

Final Project Proposal BIOS 7659

Dominic Adducci

The data for this project comes from a recently published study which included an epigenetic analysis to determine the effect of preterm birth on lung health (Cho, et al. 2023). The original study utilized samples from three different time points, cord blood, day 14 peripheral blood, and day 28 peripheral blood. Different analyses were performed at these time points to identify differentially methylated CpGs and differentially expressed genes. One analysis investigated the effects of NICU O₂ therapy on methylation for infants who did not have BPD at day 14. Increased levels of oxygen can have detrimental effects, and the purpose of the analysis was to determine how methylation responds to supplemental oxygen for infants who do not have BPD, and how these two groups (those who received oxygen and those who did not) compare against infants who were diagnosed with BPD. The project I plan on doing is an extension of this analysis, where non-BPD infants exposed to supplemental O₂ are compared against infants who received no supplemental O₂ by day 28.

DNA methylation microarray analysis was performed using Human MethylationEPIC BeadChip from Illumina, and data is provided in IDAT files. The subset of infants used in the original sub-analysis were those who did not develop BPD. The supplemental information will be checked to identify the infants who were used for this portion of the study. If this information is not available, the paper states that additional information may be available upon reasonable request. The IDAT files will need to be filtered first, which was done in the original paper by excluding arrays that have more than 5% failed probes, CpG probes on the X and Y chromosomes, any probes containing single-nucleotide polymorphism with a minor allele frequency $\geq 1\%$ within 5 nucleotides to the CpG site, and probes that were reported to hybridize to one or more non-target sites. Data cleaning for this project will attempt to filter data as close as possible to the original filtration. After filtering the IDAT files will be read into R using the minfi package and preprocessed with background and dye bias correction using the preprocessNoob method, same as the original methods, followed by using the champ.runCombat function from the ChAMP package for batch correction. In other words, I am going to try to match the original processing steps as close as possible (assuming I can figure out how to use these methods).

The analysis will focus primarily on finding hypo and hyper differentially methylated sites between these two groups. The genes associated with the most significant sites will be identified. From the paper there was a noticeable inhibition of mitochondrial function. Potential problems include not being able to fully replicate the processing and normalization steps, in which case a simpler preprocessing and normalization stage may be required. I also need to be able to identify which infants received supplementary O₂ and for how many days. This information should be listed in the supplementary material, but I have not found it yet.