**Supporting Information Appendix**

**SI Methods**

**Cell lines and growth conditions.** A YBX1-null cell line was previously generated by transiently transfecting a pX330-based plasmid expressing Cas9, a YBX1 guide RNA and a Venus reporter to sort for transfected cells (1). 48-h post-transfection Venus-positive cells were sorted as single cells, clonally expanded and screened for homozygous deletions at the *YBX1* genomic locus by PCR and for expression of YBX1 by immunoblot (2). Wild-type HEK293T and YBX1-null cells were cultured in Dulbecco's modified Eagle medium with 10% fetal bovine serum (Thermo Fisher Scientific, Waltham, MA). For EV production, cells were seeded at ~10% confluency in 150-mm CellBIND tissue culture dishes (Corning, Corning NY) containing 30 ml of growth medium and grown to 80% confluency (~48 h). Cells grown for EV production were incubated in exosome-free medium produced by ultracentrifugation at 100,000Xg (28,000 RPM) for 18 h using an SW-28 rotor (Beckman Coulter, Brea, CA) in a LE-80 ultracentrifuge (Beckman Coulter) to deplete cow EVs present in fetal calf serum. Most abundant human and bovine tRNA species have identical sequences. However, examination of RNY3 and RNY4, each of which has two nucleotide differences between humans and cows, indicated that contamination of HEK293T cell EVs/exosomes by bovine exosomal RNAs was minimal (0.3-1%, 1-1.2%, and 2.5-4% in the EVR1, GRV, and WEX datasets, respectively; Tables S1-S3).

**Extracellular vesicle and exosome purification.** Conditioned medium (210 ml) was harvested from 80% confluent HEK293T cultured cells. All subsequent manipulations were performed at 4 ºC. Cells and large debris were removed by centrifugation in a Sorvall R6+ centrifuge (Thermo Fisher Scientific) at 1,500Xg for 20 min followed by 10,000Xg for 30 min in 500-ml vessels

using a fixed angle FIBERlite F14-6X500y rotor (Thermo Fisher Scientific). The supernatant fraction was then passed through a 0.22 μM polystyrene vacuum filter (Corning) and centrifuged at ~100,000Xg (26,500 RPM) for 1.5 h using two SW-28 rotors. The pellet material was resuspended by adding 500 μl of phosphate buffered saline, pH 7.4 (PBS), to the pellet of each tube followed by trituration using a large bore pipette over a 30-min period at 4 ºC. The resuspended material was washed with ~5 ml of PBS and centrifuged at ~120,000Xg (36,500 RPM) in an SW-55 rotor (Beckman Coulter). Washed pellet material was then resuspended in 200 μl PBS as in the first centrifugation step and 1 ml of 60% sucrose buffer (20 mM Tris-HCl, pH 7.4, 137 mM NaCl) was added and vortexed to mix the sample evenly. The sucrose concentration in the PBS/sucrose mixture was measured by refractometry and, if necessary, additional 60% sucrose buffer was added until the concentration was >48%. Sucrose buffers (40%, 20% and 0% - 1 ml each) were sequentially overlaid on the sample and the tubes were centrifuged at ~150,000Xg (38,500 RPM) for 16 h in an SW-55 rotor. The 20/40% interface fraction was collected and either subjected to nuclease protection assays or, for preparation of $CD63^+$ exosomes, diluted 1:5 with PBS followed by addition of 1 μg of rabbit polyclonal anti-CD63 H-193 (Santa Cruz Biotechnology, Dallas, TX) per liter of original conditioned medium and mixed by rotation for 2 h at 4 ºC. Magvigen protein-A/G conjugated magnetic beads (Nvigen, Sunnyvale, CA) were then added to the exosome/antibody mixture and mixed by rotation for 2 h at 4 ºC. Beads with bound exosomes were washed three times in 1 ml PBS and RNA was extracted by using a Direct-Zol RNA mini-prep kit (Zymo Research, Irvine, CA). Isolated EVs were routinely assessed for quality by sizing and quantitation on a Nanosight NS300 system (Malvern Instruments, UK). EVs generally ranged from 50-150 nm with a mean of 90-100 nm. Samples showing distinct peaks above and below this range were discarded.

**Nuclease-protection experiments.** Post-flotation EV fractions from the 20/40% sucrose gradient interface (100 µl) were mixed with Triton X-100 (TX-100; 10 µl 10% in 20 mM Tris-HCl, pH 7.4, 137 mM NaCl) or buffer alone. The mixture was then briefly mixed by vortexing and incubated on ice for 30 min. Proteinase K was added to a final concentration of 10 µg/ml to some samples, incubated on ice for 30 min and inactivated by the addition of phenylmethlysufonyl fluoride (PMSF) (5 mM). RNase $I_f$ (40 U) and 11 µl of New England Biolabs Buffer 3 were then added and incubation was continued at 30 ºC for 20 min. Alternatively, Turbo DNase (2 U; Thermo Fisher Scientific) and 11 µl of Turbo DNase buffer were added and the solution was incubated at 37 ºC for 30 min. Enzymes were inactivated by the addition of 700 µl of Trizol and RNA extraction was performed with the Direct-Zol RNA mini kit (Zymo Research).

**RNA-seq library preparation using thermostable group II intron reverse transcriptase (TGIRT-seq).** TGIRT-seq libraries were prepared from 1-3 ng of EV or $CD63^+$-exosomal RNA extracted as described above using the TGIRT total RNA-seq method (3-5). Where indicated, RNAs were treated with DNase I (Zymo Research, manufacturer's conditions) prior to library construction. For the GEV-3'P dataset (Table S2), RNAs were treated with T4 polynucleotide kinase (Epicentre, Madison,WI) to remove any 3' phosphates from the RNA fragments (4). For RNA-seq library construction, TGIRT template-switching reverse transcription reactions were performed with an initial template-primer substrate consisting of a 34-nt RNA oligonucleotide (R2 RNA), which contains an Illumina Read 2 primer-binding site and a 3'-blocking group (C3 Spacer, 3SpC3; IDT), annealed to complementary 35-nt DNA primers (R2R DNAs) that have an equimolar mixture of A, C, G, or T single-nucleotide 3' overhangs. Reactions were performed in

20 µl of reaction medium containing the RNA, 100 nM template-primer substrate, 1 µM GsI-IIC RT (TGIRT-III; InGex, St. Louis, MO) or 2 µM TeI4c RT (TeI4c-ΔEn fusion protein RT (5)), and 1 mM dNTPs (an equimolar mix of dATP, dCTP, dGTP, and dTTP) in 450 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris-HCl, pH 7.5, plus 5 mM dithiothreitol (DTT) for GsI-IIC RT or 1 mM DTT for TeI4c RT. Reactions were assembled by adding all components, except dNTPs, to a sterile PCR tube containing RNAs with the TGIRT enzyme added last. After pre-incubating at room temperature for 30 min, reactions were initiated by adding dNTPs and incubated for 15 min at 60 °C. cDNA synthesis was terminated by adding 5 M NaOH to a final concentration of 0.25 M followed by incubation at 95 °C for 3 min and finally neutralized with 5 M HCl. The resulting cDNAs were purified with a MinElute Reaction Cleanup Kit (Qiagen) and ligated at their 3' ends to a 5'-adenlyated/3'-blocked (C3 spacer, 3SpC3; IDT) adapter containing the complement of an Illumina Read 1 primer binding site (R1R) using Thermostable 5' AppDNA/RNA Ligase (New England Biolabs) according to the manufacturer's recommendations. The ligated cDNA products were re-purified with a MinElute column and amplified by PCR with Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific) and 200 nM of Illumina multiplex and 200 nM of barcode primers (a 5' primer that adds a P5 capture site and a 3' primer that adds an Illumina sequencing barcode and P7 capture site). PCR was performed with an initial denaturation at 98 °C for 5 sec followed by 12 cycles of 98 °C for 5 sec, 60 °C for 10 sec and 72 °C for 10 sec. The PCR products were purified by using Agencourt AMPure XP beads (Beckman Coulter) to remove primer dimers and sequenced on an Illumina NextSeq 500 instrument to obtain the indicated number of 75-nt paired-end reads or 150-nt single-end reads; in one experiment, an Illumina HiSeq 2500 was used to obtain the indicated number of 125-nt paired-end reads (Tables S1 to S3).

The RNA-seq libraries of HEK293T cellular RNAs were constructed similarly. Cellular RNAs (1-1.5 μg) were treated with DNase I (Zymo Research) according to the manufacturer's protocol, followed by rRNA depletion using a RiboZero™ Gold Kit (Human/Mouse/Rat) (Epicentre). The resulting RNAs (50 ng) were either used directly in the TGIRT template-switching reverse transcription reaction or were fragmented to a size predominantly between 70~100 nt by using an NEBNext® Magnesium Fragmentation Module (New England Biolabs). The fragmented RNAs were then treated with T4 polynucleotide kinase (Epicentre) to remove 3' phosphates, cleaned up with an RNA Clean & Concentrator™ Kit (Zymo Research) and used for RNA-seq library construction with TGIRT enzymes as described above.

The TGIRT-seq datasets described in this manuscript have been deposited in the National Center for Biotechnology Information Sequence Read Archive (http:/www.ncbi.nlm.nih.gov/sra under accession number SRP108712).

**Bioinformatic analysis**. Illumina TruSeq adapters and PCR primer sequences were trimmed from the reads with cutadapt (sequencing quality score cut-off at 20; p-value < 0.01) and reads <15-nt after trimming were discarded. Reads were then mapped with HISAT2 v2.0.2 with default settings to the human genome reference sequence (Ensembl GRCh38 Release 76) combined with additional contigs for 5S and 45S rRNA genes and the *E. coli* genome sequence (Genebank: NC_000913) (denoted Pass 1). The additional contigs for the 5S and 45S rRNA genes included the 2.2-kb 5S rRNA repeats from the 5S rRNA cluster on chromosome 1 (1q42, GeneBank: X12811) and the 43-kb 45S rRNA repeats that contained 5.8S, 18S and 28S rRNAs from clusters on chromosomes 13,14,15,21, and 22 (GeneBank: U13369). Unmapped reads from Pass 1 were re-mapped to Ensembl GRCh38 Release 76 by Bowtie 2 v2.2.6 with local alignment

to improve the mapping rate for reads containing post-transcriptionally added 5' or 3' nucleotides (*e.g.*, CCA and poly(U)), short untrimmed adapter sequences, or non-templated nucleotides added to the 3' end of the cDNAs by TGIRT enzymes (denoted Pass 2). The uniquely mapped reads from Passes 1 and 2 were combined by using Samtools. To process multiply mapped reads, we collected up to 10 distinct alignments with the same mapping score and selected the alignment with the shortest distance between the two paired ends (*i.e.*, the shortest read span). In the case of ties between reads mapping to rRNA and non-rRNA sequences, the read was assigned to the rRNA sequence, and in other cases, the read was assigned randomly to one of the tied choices. Uniquely mapped reads and the filtered multiply mapped reads were combined and intersected with gene annotations (Ensembl GRCh38 Release 76), with RNY5 gene and its 10 pseudogenes, which are not annotated in this release, added manually to generate the counts for individual features. Coverage of each feature was calculated by Bedtools. To avoid mis-counting reads with embedded sncRNAs, reads were first intersected with sncRNA annotations and the remaining reads were then intersected with the annotations for protein-coding genes, lincRNAs, antisense, and other lncRNAs. To further improve the mapping rate for tRNAs and rRNAs, we combined reads that were uniquely or multiply mapped to tRNAs or rRNAs in the initial alignments and re-mapped them to tRNA reference sequences (Genomic tRNA Database, and UCSC genome browser website) or rRNA reference sequences (GeneBank: X12811 and U13369) using Bowtie 2 local alignment. Because similar or identical tRNAs with the same anticodon may be multiply mapped to different tRNA loci by Bowtie 2, mapped tRNA reads were combined according to their tRNA anticodon (N = 55) prior to calculating the tRNA distributions. The 55 anticodons included 48 commonly used tRNA anticodons, iMetCAT, SelCysTCA, two families of suppressive tRNAs with anticodons corresponding to stop codons

(SupCTA and SupTTA), a family of tRNAs with an undetermined anticodon and two predicted tRNAs (CysACA and SerACT). Coverage plots and read alignments were created by using Integrative Genomics Viewer (IGV) (6). Genes with >1,000 mapped reads were down sampled to 1,000 mapped reads in IGV for visualization. For correlation analysis, RNA-seq datasets were normalized for the total number of mapped reads by using DESeq2 (7) and plotted in R. To identify putative mRNAs, we remapped reads that mapped to protein-coding genes in the initial mapping pipeline to the human transcriptome reference (Ensembl GRCh38 releae 76). Reads from different transcript isoforms of the same gene were combined for count normalization in scatter plots. Reads that mapped to putative mRNAs in the human transcriptome reference were retrieved and remapped to the human genome reference using HISAT2 and analyzed by picardtools to calculate the percentage of bases in CDS, UTR, intron, and intergenic regions.

