

Cell Host & Microbe, Volume 16

Supplemental Information

Rethinking “Enterotypes”

**Dan Knights, Tonya Ward, Christopher McKinlay, Hannah Miller, Antonio Gonzalez,
Daniel McDonald, and Rob Knight**

Supplementary Methods	2
Simulation of microbiomes for visualization comparison	2
Supplementary Tables	2
Supplementary Figures	3
Supplementary References	5

Supplementary Methods

Simulation of microbiomes for visualization comparison

Simulated samples have 100 taxa, with three dominant taxa drawn from a log normal distribution with mean 4, and the remaining 97 taxa drawn from a log normal distribution of mean 1. Both distributions have standard deviation 1. Samples were normalized to sum to 1. The following R code was used to simulate the taxa.

```
nsamples <- 50
ntaxa <- 100
ndominant <- 3

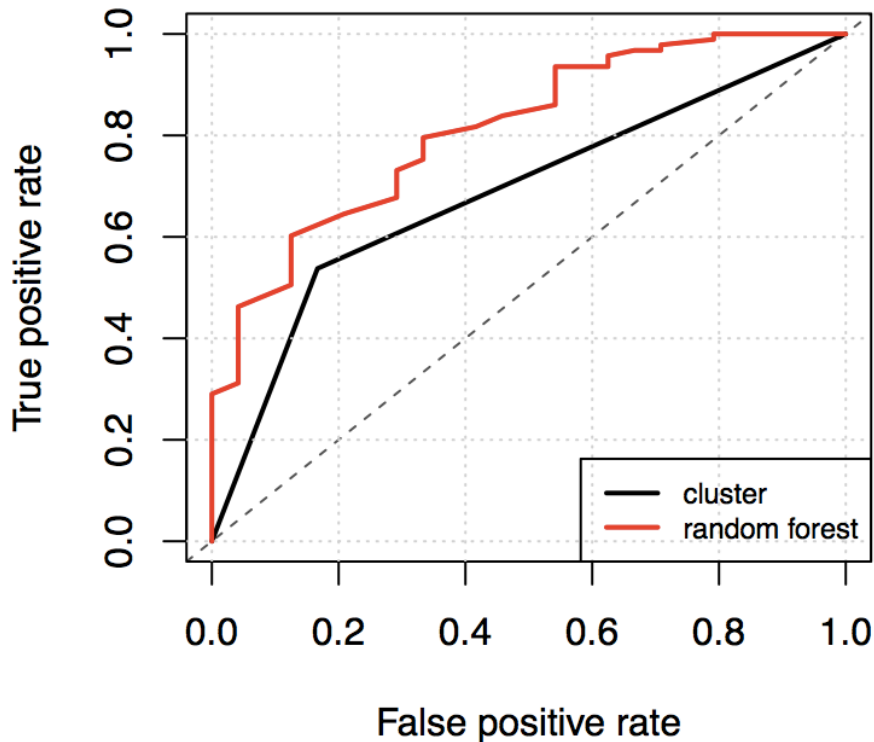
x <- matrix(rnorm(nsamples*(ntaxa-ndominant), 1), nrow=nsamples)
x <- cbind(x, matrix(rnorm(nsamples * ndominant, 4), nrow=nsamples))
x <- exp(x)
x <- sweep(x, 1, rowSums(x), '/')
```

Supplementary Tables

	Name	PMID	Sequencing platform	Region	QIIME DB #	Other notes
1	Moving pictures of the human microbiome.	21624126	GA-IIx	V4	550	
2	Structure, function and diversity of the healthy human microbiome.	22699609	Titanium	V1-3 and V3-5	968, 969	
3	Human gut microbiome viewed across age and geography.	22699611	Illumina HiSeq	V4	850	
4	Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment.	23013615	Titanium	V3-5	1460	
7	A core gut microbiome in obese and lean twins	19043404	FLX	V2	77	
8	Enterotypes of the human gut microbiome	21508958	Sanger	NA	NA	

