Broadband Light for Refractory Dry Eye Disease
Statistical Analysis Plan

The SPEED validated dry eye disease questionnaire and visual analog scale will be given to patients at time of enrollment and after three sessions, the mean score before and after treatment for all patients will be calculated along with standard deviation and a student's T test will be used to assess whether there is a significant difference ($p<0.05$).
The same will be done for numbers such as tear osmolarity, basal tear production test times, tear break-up times. The proportion of patients in each category of the oxford scale for fluorescein grading (I-V) will also be compared at baseline and post-treatment and the change in grade of individual patients will be assessed descriptively.
For the microbiome portion of the study, swabs from the ocular surface and lid margin will be obtained, in addition to facial skin and environmental control, at baseline and after treatment. 16S ribosomal DNA sequencing will be performed to generate an amplicon sequence variant (ASV) table using DADA2. Briefly, DADA2 performs quality filtering, de-noising, sample inference, and chimera removal using an empiric error model to generate exact amplicon sequences. Next, taxonomies will be assigned to each sequence variant using the RDP naïve Bayes classifier. The 'decontam' R package (https://www.biorxiv.org/content/early/2017/11/17/221499) will be used to remove contaminant sequences based on the composition of negative controls including both DNA extraction buffers and PCR negatives.
Principal coordinates analysis (PCoA) and permutational multivariate ANOVA will be used to assess for site-dependent effects of treatment on bacterial composition (eyelid margin vs. facial skin vs. environmental control). Alpha diversity will be measured using the Chao1 and Shannon indices, and beta diversity will be measured using Bray-Curtis, Jaccard, and Jenson-Shannon distances. Standard non-parametric methods (e.g. Wilcoxon rank-sum test, Kruskal-Wallis test) will be used to compare alpha and beta diversity metrics between various groups. Zero-inflated negative binomial (ZINB) regression will be used to identify specific bacterial taxa that are differentially abundant by site and treatment. We will also utilize random forest classification as an alternative approach to identify bacterial taxa that distinguish the groups. The Benjamini-Hochberg false discovery rate (FDR) method will be used to adjust all statistical comparisons for multiple hypotheses.