**Reverse transcription PCR.** EVs were isolated and left untreated or treated with RNase $I_f$ and/or detergent as described above. RNA was extracted using the Zymo RNA miniprep kit according to the manufacturer's instructions. 10 ng of EV or whole cell RNA was reverse transcribed using the SuperScript IV cDNA synthesis kit (Thermo Fisher Scientific) with the supplied oligo(dT) primer. PCR was performed using primers that spanned at least two splice junctions (MIF-F: 5'-TGCCGATGTTCATCGTAAACA, MIF-R: 5'-TTAGGCGAAGGTGGAGTTGT; RPS17-F: 5'-ATAGAAAAGTACTACACGCGC, RPS17-R: 5'-TTAGGCGAAGGTGGAGTTGT; RPS2-F: 5'-TATGCCAGTGCAGAAGCAGA, RPS2-R: 5'-TGGTCAGTGAACTCCTGATA) over 35 cycles (95 ºC for 30 sec, 55 ºC for 30 sec and 72 ºC for 30 sec) with a final extension at 72 ºC for 2 min using GoTaq DNA polymerase

(Promega). PCR products were separated on a 2.5% agarose gel and visualized using GelRed stain (Biotium Inc.)


1. Hsu PD*, et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31(9):827-832.
2. Shurtleff MJ, Temoche-Diaz MM, Karfilis KV, Ri S, Schekman R (2016) Y-box protein 1 is required to sort microRNAs into exosomes in cells and in a cell-free reaction. *eLife* 5:e19276.
3. Mohr S*, et al.* (2013) Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* 19(7):958-970.
4. Nottingham RM*, et al.* (2016) RNA-seq of human reference RNA samples using a thermostable group II intron reverse transcriptase. *RNA* 22(4):597-613.
5. Qin Y*, et al.* (2016) High-throughput sequencing of human plasma RNA by using thermostable group II intron reverse transcriptases. *RNA* 22(1):111-128.
6. Robinson JT*, et al.* (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24-26.
7. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550.

**Table S1.** Read statistics and mapping for TGIRT-seq of RNA isolated from HEK293T cell EVs after various treatments.

| Dataset | EVR1 | EVR2 | EVR3 | EVR4 | EVR5 | EVR6 | EVD1 | EVD2 | EVD3 | EVD4 | EVD5 | EVD6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample ID | 6251 | 6252 | 6253 | 6254 | 6255 | 6256 | 6141 | 6142 | 6143 | 6144 | 6145 | 6146 |
| Protease | - | - | + | + | - | - | - | - | + | + | - | - |
| Detergent | - | - | - | - | + | + | - | - | - | - | + | + |
| RNase | - | + | - | + | - | + | - | - | - | - | - | - |
| DNase | - | - | - | - | - | - | - | + | - | + | - | + |
| Total reads ($x10^6$) | 23.4 | 20.5 | 16.9 | 29.5 | 22.6 | 21.0 | 21.7 | 23.6 | 20.1 | 37.6 | 22.0 | 26.9 |
| % of reads after trimming | 79.6 | 66.3 | 69.0 | 54.4 | 61.8 | 51.6 | 96.9 | 86.4 | 94.3 | 81.1 | 94.3 | 90.9 |
| % of mapped reads | 84.0 | 81.4 | 83.2 | 83.7 | 84.4 | 78.3 | 18.1 | 57.6 | 15.9 | 34.6 | 26.6 | 43.8 |
| % reads mapped to features | 70.4 | 69.7 | 61.5 | 68.2 | 53.7 | 53.1 | 56.7 | 67.7 | 62.2 | 64.1 | 58.4 | 54.2 |

TGIRT-seq libraries were prepared by using GsI-IIC RT (TGIRT-III) from RNA extracted from HEK293T cell EVs that were either untreated or treated with protease, detergent (TX-100), RNase or DNase alone or in combination, as indicated in the Table, and then sequenced on an Illumina NextSeq 500 instrument to obtain the indicated numbers of 75-nt paired-end reads. The reads were trimmed to remove adapter sequences and low quality base calls (sequencing quality cutoff 20, $p < 0.01$), and reads <15 nt after trimming were discarded. Trimmed reads were then mapped by using HISAT2 and Bowtie2 to a human genome reference sequence (Ensembl GRCh38) modified to include additional rRNA repeats. The loss of reads after trimming for the EVR samples likely reflects incomplete removal of adapter dimers from libraries prepared from small amounts of EV RNA. The dataset name is given above and the laboratory notebook sample identification is given below.

**Table S2.** Read statistics and mapping for TGIRT-seq of RNA from HEK293T cell EVs using GsI-IIC or TeI4c RT.

| Dataset | GC | GEV | GEV-3'P | TC | TEV |
|---|---|---|---|---|---|
| Sample ID | GMS2 | GMS1 | GMS3 | TMS2 | TMS1 |
| RT | GsI-IIC | GsI-IIC | GsI-IIC | TeI4c | TeI4c |
| RNA | Whole cell | EVs | EVs | Whole cell | EVs |
| Phosphatase | - | - | + | - | - |
| Total reads ($x10^6$) | 55.0 | 38.3 | 36.1 | 52.5 | 58.7 |
| % of reads after trimming | 49.7 | 81.5 | 87.9 | 51.1 | 97.6 |
| % of mapped reads | 95.9 | 64.9 | 75.5 | 92.7 | 59.8 |
| % reads mapped to features | 70.5 | 42.2 | 46.9 | 77.3 | 40.7 |

TGIRT-seq libraries were prepared by using either GsI-IIC RT (TGIRT-III) or TeI4c RT from unfragmented HEK293T whole cell RNA or RNA extracted from EVs from the same culture. The cellular RNA libraries (GC and TC) were sequenced on an Illumina HiSeq 2500 to obtain the indicated numbers of 125-nt pair-end reads, and the EV RNAs (GEV, GEV-3'P and TEV) were sequenced on an Illumina NextSeq 500 to obtain the indicated numbers of 150-nt single-end reads. The reads were trimmed to remove adapter sequences and low quality base calls (sequencing quality cutoff 20, $p < 0.01$), and reads <15 nt after trimming were discarded. Trimmed reads were then mapped by using HISAT2 and Bowtie2 to a human genome reference sequence (Ensembl GRCh38) modified to include additional rRNA repeats. The dataset name is given above and the laboratory notebook sample identification is given below.

**Table S3.** Read statistics and mapping for TGIRT-seq of HEK293T whole cell RNA and $CD63^+$ exosomal RNA.

| Dataset | WC | WCF | WEX | YC | YCF | YEX |
|---|---|---|---|---|---|---|
| Sample ID | MSC1 | MSC1F | MSE1 | MSC2 | MSC2F | MSE2 |
| Fragmentation | - | + | - | - | + | - |
| YBX1 | + | + | + | Δ | Δ | Δ |
| Total reads ($x10^6$) | 120.3 | 104.0 | 94.6 | 103.8 | 90.8 | 96.7 |
| % of reads after trimming | 96.4 | 96.1 | 86.2 | 91.7 | 97.5 | 76.5 |
| % of mapped reads | 82.4 | 87.4 | 82.1 | 81.8 | 84.5 | 84.8 |
| % reads mapped to features | 94.5 | 93.0 | 80.8 | 94.4 | 93.2 | 76.2 |

TGIRT-seq libraries were prepared from DNase-treated HEK293T whole cell RNA (fragmented or unfragmented, as indicated in the Table, and from DNase-treated RNA extracted from $CD63^+$ exosomes from the same cultures by using TGIRT-III and sequenced on an Illumina NextSeq 500 instrument to obtain the indicated numbers of 75-nt paired-end reads. The reads were trimmed to remove adapter sequences and low quality base calls (sequencing quality cutoff 20, $p < 0.01$), and reads <15 nt after trimming were discarded. Trimmed reads were then mapped by using HISAT2 and Bowtie2 to a human genome reference sequence (Ensembl GRCh38) modified to include additional rRNA repeats. The dataset name is given above and the laboratory notebook sample identification is given below.

**Figure S1.** RNase protection of EV-associated transcripts. Pairwise scatter plots of DESeq2-normalized read counts and computed Pearson's correlation coefficients (*upper left* corner) are shown for all treatment conditions.

**Figure S2.** DNase protection of EV-associated transcripts. Pairwise scatter plots of DESeq2-normalized read counts and computed Pearson's correlation coefficients (*upper left* corner) are shown for all treatment conditions.

**Figure S3.** Scatter plots comparing the relative abundance of different tRNA species in EVs and exosomes with that in the cells from which the EVs were isolated. (*A*) Scatter plots of DESeq2-normalized reads counts in EV and unfragmented whole cell RNA (datasets GEVand GC, respectively; Table S2). (*B*) Scatter plots of DESeq2-normalized read counts in exosomes and unfragmented whole cell RNA (datasets WEX and WC, respectively; Table S3). tRNAs are grouped by anticodon and color coded according the type of amino acid. Spearman's correlation coefficients (ρ) are shown in the upper left corner of each plot.

**Figure S4.** Pie charts showing the percentage of different RNA biotypes detected in EVs (*A,B*) after treatment of EVs with RNase or protease plus RNase in the absence of detergents or (*C,D*) before and after treatment of EV RNAs with T4 polynucleotide kinase to remove 3' phosphates prior to RNA-seq library construction. The pie charts show the percentage of mapped total cellular RNA and sncRNA reads (*left* and *right*, respectively) corresponding to the indicated features. The number of genes represented for each biotype is shown in parenthesis. tRNA gene counts are for tRNA genes grouped by anticodon (see Fig. 2 legend and SI Methods), and sncRNA gene counts include pseudogenes.

