Title
Mechanisms of RNA sorting into exosomes

Author
Shurtleff, Matthew James

Publication Date
2016

Peer reviewed|Thesis/dissertation

Mechanisms of RNA sorting into exosomes

By

Matthew James Shurtleff

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Randy Schekman, Chair
Professor Arash Komeili
Professor James Hurley
Professor Lin He

Summer 2016

Abstract

Mechanisms of RNA sorting into exosomes

by

Matthew James Shurtleff

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Randy Schekman, Chair

Exosomes are vesicles that are released by cells into the extracellular environment and populate all bodily fluids. These vesicles contain molecular cargo, including RNA, proteins and lipids and therefore may serve as vehicles for intercellular communication by transferring unconventional signals between cells. Despite widespread scientific interest in the physiological role of exosomes in health and disease, little is currently known about how molecules are selectively sorted into exosomes.

In the work described herein, I used biochemical approaches to purify exosomes from cells grown in culture and identify microRNAs that are selectively sorted into exosomes. I then developed a cell-free reaction that reconstitutes the selective sorting of microRNA into exosomes *in vitro*. The reaction was then utilized to identify an RNA binding protein, Y-box Protein I (YBX1), that is required for sorting an exosomal microRNA. Next, I used the cell-free reaction as the basis for a selection strategy (termed Exo-SELEX) to identify primary RNA sequence motifs that act as positive and negative sorting signals *in vitro*. Finally, I used a high throughput RNA sequencing approach (TGIRT-seq) that allowed for a comprehensive transcriptome analysis (including highly structured or modified transcripts) of purified exosomes from normal and YBX1 knockout cells. TGIRT-seq analysis revealed that the most abundant transcript biotypes in exosomes are tRNA followed by other small non-coding RNA species. The abundant small non-coding RNA (tRNA, Y-RNA and Vault RNA) were strongly depleted in YBX1 knockout exosomes while sequences representing long non-coding RNA and protein coding genes were unaffected, indicating that these are sorted through a separate mechanism and allowing for broad classification of exosomal RNA based on YBX1-dependence or independence. In sum, this work provides a preliminary mechanistic understanding of the process of RNA sorting into exosomes and establishes multiple tools for the continued dissection of these pathways.

Table of Contents

About this dissertation

This dissertation describes work performed in the laboratory of Dr. Randy Schekman, Professor in the Department of Molecular and Cellular Biology from 2012-2016. The main text is organized in three chapters preceded by a brief introduction. Each chapter is consists of four sections (Introduction, Materials and Methods, Results and Discussion). All chapters describe primary research findings relating to how RNA molecules are specifically recognized and sorted into exosomes. Chapter 1 describes the purification of exosomes, sequencing of exosomal miRNAs, the establishment of a cell-free reaction that reconstitutes the selective sorting of exosomal microRNAs and the identification of Y-box protein I (YBX1) as a sorting factor. Chapter 2 relates the application of the cell-free packaging reaction to identify primary sequences that act as positive and negative exosomal RNA sorting motifs *in vitro*. Chapter 3 reports sequencing studies, performed in collaboration with Alan Lambowitz and colleagues at the University of Texas, that classify transcript biotypes beyond microRNAs that are secreted via exosomes in a YBX1-dependent manner.

Acknowledgements

Thank you to Mary, this is for you.

Thank you to Kate for never forgetting what is most important.

Thank you to my committee members, Jim Hurley, Arash Komeili and Lin He. Your insights, advice and support have been invaluable.

Thank you to all members of the Schekman lab during the time I spent here. I greatly benefitted from drawing on the expertise of the entire lab. Especially, thank you to Jason Lam who tolerated an independent thinking rotation student and to Morayma Temoche-Diaz who, as a rotation student, tolerated an independent thinking rotation supervisor.

Thank you to my classmates and friends for drinking beer with me to celebrate the good times and for drinking beer with me to forget the bad times.

Thank you to Nick Ingolia who provided the inspiration for Chapter 2 and to Kathy Collins who facilitated the collaboration that led to Chapter 3.

Thank you to the staff at the UC Berkeley Tissue Culture facility, Vincent J. Coates Sequencing facility, Vincent J. Coates Proteomics Facility and the Computational Genomics Resources Laboratory.

Finally, thank you to Randy Schekman. You taught me to work hard in the lab but to also stop and think. You taught me to think about cellular questions from all perspectives, from single molecules to entire organisms. You taught me to communicate my work in simple terms. You taught me to not just do the simplest experiments but to seek experiments that will tell you something new. In sum, you taught me how to be a scientist, not just a researcher.

Introduction: Secretion of RNA via exosomes

In contrast to the conventional pathways of protein secretion, the processes by which unconventional cargoes are secreted have proved diverse and enigmatic. Indeed, our understanding of unconventional secretory mechanisms is limited to a few examples of leader-less soluble and transmembrane proteins [1]. Unconventionally secreted molecules may be externalized in a soluble form by translocation across various membranes. This may include direct translocation across the plasma membrane, or across an organelle membrane followed by fusion of the organelle with the plasma membrane [2]. Alternatively, proteins and RNAs can be secreted within vesicles that bud from the plasma membrane, as in the budding of enveloped viruses such as HIV, or within vesicles internalized into a multivesicular body (MVB) that fuses with the plasma membrane [3].

RNA is actively secreted into the medium of cultured cells and can be found in all bodily fluids enclosed within vesicles or bound up in ribonucleoprotin complexes, both forms of which are resistant to exogenous ribonuclease [3-5]. Importantly, extracellular vesicle-bound RNAs appear to be enriched in specific classes of RNAs, including small RNAs and microRNAs (miRNAs) [6-8].

Exosomes are a subclass of extracellular vesicle which can be defined as 30-100 nm vesicles with a buoyant density of ~1.10-1.19 g/ml that are enriched in specific biochemical markers, including tetraspanin proteins [3]. It is often assumed that vesicles fitting this description are derived from the multivesicular body, but some evidence suggests that physically and biochemically indistinguishable vesicles bud directly from the plasma membrane [9]. Numerous studies have reported the presence of RNAs, especially microRNAs, from fractions containing exosomes, though many of these studies have relied on isolation techniques (e.g. high speed sedimentation) that do not resolve vesicles from other cellular debris or RNPs [10]. Thus, it is difficult to know in which form RNAs are secreted and even more challenging to determine which miRNAs may be specifically secreted as exosome cargo. The use of many different cell lines, bodily fluids and isolation methods to identify which miRNAs are specifically packaged into exosomes further complicates the establishment of widely accepted exosomal miRNA cargo.

Even with the crude preparations that have been characterized, it is clear that RNA profiles from exosomes are distinct from those of the producer cells. Thus RNA capture or stabilization in exosomes is likely to occur through a selective sorting mechanism. RNA packaging may occur by specific interactions with RNA binding proteins that engage the machinery necessary for membrane invagination into the interior of an MVB or by interaction of RNAs with lipid raft microdomains from which exosomes may be derived [11]. The major players involved in selectively sorting RNA into exosomes are yet to be discovered.

Chapter 1: Cell-free packaging of microRNA into exosomes reveals Y-box Protein I as a critical sorting factor

Introduction

Previous reports have indicated that some miRNA are secreted in the form of stabilized RNPs and inside of extracellular vesicles termed exosomes [3-5]. However, a conclusive analysis of which miRNAs are selectively sorted into exosomes is hampered by the lack of rigorous methods to purify vesicles from RNPs. The development and application of biochemical assays to monitor vesicle budding and membrane fusion have been crucial to elucidate the cellular mechanisms for the conventional secretory pathway, however no analogous cell-free assays yet exist for exosome biology.

In order to probe the mechanisms of exosome biogenesis, I developed procedures to refine the analysis of RNA sorting into exosomes. Using traditional means of membrane fractionation and immuno-isolation, I identified unique miRNAs highly enriched in exosomes marked by their content of CD63. This miRNA sorting process was then reproduced with a cell-free reaction reconstituted to measure the packaging of exosome-specific miRNAs into vesicles formed in incubations containing crude membrane and cytosol fractions. Among the requirements for miRNA sorting *in vitro*, I found one RNA-binding protein, YBX1, which is a known constituent of exosomes secreted from intact cells [12, 13].

Materials and Methods

Cell lines, media and general chemicals

HEK293T cells were cultured in DMEM with 10% FBS (Thermo Fisher Scientific, Waltham, MA). For exosome production, cells were seeded to ~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). I noted that confluency >80% decreased the yield of exosome RNA. Cells grown for exosome production were incubated in exosome-free medium produced by ultracentrifugation at 100,000Xg (28,000 RPM) for 18h using an SW-28 rotor (Beckman Coulter, Brea, CA) in a LE-80 ultracentrifuge (Beckman Coulter). Unless otherwise noted, all chemicals were purchased from Sigma Aldrich (St. Louis, MO).

Exosome purification

Conditioned medium (3 l for small RNA-seq and 420 ml for all other experiments) was harvested from 80% confluent HEK293T cultured cells. All subsequent manipulations were performed at 4ºC. 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 maximum rotor capacity was 210 ml, thus the small RNA-seq processing required pooling from ~15 independent centrifugations. 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 4ºC. 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 mixed with the use of a vortex to mix the sample evenly. The sucrose concentration in the PBS/sucrose mixture was measured by refractometry and, if necessary, additional 60% sucrose buffer as added until the concentration was >50%. Aliquots (1 ml) of 40%, 20% and 0% sucrose buffer were sequentially overlaid and the tubes were centrifuged at ~150,000Xg (38,500 RPM) for 16 h in an SW-55 rotor. The 20/40% interface was harvested, diluted 1:5 with phosphate buffered saline (pH 7.4) and 1 µg of rabbit polyclonal anti-CD63 H-193 (Santa Cruz Biotechnology, Dallas, TX) was added per liter of original conditioned medium and mixed by rotation for 2 h at 4C. 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 4C. Beads with bound exosomes were washed three times in 1 ml PBS and RNA was extracted using Direct-Zol RNA mini-prep (Zymo Research, Irvine, CA) or protein was extracted in 100 µl 1X Laemmli sample buffer and dispersed with the use of a vortex mixer for 2 min.

Negative staining and visualization of exosomes by electron microscopy
An aliquot (4 ul) of the resuspended 100,000Xg pellet fraction or a sample from the 20/40% interface that was diluted 10-fold with PBS, centrifuged at 100,000Xg in a TLS-55 rotor and then resuspended in 1 % glutaraldehyde, was spread onto glow discharged Formvar-coated copper mesh grids (Electron Microscopy Sciences, Hatfield, PA) and stained with 2% Uranyl acetate for 2 min. Excess staining solution was blotted off with filter paper. Post drying, grids were imaged at at 120 kV using a Tecnai 12 Transmission Electron Microscope (FEI, Hillsboro, OR) housed in the Electron Microscopy Laboratory at UC Berkeley.

Nanoparticle Tracking Analysis
Conditioned medium (1 ml) from wild type and ΔYBX1 cells was harvested and the supernatant from a 10,000Xg centrifugation was drawn into a 1 ml syringe and inserted into a Nanosight LM10 instrument (Malvern, UK). Particles were tracked for 60 s using Nanosight nanoparticle tracking analysis software. Each sample was analyzed 4 times and the counts were averaged.

Construction of inducible 293:CD63-luciferase cell line and luciferase activity assays
HEK293 cells expressing doxycycline-inducible CD63-luciferase was generated using the T-REx - 293 cell line according to the manufacturer's instructions (Life Technologies, Grand Island, NY). The open reading frame for CD63-was amplified from human cell cDNA and firefly luciferase-FLAG was amplified from a plasmid source, both using Phusion DNA Polymerase (NEB). CD63 was fused to luciferase by NotI digestion, ligation and PCR amplification. The CD63-luciferase-FLAG amplicon was then digested and ligated into pcDNA5/FRT/TO (Life Technologies) using NdeI and PstI sites. The resulting plasmid was co-transfected with pOG44 (Life Technologies) and a stable cell line was selected using hygromycin selection (100 µg/ml). CD63-luciferase expression was induced with 1 µg/ml doxycycline 48 h prior to exosome harvesting. Luciferase

activity was measured using a Promega Glowmax 20/20 luminometer (Promega, Madison, WI) with a signal collection integration time of 1 s. Luciferase reactions contained 50 µl sample, 10 µl 20X luciferase reaction buffer (500 mM Tricine, pH 7.8, 100 mM MgSO4, 2 mM EDTA), 10 µl 10 mM D-luciferin dissolved in PBS, 10 µl ATP dissolved in deionized water and 120 µl deionized water. Where indicated, samples were pre-treated with final concentrations of 1% Triton X-100 and/or 100 µg/ml trypsin for 30 min on ice. Total protein concentrations were measured using Pierce BCA protein assay according to the manufacturer's instructions.

Immunoblotting

Exosome and cell lysates were prepared by mixing in lysis buffer (10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 0.1% sodium dodecyl sulfate, 0.5% sodium deoxycholate, 1% Triton X-100, 10% glycerol). Lysates were diluted 4-fold with 4X Laemmli sample buffer, heated to 65ºC for 5 min and separated on 4-20% acrylamide Tris-Glycine gradient gels (Life Technologies). Proteins were transferred to polyvinylidene difluoride membranes (EMD Millipore, Darmstadt, Germany), blocked with 5% bovine serum albumin in TBST and incubated overnight with primary antibodies. Blots were then washed with TBST, incubated with anti-Rabbit or anti-Mouse secondary antibodies (GE Healthcare Life Sciences, Pittsbugh, PA) and detected with ECL-2 reagent (Thermo Fisher Scientific). Primary antibodies used in this study were anti-YBX1 (Cell Signaling Technology, Danvers, MA), anti-GAPDH (Santa Cruz Biotechnology), anti-TSG101 (Genetex, Irvine, CA), anti-CD9 (Santa Cruz Biotechnology), anti-Flotillin 2 (Abcam, Cambridge, MA), anti-Alix (Santa Cruz Biotechnology) and anti-Ago2 (Cell Signaling Technology). For quantitative immunoblotting (in Fig. 6), the same procedures were used, but were instead imaged using the LiCOR Odyssey imaging system.

Quantitative real-time PCR

RNA was extracted using the Direct-Zol RNA mini-prep and cDNA was synthesized either by oligo-dT priming (mRNA) or gene-specific priming (miRNA) according to the manufacturer's instructions. For miRNA, I used Taqman miRNA assays from Life Technologies (assay numbers: hsa-mir-223-3p: 000526, hsa-mir-190a-5p: 000489 and hsa-miR-144-3p: 002676). Because there is no well-accepted endogenous control transcript for exosomes, relative quantification was performed from equal amounts of total RNA. Qubit (Thermo Fisher Scientific), was used to quantify total RNA from the medium or cells: 10 ng of RNA was reverse transcribed and qPCR was performed according to manufacturer's instructions. Relative quantification was calculated from the expression $2^{-(Ct_{control}-Ct_{experimental})}$. Taqman qPCR master mix with no amperase UNG was obtained from Life Technologies and quantitative real-time PCR was performed using an ABI-7900 real-time PCR system (Life Technologies).

Cell-free biochemical assays

Preparing membranes and cytosol

HEK293T cells were harvested at ~80% confluency by gently pipetting with PBS. Cells were centrifuged at 500Xg and the pellet was weighed and frozen at -80C until use. Cells were thawed, resuspended in 2 vol of homogenization buffer (250 mM sorbitol, Tris-HCl,

pH 7.4, 137 mM NaCl) containing protease inhibitor cocktail (1mM 4-aminobenzamidine dihydrochloride, 1 μg/ml antipain dihydrochloride, 1 μg/ml aprotinin, 1 μg/ml leupeptin, 1 μg/ml chymostatin, 1 mM phenymethylsulfonly fluoride, 50 μM N-tosyl-L-phenylalanine chloromethyl ketone and 1 μg/ml pepstatin) and passed 7-15 times through a 22 gauge needle until >80% of cells were disrupted, as assessed by microscopy and trypan blue staining. The homogenized cells were then centrifuged at 1,500Xg and the post-nuclear supernatant fraction (PNS) was centrifuged at 15,000Xg using a FA-45-30-11 rotor and Eppendorf 5430 R centrifuge (Eppendorf, Hamburg, Germany). The supernatant fraction was centrifuged again at 55,000 RPM in a TLS-55 rotor and Optima Max XP ultracentrifuge (Beckman Coulter) to generate the cytosol fraction (~5 mg/ml). The 15,000Xg pellet fraction was resuspended in 2 packed cell vol homogenization buffer and an equal vol of 1 M LiCl. The membranes were then centrifuged again at 15,000Xg and resuspended in 1 original packed cell vol to generate the membrane fraction.

Cell-free exosome biogenesis assay

Membranes were prepared from HEK293:CD63-luciferase cells and cytosol from HEK293T cells. Complete biogenesis reactions (40 μl) consisted of 10 μl membranes, 17 μl cytosol (2 mg/ml final concentration) + homogenization buffer, 4 μl 10X ATP regeneration system (10mM ATP, 500 mM GDP-mannose, 400 mM creatine phosphate, 2mg/ml creatine phosphokinase, 20 mM HEPES, pH 7.2, 250 mM sorbitol, 150 mM KOAc, 5 mM MgOAc), 8 ul 5X incorporation buffer (80 mM KCl, 20 mM CaCl2, 12.5 mM HEPES-NaOH, pH 7.4, 1.5 mM, MgOAc, 1 mM DTT), 1 μl D-luciferin (10 mM in PBS). The reaction mixture was incubated at 30ºC for 20 min and membranes were sedimented at 15,000Xg at 4ºC. Post-reaction membranes were then resuspended in 500 ul PBS with 0.5 mg/ml trypsin and incubated at 4ºC for 1h to inactivate CD63-luciferase that had not been internalized during the incubation period. The trypsin-treated reactions were then incubated for 2 min at 25ºC and luciferase activity was quantified using the luminometer conditions described above. Exosome biogenesis for experimental conditions was calculated as the relative ratio compared to the complete control reaction described above ($RLU_{experimental}/RLU_{control}$).

Cell-free exosome miRNA packaging assay

RNA oligonucleotides corresponding to the targeting or passenger strand sequences of miR-223-3p or the targeting strand of miR-190a-5p were purchased from Integrated DNA Technologies (IDT, Coralville, IA). Duplex substrates were generated by incubating mature and targeting strands dissolved in annealing buffer (10 mM Tris, pH 7.5, 20 mM NaCl in RNase-free water) at 95°C for 2 min in a heat block and then removing the block from heat and allowing the samples to cool to room temperature over the course of 1-2 h on the bench top[14].

Membranes and cytosol were prepared from HEK293T cells. Complete miRNA packaging assays (40 μl) contained 10 μl membranes, ~16 μl cytosol + homogenization bufffer (2 mg/ml final concentration), 4 μl 10X ATP regenerating system, 8 μl incubation buffer, 1 μl 10 μM synthetic single stranded or duplex miRNA and 1 μl RNAsin (Promega). Reactions were incubated for 20 min at 30ºC then placed on ice and mixed with 4.3 μl of 10X NEB buffer 3 and 1 μl of RNase I_f (50,000 units/ml) NEB, Ipswich,

MA) was added to all reactions except to a no RNase control. Reactions were then incubated at 30ºC for a further 20 min. Following incubation, RNA was immediately extracted according to Direct-Zol (Zymo Research) manufacturer's instructions. First strand complementary DNA synthesis and quantitative PCR was performed using TaqMan miRNA assays (Life Technologies) for hsa-miR-223 or hsa-miR-190. Percent protection was calculated from the qPCR data by comparing the Ct of miRNA in the RNase treated samples against the no RNase control reaction ($2^{-(Ct_{experimental}-Ct_{control})}$) in which the no RNase control was set to 100 percent.

Streptavidin pull-down of miR-223 and interacting proteins

The *in vitro* packaging assay was performed as described above with miR-223-3p with biotin linked either to the 3' phosphate or internally biotinylated at position 13 (IDT). Samples were heated to 65ºC for 20 min to inactivate RNase I_f and then mixed with 4.4 ul 10% Triton X-100 for a final concentration of 1% and kept on ice for 30 min. Novagen MagPrep Streptavidin-coated beads (10 µl/reaction) (EMD Millipore) were washed 3 times with 1 ml PBS and then added to the reaction lysate. The suspension was mixed by rotation for 2 h at 4ºC, the beads were immobilized using a magnet and washed 3 times with 1 ml PBS. Proteins were eluted from bead-bound miR-223 with 50 µl 1 M KCl. In-solution liquid chromatography and mass spectrometry were performed according to standard procedures by the Vincent J. Coates Proteomics/Mass Spectroscopy laboratory (UC Berkeley).

Small RNA sequencing of cellular and exosomal RNA

RNA was prepared from cells and 3 l of HEK293T conditioned media. Sequencing libraries were generated using the Scriptminer Small RNA sequencing kit (Epicentre Biotechnologies, Madison, WI) from 1 µg total RNA from cells and 200 ng total RNA from exosomes according to the manufacturer's protocol. The libraries were amplified and index barcodes were added by 11 cycles of PCR. Libraries were sequenced by 50 bp single read massively parallel sequencing on an Illumina Hi-Seq 2000 System at the Vincent J. Coates Genomic Sequencing Laboratory (UC Berkeley).

Small RNA sequencing analysis

Preprocessing of the 50 base pair single reads was filtered for read quality (read quality >20 and percent bases in sequence that must have quality >90) and adaptor sequences were clipped using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) implementation on the GALAXY platform (usegalaxy.org) ([15-17]. Sequences were mapped to miRbase using miRdeep2 and counts tables were obtained using the quantifier program using default settings[18]. Reads were normalized by dividing the number of reads mapping to each miRNA by the number of total reads mapping to all miRNAs and the quotient was then multiplied by one million (reads per million miRNA mapped reads - RPM). To analyze miRNA species, I used the quantifier program of the miRdeep2 software suite [18]. Precursor reads were determined by subtracting the number of reads mapping to mature (either targeting or passenger strand sequences) from the total number of reads mapping to the full-length precursor transcript for each miRNA. Those miRNAs with described passenger strands (star strand) were then analyzed to determine how many mature reads mapped to either targeting or passenger strands.