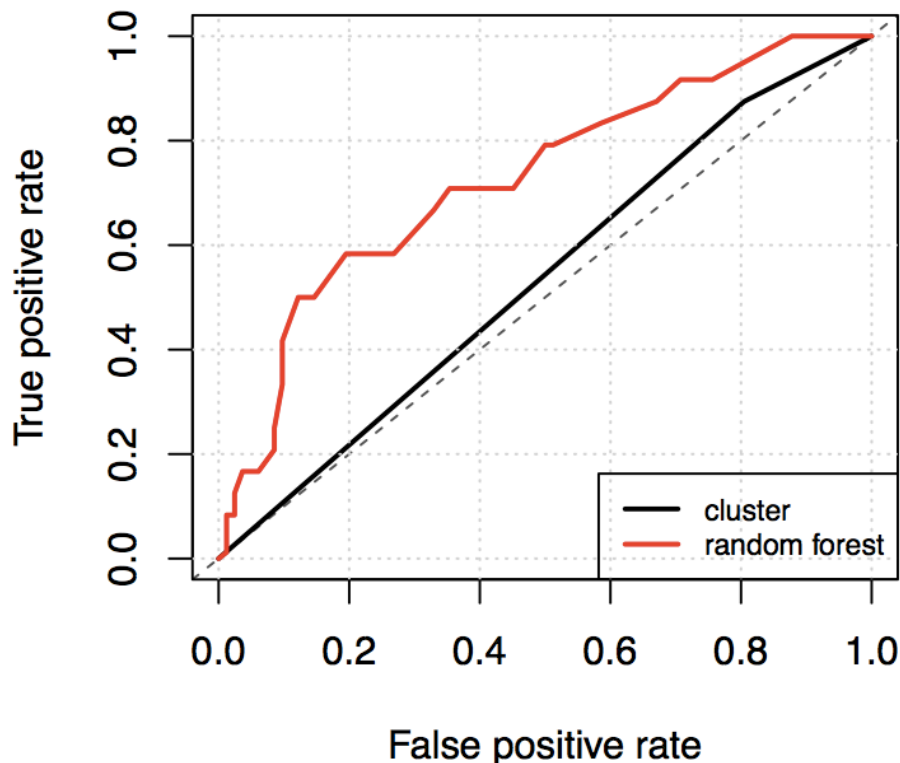
Supplementary Table 1. Data sets used in meta-analyses.

Supplementary Figures



Supplementary Figure 1. Receiver operating characteristic curve for classification of patients with Crohn's disease and healthy subjects.

To classify samples using cluster membership we first removed a single sample (the jackknife sample) from the analysis. We then performed unsupervised clustering using partitioning around medoids as described previously (Arumugam et al., 2011). We calculated the within-cluster fraction of samples belonging to each classification category (i.e. healthy vs. Crohn's disease), while excluding a single sample (the jackknife sample), and used these fractions as the estimates of the posterior probabilities for each sample in a given cluster belonging to each of the classification categories. We repeated this process with each sample being held out one time, and obtained the plotted curve as the average of the jackknifed receiver operating characteristic (ROC) curves. To obtain the random forests curve we performed the same leave-one-out jackknifing and used the out-of-bag predictions for each sample from the randomForest R package (Liaw and Wiener, 2002).



Supplementary Figure 2. Receiver operating characteristic curve for classification of obese and lean adult subjects.

To classify samples using cluster membership we first removed a single sample (the jackknife sample) from the analysis. We then performed unsupervised clustering using partitioning around medoids as described previously (Arumugam et al., 2011). We calculated the within-cluster fraction of samples belonging to each classification category (i.e. lean vs. obese), while excluding a single sample (the jackknife sample), and used these fractions as the estimates of the posterior probabilities for each sample in a given cluster belonging to each of the classification categories. We repeated this process with each sample being held out one time, and obtained the plotted curve as the average of the jackknifed receiver operating characteristic (ROC) curves. To obtain the random forests curve we performed the same leave-one-out jackknifing and used the out-of-bag predictions for each sample from the randomForest R package (Liaw and Wiener, 2002).

Supplementary References

- Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.-M., et al. (2011). Enterotypes of the human gut microbiome. *Nature* 473, 174–180.
- Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. *R News* 2, 18–22.