Machine Learning to predict disease/age/family

Studies have revealed significant diversity in the gut microbiome composition related to various phenotypes. For example, studies of lean and obese mice suggest a strong relationship between gut microbiome and obesity. Obesity has been associated with changes in the microbiota at phylum-level, reduction in bacterial diversity, and different representations of bacterial genes. There also have been some reports of a strong association between the skin microbiome and aging.

In this project, we will try to answer some questions using machine learning approaches.
1. Is there an association between gut microbiome and age. In other words, how well can we predict age using gut microbiome composition using machine learning?
2. How well can we classify different classes (obese/lean/overweight) based on gut microbiome using machine leaning?
- Check if we our model can identify important bacteria & replicate previous findings.
- Is there any systematic differences between monozygotic and dizygotic twins with regard to their microbiota composition controlling for other confounders.
3. Finally, how reliable and accurate it is to predict ancestry using gut microbiome and machine learning

Proposal

- Data:
  - Twin study (281 samples, >5000 OTUs)
  - Predictions for age, ancestry, disease class, family
- Machine Learning – Different Models
- Output:
  - Prediction
  - Visualize difference in microbiome composition based on age, ancestry, etc.
  - Phylogenetic trees
  - PCA (age, ancestry)
  - Interpreting the results
  - Add interactive component (user point of view)