Do probiotics modulate the intestinal microbiome in extremely premature infants? Statistical Analysis Plan

Principle Investigator: Dr. Harish Amin

REB16-0542

October 11, 2017

Characteristics variables will be analyzed using independent 2-sample *t* tests when continuous and Pearson χ^2 test when categorical. A P-value of less than 0.05 will be considered significant.

For the microbiome, the diversity within each sample, also known as α diversity, will be measured using the Shannon index. This is a measure of both the organismal richness of a sample and the evenness of the organisms' abundance distribution. A two-sided Mann-Whitney test (with False Discovery Rate multiple test correction) will be used to determine if there were significant differences in Shannon diversity between cases and controls.

Diversity between samples (β -diversity) will be evaluated using weighted UniFrac, a rank-based ordination method and visualized with Non-metric Multidimensional Scaling (NMDS). β-diversity represents the explicit comparison of microbial (or other) communities based on their composition or relative abundance. β diversity metrics thus assess the differences between microbial communities. To test whether there are significant differences in community structure between groups the UniFrac distance matrix will be tested with permutational multivariate ANOVA (PERMANOVA) implemented in the adonis function in the vegan R package. Differential abundance of individual OTUs between groups will be assessed using generalized linear models with a negative binomial distribution as is used commonly for RNA-seq. The similarity in data structure between RNA-seq and 16S allow us to adapt already mature methods. This analysis will be performed in DESeq2 v1.8.1 using default parameters.