UniFrac workshop

Hosted by Ruth from the Gloor lab

What

UniFrac measures the distance between two microbiome samples.

It requires

1) The count table of counts per taxa per sample
2) A phylogenetic tree

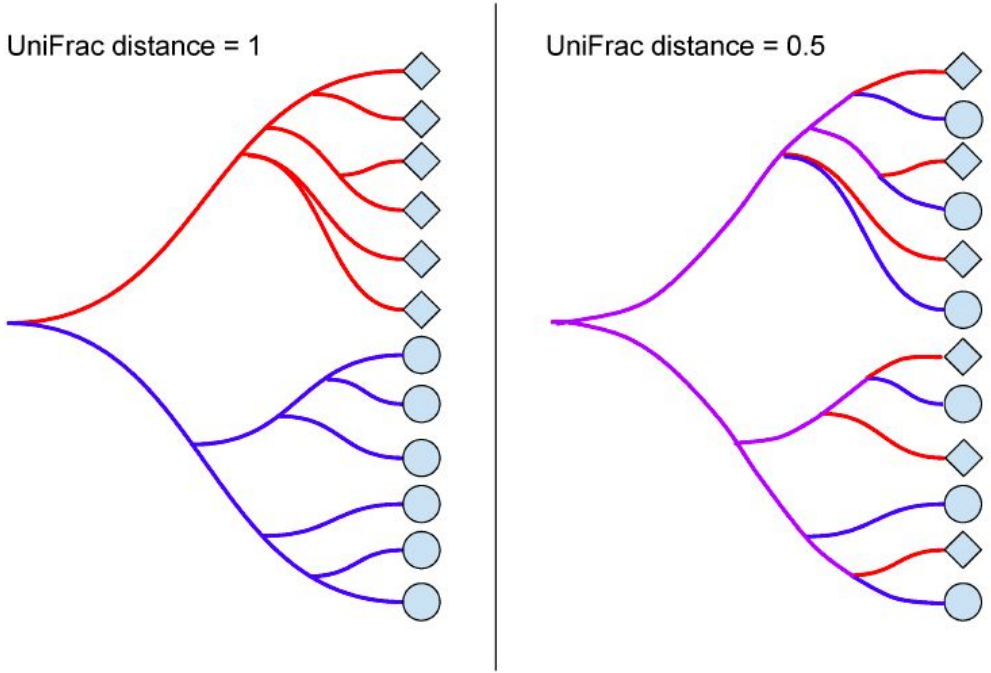
How

Unshared branch lengths

divided by

Total branch lengths

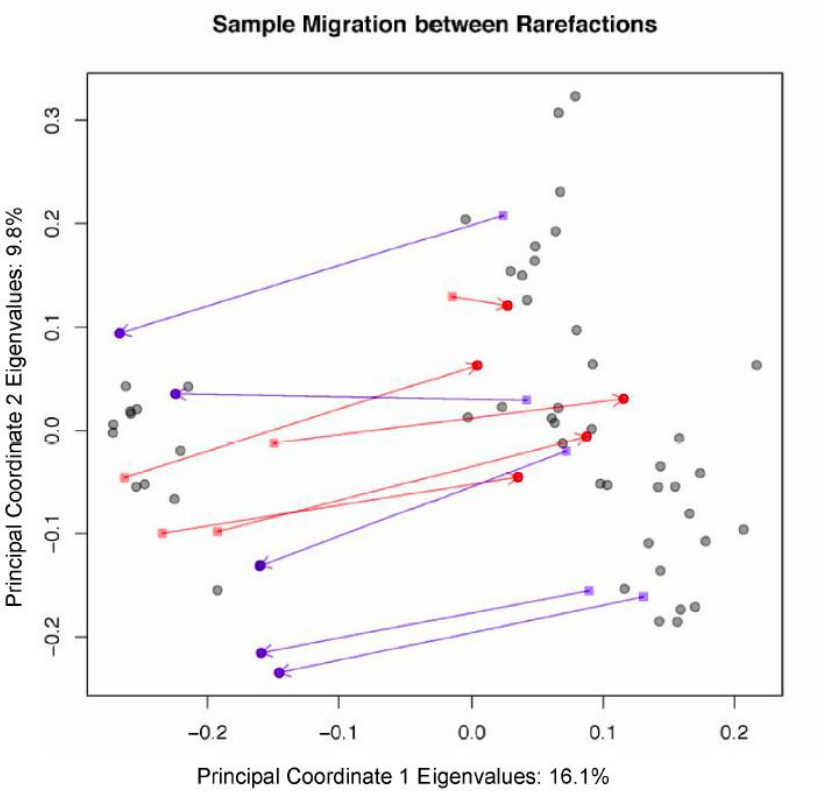
** you MUST rarefy your samples to the same sequencing depth to use Unweighted UniFrac



Weighting

- In classic unweighted UniFrac, results can be randomly skewed by low-count taxa, which are randomly detected or not detected (especially if you are rarefying!)

- The branch lengths of the phylogenetic tree can be multiplied by a ‘weight’, a number related to the children taxa, to prevent this.



Weighted UniFrac

- weights tree branches by the difference in taxa proportional abundance between the two samples.
- Low count taxa don’t significantly affect the measurement.

