Background: Recent advances in high-throughput small-RNA (sRNA) sequencing (sRNA-seq) technologies have fueled the discovery of many new classes of biologically relevant non-coding sRNAs. Accumulating evidence suggests that sRNAs are critical contributors to the pathogenesis of various diseases and play an essential role in regulating gene expression levels[1]. RNA-seq analysis has revealed diverse classes of sRNAs circulating on various lipid and protein carriers, including high-density lipoproteins (HDL) [1]. The most well characterized sRNAs are microRNAs (miRNA) and the sRNA-seq analysis tools currently available are designed to focus mostly on miRNA quantification. Due to the limitations of previous analysis tools, our lab developed a novel sRNA analysis pipeline (i.e. TIGER) which profiles many classes of host (e.g. mouse, human) and non-host (e.g. bacteria, archaea, fungal) sRNAs present on lipoproteins[2]. Using our new TIGER pipeline, we discovered that the overwhelming majority of circulating sRNAs on HDL are classified as host and non-host ribosomal RNA-derived fragments (rDF)[2].

Motivation: Although the functional role(s) of rDFs are poorly understood, mounting evidence suggests that rDFs are not products of random degradation, but regulated by specific endonucleolytic cleavage processes, similar to that of transfer RNA-derived fragments (tDFs)[3]. Indeed, various stressors were shown to induce transfer RNA cleavage events, producing stable tDRs[3]. Angiogenin is an RNase A-family enzyme that is thought to be primarily responsible for stress-induced tRNA fragmentation within mammalian cells[4]. Angiogenin has also been shown to induce rRNA fragmentation, although to a lesser extent[4]. Interestingly, the cleavage of tRNAs and rRNAs are inherently linked to their chemical modifications (i.e. m^1A and m^5C)[5]. Both tRNAs and rRNAs represent the most abundant sRNAs, and the two most heavily modified RNAs in the eukaryotic genome. Although the identification of chemical modifications on tDFs has been actively pursued, few studies address the modifications found on rDFs. However, preliminary data I have generated using 2D-thin layer chromatography (2D-TLC) identified abundant base modifications (e.g. m^5C, m^6A) on sRNAs isolated from human HDL samples.

A major limitation when exploring the sRNA world is that many base modifications can disrupt Watson/Crick base pairing and impede first-strand synthesis by reverse transcriptase (RT) [4]. These chemical modifications therefore affect the detection and quantification of sRNAs, limiting the power of discovery. Although recent improvements to our TIGER pipeline have greatly enhanced our ability to assess sRNA content on HDL, base modifications would significantly impair efficient detection of these modified sRNAs. To circumvent these issues, a relatively new method was developed called AlkB-facilitated RNA methylation sequencing (ARM-seq) which exploits the RT roadblocks created by chemical base modifications (i.e. m^1A, m^3C and m^1G) in tRNAs[6]. The *E. coli* AlkB homologs (ALKBH1 and ALKBH2) act as "eraser" proteins, catalyzing the demethylation of specific chemical base modifications[6]. This method represents a large step forward in the quantification of tRNAs, however a very limited number of studies have used ARM-seq for rDFs. Similar to tDFs and miRNAs, which were once readily discarded from RNA-seq datasets, rDFs may play important roles in the regulation of gene expression. As such, accurately quantifying rDFs and their modification status on HDL is key to gaining a more complete understanding of the biological functions of the epitranscriptome. Based on our previous studies and preliminary results **I hypothesize that: (1)** *Improved sRNA-seq methods will increase the inclusion and identification of rDRs in HDL-sRNA datasets.* **(2)** *Stress factors induce parent rRNA fragmentation leading to an increase in circulating rDRs.* I will address these hypotheses through two central aims.

Aim 1: *Enhance HDL-sRNA identification by characterizing the landscape of chemical base modifications found on sRNAs.* To achieve this goal, we must capture and identify all host

and non-host rDRs. This will include a.) Expanding bioinformatic analyses for rDRs, b.) Improving the identification of modified rDRs, and c.) Removing modifications on sRNAs for enhanced rDF inclusion in sequencing analyses. To address this aim I will first collect blood from healthy individuals and isolate their sRNAs found circulating on HDL using fast protein liquid chromatography. I will then pretreat HDL-sRNAs with the purified AlkB enzymes prior to cDNA synthesis (RT step) and library preparation. By comparing AlkB-treated and untreated samples, I will reveal the positional modification profile of HDL-sRNAs, including rDFs. The TIGER pipeline will be used to identify the diverse classes of sRNAs on HDL particles. The power of ARM-seq will be maximized by taking advantage of RNA modification databases, such as Modomics and RMBase. I expect ARM-seq to efficiently reveal chemical base modifications in the sRNA samples and increase the repertoire of rDFs.

Aim 2: *Characterize changes in parent rRNA fragmentation and cellular rDF export to HDL in response to environmental stress.* Overwhelming evidence supports the role for specific environmental stressors to induce tRNA cleavage; however, very few studies have looked at rRNA fragmentation during environmental stress[7]. To determine whether oxidative stress, heat and cold stress, or γ-irradiation promote rRNA cleavage events, and the export of rDFs to HDL, I will treat human hepatic and non-hepatic cell lines with various environmental stressors (hydrogen peroxide, cold or heat shock, or irradiation with UV). Afterwards, the cells will be fractionated into nuclear and cytoplasmic extracts, and HDL will be isolated using a FPLC. To examine stress-induced rRNA fragments within these cellular fractions, I will use improved sRNA-seq approaches and confirm candidate rDFs using northern blot techniques. Moreover, we will quantify the export of hepatic rDFs to HDL in response to stresses to using HDL-sRNA export assays. I fully expect that exposure of specific environmental stressors will induce distinct parent rRNA fragmentation patterns and alter hepatic rDF export to HDL

Broader Impact: Circulating sRNAs have been shown to be differentially altered in several diseases and hold great potential for the discovery of novel biomarkers and highly promising therapeutics. Given the value of potential biomarkers, the field of sRNA has led to cutting edge research. However, there are still gaps in our understanding of sRNA diversity on circulating HDL. My proposal helps to address this gap and may lead to the identification of yet unknown RNAs. With novel classes or sRNAs being discovered, and the validation of modified sRNAs, it is paramount that RNA-modification and sRNA databases are updated. I will disseminate my findings to web portals and servers dedicated to compiling databases for RNA modifications.

Intellectual Merit: It was not very long ago that many sRNAs were considered "junk" and often removed from RNA-sequencing data analysis. However, we now know that sRNAs can regulate several aspects of gene expression. The novel pipeline generated by our bioinformatics team allows us to discern several classes of small RNAs found in both eukaryotes and prokaryotes. This interdisciplinary proposal applies techniques from bioinformatics, transcriptomics, microbiology, and biochemistry, and represents the *first study aimed at identifying modified small RNAs on HDL.* Successful completion of this proposal will not only expand the repertoire of sRNAs and rDRs but will also show how rDRs are important biological molecules.

References: **[1]** Vickers et al. 2011. *Nature Cell Biology*. **[2]** Allen et al. 2018. *Journal of Extracellular Vesicles.* **[3]** Lambert et al. 2018. Non-coding RNA Investigation. **[4]** Su et. Al. 2019. *J Biol Chem.* **[5]** Rashad et al. 2020. *Neural Regeneration Research.* **[6]** Cozen et al. 2015. *Nature Methods.* **[7]** Thompson et al. 2009. *Cell.*