Bioinformatic analysis of endogenous and exogenous small RNAs on lipoproteins

*Ryan M. Allen[1], *Shilin Zhao[2], Marisol A. Ramirez Solano[1], Danielle L. Michell[1], Yuhuan Wang[3], Yu Shyr[2], Praveen Sethupathy[4], MacRae F. Linton[1], Greg A. Graf[3], #Quanhu Sheng[2] #Kasey C. Vickers[1]

[1]Department of Medicine, Vanderbilt Univ. Medical Center, Nashville, TN. 37232 USA
[2]Department of Biostatistics, Vanderbilt Univ. Medical Center, Nashville, TN. 37232 USA
[3]Department of Pharmacology, University of Kentucky. Lexington, KY. 40536 USA
[4]Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY. 14853 USA

*Co-first authors
#Co-corresponding authors

CORRESPONDING AUTHOR:
Kasey C. Vickers, PhD
2220 Pierce Ave.
312 Preston Research Building
Nashville, TN 37232
Ph: 1-615.936.2989
Fax: 1-615.936.1872
kasey.c.vickers@Vanderbilt.edu

ABSTRACT: 274
BODY: 11,112

ABBREVIATIONS:

exRNA, extracellular RNAs; HDL, high-density lipoproteins; HMB, human microbiome project; lncRNA, long non-coding RNA; LDL, low-density lipoproteins; miscRNA, miscellaneous sRNA; ncRNA, non-coding RNA; NIH, National Institutes of Health; nts, nucleotides; osRNA, other sRNA; rDR, rRNA-derived sRNA, RPM, Reads Per Million total reads; rRNA, ribosomal RNA; sRNA, small RNAs; snDR, snRNA-derived sRNA; snoDR, snoRNA-derived sRNA; snoRNA, small nucleolar RNA; snRNA, small nuclear RNA; SR-BI, scavenger receptor BI; sRNA-seq, small RNA sequencing, tDR, tRNA-derived sRNA; tRNA, transfer RNA; yDR, Y RNA-derived sRNA; 3' UTR, 3' untranslated regions.

Abstract

High-throughput small RNA sequencing (sRNA-seq) has facilitated the discovery of many classes of small RNAs (sRNA) and helped establish the field of extracellular RNA (exRNA). Although several tools are available for sRNA-seq analysis, exRNAs present unique analytical challenges that are not met by current software. Therefore, we developed a novel data analysis pipeline specifically for exRNAs entitled, "***T**ools for **I**ntegrative **G**enome analysis of **E**xtracellular s**R**NAs (**TIGER**)*." To demonstrate the power of this tool, sRNA-seq was performed on high-density lipoproteins (HDL), apolipoprotein B-containing particles (APOB), bile, urine, and liver samples collected from wild-type (WT) and scavenger receptor BI knockout (SR-BI KO) mice. TIGER was able to account for approximately 60% of reads on lipoproteins and >85% of reads in liver, bile, and urine, a significant advance compared to existing software, largely due to the identification of non-host sRNAs in these datasets. A key advance for the TIGER pipeline is the ability to analyze host and non-host sRNAs across many classes at the genome, parent RNA, and individual fragment levels. Moreover, disparate sample types were compared at each level using hierarchical clustering, correlations, betadispersions, principal coordinate analysis, and permutational multivariate analysis of variance. TIGER analysis was also used to quantify distinct features of exRNAs, including 5' microRNA (miRNA) variants, 3' miRNA non-templated additions, parent RNA positional coverage, and length distributions by RNA class. Results suggest that the majority of sRNAs on lipoproteins are non-host sRNAs derived from bacterial sources in the microbiome and environment, specifically rRNA-derived sRNAs from proteobacteria. Here, we report novel discoveries of lipoprotein sRNAs that were facilitated by the new sRNA-seq analysis pipeline, TIGER, which has tremendous applicability for the field of exRNA.

Introduction

High-throughput sRNA-seq is the current state-of-the-art method for profiling sRNAs, and is widely-used across many disciplines. Although many software are currently available for sRNA-seq data analysis, most fail to meet the present demands for the study of host and non-host sRNAs across diverse RNA classes. This is particularly important for the investigation of exRNAs, which are heterogeneous pools of host (e.g. human) and non-host (e.g. bacteria) sRNAs. Furthermore, individual sRNA classes harbor distinct features, e.g. miRNA 3' non-templated additions (NTA)[1-3], and these features each require unique strategies for alignments and quantification. A key objective for data analysis is to account for all reads in the sRNA-seq dataset, and current approaches to sRNA profiling now require sophisticated analysis strategies. Therefore, we developed a novel data analysis pipeline specifically for exRNAs entitled, "***T**ools for **I**ntegrative **G**enome analysis of **E**xtracellular s**R**NAs (**TIGER**)*." This pipeline integrates host and non-host sRNA analysis through both genome and database alignments, and greatly improved our ability to account for a larger number of reads in sRNA-seq datasets. The TIGER pipeline was designed for the study of lipoprotein sRNAs; however, it has great applicability to all exRNA studies.

The most extensively studied class of sRNAs is miRNAs[4] and many sRNA-seq analysis tools are limited to only miRNA quantification[5]. In addition to miRNAs, many other classes of sRNAs are present in sRNA-seq datasets[6]. These include sRNAs derived from parent transfer RNAs (tRNA), ribosomal RNAs (rRNA), small nucleolar RNAs (snoRNA), small nuclear RNAs (snRNA), long non-coding RNAs (lncRNA), Y RNAs, and several other miscellaneous non-coding RNAs[7, 8]. For consistency in nomenclature, here, we will refer to these novel sRNA classes as tRNA-derived sRNAs (tDR), rRNA-derived sRNAs (rDR), lncRNA-derived sRNAs (lncDR), snRNA-derived sRNAs (snDR), snoRNA-derived SRNAs (snoDR), Y RNA-derived sRNAs (yDR) and other miscellaneous sRNAs (miscRNA). Outside of miRNAs and tDRs, the biological function(s) of these other endogenous sRNAs are unknown[8, 9]; however, similar to miRNAs, many of these endogenous sRNAs are present in biological

fluids and hold great potential as disease biomarkers or intercellular communication signals. Nevertheless, tools for their analysis in sRNA-seq datasets are very limited[10, 11].

The key advantage of sRNA-seq over other profiling methods (e.g. microarrays) is the ability to quantify sRNAs without prior knowledge of sRNA sequences or genomic annotation, thus sRNA-seq provides the opportunity to discover novel endogenous sRNAs from unlimited host genomes and, most interestingly, exogenous sRNAs - likely from bacteria and fungi – in host tissues and biological fluids[12]. Although it is to be expected that exogenous sRNAs are present in certain biofluids, e.g. saliva[13], bacterial and fungal sRNAs have also been reported in plasma[12], and comprehensive analysis of exRNA requires their quantification in sRNA-seq datasets. Furthermore, analysis sRNA-seq data generated from cells/tissues should also assess exogenous sRNAs as non-human sRNAs have been detected within the RNA-Induced Silencing Complex (RISC) in human cells[12]. Although the functional relevance of exogenous sRNAs in RISC and their potential regulation of human gene expression remains to be determined, these initial studies support the demand for bioinformatic strategies to identify and quantify exogenous sRNAs in diverse sRNA-seq datasets.

In plasma and other biofluids, exRNAs are carried by extracellular vesicles (EV), lipoproteins, and ribonucleoproteins, which protects exRNAs against RNase-mediated degradation[14, 15]. Previously, we reported that lipoproteins - low-density lipoproteins (LDL) and high-density lipoproteins (HDL) - transport miRNAs in plasma, and lipoprotein miRNA signatures are distinct from exosomes[16]. Using real-time PCR-based TaqMan arrays to profile HDL-miRNAs, we further identified HDL-miRNAs that were significantly altered in hypercholesterolemia and atherosclerosis[16]. Currently, it is unknown if lipoproteins transport other sRNAs in addition to miRNAs. In a previous study, we reported that HDL transfer miRNAs to recipient cells and this process is regulated by HDL's receptor, scavenger receptor BI (SR-BI), in hepatocytes[16]. SR-BI is a bidirectional transporter of cholesterol and a critical factor in reverse cholesterol transport pathway in which HDL returns excess cholesterol to the liver for excretion to bile. Currently, it is unknown if miRNAs, and potentially other sRNAs, on lipoproteins follow

cholesterol and are transported to the liver for secretion to bile. Furthermore, it is unknown if SR-BI regulates miRNAs or other sRNAs on HDL, LDL, or biofluids, e.g. bile.

To demonstrate the power of the TIGER pipeline, we present results from our comprehensive analysis of lipoprotein sRNAs by high-throughput sRNA-seq. Using TIGER, we found that lipoproteins transport a wide-variety of host and non-host sRNAs, most notably, HDL and APOB particles transport non-host bacterial tDRs and rDRs. TIGER analysis was also used to demonstrate that lipoprotein sRNA signatures are distinct from liver, bile, and urine for host sRNAs and their distinct features. Moreover, TIGER analysis was used to determine the role of SR-BI in the regulation of exRNAs on lipoproteins and in biofluids. At the parent RNA level, SR-BI-deficiency had minimal impact on sRNA levels; however, by organizing sRNAs at the individual fragment level, we found that loss of SR-BI in mice resulted in significant changes to specific sRNA classes in different sample types, e.g. snDRs on HDL were found to be increased in SR-BI KO mice compared to WT mice. Collectively, the development and application of TIGER overcame many of the barriers and challenges exRNA sequencing analysis and uncovered many novel observations for sRNAs on lipoproteins and in liver and biological fluids.

Results

Lipoproteins transport distinct sRNA signatures.

miRNAs are just one of several classes of non-coding sRNAs, and many of these non-miRNA sRNAs have been detected in plasma and extracellular fluids[7-9, 17]. Moreover, many of these sRNAs are also likely to be associated with HDL and LDL in circulation; however, the full-compendium of exRNAs on lipoproteins has not been investigated. Therefore, an unbiased approach to identifying and quantifying sRNAs on lipoproteins was warranted. To address this gap in knowledge, high-throughput sRNA-seq was used to profile all sRNAs on HDL and apoB-containing particles (APOB) purified from mouse plasma by size-exclusion chromatography (SEC) (**Fig.S1A-C**). Profiling lipoprotein sRNAs by sRNA-

seq presented unique challenges in data analysis, e.g. miRNAs were only a minor fraction of all host sRNAs and we detected many non-host sRNAs by individual searches. Therefore, we designed a new pipeline (TIGER) to address these complexities. To demonstrate the unique features of sRNAs on lipoproteins, mouse HDL and APOB profiles were compared to mouse liver, bile, and urine. The TIGER pipeline is composed on multiple data analysis modules in which the priority and order of analyses can be defined. Many biofluids contain both host (e.g. mouse) and non-host (e.g. bacteria) sRNAs and the ratio between host and non-host can be used to determine the order of analysis and priority of assumptions. For example, in a recent study, sRNAs in saliva were aligned first to bacterial genomes prior to the host genome[13]. Here, we aligned reads first to the host using a combination of host genome and mature transcripts in specific databases followed by parallel alignments to non-host genomes (e.g. bacteria species) and non-host sRNA libraries (e.g. tRNA database) (**Fig.1**). For host sRNAs, the TIGER pipeline prioritized annotated sRNAs in ranking order; miRNAs, tDRs, rDRs, snDRs, snoDRs, yDRs, lncDRs, miscellaneous sRNAs (miscRNA), and unannotated host genome sRNAs. For non-host genomes, the TIGER pipeline aligned reads in parallel (equally) to genomes organized into microbiome bacteria, environment bacteria, and fungi. Using this approach, we were able to make many novel observations concerning nucleic acid cargo on lipoproteins that would not be possible using currently available software.

Distinct features of miRNAs on lipoproteins

To compare miRNA content between groups, miRNA read counts can be normalized by Reads Per Million total reads (RPM) or Reads Per Million miRNA reads (RPM miR). To determine the appropriate normalization for our data, both approaches were compared to real-time PCR (TaqMan assay) results for 9 miRNAs across all samples, and we found that normalization of miRNAs by RPM (R^2=0.45) showed a higher correlation between sequencing and PCR results than RPM miR (R^2=0.17) (**Fig.2A, Table S1**) Lipoproteins, specifically APOB particles, were found to have less miRNA content, as

reported by total miRNA counts (RPM), than livers which had the largest fraction of miRNAs per total reads (RPM) (**Fig.2B**). To compare miRNA signatures across sample types, Principal Coordinate Analysis (PCoA) was used, and we found distinct clustering of lipoproteins and biofluids separate from livers (**Fig.2C**). To quantify differences in the homogeneity of multivariable distributions (miRNAs) for the samples in each group, PERMANOVA tests were used, and we found that the miRNA profiles of lipoproteins (HDL and APOB) and biofluids (bile and urine) were significantly distinct from livers (WT mice) – APOB ($F=9.57$, $p=0.001$), HDL ($F=7.11$, $p=0.001$), bile ($F=5.56$, $p=0.001$), and urine ($F=8.42$, $p=0.001$) (**Table S2**). Betadispersions can be used to calculate the distances of individual samples within a group to the group's centroid, and lipoprotein (high-dispersions) and biofluid (high-disperisons) samples were significantly (ANOVA $P<0.05$) more dispersed than livers (low-dispersion) – APOB ($F=31.03$, $p<0.0001$), HDL ($F=23.20$, $p<0.0001$), bile ($F=17.09$, $p<0.0001$), and urine ($F=15.47$, $p<0.0001$) (**Fig.2C**). To further compare miRNA signatures between groups, high-end analyses were performed using hierarchical clustering and correlations (Spearman) of group means. Both HDL and APOB groups distinctly clustered away from liver and biofluids, and lipoproteins displayed high correlations between HDL and APOB groups and modest correlations with liver, bile, and urine groups (**Fig.2D**). These results suggest that HDL and APOB transport unique miRNA signatures that are distinct from liver, as lipoproteins showed significantly less homogeneity of miRNAs, increased sample dispersions, and clustered separately from liver.

miRNAs (19-23 nts in length) post-transcriptionally regulate gene expression through binding to and suppressing mRNA targets[4]. Recognition of mRNA target sites is conferred through a critical "seed" region (bases 2-7) on the 5' end of the miRNA[18]. During biogenesis, mature miRNAs are processed from precursor miRNA hairpins and imprecise cleavage can give rise to variations on the 5' end[19-21]. As such, one miRNA locus can produce multiple miRNA isoforms, termed isomiRs, which can differ by one or two nts at the 5' start position; therefore, the miRNA "seed" region sequence can be shifted and the recognition of mRNA targets altered[20-22]. Therefore, it is important that analysis of miRNAs in sRNA-seq

datasets includes quantification of isomiRs; however, most analysis software only assess canonical miRNAs. One feature of the TIGER pipeline is the ability to quantify both canonical miRNAs and their various isoforms (isomiRs). In our study, all samples contained 5' isomiRs, the largest fraction was found on HDL (8.42%) followed by urine (7.2%), APOB (6.53%), bile (4.54%), and liver (4.34%) (**Fig.S2A-B**). In addition, we found specific examples of miRNAs with different 5' terminal start positions than their reported canonical forms, e.g. miR-142-5p (-2), miR-133a-3p (+1) and miR-192-5p (+1), and these patterns were consistent across all sample types, suggesting these annotations may be inaccurate in the database (miRBase) (**Fig.2E**). Most interestingly, we found that lipoproteins and biofluids contained significantly more 5' (-1) isomiRs of miR-101a-3p than liver samples, which may be evidence of isomiR partitioning for miR-101a-3p between extracellular and intracellular pools (**Fig.2E**). Mature miRNAs also harbor extensive variability on their 3' terminal ends due to imprecise processing and NTAs, i.e. extra non-genomic 3' nts added by cytoplasmic nucleotidyltransferases[3, 23]. The most common NTA events are poly-uridylation and poly-adenylation, which have been reported to decrease stability and activity, respectively[2, 3, 21, 23]. In our analysis, all sample types were found to contain a substantial fraction of miRNAs that were modified with NTAs (17-32%) (**Fig.S2C**). Contrary to what was observed for canonical miRNAs and 5' isomiRs, APOB particles contained significantly more miRNAs with NTAs than liver samples (**Figs.S2B-C**). A previous study reported that miRNA poly-uridylation (NTA-U) was significantly increased on extracellular miRNAs released in exosomes, whereas miRNA poly-adenylation (NTA-A) was associated with cellular retention[24]. To determine if lipoproteins and/or biofluids are similarly enriched with NTA-U, NTA patterns were compared between groups, and we found that HDL and APOB (WT mice) were indeed significantly enriched with NTA-U compared to liver samples which were enriched with NTA-A (**Fig.2F**). Nonetheless, miRNAs in bile and urine from WT mice were not enriched with NTA-U (**Fig.2F**). Collectively, these results further demonstrate that miRNAs on lipoproteins, particularly APOB particles, are distinct for many features from hepatic miRNAs, including 5' isomiRs and 3' NTAs, and the TIGER analysis pipeline aided in these findings.