CRISPR/Cas9 genome editing

A pX330-based plasmid expressing venus fluorescent protein was kindly provided by Robert Tjian [19]. A CRISPR guide RNA targeting the first exon of the YBX1 open reading frame was selected using the CRISPR design tool [20]. The YBX1 guide RNA was introduced into pX330-Venus by oligonucleotide cloning as described [19]. HEK293T cells were transfected for 48 h at low passage number, trypsinized and sorted for single, venus positive cells in a 96 well plate using a BD Influx cell sorter. Wells containing single clones (16 clones) were allowed to expand and were screened by semi-nested PCR using primers targeting the genomic region flanking the guide RNA site. Primers for the first round of PCR (10 cycles) were: YBX1-F1 (GGTTGTAGGTCGACTGAATTA) and YBX1-R1 (ACCGATGACCTTCTTGTCC). The PCR primers from the first round were removed using DNA clean and concentrator-5 kit (Zymo Research) according to manufacturer's instructions and the second round of PCR (25 cycles) was performed with primers: YBX1-F2 (CGGCCTAGTTACCATCACA) and YBX1-R1 (ACCGATGACCTTCTTGTCC). PCR products were separated on a 2.5% agarose gel to identify products smaller than the wild type PCR product, indicating a deletion. Clones (8) showing homozygous or heterozygous mutations were then screened by immunoblot to identify those that did not express YBX1. A clone containing a single homozygous mutation at the target site and not expressing YBX1 by immunoblot was recovered and designated ΔYBX1.

siRNA Knockdown and measurement of miR-223 secretion

Predesigned siRNA oligos targeting YBX2 were obtained from Qiagen (Hs_YBX2_3:AAGCCGGTGCTGGCAATCCA). Cells were seeded at 60% confluency and siRNA was transfected using Lipofectamine 2000 reagent (Life Technologies). After 48 h the media was replaced with exosome depleted media and allowed to incubate another 24 h. An aliquot (1 ml) of medium was removed and centrifuged at 1,500 and 15,000Xg and then extracted using the Zymo RNA Prep Kit. 1 ng of RNA was reversed transcribed and qPCR was performed as described above.

Results

Purified exosomes contain RNA

We first sought to purify exosomes from other extracellular vesicles and contaminating particles containing RNA (e.g. aggregates, ribonucleoprotein complexes) that sediment at high speed. I define exosomes as ~30-100 nm vesicles with a density of 1.08-1.18 g/ml and containing the tetraspanin protein CD63. Based on these criteria, purified exosomes were recovered using a three-stage purification procedure (Fig. 1-1A). First, large contaminating cellular debris was removed during low and medium speed centrifugation and exosomes were concentrated by high-speed sedimentation from conditioned medium. Next, to eliminate non-vesicle contaminants, the high-speed pellet fraction was suspended in 60% sucrose buffer and overlaid with layers of lower concentrations of sucrose buffer followed by centrifugation to float vesicles to an interface between 20 and 40% sucrose. Analysis of this partially purified material by electron microscopy showed vesicles of the expected size and morphology with fewer profiles of larger (>200 nm) membranes and reduced appearance of protein aggregates (Fig. 1-1B,C compared to Fig. 1-1D,E). Finally, sucrose gradient fractions were mixed with CD63 antibody-immobilized beads to recover vesicles enriched in this exosome marker protein.

To monitor and quantify the exosome purification, I generated a stable, inducible HEK293 cell line expressing a CD63-luciferase fusion. Tetraspanin proteins share a common topology in which the amino- and carboxyl-termini face the cytoplasm resulting in a predicted orientation inside the lumen of an exosomal vesicle. Although an intact C-terminal sequence is reported to be required for the proper localization of CD63 to the cell surface [21], I found that the overexpressed CD63-luciferase fusion was localized to a variety of cell surface and intracellular membranes (Fig. 1-2). Using isolated exosome fractions, I confirmed that the CD63-luciferase fusion maintained the expected topology. Luciferase activity was stimulated by the addition of detergent to disrupt the membrane and allow access to the membrane impermeable substrates luciferin and ATP, and to trypsin, which inactivated luciferase activity in the presence but not in the absence of detergent (Fig. 1-1F). The CD63-luciferase cell line was then used to monitor exosome purification. CD63-luciferase specific activity increased at each step of the purification, yielding a 5-fold purification of exosomes from the starting 100,000Xg pellet (Fig. 1-1G). Additionally, following immunoisolation, most of the RNA was found in the CD63 positive bound (B) fraction, showing that RNA is associated with purified exosomes from HEK293T cells (Fig. 1-1H). These results established that the RNA is associated with CD63-containing exosomes, but not necessarily enclosed within exosomes.

Exosomes contain selectively packaged microRNAs

Previous reports indicated the presence of microRNAs (miRNAs) in fractions containing exosomes but which also contain contaminating particles [6, 7, 10]. To identify the specific miRNAs that are enriched in CD63-positive exosomes from 293T cells, I performed Illumina-based small RNA sequencing on libraries prepared from purified exosomes and from cells. I obtained a total of 123,679 miRNA reads (4.4% of total mapped reads - Table 1-1) in the exosome library representing 502 distinct miRNAs and 880,093 reads (7.3% of total mapped reads - Fig. 2 supplement) representing 637 miRNAs in the cell library (Fig. 1-3A). To determine if a particular miRNA species was over-represented in exosomes, I analyzed the datasets for reads mapping to miRNA precursors and the targeting or passenger strand of mature miRNAs (Fig. 1-3B). Exosomes were slightly enriched in reads mapping to precursor and passenger strand transcripts, however, the vast majority of miRNAs (91% from cells and 88% from exosomes) mapped to the mature targeting strand. The relative abundance of each miRNA was estimated by normalizing to the total number of miRNA-mapped reads (i.e. the number of reads mapped to a miRNA locus divided by the total number of miRNA mapped reads for each dataset - RPM). Of these, 134 and 269 miRNAs were uniquely found in the exosome and cell datasets respectively (Fig. 1-3A). Most of the miRNAs uniquely found in exosomes were of very low abundance, with only a few counts for each miRNA. A notable exception was miR-223-3p, which was in the 72nd percentile for normalized reads in exosomes (Fig. 1-3C,D - red). The relatively high abundance of miR-223 in exosomes and its low level in the cellular library indicated that miR-223 was very efficiently packaged and secreted via exosomes.

We also identified miRNAs that were found in both libraries but were highly enriched in exosomes. A total of 368 miRNAs were detected in both the exosome and cell libraries. Most of these miRNAs were more enriched in cells than exosomes; however, some (e.g. miR144-3p, miR150-5p, miR142-3p) were highly enriched in exosomes and, like

miR223, were moderately abundant in exosomes (Fig. 1-3C,D - yellow). To summarize, small RNA sequencing from purified exosomes and subsequent miRNA analysis identified several putative exosomal miRNAs.

miR-223 and miR-144 are selectively packaged exosomal miRNAs

We selected miR-223 and miR-144 from the group of unique and enriched miRNAs for further analysis, as these were the most abundant species with documented functions that mature by the normal pathway of miRNA biogenesis. I performed quantitative reverse transcriptase PCR (qRT-PCR) for each miRNA target during the course of exosome purification from conditioned medium. my results showed a selective enrichment of both miR-223 and miR-144 at each stage of the purification (Fig. 1-3E and 1-3F). Thus, these miRNAs are associated with CD63 exosomes. To determine if these exosome associated RNAs are contained within exosomes and not simply bound to the surface, I performed an RNase protection experiment. Both miRNAs were protected from RNase I digestion, unless detergent was added to disrupt the membrane (Fig. 1-3G). These results confirm that miRNAs are selectively packaged into exosomes purified from HEK293T conditioned media and establish miR-223 and miR-144 as specific exosomal miRNAs.

Cell-free assays for exosome biogenesis and miRNA packaging

The mechanism of exosome biogenesis has been probed in mammalian cell culture using the tools of gene knockdown, knockout and overexpression where it is often difficult to distinguish a primary or indirect role for a gene product. I sought to minimize these challenges by developing simple biochemical assays that reproduce an aspect of exosome biogenesis and miRNA packing in a cell-free reaction.

Exosome biogenesis *in vitro*

Since packaging of cargo into newly-formed vesicles presumably occurs concurrently with membrane budding, I sought to develop an assay to monitor the incorporation of an exosome membrane cargo protein into a detergent sensitive membrane formed in an incubation containing membranes and cytsolic proteins obtained from HEK293 cells (Fig. 1-4A). Previous studies have reported the use of cell-free assays to monitor multivesicular body biogenesis and sorting of ubiquitinated membrane proteins into intraluminal vesicles [22-24]. To specifically monitor exosome biogenesis, I measured the protection over time of luciferase fused to CD63. The fusion protein displayed luciferase on the cytoplasmic face of a membrane such that its incorporation into a vesicle by budding into the interior of an endosome (or into a vesicle that buds from the cell surface) would render the enzyme sequestered and inaccessible to exogenous luciferin and ATP, the substrates of catalysis (Fig. 1-1F). During the incubation, substrate would have access to luciferase exposed on the cytoplasmic face of a membrane or to enzyme about to be internalized into a bud, but not to luciferase that had already become sequestered within vesicles in cells prior to rupture (Fig. 1-4A). In order to focus only on luciferase that became sequestered during the cell-free reaction, aliquots taken after a 20 min incubation of membranes, cytosol and substrate at 30ºC were sedimented, resuspended in buffer and sedimented again to remove excess substrate, and CD63-luciferase remaining exposed on the surface of membranes was inactivated by treatment with trypsin (0.5mg/ml) for 1h at 4ºC. Remaining luciferase activity was then monitored in a luminometer. Since the substrates D-luciferin salt and ATP are membrane

impermeable, the residual luminescence measured in trypsin-treated samples should derive from luciferase that became segregated along with substrate during the cell-free incubation. Relative protection was quantified as the ratio of relative light units (RLU) for reactions incubated without cytosol, at 4ºC and in the presence of detergent (Triton X-100) divided by the RLU for a complete reaction incubated at 30ºC. I observed that the formation of sequestered luciferase required cytosol and incubation at 30ºC and was disrupted in incubations containing detergent (Triton X-100) (Fig. 1-4B).

To examine the connection between the formation of sequestered luciferase and exosome biogenesis, I performed my cell-free reaction in the presence of an inhibitor (GW4869) of neutral sphingomyelinase, an enzyme that cleaves sphingolipid to form ceremide [25]. Treatment of cells with an inhibitor of neutral sphingomyelinase 2 (NS2) reduces the secretion of exosomes and exosome-associated miRNAs [26-28]. GW4869 inhibited the protection of CD63-luciferase protection in my cell-free assay at concentrations required to inhibit NS2 activity in partially purified fractions of the enzyme (Fig. 1-5D) [25]. Thus, my cell-free reaction may recapitulate an aspect of exosome biogenesis.

MicroRNA packaging into vesicles in the cell-free reaction

Having identified miRNAs that are selectively packaged into exosomes in cultured cells, I examined my cell-free reaction for the RNA-selective segregation of miRNAs into RNase resistant and detergent sensitive vesicles (Fig. 1-5A). As in the biogenesis reaction described above, crude membranes and cytosol from broken cells were mixed in buffer containing an ATP regenerating system, but in this case supplemented with synthetic miRNA, specifically miR-223, the miRNA that was most highly enriched in exosomes isolated from the medium of cultured HEK293T cells (Fig. 1-3d). Given the low relative abundance of miR-223 in HEK293T cells, I set the exogenous concentration in the incubation to be ca. 1000 fold in excess to ensure that the chemically synthetic material predominated in any packaged signal. After a 20 min incubation at 30ºC, aliquots were treated with RNase I to digest any unpackaged miRNA. RNA was then purified, reverse transcribed using a miRNA specific primer and the amount of miRNA that became protected during the incubation was measured by quantitative PCR. Packaged RNA was quantified as the percentage of miR-223 RNA molecules protected from RNase during the course of incubation. Packaging of miR-223 required membranes, cytosol and incubation at physiologic temperature (Fig. 1-5B). As expected for the segregation of miR-223 into a membrane bound compartment, addition of TX-100 during the RNase incubation abrogated protection (Fig. 1-5B). Furthermore, at a minimal concentration of cytosol (0.5 mg/ml), protection was stimulated two-fold over reactions performed in the in the absence of ATP, or in the presence of apyrase or a non-hydrolyzable analog of ATP (Fig. 1-5C). The sphingomyelinase inhibitor GW4869 inhibited miR-223 packaging to the same extent and at similar concentrations to its effect on the formation of sequestered luciferase in the biogenesis reaction (Fig. 1-5D). Thus, the biogenesis and miRNA packaging reactions display similar biochemical requirements and may reflect the same process of exosome formation in my cell-free reaction.

MicroRNA-223 is selectively packaged into vesicles *in vitro*

Having shown that specific miRNAs are enriched in exosomes produced *in vivo*, I next determined if selective sorting of miRNAs into vesicles could be reconstituted *in vitro*. I

used the cell-free packaging assay to compare the efficiency of incorporation of synthetic miR-223 and a relatively abundant cellular miRNA that is not found in exosomes (miR-190a-5p - miR-190). Exosomal miR-223 was more efficiently packaged into vesicles (9%) than cellular miR-190 (1.5%) (Fig. 1-5E). Furthermore, the rate of miR-223 packaging mirrored the rate at which luciferase became sequestered in the biogenesis reaction whereas the rate of miR-190 protection in a 30ºC incubation reflected the low rate of formation of sequestered luciferase in an incubation held on ice (Fig. 1-5F). Based on these experiments, I conclude that the cell-free packaging assay reconstitutes the selective sorting of exosomal miR-223 over cellular miR-190 into vesicles, possibly exosomes, formed *in vitro*.

Identifying candidate proteins involved in miRNA sorting into exosomes

To identify proteins that may be involved in miRNA packaging into exosomes, I employed a proteomics approach utilizing the *in vitro* packaging assay to capture RNA binding proteins. MiRNA sorting may require an RNA binding protein to segregate an RNP into a nascent budded vesicle. Synthetic 3' biotinylated miR-223 was substituted for unmodified miRNA in the cell-free reaction. Samples were treated with RNase, quenched with RNase inhibitor and solubilized with Triton X-100. miR-223-biotin was captured on streptavidin-coated beads and interacting proteins were eluted with high salt buffer. Mir-223-interacting proteins were identified by in-solution liquid chromatography/mass spectrometry (Fig. 1-6A). Based on peptide count and coverage, the most highly represented protein was Y-box binding protein I (YBX1) (Fig. 1-6B). Peptides representing >45% YBX1 of the protein were identified stretching from the cold shock domain to the C-terminus (Fig. 1-6C).

YBX1 is a multi-functional RNA binding protein that shuttles between the nucleus, where it plays a role and splice site selection [29, 30], and the cytoplasm where it is required for the recruitment of RNAs into cytoplasmic ribonucleoprotein granules containing untranslated mRNAs and plays a role in mRNA stability [31]. YBX1 also co-localizes with cytoplasmic P-bodies containing members of the RISC complex, including GW182 which can be found in exosomes [32, 33]. Interestingly, YBX1 is secreted in a form that resists trypsin in the absence but not in the presence of a non-ionic detergent (Triton X-100), consistent with a location in vesicles, perhaps exosomes [34, 35]. Furthermore, YBX1 has been detected by mass spectrometry in isolated exosomes [12, 13]. I first determined if YBX1 co-purifies with exosomes. I purified exosomes as in Fig. 1a and found that YBX1 was primarily associated with the CD63-bound fraction containing known exosome markers (TSG101, Alix, CD9), as opposed to flotillin 2, which was predominantly found in the unbound fraction (Fig. 1-6D). The CD63 positive (exosome) fraction contained most of the RNA (Fig. 1-1F). These results show that HEK293T cells release at least two vesicle types (CD63 positive and negative).

We next examined the biochemical requirements for co-packaging of miR-223 and YBX1 using the biotin-miR-223 packaging reaction described in Fig. 1-6A. An immunoblot showed YBX1 bound to biotin-mi223 was recovered in exosomes in a complete reaction, while no detectable YBX1 was recovered in exosomes in incubations that lacked cytosol or membranes and much reduced signals in control incubations held at

4C or conducted in the absence of biotin-mi223 (Fig. 1-6E). These conditions mirror those required for the packaging of miR-223 in my cell-free reaction.

Because YBX1 is a known RNA binding protein, is secreted by cells in exosomes and physically interacts with miR-223 during the *in vitro* packaging assay, it met my criteria for a potential exosomal miRNA sorting factor.

Lack of evidence for a specific role for Ago2 in sorting miR-223 into exosomes
Given that mature miRNA guide sequences in the cell are bound to an argonaute family protein, I were surprised that I did not detect any of these proteins associated with miR-223 in the streptavidin-bound fraction. Argonaute proteins have been variously described as being inside of extracellular vesicles [4, 36] or released as free proteins independent of vesicles [37, 38]. I first performed immunoblot on 100,000Xg pellet fractions (which should contain vesicle associated and non-vesicle associated Ago2) and vesicle fractions purified on a buoyant density gradient (which should eliminate non-vesicle Ago2) to determine if I could detect Ago2 in fractions that contain exosomes. As expected, after flotation the exosome marker proteins were enriched compared to the pellet fraction (Fig. 1-7A). In contrast, Ago2 was not detected in the density gradient purified vesicle fraction (Fig. 1-7A). The apparent lack of Ago2 in vesicles could be due to its absence in vesicles or a relatively low chemical abundance relative to other molecules. Regardless, my results show that in HEK293T cells the large majority of extracellular Ago2 exists as a non-vesicle associated species. These results support previously published evidence that Ago is not associated with density gradient purified exosomes isolated from breast cancer (MCF7) cells [39]

To further investigate the role of Ago2 in sorting miR-223 into exosomes, I employed the cell-free packaging assay. The inability to detect Ago2 in my mass spectrometry results from *in vitro* packaged miR-223 could be due to technical limitations of my reaction. I considered two potential technical explanations. First, I used the single stranded guide RNA in my reaction rather than a duplex RNA molecule. Although Ago2 associates with guide RNA in a reaction containing purified components [40, 41], single-stranded miRNA may be rapidly degraded in a crude extract [41]. Thus, it was possible that the duplex molecule might represent a more relevant packaging substrate. Second, 3' biotinylated RNA was used as the substrate for the packaging reaction. A recent report suggested that, in some cases, 3' biotinylated RNA could prevent the proper loading of the miRNA into complex with Ago2 [42]. To address these concerns I synthesized miR-223 passenger strand and guide strands biotinylated at the 3' end or internally at position 13 (Fig. 1-7B). The internal position was chosen because in crystal structures of miRNA-protein complexes, middle positions of guide RNAs do not appear to be in direct contact with Ago2 [42-44]. I then annealed the guide and passenger strands to form the miR-223 duplexes. To determine if the guide or passenger strands can be efficiently loaded into Ago2 in my *in vitro* reaction conditions, I first mixed the biotinylated substrates with cytosol alone and evaluated complexes that associated with streptavidin beads. In my reaction conditions, both the single stranded guide and duplex oligonucleotides bound apparently equally to Ago2 in cytosol alone and there was no discernible difference in the association comparing 3' or internally biotinylated oligonucleotides (Fig. 1-7C). I then tested the various substrates in my complete packaging reaction including membranes,

cytosol and an ATP regenerating system. Duplex substrates were packaged ~2-fold more efficiently than the single stranded guide RNA, irrespective of the position of the biotin group (Fig. 1-7D). Interestingly, in reactions programmed with duplex RNA substrate, only the guide RNA and not the passenger RNA was sorted into a protected compartment (Fig. 1-7D). In similar incubations, the YBX1 protein was ~2X more efficiently packaged in reactions programmed with duplex RNA but Ago2 was not detected associated with miR-223 in any of the complete reactions (Fig. 1-7E,F). This suggests that whereas the RNA substrates are capable of being bound by Ago2 in the cytosol, in the complete reaction containing membranes and ATP, YBX1 is the predominant binding factor. These results explain why Ago2 was not detected in the mass spectrometry data and are consistent with my failure to detect Ago2 in buoyant density-fractionated extracellular vesicles. Similarly, Van Deun et al. (2014) found no evidence of Ago2 in density gradient-isolated exosomes from MCF7 cells [39].

Y-box protein 1 is involved in sorting miR-223 into exosomes

Given the absence of Ago2 in exosomes or associated with miR-223 in my cell-free RNA sorting reaction, I focused on the primary candidate RNA binding protein found in my mass spectrometry results. I next evaluated the requirement for YBX1 in packaging exosomal miRNAs in cells and in the cell-free reaction. To address this question, I generated a YBX1 knockout HEK293T cell line with CRISPR/Cas9 using a guide RNA targeting the YBX1 locus [19, 45, 46]. Clones were screened by genomic PCR and immunoblot for YBX1. I recovered a homozygous mutant clone (ΔYBX1) that had been targeted at the YBX1 locus and no longer expressed YBX1 protein (Fig. 1-8A). The homozygous mutant cells grew normally under the conditions used to propagate HEK293T cells and released an approximately equal number of particles into the medium after 48 h of growth ($2.38X10^7$ and $2.42X10^7$ particles/ml for wild type and ΔYBX1) as determined by Nanosight nanoparticle tracking analysis.

To determine if YBX1 was required for miRNA packaging, I prepared cytosol from ΔYBX1 cells and tested miR-223 incorporation using the *in vitro* packaging assay. Cytosol from ΔYBX1 cells did not support miR-223 protection *in vitro* but activity was largely restored in reactions containing cytosol from a ΔYBX1 line transfected with plasmid encoding YBX1 (Fig. 1-8B). I also evaluated the role of YBX1 in my biogenesis reaction (Fig. 3a) and found that cytosol from wt and ΔYBX1 were indistinguishable in the formation of latent luciferase activity (Fig. 1-8C). Thus, YBX1 is required for exosomal miRNA packaging *in vitro* but is not required for the sorting of an exosome membrane cargo protein into vesicles in my cell-free reaction.

In an effort to connect the results of my cell-free reaction to the mechanism of sorting of miRNAs into exosomes secreted by HEK293T cells, I examined the secretion of miR-223 and of another miRNA, miR-144, which was also highly enriched in my purified exosome fraction (Fig. 1-3D). I measured the amount of miR-144 and miR-223 secreted into the medium and retained in cells by qRT-PCR. ΔYBX1 cells showed a significant decrease in secretion of both miRNAs, though more notably of miR-144, during a 24 h incubation in fresh medium (Fig. 1-8D). When the YBX1 paralog YBX2 was knocked down in ΔYBX1 cells, miR-223 secretion was diminished to the baseline level, suggesting partial functional redundancy for YBX paralogs in the secretion of miR-223

(Fig. 1-9). The defect in miR-223 secretion was less substantial in cells compared to the cell-free reaction (Fig. 1-8B), possibly reflecting a rate effect that distinguishes the magnitude of the defect in a 20 min cell-free incubation at 30C vs. a 24h incubation of cells at 37C. Nonetheless, the secretion defect was accompanied by an ~3-4 fold accumulation of each miRNA in cells. No accumulation was observed for another miRNA (miR-190), which is not released in exosomes (Fig. 1-8D). These results show that YBX1 controls secretion of select exosomal miRNAs *in vitro* and *in vivo*.

Discussion

Our results establish miR-223 and miR-144 as specific exosome cargo in HEK293T cells. In order to probe the mechanism of RNA sorting into exosomes, I developed biochemical assays that measure the capture of an exosome membrane protein and miRNA into vesicles formed in a cell-free reaction. Using this approach, I identify YBX1 as an RNA-binding protein that is critical for the efficient packaging of miR-223 *in vitro* and its secretion in cultured human cells.

