



Question

Michael Agronah

2023-01-21

Research goals:

My goal is to estimate the power of differential abundance microbiome studies for a given range of abundance and effect sizes. For example, on average, what is the power of an OTUs with mean abundances [100,200] and fold changes [1,2]. I show illustrations with the heatmaps in figure 1.

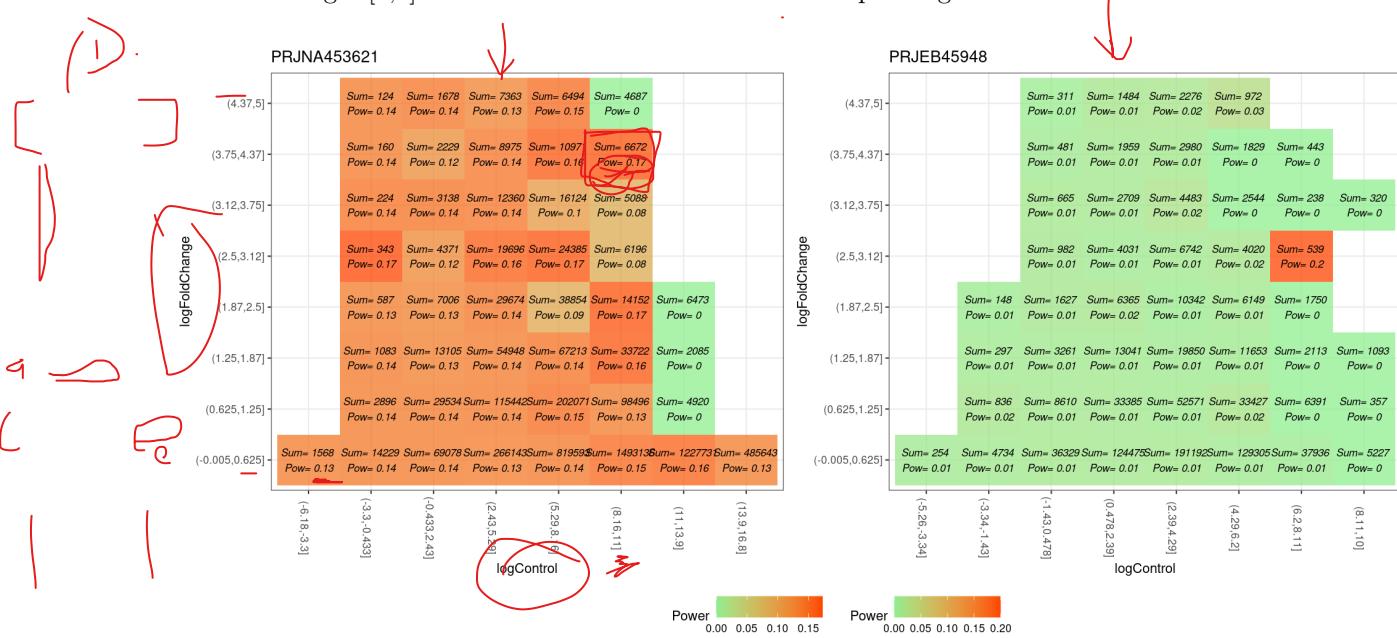


Figure 1: Heatmap showing the power and total control abundances in grids control abundance and fold changes

Relationship between control abundances and foldchanges for 4 datasets

1. Plot of the relationship between fold changes and control abundance is shown in figure 2
2. Investigating the scale-location plot shown in figure 3, I found I can model the variance in relationship by some quadratic function.
3. The abundances follow a skew normal distribution and I fitted a truncated Cauchy to fold changes (figure 4 - 5).