Lipoproteins transport many classes of host sRNAs

Most, if not all, non-coding RNAs are processed to smaller fragments creating an enormously diverse pool of sRNAs in cells and extracellular fluids[9]. To determine if HDL and APOB particles also transport non-miRNA sRNAs and to compare annotated host sRNAs across sample types, reads were aligned to the host (mouse) genome, as well as to mature transcripts for specific classes of RNAs with genes harboring introns, e.g tRNAs and rRNAs. The largest class of host sRNAs detected in livers was rDRs 42-45 nts in length (**Figs.3A,B**). rDRs were also present on HDL and APOB particles; however, their lengths were variable (**Figs.3A,C,D**). We also detected snoDRs (57-64 nts in length) in livers; however, snoDRs were largely absent from lipoproteins and biofluids, suggesting that the liver and other tissues may not export this class of sRNAs to lipoproteins or into bile or urine (**Figs.3A-F**). Both lipoproteins and biofluids contained tDRs 28-36 nts in length, which suggests that these reads are likely tRNA-derived halves (tRHs), a sub-class of tDRs approximately 31-35 nts in length (**Figs.3A,C,D**)[42, 47]. Most tDRs on lipoproteins and in biofluids aligned to the 5' halves of parent tRNAs, particularly anti-codons representing glutamate (GluCTC), glycine (GlyGCC), aspartate (AspGTC), and valine (ValCAC) (**Figs.4A,S3**). Strikingly, 68.9% of tDR reads on WT HDL and APOB particles aligned to the parent tRNA GluCTC (**Figs.4A,S4A,B**). At the parent tRNA level, tDR signatures demonstrated considerable overlap of all groups by PCoA (**Fig.4B**). Nevertheless, at the fragment level, lipoprotein tDR signatures were demarcated from livers and biofluids (**Fig.4C**). PERMANOVA analysis found that lipoprotein and biofluids were significantly distinct from liver signatures at the fragment level: WT APOB ($F=5.32$, $p=0.001$), HDL ($F=2.94$, $p=0.014$), bile ($F=10.22$, $p=0.001$), and urine ($F=7.08$, $p=0.001$) (**Table S2**). Hierarchical clustering and correlation analyses further supports that individual tDR fragments, not parent tRNAs, define tDR profiles across sample types (**Figs.S5A-B**). Most interestingly, this observed pattern of overlap at the parent level and definition at the fragment level was consistent for other host sRNAs, including rDRs and snDRs (**Figs.S5C-F,S6A-F, Table S2**).

To validate candidate tDRs on lipoproteins and in biofluids that were identified by sRNA-seq, real-time PCR using custom locked-nucleic acid (LNA)-based assays (Exiqon) were completed. Both tDR-GluCTC (38 nts in length) and tDR-GlyGCC (32 nts in length) were confirmed to be highly-abundant on lipoproteins; however, they were also readily detected livers, bile, and urine, but not detected in the negative control (buffer) solution used to isolate the lipoproteins (**Fig.4D,E**). Furthermore, real-time PCR was used to validate other sRNA candidates on lipoproteins representing other classes of RNAs. For example, the abundance of two distinct snDRs and a candidate sRNA cleaved from a ribozyme (miscRNA) were detected by PCR on lipoproteins similarly to a previously reported miRNA on lipoproteins (miR-223-3p) (**Figs.S7A-D**). Although these PCR assays detected single products, as determined by melting curves, sRNA-seq datasets contained many sequences that were very similar to the candidate sRNAs. Moreover, although the general, regional cleavage patterns for specific parent RNAs were consistent for tRNAs (**Fig.S3**) and snRNAs (**Fig.S8**), specific fragmentation and exact sRNA sequences were variable across samples, e.g. lengths of related sRNAs. Therefore, to compare between samples within a group, correlations were performed at both the parent and fragment levels. For tDRs (**Fig.4F**) and other RNA classes (**Fig.S9**), we found high correlation between samples at the parent level and poor correlation across samples at the fragment (read) level for lipoproteins and biofluids. For liver samples, high-correlation was detected for sRNAs at both the parent and fragment levels (**Figs.4F,S9**). These results suggest that, although individual fragments define sRNA classes across groups, further investigation of individual candidate sRNAs (fragments) may be challenging due to variability across samples.

Lipoproteins are highly-enriched in exogenous sRNAs

Reads aligning to non-human transcripts have previously been detected in human plasma samples[12]; however, it is unknown which carriers transport non-host sRNAs in host circulation. To determine if

lipoproteins carry exogenous bacterial and fungal sRNAs, reads >20 nts in length that failed to map to the host (mouse) genome were aligned in parallel to A.) Annotated non-host transcripts curated in GtRNAdb (tRNA), SILVA (rRNA), and miRBase (miRNA) databases, and B.) Genomes of bacteria and fungi present in the microbiome (human microbiome, HMB) or environment (ENV) (**Fig.1**). To identify exogenous miRNAs (xenomiRs), reads were aligned with perfect match (PM) to non-host mature miRNA sequences (miRBase.org); however, only a few xenomiRs were detected on mouse lipoproteins or in liver, bile, or urine datasets (**Table S3**). To determine the levels of exogenous tDRs on lipoproteins, non-host reads were aligned to parent tRNAs curated in the GtRNAdb library. Both HDL and APOB particles were found to transport a diverse set of exogenous tDRs across multiple kingdoms, which accounted for approximately 2.5% of the sRNAs (total reads) circulating on each lipoprotein class (**Figs.5A, Table S4**). Based normalized read counts, bacteria was the most represented kingdom, and the bacterial species with the highest normalized read counts were Pseudomonas fluorescens, Pseudomonas aeruginosa, and Acinetobacter baumanni (**Fig.S10A,B**). The parent tRNA (amino acid anti-codons) with the highest normalized read counts were fMetCAT, GluTTC, AspGTG, and AsnGTT (**Figs.S10C, Table S4**). Positional coverage analyses of bacterial tDRs found that bacterial tDRs aligned to both the 5' and 3' halves of parent tRNAs (**Figs.5B,S11,S12**), which differed from host tDRs which predominantly aligned to the 5' halves of parent tRNAs (**Figs.4A,S3**). To determine if lipoproteins also transport exogenous rDRs, non-host reads were also aligned to known rRNA transcripts curated in the SILVA database, and remarkably, reads aligned to non-host rDRs accounted for approximately 25% of the total reads in each of the HDL and APOB datasets (**Figs.5C, Table S5**). Strikingly, rDRs from every taxonomical kingdom were present on lipoproteins; however, rDRs from bacteria were the most abundant on HDL and APOB particles (**Figs.5C,S13, Table S5**). Although the overall content of non-host sRNAs on HDL and APOB particles were similar, HDL were found to be enriched for shorter length non-host tDRs and rDRs compared to APOB particles (**Figs.5D,E**). Collectively, these results suggest that lipoproteins transport exogenous tDR and rDRs, most of which are likely bacterial in origin.

Aligning reads to transcripts (databases) is biased in that only known (annotated) RNAs are queried, and thus, limits the power of discovery in sRNA datasets as many non-host genomes are poorly annotated. Furthermore, non-host reads on lipoproteins, and in biological fluids, are not likely to be restricted to only tRNAs and rRNAs. To overcome these limitations and comprehensively analyze exogenous sRNAs on lipoproteins and in biofluids and livers, non-host reads were also aligned to bacterial genomes within the NIH HMB Project (hmpdacc.org). The HMB database currently holds 3,055 genomes, many of which are highly similar; therefore, to address potential multi-mapping issues, we collapsed these species into 206 representative genomes that spanned 11 phyla and accounted for every genera within the HMB. Alignment of non-host reads to HMB genomes resulted identified many bacterial sRNAs on lipoproteins and in biofluids, reported as summarized genome read counts per million total reads (RPM) **(Figs.S14A-B, Table S6)**. The HMB genomes with the largest read counts for sRNAs on HDL from WT mice were Pseudomonas 2 1 26 uid40037, Micrococcus luteus SK58 uid34071, Acinetobacter ATCC 27244 uid30949 (**Figs.S14A-B, Table S6**). To perform a taxonomical analysis of lipoprotein-associated bacterial sRNAs, circular tree maps were generated. As shown by concentric rings in the tree maps, the vast majority of both HDL and APOB bacterial reads mapped to the Proteobacteria phylum (green), followed by the Actinobacteria (blue) and Firmicutes (yellow) phylums (**Figs.6A,S15A**). Within the Proteobacteria phylum, a majority of the reads aligned to genomes in the Gammaproteobacteria class, particularly the orders of Pseudomonadales and Enterobacteriales, and the family of Enterobacteriaceae. Among individual genera (inner-most circles), counts for the genus Pseudomonas (Proteobacteria phylum) were consistently the high, as were Micrococcus (Actinobacteria phylum) (**Figs.6A,S15A**).

We also observed that many reads that aligned to bacterial rRNA transcripts, failed to align to HMB genomes, thus suggesting that these reads may be derived from bacteria not presently curated in the HMB database. Using BLASTn (NCBI), many highly abundant reads were found to be perfect matches to genomes of bacterial species present in the environment, e.g. soil bacteria. Therefore, to increase

our bacterial coverage, 167 additional bacterial genomes representing non-redundant genera of 8 taxonomical phyla were added, termed here as environmental bacteria (ENV). The bacterial species with the most ENV genome counts for lipoproteins from WT mice were Pseudomonas fluorescens, Pseudomonas putida, Propionibacterium acnes, and Stenotrophomonas maltophilia (**Figs.S14C-D,S15B-C, Table S7**). Although many bacterial reads (sequences) were shared between HMB and ENV bacterial genomes, greater than half of all non-host (genome) bacterial reads could be assigned exclusively to only one database, suggesting a complex origin for bacterial sRNAs associated with lipoproteins (**Fig.S16**). Although most reads identified through BLASTn analysis were bacterial, we also identified several fungal species, including the genus *Fusarium*. The highest genome counts for fungal species on WT HDL were Fusarium oxysporum, Histoplasma capsulatum, Cryptococcus neoformans (**Figs.S17A-B, Table S8**).

To assess bacterial sRNAs across samples, non-host sRNAs (HMB and ENV) signatures on lipoproteins were correlated (Spearman) between samples at both the genome and fragment levels. For both databases, we identified high correlations between lipoprotein samples at the genome level and low correlations at the fragment level (**Figs.6B,C**). These data suggest that similar bacteria are contributing sRNAs to circulating lipoproteins across all mice; however, these bacteria are likely contributing different sRNAs (sequences) to HDL and APOB particles in different mice. Most interestingly, a key difference between HDL and APOB bacterial sRNAs was length, as HDL were enriched for shorter sRNAs than APOB particles; this pattern was evident for both HMB and ENV sRNAs (**Figs.6D,E**). A similar trend was observed for reads mapping to fungal genomes (**Fig.S18**). For host sRNAs, HDL and APOB particles were found to transport very similar profiles and the lipoprotein samples clustered together with considerable overlap at both the parent and fragment levels (**Figs.2C,S6**). To determine if HDL and APOB particles transport different exogenous (non-host) sRNA signatures, PCoA and PERMANOVA analyses were completed. At the genome level, the HDL and APOB particles were indistinguishable for HMB and ENV bacteria (**Figs.S19A-B**). At the fragment level,

HDL and APOB profiles clustered separately and HDL and APOB profiles were significantly distinct (F=1.7, p=0.048) for ENV bacterial sRNAs by PERMANOVA (**Figs.S19C-D, Table S9**)

The lack of strong correlation at the fragment level for non-host sRNAs is likely due to differences in read lengths and sequences (e.g. terminal nts) for similar reads, and thus variable read counts across samples. These observations present unique challenges to study individual sRNAs for biological function; however, many candidate sRNAs do exist within the very large pool of non-host reads. Real-time PCR was used to quantify candidate bacterial sRNAs on lipoproteins, and we confirmed that HDL and APOB particles transport a 22 nt rDR (5'-AGAGAACUCGGGUGAAGGAACU-3') likely from bacteria in the Proteobacteria phyla (**Figs.6F,S20**). Likewise, HDL and APOB were also found to transport another rDR in the Proteobacteria phyla, likely from the order of Burkholderiales (33 nts, 5'-GACCAGGACGUUGAUAGGCUGGGUGUGGAAGUG-3') (**Figs.6G,S21**). In addition to bacterial sRNAs, real-time PCR was also used to confirm that lipoproteins also transport a fungal rDR from the Verticillium genus (21 nts 5'-UGGGUGUGACGGGGAAGCAGG-3') (**Fig.S22**). Collectively, these results suggest that HDL and APOB transport non-host sRNAs derived from bacterial and fungal sources in the microbiome and environment, and that analysis of lipoprotein sRNA-seq data should include their identification and quantification.

Class-Independent Analysis of Lipoprotein sRNAs

To determine which RNA class and species contribute to the most abundant sRNAs in each sample type, the top 100 ranked reads for each sample were filtered and redundant reads were removed for each group. The top abundant sequences were then linked to reads identified in the host and non-host modules. For liver samples, the top ranked reads were entirely host sRNAs (**Fig.7A**). Although lipoproteins likely transport more non-host sRNAs than host sRNAs, circos plots demonstrated that the top most abundant reads on lipoproteins are comprised of both host and non-host sRNAs (**Figs.7B,C**).

The top ranked reads in urine samples were found to be mainly host sRNAs (tDRs); however, many links to exogenous bacterial sRNAs were identified (**Fig.7D)**. For bile, the most abundant reads were entirely host sRNAs (**Fig.7E**). Nonetheless, some of the top ranked sequences were not identified through our host and non-host analyses; therefore, we sought to further analyze lipoproteins sRNAs using a class-independent strategy.

Many sRNA-seq data analysis pipelines are limited to only miRNAs and most of the recent advances in sRNA-seq data analysis for other host RNAs are designed to quantify and categorize sRNAs based on the contributing parent RNA class. However, the biological functions of individual sRNAs, although currently unknown, are not likely to be conferred by specific groupings or classifications based on the parent RNAs. Their biological roles and activities are most likely to be influenced by their individual sequence, length, base chemistry, terminal nts, and most importantly, abundance. Furthermore, exRNAs are a new class of disease biomarkers and their value as representative signals is not solely dependent on parent RNAs, if at all. Therefore, a key advantage of the TIGER pipeline is the ability to assess the most abundant reads in sRNA datasets independent of parent RNA class and/or contributing species (host or non-host). To assess the similarity of profiles between groups for the top ranked sRNAs, hierarchical clustering and correlations were performed, and lipoproteins were highly correlated between groups, and HDL and APOB profiles clustered separately from livers, bile, and urine (**Fig.S23**). These observations were confirmed by PCoA, as lipoprotein samples overlapped and clustered together, separately from bile, urine and liver samples (**Fig.7F**). PERMANOVA analysis found that APOB ($F=19.56$, $p=0.002$), HDL ($F=15.71$, $p=0.001$), bile ($F=49.74$, $p=0.003$), and urine ($F=22.07$, $p=0.002$) were significantly distinct from liver profiles (**Table S10**). Most interestingly, every group was significantly distinct from each other based on the most abundant sRNAs (Top 100) in each WT group, as determined by PERMANOVA (**Table S10**). These results suggest that each sample type can be defined by their most abundant sRNAs independent of parent RNA class or contributing host or non-host species which is highly appropriate for the study of exRNAs.

Comparison of TIGER to Other Pipelines.