$$u = \sum_{i}^{n} b_i \times \left| \frac{A_i}{A_T} - \frac{B_i}{B_T} \right|$$

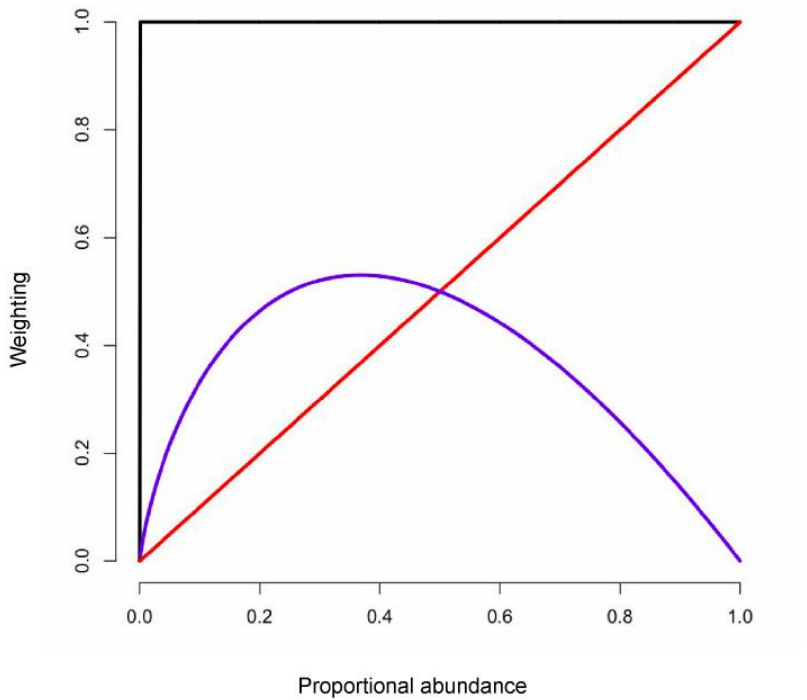
Weighted UniFrac

$$D = \sum_{j}^{n} d_j \times \left(\frac{A_j}{A_T} + \frac{B_j}{B_T} \right)$$

Weighted UniFrac scaled between 0 and 1

Information UniFrac

- weights tree branches by uncertainty
- a 50/50 composition is more uncertain than a 90/10 composition
- incorporates the taxa abundance evenness



black is unweighted UniFrac

red is weighted UniFrac

blue is information UniFrac

Ratio UniFrac

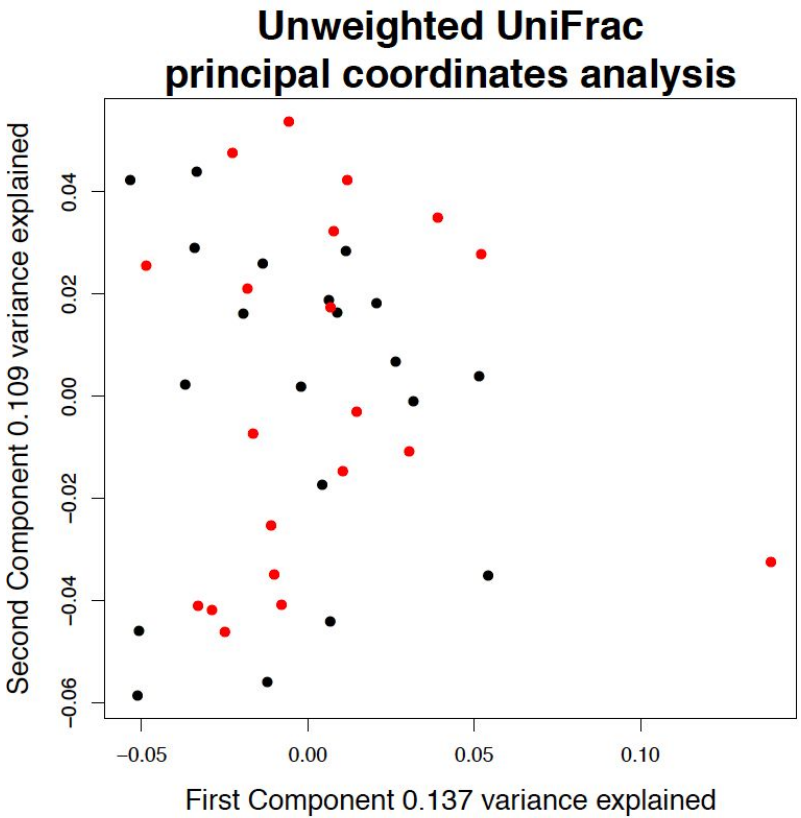
- weighted by taxa abundance, except that the taxa abundance are divided by the geometric mean
- geometric mean serves as a baseline taxa abundance
  - This is what we use in ALDEx, but without the logarithm

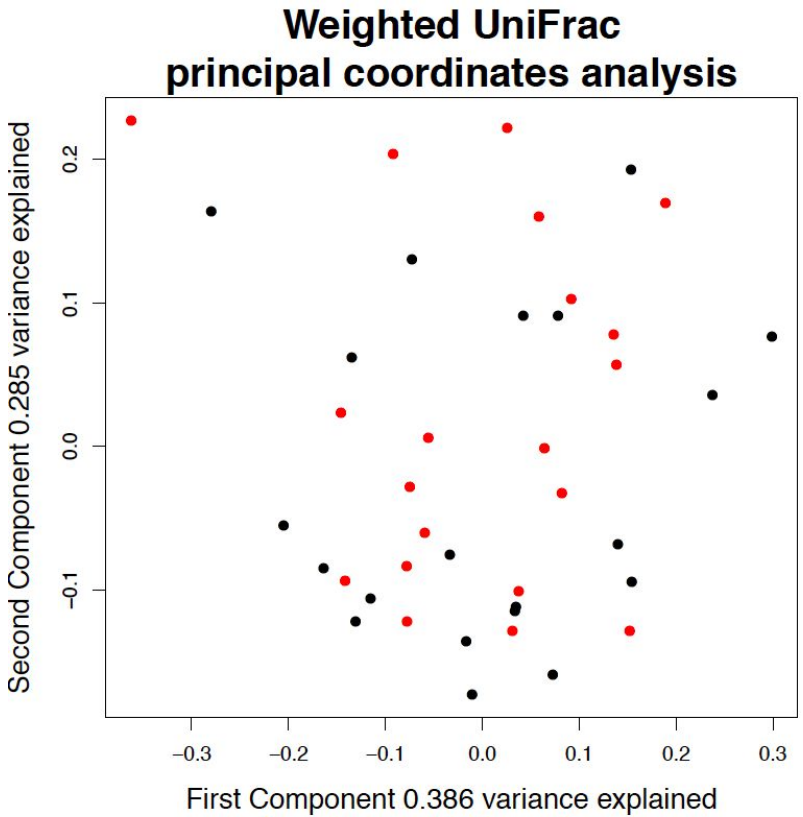
$$\sum_{i}^{n} b_i \times \left| \frac{\frac{A_i}{A_T}}{gm(A_i)} - \frac{\frac{B_i}{B_T}}{gm(B_i)} \right|$$

