STATISTICAL ANALYSIS COVER PAGE

TITLE: Plasma Exosomal MicroRNAs as Promising Novel Biomarkers for Suicidality and Treatment Outcome

NCT NUMBER: NCT02418195

DOCUMENT DATE: February 24, 2022

Aim 1: Examine whether suicidality is associated with differences in the expression of neural-derived plasma exosomal miRNAs.

Analysis of differentially expressed miRNAs for human brain studies: To detect differentially expressed miRNAs, 'Sequencing Capacity' package in the 'Basic Analysis' module of SeqBuster will be used. The Willcoxon sign-rank test will be applied to determine statistically significant differences in the frequency distribution between MDD, MDD suicide and control subjects. Two types of differential expression analyses will be performed: 1) different miRNA loci containing all the sequences mapped to each locus, grouped including all precursors, mature miRNA, miRNA*, and isomiRs; and 2) IsomiRs for only 5'-trimming and nucleotide substitution that affect the seed region of the miRNA. To compare abundance across various groups, raw small RNA sequence counts in each sample will be normalized by endogenous miRNAs and synthetic miRNA added as spike-in control. The Z-test will be applied to show statistical significance in the differential expression. The Benjamini and Hochberg method[121] will be applied to correct the p values assigned by the Z-test. We will apply the Z-score option to exclude sequencing errors as the possible cause of the nucleotide changes observed in the isomiR. miRNAs, which show significant changes, will be further evaluated by correlating with variables such as age, PMI, gender, and brain pH. Duration of medication, life-time antidepressant treatment history, and prior suicide attempt (MDD cases), will be used as covariate.

Analysis of differentially expressed miRNAs for clinical studies: Data will first be cleansed to remove outliers and then normalized using spike-in RNAs if appropriate. Sequences will be analyzed to identify those which show mean group differences (>1.5 fold) across various groups.

Models and Hypotheses Testing: Proc Mixed (SASv.9.3) will be used for this data analysis. A mixed-effects model with random subject intercept to incorporate correlations of observations (i.e. expression of miRNAs) nested within the same subject will be used to analyze miRNA data. This analysis will borrow strength from all groups to estimate common parameters such as variance component of random effect and error variance. The model will include miRNA specific fixed parameter that depends on the group under consideration. Thus for a given miRNA, there will be four fixed parameters corresponding to four groups such as MDD suicide attempter (MDD-SA), MDD suicidal ideation (MDD-SI), MDD non-suicidal (MDD-NS), and healthy controls (HC). To compare mean expressions of a miRNA of any two groups, the corresponding fixed parameters of that particular miRNA of those two groups will be tested. For Aims 1a and b, the difference of miRNA specific mean parameters of MDD-SA patients and MDD-NS patients and those of MDD-SA patients and HC will be tested to see if mean expressions of each of the miRNAs of MDD-SA and MDD-NS, and MDD-SA and HC differ significantly. For Aim 1c, the primary covariate analysis will use a mixed-effects model to test for significant effects of impulsivity (BIS-11), hostility (BDHI), and aggression (AGHA) as fixed covariates to analyze miRNAs differentially-expressed in MDD-SA and MDD-SI groups. A secondary mixed-effects model test will be conducted for the remainder of items listed in **Table 5**, including subscale scores on the BDHI, BIS-11, and CTQ. Parameters corresponding to fixed covariates will be tested individually for significance.

Table 5. Covariate analysis variables

Variables	Measure	Type
Impulsivity[a]	BIS-11	Continuous
Hostility[a]	BDHI	Continuous
Aggression[a]	BGHA	Continuous
Hopelessness	HS	Continuous
Borderline PD	PAI-B	Continuous
Depression	BDI	Continuous
Anxiety	BAI	Continuous
Neuroticism	NEO-N	Continuous
Extraversion	NEO-E	Continuous
Childhood trauma	CTQ	Continuous
Recent stress	PSS	Continuous
Age	Demog.	Continuous
Sex	Demog.	Categorical
Substance dep.	MINI	Categorical
Prior attempt	History	Categorical
Hospitalizations	History	Categorical
Attempt type	History	Categorical
# of attempts	History	Ordinal

[a]Primary covariate analysis.

Multiple Comparisons: To address the multiple comparison (i.e., comparison of multiple miRNAs), we will employ FDR approach.[121] FDR method is adaptive, in the sense that the thresholds that are chosen are automatically adjusted to the strength of the signal in the data. The parameter q has a definite and clear meaning that is comparable across studies. Finally, since the procedures work with p-values, and do not test statistics, FDR methods can be applied

with any valid statistical test. For testing *m* hypotheses (denoted by *H1, H2, Hi, Hm*), using FDR procedure we will: 1) set the desired FDR bound *q=.20.* This is the maximum proportion of false discoveries that we are willing to tolerate on average. Note that *q* is not comparable to α, the type I error rate in traditional hypothesis testing which assumes no true discoveries in null hypothesis, whereas FDR recognizes the existence of true discoveries; 2) order the p-values from smallest to largest: *p1 <= p2 <= ... <= pm,* where *pi* is the p-value corresponding to *Hi; 3)* find *r,* the largest *i* for which *pi <= q × i/m; 4)* Reject the hypotheses *....H1, ...H2, ...,Hr.*

Aim 2: Examine whether anti-suicidal/antidepressant effects of ketamine is associated with miRNAs changes.

