

C Diff Analysis

Feb. 5, 2020

Contents

| | |
|--|----|
| Figure 1 Calprotectin | 1 |
| Figure 1 qPCR | 2 |
| Figure 2 A-C | 3 |
| Figure 2 DEF | 3 |
| Figure 2 G | 4 |
| Figure 2 H | 4 |
| Figure 3 PCA | 5 |
| Figure 3 selected metabolites (chenodeoxycholate, cholate, deoxycholate, lithocholate) | 6 |
| Figure 4 A | 7 |
| Figure 4 B | 7 |
| Figure 4C | 8 |
| Figure 4D | 8 |
| Figure 4 EF (isocaproate (i6:0), isocaproyltaurine) | 10 |
| Figure 5 B | 10 |
| Figure 5 C | 11 |
| Figure 5 D | 11 |
| Supplementary Figure 2 | 12 |
| Supplementary Figure 3 | 12 |

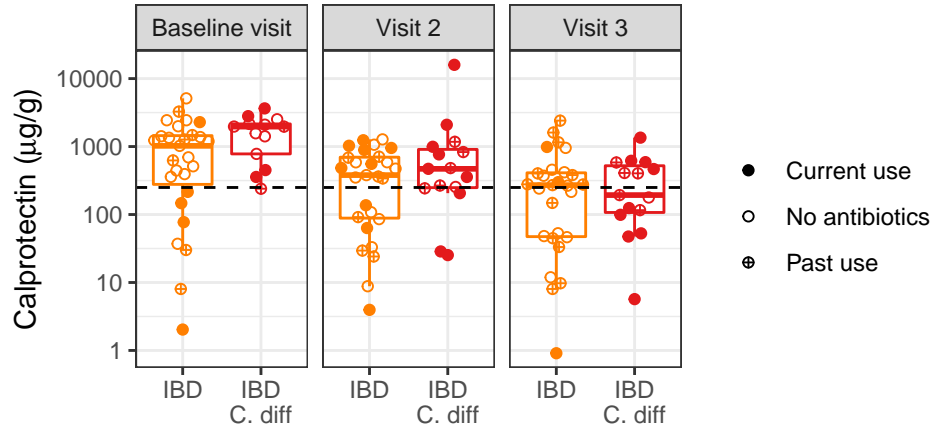
Figure 1 Calprotectin

This is the plot of FCP values for fecal samples at 3 time points.

```
##  
## Fligner-Killeen test of homogeneity of variances  
##  
## data: log10_calprotectin_ave by as.factor(Time)  
## Fligner-Killeen:med chi-squared = 0.22165, df = 2, p-value =  
## 0.8951
```

Table 1: Linear mixed-effects model fit by maximum likelihood : $\log_{10} \text{calprotectin_ave} \sim \text{age_yrs} + \text{antibiotics} + \text{study_group_new} * \text{Time}$

| | Value | Std.Error | DF | t-value | p-value |
|--------------------------------|----------|-----------|----|---------|-----------|
| (Intercept) | 2.522 | 0.2943 | 74 | 8.568 | 1.09e-12 |
| age_yrs | 0.01496 | 0.01982 | 74 | 0.7547 | 0.4528 |
| antibioticsPast use | -0.02673 | 0.1722 | 74 | -0.1552 | 0.8771 |
| antibioticsYes | -0.2318 | 0.1619 | 74 | -1.432 | 0.1564 |
| study_group_newIBD_C.Diff | 0.4384 | 0.2197 | 43 | 1.995 | 0.05236 |
| Time | -0.06349 | 0.0165 | 74 | -3.847 | 0.0002514 |
| study_group_newIBD_C.Diff:Time | -0.02186 | 0.02868 | 74 | -0.762 | 0.4485 |



There is no significant difference between the FCP of IBD and IBD C.diff group. The p value is 0.05236.

Figure 1 qPCR

This is the plot of qPCR values for fecal samples at 3 time points.

The undetermined equals 40.

| study_group_new | Baseline Visit | Visit 2 | Visit 3 |
|-----------------|----------------|---------|---------|
| Healthy | 15 | 15 | 15 |
| IBD | 6 | 6 | 6 |
| IBD_C.Diff | 17 | 15 | 15 |
| ONC_C.Diff | 9 | 3 | 2 |

```
##
## Fligner-Killeen test of homogeneity of variances
##
## data: tcdB_ct_ave by as.factor(current_antibiotics)
## Fligner-Killeen:med chi-squared = 1.0492, df = 1, p-value = 0.3057
```

Table 3: Linear mixed-effects model fit by maximum likelihood :
 $\text{tcdB_ct_ave} \sim \text{current_antibiotics} + \text{Time}$

| | Value | Std.Error | DF | t-value | p-value |
|------------------------|-------|-----------|----|---------|-----------|
| (Intercept) | 30.91 | 1.226 | 25 | 25.22 | 2.682e-19 |
| current_antibioticsYes | 2.579 | 1.498 | 25 | 1.721 | 0.09755 |
| Time | 0.653 | 0.2269 | 25 | 2.878 | 0.008086 |

Shape by antibiotics.