How are miRNAs recognized for sorting into exosomes?

We find that synthetic miR-223 is sequestered into vesicles more efficiently than miR-190, consistent with the possibility of a primary RNA sequence or secondary structure, perhaps stabilized by an RNA binding protein such as YBX1, that directs RNA sorting. One possible sorting motif – GGAG – is enriched in miRNAs secreted in exosomes from T-cells [47]. This motif is recognized by hRNPA2B1, a T-cell exosome RNA-binding protein, which requires sumoylation for efficient secretion via exosomes. It was therefore suggested that binding of GGAG containing miRNAs by sumoylated hRNPA2B1 was a sorting mechanism for miRNAs into T-cell-derived exosomes.

We were unable to identify any statistically significant primary sequence motifs for miRNAs by either multiple alignment (ClustalW) or multiple Em for motif elicitation (MEME) in HEK293T-derived exosomes [48, 49]. Furthermore, the mature targeting strand of miR-223 packaged into exosomes contains no guanine nucleotides and hRNPA2B1 was not detected in my mass spectrometry results for proteins bound to miR-223-biotin isolated from vesicles formed in my cell-free reaction. The human genome encodes more than 1,000 experimentally determined and predicted RNA-binding proteins [50, 51]. I therefore propose that different cell types may use RNA binding proteins with distinct binding preferences to secrete miRNAs, and perhaps other RNA classes, in exosomes. In addition, some cell types may deploy multiple RNA-binding proteins to sort RNAs into exosomes, in which case motif discovery would be challenging, even in highly purified vesicles, due to diverse motif preferences from distinct proteins.

We identified YBX1 as the dominant RNA-binding protein physically interacting with miR-223 *in vitro* and confirmed its role in miR-223 packaging into exosomes both *in vitro* and in cultured cells. YBX1 is found within mammalian P-bodies (GW bodies) containing untranslated RNAs [52] and in the nucleus where it plays a role in RNA splice site selection by binding short sequence motifs [29, 30]. YBX1 binds RNA via an internal cold shock domain and an inherently disordered, highly charged C-terminus [31]. Interestingly, another cold shock domain containing protein, Lin28, binds pre-miRNAs of the Let7 family via hairpin-loop structures [53]. YBX1 also binds hairpin-loops in a

murine retrovirus, leading to stabilization of the viral RNA genome and increased particle production [54]. YBX1 binding of viral RNA also increases production of other retroviruses, including HIV [54-56]. This raises the possibility that the recognition motif for sorting into exosomes may be based on secondary rather than primary RNA structure and that YBX1 may act as an RNA co-factor to escort exosomal RNAs into exosomes.

Our studies focused on miRNAs, however it is possible that YBX1 is responsible for the secretion of other RNA classes in exosomes. Several recent reports indicate a role for YBX1 in binding various small RNAs, including miRNAs, tRNA fragments and snoRNAs [57-59]. It is interesting to note that most miRNAs present in exosomes in my study are not highly enriched compared to their relative abundance in cells. This raises the possibility that highly enriched exosomal miRNAs mimic other classes of RNAs that are more efficiently packaged in a YBX1-dependent manner.

Chaperone-mediated sorting of miRNAs into exosomes

A surprising finding from my study is the lack of evidence for the argonaute proteins in isolated exosomes or sequestered with miR-223 in my cell-free RNA sorting reaction. This in spite of my observation that cell-free reaction reconstitutes sorting of the mature strand from a duplex RNA. Some recent evidence suggests that Ago2 may be sorted along with miRNAs into exosomes as a result of aberrant KRAS signaling. Analysis of extracellular RNA in isogenic cell lines differing only in KRAS status revealed that secretion of a sub-population of miRNAs is decreased in colorectal cancer cells harboring an activating KRAS mutation whereas other miRNAs are secreted at equivalent levels irrespective of KRAS status [60]. Subsequent experiments showed that KRAS mutation results in phosphorylation of Ago2 causing its re-localization from multivesicular bodies to P-bodies leading to decreased secretion of select miRNAs [61]. Notably, the miRNA (miR-223-3p) shown here to be dependent on YBX1 is among the cohort of miRNAs that were not affected by KRAS status. These results combined with ours suggest two possible routes for miRNA egress via exosomes, an Ago2-associated pathway and a RNA-binding protein-dependent pathway that I term chaperone-mediated sorting. The chaperone-mediated pathway would include the previously identified hnRNPA2B1 in T-cells[47] and YBX1 in HEK293T cells. Interestingly, one other highly enriched miRNA (miR-328-5p) has been previously shown to conditionally associate with either Ago2 or with the RNA-binding protein hnRNPE2[62], suggesting that these two pathways may not be mutually exclusive. A notable feature of the chaperone-mediated pathway is that the RNA content of exosomes may be manipulated by altering the expression of individual RNA binding proteins involved in RNA export with distinct nucleic acid binding specificities. Further characterization of the chaperone-mediated pathway may then allow for targeted sorting of engineered RNA species into exosomes.

The physiological role of exosomal miRNA

The functional role of secreted miRNAs has been a matter of discussion since the first reports of extracellular RNA [63]. Numerous studies have shown that miRNAs can be transferred to neighboring cells in experimental settings [6-8, 60, 64-66]. However, the transfer of miRNAs in biologically significant quantities for function in a physiological context is far from proven. Indeed, a recent study reported a stoichiometry of less than one specific miRNA per exosome, with the caveat that this study characterized crude,

high-speed pellet fractions from conditioned medium [67]. Functional miR-223 transferred between macrophages and miR-223-containing exosomes can induce macrophage differentiation, however, it has yet to be shown that miR-223 transfer plays a direct role in the differentiation [68]. Indeed, direct and convincing evidence for a physiological role of miRNAs secreted via exosomes has so far proven elusive. Alternatively, exosomes may be a convenient carrier to purge unnecessary or inhibitory RNAs from cells. A recent report provided evidence for both alternative views with the demonstration that target transcript levels for miRNAs in the cell modulate the abundance of miRNAs in macrophage exosomes, and this in turn dictates which miRNAs are transferred to repress transcripts in recipient cells [37]. Because YBX1 and the RISC machinery have both been shown to localize to P-bodies and P-bodies are closely juxtaposed to multivesicular bodies, all of the necessary machinery is poised to efficiently secrete miRNAs in exosomes [38]. YBX1 may complex with miRNAs whose mRNA targets are not expressed, and sort them into the intralumenal vesicles of a multivesicular body for export by unconventional secretion. Physiological studies of the function miRNAs that are secreted via the Ago2-associated vs. chaperone-mediated pathways may explain contradictory results for different miRNAs and provide general rules for extracellular miRNA function.

Figures

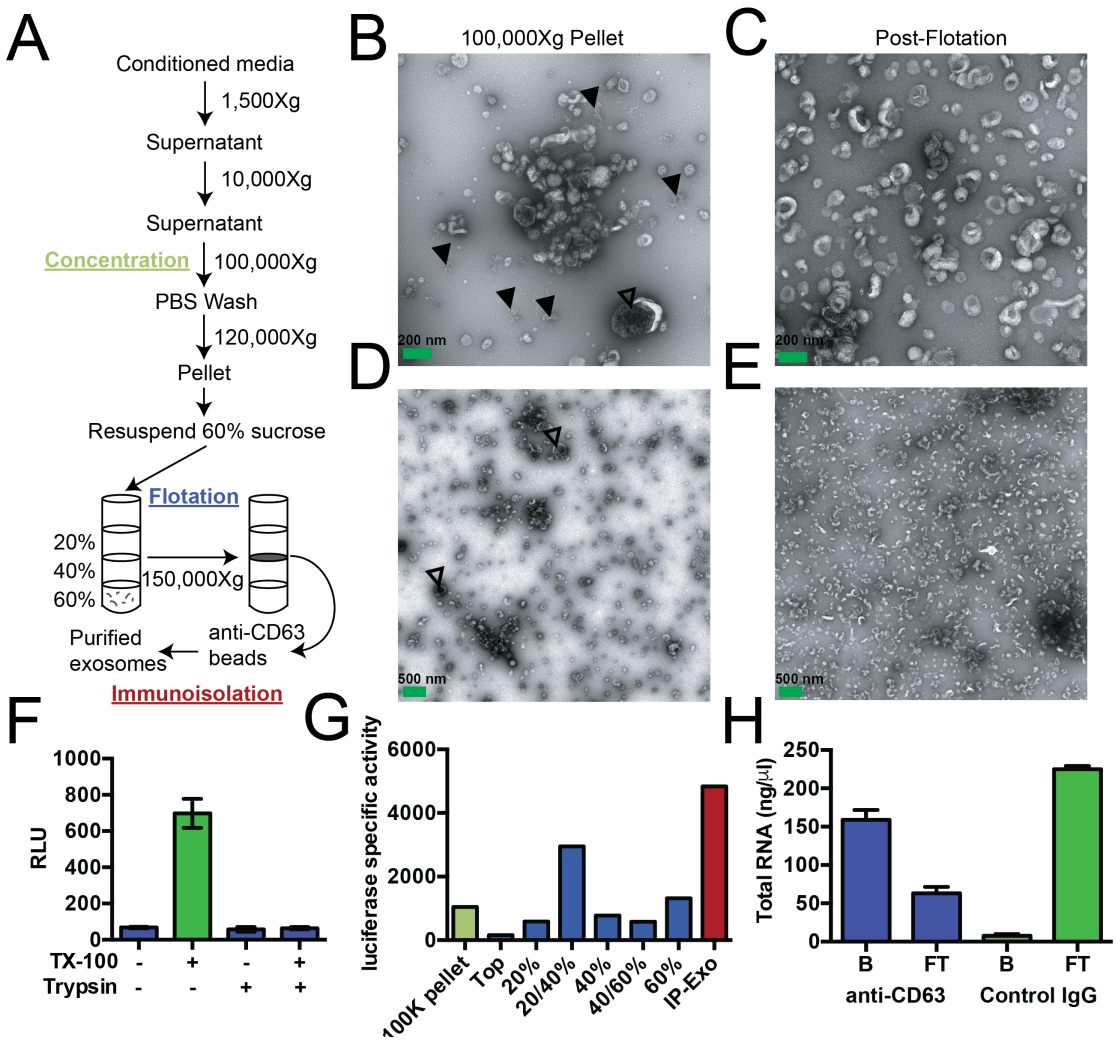


Fig. 1-1: Purified CD63-positive exosomes contain RNA

(A) Exosome purification schematic. (B-E) Representative electron micrographs of negative stained samples from the 100,000Xg pellet fraction (B,D) and post-flotation fractions (C,E) at either 9,300X (A,B) or 1,900X (C,E) magnification. Open arrows indicate large (>200 nm) vesicle contaminants and closed arrows indicate protein aggregates. (F) CD63-luciferase activity in purified exosomes after treatment with 1% Triton X-100 (TX-100) and/or 100 µg/ml trypsin for 30 min at 4ºC. Error bars represent standard deviations from 3 independent samples. (G) Specific activity of CD63-luciferase (RLU/µg of total protein) at each stage of purification (green: 100,000Xg pellet, purple: post-flotation, red: post-immunoisolation α-CD63 beads). (H) Total RNA recovered from conditioned medium after immuno-isolation with α-CD63 or an IgG control. B – bound to beads, FT – flow-through not bound to beads. Error bars represent standard deviations from 3 separate purifications (biological replicates).

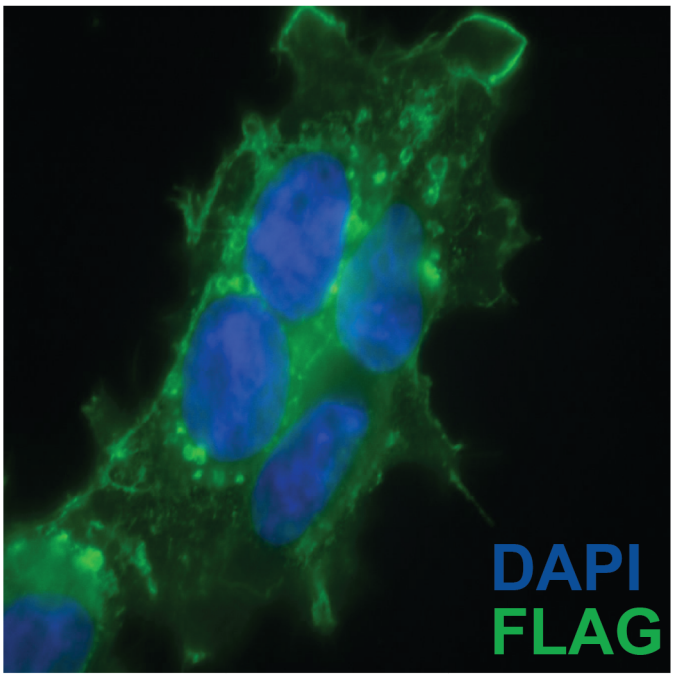
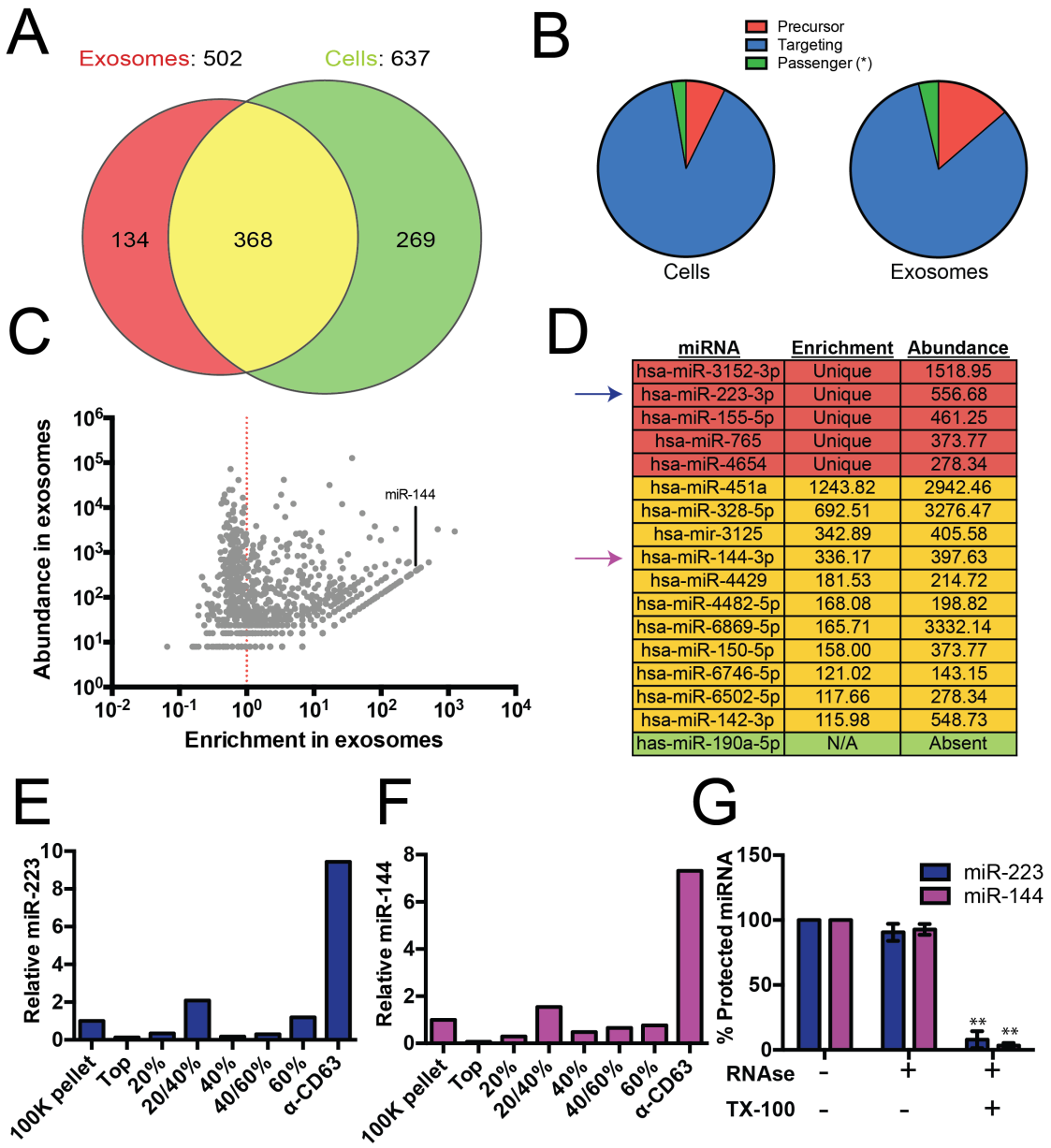


Fig. 1-2: Sub-cellular localization of C-terminal CD63-luciferase-FLAG fusion
CD63-luciferase-FLAG cells induced for 48 hours were fixed with 4% paraformaldehyde blocked with 5% BSA in PBS and stained with M2-Flag antibody (1:500) and then Alexa-488 conjugated anti-mouse secondary. Cells were mounted with prolong gold (containing DAPI stain) and imaged at 400X total magnification.



miRNA	Enrichment	Abundance
hsa-miR-3152-3p	Unique	1518.95
hsa-miR-223-3p	Unique	556.68
hsa-miR-155-5p	Unique	461.25
hsa-miR-765	Unique	373.77
hsa-miR-4654	Unique	278.34
hsa-miR-451a	1243.82	2942.46
hsa-miR-328-5p	692.51	3276.47
hsa-mir-3125	342.89	405.58
hsa-miR-144-3p	336.17	397.63
hsa-miR-4429	181.53	214.72
hsa-miR-4482-5p	168.08	198.82
hsa-miR-6869-5p	165.71	3332.14
hsa-miR-150-5p	158.00	373.77
hsa-miR-6746-5p	121.02	143.15
hsa-miR-6502-5p	117.66	278.34
hsa-miR-142-3p	115.98	548.73
has-miR-190a-5p	N/A	Absent

Fig. 1-3: Enrichment of select microRNAs in exosomes

(A) Venn diagram showing the number of total (above diagram), unique (inside red or green circles) and shared miRNAs (inside yellow) from each library. (B) Pie charts showing the relative proportion of reads mapping to each miRNA species (Precursor - red, passenger strand - green, targeting strand - blue) in cellular and exosome small RNA libraries. (C) Scatterplot showing the enrichment (reads per million miRNA mapped reads (RPM) in exosomes/cells) and relative abundance in exosomes (RPM) of all miRNAs found in both libraries. (D) Table showing the enrichment and abundance (RPM) of relevant miRNAs in exosomes. (Red - unique to exosomes, Yellow - highly enriched in exosomes, Green - unique to cells) (E,F) Relative miR-223 (E) and miR-144 (F) per ng of RNA as quantified by qRT-PCR during each stage of the purification. The 100K pellet

was set to 1. (G) RNase protection of exosomal miRNAs quantifed by qRT-PCR. Purified exosomes treated with or without RNase I and/or Triton X-100. Statistical significance was performed using Student's t-test (*$p<0.05$, **$p<0.01$).

Table 1: Mapping statistics for small RNA-seq libraries

Source	Mapped reads (hg19)	rRNA	miRNA	
293T Cells	12089110 (100%)	7905728 (65.4%)	880093	
			w/ rRNA (7.3%)	w/o rRNA 21.0%)
293T Exosomes	2811327 (100%)	2242164 (79.8%)	123679	
			w/ rRNA (4.4%)	w/o rRNA (21.7)

Reads were processed (see materials and methods) and mapped to the human genome (hg19) using Bowtie 2. Total counts for reads mapped to the genome, to rRNA and to miRNA (using miRdeep2 - see materials and methods) are shown. Percent of total reads are shown in parenthesis.

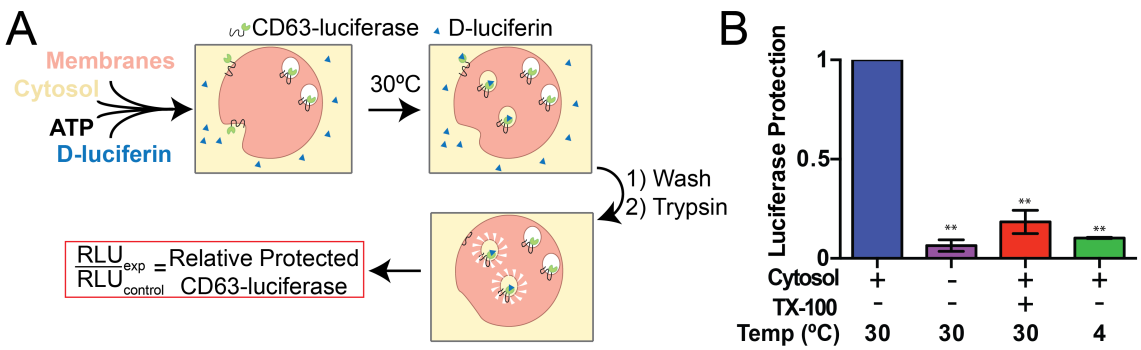


Fig. 1-4: Cell-free exosome biogenesis reaction
(A) Schematic illustrating the *in vitro* biogenesis reaction. (B) Exosome biogenesis measured by relative protected CD63-luciferase. Reactions with or without cytosol, 1%Tx-100 and incubation temperature are indicated. Statistical significance was performed using Student's t-test (**p<0.01).