Statistical Analysis: We will assess mediating and moderating effects of ketamine on study outcomes, i.e., suicidal ideation/attempt and MDD. Simple cross-sectional designs cannot truly assess causal relationships between factors (e.g., miRNAs and suicide). Causal inference is strengthened considerably by conducting moderating and mediating effects in the context of an intervention. A moderator variable determines on whom or under what conditions a factor (in this case miRNA levels) is associated with change in a primary outcome measure (here ketamine effects on suicidal ideation or MDD). A mediator variable determines how change in another factor results in change in an outcome measure. The MacArthur Foundation Network[129-132] refined the original concepts of the Baron and Kenny model to emphasized the fact that the identification of moderators and mediators cannot occur exclusively in the context of cross-sectional studies and the importance of temporal precedence in identifying moderators and mediators. To assess these models, we will test i) significant effects of ketamine on suicidal ideation (BSSI) and MDD (MADRS); ii) significant effects of ketamine on miRNA expressions; iii) significance of baseline miRNA expressions on suicidal ideation and MDD; and iv) check if the residual direct effect of ketamine on suicidal ideation or MDD is smaller than the overall effect to determine if changes in miRNA are specific to change in suicidal ideation or depression or are non-specific effects to ketamine. We will follow procedures first outlined by Baron and Kenny[(133)] and more recently refined by Zhao et al.[134] In addition, for both outcome measures suicidal ideation and MDD, we will test if baseline miRNA levels are moderating the effect of ketamine. For Aim 2b, we will determine miRNAs that expressed significantly by comparing before and after treatment of ketamine in each group (suicidal and non-suicidal MDD). Significantly expressed miRNAs belonging to the intersection of these two groups will be considered as causing to both MDD and suicidal behavior; those significantly expressed miRNAs belonging to non-suicidal MDD will be considered as causal to MDD; and those significantly expressed miRNAs belonging to the complement of MDD to suicidal MDD group will be considered as causing to suicidal behavior only. FDR will be employed for multiple comparison (multiple miRNAs).

Aim 3: Deduce specific miRNA-regulatory pathway(s) that contribute to suicide pathogenesis and treatment response.

RNA-Seq data analysis: The raw sequencing reads will be processed using the Illumina CASAVA package. This analysis includes: 1) alignment of reads directly to the reference human genome; 2) identification of expressed genes and isoforms; and 3) estimation of gene and isoform abundance and analysis of differential expression among various experimental groups. Short reads FASTQ format will be processed and aligned to the UCSC Human reference genome using PALMapper, which uses seed extend method for mapping reads.[137] The output SAM file will be used for transcript assembly and abundance analysis using Cufflinks.[138] Cufflinks constructs a minimum set of transcripts that best describes the reads in the dataset and uses the normalized RNA-Seq fragment counts to measure the relative abundances of transcripts. The unit of measurement is fragments per kilobase of exon per million mapped reads (FPKM). Confidence intervals for FPKM estimates will be calculated using a Bayesian inference method.[139] The output GTF file will be used for reference annotation and differential expression analysis. Cuffcompare,

a part of Cufflinks, will be used to analyze the GTF file to classify each transcript, followed by the use of Cuffdiff, also a part of Cufflinks, which will compare expression, splicing, and promoter use between various groups. By tracking changes in the relative abundance of transcripts with a common transcription start site, Cuffdiff identifies changes in splicing.

Network analysis to identify miRNA-mRNA co-expression modules: To identify modules containing miRNA-mRNA regulation clusters, Weighted Co-expression WCGNA will be performed by pooling mRNA and miRNA data.[140] The absolute values of Pearson correlation coefficients will be calculated for all possible pairs. Values will be entered into a matrix and the data will be transformed so that the matrix will follow an approximate scale-free topology. A dynamic tree cut algorithm[141] will be used to detect network modules. For brain studies, the WCGNA will be first performed in cases and controls separately. Then, module preservation statistic $Z_{summary}$ will be used to assess the module preservation among the two brain regions (dlPFC and hippocampus), comparing modules detected in cases and controls.[142] $Z_{summary}$ takes the overlap in module membership, the density, and connectivity patterns of modules into account. As in a previous study,[143] the cases and controls will have consistent modules. By that, we could combine case-control samples together for WCGNA so that sample size will be doubled for detection of correlation. In addition, we will test eigengenes (representative nodes)[144] of the miRNA-mRNA modules differential expression comparing MDD non-suicide, MDD suicide, and normal controls. Multiple testing corrections will be controlled by FDR ($q<0.05$).

Target prediction from mRNA-miRNA pair analysis: The identified mRNA-miRNA pairs will be analyzed using bioinformatics tools: MicroInspector, TargetScan, miRBase, and miRTarBase. MicroInspector spots the correct sites for miRNA-interaction in known target mRNAs. This program allows variation of temperature, the setting of energy values as well as the selection of different miRNA databases to identify miRNA-binding sites of different strength. TargetScan predicts biological targets of miRNAs by searching for the presence of conserved 8- and 7-mer sites that match the seed region of each miRNA. The miRBase is a database of published miRNA sequences and annotation. Each entry represents a predicted hairpin portion of a miRNA transcript with information on the location and sequence of the mature miRNA. miRTarBase has the largest collection of more than 3500 manually curated miRNA-target interactions from 985 publications, all of which are experimentally validated by luciferase reporter assay, western blot, qRT-PCR, microarray experiments with overexpression or knockdown of miRNAs, or pulsed stable isotope labeling with amino acids in culture (pSILAC).[145] Targets which are predicted by multiple algorithms, hit by the same miRNA at multiple sites, and hit by multiple different affected miRNAs at sites close to mRNA, will be considered as likely candidates. Since targets are of no use if not expressed in the human brain, we will verify the expression of target genes in the human brain at mRNA levels by analyzing them on Uni-Gene (www.ncbi.nlm.nih.gov/unigene).