To compare the TIGER pipeline to other sRNA-seq analysis software, APOB, HDL, and liver samples from WT mice were analyzed by Chimira[5], Oasis[11], ExceRpt[13], and miRge[10] software (**Table S11**). Although each pipeline is designed for different outputs, each can quantify host miRNAs for which we used to compare analyses, and we found that all the pipelines were comparable in their ability to quantify host canonical miRNAs for different sample types (**Fig.24A**) and the pipelines were highly correlated for miRNAs (**Fig.S24B**). Most available software for sRNA-seq data analysis are restricted to miRNAs or endogenous (host) sRNAs, including Chimira, Oasis, and miRge (**Table S11**). This may be suitable for liver samples (red circles), but HDL (blue circles) and APOB (green circles) samples remain largely unexplained using this approach, as demonstrated by ternary plots (**Fig.8A**). Incorporation of both endogenous and exogenous sRNAs, a key feature of the TIGER pipeline, is essential to studying HDL and APOB sRNAs, as this strategy accounts for more reads in the datasets, as depicted by the left shifts of blue and green circles in the ternary plot (**Fig.8B**). The main feature of the TIGER pipeline is the ability to explain a large amount of data in exRNA datasets, as demonstrated by the summary output **Table S12**. A key metric for comparing pipelines/software is the percent of assigned quality reads, i.e. the amount of (useable) information extracted from the data by the software. Remarkably, the TIGER pipeline accounted for 87.95% bile, 87.9% of liver, 85.3% urine, 71.5% HDL, and 62.2% APOB reads in WT mice (**Fig.8C, Table S13**). In comparison to other pipelines, that TIGER pipeline significantly increased the assignment % of total reads for HDL and APOB (**Fig.8D**). Remarkably, the TIGER pipeline also explained significantly more reads than Chimira, Oasis, and ExceRpt in liver datasets which contain a large fraction of host sRNAs (**Fig.8D**). Furthermore, after the TIGER pipeline aligns non-host reads to RNA transcripts in the databases and the different collections of bacterial and fungal genomes, the top 100 ranked sequences of the remaining unexplained reads are filtered and searched using BLASTn (**Fig.1**). This is an added feature of the TIGER pipeline that is designed to

identify the potential sources and species of non-host reads that were not accounted for by the initial alignment strategy. Collectively, the TIGER pipeline provides an opportunity to analyze sRNA-seq with increased depth and detail which is particularly suited for analysis of exRNA and sRNAs on lipoproteins.

SR-BI Regulation of Lipoprotein sRNAs

SR-BI is highly-expressed in the liver and plays a fundamental role in reverse cholesterol transport mediating hepatic uptake of HDL-cholesteryl esters and biliary cholesterol secretion[25-27]. Loss-of-function variants in human *SCARB1* (SR-BI) were associated with increased in circulating HDL-C levels[28]. Likewise, *Scarb1* mutations in mice also resulted in increased HDL-C levels[29]. We have previously reported that HDL-delivery of miRNAs to hepatocytes requires SR-BI[16]. Based on these observations, SR-BI may regulate sRNA levels on lipoproteins and in liver and bile. To quantify the impact of SRBI-deficiency on exRNAs *in vivo*, host sRNAs were compared at both the parent and fragment levels. For miRNAs, loss of SR-BI in mice did not alter miRNA content in liver, urine, bile, or APOB particles at the parent level, and only one miRNA (mmu-miR-143-3p, 0.199-fold, adjp= 0.00042) was significantly altered in SR-BI KO mice compared to WT mice (**Fig.9A, Table S13**), and the overall miRNA profiles were highly correlated between genotypes for all groups (**Fig.2D**). To determine if SR-BI regulates distinct features of miRNAs, 5' isomiRs or 3' NTA counts were compared between genotypes, and SR-BI-deficiency did not alter miRNA isomiRs or NTA counts for HDL, APOB, liver, or bile (**Figs.S2B-C**). Nevertheless, SR-BI may regulate urinary miRNA NTAs as SR-BI KO mice were found to have a significant increase in urinary miRNA NTAs ($p<0.001$) compared to WT mice (**Fig.S2C**). Moreover, we found a significant ($p=0.0021$) change in NTA-A/U ratios in urine from SR-BI KO mice compared to WT mice, as urine samples from WT mice were enriched for poly-adenylated miRNAs (NTA-A) and samples from SRBI KO mice were enriched for poly-uridylated miRNAs (NTA-U) (**Fig.2F**).

To determine if SR-BI regulates non-miRNA host sRNAs at the parent level, differential expression analyses were performed, and we identified a limited number of significantly altered host sRNAs by parent in SR-BI KO mice compared to WT mice (**Table S13**). For APOB, 1 tDR (tDR-AsnGTT, 14.09-fold, adjp=1.22E-02) and 3 miscRNAs (Vaultrc5, 6.02-fold, adjp=5.70E-05; lincDR-Malat1, 11.25-fold, adjp=4.61E-02) were significantly increased on APOB particles (**Fig.S25, Table S13**). Likewise, 1 snoDR (Snord22, 9.23-fold, adjp=2.40E-02) and 2 snDRs (Gm25587, 3.54-fold, adjp=2.92E-02; Gm22866, 3.81-fold, adjp=2.92E-02) were significantly increased on HDL from SR-BI KO mice compared to WT mice (**Figs.9,S25, Table S13**). For liver, 1 snDR (Snord64, 2.26-fold, adjp=2.23E-02) was significantly increased; and 1 snoDR (Gm22270, 0.39-fold, adjp=8.04E-03), 1 snDR (Gm23686, 0.43-fold, adjp=2.38E-02), and 1 miscRNA (lncDR-Gm26904, 0.31-fold, adjp=1.61E-02) were significantly decreased in SR-BI KO mice (**Figs.9,S25, Table S13).** For biofluids, 1 rDR (n-R5s2, 0.19-fold, adjp=3.28E-02) was found to be significantly decreased in urine samples and 1 snDR (Gm24621, 4.95-fold, 2.99E-02) was found to be significantly increased in bile from SR-BI KO mice compared to WT mice (**Figs.9,S25, Table S13**)**.**

To perform differential expression analysis at the parent level, many closely related sequences were grouped together to summarize parent RNA counts. Nonetheless, SR-BI may regulate sRNA flux between cells and extracellular carriers; therefore, the impact of SR-BI-deficiency on lipoprotein sRNAs may not be evident by grouping individual sRNAs. Therefore, to determine if SR-BI regulates lipoprotein sRNAs, or hepatic and biliary sRNAs, host sRNAs were filtered based on parent RNA mapping and then analyzed at the individual fragment level for differential expression. Strikingly, the abundance of many individual fragments were found to be significantly altered in SR-BI KO mice compared to WT mice and distinct patterns were detected (**Figs.9,S25**). For example, SR-BI-deficiency resulted in a significant decrease to 21 individual miRNA fragments (**Fig.9, Table S14**). Conversely, we found 57 snDR fragments that were significantly increased on HDL from SR-BI KO mice compared to WT mice; many of these fragments are very similar in sequence and length **(Fig.9, Table S14)**. Likewise, we

found 8 HDL-rDR fragments that were significantly increased in SR-BI KO mice (**Fig.9, Table S14**). In livers, we found 14 snDRs and 16 rDRs that were significantly increased at the fragment level; however, these were not identical sequences to fragments found to be decreased on HDL for these classes (**Fig.9, Table S14**). These results suggest that SR-BI may play a limited role in regulating sRNAs circulating on HDL and in livers. Nevertheless, these results strongly support the need to analyze host sRNAs not just at the parent level, but also the fragment level, as many critical observations may be lost in the grouping of similar sequences for parent analysis.

In addition to the liver, SR-BI is also expressed in the intestine and commensal bacteria likely regulate SR-BI, as both intestinal and hepatic SR-BI expression was reported to be increased in germ-free mice compared to control specific pathogen-free mice[30]. Nevertheless, SR-BI regulation the gut microbiome is unclear and the role of SR-BI in regulating circulating non-host bacterial sRNAs on lipoproteins is completely unknown. To determine if SR-BI contributes to exogenous sRNAs on lipoproteins and in biofluids, differential expression analysis was performed at both the genome and fragment levels. Only one bacterial species was found to be significantly altered in urine between SR-BI KO and WT mice, as determined by genome counts (**Fig.S26A, Table S16**). Likewise, only 3 individual sRNAs that aligned to bacterial genomes in the environment were significantly affected in by SR-BI-deficiency in mice; one each in APOB, bile, and urine samples (**Fig.S26B, Table S17**)**.** These results suggest that SR-BI does not likely regulate non-host bacterial sRNAs on lipoproteins or in biofluids. Conversely, we found that SR-BI-deficiency resulted in a significant increase in all fungal genome counts in SR-BI KO mice compared to WT mice (**Fig.S26A, Table S16**). These observations were not likely the result of a few reads that were shared across all fungal genomes as we failed to find any individual fungal sRNAs that were significantly affected by loss of SR-BI (**Fig.S26B**). To determine if SR-BI-deficiency in mice results in changes to the most abundant sequences in each group, independent of RNA class or genotype, differential expression analysis was performed for the top 100 reads filtered in the class-independent analysis, as described above. For APOB and HDL, 8-9 highly

abundant reads were found to be significantly altered in SR-BI KO mice compared to WT mice; however, we failed to find any significant changes in the expression of the most abundant reads in liver, bile, and urine samples (**Fig.S27, Table S17**).

Discussion

High-throughput sequencing of sRNAs has revealed a complex landscape of various types of sRNAs in cells and extracellular fluids, many of which have not been studied. Currently, there is a great need for tools that can extract novel sRNAs and distinct features from sequencing datasets. Previously, we have reported that HDL and LDL transport specific miRNAs, as quantified by real-time PCR-based TaqMan arrays[16]. Here, we used sRNA-seq and the TIGER pipeline to profile all sRNA classes on HDL and APOB particles and compared these profiles to liver, bile, and urine. Using this approach, we found that HDL and APOB particles transport a wide-variety of host sRNAs, including tDRs, rDRs, snDRs, and many other miscRNAs. Moreover, we found that exRNAs on lipoproteins harbored unique features, e.g. enrichment of poly-uridylation NTA events on miRNAs and discrete length distributions for HDL and APOB particles. Moreover, lipoproteins were found to transport a multitude of non-host sRNAs from exogenous bacterial and fungal species likely the microbiome and environment. Many of these non-host sRNAs were found to be likely processed from parent tRNAs and rRNAs. Using TIGER, we were also able to define each sample type by their most abundant sRNAs independent of class or species which is particularly suited for the study of exRNA. Furthermore, the TIGER pipeline allows for the quantification and differential expression analysis of sRNAs at both the parent and fragment levels. This strategy allowed our determination that SR-BI has a limited role in regulating cellular and extracellular sRNAs, which would not have been feasible with other analysis strategies focused solely on the parent RNA organization. Overall, this study demonstrates the power of expanding sRNA-seq analysis beyond canonical miRNAs and exploring the full breadth of host and non-host sRNAs in every dataset.

Although many researchers are using high-throughput sequencing to quantify sRNAs, many investigators do not take advantage of the enormous amount of information contained within sRNA-seq datasets. The mammalian transcriptome is immensely diverse and complex, and thus, requires new analytical tools and novel strategies to address the many distinct features of different sRNA classes and contributing species[7, 9, 31]. The TIGER pipeline is designed to incorporate both host and non-host sRNA analysis into a modular design that allows for custom prioritization and parallel alignments to both genomes and transcripts (libraries), and organizes data at the parent RNA, fragment, and class-independent levels. The 7 modules include preprocessing, host genome and database, non-host library, non-host genome, class-independent, summary, and unmapped. For host miRNAs, we expanded miRNA analysis to include 5' and 3' terminal isomiRs and 3' NTAs. Furthermore, we extended our analysis of annotated host sRNAs to include tDRs, rDRs, snDRs, snoDRs, lncDRs, and many other less studied classes, e.g. yDRs. A key feature of the TIGER pipeline is the alignment strategy for host tDRs and rDRs which includes mapping to the host genome and mature transcripts in corresponding databases, which overcomes specific issues, e.g. introns[32, 33]. Another key advance in our pipeline is the parallel analysis of host sRNAs at the parent and individual fragment levels. Organization of sRNAs at the parent level allows for categorical analysis and positional coverage alignments which provides information on parent RNA processing (cleavage). Conversely, analysis of sRNAs at the individual sequence (fragment) level aids biomarker discovery and is critical to determining biological functions. Collectively, these features represent a substantial advance for the analysis of endogenous host sRNAs.

The TIGER pipeline allowed for extensive comparisons between lipoproteins, biofluids, and liver samples across many different levels and features. We found many examples where sRNAs on lipoproteins were different than sRNAs in liver, bile, or urine. For example, host sRNAs on lipoproteins differ dramatically from liver sRNAs based on class, as demonstrated by hierarchical clustering and PCoA plots for every RNA class at the individual fragment level. Notably, lipoproteins are also distinct

from liver profiles independent of class organization, as observed within the class-independent module in the TIGER pipeline. Nonetheless, the breakdown of sRNA classes on lipoproteins is different than liver, bile, and urine. For example, although lipoproteins transport miRNAs, the fraction of miRNAs per total reads (2.2% APOB, 6.2% HDL) was less than what was observed for liver samples (16.2%). The distribution of sRNA lengths within each class likely contributes to the overall differences between sRNAs in lipoproteins, livers, and biofluids. Lipoproteins were found to transport tDRs 30-36 nts in length, likely tRHs, which were only a minor fraction of total reads in liver samples. Conversely, sRNAs in bile and urine sRNA were predominantly tDRs. All sample types contained rDRs; however, liver rDRs were primarily 42-46 nts in length, whereas rDRs in lipoproteins and biofluids were variable in length. These results suggest that rRNA processing in the liver may be highly-regulated to produce these specific length fragments. Another observation was that liver samples contained snoDRs >50 nts in length which were largely absent from HDL or APOB profiles. These results suggest that snoDRs are likely retained in cells and not exported to lipoproteins, bile, or urine. In addition to differences in sRNA classes and length, liver samples were found to have more miRNA 5' isomiRs than APOB samples, but less 3' NTA events. Moreover, liver miRNAs were heavily poly-adenylated (NTA-A), where miRNAs on HDL and APOB were enriched for poly-uridylation (NTA-U) events. These results suggest that miRNAs on HDL and APOB particles are distinct from liver miRNAs and may be evidence of partitioning miRNAs between cellular and extracellular pools based on NTAs. In comparing lipoproteins to other sample types, we found evidence that suggests studying extracellular non-miRNAs may be more challenging than previous research into miRNAs, as non-miRNAs displayed large variability in sequences on lipoproteins and biofluids. For example, all samples within groups were highly correlated when compared at the parent RNA level; however, there is only a limited number of parent RNAs to align individual sRNAs to for each RNA class. When samples were compared for individual fragments, we found low correlations amongst samples within each group for lipoproteins and biofluids. This was not the case for liver where all RNA classes were highly correlated at both the parent and fragment levels. In comparison, miRNAs were modestly to highly correlated in both lipoprotein and biofluid

datasets at the parent and fragment levels. These results are likely due to precise cleavage and processing for miRNAs and less precision for non-miRNA sRNAs. The observed high variability of sequences for exRNAs, i.e. sRNAs on lipoproteins and in biofluids, may be due to the contribution of many heterogeneous cell-types to these compartments. Moreover, sRNAs on lipoproteins may encounter further nucleotide hydrolysis from circulating RNases, thus producing variable sRNA sequences (lengths) between samples. Most interestingly, sRNA length was a key difference between HDL and APOB samples, as shorter length fragments were enriched on HDL compared to APOB particles for tDRs and rDRs. This may be due to differences in lipoprotein particle sizes -- HDL is smaller in diameter (12 nm) than APOB particles (22-70 nm) -- and particle size may confer protection from ribonucleases. The unique distribution of lengths between HDL and APOB particles was also evident for non-host bacterial and fungal sRNAs.