**Figure S5.** Characterization of additional EV-RNA biotypes by read-span analysis. (*A-F*) Read spans from paired-end sequencing for 7SL RNA (A), snoRNA (B), 7SK RNA (C), Vault RNA (D), 5S and 5.8S rRNA (E) and 18 and 28S rRNA (F) before and after the indicated treatment conditions (keys at *upper right*). (*G,H*) Read spans for tRNAs (G*)* and Y RNAs (H) in EV RNA preparations that were untreated or treated with T4 polynucleotide kinase to remove 3' phosphates prior to RNA-seq. (*I-K*) Reads spans for lincRNAs (I), annotated antisense RNAs (J) and other lncRNAs (K) isolated from EVs before and after the indicated treatments. The plots show the % of total reads of different read spans for the mapped paired-end reads for each biotype. Peaks for different subspecies of snoRNA, Y RNAs, and Vault RNA are indicated on the plots. tRNA* indicates tRNA reads with a stop at position 16 in the D-loop.

**Figure S6.** Integrative Genomics Viewer (IGV) screen shots showing read coverage across the tRNA or sncRNA coding sequence for representative tRNAs and other sncRNAs. The arrow at the top indicates the 5' to 3' orientation of the RNA. Coverage plots are above read alignments with reads sorted by top strand position. The numbers of reads are indicated to the left of the coverage plot and were down sampled to 1,000 reads in IGV for visualization. Nucleotides in reads matching nucleotides in the annotated reference are colored gray. Nucleotides in reads that do not match the annotated reference are color coded by nucleotide (A, green; C, blue; G, brown; and T, red) and may correspond to the complementary nucleotide depending on gene orientation. The 1-methyladenine ($m^1A$), $m^1G$, $m^5C$/Cm, non-coded 5' G residue of HisGTG, and 3' CCA post-transcriptional tRNA modifications are indicated. The D-loop stop at position U16, a non-

coded poly(U) tail at the 3' end of some VTRNA1-1 RNAs, and a previously annotated SNP (dbSNP, NCBI) found in snRNA U5D-1 are also indicated.

**Figure S7.** (*A*) Stacked bar graphs of the percentage of reads mapped to the mRNA transcriptome corresponding to coding sequences (CDS), untranslated regions (UTR), introns, and intergenic regions. (*B*) Scatter plot comparing normalized read counts for mRNAs from DNase-treated exosomal RNA and unfragmented whole cell RNAs. (*C*) Scatter plot comparing normalized read counts for mRNAs from DNase-treated unfragmented and fragmented whole cell RNA. 5' TOP-containing mRNAs are denoted with blue dots and aminoacyl-tRNA synthetase mRNAs are denoted with red dots. *r* = Pearson's correlation coefficient.

**Figure S8.** (*A-D*) Plots showing the normalized read count for mRNAs versus transcript length for exosomal RNA (A), fragmented whole cell RNA (B), unfragmented whole cell RNA (C) and an overlay of whole cell (fragmented) and exosomal RNA (D). (*E*) Total mRNA mapped reads binned by transcript length for exosome and whole cell RNA. (*F*) Ratio (exosomes/whole cell RNA) of total mRNA mapped reads for the indicated transcript length bins.

**Figure S9.** IGV screen shots showing read coverage across the *EEF1A1* and *RPS2* genes in untreated EVs compared to EVs treated with DNase or RNase in the absence or presence of detergent. Coverage plots show the read coverage along the gene body. The number of mapped reads for each condition is indicated to the left of the coverage plot. Read alignments show the mapped reads sorted by genomic coordinates. Nucleotides in reads matching nucleotides in the annotated reference are colored gray. Nucleotides in reads that do not match the annotated reference are color coded by nucleotide (A, green; C, blue; G, brown; and T, red) and correspond to the complementary nucleotide in the mRNA because of the gene orientations in IGV. Splice junctions are depicted in the alignments as thin blue bars between gray reads.

**Figure S10.** (*A*) Scatter plots comparing DESeq2-normalized read counts for different RNA biotypes in DNase-treated RNAs isolated from WT and YBX1-null exosomes. sncRNAs include pseudogenes. mRNA counts were generated by collecting reads that mapped to protein-coding genes in the initial mapping and then remapping to the human transcriptome reference sequence (Ensemble GRCh38 Release 76). (*B*) Scatter plots compare normalized read counts for mitochondrial (Mt) tRNAs in WT and YBX1-null whole cells and exosomes.

**Figure S11.** (*A,B*) Bar graphs showing normalized read counts mapping to different subspecies of Y RNA (A) and Vault RNA (B) for libraries prepared from DNase-treated WT and YBX1-null whole cell RNA (unfragmented) and exosomal RNA. The bar graphs show DESeq2-normalized read counts summed for all transcripts annotated for each indicated biotype in the GENCODE gene set.

**Figure S12.** Read-span analysis of sncRNAs in DNase-treated wild-type and YBX1-null exosomal RNA preparations. (*A-G*) Read spans from paired-end sequencing of tRNA (A), Y RNA (B), 7SL RNA (C), snoRNA (D), snRNA (E), 7SK RNA (F), and Vault RNA (G). The plots show the % of total reads of different read spans for the mapped paired-end reads for each biotype. Peaks for different subspecies of Y RNA, snoRNA, snRNA and Vault RNA are indicated in the plots.

**Figure S13.** Bar graphs showing the percentage of tRNA* with a stop at position 16 in the D-loop for each tRNA isoacceptor species (N=49) in DNase-treated wild-type and YBX1-null whole cell and exosomal RNAs. The percentages are the proportion of reads for each tRNA species that start between positions 15 and17.

**Figure S14.** IGV screen shots showing (*A*) read coverage across the tRNA coding sequence for valine isoacceptor TAC from TGIRT-seq of EV and unfragmented whole cell RNA using GsI-IIC RT or TeI4c RT and (*B*) read coverage across the genomic sequence for precursor tRNAs for LysCTT (Fig. 7*B*) and ValTAC (this Figure, panel *A*) from TGIRT-seq of unfragmented whole cell RNA using TeI4c RT. The genomic sequences in panel B consist of the tRNA coding sequence plus 15 bp of flanking genomic sequences retrieved from the UCCS genome browser for tRNA gene loci indicated below the alignments. Coverage plots are above the read alignments with reads sorted by top strand position. The numbers of reads are indicated to the left of the coverage plot and were down sampled to 1,000 reads in IGV for visualization. Nucleotides in reads matching nucleotides in the annotated reference are colored gray. Nucleotides in reads that do not match the annotated reference are color coded by nucleotide (A, green; C, blue; G, brown; and T, red). The 1-methyladenine ($m^1A$) and the 3' CCA post-transcriptional modifications, extra non-templated nucleotides added to the 3' end of the cDNA by the RT, and the D-loop truncation (U16) are indicated. Reads identified as tRNA precursors have short 5'- and 3'-end extensions that appear as additional unmapped sequences at both ends of the tRNA sequence when mapped to the tRNA coding sequence (Fig. 7*B*, and this Figure, panel *A*), but match the genomic sequence on either side of the tRNA.