General Usage

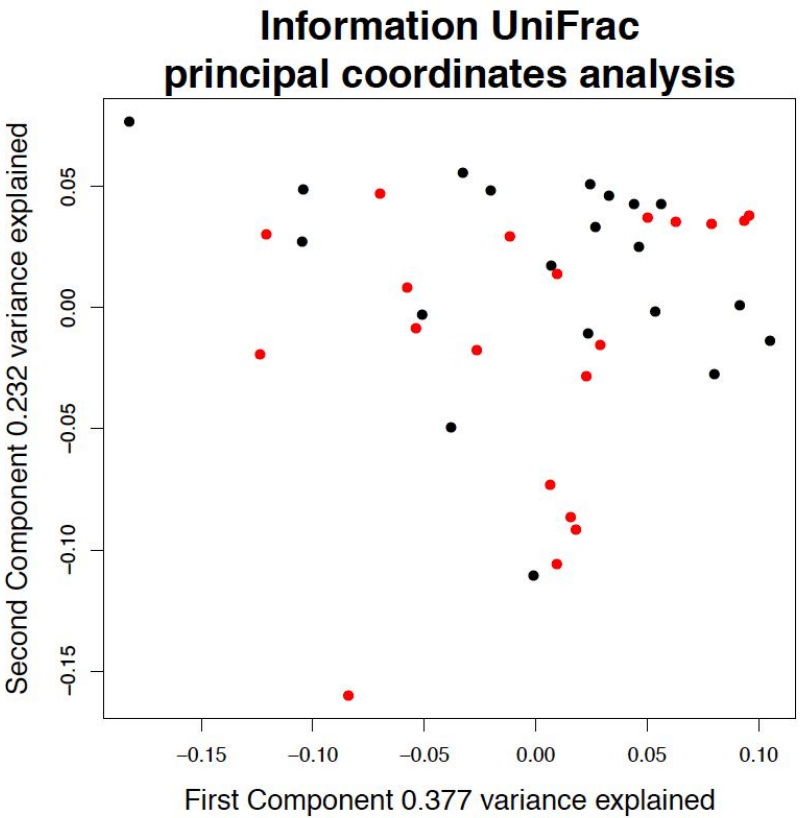
- Use UniFrac to find the difference between each pair of samples
  - distance matrix
- Throw it into a PCoA plot
- Look at the variance explained
  - Higher is better
  - More on PC1 vs PC2 is better (separation vs. ball of points)
- Look at how the data separates
  - Color by different metadata