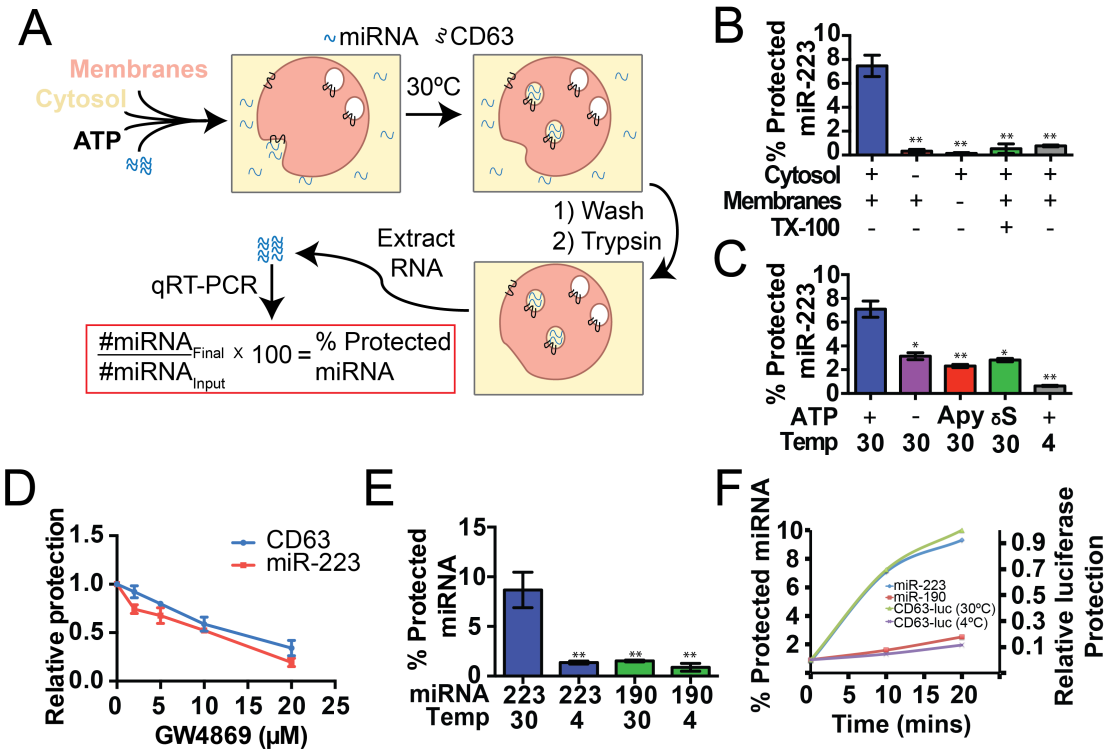
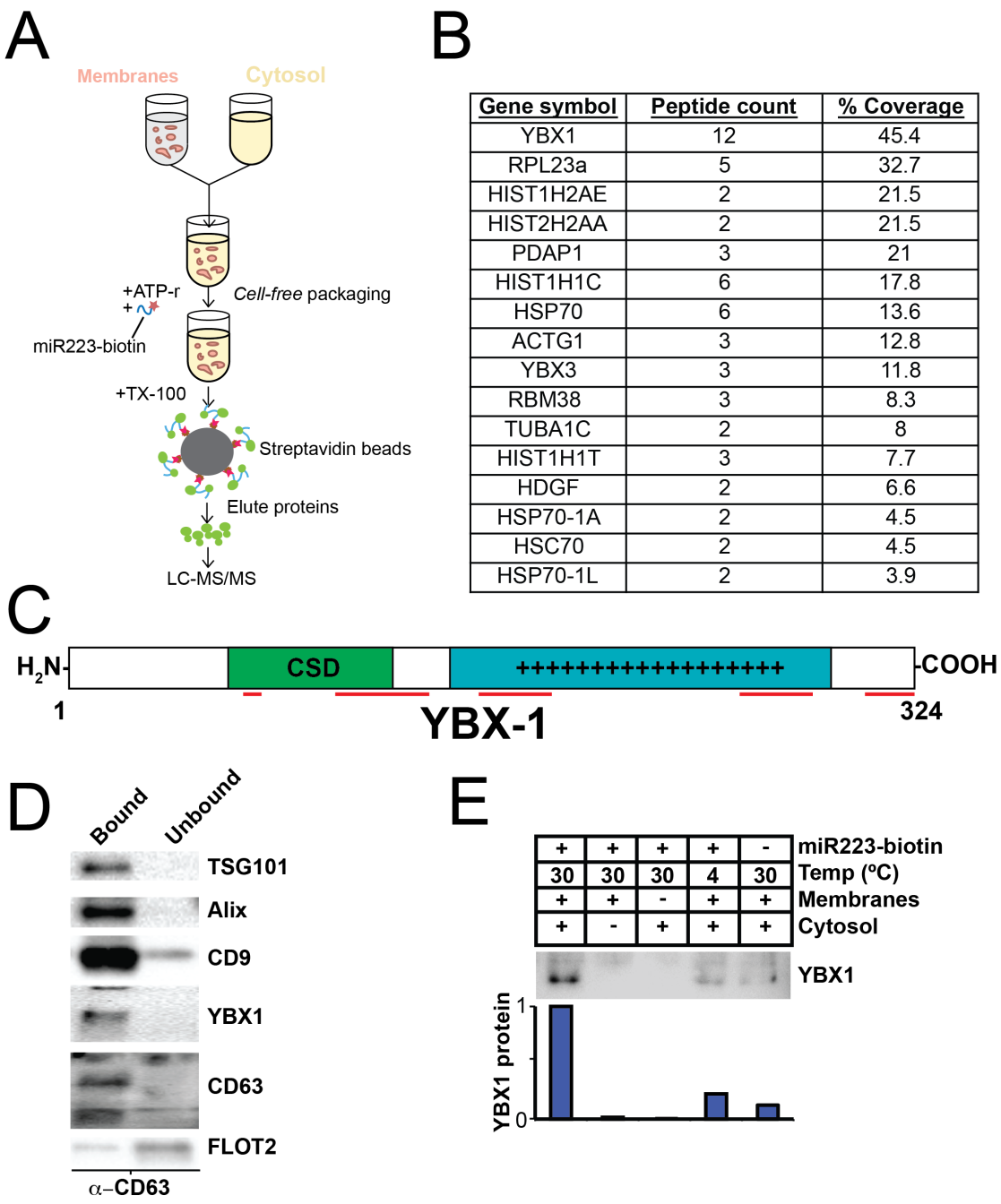


Fig. 1-5: Cell-free selective sorting of miRNA into exosomes

(A) Schematic illustrating the *in vitro* packaging reaction. (B) Cell-free packaging of miR-223 measured as percent protected by qRT-PCR. Reactions with or without membranes (15,000Xg pellet), cytosol (100,000Xg supernatant) and 1% Triton X-100 (TX-100), and incubated at 4 or 30 C are indicated. (C) ATP requirements for miR-223 packaging (Apy - Apyrase, (1 U/ml) γS – ATPγS (10 mM)). (D) Dose dependent effect of neutral sphingomyelinase 2 inhibitor (GW4869). Measured as relative protection of miRNA and CD63-luciferase normalized to vehicle only control (DMSO). (E) Protection of miR-223 or miR-190 measured as a percent protected by qRT-PCR. (F) Relative CD63-luciferase (right axis) and percent miRNA protection (left axis) measured over a 20-min time course using the indicated miRNA cargo and incubation temperature. Statistical significance was performed using Student's t-test (*p<0.05, **p<0.01).



Gene symbol	Peptide count	% Coverage
YBX1	12	45.4
RPL23a	5	32.7
HIST1H2AE	2	21.5
HIST2H2AA	2	21.5
PDAP1	3	21
HIST1H1C	6	17.8
HSP70	6	13.6
ACTG1	3	12.8
YBX3	3	11.8
RBM38	3	8.3
TUBA1C	2	8
HIST1H1T	3	7.7
HDGF	2	6.6
HSP70-1A	2	4.5
HSC70	2	4.5
HSP70-1L	2	3.9

Fig. 1-6: Identification of YBX1 as a candidate exosomal miRNA sorting protein
(A) Scheme to identify candidate miRNA sorting proteins (B) Proteins identified by tandem mass spectroscopy from the experiment illustrated in (A). (C) Schematic of YBX1 protein. The cold-shock domain (green) and positively charged low-complexity region (blue) are highlighted. Red lines indicate detected unique peptides from mass spectroscopy. (D) Immunoblots for the indicated protein markers in the CD63 immuno-isolated (bound) or unbound fractions. Exosomes were purified as in Fig. 1-1A. (E) Immunoblot for YBX1 following cell-free packaging reactions performed according to

the conditions indicated and immobilized with streptavidin beads as shown in (Fig. 1-6A). Bar graph represents densitometry values for the blot shown.

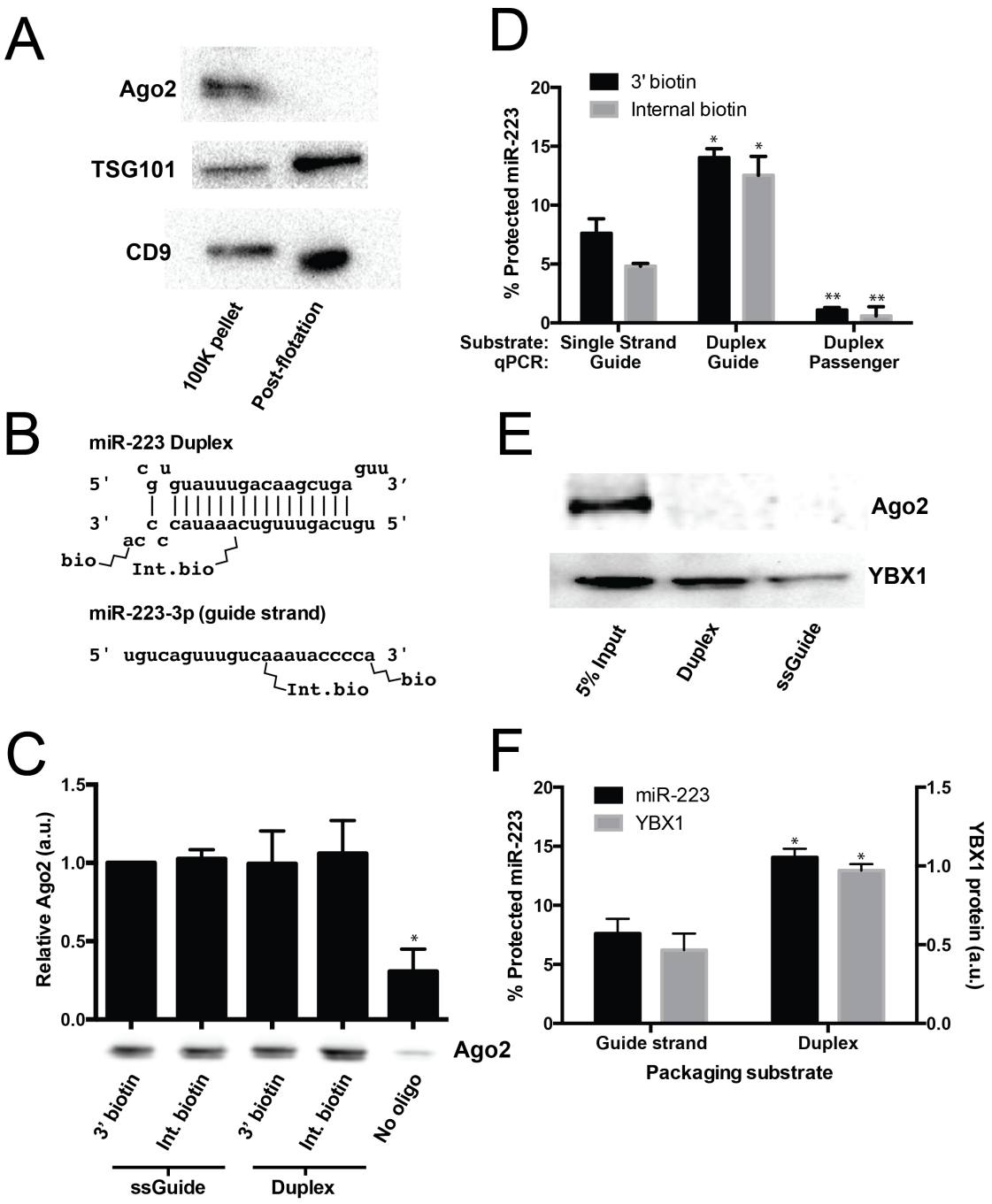


Fig. 1-7: Lack of evidence for a specific role for Ago2 in sorting miR-223 into exosomes

(A) Immunoblots for Ago2 and exosome markers TSG101 and CD9 in 100,000Xg (100K) pellet and the 20/40% sucrose interface fractions. (B) Schematics showing 3' and internally biotinylated miR-223 duplex and mature guide strand substrates. (C) Immunoblots for Ago2 from substrates mixed with cytosol alone for 30 min at 30°C and then absorbed on streptavidin-conjugated beads. (D) Percent protected miR-223 (either

guide strand or passenger strand) from 3' biotinylated or internally biotinylated single strand or duplex substrates. (E) Immunoblots for Ago2 and YBX1 from substrates packaged in the complete *in vitro* reaction and then absorbed on streptavidin-conjugated beads. (F) Percent RNAse protected miR-223 and relative level of streptavidin-absorbed YBX1 protein (normalized to duplex). MiR-223 and YBX1 quantification comes from data in (D) and (E), respectively. Statistical significance was performed using Student's t-test (*$p<0.05$, **$p<0.01$).

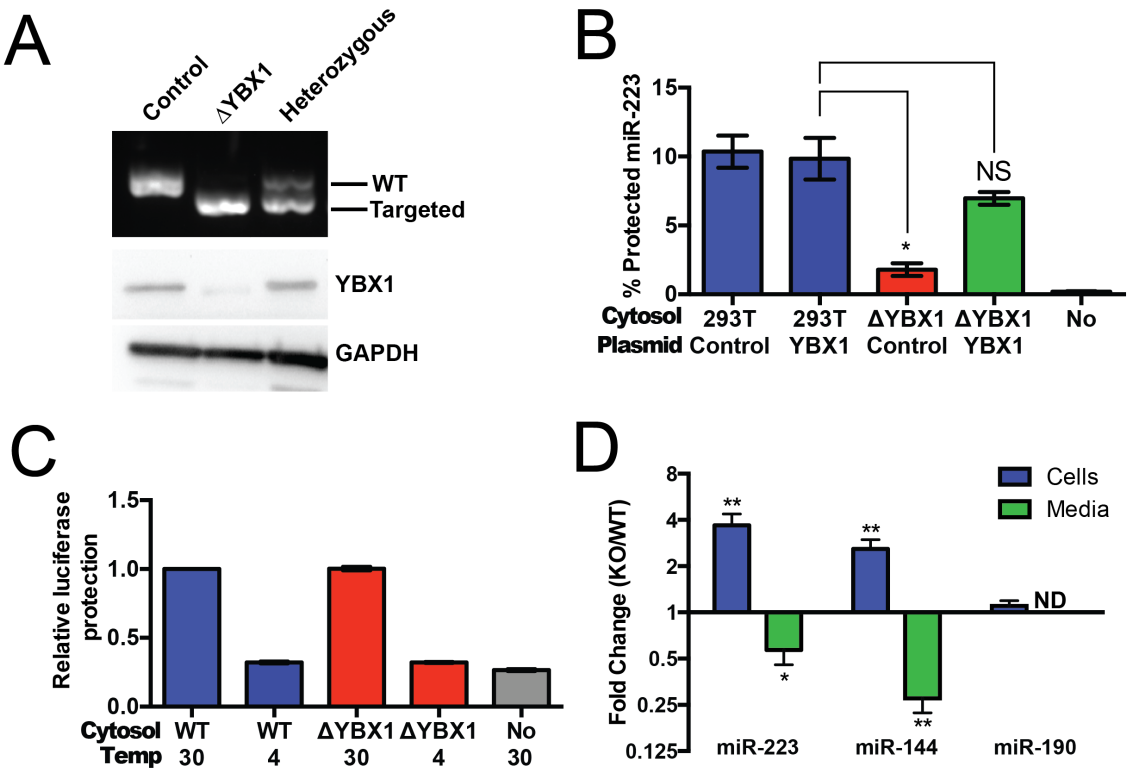


Fig. 1-8: YBX1 is necessary for exosomal miRNA packaging and secretion
(A) Analysis of wild-type and CRISPR/Cas9 genome edited HEK293T clones by PCR flanking the genomic target site (top) and immunoblot for YBX1 (middle) and GAPDH (bottom). (B) *In vitro* miR-223 packaging into exosomes from ΔYBX1 or WT cytosol transfected with control (pCAG) or YBX1 plasmid. (C) Cell-free exosome biogenesis with cytosol from ΔYBX1 or WT cells and membranes from CD63-luciferase cells. (D) Fold change of miR-223 and miR-144 in cells and media from by ΔYBX1 (KO) and WT cells (KO/WT) ND = Not detected. All quantifications represent means from three independent experiments and error bars represent standard deviations. Statistical significance was performed using Student's t-test (*$p<0.05$, **$p<0.01$ and NS = not significant).

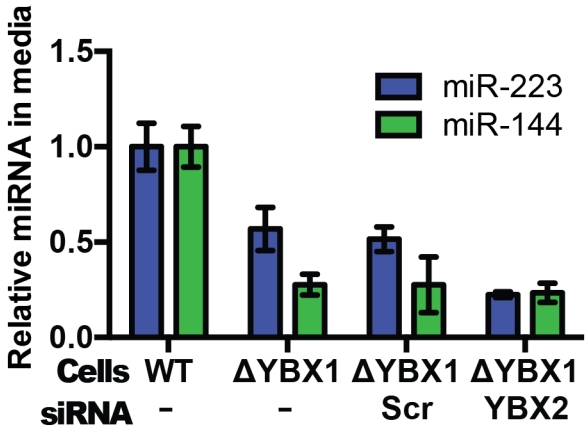


Fig. 1-9: Partial redundancy for YBX2 for the secretion of miR-223 in cells
Relative quantity of miR-223 secreted into the medium by WT and ΔYBX1 cells after 24 h with or without transfection with control or YBX2 siRNA.

Chapter 2: *In vitro* discovery of an exosomal RNA sorting motif

Introduction

An Analysis of enriched miRNAs from purified exosomes in Chapter 1 failed to identify specific motifs that direct miRNAs to be sorted into exosomes possibly because other RNA-binding proteins besides YBX1 could contribute to the complete exosome miRNA profile. Since miRNAs are likely too small to maintain significant secondary structure, I reasoned that a primary sequence motif is more likely than a structural motif and sought to develop an experimental method to identify primary sequence sorting motifs *de novo*. *In vitro* evolution experiments are often performed to determine the binding specificities of nucleic acid binding proteins using the SELEX (systematic evolution of ligands by exponential enrichment) approach. SELEX involves the mixing of purified protein with a large randomized pool of DNA or RNA, removing unbound nucleic acid, regenerating a pool of nucleic acid from the first round of selection and then repeating the mixing, removal of unbound nucleic acid sequentially for over several cycles[69, 70]. Traditionally SELEX was performed using many cycles of enrichment (~20-30 cycles) to enrich for very few sequences with high affinities. Recently, high throughput sequencing has been used to simplify the SELEX process by allowing identification of high affinity motifs after just a few cycles of enrichment[71, 72].

In this chapter I describe the elucidation of positive and negative exosome sorting motifs using an approach that combines the SELEX rationale with the cell-free packaging assay introduced in Chapter 1.

Materials and Methods

Generation of a randomized RNA library

A 25 nucleotide randomized sequence DNA template library was generated according to a previously published protocol [73]. Briefly, the library was synthesized by oligonucleotide extension by mixing primers SELEX-F (CGC GAA TTC TAA TAC GAC TCA CTA TAG GGG CCA CCA ACG ACA TT) and SELEX-library (TTA CAG CAA CCA CCG GGG ATC CAT GGG CACTAT TTATAT CAA CNN NNN NNN NNN NNN NNN NNN NNN NNA ATG TCG TTG GTG GCC C) using Klenow polymerase (New England Biolabs - NEB). The DNA template library was gel purified, diluted to 1.7 nmol/100µl and stored at -20°C. The first round RNA pool was generated by *in vitro* transcription in a reaction containing: 10,000U T7 RNA polymerase (Promega), 1 ml 10X T7 transcription buffer, 100 µl 100 mM ATP, 100 µl 100 mM GTP, 100 µl 100 mM CTP, 100 µl 100 mM UTP, 1.7 nmol DNA template library (in 100 µl) and 8.6 ml nuclease-free water. The reaction was incubated overnight in a 37°C water bath, treated with 1,000U DNAse I (NEB) for 1 hr at 37°C and gel purified over a 15% polyacrylamide TBE RNA gel (Invitrogen). RNA was diluted to 17 nmol/100 ul and stored at -80°C.

Exosome systematic evolution of ligands by exponential enrichment (Exo-SELEX)

Exo-SELEX process was performed in duplicate. Scaled up *in vitro* packaging reactions (see Chapter 1 for membrane and cytosol preparation) was used for the first cycle of Exo-SELEX. Each reaction consisted of 100 µl of cytosol (4 mg/ml), 150 µl of LiCl washed

membranes, 50 µl of 10X ATP regenerating system, 100 µl 5X incorporation buffer, 17 nM RNA library (in 100 µl), 200U RNAsin. The reaction was incubated at 30°C for 45 min. NEB buffer 3 (56 µl) of and 250U of RNase I_f (NEB) was added and incubated at 30°C for 30 min The membranes containing packaged RNA were then sedimented by centrifugation for 15 mins at 15,000Xg. RNA was extracted using the Zymo Research RNA Miniprep Kit and eluted in 30 µl. 10 µl of RNA was then reverse transcribed by mixing with 1 µl of 100 µM SELEX-R primer (CCC GAC ACC GCG GGA TCC ATG GGC ACT ATT TAT ATC AA) and 2 µl of 10 mM dNTP, incubating at 65°C for 5 min adding 4 µl 5X RT buffer (Invitrogen) 10U of RNAseout (Invitrogen) and 15U of AMV reverse transcriptase (Invitrogen). Reactions were incubated at 45°C for 1 hr and then 85°C for 5 min. cDNA (6 µl) was then used as template in a 100 µl PCR to regenerate the double stranded DNA template by mixing with 20 µl 5X GoTaq Green buffer (Promega), 1 µl 10 mM each dNTPs, 3 µl SELEX-F (10 µM), 3 µl SELEX-R (10 µM), 5U GoTaq polymerase (Promega) and 66 µl water. PCR was performed by incubating at 95°C for 5 min then 15 cycles at 95°C for 30 s, 50°C for 30s and 72°C for 30s and a final extension at 72°C for 10 min. BamHI (10U) was added and incubated for 1 hr and the DNA was purified using the Zymo Research DNA Clean and Concentrator - 5 Kit (eluted in 15 µl). A new RNA library for the next round of Exo-SELEX was then *in vitro* transcribed by mixing 10 µl of regenerated library, 4 µl of 5X T7 *in vitro* transcription buffer (Promega), 2 µl 10mM each rNTP, 40U RNAsin, 1 µl water and 50U T7 RNA polymerase (Promega) and incubated for 2 hr at 37°C. DNAse I (10U) was added and incubated for 15 min at 37°C to remove the DNA template. The new RNA library was purified using the Zymo Clean and Concentrator -5 kit (eluted in 30 ul). The subsequent rounds (2-4) of Exo-SELEX were performed as described for round 1 except that the reactions were scaled down 5-fold (100 µl total reactions).