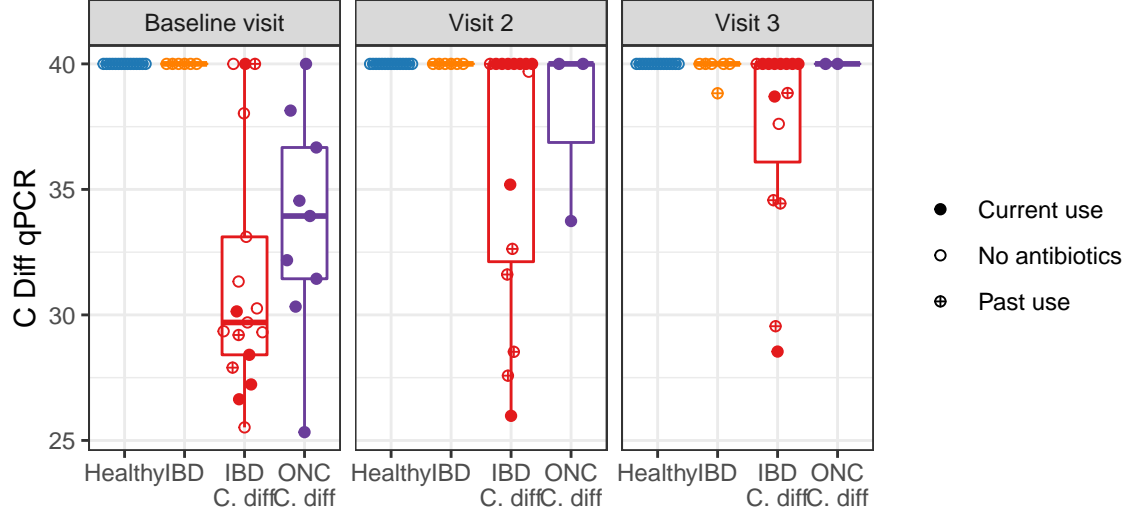


Figure 2 A-C

This is the PCoA plot based on Bray-Curtis Distance using kraken result.

The centroid is defined by using all healthy controls (including Healthy ≥ 6 and Healthy <6).

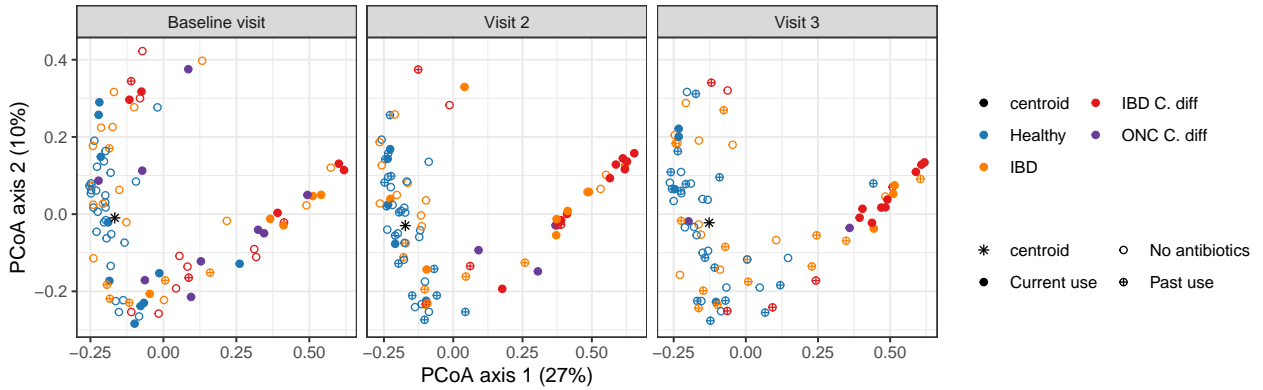


Figure 2 DEF

This is the dendro bar plot for fecal samples.

