Final Project Prospectus

Evan Ott and Raghav Shroff

November 1, 2016

Problem set-up

The human microbiome underscores an important role in human health, often producing a disease phenotype when disrupted or during compositional changes. We seek to further understand the genetic changes within the microbiome in response to the development of periodontal disease. We will fit a statistical model to delineate control and diseased microbiomes and identify potential markers for periodontal disease.

The data we will use comprises of 6 patients with stratified control and diseased samples within each patient (i.e. one set of gums developed periodontal disease while the other set did not). Data was collected at least two time points for each patient, for a total of 27 data points.

For this project, we will incorporate the following types of data:

(1) Gene expression data aligned to the human oral microbiome (463 individual genomes) from RNA-seq (raw read counts per gene)

(2) Abundance of each microbiome species obtained from 16S sequencing (raw read counts per species)

(3) Gingival crevicular fluid cytokine analysis providing the immunological response elicited in each patient (protein level expressed as pg/ml)

(4) Clinical scoring of each data point to indicate extent of periodontal disease (scale of 1 to 5)

Using this data, we plan to use a logistic regression model to predict: a) control/disease and b) clinical scoring of the individuals given the gene expression data. The gene expression data is extremely sparse: there are approximately 180,000 metabolic genes (the subset of interest) and each sample has on average 40,000 present.

There are some interesting additional components to the model in this example. Each person in the study has data represented at least four times (each set of gums, measured twice or more), so we would certainly expect their samples to be correlated with one another. Additionally, we have two sets of response variables: 1) the ground truth for each set of gums in terms of being the control or treatment and 2) the