**Figure S1**

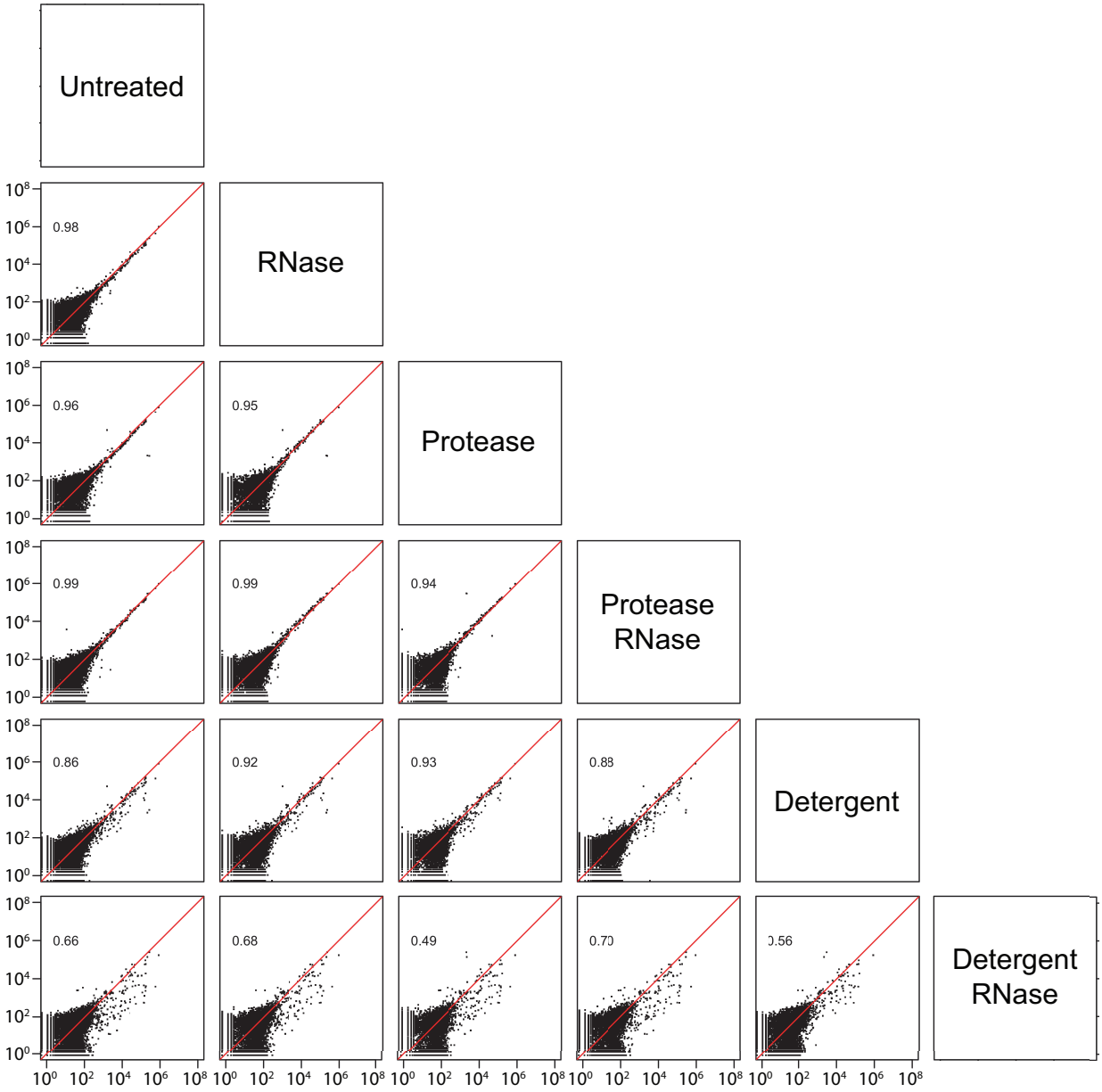
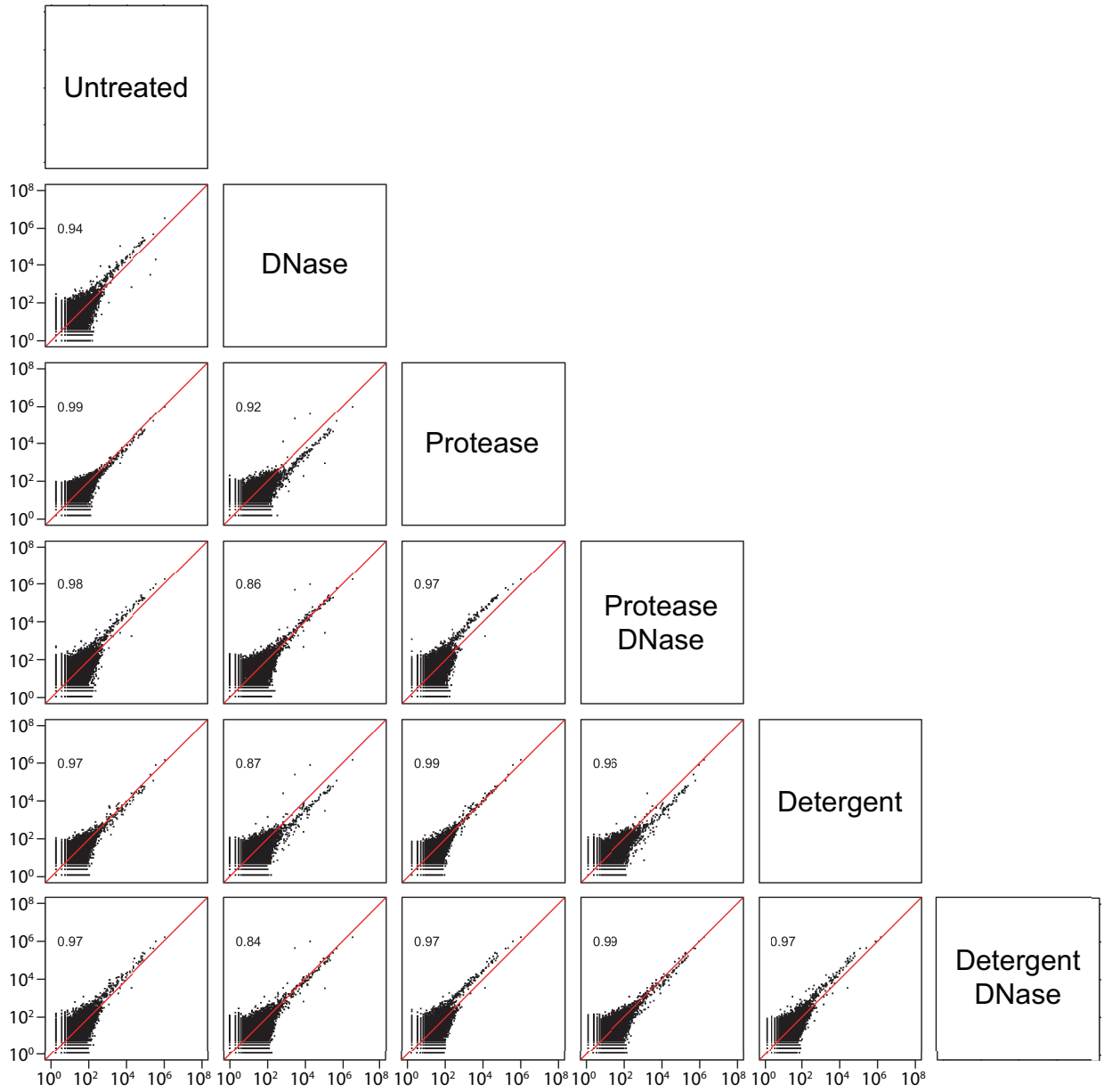
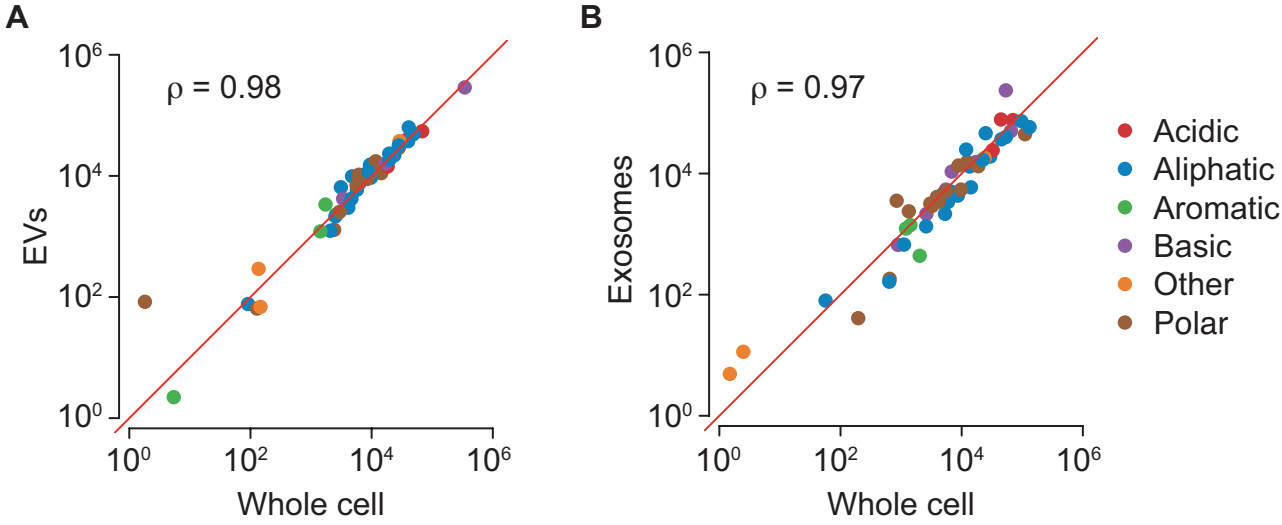

Untreated
RNase
Protease
Protease RNase
Detergent
Detergent RNase
0.98
0.96
0.95
0.99
0.99
0.94
0.86
0.92
0.93
0.88
0.66
0.68
0.49
0.70
0.56
$10^0$ $10^2$ $10^4$ $10^6$ $10^8$

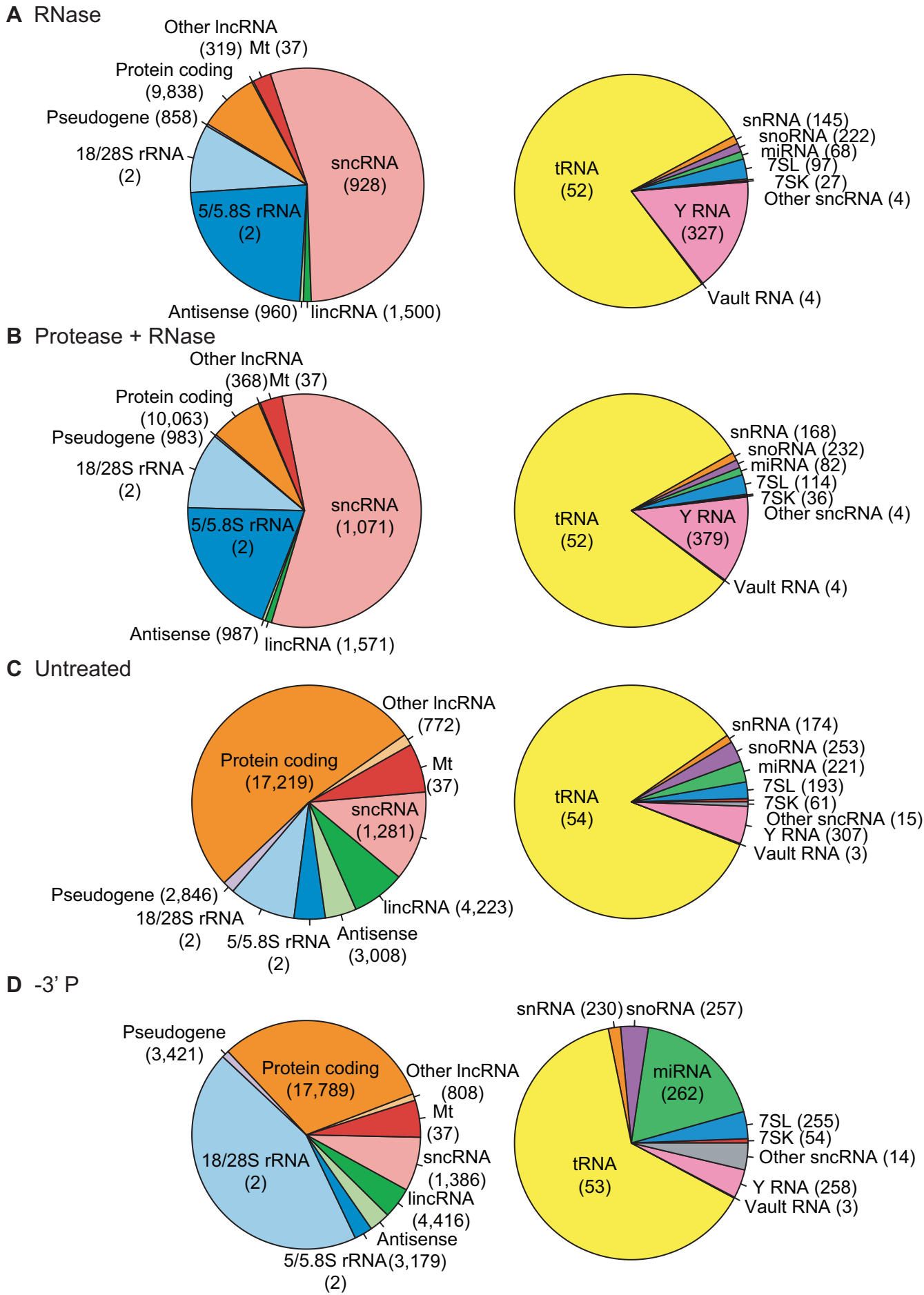
**Figure S2**


Untreated
DNase
Protease
Protease DNase
Detergent
Detergent DNase
0.94
0.99
0.92
0.98
0.86
0.97
0.97
0.87
0.99
0.95
0.97
0.84
0.97
0.99
0.97
$10^0$ $10^2$ $10^4$ $10^6$ $10^8$

# Figure S3


A
ρ = 0.98
EVs
Whole cell
B
ρ = 0.97
Exosomes
Whole cell
Acidic
Aliphatic
Aromatic
Basic
Other
Polar

# Figure S4


A RNase
Other lncRNA (319)
Mt (37)
Protein coding (9,838)
Pseudogene (858)
18/28S rRNA (2)
5/5.8S rRNA (2)
sncRNA (928)
Antisense (960)
lincRNA (1,500)
tRNA (52)
snRNA (145)
snoRNA (222)
miRNA (68)
7SL (97)
7SK (27)
Other sncRNA (4)
Y RNA (327)
Vault RNA (4)
B Protease + RNase
Other lncRNA (368)
Mt (37)
Protein coding (10,063)
Pseudogene (983)
18/28S rRNA (2)
5/5.8S rRNA (2)
sncRNA (1,071)
Antisense (987)
lincRNA (1,571)
tRNA (52)
snRNA (168)
snoRNA (232)
miRNA (82)
7SL (114)
7SK (36)
Other sncRNA (4)
Y RNA (379)
Vault RNA (4)
C Untreated
Other lncRNA (772)
Protein coding (17,219)
Mt (37)
sncRNA (1,281)
Pseudogene (2,846)
18/28S rRNA (2)
5/5.8S rRNA (2)
Antisense (3,008)
lincRNA (4,223)
tRNA (54)
snRNA (174)
snoRNA (253)
miRNA (221)
7SL (193)
7SK (61)
Other sncRNA (15)
Y RNA (307)
Vault RNA (3)
D -3' P
Pseudogene (3,421)
Protein coding (17,789)
Other lncRNA (808)
Mt (37)
18/28S rRNA (2)
sncRNA (1,386)
lincRNA (4,416)
Antisense (3,179)
5/5.8S rRNA (2)
snRNA (230)
snoRNA (257)
miRNA (262)
tRNA (53)
7SL (255)
7SK (54)
Other sncRNA (14)
Y RNA (258)
Vault RNA (3)