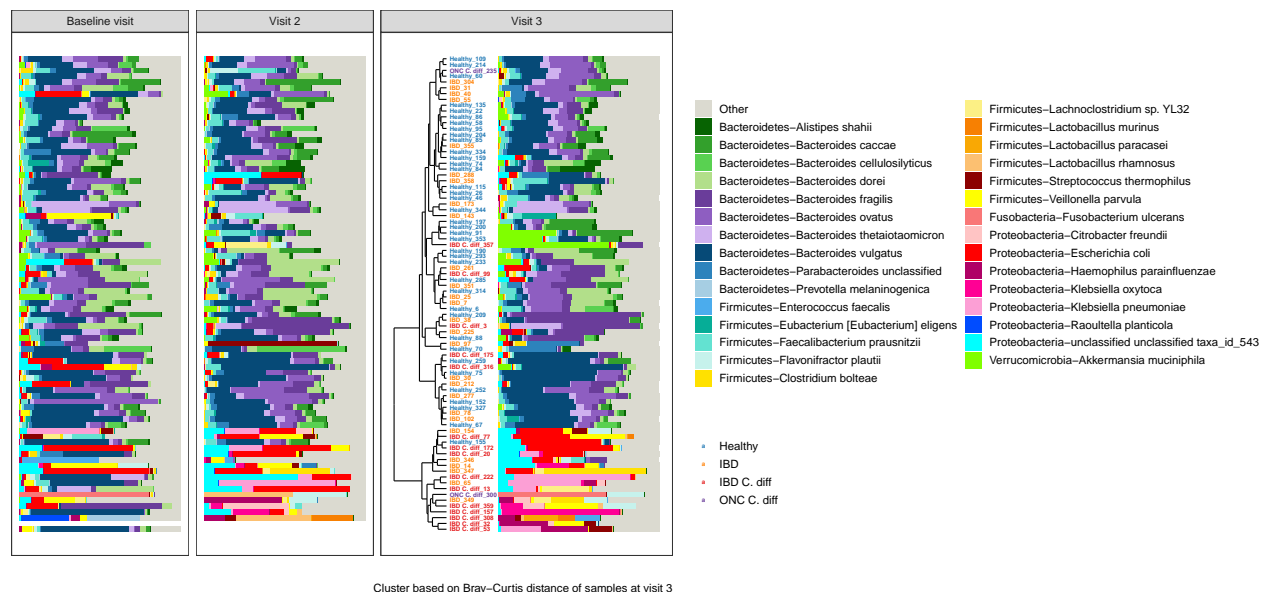


Figure 2 G

This is the boxplot of rarefied richness for fecal samples at the baseline visit.

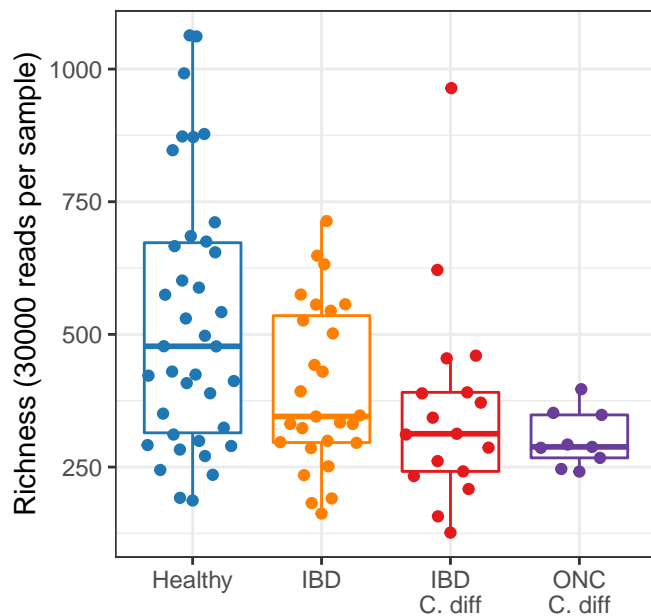


Figure 2 H

This is the boxplot of human DNA percentage for fecal samples at 3 time points.

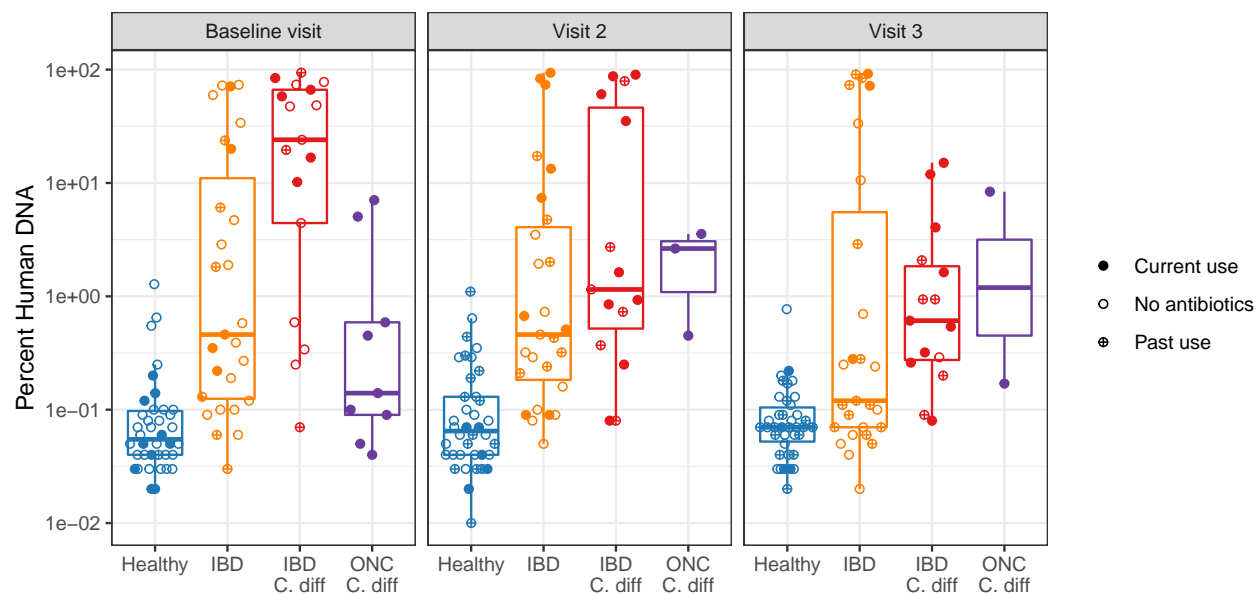


Figure 3 PCA

This is a PCA plot based on metabolomics data.