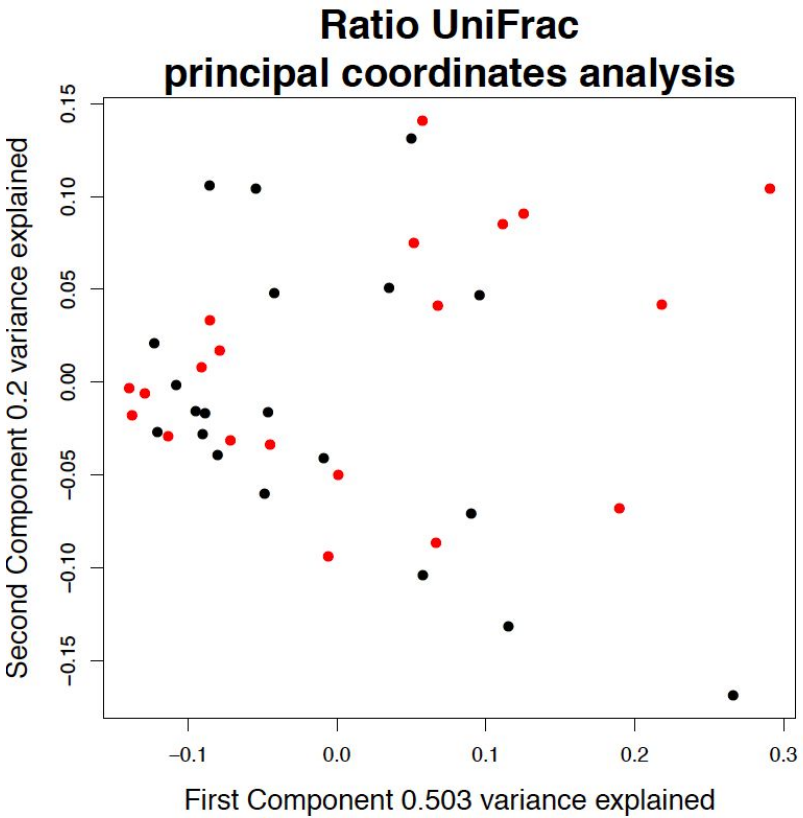
Comparison: No difference





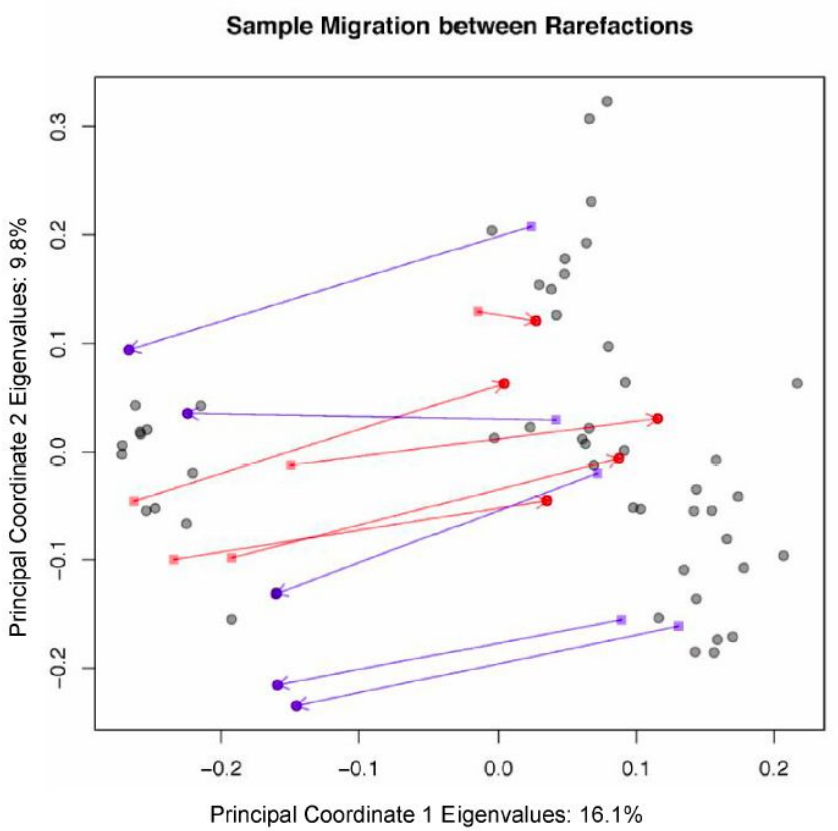
Comparison: No difference



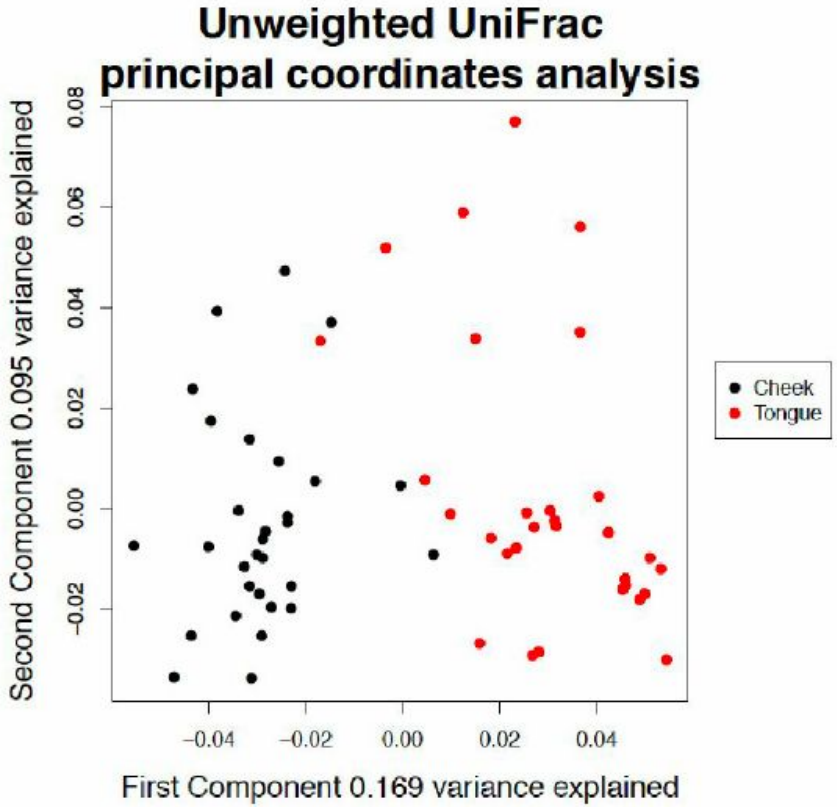


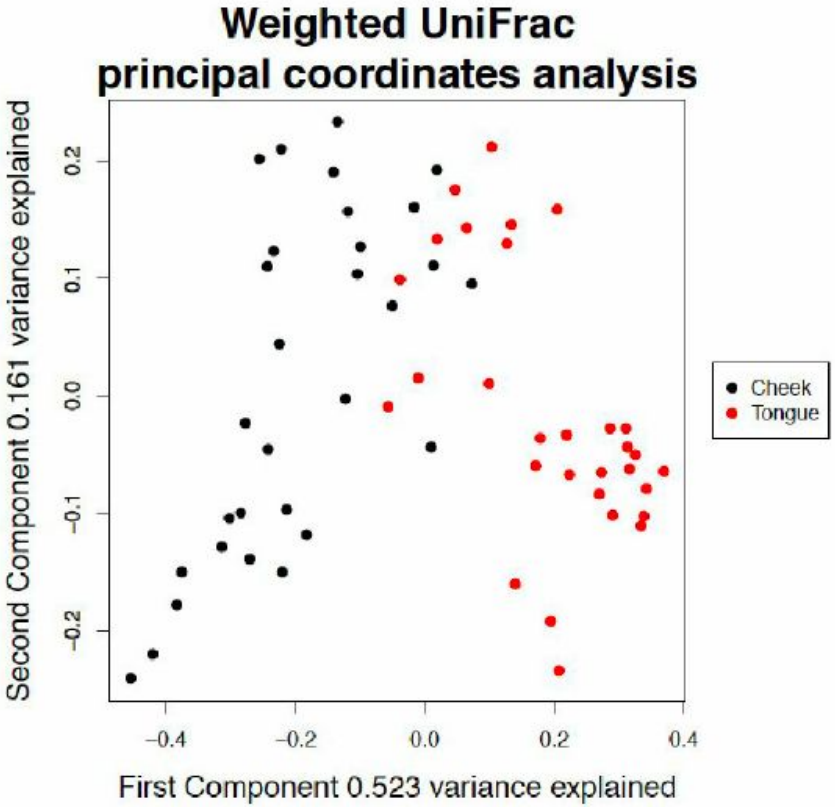
Comparison: No difference - CAVEAT

Unweighted UniFrac can give you vastly different results in different rarefaction instances.