A critical difference between cellular RNA and exRNA profiles is the presence of non-host sRNAs present in exRNA samples[12, 34, 35]. ExRNAs hold great potential as disease biomarkers, indicators of specific cell phenotypes and damage, intercellular communication signals, and drug targets for future therapies[36-38]. Current sRNA-seq analysis pipelines are not particularly suitable for the study of exRNAs as many are restricted to only canonical miRNAs, or a limited number of host sRNAs, and lack analysis of non-host sRNAs which will likely be a major focus of future investigations. Based on a previous study reporting that bacterial sRNAs are present in human plasma, the TIGER pipeline was designed to identify exogenous bacterial and fungal sRNAs. Strikingly, we found that the majority of sRNAs on HDL and APOB particles are likely from bacteria present in the microbiome and environment. This was achieved through mapping non-host reads to bacterial and fungal genomes, as well as to mature transcripts of non-host miRNAs, tDRs, and rDRs across all kingdoms. Exogenous bacterial and fungal sRNAs on lipoproteins are not likely contamination products due to several observations. First, we were not able to detect candidate bacterial sRNAs in control buffer used to isolate the lipoproteins by real-time PCR. Moreover, reads aligning to bacterial and fungal genomes were not likely contamination of

reagents used for sequencing preparation as most of these reads were not present in liver or bile datasets. Next, we found very low correlation between lipoprotein samples for non-host bacterial and fungal sRNAs suggesting that there was not a common source of bacterial or fungal RNA in the preparation reagents. In addition, we found that bacterial and fungal sRNAs on HDL were enriched for short length sRNAs as compared to APOB particles, a pattern that was also observed for host sRNAs, thus supporting a common mechanism of loading or association for sRNAs that is different for HDL and APOB particles. Moreover, we found that non-host bacterial sRNA profiles were distinct for HDL and APOB as demonstrated by PCoA and PERMANOVA. Collectively, these results strongly support that HDL and APOB particles transport distinct sets of exogenous (non-host) sRNAs that are not likely due to bacterial and fungal contamination or foreign RNA in reagents or the research environment. Currently, the biological functions of non-host sRNAs in host circulation are unknown; however, research into potential cross-kingdom gene regulation has been proposed, albeit met with controversy. For example, sRNA-seq has been applied to study dietary miRNAs (xenomiRs), exogenous miRNAs absorbed through the gut and proposed to regulate host gene expression and phenotype[39]. Studies from at least two independent labs have provided evidence for such cross-kingdom gene regulatory networks[39-41]; however, other groups have argued against these claims[42-45]. Most of the investigation into exogenous sRNAs regulating host gene expression is limited to known plant or animal miRNAs, and lost in this controversy is the potential for non-miRNA sRNAs mediating cross-kingdom gene regulation. In fact, there have been reports of exogenous non-miRNA sRNAs stably circulating in human plasma that are likely derived from organisms across several kingdoms[12, 46, 47]. If exogenous sRNAs, e.g. bacterial tDRs and rDRs, do indeed regulate gene expression through post-transcriptional or other mechanisms in human cells, this would represent an important link between humans and their environment at the gene regulation level. Moreover, this would suggest that the initial studies of dietary xenomiRs are thus, very limited in their scope by only studying annotated miRNAs. Nevertheless, cross-kingdom gene regulation is not a new concept as plants, bacteria, and fungi have been

extensively studied for their ability to utilize both miRNAs and non-miRNA sRNAs to regulate gene expression in cross-kingdom networks[48-50].

The inclusion of non-host reads in our analysis greatly increased our ability to account for reads in lipoprotein datasets. Nevertheless, there are many exogenous sRNAs in biological fluids that are neither processed from annotated transcripts in databases nor originate from species currently represented in the HMB project. Therefore, another key feature of the TIGER pipeline is the ability to analyze data independent of species identification or library annotation. This is critically important for the study of exRNAs as non-host sRNAs in these samples are very diverse and many of the most abundant sRNAs are not processed from transcripts in RNA databases or originate from species currently collected in the HMB database. As such, class-independent analysis extracts more data and eliminates a major barrier to the discovery of biomarkers and intercellular communication signals. Notably, class-independent analysis of exRNAs captures sRNA sequence, length, and abundance which are the important defining characteristics of biomarkers in extracellular fluids and bioactivity in recipient cells. The TIGER pipeline also advances sRNA-seq analysis through the incorporation of high-end comparative analyses and data visualizations, including PCoA, PERMANOVA, hierarchical clustering and correlations, positional coverage maps, circular tree maps, circos linkage maps, and ternary plots. The TIGER pipeline addresses many issues in sRNA-seq analysis; however, we have identified a few limitations to the software. Although the TIGER pipeline is designed to quantify 5' and 3' variants, it does not currently identify internal modifications, ADAR editing events, or single nucleotide polymorphisms. This feature would aid in the study of tDRs, which are heavily modified, and would potentially improve analysis of non-host bacterial sRNAs where reference genomes may be lacking. The ability to quantify internal variance is a key feature of Chimira, as well as other software, including UEA workbench[51], and MAGI[52]. Furthermore, the TIGER pipeline does not include the analysis of PIWI-Interacting RNAs (piRNA) and a few other sRNAs, including promoter-associated sRNAs, which present unique challenges in alignments, quantification, and nomenclature[53]. Future versions of the

pipeline will include less studied sRNA classes and the ability to discover new host sRNAs, as the current pipeline does not have the feature to identify novel miRNAs based on adjacent genomic sequences which is an output of other pipelines[54, 55]. Despite these limitations, the TIGER pipeline sets forth many improvements to sRNA-seq analysis.

Here, we demonstrated the advanced features of the TIGER pipeline for sRNA-seq analysis to determine SR-BI's regulatory role of exRNA. SR-BI is a bidirectional transporter of cholesterol and HDL's primary receptor[56]. Furthermore, SR-BI is also a strong contributor of biliary cholesterol secretion[27]. The ability of the TIGER pipeline to analyze sRNAs at the parent and fragment levels allowed for us to determine that the impact of SR-BI-deficiency occurs at the fragment level, but not the parent level. However, due to the number of individual reads, the chances of observing false positives are much greater at the fragment level than the parent level. Results suggest that SR-BI likely does not regulate the levels of extracellular miRNAs circulating on lipoproteins, in liver or bile when organized at the parent level. Nevertheless, loss of SR-BI resulted in a significant decrease of individual fragments that aligned to miRNA coordinates. SR-BI-deficiency in mice also resulted in a significant increase in fragments that aligned to parent snRNAs and rRNAs; however, these changes were not inversely altered in liver or other sample types suggesting that these changes on HDL were not likely due to inhibition of a potential systemic clearance mechanism for lipoprotein sRNAs, e.g. HDL-liver-bile. Most interestingly, SR-BI-deficiency in mice resulted in changes to miRNAs in urine. We found that urine samples from SR-BI KO mice contained a significant increase in miRNA 3' NTA events and a concomitant increase in NTA-U/A ratio in absence of a significant increase in total miRNA counts. These results suggest that SR-BI specifically inhibits the secretion of miRNAs harboring 3' poly-uridylation to urine. To determine if SR-BI regulates exogenous sRNAs on lipoproteins, liver or biofluids, differential expression analyses were performed for the different non-host sRNA groupings and features. A previous report has demonstrated that SR-BI expression in the intestine and liver may be inhibited be commensal bacteria, as SR-BI expression was demonstrated to be increased in germ-

free mice[30]; however, it is currently unknown if SR-BI regulates bacteria in the microbiome. Results from our study suggest that SR-BI does not regulate bacterial sRNAs on lipoproteins, liver, or biofluids. Nevertheless, SR-BI may regulate fungal sRNAs in bile, as genome counts for all fungal species analyzed were significantly increased in bile from SR-BI KO mice compared to WT mice.

In summary, the value of any sequencing data analysis pipeline, ultimately, is the ability to extract the most useable information from the generated data; therefore, the goal of the TIGER pipeline was to assess both host and non-host sRNAs which greatly improved the ability to account for more reads in our sRNA-seq datasets, particularly exRNAs. The TIGER pipeline also advances the field in its ability to analyze host sRNAs at the parent and fragment levels and non-host sRNAs at the genome and fragment levels. This approach may be critical to discovering novel biomarkers and intercellular communication signals that would be masked by analyzing the sRNAs by their parent RNAs and nomenclature. Likewise, the TIGER pipeline analyzes sRNAs by class and species (genome) as well as class-independent approaches. This is very important for exRNAs where the contributing exogenous species for sRNAs may not be curated in bacterial or fungal genome databases, or the contributing parent RNA may not be annotated for the host genome. The TIGER pipeline is particularly suited for lipoprotein sRNAs which are predominantly rRNA-derived fragments of bacterial origin. Using TIGER, we were able to make critical observations comparing lipoprotein sRNAs to liver and biofluids that would not be observed by existing pipelines. Therefore, this tool is well-suited for the analysis of exRNA.

Materials and Methods:

Animal Studies: Plasma, basal bile, urine, and livers were collected from wild-type (WT) and SR-BI-deficient (*B6;129S2-Scarb1tm1Kri/J,* SR-BI KO) mice, as previously described[57]. Mice were anesthetized with urethane (1g/kg, i.p.). The common bile duct was ligated and the gall bladder

cannulated to divert bile into collection tubes. Basal bile was collected for a period of 30 min. Mice were then exsanguinated, blood was collected from the abdominal aorta in EDTA coated tubes and placed on wet ice, and tissues were dissected and snap frozen in liquid nitrogen. Plasma and tissues were stored at -80^{o}C prior to analysis. All animal procedures were completed under active and approved IACUC protocols.

Lipoprotein isolation: To separate HDL and apolipoprotein B (APOB)-containing lipoproteins from mouse plasma, 200 µL of 0.22-µm filtered-plasma samples were diluted to 500 µL in size-exclusion chromatography (SEC) running buffer (10 mM Tris-HCl, 0.15 M NaCl, 0.2% NaN_3) and injected an ÄKTA SEC system (GE Healthcare) with three in-series Superdex-200 Increase gel filtration columns (10/300 GL; GE Healthcare). Samples were applied to the column with a flow rate of 0.3 mL/min at room temperature and eluate collected as 72 x 1.5 mL fractions using a F9-C 96-well plate fraction collector (GE Healthcare). Each fraction was analyzed for total protein (BCA; Pierce), total cholesterol (Raichem), and triglycerides (Raichem) to identify fractions corresponding with HDL and APOB particles. Due to the SEC set-up, we were not able to separate VLDL from LDL particles, and thus, we collected fractions covering both lipoprotein classes, referred to here as APOB. Fractions corresponding with each lipoprotein group were pooled, concentrated with Amicon Ultra-4 10 kDa centrifugal filters (Millipore) to <200 µL volume, and protein concentrations were quantified by BCA assays (Pierce). Based on the distribution of total cholesterol, triglycerides, and protein, fractions corresponding to HDL and APOB were collected, pooled, and concentrated.

RNA Isolation: To differentiate lipoprotein sRNA signatures from liver and biofluids, and determine the impact of SR-BI-deficiency, samples were collected from *Scarb1*$^{-/-}$ (SR-BI KO) and wild-type (WT) mice. Total RNA was extracted from HDL (WT N=7, SR-BI KO N=7) and APOB (WT N=7, SR-BI KO N=7)

particles, as well as livers (WT N=7, SR-BI KO N=7), bile (WT N=7, SR-BI KO N=6), and urine (WT N=5, SR-BI KO N=6). RNA was isolated from equal inputs of either bile (volume), liver (mg), HDL (protein) or APOB (protein) using miRNAEasy Mini kits (Qiagen). Specifically, 30 µL of primary bile, 120 µg of APOB, 180 µg of HDL or 20 mg of liver were added to 1 mL of Qiazol. Livers were homogenized in Qiazol with High-Impact Zirconium beads using a Bead Bug Homogenizer (Benchmark Scientific). After removal of beads, subsequent steps for liver RNA extraction were followed according to manufacturer's protocol. Bile, APOB and HDL RNA isolations were processed according to manufacturer's protocol, except that after addition of ethanol, samples were incubated at -80^{o}C overnight before application to isolation columns, and were eluted with a volume of 50 µL. Liver RNA samples were quantified by Take3 plates (BioTek).

Real-Time PCR: Total RNA from equimolar amounts of HDL or APOB protein and equivolume amounts of bile or urine samples were diluted 1:10; 50 ng of total RNA from liver was used for reverse transcription using either miRCURY LNA universal RT kit (Exiqon) or TaqMan miRNA Reverse Transcription kit, as per manufacturer's instructions. Real-time PCR was performed with the QuantStudio 12K Flex Real-Time PCR System (Life Technologies) using either: A) miRCURY LNA SYBR Green PCR kit (Exiqon) and either miRNA-specific or custom-sequence specific LNA probes (Exiqon; Table S19) or B) TaqMan miRNA-specific probes. Relative quantitative values (RQV) were determined for both HDL and cellular miRNA analyses. RQV = 2^{-dCt}. For HDL, APOB, bile, and urine samples, an arbitrary housekeeping Ct = 32 was applied, and RQVs for liver sRNAs were normalized by U6.

Small RNA sequencing: NEXTflex Small RNA Library Preparation Kits v3 for Illumina® Platforms (BioO Scientific) were used to generate cDNA libraries for sRNA-seq. Briefly, 1 µg of liver total RNA was used as input for adapter ligation, as per manufacturer's protocol. For bile, APOB and HDL RNA, 10.5 µL

(21%) of the RNA isolation eluate was used as input for adapter ligation. Library generation was performed according to manufacturer's protocol (BioO Scientific) with a modification to the amplification step, as liver libraries received 18 cycles and bile, APOB and HDL libraries received 27 cycles. After amplification, samples were size-selected using a Pippin-Prep (Sage Science) -- set for a range of 135-200 nts in length -- and subsequently purified and concentrated using DNA Clean and Concentrator 5 kit (Zymo). Individual libraries were then screened for quality by High-Sensitivity DNA chips using a 2100 Bioanalyzer (Agilent) and quantified using High-Sensitivity DNA assays with Qubit (Life Technologies). Equal concentrations of all individual libraries were pooled for multiplex sequencing runs, and concentrated using DNA Clean and Concentrator 5 kit (Zymo). For rigor in down-stream comparisons, all 66 sequencing libraries were randomized and run independently on three individual sequencing lanes. Single-end sequencing (75 cycles) of multiplexed libraries were performed on an Illumina NextSEQ 500 at the Vanderbilt Technologies for Advanced Genomics (VANTAGE) core laboratory. Each library was sequenced at an average depth of 16.28 million reads/sample.

Data analysis: The TIGER pipeline has many unique analysis features built into seven modules for low-level and high-level analyses with data visualization packages. The first module contains pre-processing steps (green) prior to data analysis (**Fig.1**). To assess raw data quality, FastQC was performed at the raw read level to check for base quality, total read counts, and adapter identification. Cutadapt was then used to trim 3' adapters from processed reads (-a TGGAATTCTCGGGTGCCAAGG). Although this pipeline can analyze sRNA-seq data prepared by different library generation methods, TIGER was optimized to analyze sRNA-seq data prepared by ligation of adapters containing 4 terminal degenerate bases, which reduce ligation bias (e.g. BioO Scientific NEXTflex Small RNA-seq kit v3). Cutadapt was then used to remove the first and last 4 bases from the trimmed reads and all trimmed reads <16 nts in length were removed (-m 16 -u 4 -u -4). After trimming, read length distributions were plotted and FastQC was performed on trimmed reads to

validate the efficiency of adapter trimming. The processed reads were then summarized and plotted. To generate identical read files, trimmed reads in each sample were collapsed into non-redundant "identical" reads in FASTQ format and copy numbers were recorded for downstream analysis. Preprocessed reads were then analyzed by the Host Genome & Database (blue) and Class-Independent (red) modules in parallel (**Fig.1**). In the Host Genome & Database alignment module (blue), bowtie (v1.1.2) was used to map reads to a costumed database with option (-a -m 100 --best -strata -v 1) which allows 1 mismatch (MM) and 100 multi-mapped loci, and only the best matches were recorded. The costumed database was constructed by the host genome and known sequences of host mature transcripts curated in specific library databases – tRNAs (http://gtrnadb.ucsc.edu/GtRNAdb2/) and rRNA (http://archive.broadinstitute.org/cancer/cga/rnaseqc_download). A small number of parent tRNA genes contain introns and the mature transcript differs from the genomic sequence; therefore, the incorporation of mature tRNA transcripts from GtRNAdb database into the genomic alignment overcame these limitations. This approach allows for the detection of tDRs spanning exon junctions and allows reads the chance to be mapped to other non-tRNA loci in the genome with best alignment score which reduces the false positive tDR reads that can result from database only alignment strategies. Counting and differential expression analysis of miRNAs, tDRs, rDRs, snDRs, snoDRs, and other miscellaneous sRNAs (miscRNA), including yDRs and lincDRs, were performed. The pipeline does not quantify Piwi-interacting RNA (piRNA) or circular RNAs (circRNA), but this function can be amended. All prepossessed quality reads were assigned to different classes of annotated sRNAs using distinct rules -- miRNA: 1 MM, ≥16nt, offset -2, -1, 0, 1, 2 and tDR, snDRs, snoDRs, yDRs, and miscRNAs: 1 MM, ≥16nt, overlap ≥0.9 overlap. Based on the extensive genomic coverage of lncRNAs and repetitive elements and conservation of rRNAs, the TIGER pipeline applies more stringent assignment rules for lncDRs and rDRs – perfect match, ≥20 nt, and ≥90% overlap with parent lncRNAs or rRNAs. Furthermore, reads assigned to lncDRs must only be aligned to lncRNA coordinates and not to any other loci in the genome. All reads ≥20 nts in length and not aligned to the costumed database were extracted and tested for alignment as non-host reads. After tabulation of read counts, high-end

