**Meetings with Lauren Vanderlinden**

**5/26/2020**

**Discussing GSEAs volcano plots**

1. **The plots need to have actual values but they’re hard to acquire**
   1. I tried increasing the permutation which helped but not fully.
   2. 20,000 was the max. This gave us about 50-100 more Nominal P-values
   3. Still not enough so I met with Lauren for advice
2. **Lauren’s suggestions**
   1. Try reducing the outputted genesets or rather, isolate maybe top 100 genesets which may provide us with what we want. Then overlay the results onto the previous figure.
      1. Hesitant about this and believe that this would either give unequivalent p-values that aren’t congruent with the previous results ‘
      2. **OR** the values still won’t be available.
   2. **In the event that the above method does not work** 
      1. Manually permute the data before submitting to GSEA, extract the test statistics and generate the permuted Pvalue by hand.
      2. “we should perform permutations ourselves and submit each permutation to GSEA individually and collect the resulting test statistics. Then we can generate the permuted p-values by hand. This would be a lot of coding work. “
         1. Not sure how this would work but I’m attempting the first method and will review the second permute by P value method.

**5/28/20**

1. N= 7 is why its difficult to resolve low p values. The sample size is small
   1. Sample sizes are very important and the larger they are the more power we have
2. Create the 2x2 table indicating results that the volcanos try to visualize
3. Chi square test on the 2x2 data
   1. Should be an R function for this
4. Something about mixed models of repeated measures (ask Lauren for clarification)

**6/1/20**

1. The plots should include a legend
2. Remove the pvalue significance line and geneset labels from plot
3. Show the <.00005 plots by giving them .00005 instead of 0
4. Add additional details to the methods section
5. Instead of chi square lets use the fisher exact test on the 2x2 tables sindicating genesets of interest among the entire cohort of genesets

**6/10/20**

1. **Kyrsten’s Project (Harry)**
   1. **Pathways of particular interest**
   2. **Summarize expression for the pathway (PCA)**
   3. **Redo the excel documents with the gene names**
   4. **Redo the plots with correct PCs and axis percentages**
   5. **Linear Regression model (bcuz deseq is NBN)**
   6. **Summarize genesets with PCA (get descriptive statistics)**
   7. **Skree plot**
   8. **Email Kyrsten for deadlines (shes in obgyn)**
   9. **Have different color for each variable**
   10. **3 dimenensional components could be helpful**
2. **GSEA** 
   1. **Switch colors**
   2. **Add everything to Tims server**
   3. **Send code to Lauren**

**6/16/20**

1. **Givinostat project nearly complete**
   1. **Waiting on additional edit requests**
2. **Krysten Boyle’s project (rc2 building)**
   1. **Pediatrics and nutrition**
   2. **Gestation diabetes**
   3. **Umbilical cord data**
   4. **Just going to wait for now**
3. **Laurens leaving in the fall** 
   1. **Grant with jill noris for epidemiology**
   2. **Leaving CIDA**
   3. **Grad school? PhD**
4. **General info**
   1. **Small sample size dont need batch effect correction**
   2. **Large 40+ may be from several sources and probably needs a reduction from batch effects**
   3. **Longitudinal data -> monitoring a subject over period of time**
      1. **Rlog transform and do a linear model**
   4. **For EoE project**
      1. **Do histograms of p values**
      2. **Q plots of p vales**
   5. **Next week shes going on vacation 24-4th**
   6. **Grant season is hectic**