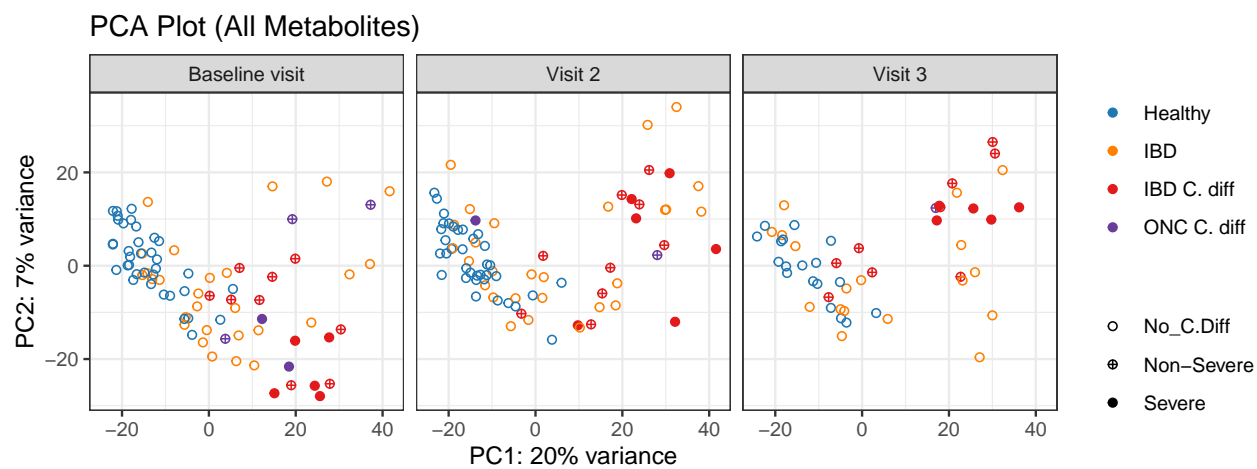
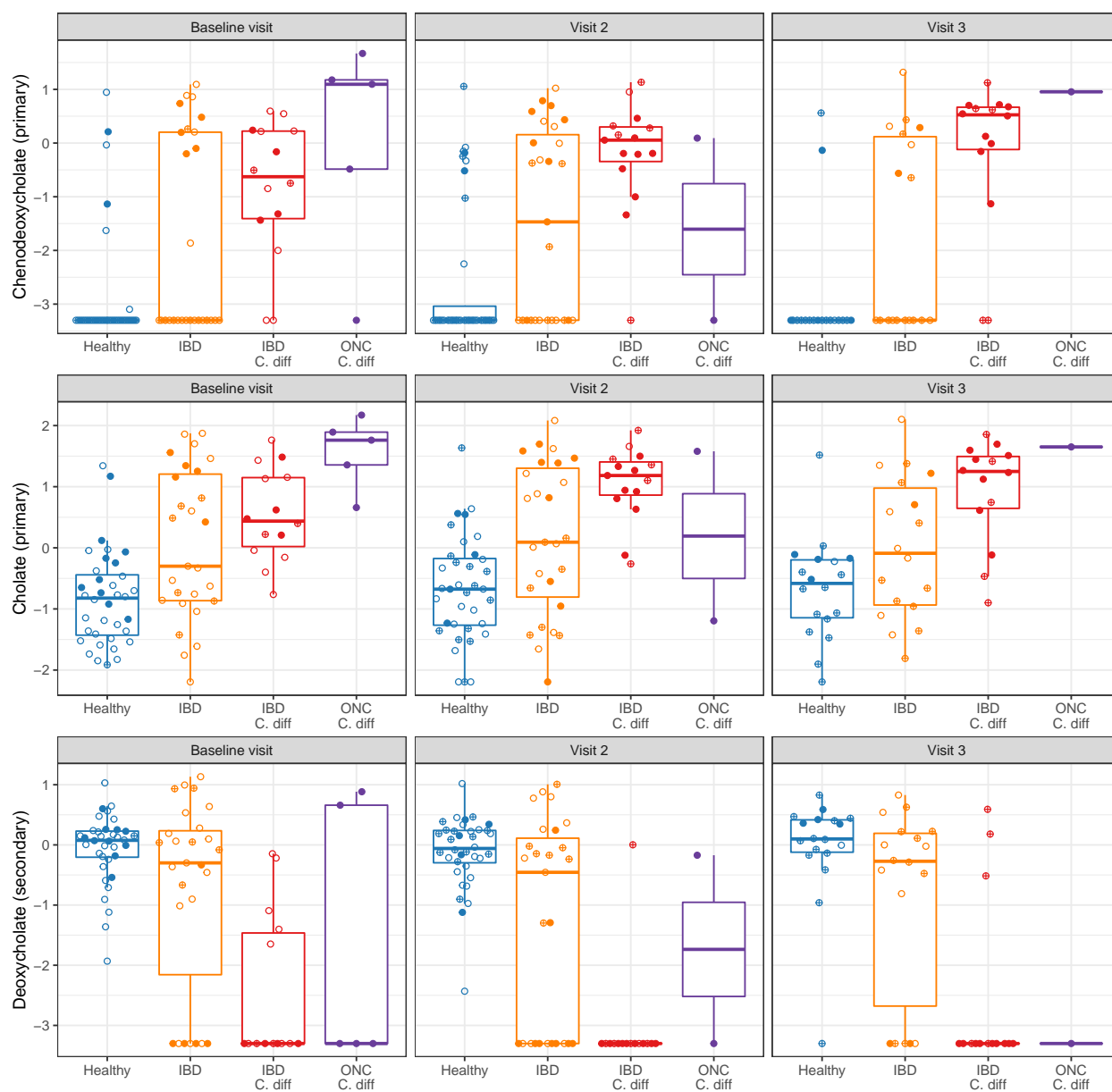


Figure 3 selected metabolites (chenodeoxycholate, cholate, deoxycholate, lithocholate)