Illumina sequencing library preparation

Regenerated libraries (100 ng) of following each round of Exo-SELEX were reverse transcribed as described above. The library was amplified using a forward primer to include the Illumina adaptor sequence (Adaptor-F: TTA CTA TGC CGC TGG TGG CTC TAG ATG TGA GAA AGG GAT GTG CTG CGA GAA GG CT AGA AGG GGC CAC CAA CGA CAT T) and reverse primers to introduce unique indexes for each library (Indexes 1-4: Replicate 1 Cycle 1-4 and Indexes 5,6,12 and 19: Replicate 2 Cycle 5-8):

Index1:GTTCGTCTTCTGCCGTATGCTCTAGCACTACACTGACCTCAAGTCTGCA
CACGAGAAGGCTAGATCCATGGGCACTATTTATATCAA

Index2:GTTCGTCTTCTGCCGTATGCTCTATGTAGCCACTGACCTCAAGTCTGCA
CACGAGAAGGCTAGATCCATGGGCACTATTTATATCAA

Index3:GTTCGTCTTCTGCCGTATGCTCTACGGATTCACTGACCTCAAGTCTGCA
CACGAGAAGGCTAGATCCATGGGCACTATTTATATCAA

Index4:GTTCGTCTTCTGCCGTATGCTCTAACCAGTCACTGACCTCAAGTCTGCA
CACGAGAAGGCTAGATCCATGGGCACTATTTATATCAA

Index5:GTTCGTCTTCTGCCGTATGCTCTAGTGACACACTGACCTCAAGTCTGCA
CACGAGAAGGCTAGATCCATGGGCACTATTTATATCAA

Index6:GTTCGTCTTCTGCCGTATGCTCTATAACCGCACTGACCTCAAGTCTGCA
CACGAGAAGGCTAGATCCATGGGCACTATTTATATCAA

Index12:GTTCGTCTTCTGCCGTATGCTCTAATGTTCCACTGACCTCAAGTCTGC
ACACGAGAAGGCTAGATCCATGGGCACTATTTATATCAA

Index19:GTTCGTCTTCTGCCGTATGCTCTAAAAGTGCACTGACCTCAAGTCTGC
ACACGAGAAGGCTAGATCCATGGGCACTATTTATATCAA

PCR was performed over 12 cycles (using the thermocycling conditions described above) by mixing 20 µl cDNA, 10 µl HF buffer (Thermo Fisher Scientific), 1 µl 10 µM Adaptor-F, 1 µl 10 µM Index primer, 1U Phusion DNA polymerase, 1 µl 10 mM each dNTPs and 16.5 µl water. The resulting library was gel purified and submitted for sequencing to the Vincent J. Coates High Throughput Sequencing Facility at UC-Berkeley using the HiSeq-2000 platform with a custom Read 1 primer (ATC TAG GGG CCA CCA ACG ACA TT).

Sequence analysis

Preprocessing of the 50 base pair single reads was filtered for read quality (read quality >20 and percent bases in sequence that must have quality >90) using the FASTX toolkit as described in Chapter 1 and sequences were clipped and trimmed using the Python package cutadapt 1.10 [74]. Kmer analysis was performed using an in house script written by David Melville, Ph. D.. Motif analysis of Kmers was performed using Multiple Em for Motif Elucidation (MEME) using a zero or one occurrence per sequence model [49].

Cell-Free packaging reactions

In vitro packaging reactions were performed as described in Chapter 1 with the exception that post-nuclear supernatant (PNS) was used rather than isolating membranes and cytosol separately from PNS and then mixing them back together. Complete reactions (40 µl) consisted of 15 µl PNS (4 mg/ml), 4 µl 10X ATP regenerating system, 8 µl 5X incorporation buffer, 1 µl 10 µM synthetic RNA and 40U RNAsin (Promega). Reactions were incubated for 20 min at 30ºC then placed on ice and mixed with 4.3 µl of 10X NEB buffer 3 and 50U of RNase I_f was added to all reactions except to a no RNase control. Reactions were then incubated at 30ºC for a further 20 min. RNA was extracted, reverse transcribed and quantified by qRT-PCR using the stem-loop TaqMan approach[75]. Commercial primers and probes were used for hsa-miR-223-3' and miR-190a-5' (Life Technologies) and custom primers and probes (Fam/BHQ) were used for mutants (Integrated DNA Technologies) (miR-223mut RT: GTC GTA TCC AGT GCA GGG TCC GAG GTA TTC GCA CTG GAT ACG ACT CCA AC, F-primer: CAC ACG CAT GTC AGT TTG TCA A, Probe: TCG CAC TGG ATA CGA CTC CAA CA and miR-190mut RT: GTC GTA TCC AGT GCA GGG TCC GAG GTA TTC GCA CTG GAT ACG ACT GGG TA, F-primer: CTC ACA CGC ATG ATA TGT TTG ATA, Probe: TCG CAC TGG ATA CGA CTG GGT AT, Universal-R: CCA GTG CAG GGT CCG AGG TA).

Results

The Exo-SELEX strategy to elucidate exosomal RNA sorting motifs

Using the cell-free packaging assay as a selection method, I devised a SELEX-like strategy, termed Exo-SELEX, to uncover specific primary sequence exosome sorting motifs (Fig. 2-1). Using a DNA template library with a 25 nucleotide randomized base sequence flanked the T7 RNA polymerase promoter sequence and a constant region for reverse transcription I generated a sizeable RNA pool which, theoretically, contained 10 copies of all possible 25 nucleotide sequences. This RNA pool was then used as the input for a large-scale *in vitro* packaging reaction that included cytosol, membranes and ATP. After incubation at room temperature, the reaction was treated with ribonuclease (RNAse I) to remove unpackaged RNA and the selected RNA was purified, reverse transcribed, amplified by PCR and then a new input library was generated using T7 RNA polymerase. The selection was performed over 4 cycles. Following each cycle, a sample of the RNA output was saved for high-throughput libary preparation and Illumina sequencing.

Exo-SELEX enriches for specific sequences

RNA following each round of selection (4 cycles) was sequenced and two separate Exo-SELEX experiments were performed for a total of 8 libraries. After adapter clipping and trimming the total reads between cycles was similar for each replicate (Rep. 1 cycle (R1C1) = 24M, R1C2 = 23M, R1C3 = 29M, R1C4 = 24M, R2C1 = 20M, R2C2 = 20M, R2C3 = 16M and R2C4 = 17M). As very few duplicate 25 nucleotide sequence reads were detected in all libraries, even after 4 rounds of selection, I analyzed the libraries to determine if substrings within the randomized 25 base libraries were enriched over the selection period. To facilitate this search, David Melville wrote a simple program to search all reads within a library for sub-sequences (Kmers) of varying length (K). Through trial and error K = 8 was chosen as the sliding window by which to search for sequence motifs as this length was likely long enough to cover a primary sequence motif but short enough so that the number of possible Kmers (4^8) was limited such that the program could be run on a desktop computer over a reasonable time period. As expected, most Kmers were selected against during the 4-cycle selection process while a small subset of Kmers became enriched (Fig. 2-2A). Of the 65,536 possible 8mers, 677 became enriched beyond 3 standard deviations of the mean and 164 of these were progressively enriched after each cycle (4>3>2>1) (Fig. 2-2B). These results indicated that Exo-SELEX enriches for short sequences over the course of selection.

Exo-SELEX identifies positive and negative exosome sorting motifs

I next sought to determine if specific sequence motifs became enriched over the selection process. Visual inspection of the most enriched Kmers in both replicates suggested that cytosine-rich sequences became specifically enriched (Fig. 2-3A,B). The top 100 most enriched Kmers from replicate 1 were further analyzed to identify specific sorting motifs using multiple Em for motif elucidation (MEME). This analysis identified a single, highly statistically significant ($E = 1E^{-55}$) cytosine-rich motif (A/UCCC) (Fig. 2-3C).

I next analyzed the datasets to identify Kmers that were highly selected against to see if a specific motif is associated with exclusion from exosomes, likely serving as a cellular retention signal. The top 20 most depleted Kmers appeared to be highly enriched in guanosine bases (Fig. 2-4A,B). Indeed, MEME analysis of the 100 most depleted Kmers

identified a highly significant ($E = 1E^{-92}$) guanosine-rich motif (GUUGG) (Fig. 2-4C). Exo-SELEX therefore uncovered two putative primary sequence sorting motifs, A/UCCC and GUUGG, the former acting as a positive packaging signal and the latter as an exclusion or cellular retention signal.

The cytosine-rich motif is necessary and sufficient for miRNA sorting into exosomes *in vitro*

In Chapter 1 I established that the *in vitro* packaging reaction reconstitutes the sorting of an exosomal miRNA (miR-223) but does not support the recruitment of a miRNA (miR-190a) that is not secreted in exosomes. I next checked to see if the motifs identified in the Exo-SELEX experiments were present in these miRNAs. Corroborating the Exo-SELEX results, miR-223 contains an ACCC motif near the 3' end and miR-190a contains a motif similar to GUUGG (UUGG) motif also at the 3' end. The sequences of the two miRNAs 5' of the motifs are highly similar, which combined with the Exo-SELEX results, strongly suggested that these motifs are responsible for the difference in packaging efficiency. To test this directly, I mutated the ACCCC sequence in miR-223 to GUUGG and the UUGGU in miR-190a to ACCCA (Fig. 2-5A). The ACCCC --> GUUGG mutation in miR-223 abolished packaging while the UUGGU --> ACCCA mutation rendered miR-190a competent for packaging. Taken together, these results indicated that the ACCC motif is both necessary and sufficient for miRNA packaging *in vitro* and that guanine-rich sequences likely act as cellular retention signals.

Discussion

By employing the cell-free reaction to select from a large randomized library of sequences I identified motifs that act as positive and negative RNA motifs for sorting into exosomes *in vitro*. A caveat to these *in vitro* studies is that they require validation in vivo. This could be done by transfecting plasmids to express transcripts containing each motif and measuring their comparative abundance in cells and exosomes. Alternatively, the sequences used here (miR-223 and miR-190a and their mutant versions) can be transfected and their release in exosomes compared.

In Chapter 1 I identified YBX1 as a required sorting factor for miR-223 *in vitro* and in vivo. Two previous studies have investigated the binding preferences of YBX1. The first used the SELEX approach with purified protein to identify CAUC and CACC as binding motifs[30]. The second used an assay based on the prevention of intron excision when an RNA binding protein is bound to identify CACCA as a binding sequence for YBX1 in vivo[29]. As the motif identified here (A/UCCC) is highly similar to the motifs previously identified, my results corroborate the prior studies and provide further evidence for YBX1 as the primary exosomal RNA sorting factor in HEK293T cells.

A previous study in T-cells identified a short tetranucleotide motif (GGAG) that is enriched in exosomes and bound by the the sumoylated form of the RNA-binding protein hnRNPA2B1[47]. According to the results described in this chapter, GGAG would like be a very poor packaging signal *in vitro*. This apparent contradiction can be explained by the likelihood that different cell types might use distinct RNA-binding proteins with different binding specificities to modulate their exosomal RNA profile. The Exo-SELEX

approach described here for HEK293T cells could be broadly applied to experimentally determine sorting motifs for various cell types.

Figures

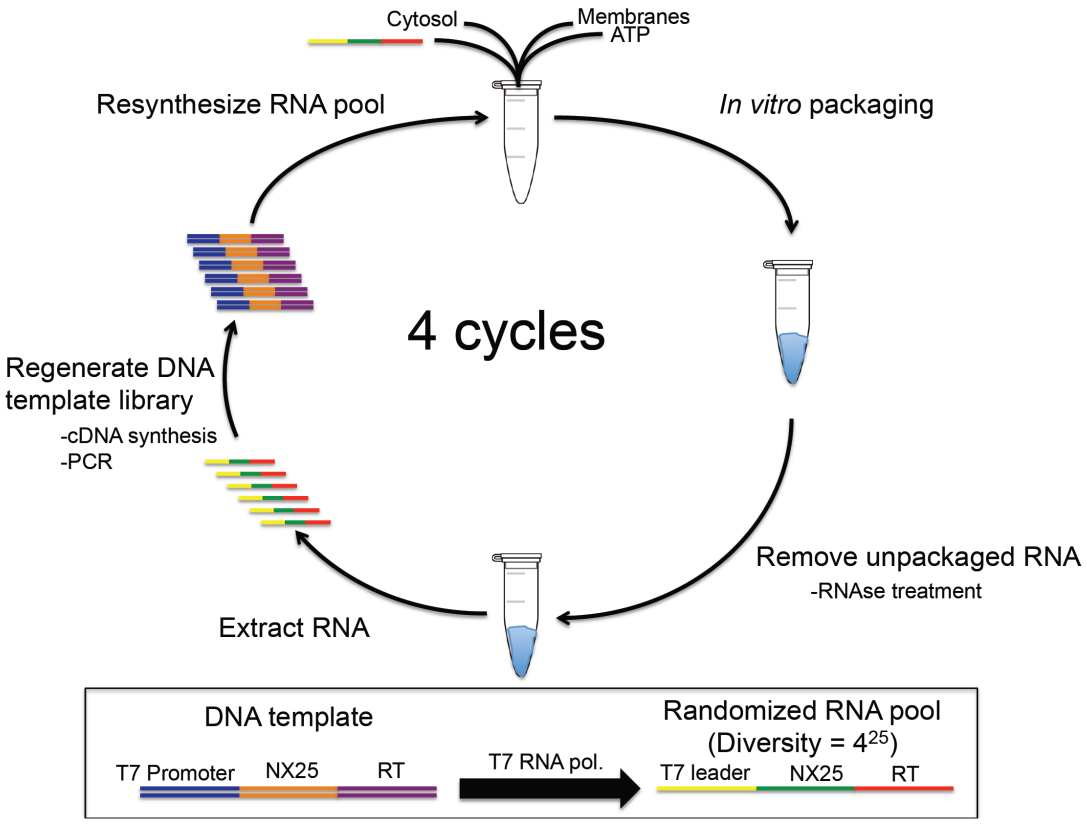


Fig. 2-1: Exo-SELEX schematic
Each step in the exosome selective evolution of ligands by exponential enrichment (Exo-SELEX) procedure depicted graphically. (Inset) Features of the DNA *in vitro* transcription template library and resulting RNA pool. The T7 promoter/leader, 25 nucleotide randomized sequence (NX25) and constant sequence used for reverse transcription to regenerate the library after each round of selection (RT) are indicated.

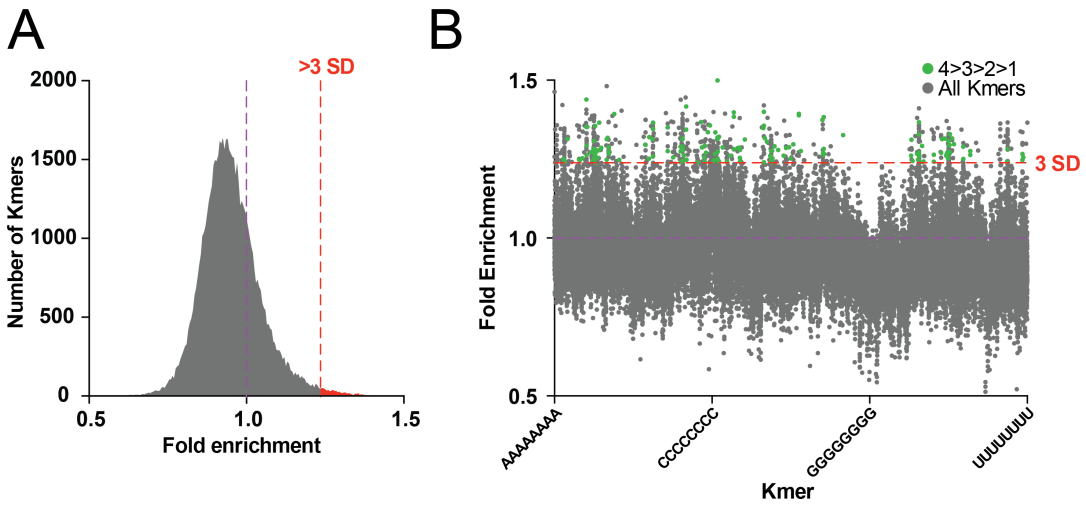


Fig. 2-2: Enrichment of Kmer sequences following Exo-SELEX
(A) Histogram showing the enrichment distribution of all Kmers (K=8) after 4 cycles of Exo-SELEX. The violet line indicates no enrichment and the red line and shading denotes the enrichment beyond three standard deviations (3 SD) of the mean. (B) The fold enrichment of all Kmers (K=8) after 4 cycles of Exo-SELEX (grey circles) and the Kmers > 3 SD (red line) that demonstrated enrichment after each of the 4 cycles compared to the previous cycle (green circles).

A

Replicate 1

Sequence	Enrichment
CCGACACC	1.499
ACACCCCU	1.439
CACUCCCC	1.416
CUACCCCU	1.400
CGACACCC	1.398
GACACCCC	1.393
AGAACCCC	1.393
GACACCGC	1.389
CGACACCG	1.384
GCACUCCC	1.383
GCACACCC	1.374
CAUUCCCC	1.368
UACCCCAC	1.367
AUCCCGAC	1.366
ACGACACC	1.365
CCCACACC	1.357
CUACCCCG	1.356
ACCGACAC	1.355
ACACUCCC	1.352
ACACCCCG	1.351

B

Replicate 2

Sequence	Enrichment
AUCCCCAC	1.181
ACACCCAC	1.174
ACACACCC	1.171
CACCCACA	1.169
ACCCACAC	1.168
CCCACCAC	1.167
UCCCCACA	1.165
CCACACCA	1.163
CACAUCCC	1.163
CCCGCCAA	1.163
CCACAUAC	1.162
ACAUACCC	1.162
ACCACACC	1.161
CACACCCA	1.161
CCACAACC	1.161
CACAACCC	1.161
CCCACAUC	1.160
CACCACAC	1.160
CAUUUCCC	1.159
CGCCCACA	1.156

C

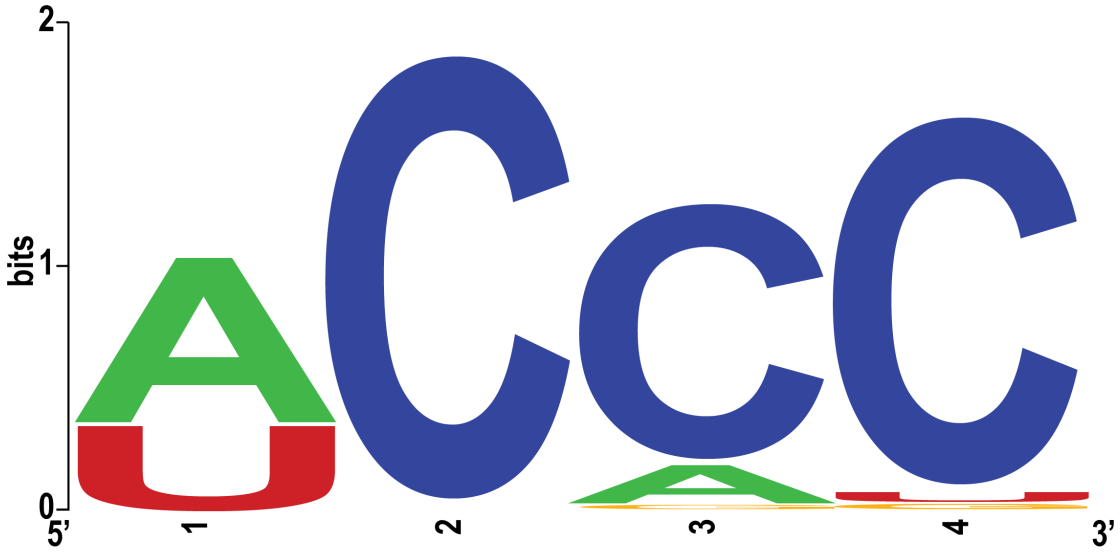


Fig. 2-3: Exo-SELEX reveals a cytosine-rich exosome sorting motif
(A,B) The top 20 most enriched Kmers (K=8) from each of two replicate Exo-SELEX experiments. Kmers were curated for those that were enriched after each cycle compared

to the previous cycle (4>3>2>1). (C) Exosomal RNA sorting motif computed using the top 100 most enriched Kmers (4>3>2>1) using MEME (multiple Em for motif elucidation) under standard conditions ($E_{value}=1E^{-55}$).

A

Replicate 1

Sequence	Enrichment
GUUGGGGC	0.621
CGUGGGUG	0.622
CGUGGUGG	0.623
GGGCCCCA	0.627
CGUUGGGG	0.637
GUUGGGGU	0.642
GUUGGAGG	0.649
UCGUUGGG	0.650
GUUGGGCC	0.653
GGCCCCGG	0.660
GUCGUUGG	0.661
UUGAUGGU	0.666
GUUUGGUG	0.667
GGUGGGGG	0.669
UAGGUGGG	0.670
UUGGUGUC	0.670
GUGGGUGC	0.671
UGUUGGUU	0.674
CGUUGGUC	0.675
CUGGGGGC	0.676

B

Replicate 2

Sequence	Enrichment
GGGGGGGG	0.598
GUCGUGGG	0.623
GGGGCCCC	0.629
GGUGGUCC	0.634
UGCGGUGG	0.650
AUUGGUGG	0.653
UGGGCCCC	0.660
UUGGUGGG	0.661
AUGGCCCC	0.665
UUGGUGGU	0.665
AAGGGGGG	0.667
GGGCCCCA	0.668
UGGUGGUC	0.669
UGGGUGGC	0.669
GGGGGGGU	0.672
UUGGUGGC	0.673
UGAGGUGG	0.677
CGGGGGGG	0.681
GGGCCCCG	0.682
GGUGGGGG	0.682

C

Fig. 2-4: Guanine-rich sequences are negative exosome sorting motifs
(A,B) The 20 most depleted Kmers (K=8) from each of two replicate Exo-SELEX experiments. Kmers were curated for those that were enriched after each cycle compared to the previous cycle (4>3>2>1). (C) Negative exosomal RNA sorting motif computed

using the 100 most depleted Kmers (4>3>2>1) using MEME (multiple Em for motif elucidation) under standard conditions ($E_{value}=1E^{-92}$).