# Figure S5

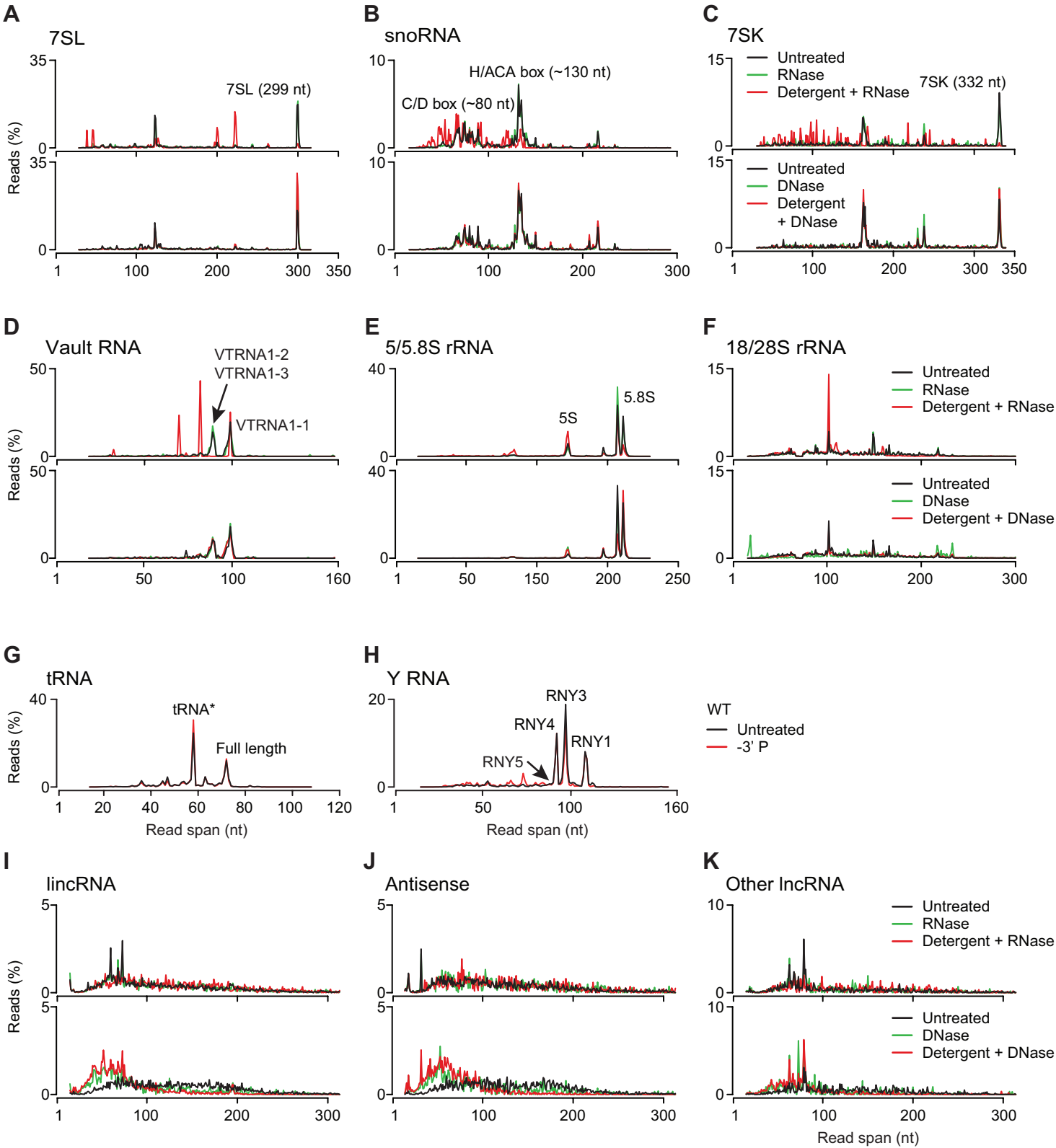
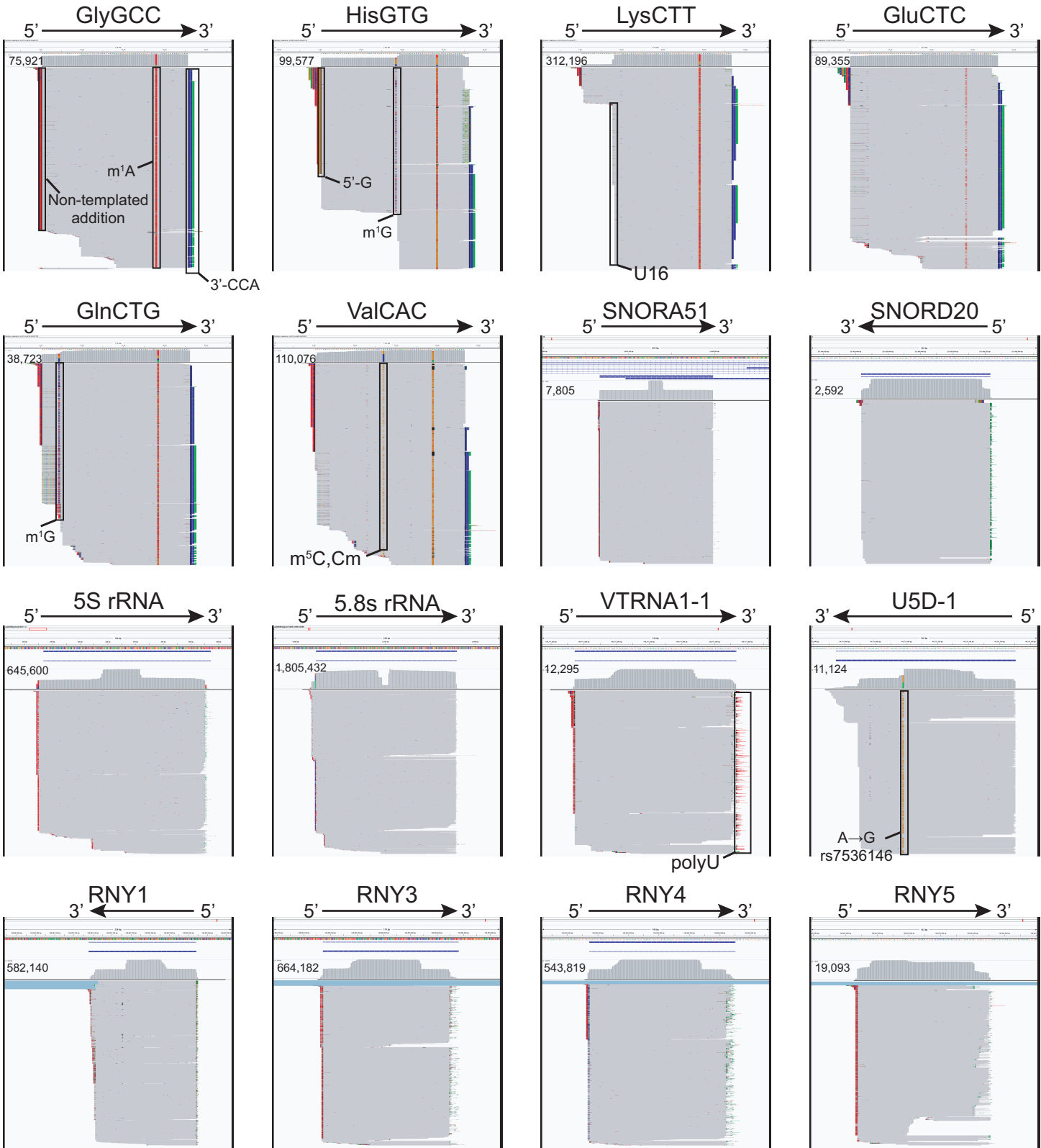
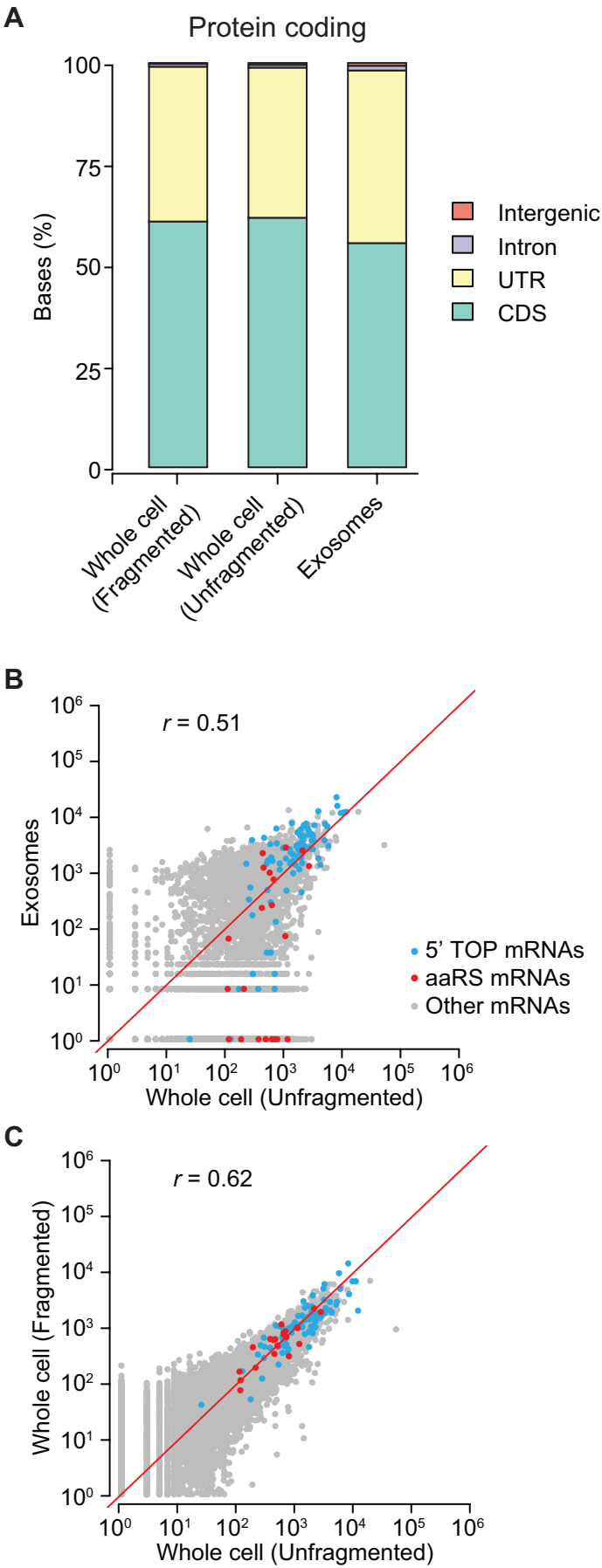