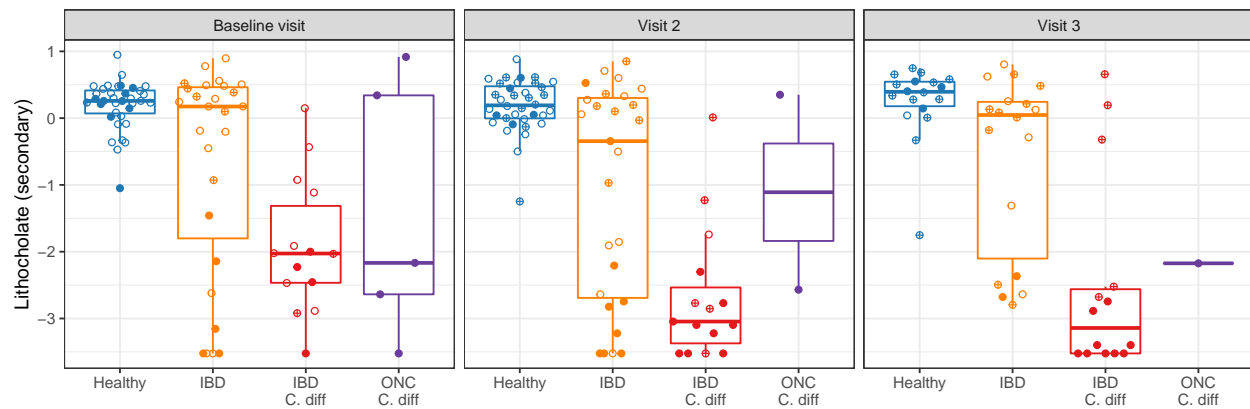


Figure 4 A

We used the randomForest model to compare IBD C.Diff and Healthy based on the metabolomics data at the baseline visit.

This is the heatmap of the top 30 important features.

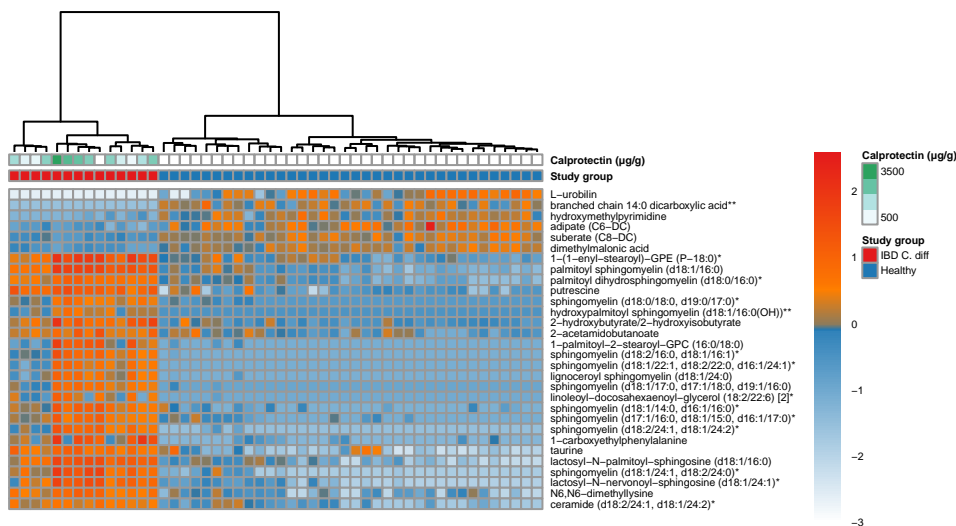


Figure 4 B

We used the randomForest model to compare IBD C.Diff and IBD based on the metabolomics data at the baseline visit.

This is the heatmap of the top 30 important features.

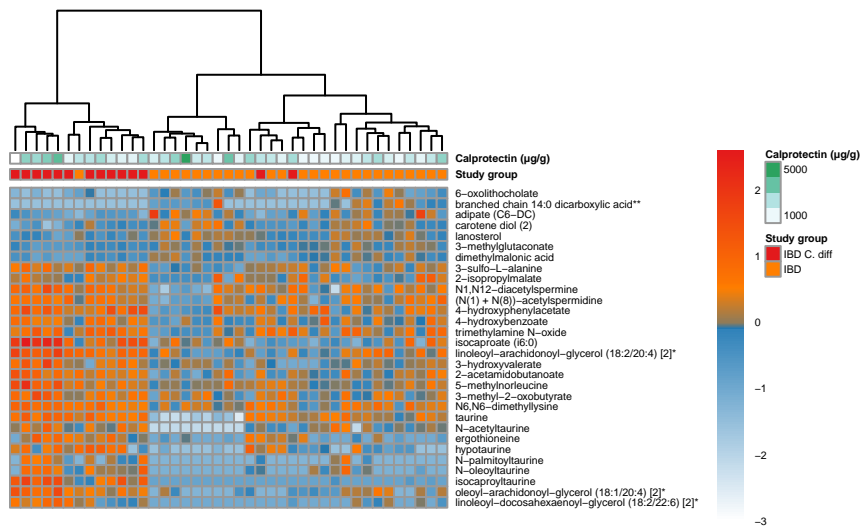


Figure 4C

We used the randomForest model to compare IBD C.Diff and IBD based on the metabolomics data at the baseline visit.

This is a barplot of the top 30 important features.