analyses were performed on host sRNAs. These include principal component analysis, hierarchical clustering and correlation of samples and groups at the parent and individual fragment levels. Differential expression of tabulated read counts were performed by DEseq2[58]. Differential expression result was plotted as volcano plot, venn diagram, and heatmap. Categorical analyses of tDRs based on amino acid and anti-codons of the parent tRNAs were also quantified and plotted. Likewise, categorical analysis of snDRs based on U class were analyzed and plotted. In addition, miRNAs were analyzed at the canonical, isomiR, NTA, NTA base, and isomiR NTA levels. Non-host reads were then analyzed using the Non-Host Genome (Purple) and Non-Host Library (Gold) modules in parallel (**Fig.1**). In the Non-Host Genome module, reads were aligned in parallel to two collections of bacterial genomes: a human microbiome (HMB) collection and a hand-curated list of environmental bacteria observed during sequencing of human and mouse lipoproteins. The HMB list was compiled by reducing 3,055 bacterial genomes available from the Human Microbiome Project (www.hmpdacc.org) to single non-redundant genera, and extracting the largest available complete genome for each genera. Conversely, to generate the environmental bacteria list, the top 100 most abundant sequences in a control HDL cohort, that were not mapped to the host genome, were submitted to NCBI for BLASTn. All hits that showed 100% coverage and 100% identity were then compiled; non-redundant genera were extracted; redundant genera to the HMB were removed. Representative genomes from the remaining species were then compiled to the environmental bacteria list (ENV). Additionally, a small group of fungal genomes associated with the human pathology were also collected. The HMB, ENV, and fungal modules contain 206, 167, 8 representative genomes, respectively. Due to high conservation between bacterial genomes and multi-mapping issues, a different bowtie option (-a -m 1000 --best -strata -v 0) was used which allowed perfect match only and 1000 multi-mapped loci. Reads were aligned to the HMB, ENV, and fungal groups in parallel and, thus, the same reads could have been counted in multiple groups. The fraction of reads that align to both databases (HMB, ENV) and the reads that are unique to specific databases were plotted. Differential expression and high-end analyses, as described above, were performed at the genome level (total normalized read count for each genome) and at the individual read

level. In parallel, non-host reads were also analyzed by the Non-Host Library (Gold) module where they were aligned to non-coding RNA databases with same bowtie option as non-host genome analysis. To identify possible non-host miRNAs (xenomiRs) in sRNA-seq datasets, all non-host reads were aligned perfectly to annotated miRNAs in miRBase (miRBase.org) and tabulated. Similarly, non-host reads were aligned to all tRNAs in the GtRNAdb database (GtRNAdb2). Extensive categorical analysis of parent non-host tRNAs were performed at the kingdom, genome (species), amino acid, anti-codon, and fragment (read) levels. All assigned non-host tDRs underwent differential expression analysis, high-end analysis, and data visualization, as described above. Non-host reads were also aligned to prokaryotic and eukaryotic rRNA transcripts in SILVA database (https://www.arb-silva.de). TIGER limits the analysis of non-host rDR to the kingdom level for counting, differential expression analysis and high-end analysis.

The TIGER pipeline also analyzed the top most abundant reads independent of class or annotation in parallel of the host genome, non-host genome, and database modules. The Class-Independent module (red) ranked and filtered the top 100 most abundant reads in each sample independent of genomic annotation. The list of top 100 reads from all samples were combined, a count file table was generated and top 100 overall reads were used to perform hierarchical clustering and correlations at the individual sample and group levels. Differential expression analyses were performed by DEseq2, and significantly altered sequences were searched in NCBI nucleotide database using BLASTn to identify possible sources (species). All results from the host genome, class-independent, non-host genome, and non-host database modules were then analyzed by the Summary & Data Visualization (dark blue) module (**Fig.1**). In this module, TIGER summarized and organized many of the individual comparisons. For example, individual volcano plots were graphed into larger matrices grouping different classes of sRNAs and/or genomic groups (e.g. bacteria and fungi). This module also generated a comprehensive table for all mapped reads listing the assignments for each read across modules. Moreover, positional coverage of sRNAs against host parent RNAs were plotted for miRNAs,

tDRs, snDRs, and rDRs. Positional base coverages were also plotted for individual samples, groups, and significantly altered tDRs and snDRs. For groups, the means of normalized positional coverage counts (base positional counts per million mapped total reads) for individual samples in the groups were plotted. Furthermore, this module identified sRNA classes and genomes for the top 100 ranked reads (analyzed earlier in the Class-Independent module) and graphed the linkages by circos plots. Finally, this module summarized the read counts in each task and determined the fraction of total reads that were assigned to any module, genome, or database. For example, pie charts and stacked bar charts were generated to illustrate the fraction of reads mapped to the host genome and non-host genome and the fraction of unmapped reads. All unmapped and unaccounted for reads entered the Final Unmapped Analysis (orange) module (**Fig.1**). In this module, the top 100 analysis was reapplied to all unmapped and unaccounted reads, as described above. After ranking, filtering, and tabulation, differential expression analysis was performed and the significantly altered unmapped reads were searched in BLASTn to identify possible genomes not contained in the TIGER analysis. These unique features were designed to extensively and exhaustively analyze sRNA-seq data on lipoproteins (e.g. HDL and apoB particles) and extracellular fluids (e.g. bile and urine) which have many different types of sRNAs and diverse species.

Data Visualization: Read counts were reported as both raw counts and normalized count per million total counts (RPM). RPMs were used for stacked bar plots in each module. Cluster analysis by heatmap3[59]. Principle component analysis were performed based on normalized expression value calculated by the variance stabilizing transformation in DESeq2. DESeq2 was used to perform miRNA, tDR and other sRNA differential expression analyses. Significantly differential expressed sRNAs with adjusted FDR less than 0.05 and absolute fold change larger than 1.5 will be highlighted in volcano plot (red, increased; blue, decreased) and outputted as tabulated file for further validation. Non-metric multidimensional scaling of Bray-Curtis dissimilarity indexes, homogeneity analysis of group

dispersions, and principal coordinate analysis visualization was performed using R package "vegan". R Packages ggplot2, vegan, ggraph, igraph, reshape2, data.table, RColorBrewer, circlize, ggtern, and XML were used for data visualization.

Statistics: For continuous variables, mean and standard error of the mean (S.E.M.) were used. Comparisons with two variables were calculated using Welch two sample t-tests, two-way Student's t-tests, or Mann-Whitney nonparametric tests. For comparisons with more than two variables, linear one-way analysis of variance (ANOVA) were used. Spearman ranked method was used for calculating the correlation coefficient (R). Two-sided p values ≤0.05 were considered statistically significant. Statistical analyses were performed using R version 3.4.3.

Acknowledgments: We would like to acknowledge Carolin Besenboeck for her assistance in RNA methods.

Funding: This work was supported by awards from the National Institutes of Health, National Heart, Lung and Blood Institute to K.C.V. HL128996, K.C.V. and M.F.L. HL127173, and M.F.L. HL116263. This work was also supported by awards from the American Heart Association to K.C.V. and P.S. CSA2066001 and R.M.A. POST25710170.

Figure Legends:

Figure 1. Schematic of the TIGER sRNA-seq analysis workflow.

Figure 2. Host miRNAs on lipoproteins have distinct features compared to liver. WT, wild-type mice; SR-BI KO, Scavenger receptor BI Knockout mice (*Scarb1*$^{-/-}$). (**A**) Correlation of sRNA-seq reads per million total reads (RPM, blue) and miRNA reads (RPM miR, gray) to real-time PCR relative quantitative values (RQV). Spearman correlation. HDL, APOB, liver, bile, and urine samples, N=66. (**B-F**) Results from sRNA-seq analysis. HDL WT, N=7; HDL SR-BI KO N=7; APOB WT, N=7, APOB SR-BI KO N=7; Liver WT, N=7; Liver SR-BI KO, N=7; Bile WT, N=7; Bile SR-BI KO, N=6; Urine WT, N=5; Urine SR-BI KO, N=6. (**B**) Abundance of canonical miRNAs. Mean ±S.E.M. (**C**) Principal Coordinate Analysis (PCoA) of canonical miRNA profiles for samples from WT (empty circles) and SR-BI KO (filled circles) mice. NMDS1, Non-metric multidimensional scaling. (**D**) Heatmap of hierarchical clustered pairwise correlation (Spearman, R) coefficients between group means for canonical miRNAs. (**E**) Start position analysis of 5' miRNA variants (isomiR) for combined (WT and SR-BI KO) mouse samples. (**F**) Ratio of non-templated U (poly-uridylation) to A (poly-adenylation) for miRNAs. Mean ±S.E.M. One-way ANOVA. *p<0.05; **p<0.01

Figure 3. Host sRNAs account for a minor fraction of total reads in lipoprotein sRNA-seq datasets. WT, wild-type mice; SR-BI KO, Scavenger receptor BI Knockout mice (*Scarb1*$^{-/-}$). (**A-F**) Results from sRNA-seq analysis. HDL WT, N=7; HDL SR-BI KO N=7; APOB WT, N=7, APOB SR-BI KO N=7; Liver WT, N=7; Liver SR-BI KO, N=7; Bile WT, N=7; Bile SR-BI KO, N=6; Urine WT, N=5; Urine SR-BI KO, N=6. Host tDRs (yellow), rDRs (red), miRNAs (blue), snoDRs (purple), snDRs (green), miscellaneous RNA (pink), and unannotated genome (black). (**A**) Percent of total reads for host sRNA classes. Mean ±S.E.M. (**B-F**) Distribution of read length by host sRNA classes (colors) and total reads (gray), as reported by percent of total reads. Mean ±S.E.M. (**B**) Liver. (**C**) APOB particles. (**D**) HDL. (**E**) Bile. (**F**) Urine.

Figure 4. Lipoproteins, bile, and urine contain distinct tDR profiles. WT, wild-type mice; SR-BI KO, Scavenger receptor BI Knockout mice (*Scarb1*$^{-/-}$). (**A-C,F**) Results from sRNA-seq analysis. (**A**)

Positional coverage maps of tDRs for parent tRNA amino acid anti-codons, as reported as mean cumulative read fractions (read counts / total counts). (**B-C**) Principal Coordinate Analysis (PCoA) of tDR profiles based on (**B**) parent tRNAs and (**C**) individual fragments for samples from WT (empty circles) and SR-BI KO (filled circles) mice. NMDS1, Non-metric multidimensional scaling. (**D-F**) Real-time PCR analysis of candidate tDRs with predicted folding structures and sequences for (**D**) tDR-GluCTC and (**E**) tDR-GlyGCC. WT, white circles; SR-BI KO, red circles. (**F**) Heatmaps of correlation coefficients (Spearman, R) for tRNA parents and individual tDR fragments across samples within each group. HDL WT, N=7; HDL SR-BI KO N=7; APOB WT, N=7, APOB SR-BI KO N=7; Liver WT, N=7; Liver SR-BI KO, N=7; Bile WT, N=7; Bile SR-BI KO, N=6; Urine WT, N=5; Urine SR-BI KO, N=6.

Figure 5. Lipoproteins transport exogenous non-host tDRs and rDRs. WT, wild-type mice; SR-BI KO, Scavenger receptor BI Knockout mice (*Scarb1*$^{-/-}$). (**A**) Stacked bar plots of non-host tDRs aligned to parent tRNAs across kingdoms and higher organizations – bacteria, blue; eukaryota, yellow; fungi, red; embryophyta, orange; vertebrata, purple; archaea, green – as reported as percent of total reads. (**B**) Positional coverage maps of non-host tDRs for parent tRNA amino acid anti-codons, as reported as mean cumulative read fractions (read counts / total counts) for HDL and APOB particles. (**C**) Stacked bar plots of non-host rDRs aligned to parent rRNAs across kingdoms and higher organizations – bacteria, yellow; eukaryota, red; fungi, white; protists, purple; archaeplastida, dark blue; embryophyta, light blue; archaea, green – as reported as percent of total reads. (**D-F**) Distribution of read lengths, as reported as percent of total reads, for all non-host (**D**) tDRs and (**F**) rDRs. Two-tailed Student's t-tests. *$p<0.05$. HDL WT, N=7; HDL SR-BI KO N=7; APOB WT, N=7, APOB SR-BI KO N=7; Liver WT, N=7; Liver SR-BI KO, N=7; Bile WT, N=7; Bile SR-BI KO, N=6; Urine WT, N=5; Urine SR-BI KO, N=6.

Figure 6. Lipoproteins are enriched for sRNAs derived from proteobacteria in the microbiome and environment. WT, wild-type mice; SR-BI KO, Scavenger receptor BI Knockout mice (*Scarb1*$^{-/-}$). (**A**) Circular tree maps for non-host bacterial sRNAs on HDL from WT mice, as organized by taxonomy

– proteobacteria, green; actinobacteria, blue; firmicutes, yellow; bacteroidetes, red. Diameter is proportional to the mean number of reads at the genome level (counts). (**B**-**C**) Heatmaps of correlation coefficients (Spearman, R) for non-host sRNAs (on HDL and APOB particles) for bacterial genomes and individual bacterial fragments across samples grouped by (**B**) human microbiome (HMB) and (**C**) environment (ENV) species. (**D**-**E**) Distribution of read lengths, as reported as percent of total reads, for non-host bacterial sRNAs grouped by (**D**) HMB and (**E**) ENV species. Two-tailed Student's t-tests. *p<0.05. (**F**-**G**) Real-time PCR analysis of candidate non-host bacterial sRNAs for (**F**) exogenous rDR Pseudomonas fluorescens 23S (exo_rDR_Pflo23S) and (**G**) exogenous rDR Janthinobacterium lividum 23S (exo_rDR_Jliv). WT, white circles; SR-BI KO, red circles. HDL WT, N=7; HDL SR-BI KO N=7; APOB WT, N=7, APOB SR-BI KO N=7; Liver WT, N=7; Liver SR-BI KO, N=7; Bile WT, N=7; Bile SR-BI KO, N=6; Urine WT, N=5; Urine SR-BI KO, N=6.

Figure 7. The most abundant sRNAs on lipoproteins are bacterial rDRs. (**A-E**) Circos plots linking the most abundant (top 100) sequences to assigned groups for non-host libraries (rRNA lib, tRNA lib), host sRNAs (rDR, osRNA, tDRs, snDRs, snoDRs, miRNAs) and non-host genomes (fungi, environment, and microbiome) for (**A**) liver, (**B**) APOB, (**C**) HDL, (**D**) urine, and (**E**) bile. (**F**) Principal Coordinate Analysis (PCoA) of sRNA profiles based on class-independent analyses. Wild-type mice, WT (open circles); Scavenger receptor BI Knockout mice (*Scarb1*$^{-/-}$), SR-BI KO (filled circles). HDL WT, N=7; HDL SR-BI KO N=7; APOB WT, N=7, APOB SR-BI KO N=7; Liver WT, N=7; Liver SR-BI KO, N=7; Bile WT, N=7; Bile SR-BI KO, N=6; Urine WT, N=5; Urine SR-BI KO, N=6.

Figure 8. TIGER analysis pipeline accounts for significantly more reads than other software for lipoprotein sRNA-seq data. (**A-B**) Ternary plots of sRNA profiles for all samples displayed as (**A**) percent unexplained (blue), miRNAs (green), and non-miRNA host sRNAs (red); (**B**) percent unexplained (blue), exogenous sRNAs (green), and host genome (red). WT, wild-type mice; SR-BI KO,

Scavenger receptor BI Knockout mice (*Scarb1*$^{-/-}$). (**C**) Pie charts illustrating the mean fraction of reads assigned to host sRNA (red), host genome (blue), non-host (purple), too short for mapping (green), and unmapped (orange). HDL WT, N=7; HDL SR-BI KO N=7; APOB WT, N=7, APOB SR-BI KO N=7; Liver WT, N=7; Liver SR-BI KO, N=7; Bile WT, N=7; Bile SR-BI KO, N=6; Urine WT, N=5; Urine SR-BI KO, N=6. (**D**) Comparisons of sRNA-seq data analysis pipelines, as reported as percent assigned per total reads for TIGER (black), Chimira (blue), Oasis (red), ExceRpt (green), and miRge (yellow) for HDL, APOB, and liver samples from WT mice. HDL WT, N=7; APOB WT, N=7, Liver WT, N=7. Mann-Whitney non-parametric tests. *p<0.05.

Figure 9. SR-BI regulates HDL-sRNAs at the individual fragment level, not parent level. Differential expression analysis by DEseq2. Volcano plots of demonstrating significant (adjusted p>0.05) differential (>1.5-absolute fold change) abundances for miRNAs, snDRs, and rDRs at the parent and individual fragment levels - red, increased; blue, decreased. HDL WT, N=7; HDL SR-BI KO N=7; APOB WT, N=7, APOB SR-BI KO N=7; Liver WT, N=7; Liver SR-BI KO, N=7; Bile WT, N=7; Bile SR-BI KO, N=6; Urine WT, N=5; Urine SR-BI KO, N=6.

