

1 Load raw data

- 1.1 Load bacterial data
 - 1.1.1 Merge PLEASE data and COMBO data.
 - 1.1.2 Distribution of total non-human reads
 - 1.1.3 Filter low depth samples
 - 1.1.4 Filter low abundant bacterial data
 - 1.1.5 Distribution of identified taxa
- 1.2 Load rarefaction data
- 1.3 Load phylogenetic tree
- 1.4 Load kmer data
- 1.5 Load kmer data
- 1.6 Load fungi data
- 1.7 Load gene/pathway/module data
 - 1.7.1 Filter gene/pathway/module data
- 1.8 Load metabolite data and preprocess
 - 1.8.1 Load metabolite data
 - 1.8.2 Normalize metabolome data
- 1.9 Load clinical outcome information
 - 1.9.1 Merge outcome tables
 - 1.9.2 Correlation of FCP and PCDAI
 - 1.9.3 FCP change across time
 - 1.9.4 PCDAI change across time
 - 1.9.5 Response rate by treatment

2 Analyze bacterial data

- 2.1 Normal vs. crohn-T1
 - 2.1.1 Calculate distance
 - 2.1.2 Find the best clustering
 - 2.1.3 Find the best clustering (numerical jaccard)
 - 2.1.4 MDS analysis
 - 2.1.5 PERMANOVA of response vs. non-response at T1
 - 2.1.6 Wilcoxon rank test of normal vs. Crohn
 - 2.1.7 Wilcoxon rank test of normal vs. Crohn without antibiotics use
 - 2.1.8 Heatmap of clustering results
 - 2.1.9 Random Forest
- 2.2 Antibiotics use
 - 2.2.1 Wilcoxon rank test
 - 2.2.2 Random Forest
- 2.3 Cluster 1 vs. cluster 2
 - 2.3.1 Wilcoxon rank test
 - 2.3.2 Random Forest

```
p <- subset(sample.info, Type=='PLEASE-T1' | Type=='COMBO',
             select=c(Disease, Cluster, Human.Per, Fungi.Per)) %>%
na.omit() %>%
ggplot(aes(log10(Human.Per), log10(Fungi.Per))) +
geom_point(aes(colour = Cluster, shape = Disease), size=5) +
  scale_color_manual(values=c("#0066CC", "#FF3333"), name="") +
  scale_shape_manual(values=c(20, 1), name="") +
  #scale_size_area(range = c(1, 10)) +
  #scale_size_continuous(range = c(3, 10)) +
  #labs(title = 'Ordination bacterial taxa') +
  #stat_smooth_func(geom="text", method="lm", hjust=0, parse=TRUE) +
  geom_smooth(method="lm", se=FALSE) +
  theme_classic()
print(p)
```