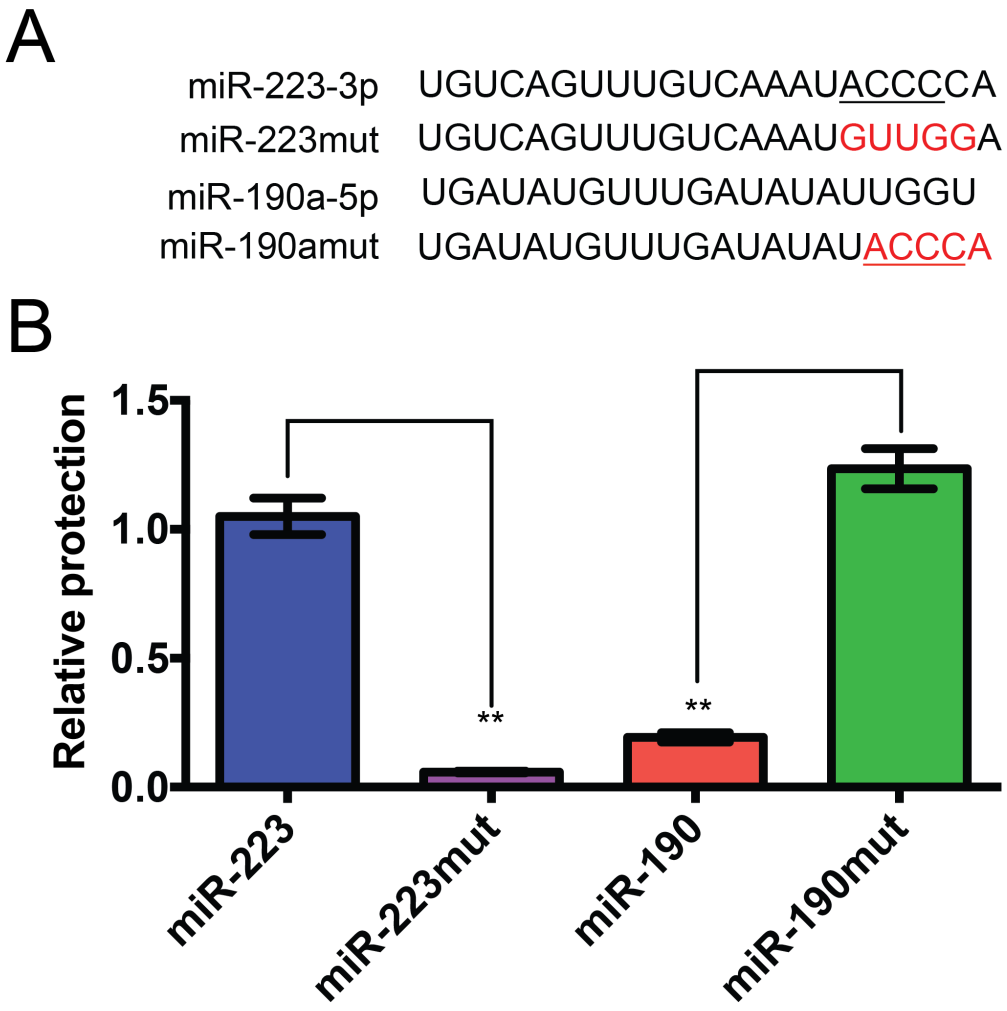


Fig. 2-5: Necessity and sufficiency of the Exo-SELEX motif in packaging miRNA into exosomes *in vitro*

(A) RNA oligonucleotides corresponding to miR-223, miR-190 and versions with mutated Exo-SELEX motif (miR-223mut) or mutation to introduce the motif (miR-190mut). The Exo-SELEX motifs are underlined and mutations are indicated in red. (B) Relative protection (standardized to miR-223) in the *in vitro* packaging assay for each of the RNA sequences depicted in (A). Error bars indicate standard deviations from the means of 3 independent experiments. Statistical significance was tested using Student's t-test (** $p<0.01$)

Chapter 3: A broad role for Y-box Protein I in defining the small non-coding RNA content of exosomes

Introduction

Chapters 1 and 2 employed cell-free biochemical reactions to identify a protein involved in sorting miRNAs into exosomes (Chapter 1) and to identify positive and negative RNA sorting motifs. Chapter 1 showed that miR-223 is secreted via YBX1 *in vitro* and confirmed the role of YBX1 in vivo, however to comprehensively define the YBX1-dependent exosomal RNA I decided to use a high throughput sequencing approach to characterize the RNA composition of wild-type and YBX1-null cells.

Metazoan cells grown in culture release extracellular vesicles (EVs) into the surrounding medium and free vesicles can be found in all bodily fluids [3]. EVs can be categorized into multiple different classes based on their size, shape and presumed membrane origin. Exosomes are defined as ~30-100 nm vesicles that originate from the multivesicular body and contain late endosome markers [10], but there is evidence that biochemically indistinguishable vesicles can bud directly from the plasma membrane [9]. Microvesicles or shedding vesicles are generally larger (>200 nm), more variable in shape and density and likely originate from the plasma membrane [3, 6]. EVs contain various molecular cargoes, including proteins, lipids and RNAs, but how these cargos are sorted into EVs remains unclear.

Since the initial description of RNA in EVs [7] and the identification vesicle-associated circulating transcripts in plasma [5], there has been widespread interest in extracellular RNA. Seminal EV RNA studies showed the presence of ribonuclease (RNase)-protected mRNA and microRNA (miRNA) species in blood that might be exploited as biomarkers and suggested a role for EVs in the horizontal transfer of genetic information between cells [3, 4, 6, 7, 76]. Several studies implicated miRNAs, in particular, as a mode of intercellular communication between cells based on reporter-based assays for miRNA function in recipient cells [8, 37, 64]. Recent studies using sensitive Cre recombinase-based genetic systems in mice have indicated that local and long-range RNA transfer via EVs can occur *in vivo* [77, 78]. High-throughput sequencing of extracellular RNA has identified many other exosomal RNA biotypes, including various small non-coding RNAs (snoRNA, tRNA, Y-RNA, vault RNA) [79-82], long non-coding RNA (lncRNA)[80, 83] and protein-coding transcripts [80, 82]. The recent application of thermostable group II intron reverse transcriptases (TGIRTs) to sequence low input RNA from human plasma samples revealed that many circulating small ncRNA transcripts are full-length and thus apparently resistant to serum ribonucleases [84]. In addition to vesicles, extracellular RNA can be found associated with circulating ribonucleoproteins, which co-sediment with vesicles during standard EV isolation protocols that culminate in high-speed (>100,000Xg) ultracentifugation [4]. The co-isolation of RNPs and multiple EV sub-populations has made if difficult to identify which RNA biotypes are exported via which pathway and through which mechanisms RNA transfer between cells might occur.

Understanding the mechanisms by which transcripts are sorted into EVs has proven challenging. A sumoylated form of the RNA binding protein hnRNPA2B1 was shown to bind miRNAs containing a tetranucleotide motif (GGAG) that are exported from T-cells [47]. In Chapter 1, I used a cell-free reconstitution approach to identify a requirement for the RNA-binding protein YBX1 in the specific sorting of a miRNA (miR-223) into HEK293 exosomes [85]. However, whether these RNA-binding proteins play a broad role in defining the RNA composition of exosomes or package only a select few transcripts has not been addressed. Indeed, to date there has been no description of the role that RNA-binding proteins play in broadly determining the RNA content of EVs.

In the present study, collaborators in the laboratory of Alan Lambowitz at the University of Texas and I sought to determine the EV-RNA composition in HEK293 conditioned media by combining nuclease protection assays and biochemical fractionation methods to purify vesicles from RNPs with TGIRT-seq, which makes it possible to obtain full-length reads of highly structured small ncRNAs. We also sought to comprehensively define the transcripts that utilize YBX1-dependent and -independent sorting pathways for cellular export.

Materials and Methods

Cell lines and growth conditions

HEK293T and YBX1-null cells were cultured in DMEM with 10% FBS (Thermo Fisher Scientific, Waltham, MA). For exosome production, cells were seeded to ~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 exosome 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).

Extracellular vesicle and exosome purification

Conditioned medium (3 l for small RNA-seq and 420 ml for all other experiments) 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 4ºC. 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 >50%. Sucrose buffers (40%, 20% and 0% - 1 ml each) were sequentially overlaid and the tubes

were ultracentrifuged at ~150,000Xg (38,500 RPM) for 16 h in an SW-55 rotor. The 20/40% interface was collected and either subjected to nuclease protection assays or for WT and YBX1-null exosomes comparisons diluted 1:5 with phosphate buffered saline (pH 7.4) 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 4ºC. 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 4ºC. 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).

Nuclease-protection experiments
Post-flotation EV fractions from the 20/40% sucrose gradient interface (100 µl) were mixed with Triton X-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 mins. Proteinase K was added to a final concentration of 20 µg/ml to some reactions and incubated on ice for 30 min and inactivated by the addition of PMSF (5 mM). RNase I_f (40U) and 11 ul of NEB Buffer 3 was added and incubated at 30º C for 20 min. Alternatively, 2U Turbo DNase (Thermo Fisher Scientific) and 11 µl of Turbo DNase buffer was added and 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 extraction kit.

Library preparation using thermostable group II intron reverse transcriptase (TGIRT-seq)
Library preparation and sequencing was performed according to previously published protocols by collaborator Yidan Qin, Ph.D. in the laboratory of Alan Lambowitz at the University of Texas - Austin[84, 86].

Sequencing and bioinformatic analysis
Bioinformatic analysis was performed by Jun Yao, Ph.D. in the laboratory of Alan Lambowitz. Briefly, RNA-Seq libraries prepared from ribo-depleted whole cell RNA (fragmented then phosphatase treated or unfragmented) and purified exosome RNA using GsI-IIC RT and sequenced on an Illumina NextSeq instrument to obtain 150-nt single-end reads. The reads were trimmed to remove adaptor sequences and low quality base calls (sequencing quality cut of 20 ($p < 0.01$)), and reads <18 nt after trimming were discarded. Trimmed reads were then mapped by using Tophat and Bowtie2 to a human genome reference sequence (Ensembl GRCh38) modified to include additional rRNA repeats. tRNA reads were retrieved from the initial mapping results and remapped to human tRNA reference using Bowtie2.

Results

Most extracellular vesicle-associated transcripts are encapsulated within EVs
Extracellular vesicles encapsulate RNAs rendering the transcripts resistant to degradation by exogenous ribonucleases. However, prior studies have primarily focused on the RNase protection of individual transcripts rather than the global RNA content. To broadly define the RNA composition of EVs, we employed TGIRT-seq in combination with RNase

protection assays in the presence or absence of a detergent to discriminate transcripts inside or outside of membrane vesicles. EVs were isolated from HEK293T cells by differential centrifugation followed by flotation gradient ultracentrifugation into a discontinuous sucrose gradient to separate vesicles from ribonucleoprotein particles (Fig. 3-1A). After flotation, vesicles were either left untreated or treated with protease (10 ug/ml proteinase K) or non-ionic detergent (1% Triton X-100) in the presence or absence of RNase (RNase I). Small RNA Bioanalyzer analysis of the resulting RNA after each condition showed the predominance of a large ~60-70 bp peak that was resistant to RNase in all treatment conditions except in the presence of detergent (Fig. 3-1B), suggesting that this predominant EV-RNA size class is contained within EVs. These results corroborate many previous studies suggesting that the predominant species in EVs, according to fragment analysis, is ~60-70 nts. However, previously published small RNA-seq datasets failed to identify the high abundance transcripts corresponding to this size distribution, indicating that conventional RNA-seq methodology does not efficiently capture this species.

Since the retroviral reverse transcriptases (RTs) used routinely to generate RNA-seq libraries are unable to include species with significant secondary structure and are impeded by specific post-transcriptional modifications, we utilized a thermostable group II intron reverse transcriptase (TGIRT-III RT) to generate cellular and EV libraries for high-throughput sequencing (TGIRT-seq) [84, 86]. RNase treatment of fractions that were not treated with protease or detergent had very little effect on global transcript relative abundances (Fig. 3-1C). In contrast, upon the addition of detergent, the relative abundance of most transcripts was substantially decreased by RNase treatment (Fig. 3-1D). Furthermore, scatterplots comparing the pairwise analysis of relative transcript abundances for each dataset showed little change under most treatment conditions (Pearson's coefficient (r) > 0.98), except when detergent was present (r = .76-78) and especially when detergent was added in conjunction with RNase (r = 0.56-0.59) (Fig. 3-2). The decreased correlation between untreated and detergent treated is consistent with the presence of endogenous RNases in the EV fractions.

To assess DNA contamination in the EV-RNA isolation, the same treatment conditions used for RNase protection were replicated for DNase protection. As expected, there was no effect of DNase treatment on the relative abundance of transcripts under all treatment conditions (Pearson's coefficient > 0.85), indicating that DNA contamination does not confound the RNA-seq results (Fig. 3-3). These results demonstrate that the vast majority of transcripts recovered from a purification employing a flotation gradient ultracentrifugation step are encapsulated within extracellular vesicles. This flotation step is sufficient to separate vesicles from RNPs and transcripts bound to the outside of vesicles or attributable to DNA are not major contributors to total mapped reads.

RNase sensitivity categorizes abundant transcripts from the most represented RNA classes into distinct categories

TGIRT-seq revealed that the most abundant RNA classes in EVs are: small non-coding RNA, protein coding transcripts, ribosomal RNA and long non-coding RNAs (lncRNA), including antisense RNA and long intergenic RNAs (lincRNA) (Fig. 3-4A). The most abundant RNA species in exosomes were small non-coding RNAs, defined as non-coding

RNA < 200 nucleotides (74%). Among, small non-coding RNA, the most abundant species was tRNAs (79%) followed by Y-RNA (12%) and 5S and 5.8S rRNA (3.9%). The remaining sncRNA species constituted <6% of the total reads (Fig 3-4B). Surprisingly, TGIRT-seq revealed very low relative levels of miRNAs in exosomes (<0.1%) compared to highly abundant small non-coding RNA transcripts 5.8S rRNA, tRNA, Y-RNA (RNY3, RNY1 and RNY4) and Vault RNAs. Due to secondary structure and post-transcriptional modifications, these transcripts may have been underestimated in previous exosome small RNA-seq libraries generated by retroviral RTs, thus causing miRNAs to be over-represented. In addition, previous studies often used small RNA extraction methods and/or size selection that enriches for miRNAs.

Sequencing reads from EV-associated RNA could be broadly assigned to three categories based on their nuclease sensitivity. The large majority (80%) of reads came from transcripts that were protected from RNase unless detergent was added with RNase (Fig. 3-4C - Class I). This class included the most abundant transcripts - 5.8S rRNA, tRNAs and Y-RNAs (Fig. 3-4D), suggesting that these transcripts are enclosed within EVs. This class also included abundant protein coding sequencing mapping reads and most abundant long non-coding RNA mapped reads. Reads mapping to some transcripts (8.8%) increased in relative abundance only after the detergent and RNase treatment, suggesting that they remain RNase resistant when the majority of other transcripts are degraded in this condition (Fig. 3-4C - Class II). These reads mapped primarily to non-coding transcripts that are bound within ribonucleoprotein complexes including: the 7SL RNA component of the signal recognition particle, the 7SK RNA associated with the transcription elongation factor P-TEFb, the U2 RNA component of the spliceosome, the RNase-P associated RNA. These results are consistent with select transcripts being sorted into EVs along with these or other protein partners. Finally, some transcripts appeared to be resistant to RNase in all conditions (11%), these reads were attributed to the 18S, 28S and 5S rRNA (Fig. 3-4C - Class III). Though it is not possible to conclusively determine from these results the relationship of Class III transcripts with EVs, it is possible that these transcripts reflect ribosomal RNA contaminants in the EV preparations.

To further classify reads according to nuclease sensitivity and to quantitatively evaluate the contribution of contaminating DNA to the sequencing reads, we next normalized relative reads abundances between treatment conditions using references that should not be degraded by DNase or RNase (Fig. 3-5). Reads mapping to tRNAs were used as a reference for normalization in the DNase treatment experiments. Since there were no reads that we could confidently attribute to DNA de novo for normalization, we first analyzed the DNase datasets to identify reads that were decreased >90% after DNase treatment and then used these for normalization in the RNAse protection treatments. After normalization, the percentage of reads mapping to each biotype was analyzed to determine the relative contribution of reads that were lost during DNAse + detergent (DNA), RNase + detergent (RNA) and remained protected under either condition (Protected). We found that the majority of reads for most biotypes could be attributed to either RNA or were protected. Some biotypes showed evidence of DNA contamination, though most of these biotypes are of very low abundance in EVs. However, protein-coding reads represented ~12% of the total reads (Fig. 3-2A) and ~ 18% of these reads were attributable to DNA. The DNA contamination of protein coding sequence reads was

mostly attributed to reads mapping to non-coding regions of protein coding genes (introns and 5' UTR). Additional categories also showed some evidence of DNA contamination, reads mapping to the antisense of protein coding genes (AS_P) (44%), pseudogenes (50%), and lincRNA (16%). Overall, this analysis showed very low levels of reads attributed to DNA in the total library (~1-2%) and in combination with the correlation analysis (Fig. 3-3) we conclude that DNA is not a major source of TGIRT-seq reads in EV libraries and most transcript biotypes are not affected by DNA contamination.

YBX1-dependent exosomal RNA transcripts

I previously identified YBX1 as an RNA-binding protein that directs selective miRNA sorting into exosomes (Chapter 1). We therefore sought to more comprehensively define the YBX1-dependent exosomal RNA composition by performing TGIRT-seq on purified exosomes from wild-type and YBX1-null cells. Since cells release multiple populations of extracellular vesicles, we focused only on a sub-population of CD63-positive EV that we term exosomes (Fig. 3-6A). HEK293 exosomes contain some miRNAs that were previously shown to be packaged in a YBX1-dependent manner. We found very little difference in the RNA composition of wild-type and YBX1-null cells (Fig. 3-6B,C). In contrast, YBX1-null exosomes had fewer reads mapping to some small non-coding RNA classes. Most dramatically affected in the YBX1-null exosomes were tRNAs which accounted for 47.2% of total mapped reads in wild-type exosomes, but only 10.5% in YBX1-null exosomes (Fig. 3-6B).

Scatter plots comparing normalized mapped reads for WT and YBX1-null cells did not show dramatic changes in the relative abundance of major exosome RNA biotypes, except for an increase in moderately expressed tRNAs (Fig. 3-6C). In contrast, YBX1-null exosomes are substantially decreased for the large majority of tRNAs (Fig. 3-6D). Thus, the substantial decrease of total reads mapping to tRNAs (Fig. 3-6B) was not a result of a small group of several highly expressed tRNA genes, but rather, suggests a broad role for YBX1 in packaging tRNAs. In addition to tRNAs (78.8% decreased), other abundant species of small non-coding RNAs (Y-RNAs (41% decreased) and Vault RNAs (89% decreased) are also affected in YBX1-null exosomes (Fig. 3-6E). For tRNAs, vault RNA and Y-RNA, YBX1-null cells showed a slight increase in reads mapping to these transcripts (Figure 3-6E), suggesting that blocking export via the exosome pathway can result in accumulation of these transcripts in cells.

Full-length tRNA sorting into exosomes is YBX1-dependent

Previous reports have identified tRNAs in exosomes, however, these have been attributed to tRNA fragments rather than full length tRNAs. We therefore explored which tRNAs species are most abundant in exosomes and secreted in a YBX1-dependent manner. In wild type exosomes, the majority of tRNA reads mapped across the locus and were ~72 bases, suggesting that primarily full-length tRNAs are packaged into HEK293T exosomes under standard growth conditions (Fig. 3-7A,E). Full-length tRNAs reads were decreased by ~80% in YBX1-null exosomes (Fig. 3-7A). Nearly all tRNA reads terminated at the 3' end of the gene (~72-75 bases) (Figure 3-7B). Most of the remaining smaller peaks identified in the read length (Figure 3-7A), represent 5' truncations of full-length tRNAs.

Analysis of tRNA fragments by tRNA read start site distribution identified several minor tRNA fragments that were sorted into exosomes, albeit at lower abundance than full-length tRNAs. As for full length tRNAs, most tRNAs fragments were also decreased in the YBX1-null exosomes to a similar magnitude (~80% decreased). However, one fragment which represented an ~16 base 5' truncated tRNA fragment was present at equal relative abundance in both WT and YBX1-null exosomes (Fig. 3-7C). This truncation represents an apparent D-loop cleavage fragment or post-transcritional modification, which to our knowledge, has not been described before (Fig. 3-7C inset and 3-7E).

As suggested by the scatter plot (Fig. 3-6D) the depletion of tRNAs seen in the YBX1-null exosomes was not simply a result of several highly abundant tRNAs being diminished. Indeed, reads mapping to nearly all tRNA genes grouped by anticodon were decreased by 50% or more with most decreased by 70-90% (mean proportion KO/WT = 0.25) (Fig. 3-7D). In contrast, tRNA levels in cells were slightly increased (mean proportion KO/WT = 1.3) (Figure 3-7D), suggesting that blocking YBX1-dependent secretion of tRNAs results in tRNA accumulation in cells.

YBX1-independent exosome transcripts

The total RNA levels of wild-type and YBX1-null exosomes were equivalent (3.3 μg and 3.4 μg respectively from 420 ml of conditioned media) and an approximately equal number of reads were obtained from each library. This suggested that some RNAs are secreted independently of YBX1 and these transcripts are represented in the datasets as those that fall along the dotted diagonal line in the scatterplots. In general, the total number of reads mapping to long transcripts (as opposed to short non-coding RNA transcripts) were not affected in the YBX1-null exosomes and were in fact slightly increased in relative abundance compared to wild type exosomes due to the decrease in small non-coding RNAs (Fig. 3-6C).

YBX1-independent classes included protein-coding and long non-coding RNA transcripts (both antisense and lincRNA). When compared the relative abundance of individual transcripts, the vast majority of these transcripts are slightly increased in relative abundance in YBX1-null exosomes, however, a subset of transcripts are decreased (Fig. 3-8A and Fig. 3-9). TGIRT-seq was performed on unfragmented RNA from cells and exosomes as well as fragmented RNA from cells. As expected, coding sequences for the fragmented libraries were higher in the fragmented cellular library compared to the unfragmented library (Figure 3-8B). When comparing protein-coding reads from unfragmented cellular and exosome libraries, exosomes are highly enriched for intronic reads (86.0% and 86.6% for WT and YBX1-null exosomes respectively compared to 43.6% and 40.4% for WT and YBX1-null cells) and have substantially fewer reads mapping to untranslated regions compared to cells (9.0% and 8.5% for WT and YBX1-null exosomes respectively compared to 52.2% and 55.4% for WT and YBX1-null cells) (Fig. 3-8B).

Discussion

Our results indicate that many small non-coding RNA biotypes are secreted in a YBX1-dependent manner including abundant, mostly full-length, tRNAs and Y-RNAs. Many of these RNAs are highly structured, suggesting that YBX1 may recognize the secondary structure of these transcripts. Lin28, another cold shock domain containing RNA binding

protein, binds the hairpin loop structure of Let7 family pre-miRNAs to prevent cleavage by Dicer to form the mature species[53]. Recently, YBX1 was shown to play an analogous role in the maturation of miR-29b[87]. However, YBX1 has also been shown to be important for the secretion of mature miRNAs, which are likely too short to maintain significant secondary structure[85], suggesting that in some cases YBX1 can bind linear sequence motifs. Consistent with recognition of a non-structural motif, previous studies have found that YBX1 binds short pyrimidine containing sequence motifs[29, 30, 59, 87]. Thus, YBX1 may utilize diverse binding preferences resulting in the sorting of multiple RNA species into exosome. Furthermore, whether YBX1 shuttles sncRNAs into exosomes or stabilizes RNAs inside of exosomes cannot be determined from these endpoint analyses.