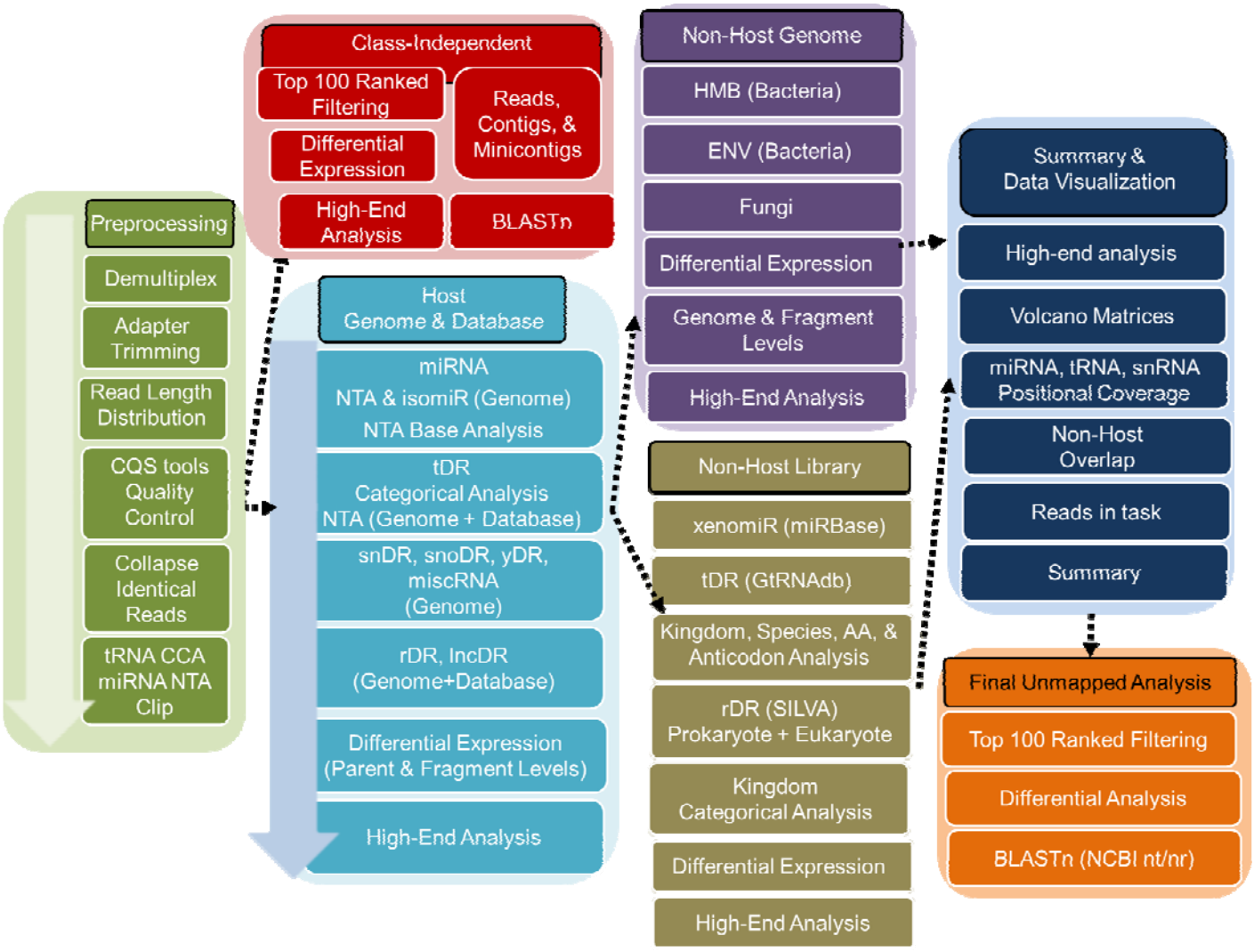
References:

1. Vickers KC, Sethupathy P, Baran-Gale J and Remaley AT. Complexity of microRNA function and the role of isomiRs in lipid homeostasis. *J Lipid Res*. 2013;54:1182-91.
2. Scott DD and Norbury CJ. RNA decay via 3' uridylation. *Biochim Biophys Acta*. 2013;1829:654-65.
3. Knouf EC, Wyman SK and Tewari M. The human TUT1 nucleotidyl transferase as a global regulator of microRNA abundance. *PLoS One*. 2013;8:e69630.
4. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004;116:281-97.
5. Vitsios DM and Enright AJ. Chimira: analysis of small RNA sequencing data and microRNA modifications. *Bioinformatics*. 2015;31:3365-7.
6. Vickers KC, Roteta LA, Hucheson-Dilks H, Han L and Guo Y. Mining diverse small RNA species in the deep transcriptome. *Trends in biochemical sciences*. 2015;40:4-7.
7. Chen CJ and Heard E. Small RNAs derived from structural non-coding RNAs. *Methods*. 2013;63:76-84.
8. Li Z, Ender C, Meister G, Moore PS, Chang Y and John B. Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. *Nucleic Acids Res*. 2012;40:6787-99.

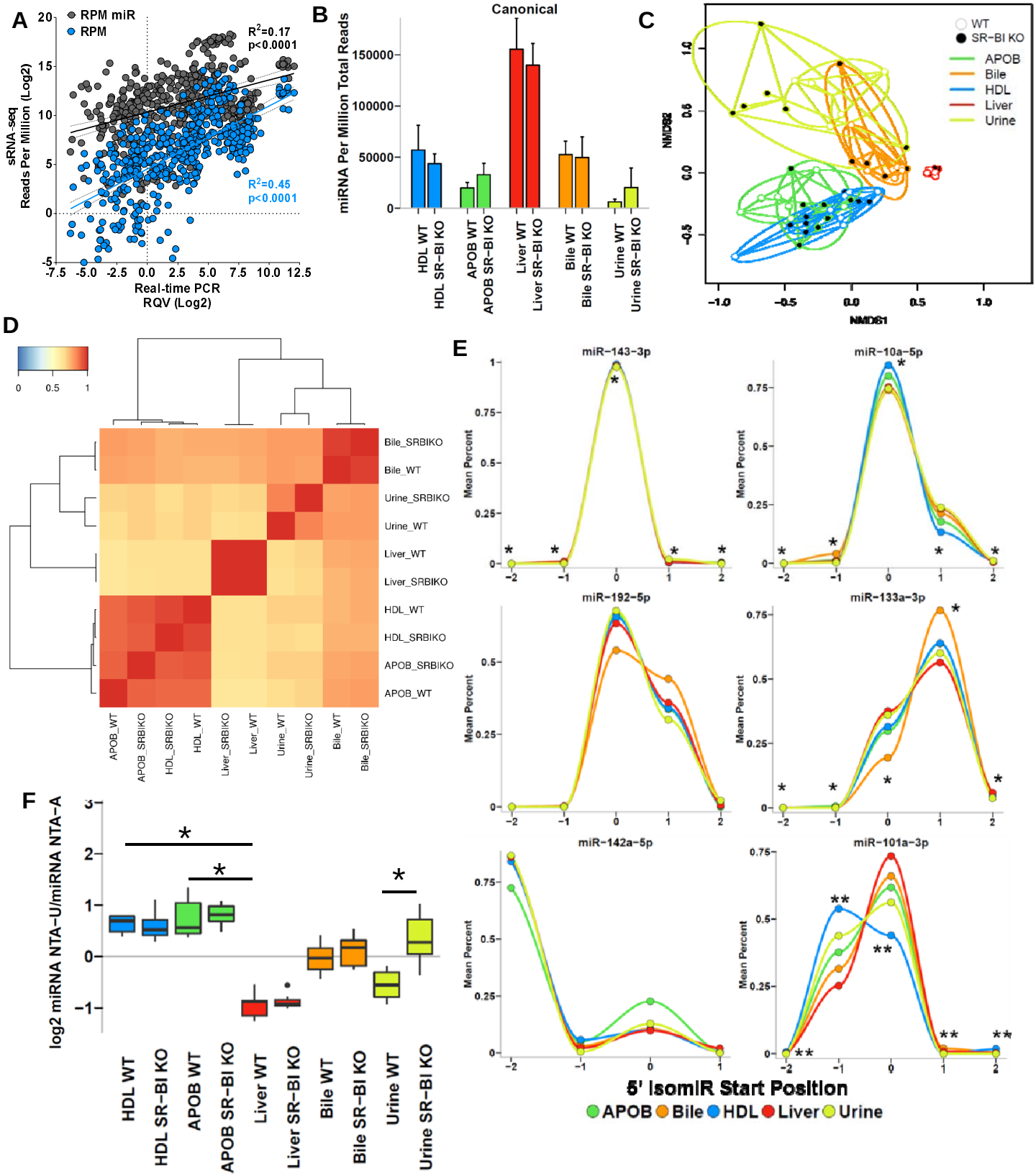
9. Vickers KC, Roteta LA, Hucheson-Dilks H, Han L and Guo Y. Mining diverse small RNA species in the deep transcriptome. *Trends in biochemical sciences*. 2015;40:4-7.
10. Baras AS, Mitchell CJ, Myers JR, Gupta S, Weng LC, Ashton JM, Cornish TC, Pandey A and Halushka MK. miRge - A Multiplexed Method of Processing Small RNA-Seq Data to Determine MicroRNA Entropy. *PLoS One*. 2015;10:e0143066.
11. Capece V, Garcia Vizcaino JC, Vidal R, Rahman RU, Pena Centeno T, Shomroni O, Suberviola I, Fischer A and Bonn S. Oasis: online analysis of small RNA deep sequencing data. *Bioinformatics*. 2015;31:2205-7.
12. Wang K, Li H, Yuan Y, Etheridge A, Zhou Y, Huang D, Wilmes P and Galas D. The complex exogenous RNA spectra in human plasma: an interface with human gut biota? *PLoS One*. 2012;7:e51009.
13. Kaczor-Urbanowicz KE, Kim Y, Li F, Galeev T, Kitchen RR, Koyano K, Jeong SH, Wang X, Elashoff D, Kang SY, Kim SM, Kim K, Kim S, Chia D, Xiao X, Rozowsky J and Wong DTW. Novel approaches for bioinformatic analysis of salivary RNA Sequencing data in the biomarker development process. *Bioinformatics*. 2017.
14. Boon RA and Vickers KC. Intercellular transport of microRNAs. *Arterioscler Thromb Vasc Biol*. 2013;33:186-92.
15. Vickers KC and Remaley AT. Lipid-based carriers of microRNAs and intercellular communication. *Curr Opin Lipidol*. 2012;23:91-7.
16. Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD and Remaley AT. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat Cell Biol*. 2011;13:423-33.
17. Dhahbi JM, Spindler SR, Atamna H, Yamakawa A, Boffelli D, Mote P and Martin DI. 5' tRNA halves are present as abundant complexes in serum, concentrated in blood cells, and modulated by aging and calorie restriction. *BMC Genomics*. 2013;14:298.
18. Lewis BP, Burge CB and Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005;120:15-20.
19. Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S, Barbacioru C, Steptoe AL, Martin HC, Nourbakhsh E, Krishnan K, Gardiner B, Wang X, Nones K, Steen JA, Matigian NA, Wood DL, Kassahn KS, Waddell N, Shepherd J, Lee C, Ichikawa J, McKernan K, Bramlett K, Kuersten S and Grimmond SM. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol*. 2011;12:R126.
20. Neilsen CT, Goodall GJ and Bracken CP. IsomiRs--the overlooked repertoire in the dynamic microRNAome. *Trends Genet*. 2012;28:544-9.
21. Vickers KC, Sethupathy P, Baran-Gale J and Remaley AT. The Complexity of microRNA Function and the Role of IsomiRs in Lipid Homeostasis. *J Lipid Res*. 2013.
22. Baran-Gale J, Fannin EE, Kurtz CL and Sethupathy P. Beta cell 5'-shifted isomiRs are candidate regulatory hubs in type 2 diabetes. *PLoS One*. 2013;8:e73240.
23. Burroughs AM, Ando Y, de Hoon MJ, Tomaru Y, Nishibu T, Ukekawa R, Funakoshi T, Kurokawa T, Suzuki H, Hayashizaki Y and Daub CO. A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. *Genome Res*. 2010;20:1398-410.
24. Koppers-Lalic D, Hackenberg M, Bijnsdorp IV, van Eijndhoven MAJ, Sadek P, Sie D, Zini N, Middeldorp JM, Ylstra B, de Menezes RX, Wurdinger T, Meijer GA and Pegtel DM. Nontemplated nucleotide additions distinguish the small RNA composition in cells from exosomes. *Cell reports*. 2014;8:1649-1658.
25. Acton S, Rigotti A, Landschulz KT, Xu S, Hobbs HH and Krieger M. Identification of scavenger receptor SR-BI as a high density lipoprotein receptor. *Science*. 1996;271:518-20.
26. Zhang Y, Da Silva JR, Reilly M, Billheimer JT, Rothblat GH and Rader DJ. Hepatic expression of scavenger receptor class B type I (SR-BI) is a positive regulator of macrophage reverse cholesterol transport in vivo. *J Clin Invest*. 2005;115:2870-4.
27. Wiersma H, Gatti A, Nijstad N, Oude Elferink RP, Kuipers F and Tietge UJ. Scavenger receptor class B type I mediates biliary cholesterol secretion independent of ATP-binding cassette transporter g5/g8 in mice. *Hepatology*. 2009;50:1263-72.

28. Zanoni P, Khetarpal SA, Larach DB, Hancock-Cerutti WF, Millar JS, Cuchel M, DerOhannessian S, Kontush A, Surendran P, Saleheen D, Trompet S, Jukema JW, De Craen A, Deloukas P, Sattar N, Ford I, Packard C, Majumder A, Alam DS, Di Angelantonio E, Abecasis G, Chowdhury R, Erdmann J, Nordestgaard BG, Nielsen SF, Tybjaerg-Hansen A, Schmidt RF, Kuulasmaa K, Liu DJ, Perola M, Blankenberg S, Salomaa V, Mannisto S, Amouyel P, Arveiler D, Ferrieres J, Muller-Nurasyid M, Ferrario M, Kee F, Willer CJ, Samani N, Schunkert H, Butterworth AS, Howson JM, Peloso GM, Stitziel NO, Danesh J, Kathiresan S, Rader DJ, Consortium CHDE, Consortium CAE and Global Lipids Genetics C. Rare variant in scavenger receptor BI raises HDL cholesterol and increases risk of coronary heart disease. *Science*. 2016;351:1166-71.
29. Varban ML, Rinninger F, Wang N, Fairchild-Huntress V, Dunmore JH, Fang Q, Gosselin ML, Dixon KL, Deeds JD, Acton SL, Tall AR and Huszar D. Targeted mutation reveals a central role for SR-BI in hepatic selective uptake of high density lipoprotein cholesterol. *Proc Natl Acad Sci U S A*. 1998;95:4619-24.
30. Zhong CY, Sun WW, Ma Y, Zhu H, Yang P, Wei H, Zeng BH, Zhang Q, Liu Y, Li WX, Chen Y, Yu L and Song ZY. Microbiota prevents cholesterol loss from the body by regulating host gene expression in mice. *Scientific reports*. 2015;5:10512.
31. Zhang X, Cozen AE, Liu Y, Chen Q and Lowe TM. Small RNA Modifications: Integral to Function and Disease. *Trends in molecular medicine*. 2016;22:1025-1034.
32. Telonis AG, Loher P, Kirino Y and Rigoutsos I. Consequential considerations when mapping tRNA fragments. *BMC Bioinformatics*. 2016;17:123.
33. Selitsky SR and Sethupathy P. tDRmapper: challenges and solutions to mapping, naming, and quantifying tRNA-derived RNAs from human small RNA-sequencing data. *BMC Bioinformatics*. 2015;16:354.
34. Wei Z, Batagov AO, Schinelli S, Wang J, Wang Y, El Fatimy R, Rabinovsky R, Balaj L, Chen CC, Hochberg F, Carter B, Breakefield XO and Krichevsky AM. Coding and noncoding landscape of extracellular RNA released by human glioma stem cells. *Nature communications*. 2017;8:1145.
35. Yeri A, Courtright A, Reiman R, Carlson E, Beecroft T, Janss A, Siniard A, Richholt R, Balak C, Rozowsky J, Kitchen R, Hutchins E, Winarta J, McCoy R, Anastasi M, Kim S, Huentelman M and Van Keuren-Jensen K. Total Extracellular Small RNA Profiles from Plasma, Saliva, and Urine of Healthy Subjects. *Scientific reports*. 2017;7:44061.
36. Quinn JF, Patel T, Wong D, Das S, Freedman JE, Laurent LC, Carter BS, Hochberg F, Van Keuren-Jensen K, Huentelman M, Spetzler R, Kalani MY, Arango J, Adelson PD, Weiner HL, Gandhi R, Goilav B, Putterman C and Saugstad JA. Extracellular RNAs: development as biomarkers of human disease. *Journal of extracellular vesicles*. 2015;4:27495.
37. Zernecke A and Preissner KT. Extracellular Ribonucleic Acids (RNA) Enter the Stage in Cardiovascular Disease. *Circ Res*. 2016;118:469-79.
38. Willeit P, Skroblin P, Moschen AR, Yin X, Kaudewitz D, Zampetaki A, Barwari T, Whitehead M, Ramirez CM, Goedeke L, Rotllan N, Bonora E, Hughes AD, Santer P, Fernandez-Hernando C, Tilg H, Willeit J, Kiechl S and Mayr M. Circulating MicroRNA-122 Is Associated With the Risk of New-Onset Metabolic Syndrome and Type 2 Diabetes. *Diabetes*. 2017;66:347-357.
39. Zhang L, Hou D, Chen X, Li D, Zhu L, Zhang Y, Li J, Bian Z, Liang X, Cai X, Yin Y, Wang C, Zhang T, Zhu D, Zhang D, Xu J, Chen Q, Ba Y, Liu J, Wang Q, Chen J, Wang J, Wang M, Zhang Q, Zhang J, Zen K and Zhang CY. Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. *Cell Res*. 2012;22:107-26.
40. Chin AR, Fong MY, Somlo G, Wu J, Swiderski P, Wu X and Wang SE. Cross-kingdom inhibition of breast cancer growth by plant miR159. *Cell Res*. 2016;26:217-28.
41. Liang H, Zhang S, Fu Z, Wang Y, Wang N, Liu Y, Zhao C, Wu J, Hu Y, Zhang J, Chen X, Zen K and Zhang CY. Effective detection and quantification of dietetically absorbed plant microRNAs in human plasma. *J Nutr Biochem*. 2015;26:505-12.
42. Dickinson B, Zhang Y, Petrick JS, Heck G, Ivashuta S and Marshall WS. Lack of detectable oral bioavailability of plant microRNAs after feeding in mice. *Nat Biotechnol*. 2013;31:965-7.