A
7SL
7SL (299 nt)
Reads (%)
B
snoRNA
H/ACA box (~130 nt)
C/D box (~80 nt)
C
7SK
Untreated
RNase
Detergent + RNase
7SK (332 nt)
Untreated
DNase
Detergent
+ DNase
D
Vault RNA
VTRNA1-2
VTRNA1-3
VTRNA1-1
E
5/5.8S rRNA
5S
5.8S
F
18/28S rRNA
Untreated
RNase
Detergent + RNase
Untreated
DNase
Detergent + DNase
G
tRNA
tRNA*
Full length
Read span (nt)
H
Y RNA
RNY3
RNY4
RNY1
RNY5
Read span (nt)
WT
Untreated
-3' P
I
lincRNA
J
Antisense
K
Other lncRNA
Untreated
RNase
Detergent + RNase
Untreated
DNase
Detergent + DNase
Read span (nt)

**Figure S6**


GlyGCC
5'
3'
75,921
m¹A
Non-templated addition
3'-CCA
HisGTG
99,577
5'-G
m¹G
LysCTT
312,196
U16
GluCTC
89,355
GlnCTG
38,723
m¹G
ValCAC
110,076
m⁵C,Cm
SNORA51
7,805
SNORD20
2,592
5S rRNA
645,600
5.8s rRNA
1,805,432
VTRNA1-1
12,295
polyU
U5D-1
11,124
A→G
rs7536146
RNY1
582,140
RNY3
664,182
RNY4
543,819
RNY5
19,093

# Figure S7


A
Protein coding
100
75
50
25
0
Bases (%)
Intergenic
Intron
UTR
CDS
Whole cell (Fragmented)
Whole cell (Unfragmented)
Exosomes
B
r = 0.51
Exosomes
Whole cell (Unfragmented)
5' TOP mRNAs
aaRS mRNAs
Other mRNAs
C
r = 0.62
Whole cell (Fragmented)
Whole cell (Unfragmented)

## Figure S8

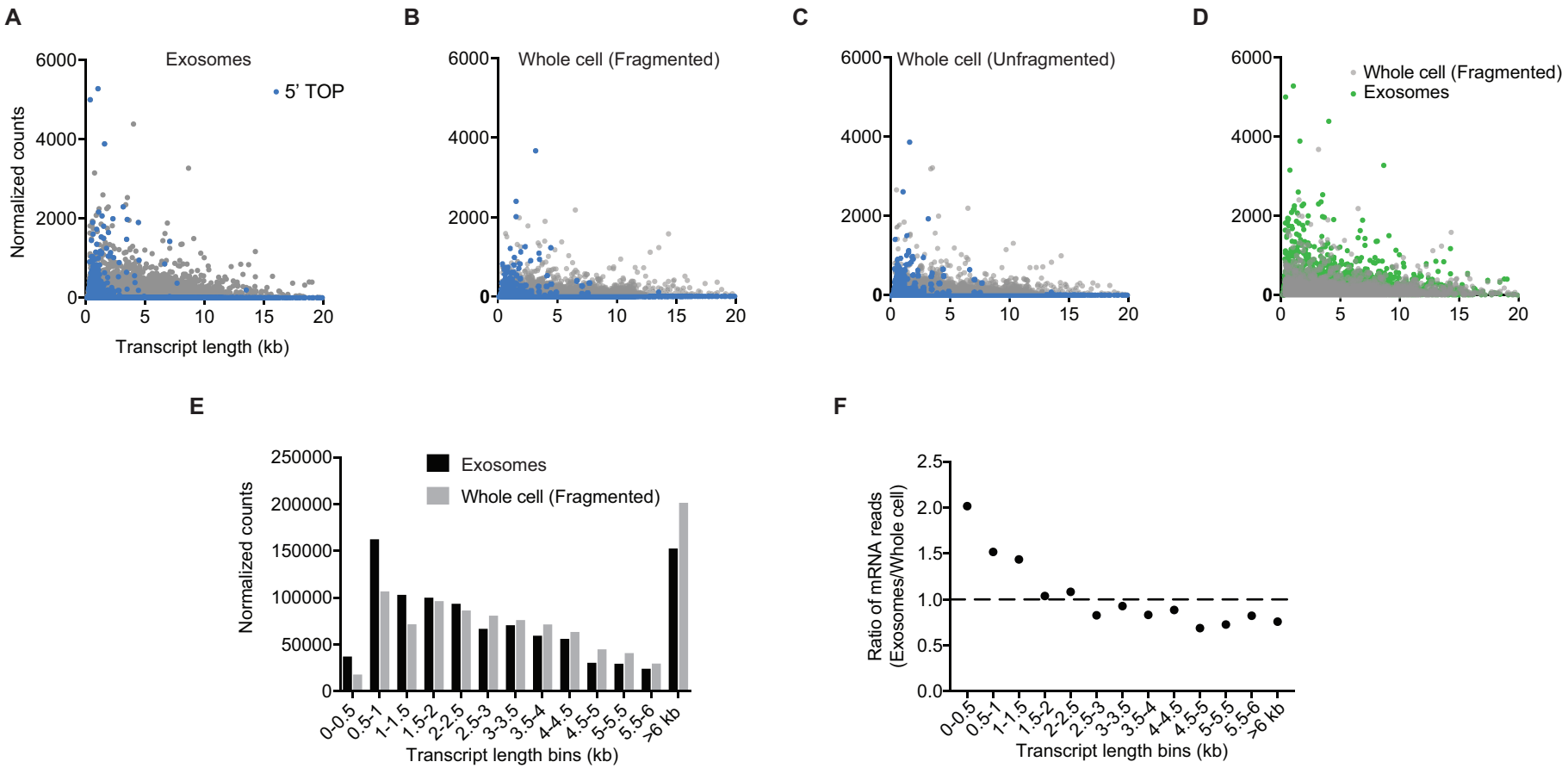
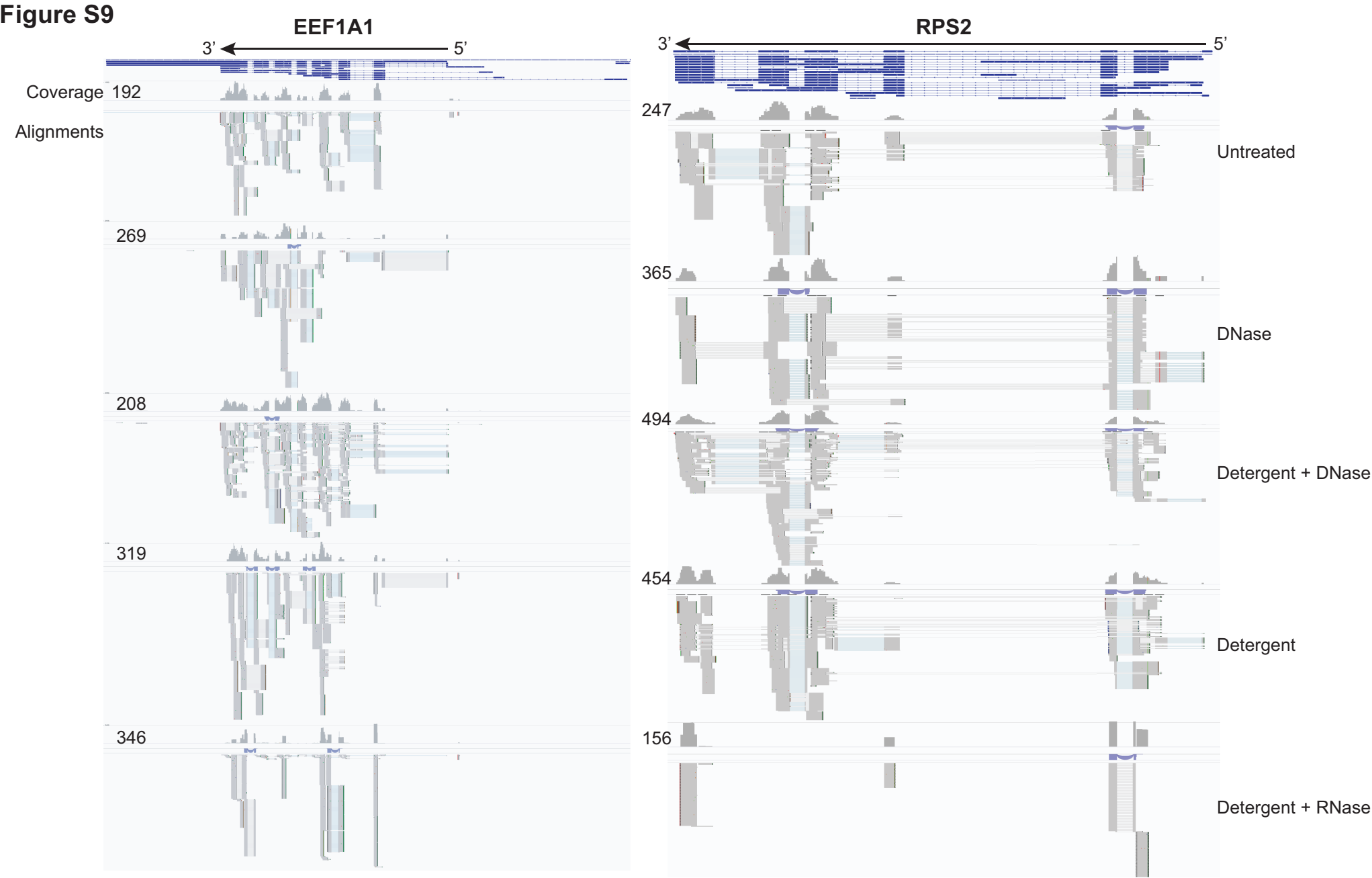
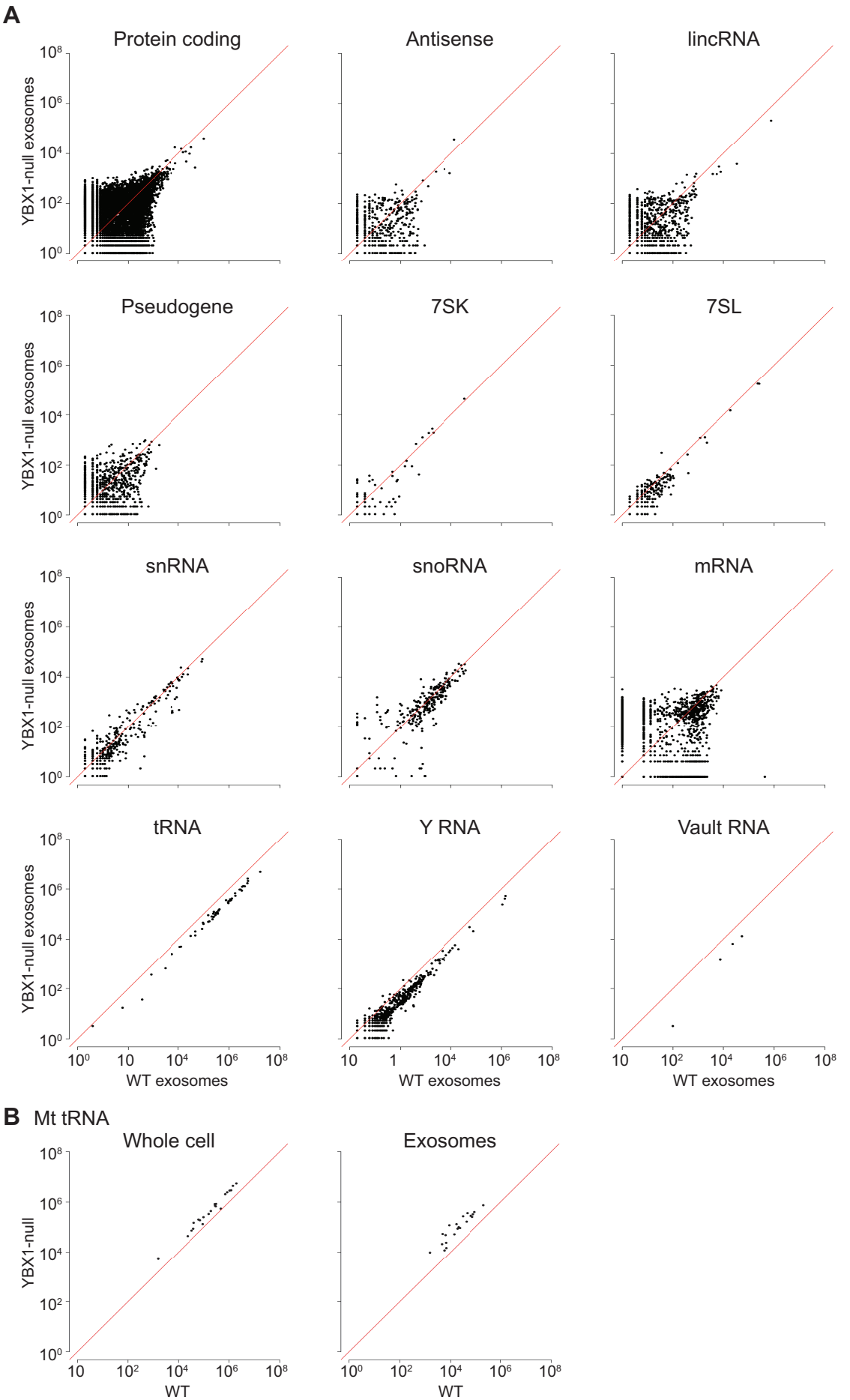