We find that most RNAs derived from longer RNA biotypes (antisense RNA, lincRNA, protein coding transcripts) are secreted in a YBX1-independent manner. Whether YBX1-independent transcripts are co-incident within the same vesicles as YBX1-dependent transcripts has not been evaluated. In an attempt to restrict our analysis to a single vesicle population, our purification method selects for CD63-positive vesicles of the size and density attributed to exosomes[3]. However, it is formally possible that our isolation procedure co-purifies multiple vesicle populations containing CD63. In this case, it would be informative to perform more rigorous purification methods, perhaps by tracing YBX1-independent and dependent transcripts to see if these can be biochemically separated. Alternatively, YBX1-independent transcripts might be coincident within the same vesicle as YBX1-dependent transcripts, but are sorted via other RNA binding proteins. Mass spectroscopy studies of exosomes have identified numerous RNA binding proteins in exosomes[88] and one protein, hnRNPA2B1, has already been shown to sort specific miRNAs in T-cells[47]. Thus, while small non-coding RNA sorting into exosomes in HEK293 cells is dependent on YBX1, studies of other transcripts and protein binding partners may reveal a distinct route for long non-coding and protein coding transcripts.

Finally, exosomes were purified from cells that had been growing in culture for 48 h and therefore represent a population that had accumulated from cells passing through several rounds of cell division. This presents the intriguing possibility that YBX1-dependent/independent export is a function of cell cycle progression. It is notable that many of the RNA biotypes identified here as YBX1-independent are generally thought to localize and function in the nucleus (antisense RNA, lincRNA, intonic segments of protein coding transcripts) as opposed to the cytoplasm where sorting into EVs would presumably occur. It is possible that following nuclear breakdown during cell division, some transcripts are released into the cytoplasm where they are recognized as mis-localized and then purged via EVs. Identification of the cellular machinery underlying YBX1-indepedent RNA export should shed light on why these transcripts are being released.

Figures

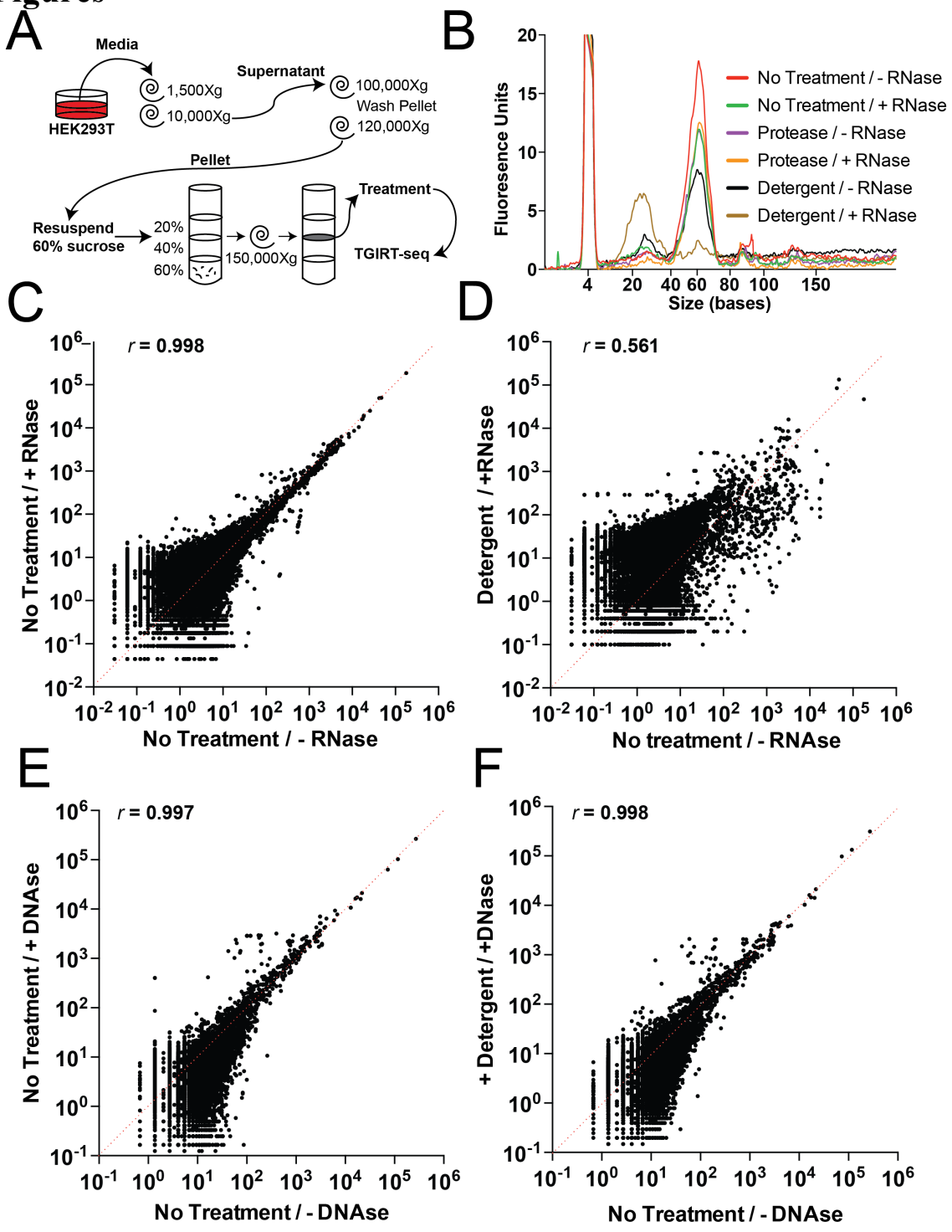


Fig. 3-1: Extracellular vesicle-associated transcripts with nuclease and detergent or protease treatment

(A) Schematic of EV isolation procedure, enzyme/detergent treatment and TGIRT-seq workflow. (B) RNA fragment size analysis by small RNA Bioanalyzer for each treatment

condition. (C-E) Scatterplots comparing nomalized read counts (reads per million mapped reads) between the no treatment control (no treatment and no nuclease) and no treatment with nuclease (C,F) or detergent with indicated nuclease (D,F) for all genes. Correlations were computed as Pearson correlation coefficients (r).

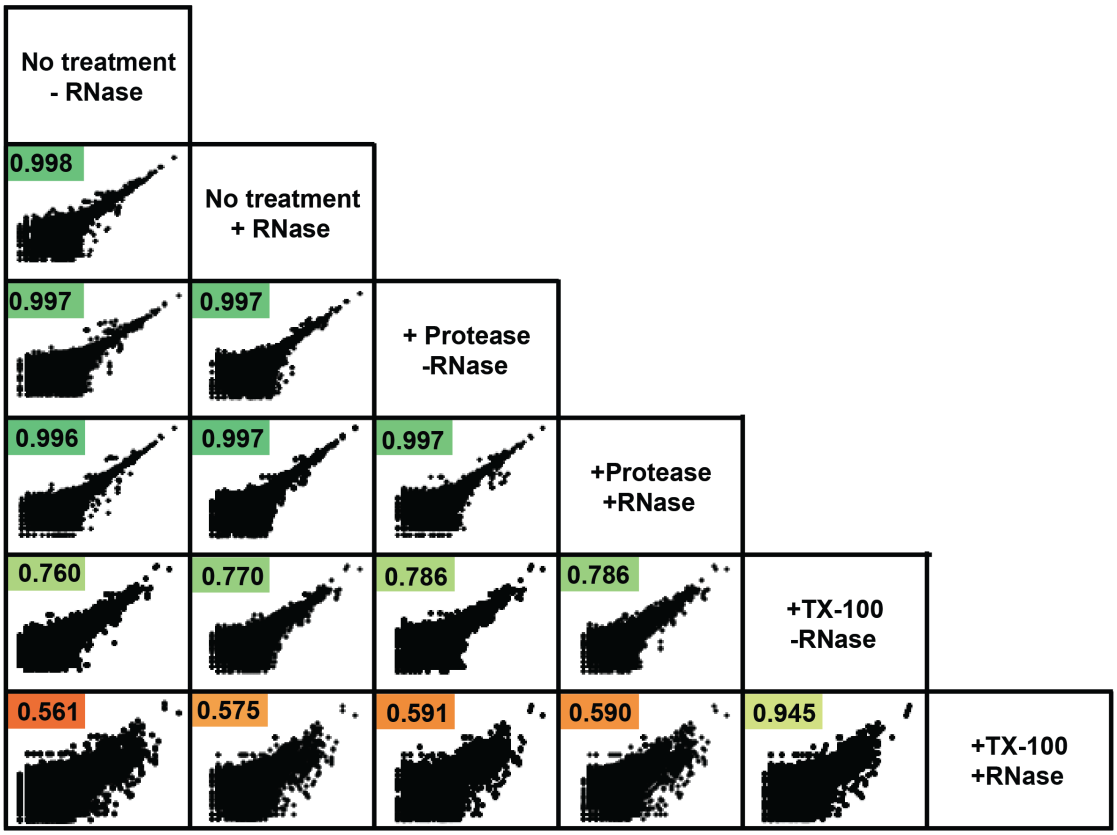


Fig. 3-2: RNase protection of EV-associated transcripts
Pairwise scatter plots and computed sample Pearson correlation coefficients for all RNase treatment conditions.

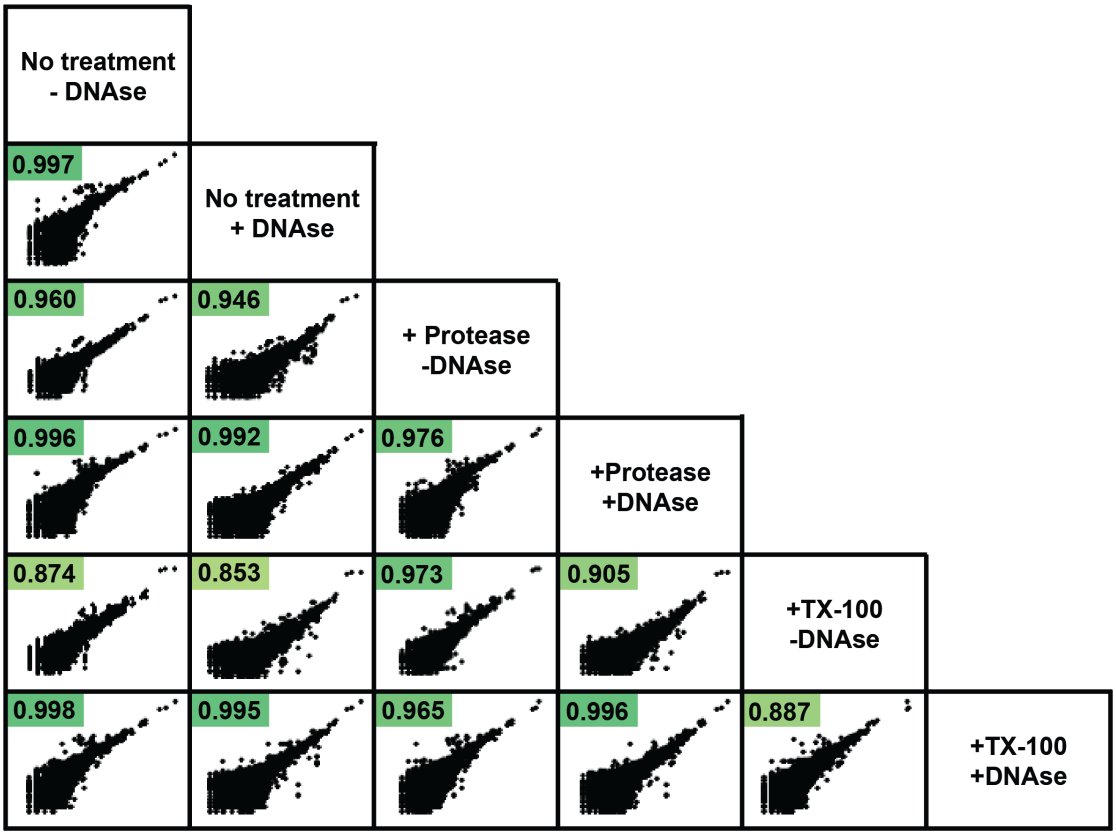


Fig. 3-3: DNAse protection of EV-associated transcripts
Pairwise scatter plots and computed sample Pearson correlation coefficients for all treatment conditions.

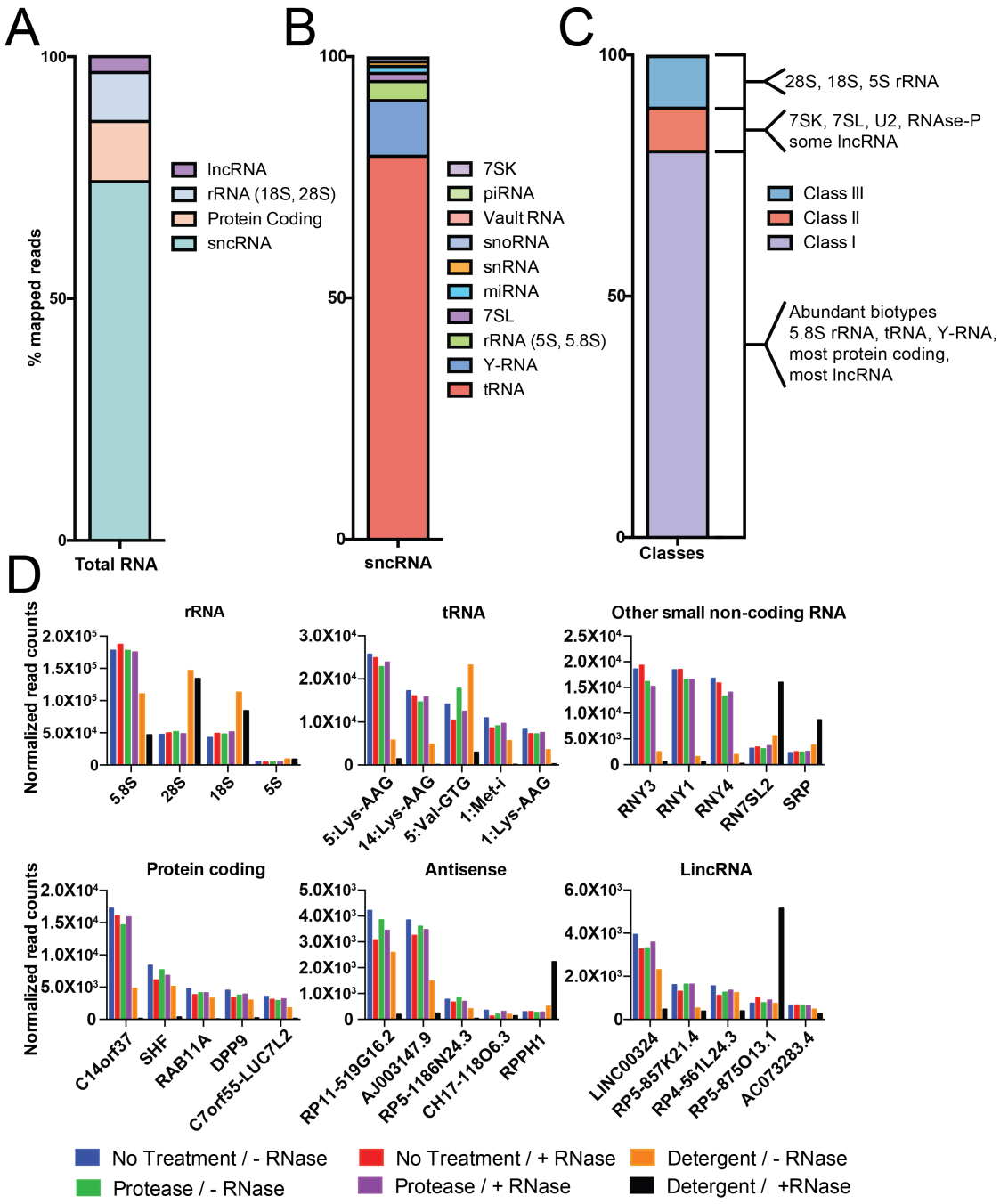


Fig. 3-4: Effect of nuclease treatment on EV-associated biotypes

(A) Stacked bar graphs of percentage of total reads mapping to the indicated features for untreated EVs. (B) Stacked bar graph of the percentage of small non-coding RNA reads mapping to the indicated biotypes. (C) Classification of EV-associated transcripts based on RNase sensitivity (Class I - decreased > 5-fold only in RNase + detergent, Class II - increased more than 5-fold in RNase + detergent, Class III - rRNA reads that remain abundant in all treatment conditions). (D) Normalized read counts (reads per million total mapped reads) of the most highly abundant transcripts across all treatment conditions for the six predominant RNA biotypes: ribosomal RNA, transfer RNA (tRNA), other small

non-coding RNA, protein coding transcripts, antisense RNA and long intergenic non-coding RNA (lincRNA).

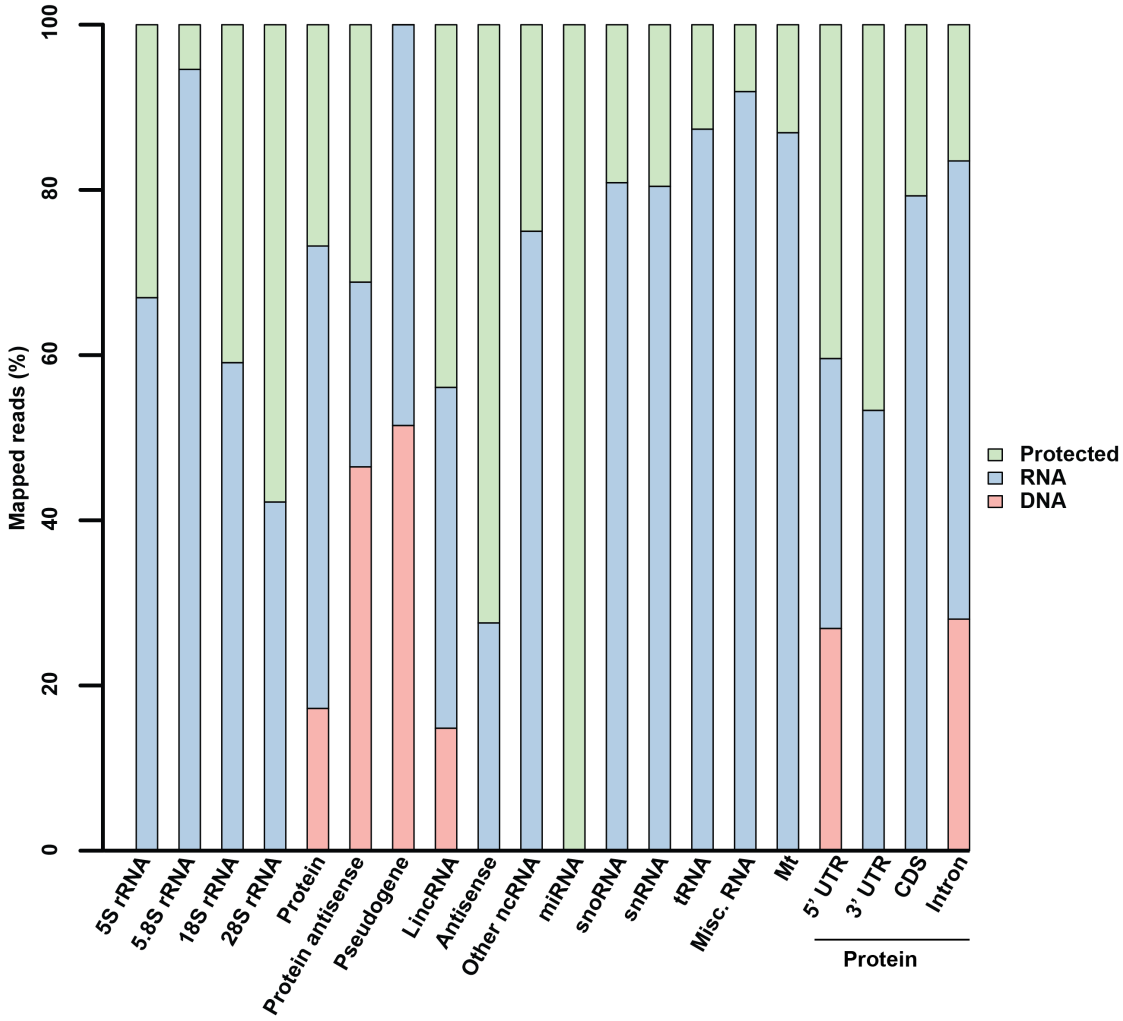


Fig. 3-5: Estimation of DNA contribution to mapped reads for each transcript biotype

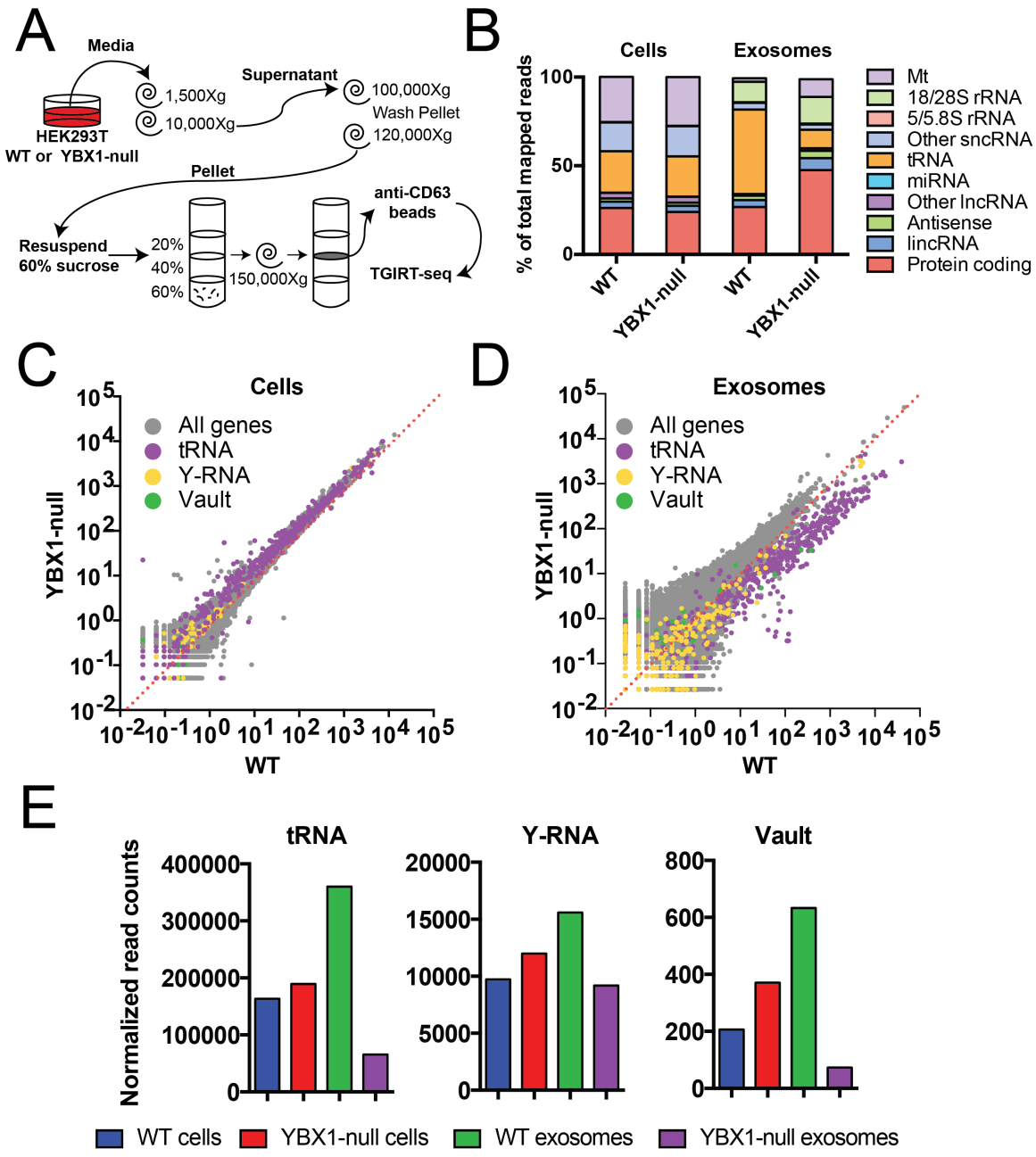


Fig. 3-6: YB1-dependent exosomal RNA

(A) Exosome (CD63-positive EV) purification and TGIRT-seq schematic. (B) Relative proportion of reads mapping to each RNA biotype as a percentage of the total mapped reads. (C,D) Scatter plots comparing normalized mapped read counts for all transcripts, tRNA, Y-RNA and Vault RNA for WT and YBX1-null cells (C) and exosomes (D). (E) Total normalized read counts mapping to tRNA, Y-RNA, Vault RNA. Normalized read counts represent read per million mapped reads summed for all transcripts annotated for each biotype in the GENCODE gene set.