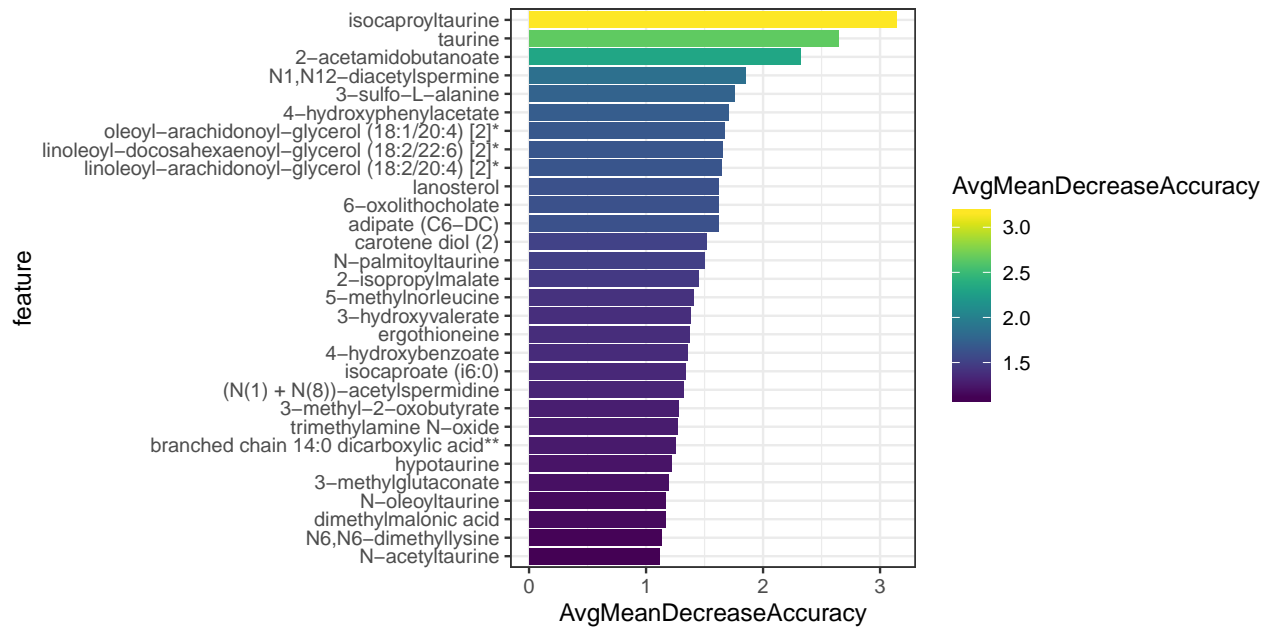


Figure 4D

Table 4: Spearman's rank correlation rho:
`sampleData$log10HumanPer` and `sampleData$putrescine`

| Test statistic | P value | Alternative hypothesis | rho |
|----------------|-----------------|------------------------|--------|
| 39580 | 1.201e-08 * * * | greater | 0.5692 |



Figure 4 EF (isocaproate (i6:0), isocaproyltaurine)