A
Exosomes
5' TOP
Normalized counts
6000
4000
2000
0
0
5
10
15
20
Transcript length (kb)
B
Whole cell (Fragmented)
C
Whole cell (Unfragmented)
D
Whole cell (Fragmented)
Exosomes
E
Exosomes
Whole cell (Fragmented)
Normalized counts
250000
200000
150000
100000
50000
0
0-0.5
0.5-1
1-1.5
1.5-2
2-2.5
2.5-3
3-3.5
3.5-4
4-4.5
4.5-5
5-5.5
5.5-6
>6 kb
Transcript length bins (kb)
F
Ratio of mRNA reads
(Exosomes/Whole cell)
2.5
2.0
1.5
1.0
0.5
0.0
Transcript length bins (kb)

Figure S9


EEF1A1
3'
5'
Coverage
Alignments
192
269
208
319
346
RPS2
3'
5'
247
365
494
454
156
Untreated
DNase
Detergent + DNase
Detergent
Detergent + RNase

# Figure S10


A
Protein coding
Antisense
lincRNA
Pseudogene
7SK
7SL
snRNA
snoRNA
mRNA
tRNA
Y RNA
Vault RNA
YBX1-null exosomes
WT exosomes
B Mt tRNA
Whole cell
Exosomes
YBX1-null
WT

## Figure S11

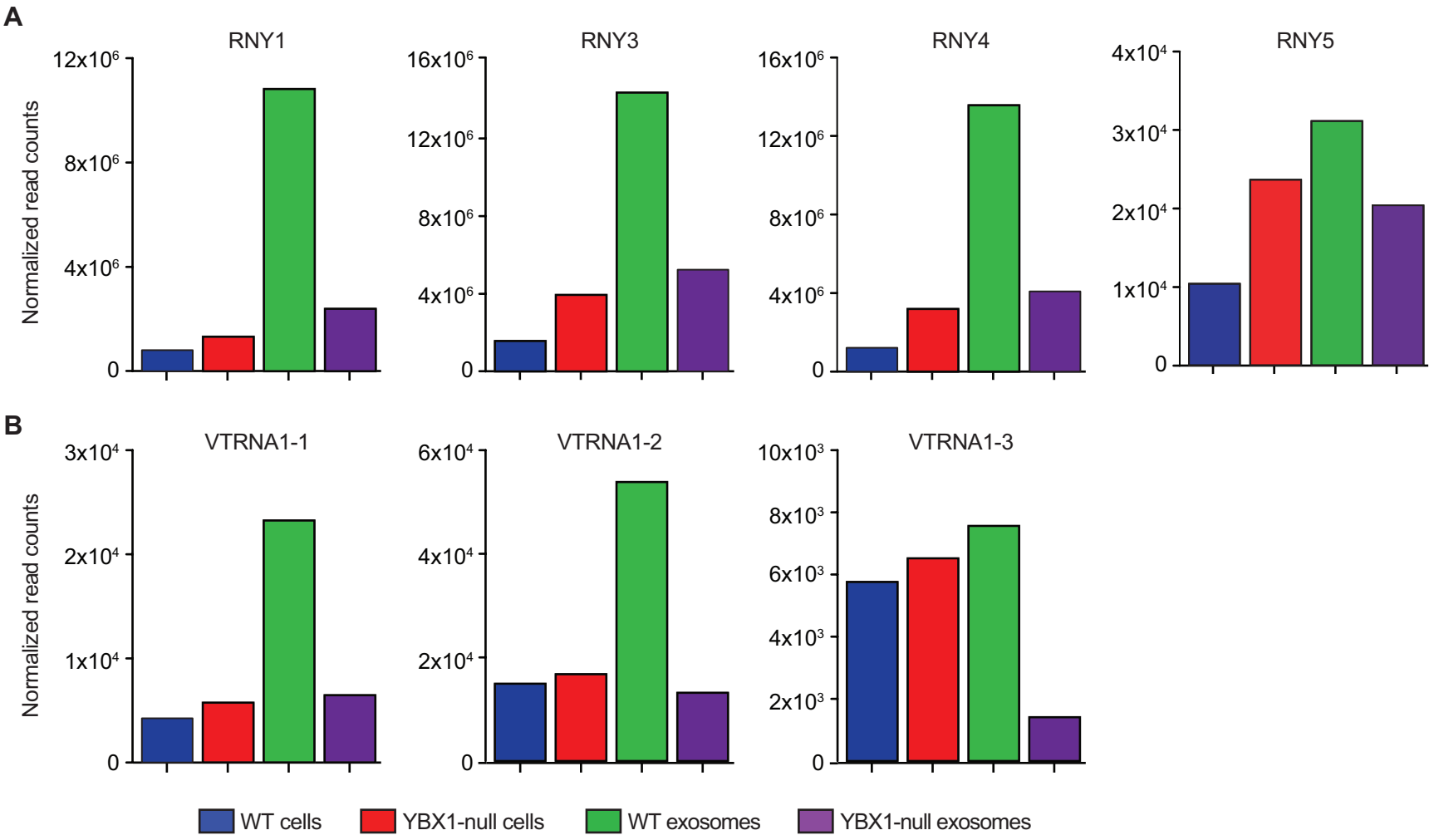

A
RNY1
RNY3
RNY4
RNY5
Normalized read counts
12x10^6
8x10^6
4x10^6
0
16x10^6
12x10^6
8x10^6
4x10^6
0
16x10^6
12x10^6
8x10^6
4x10^6
0
4x10^4
3x10^4
2x10^4
1x10^4
0
B
VTRNA1-1
VTRNA1-2
VTRNA1-3
Normalized read counts
3x10^4
2x10^4
1x10^4
0
6x10^4
4x10^4
2x10^4
0
10x10^3
8x10^3
6x10^3
4x10^3
2x10^3
0
WT cells
YBX1-null cells
WT exosomes
YBX1-null exosomes

**Figure S12**

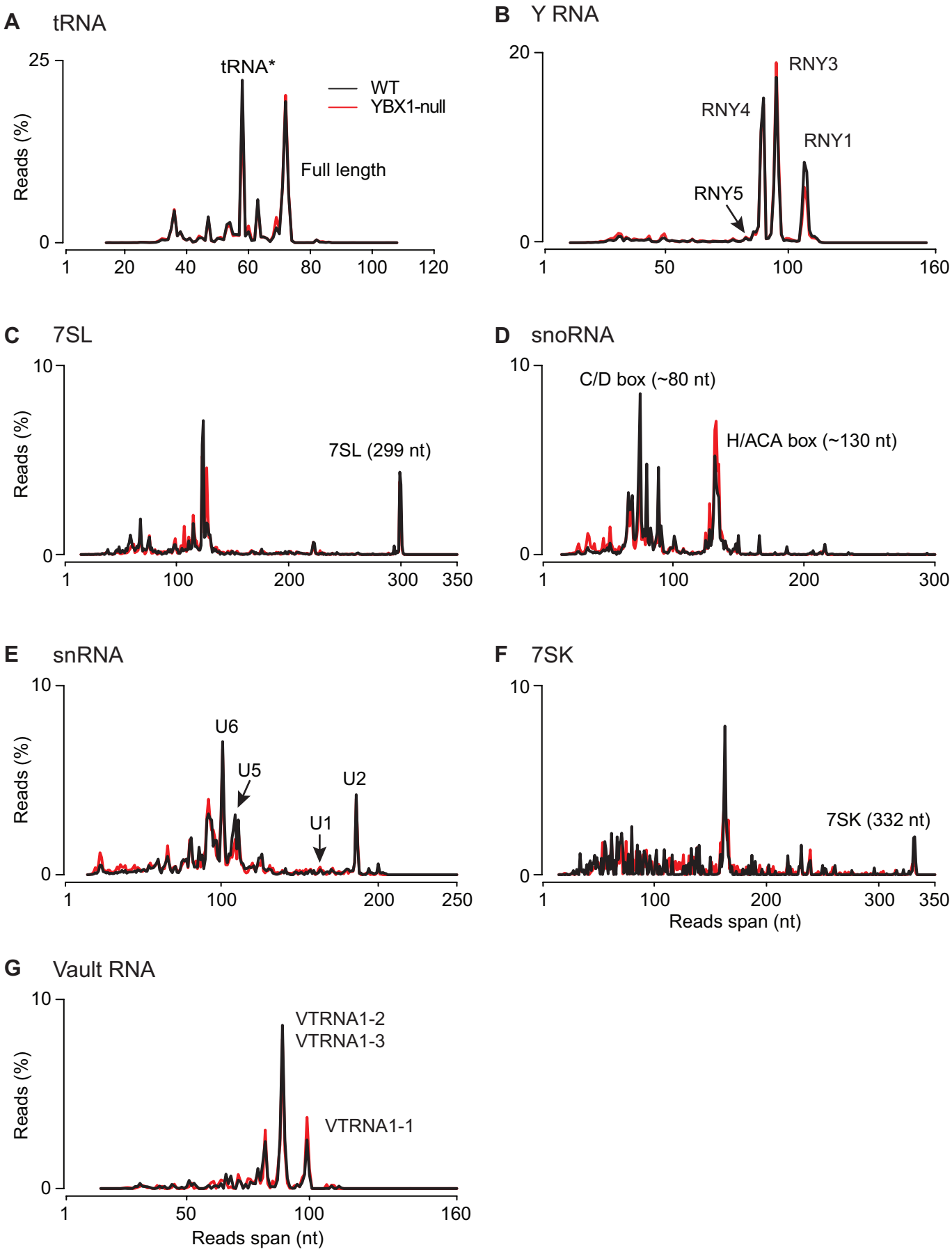

A tRNA
tRNA*
WT
YBX1-null
Full length
Reads (%)
25
0
1
20
40
60
80
100
120
B Y RNA
RNY3
RNY4
RNY1
RNY5
20
0
1
50
100
160
C 7SL
7SL (299 nt)
Reads (%)
10
0
1
100
200
300
350
D snoRNA
C/D box (~80 nt)
H/ACA box (~130 nt)
10
0
1
100
200
300
E snRNA
U6
U5
U2
U1
Reads (%)
10
0
1
100
200
250
F 7SK
7SK (332 nt)
10
0
1
100
200
300
350
Reads span (nt)
G Vault RNA
VTRNA1-2
VTRNA1-3
VTRNA1-1
Reads (%)
10
0
1
50
100
160
Reads span (nt)