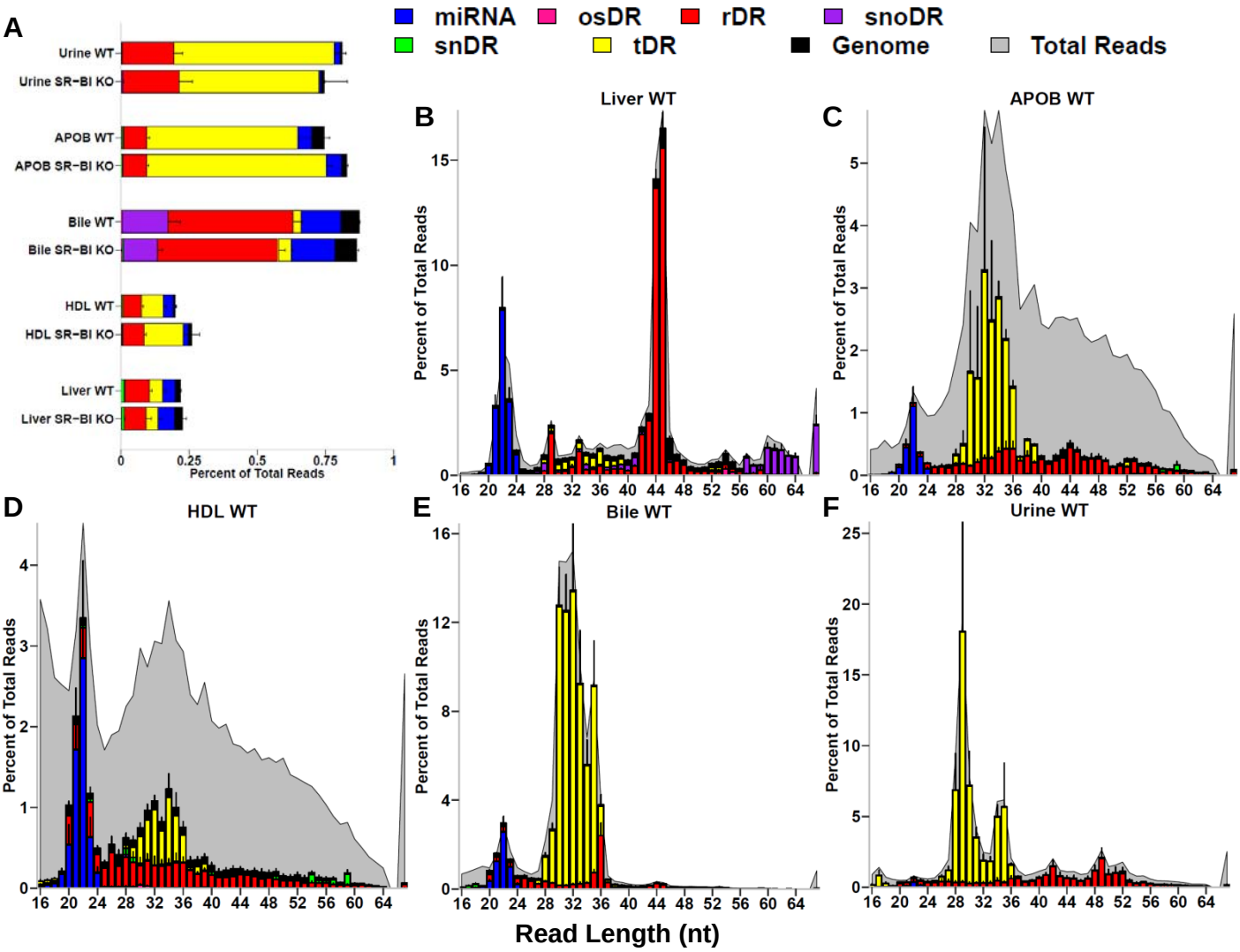
43. Masood M, Everett CP, Chan SY and Snow JW. Negligible uptake and transfer of diet-derived pollen microRNAs in adult honey bees. *RNA Biol*. 2016;13:109-18.
44. Mico V, Martin R, Lasuncion MA, Ordovas JM and Daimiel L. Unsuccessful Detection of Plant MicroRNAs in Beer, Extra Virgin Olive Oil and Human Plasma After an Acute Ingestion of Extra Virgin Olive Oil. *Plant Foods Hum Nutr*. 2016;71:102-8.
45. Witwer KW, McAlexander MA, Queen SE and Adams RJ. Real-time quantitative PCR and droplet digital PCR for plant miRNAs in mammalian blood provide little evidence for general uptake of dietary miRNAs: limited evidence for general uptake of dietary plant xenomiRs. *RNA Biol*. 2013;10:1080-6.
46. Beatty M, Guduric-Fuchs J, Brown E, Bridgett S, Chakravarthy U, Hogg RE and Simpson DA. Small RNAs from plants, bacteria and fungi within the order Hypocreales are ubiquitous in human plasma. *BMC Genomics*. 2014;15:933.
47. Qin Y, Yao J, Wu DC, Nottingham RM, Mohr S, Hunicke-Smith S and Lambowitz AM. High-throughput sequencing of human plasma RNA by using thermostable group II intron reverse transcriptases. *RNA*. 2016;22:111-28.
48. Wang M, Weiberg A, Lin FM, Thomma BP, Huang HD and Jin H. Bidirectional cross-kingdom RNAi and fungal uptake of external RNAs confer plant protection. *Nat Plants*. 2016;2:16151.
49. Weiberg A, Wang M, Lin FM, Zhao H, Zhang Z, Kaloshian I, Huang HD and Jin H. Fungal small RNAs suppress plant immunity by hijacking host RNA interference pathways. *Science*. 2013;342:118-23.
50. Zhang T, Zhao YL, Zhao JH, Wang S, Jin Y, Chen ZQ, Fang YY, Hua CL, Ding SW and Guo HS. Cotton plants export microRNAs to inhibit virulence gene expression in a fungal pathogen. *Nat Plants*. 2016;2:16153.
51. Stocks MB, Moxon S, Mapleson D, Woolfenden HC, Mohorianu I, Folkes L, Schwach F, Dalmay T and Moulton V. The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics*. 2012;28:2059-61.
52. Kim J, Levy E, Ferbrache A, Stepanowsky P, Farcas C, Wang S, Brunner S, Bath T, Wu Y and Ohno-Machado L. MAGI: a Node.js web service for fast microRNA-Seq analysis in a GPU infrastructure. *Bioinformatics*. 2014;30:2826-7.
53. Agirre E and Eyras E. Databases and resources for human small non-coding RNAs. *Hum Genomics*. 2011;5:192-9.
54. An J, Lai J, Lehman ML and Nelson CC. miRDeep*: an integrated application tool for miRNA identification from RNA sequencing data. *Nucleic Acids Res*. 2013;41:727-37.
55. Friedlander MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S and Rajewsky N. Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol*. 2008;26:407-15.
56. de La Llera-Moya M, Connelly MA, Drazul D, Klein SM, Favari E, Yancey PG, Williams DL and Rothblat GH. Scavenger receptor class B type I affects cholesterol homeostasis by magnifying cholesterol flux between cells and HDL. *J Lipid Res*. 2001;42:1969-78.
57. Wang Y, Liu X, Pijut SS, Li J, Horn J, Bradford EM, Leggas M, Barrett TA and Graf GA. The combination of ezetimibe and ursodiol promotes fecal sterol excretion and reveals a G5G8-independent pathway for cholesterol elimination. *J Lipid Res*. 2015;56:810-20.
58. Love MI, Huber W and Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
59. Zhao S, Guo Y, Sheng Q and Shyr Y. Advanced heat map and clustering analysis using heatmap3. *BioMed research international*. 2014;2014:986048.