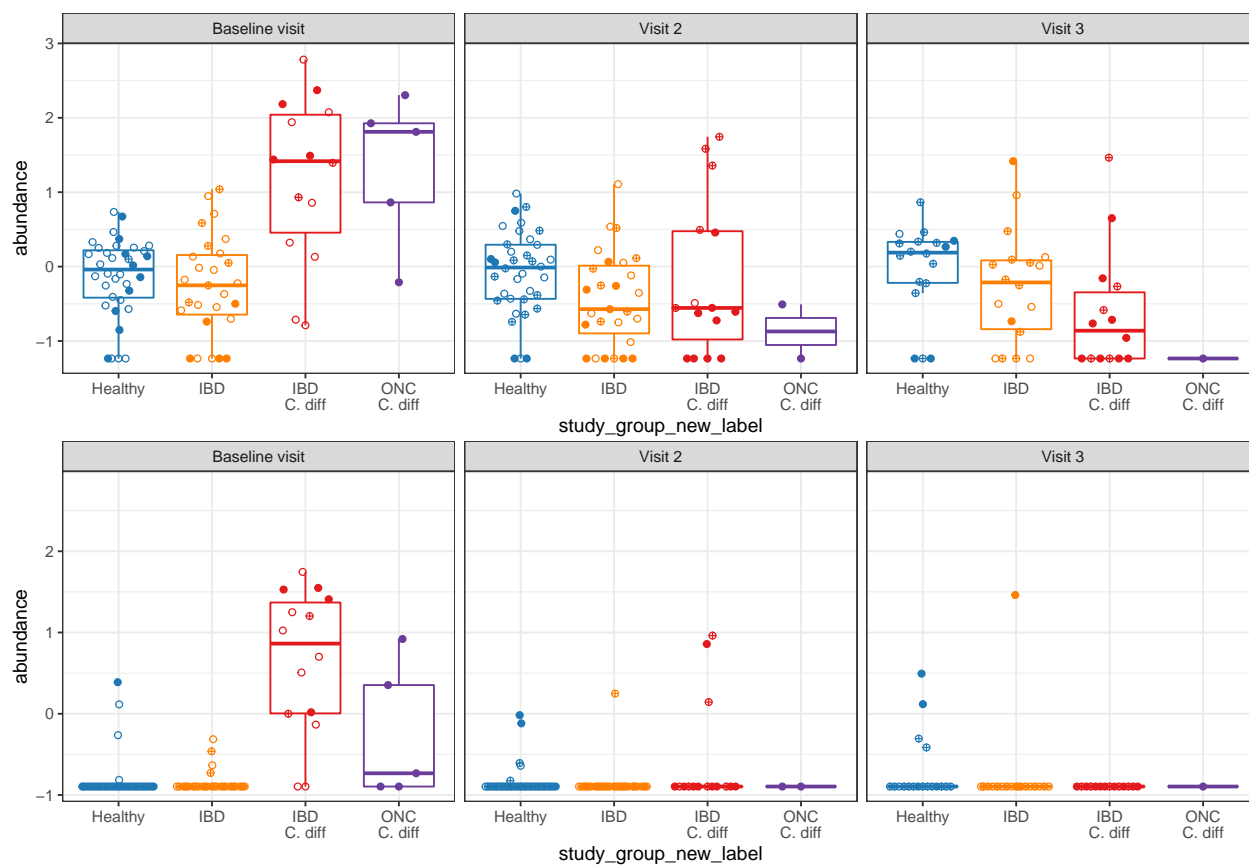


Figure 5 B

We used the randomForest model to compare severe IBD C.Diff and non-severe IBD C.Diff using the metabolomics data at the baseline visit.

This is the heatmap of the top 30 important features.

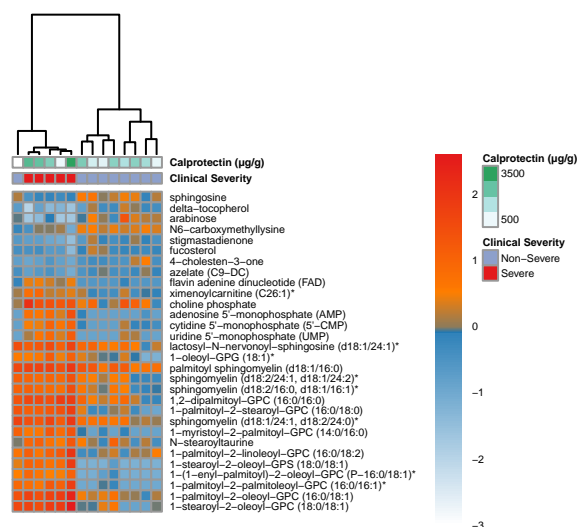


Figure 5 C

This is a boxplot to compare human DNA percentage between severe C.Diff group and non-severe C.Diff group at the baseline visit.

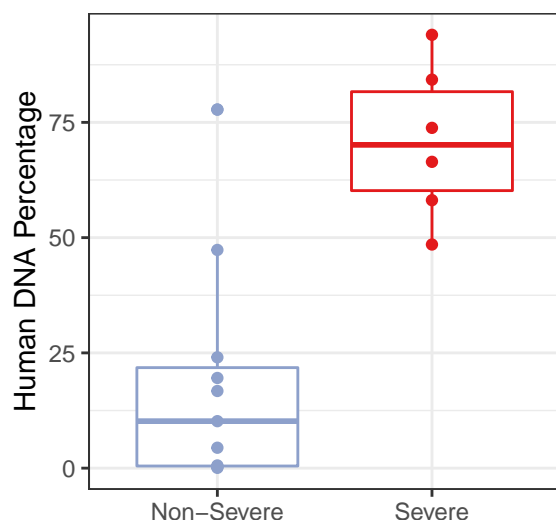
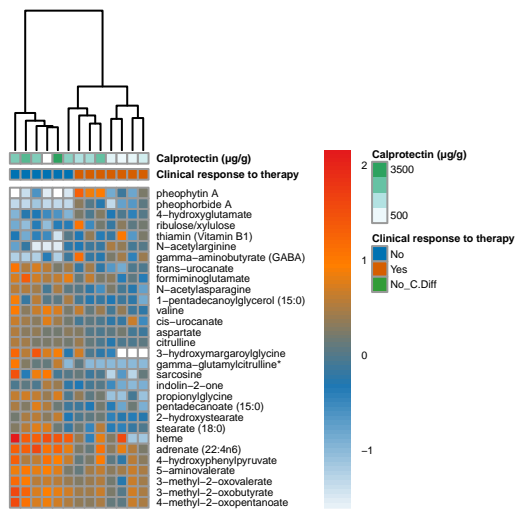


Figure 5 D

We used the linear regression to compare subjects responded to C.Diff therapy and those not responded using the metabolomics datat at the baseline visit.