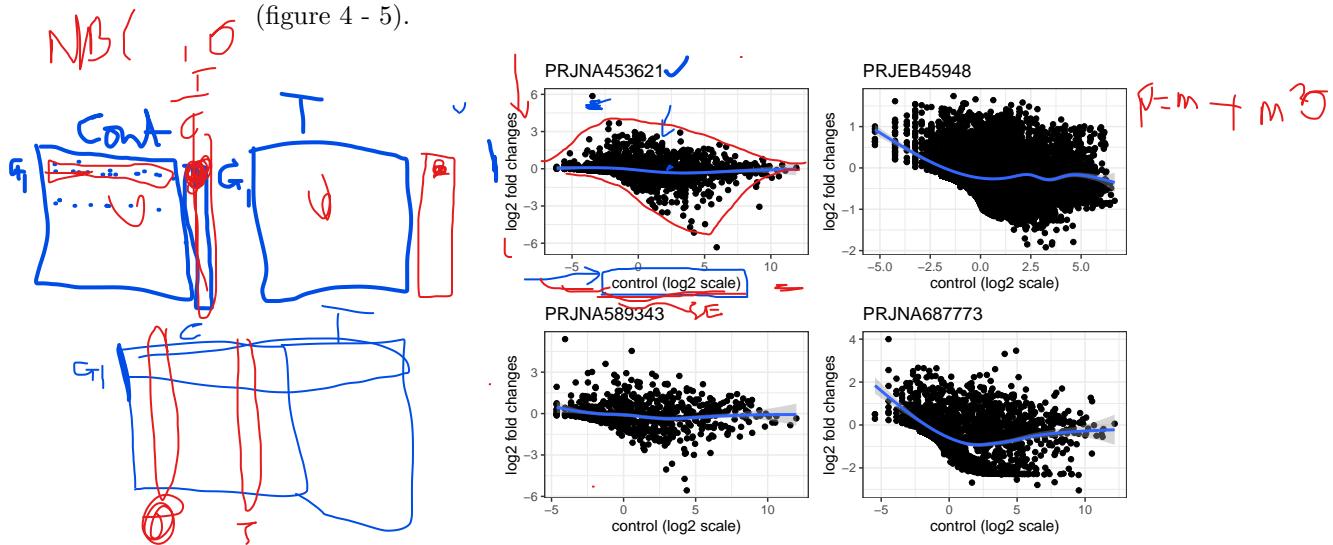


Figure 2: Plot of control against effect sizes for datasets with assessment numbers PRJNA453621, PRJEB45948, PRJNA589343 and PRJNA687773

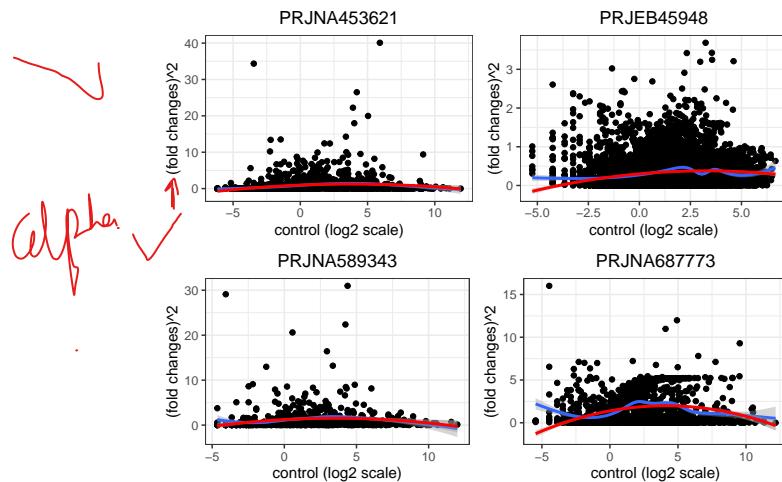


Figure 3: Scale-location plot of control against squared fold changes for datasets with assessment numbers PRJNA453621, PRJEB45948, PRJNA589343 and PRJNA687773. (red curve is a fitted polynomial of degree 2)