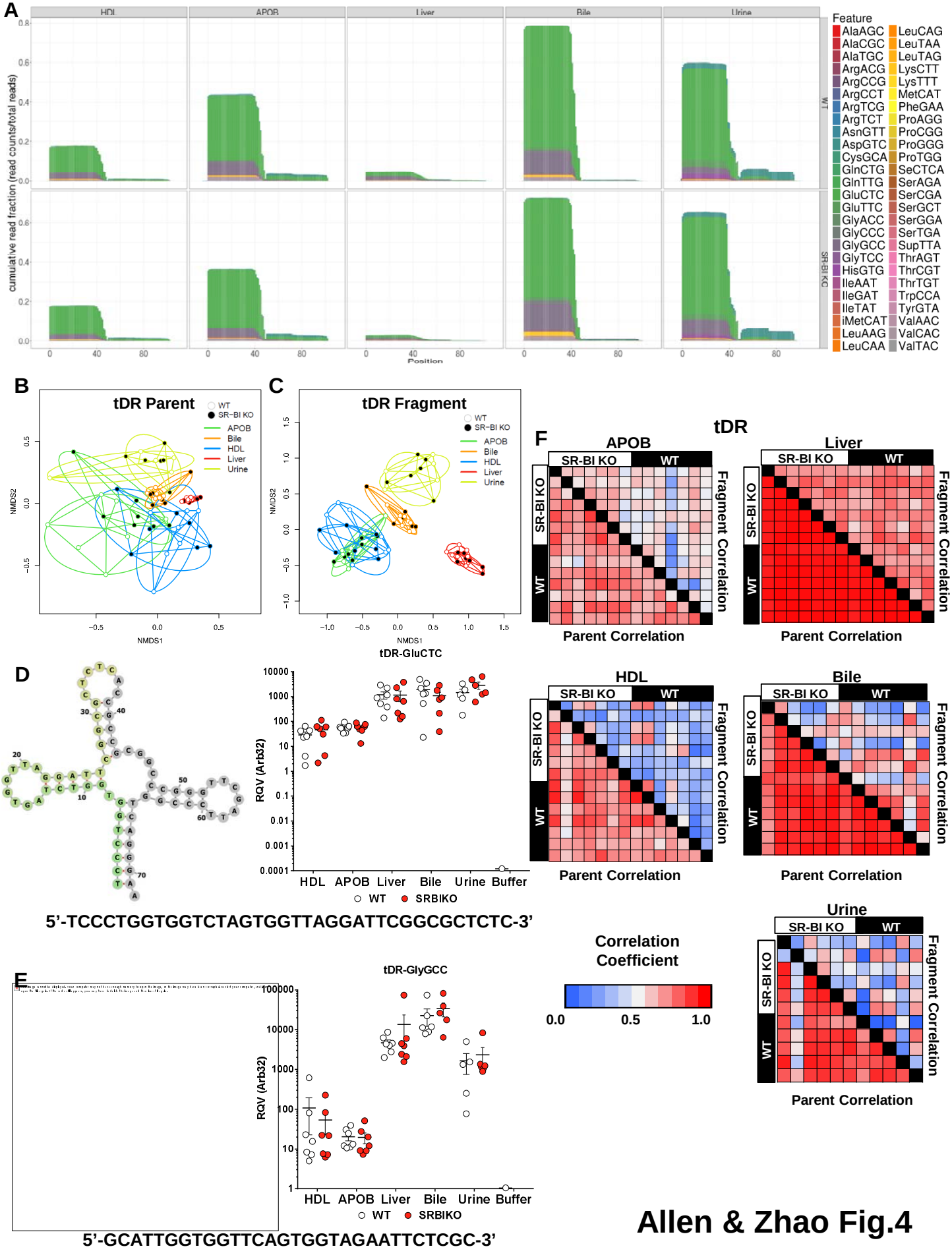
Allen & Zhao Fig.1



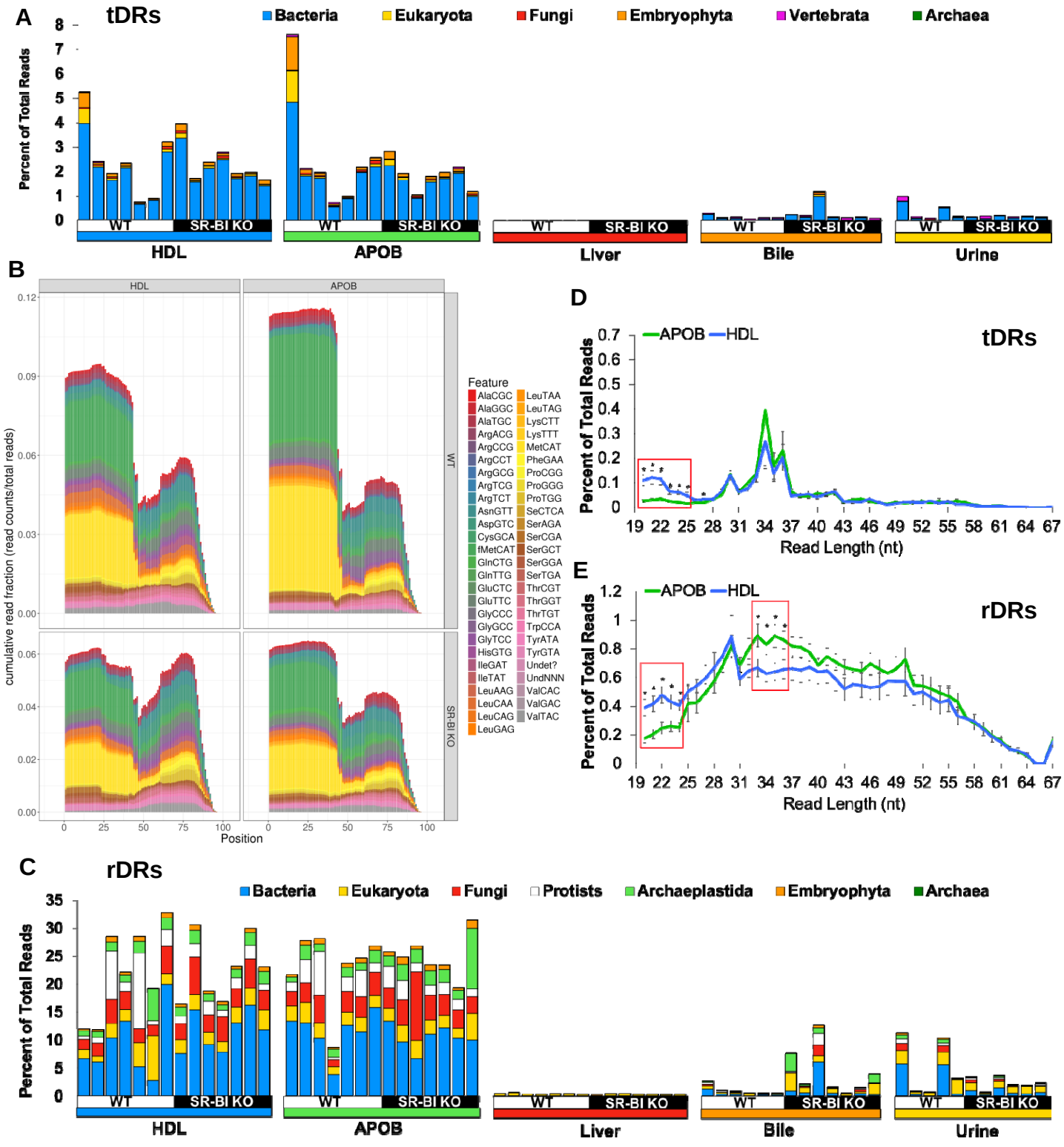
Allen & Zhao Fig.2



Allen & Zhao Fig.3

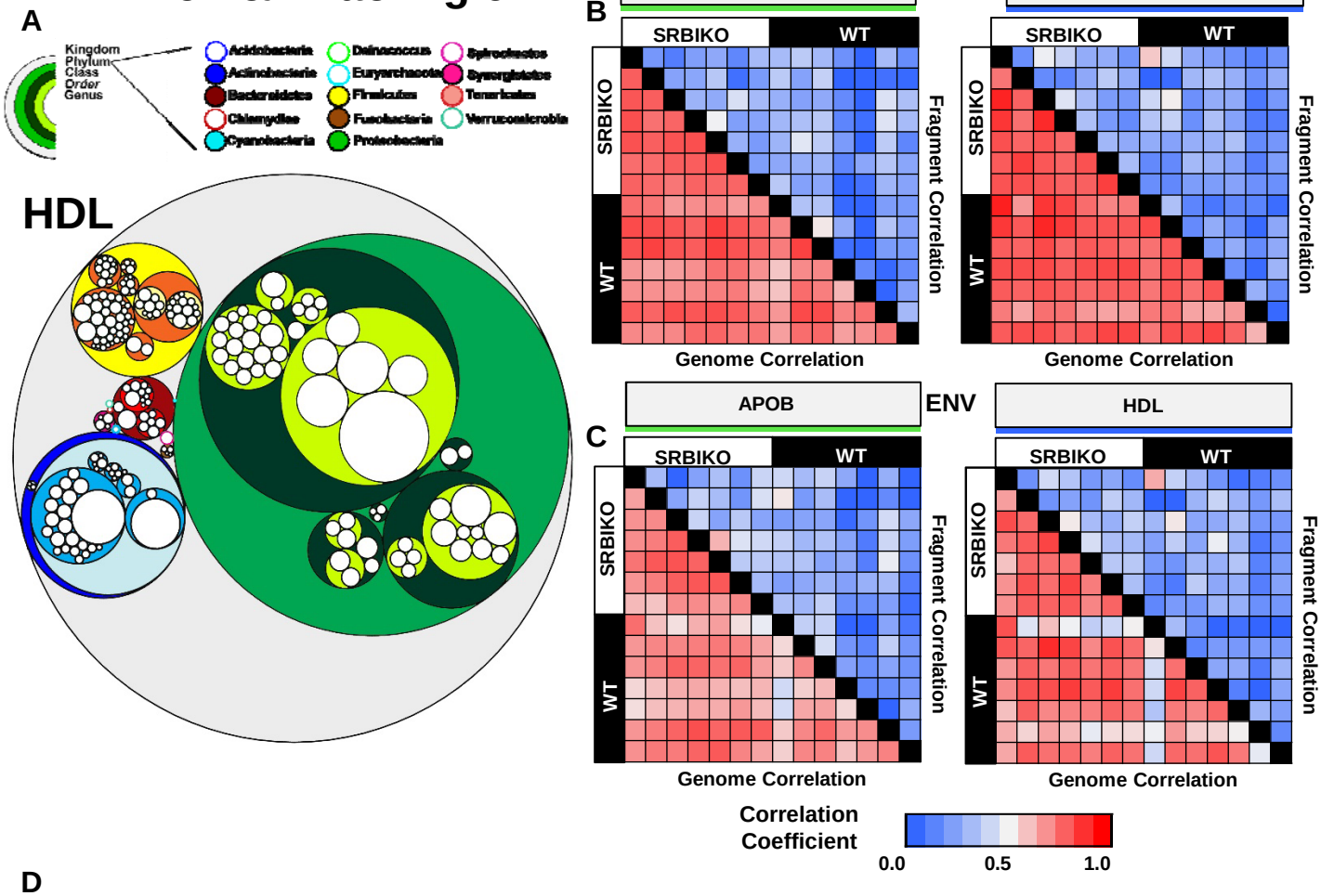


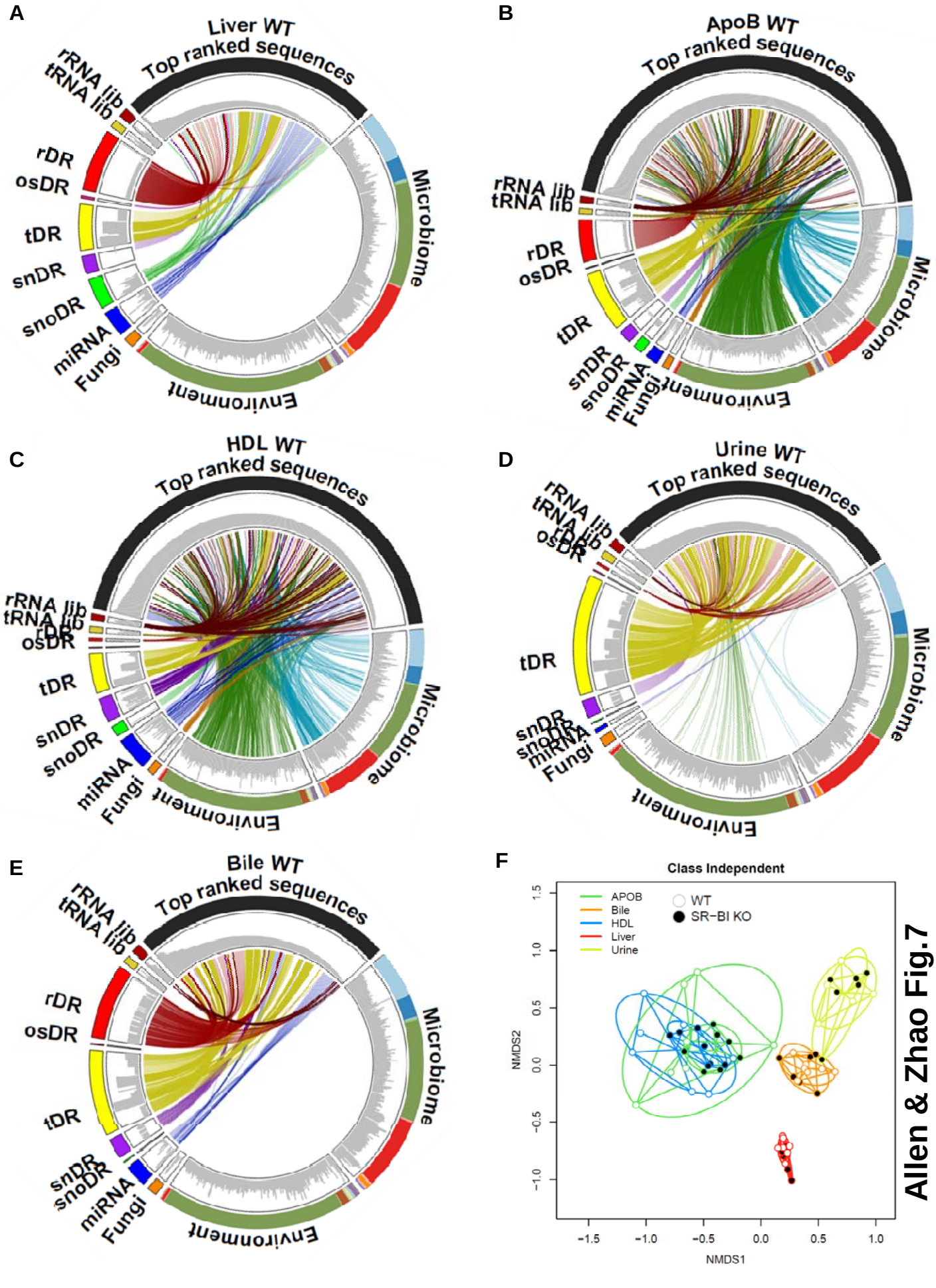
Allen & Zhao Fig.4

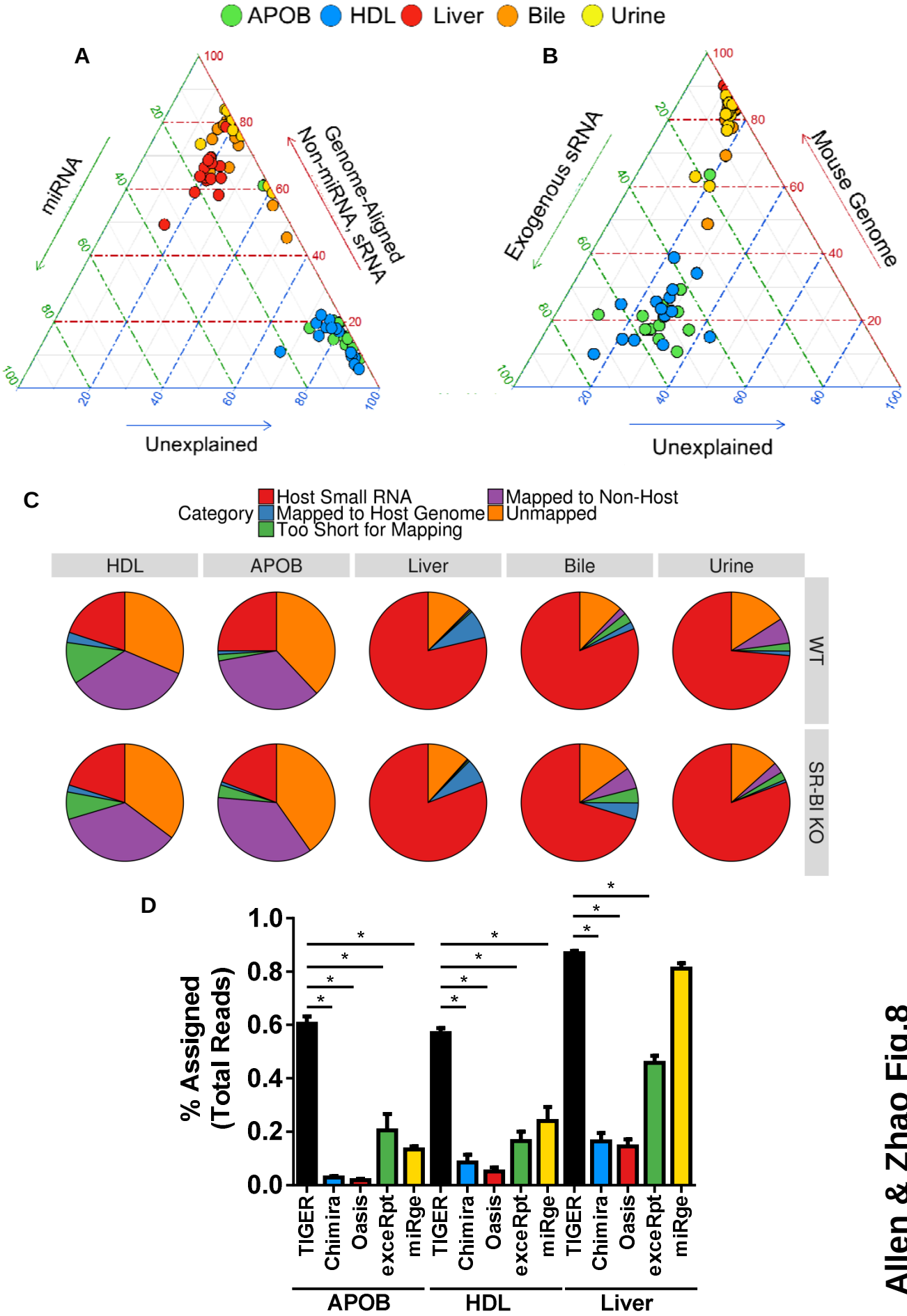


Allen & Zhao Fig.5

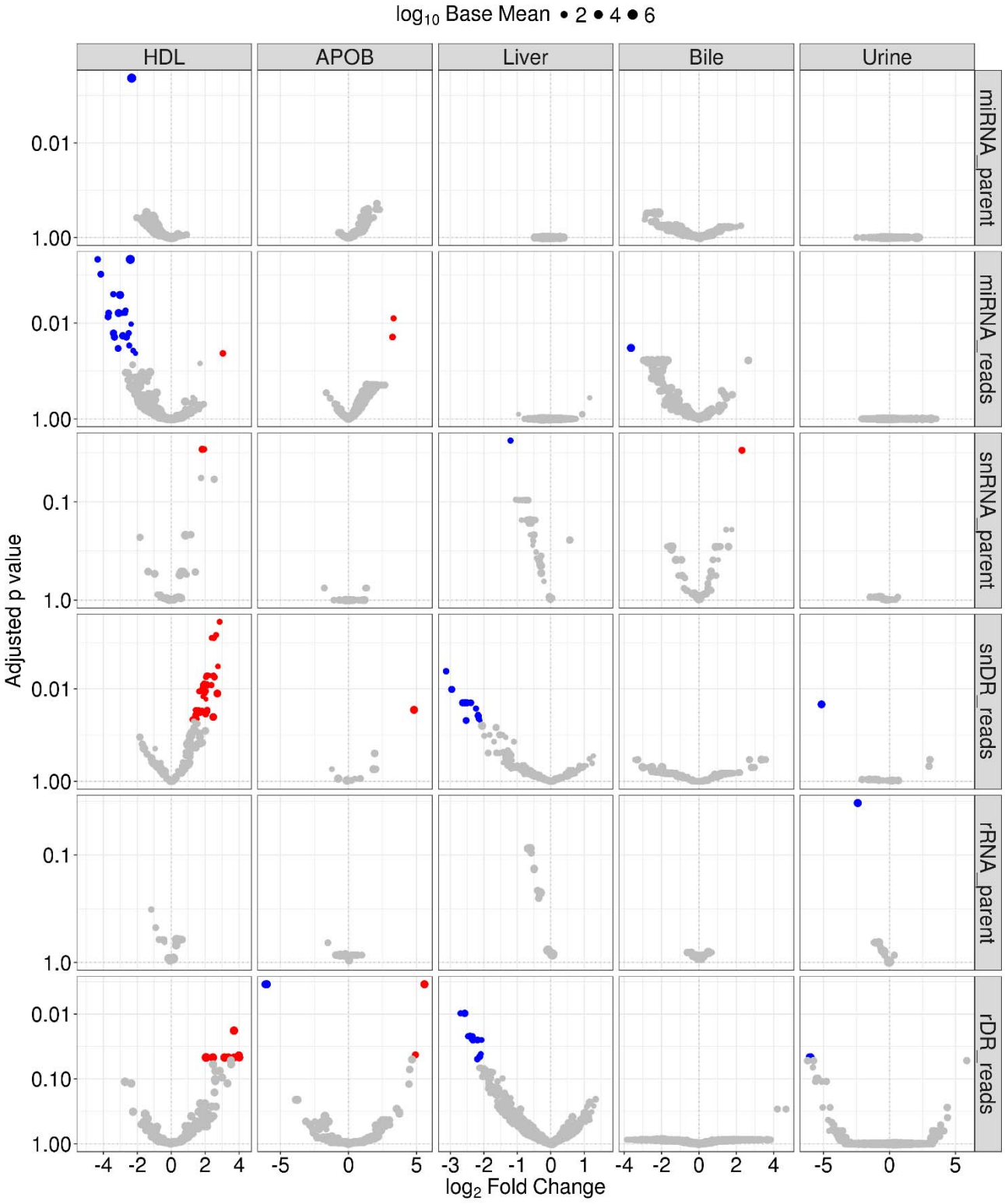
Allen & Zhao Fig.6







Allen & Zhao Fig.8



Allen & Zhao Fig.9