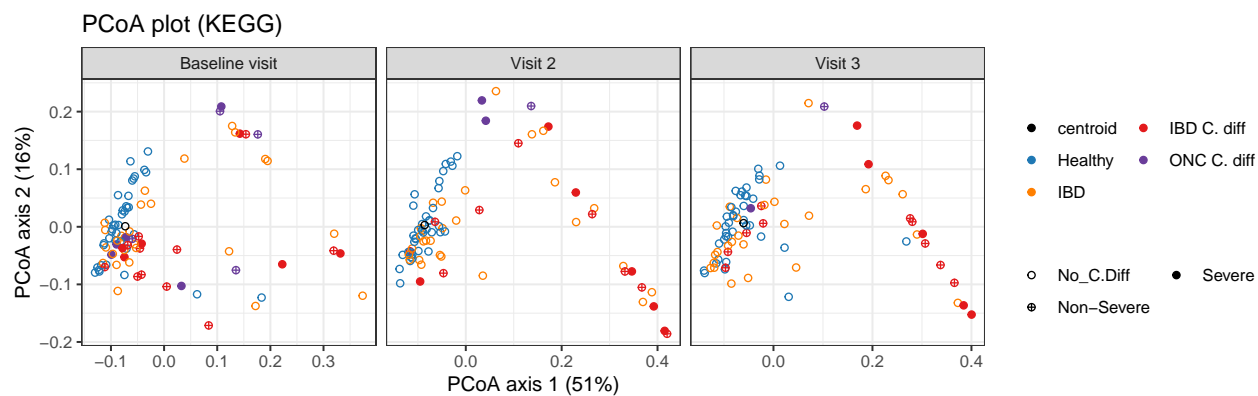
This is a heatmap for the top 30 metabolites that have raw p value < 0.05.



Supplementary Figure 2

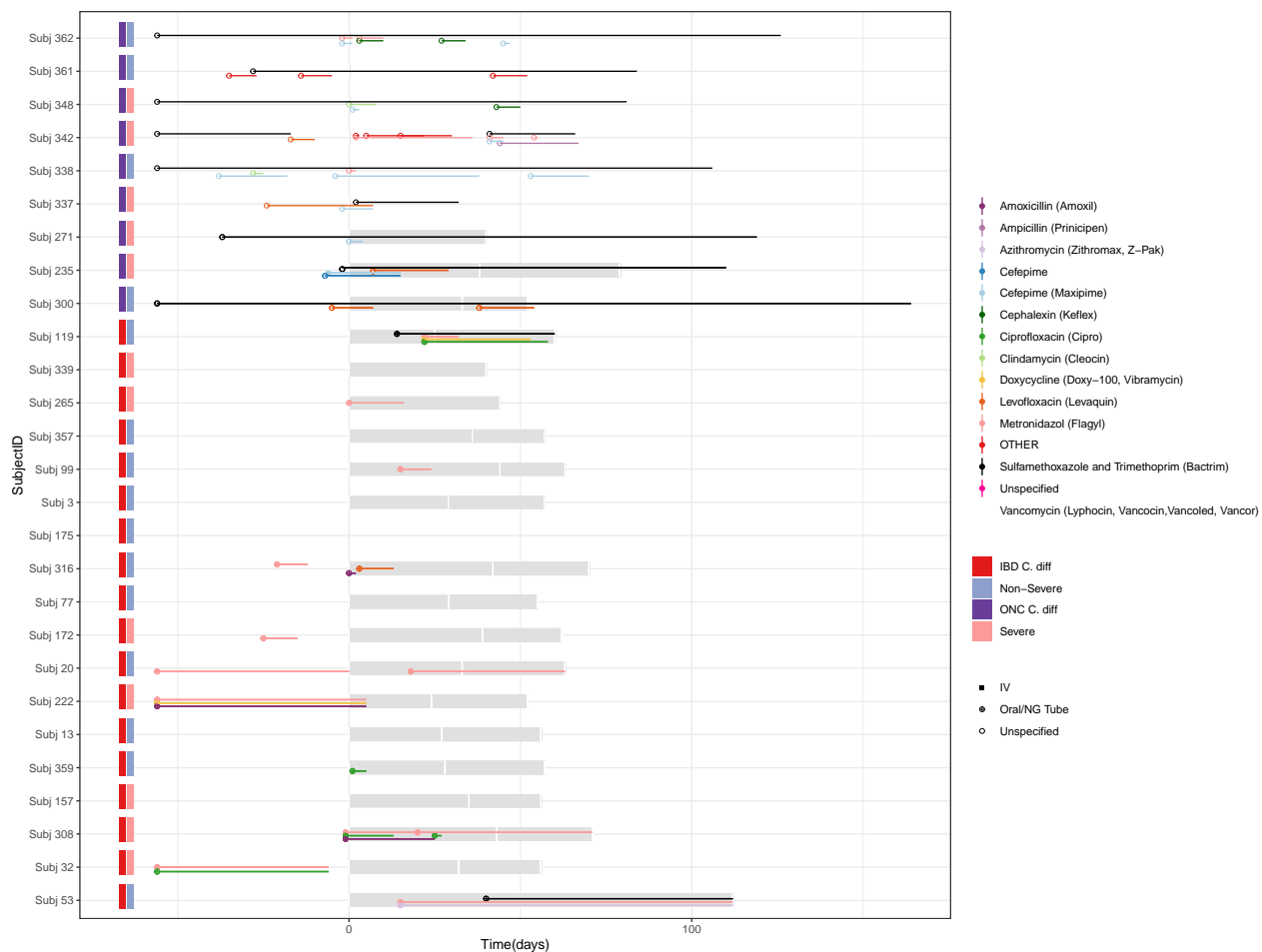
This is the PCoA plot based on Bray-Curtis distance using kegg pathway data.

The centroid is defined using all healthy controls (including Healthy \geq 6 and Healthy $<$ 6).



Supplementary Figure 3

This is a plot to summarise the antibiotics information for oncology C.Diff and IBD C.Diff patients.



The start date of the antibiotics that had started more than 56 days before the 1st sample collection is set to -56.
The relative abundance of one sample adds up to 1.
The unit of c. diff burden is CFUs/ gram of stool.