**Figure S13**

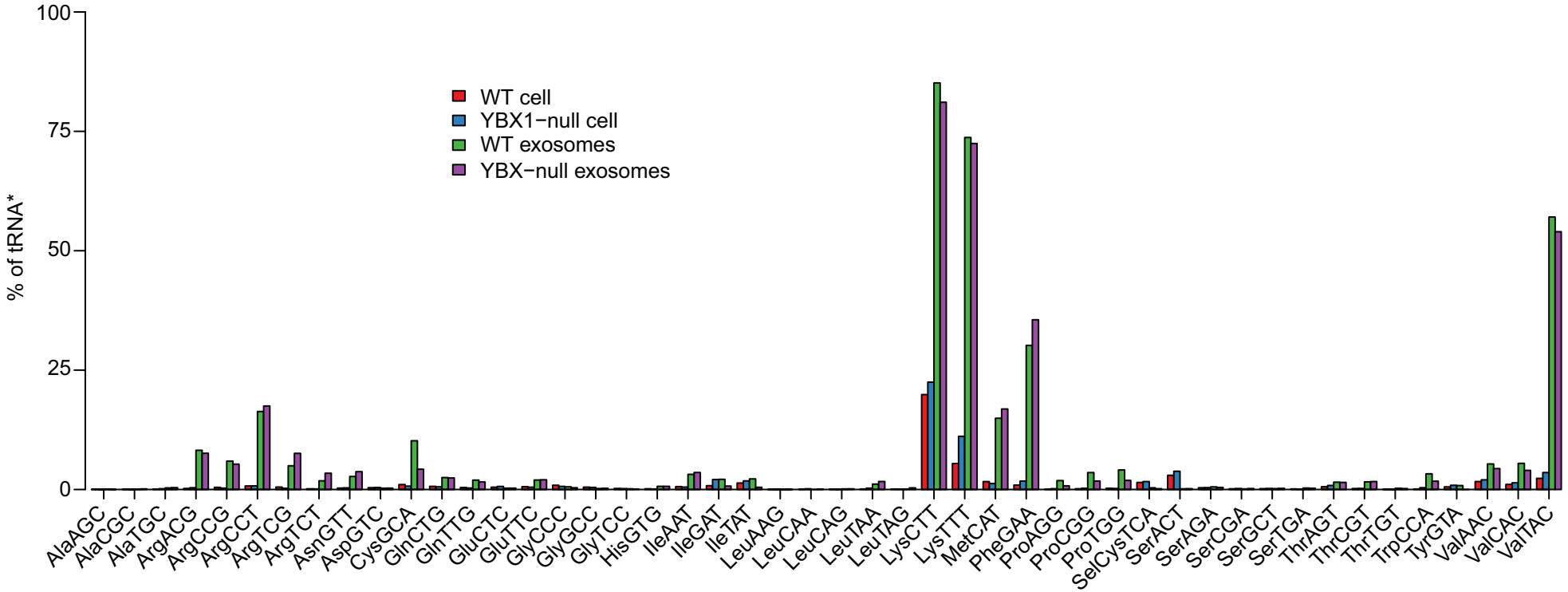
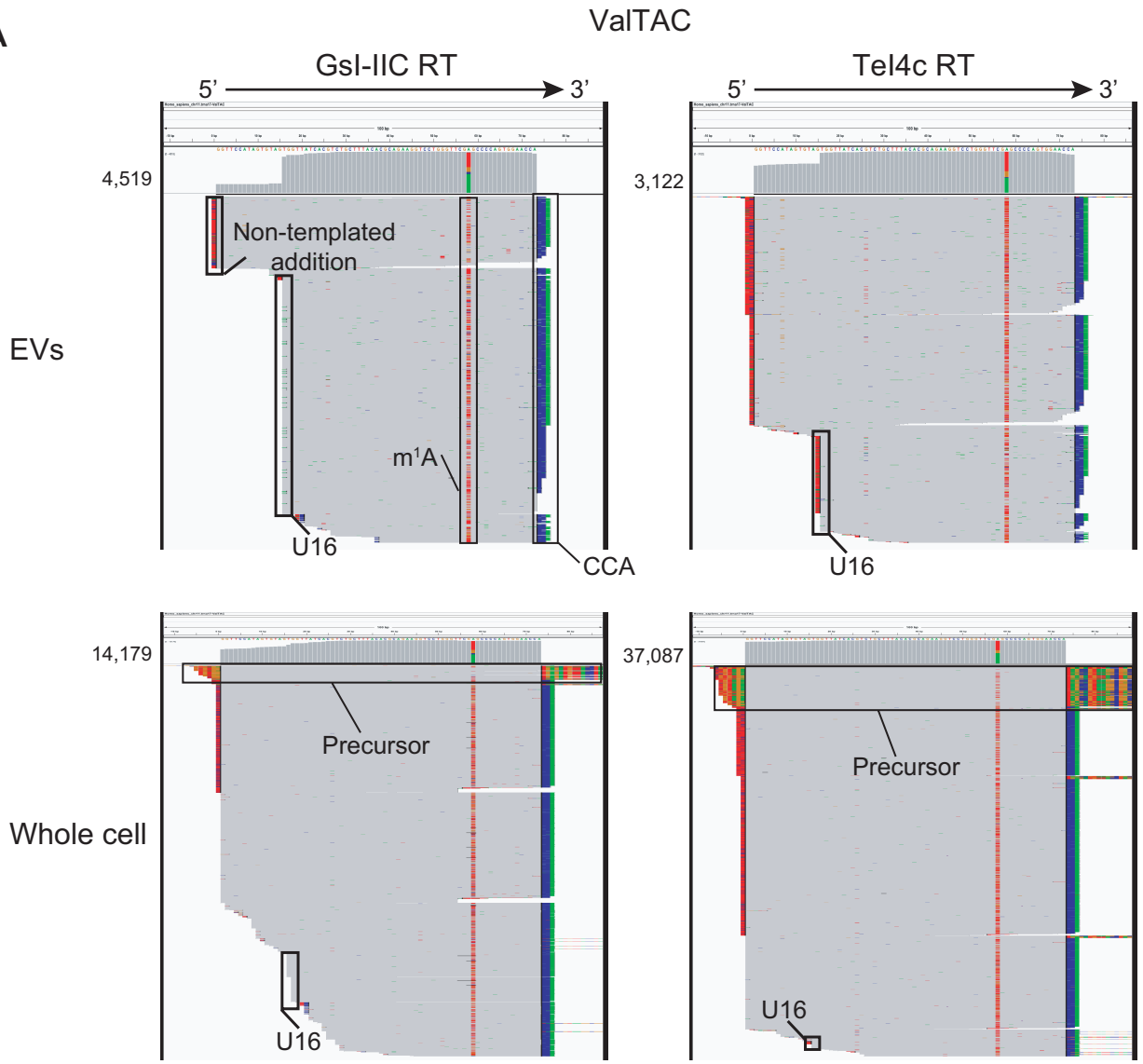
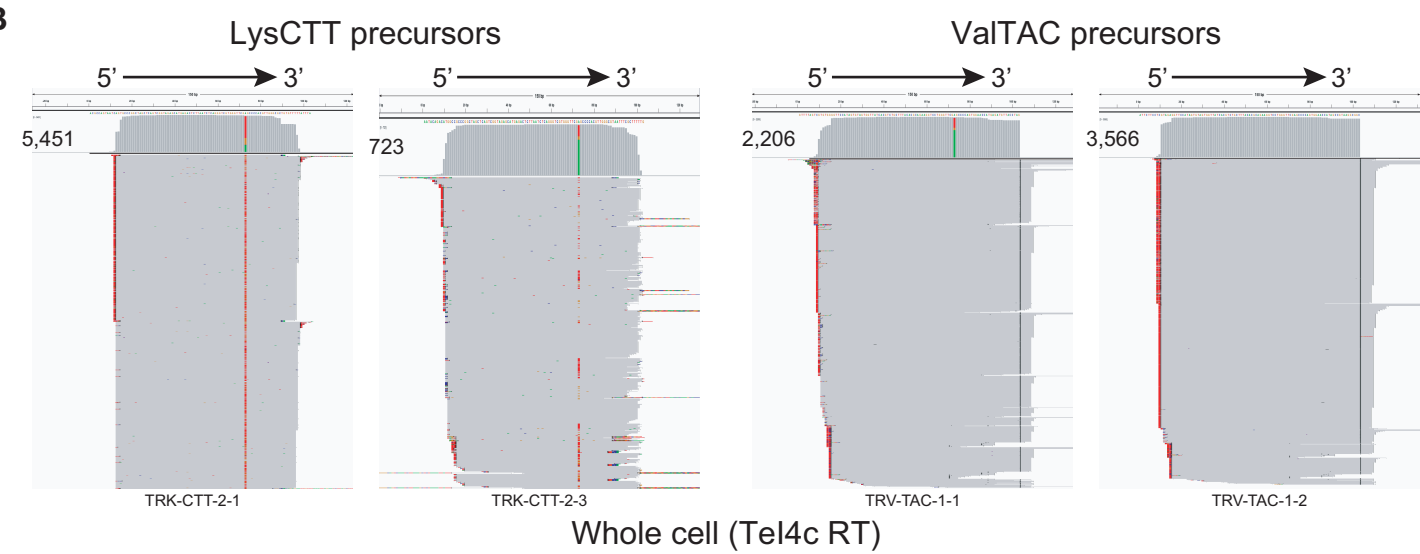

100
75
50
25
0
% of tRNA*
WT cell
YBX1−null cell
WT exosomes
YBX−null exosomes
AlaAGC
AlaCGC
AlaTGC
ArgACG
ArgCCG
ArgCCT
ArgTCG
ArgTCT
AsnGTT
AspGTC
CysGCA
GlnCTG
GlnTTG
GluCTC
GluTTC
GlyCCC
GlyGCC
GlyTCC
HisGTG
IleAAT
IleGAT
IleTAT
LeuAAG
LeuCAA
LeuCAG
LeuTAA
LeuTAG
LysCTT
LysTTT
MetCAT
PheGAA
ProAGG
ProCGG
ProTGG
SelCysTCA
SerACT
SerAGA
SerCGA
SerGCT
SerTGA
ThrAGT
ThrCGT
ThrTGT
TrpCCA
TyrGTA
ValAAC
ValCAC
ValTAC

**Figure S14**


A
ValTAC
GsI-IIC RT
Tel4c RT
5'
3'
4,519
3,122
Non-templated addition
m1A
U16
CCA
EVs
14,179
37,087
Precursor
Whole cell


B
LysCTT precursors
ValTAC precursors
5'
3'
5,451
723
2,206
3,566
TRK-CTT-2-1
TRK-CTT-2-3
TRV-TAC-1-1
TRV-TAC-1-2
Whole cell (Tel4c RT)