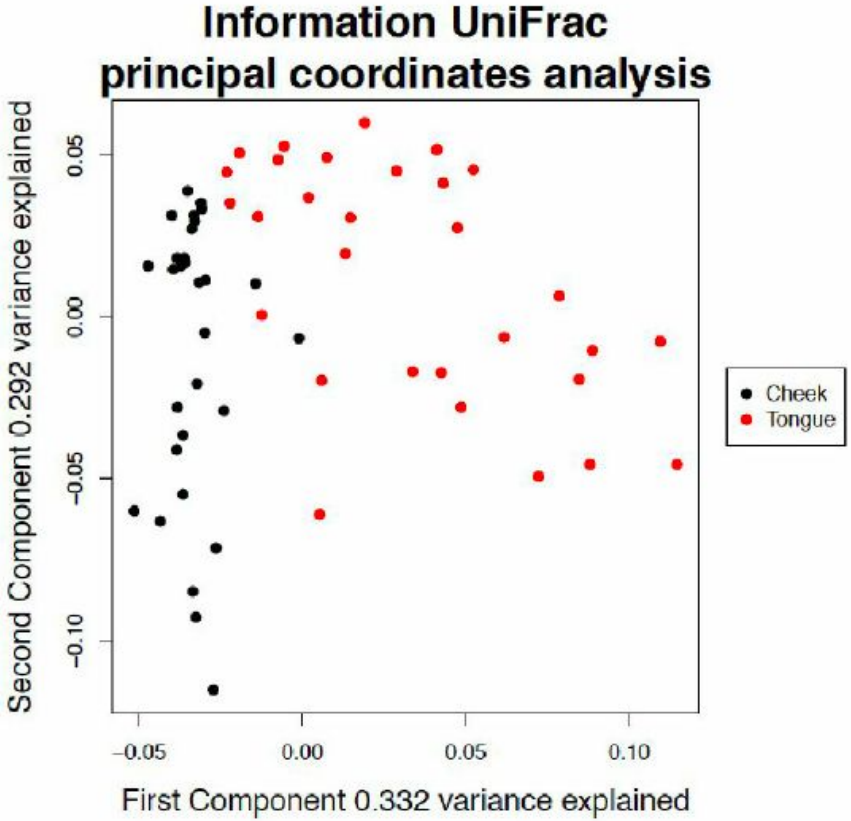


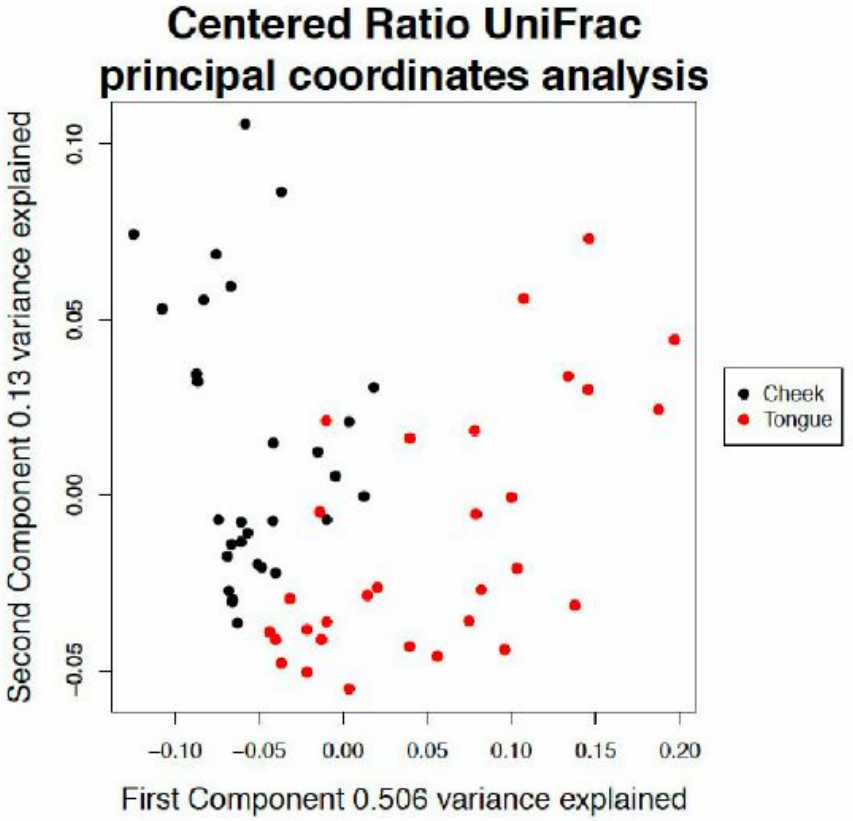
Comparison: Obvious difference



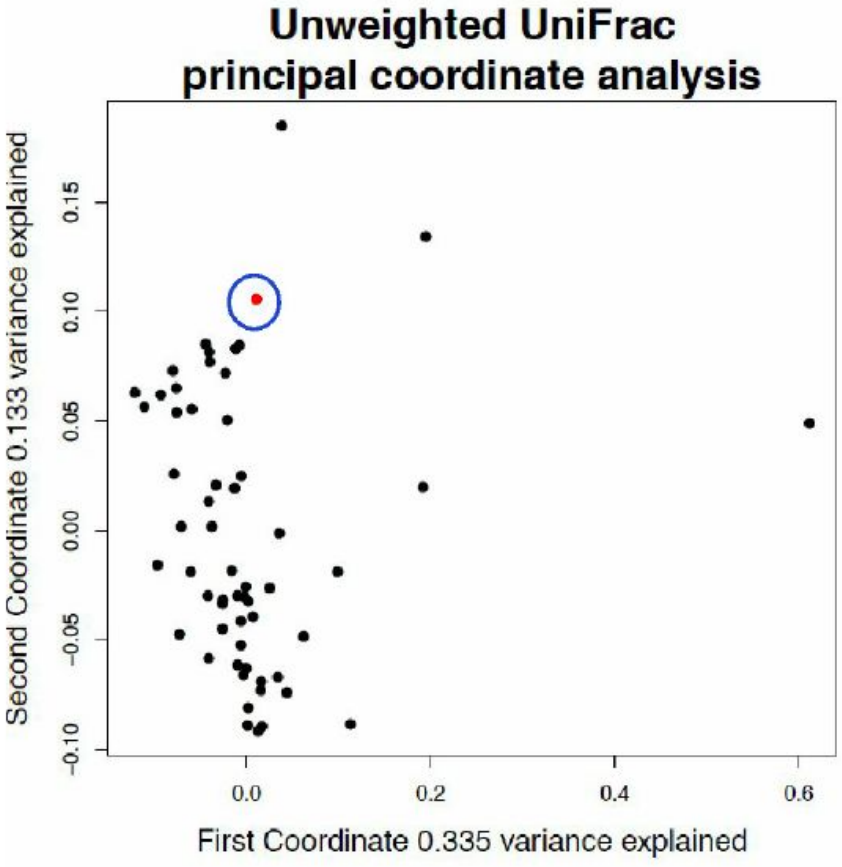


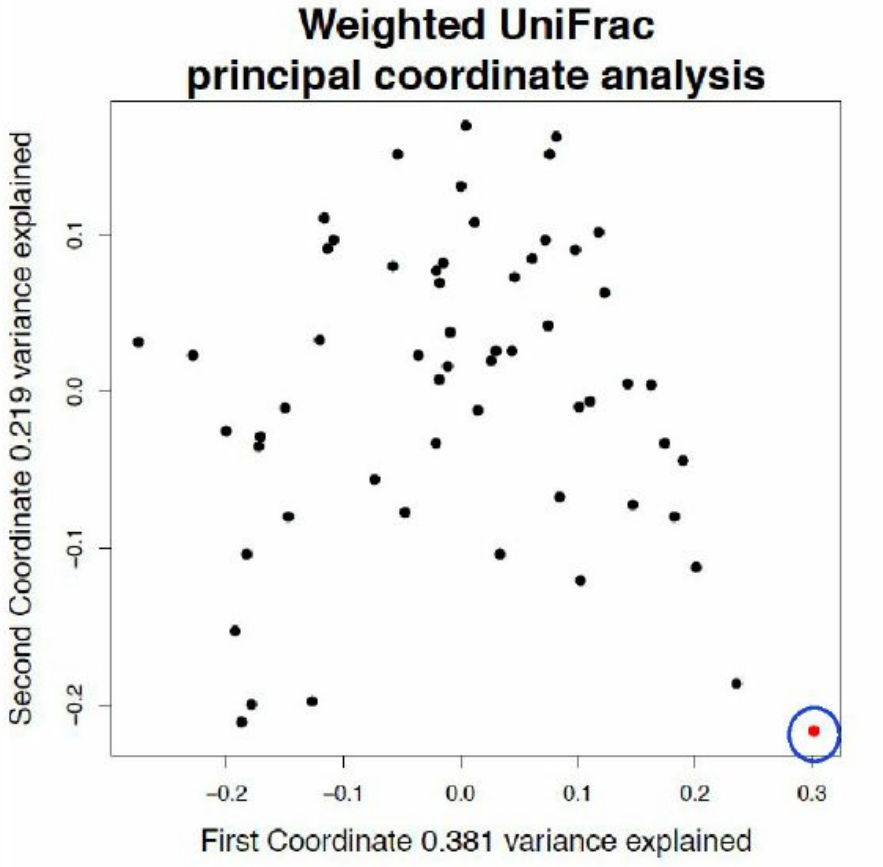
Comparison: Obvious difference



