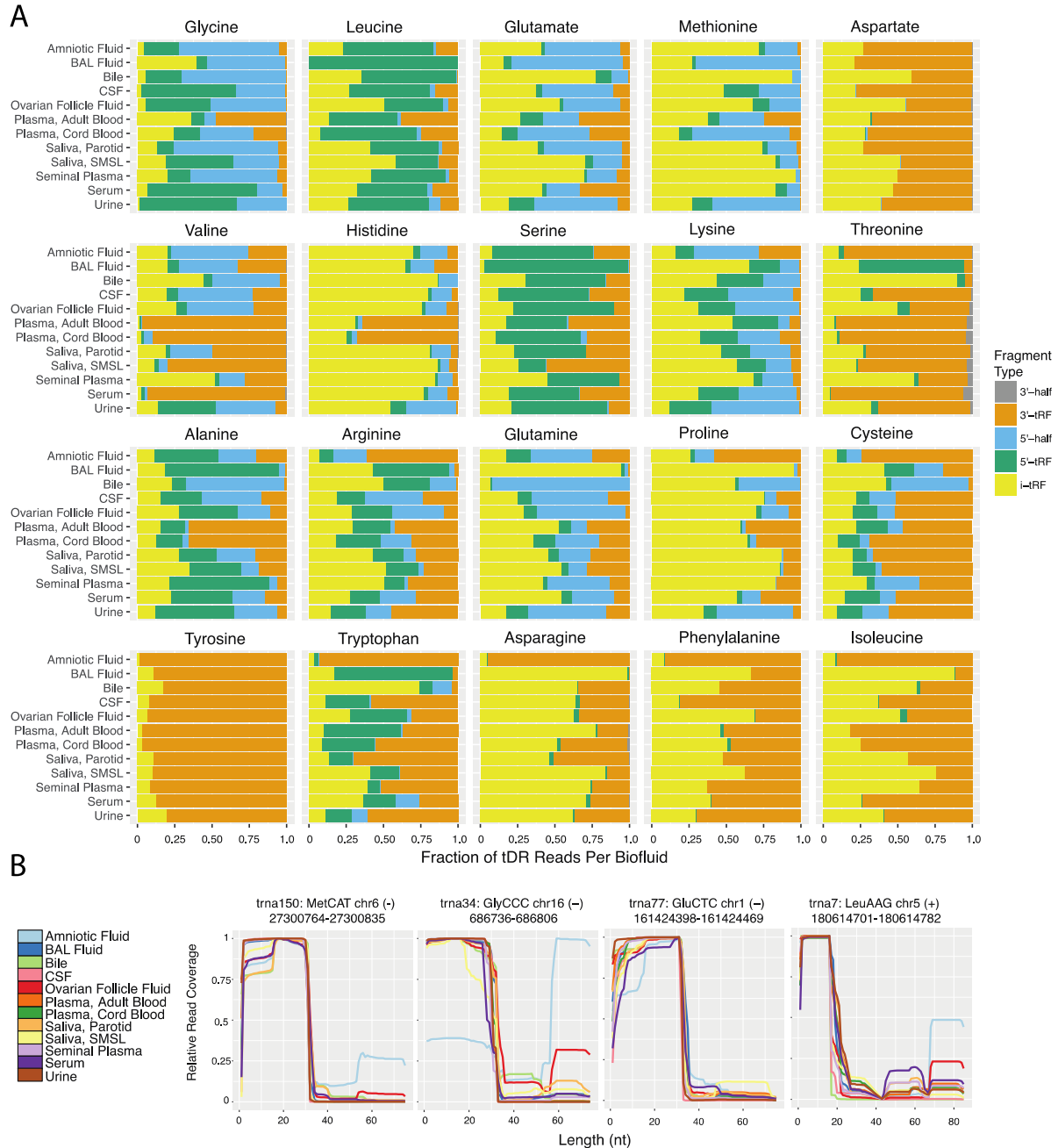
Figure S7. tDR Abundance by Fragment Type and tDR Coverage Maps, related to Figure 5.

(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 bioflu

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
Highest in amniotic fluid												
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
Highest in BAL												
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
Highest in bile												
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
Highest in CSF												
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
Highest in ovarian follicle fluid												
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
Highest in seminal plasma												
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
Highest in urine												
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

[a]miRNAs 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

[a]All 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., Cuison, 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., Eisley, 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.