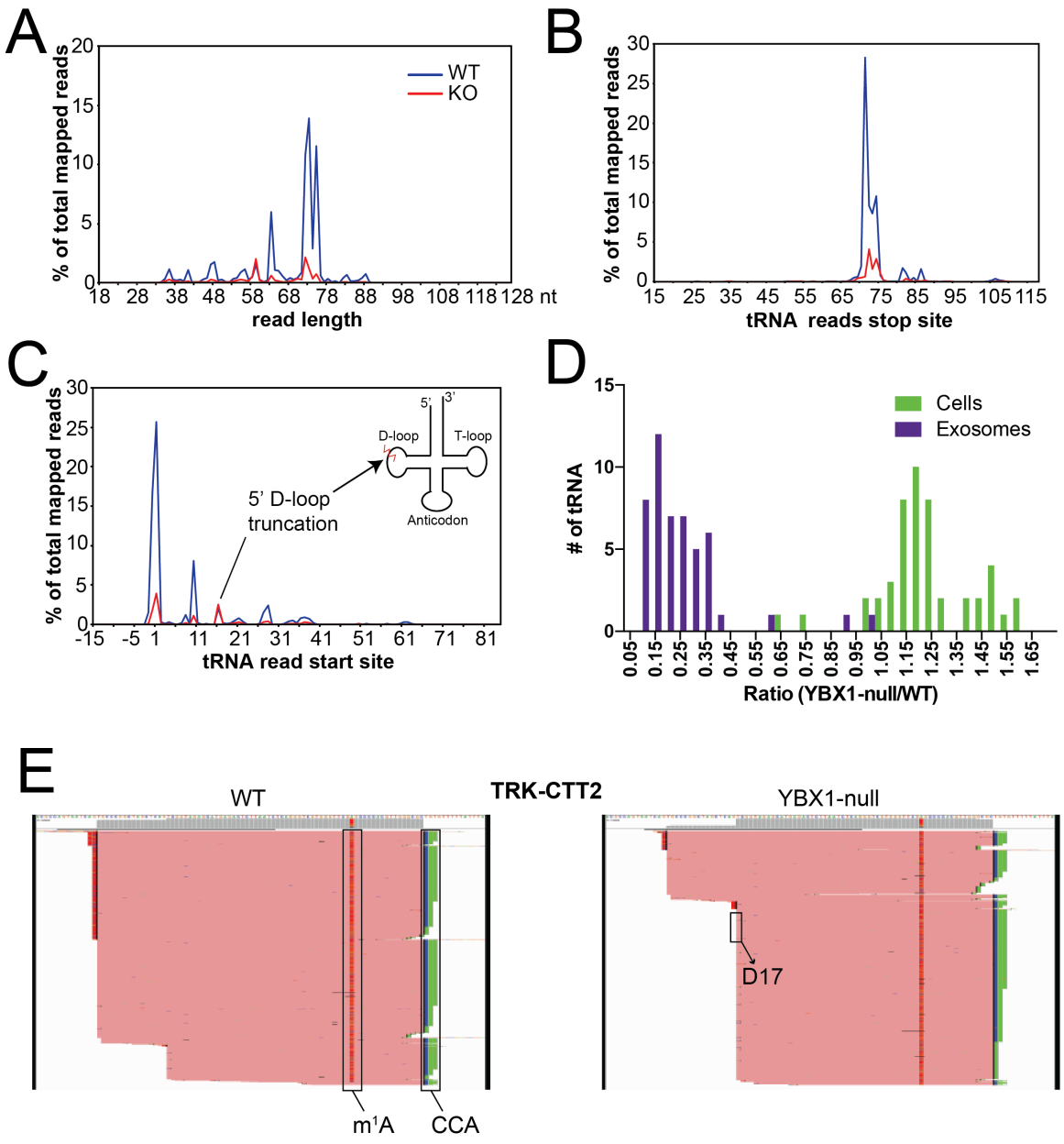


Fig. 3-7: Full-length tRNAs and tRNA fragments in WT and YBX1-null exosomes
(A-C) read length (A), stop site (B) and start site (C) distributions for tRNA mapped reads as a percentage of total mapped reads. (D) Histogram of ratios of (YBX1-null/WT) reads mapping to each tRNA gene grouped by anticodon (N=49). (E) Read coverage across the gene locus for a tRNA lysine isoacceptor (CTT). 1-methyladenine (m1A) and CCA addition post-trascriptional modifications and the D-loop truncation postion (D17) are indicated.

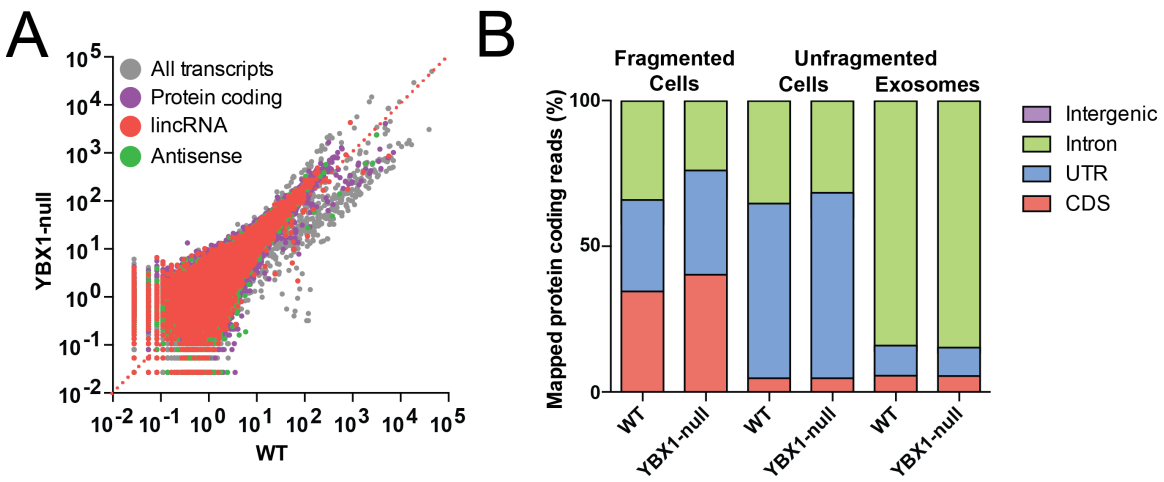


Fig. 3-8: YBX1-independent exosome transcripts

(A) Scatter plot comparing normalized mapped reads for WT and YBX1-null exosomes for all transcripts with protein coding, antisense and lincRNA reads overlaid. (B) Proportion of protein coding reads that map to intergenic, intronic, 5' or 3' untranslated region (UTR) or coding sequence (CDS).

References

1. Malhotra, V., *Unconventional protein secretion: an evolving mechanism.* EMBO J, 2013. **32**(12): p. 1660-4.
2. Zhang, M. and R. Schekman, *Cell biology. Unconventional secretion, unconventional solutions.* Science, 2013. **340**(6132): p. 559-61.
3. Colombo, M., G. Raposo, and C. Thery, *Biogenesis, secretion, and intercellular interactions of exosomes and other extracellular vesicles.* Annu Rev Cell Dev Biol, 2014. **30**: p. 255-89.
4. Arroyo, J.D., et al., *Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma.* Proc Natl Acad Sci U S A, 2011. **108**(12): p. 5003-8.
5. Mitchell, P.S., et al., *Circulating microRNAs as stable blood-based markers for cancer detection.* Proc Natl Acad Sci U S A, 2008. **105**(30): p. 10513-8.
6. Skog, J., et al., *Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers.* Nat Cell Biol, 2008. **10**(12): p. 1470-6.
7. Valadi, H., et al., *Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells.* Nat Cell Biol, 2007. **9**(6): p. 654-9.
8. Kosaka, N., et al., *Secretory mechanisms and intercellular transfer of microRNAs in living cells.* J Biol Chem, 2010. **285**(23): p. 17442-52.
9. Booth, A.M., et al., *Exosomes and HIV Gag bud from endosome-like domains of the T cell plasma membrane.* J Cell Biol, 2006. **172**(6): p. 923-35.
10. Bobrie, A., et al., *Diverse subpopulations of vesicles secreted by different intracellular mechanisms are present in exosome preparations obtained by differential ultracentrifugation.* J Extracell Vesicles, 2012. **1**.
11. Janas, T., et al., *Mechanisms of RNA loading into exosomes.* FEBS Lett, 2015.
12. Ung, T.H., et al., *Exosome proteomics reveals transcriptional regulator proteins with potential to mediate downstream pathways.* Cancer Sci, 2014. **105**(11): p. 1384-92.
13. Buschow, S.I., et al., *MHC class II-associated proteins in B-cell exosomes and potential functional implications for exosome biogenesis.* Immunol Cell Biol, 2010. **88**(8): p. 851-6.
14. Tuschl, T., et al., *Targeted mRNA degradation by double-stranded RNA in vitro.* Genes Dev, 1999. **13**(24): p. 3191-7.
15. Blankenberg, D., et al., *Galaxy: a web-based genome analysis tool for experimentalists.* Curr Protoc Mol Biol, 2010. **Chapter 19**: p. Unit 19 10 1-21.
16. Goecks, J., et al., *Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.* Genome Biol, 2010. **11**(8): p. R86.
17. Giardine, B., et al., *Galaxy: a platform for interactive large-scale genome analysis.* Genome Res, 2005. **15**(10): p. 1451-5.
18. Friedlander, M.R., et al., *miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades.* Nucleic Acids Res, 2012. **40**(1): p. 37-52.

19. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.
20. Hsu, P.D., et al., *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nat Biotechnol, 2013. **31**(9): p. 827-32.
21. Rous, B.A., et al., *Role of adaptor complex AP-3 in targeting wild-type and mutated CD63 to lysosomes.* Mol Biol Cell, 2002. **13**(3): p. 1071-82.
22. Tran, J.H., et al., *Cargo sorting into multivesicular bodies in vitro.* Proc Natl Acad Sci U S A, 2009. **106**(41): p. 17395-400.
23. Falguieres, T., et al., *In vitro budding of intralumenal vesicles into late endosomes is regulated by Alix and Tsg101.* Mol Biol Cell, 2008. **19**(11): p. 4942-55.
24. Wollert, T. and J.H. Hurley, *Molecular mechanism of multivesicular body biogenesis by ESCRT complexes.* Nature, 2010. **464**(7290): p. 864-9.
25. Luberto, C., et al., *Inhibition of tumor necrosis factor-induced cell death in MCF7 by a novel inhibitor of neutral sphingomyelinase.* J Biol Chem, 2002. **277**(43): p. 41128-39.
26. Li, J., et al., *Exosomes mediate the cell-to-cell transmission of IFN-alpha-induced antiviral activity.* Nat Immunol, 2013. **14**(8): p. 793-803.
27. Yuyama, K., et al., *Sphingolipid-modulated exosome secretion promotes clearance of amyloid-beta by microglia.* J Biol Chem, 2012. **287**(14): p. 10977-89.
28. Trajkovic, K., et al., *Ceramide triggers budding of exosome vesicles into multivesicular endosomes.* Science, 2008. **319**(5867): p. 1244-7.
29. Wang, Y., et al., *A complex network of factors with overlapping affinities represses splicing through intronic elements.* Nat Struct Mol Biol, 2013. **20**(1): p. 36-45.
30. Wei, W.J., et al., *YB-1 binds to CAUC motifs and stimulates exon inclusion by enhancing the recruitment of U2AF to weak polypyrimidine tracts.* Nucleic Acids Res, 2012. **40**(17): p. 8622-36.
31. Lyabin, D.N., I.A. Eliseeva, and L.P. Ovchinnikov, *YB-1 protein: functions and regulation.* Wiley Interdiscip Rev RNA, 2014. **5**(1): p. 95-110.
32. Goodier, J.L., et al., *LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex.* Mol Cell Biol, 2007. **27**(18): p. 6469-83.
33. Gallois-Montbrun, S., et al., *Antiviral protein APOBEC3G localizes to ribonucleoprotein complexes found in P bodies and stress granules.* J Virol, 2007. **81**(5): p. 2165-78.
34. Frye, B.C., et al., *Y-box protein-1 is actively secreted through a non-classical pathway and acts as an extracellular mitogen.* EMBO Rep, 2009. **10**(7): p. 783-9.
35. Rauen, T., et al., *YB-1 acts as a ligand for Notch-3 receptors and modulates receptor activation.* J Biol Chem, 2009. **284**(39): p. 26928-40.
36. Turchinovich, A., et al., *Characterization of extracellular circulating microRNA.* Nucleic Acids Res, 2011. **39**(16): p. 7223-33.
37. Squadrito, M.L., et al., *Endogenous RNAs Modulate MicroRNA Sorting to Exosomes and Transfer to Acceptor Cells.* Cell Rep, 2014. **8**(5): p. 1432-46.

38. Gibbings, D.J., et al., *Multivesicular bodies associate with components of miRNA effector complexes and modulate miRNA activity.* Nat Cell Biol, 2009. **11**(9): p. 1143-9.
39. Van Deun, J., et al., *The impact of disparate isolation methods for extracellular vesicles on downstream RNA profiling.* J Extracell Vesicles, 2014. **3**.
40. MacRae, I.J., et al., *In vitro reconstitution of the human RISC-loading complex.* Proc Natl Acad Sci U S A, 2008. **105**(2): p. 512-7.
41. Gagnon, K.T., et al., *Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading.* Nat Protoc, 2014. **9**(9): p. 2045-60.
42. Guo, Y.E. and J.A. Steitz, *3'-Biotin-tagged microRNA-27 does not associate with Argonaute proteins in cells.* RNA, 2014. **20**(7): p. 985-8.
43. Elkayam, E., et al., *The structure of human argonaute-2 in complex with miR-20a.* Cell, 2012. **150**(1): p. 100-10.
44. Schirle, N.T. and I.J. MacRae, *The crystal structure of human Argonaute2.* Science, 2012. **336**(6084): p. 1037-40.
45. Jinek, M., et al., *RNA-programmed genome editing in human cells.* Elife, 2013. **2**: p. e00471.
46. Mali, P., et al., *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-6.
47. Villarroya-Beltri, C., et al., *Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs.* Nat Commun, 2013. **4**: p. 2980.
48. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-8.
49. Bailey, T.L., et al., *MEME SUITE: tools for motif discovery and searching.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W202-8.
50. Cook, K.B., et al., *RBPDB: a database of RNA-binding specificities.* Nucleic Acids Res, 2011. **39**(Database issue): p. D301-8.
51. Gerstberger, S., M. Hafner, and T. Tuschl, *A census of human RNA-binding proteins.* Nat Rev Genet, 2014. **15**(12): p. 829-45.
52. Kedersha, N. and P. Anderson, *Mammalian stress granules and processing bodies.* Methods Enzymol, 2007. **431**: p. 61-81.
53. Nam, Y., et al., *Molecular basis for interaction of let-7 microRNAs with Lin28.* Cell, 2011. **147**(5): p. 1080-91.
54. Bann, D.V., A.R. Beyer, and L.J. Parent, *A murine retrovirus co-Opts YB-1, a translational regulator and stress granule-associated protein, to facilitate virus assembly.* J Virol, 2014. **88**(8): p. 4434-50.
55. Mu, X., et al., *YB-1 stabilizes HIV-1 genomic RNA and enhances viral production.* Protein Cell, 2013. **4**(8): p. 591-7.
56. Li, W., X. Wang, and G. Gao, *Expression of YB-1 enhances production of murine leukemia virus vectors by stabilizing genomic viral RNA.* Protein Cell, 2012. **3**(12): p. 943-9.
57. Blenkiron, C., et al., *Links between the oncoprotein YB-1 and small non-coding RNAs in breast cancer.* PLoS One, 2013. **8**(11): p. e80171.

58. Liu, T.T., et al., *Noncoding RNAs that associate with YB-1 alter proliferation in prostate cancer cells.* RNA, 2015.
59. Goodarzi, H., et al., *Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement.* Cell, 2015. **161**(4): p. 790-802.
60. Cha, D.J., et al., *KRAS-dependent sorting of miRNA to exosomes.* Elife, 2015. **4**.
61. McKenzie, A.J., et al., *KRAS-MEK Signaling Controls Ago2 Sorting into Exosomes.* Cell Rep, 2016. **15**(5): p. 978-87.
62. Eiring, A.M., et al., *miR-328 functions as an RNA decoy to modulate hnRNP E2 regulation of mRNA translation in leukemic blasts.* Cell, 2010. **140**(5): p. 652-65.
63. Thery, C., *Exosomes: secreted vesicles and intercellular communications.* F1000 Biol Rep, 2011. **3**: p. 15.
64. Pegtel, D.M., et al., *Functional delivery of viral miRNAs via exosomes.* Proc Natl Acad Sci U S A, 2010. **107**(14): p. 6328-33.
65. Mittelbrunn, M., et al., *Unidirectional transfer of microRNA-loaded exosomes from T cells to antigen-presenting cells.* Nat Commun, 2011. **2**: p. 282.
66. Rana, S., K. Malinowska, and M. Zoller, *Exosomal tumor microRNA modulates premetastatic organ cells.* Neoplasia, 2013. **15**(3): p. 281-95.
67. Chevillet, J.R., et al., *Quantitative and stoichiometric analysis of the microRNA content of exosomes.* Proc Natl Acad Sci U S A, 2014.
68. Ismail, N., et al., *Macrophage microvesicles induce macrophage differentiation and miR-223 transfer.* Blood, 2013. **121**(6): p. 984-95.
69. Tuerk, C. and L. Gold, *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.* Science, 1990. **249**(4968): p. 505-10.
70. Ellington, A.D. and J.W. Szostak, *In vitro selection of RNA molecules that bind specific ligands.* Nature, 1990. **346**(6287): p. 818-22.
71. Schutze, T., et al., *Probing the SELEX process with next-generation sequencing.* PLoS One, 2011. **6**(12): p. e29604.
72. Jolma, A., et al., *Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities.* Genome Res, 2010. **20**(6): p. 861-73.
73. Manley, J.L., *SELEX to identify protein-binding sites on RNA.* Cold Spring Harb Protoc, 2013. **2013**(2): p. 156-63.
74. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads.* 2011, 2011. **17**(1).
75. Kramer, M.F., *Stem-loop RT-qPCR for miRNAs.* Curr Protoc Mol Biol, 2011. **Chapter 15**: p. Unit 15 10.
76. Ratajczak, J., et al., *Embryonic stem cell-derived microvesicles reprogram hematopoietic progenitors: evidence for horizontal transfer of mRNA and protein delivery.* Leukemia, 2006. **20**(5): p. 847-56.
77. Zomer, A., et al., *In Vivo imaging reveals extracellular vesicle-mediated phenocopying of metastatic behavior.* Cell, 2015. **161**(5): p. 1046-57.
78. Ridder, K., et al., *Extracellular vesicle-mediated transfer of genetic information between the hematopoietic system and the brain in response to inflammation.* PLoS Biol, 2014. **12**(6): p. e1001874.

79. Bellingham, S.A., B.M. Coleman, and A.F. Hill, *Small RNA deep sequencing reveals a distinct miRNA signature released in exosomes from prion-infected neuronal cells.* Nucleic Acids Res, 2012. **40**(21): p. 10937-49.
80. Nolte-'t Hoen, E.N., et al., *Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions.* Nucleic Acids Res, 2012. **40**(18): p. 9272-85.
81. Ogawa, Y., et al., *Small RNA transcriptomes of two types of exosomes in human whole saliva determined by next generation sequencing.* Biol Pharm Bull, 2013. **36**(1): p. 66-75.
82. Vojtech, L., et al., *Exosomes in human semen carry a distinctive repertoire of small non-coding RNAs with potential regulatory functions.* Nucleic Acids Res, 2014. **42**(11): p. 7290-304.
83. Batagov, A.O., V.A. Kuznetsov, and I.V. Kurochkin, *Identification of nucleotide patterns enriched in secreted RNAs as putative cis-acting elements targeting them to exosome nano-vesicles.* BMC Genomics, 2011. **12 Suppl 3**: p. S18.
84. Qin, Y., et al., *High-throughput sequencing of human plasma RNA by using thermostable group II intron reverse transcriptases.* RNA, 2016. **22**(1): p. 111-28.
85. Shurtleff, M.J., Karfilis, K.V., Temoche-Diaz, M., Ri, S., Schekman, R., *Y-box protein I is required to sort microRNAs into exosomes in cells and in a cell-free reaction.* Submitted, 2016.
86. Mohr, S., et al., *Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing.* RNA, 2013. **19**(7): p. 958-70.
87. Wu, S.L., et al., *Genome-wide analysis of YB-1-RNA interactions reveals a novel role of YB-1 in miRNA processing in glioblastoma multiforme.* Nucleic Acids Res, 2015. **43**(17): p. 8516-28.
88. Kalra, H., et al., *Vesiclepedia: a compendium for extracellular vesicles with continuous community annotation.* PLoS Biol, 2012. **10**(12): p. e1001450.