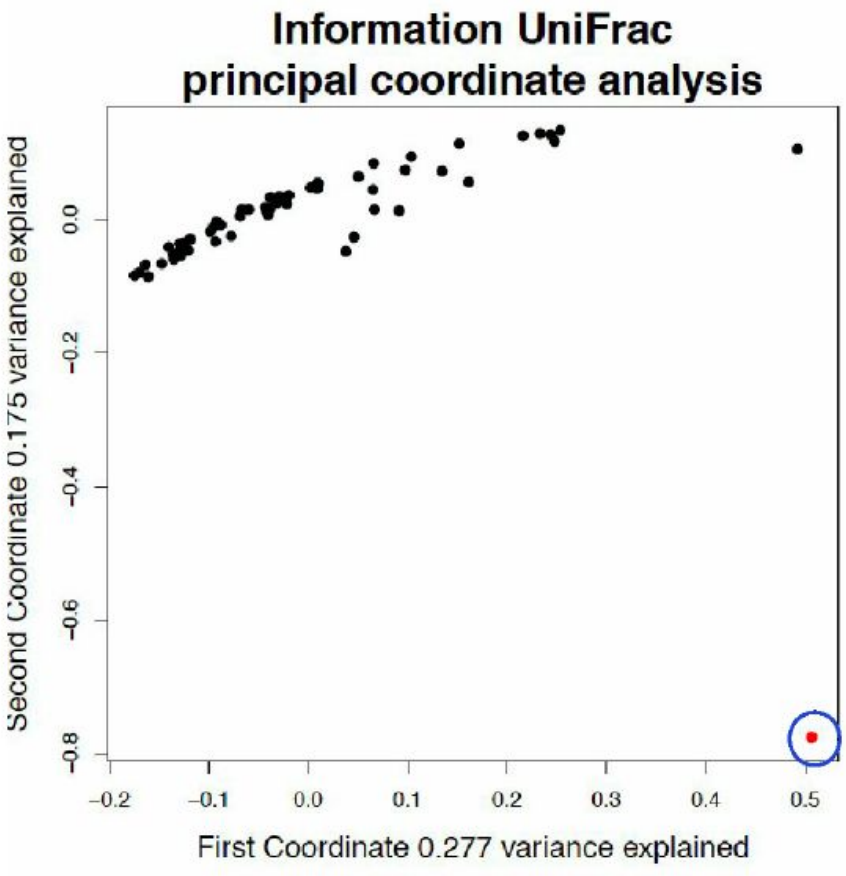


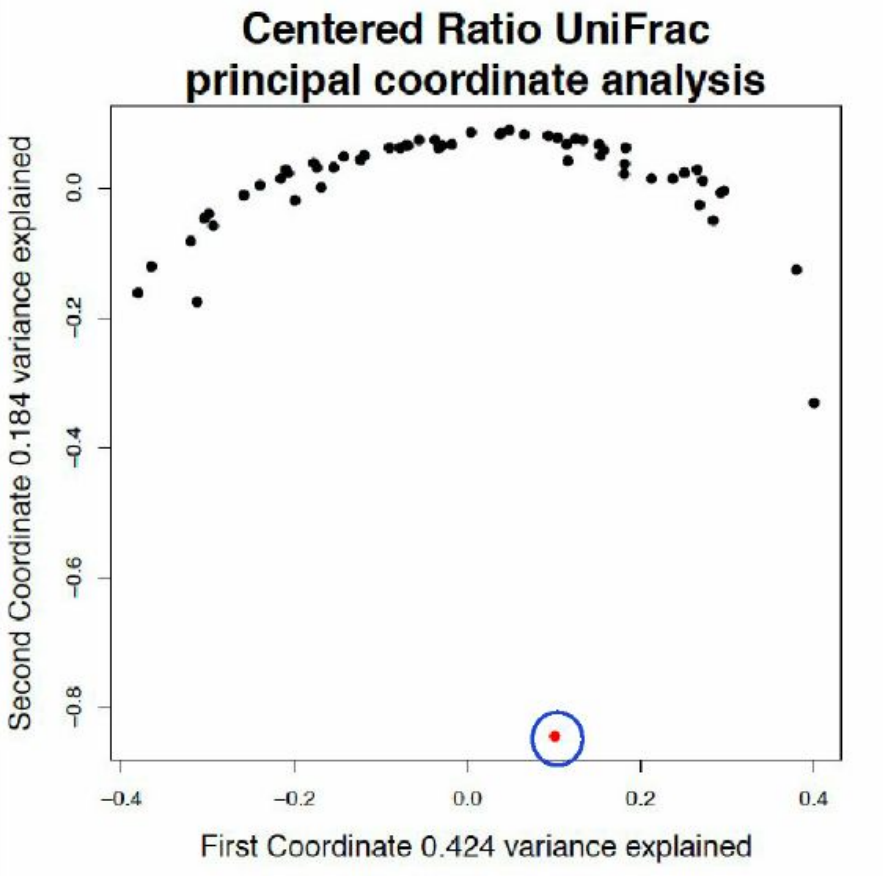
Comparison: Outlier





Comparison: Outlier

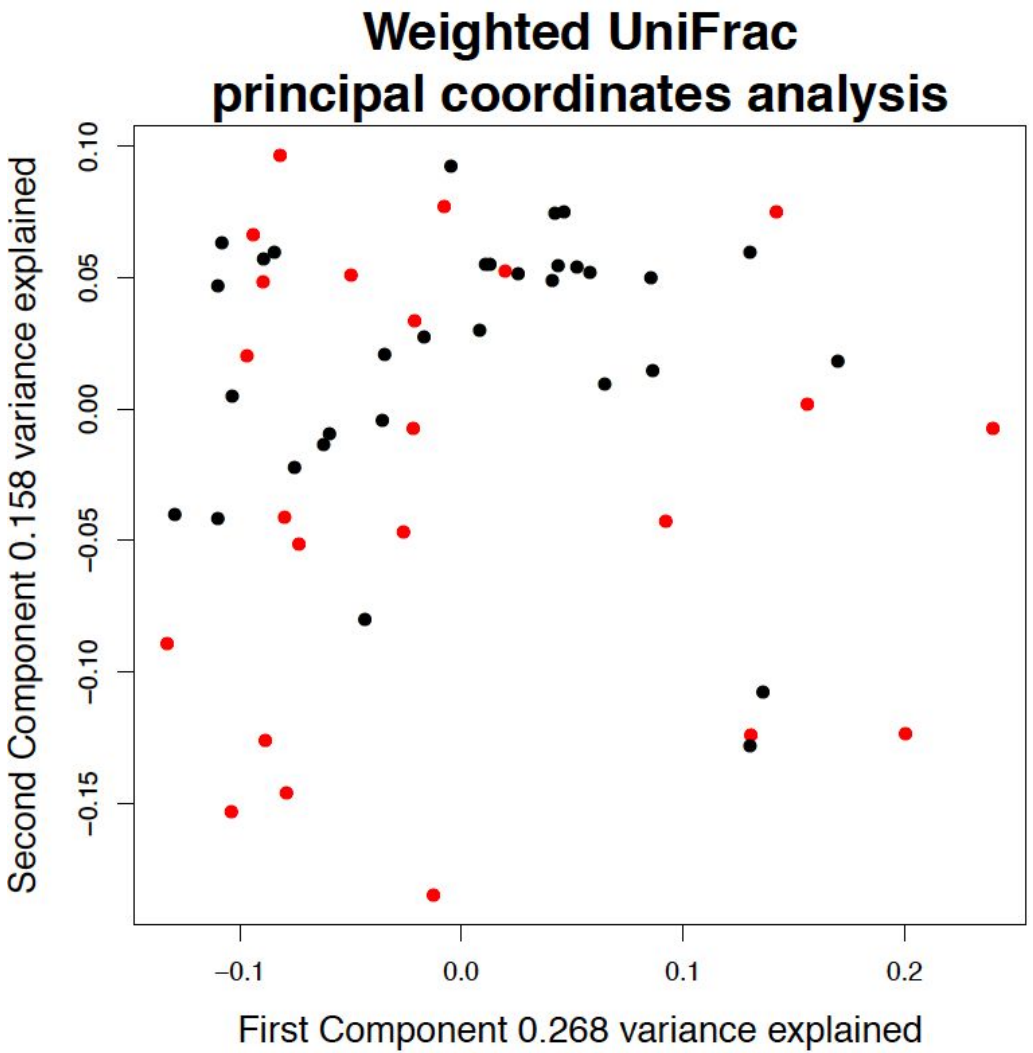




Using metadata

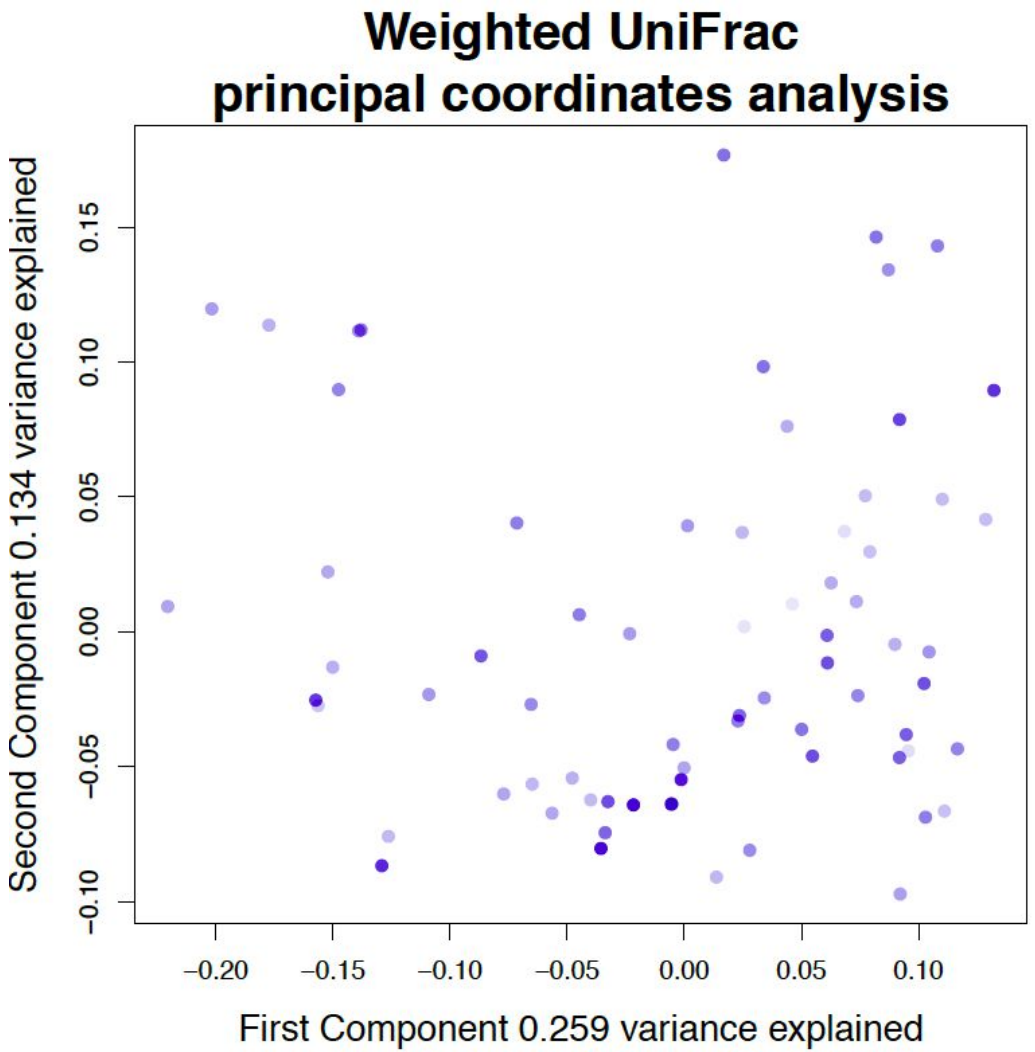
Healthy vs. NASH

- run ALDEx on this too



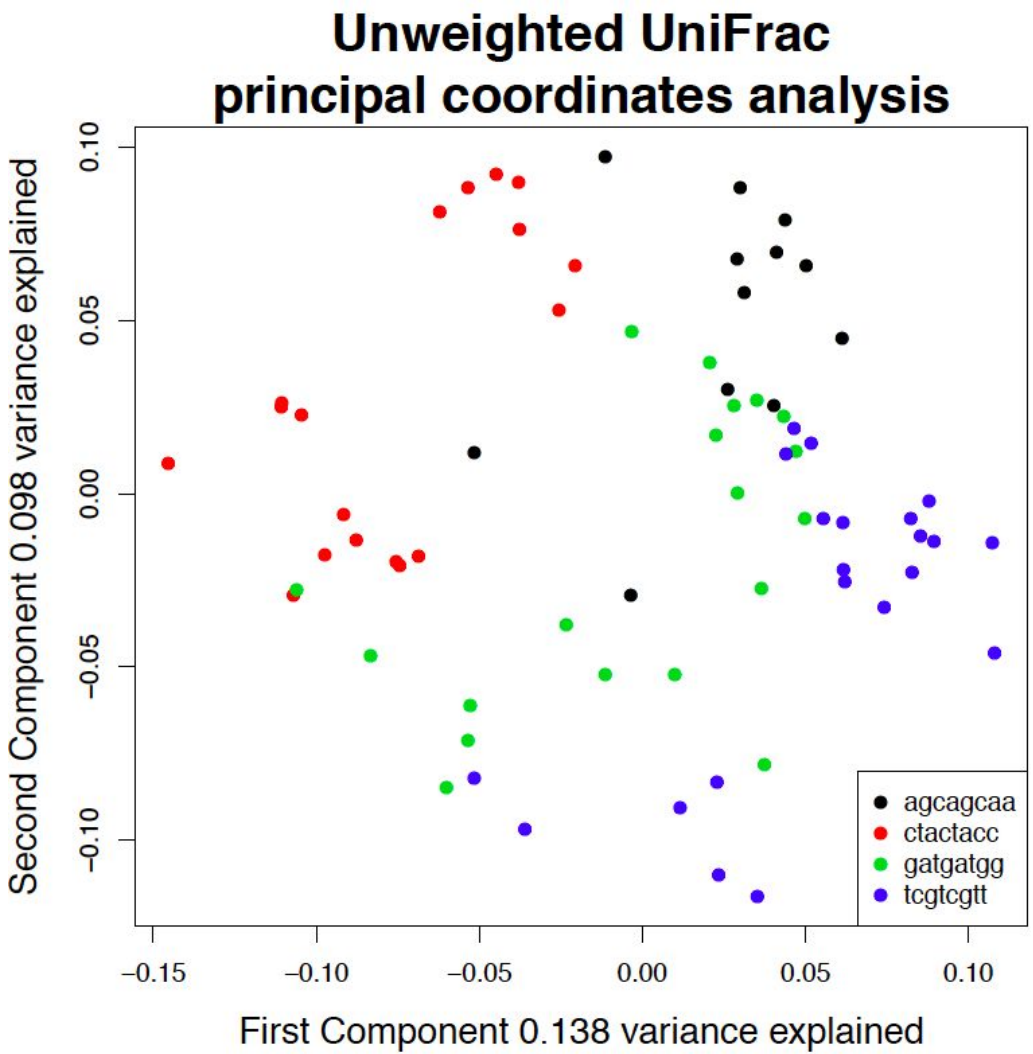
Using metadata

BMI



Using metadata

left barcodes



Exploration

- LOOK AT YOUR DATA
- Run all the different weightings!
- Figure out why you have outliers
  - Should you exclude them?
- Make sure you have real differences, not artefacts

Talk to Greg, Jean, or Ruth when you are

- Planning out your study
- Analyzing your data

There are a lot of nuances about sample collection, extraction, and study design that Greg and Jean can prevent you from messing up