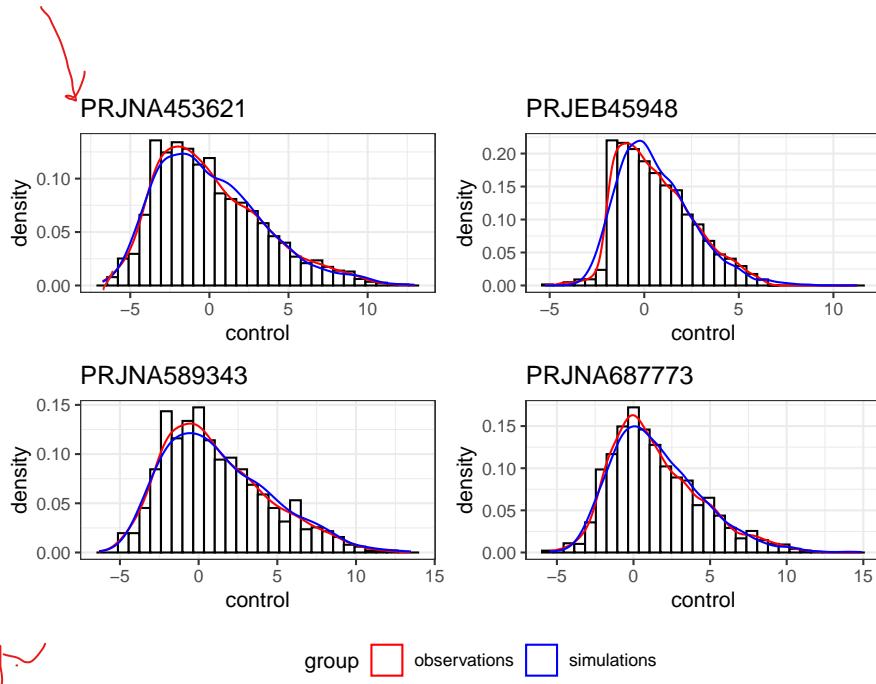


Figure 4: Histogram and density plots for control abundances and density from a simulated sample from fitting a truncated normal distribution

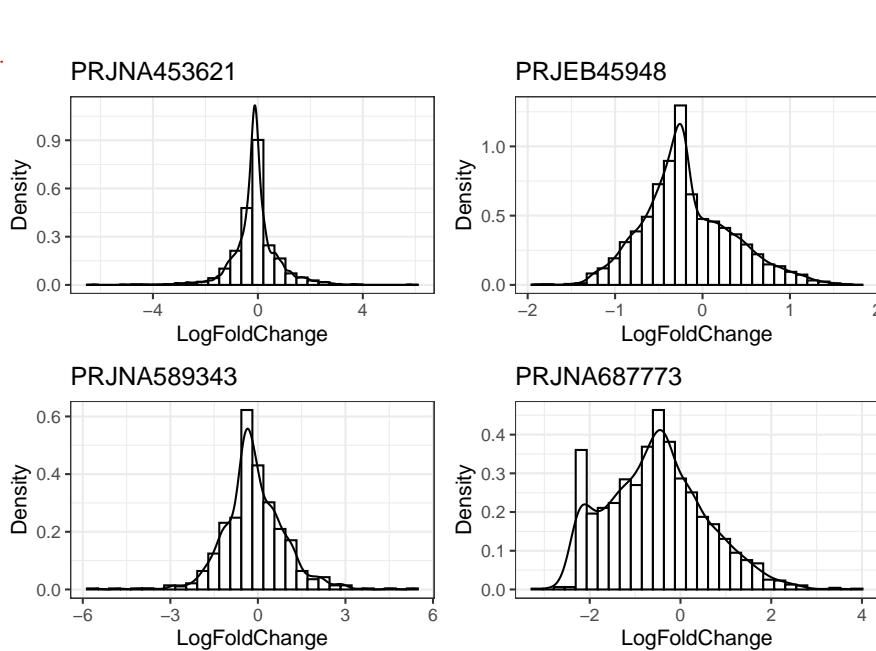


Figure 5: Histogram and density plots for control abundances and density from a simulated sample from fitting a Cauchy distribution to the data

My Question

1. Figure 6 shows the results from simulating control abundances and fold changes and calculating power.
2. I define power as the number of p-values (with multiple hypothesis correction accounted for by the Benjamin-Hochberg correction) that are less than a significant level of 0.1.
3. I don't seem to be making sense of the plots as there are no clear patterns. Wanted to check up with you on your thoughts about this plots.
4. I standardised the fold changes by dividing by the an estimated standard error to attempt normalising it but that doesn't solve the lack of pattern.
5. Plot with standardised fold changes are on the second column and non-standardised fold changes are on the first column.

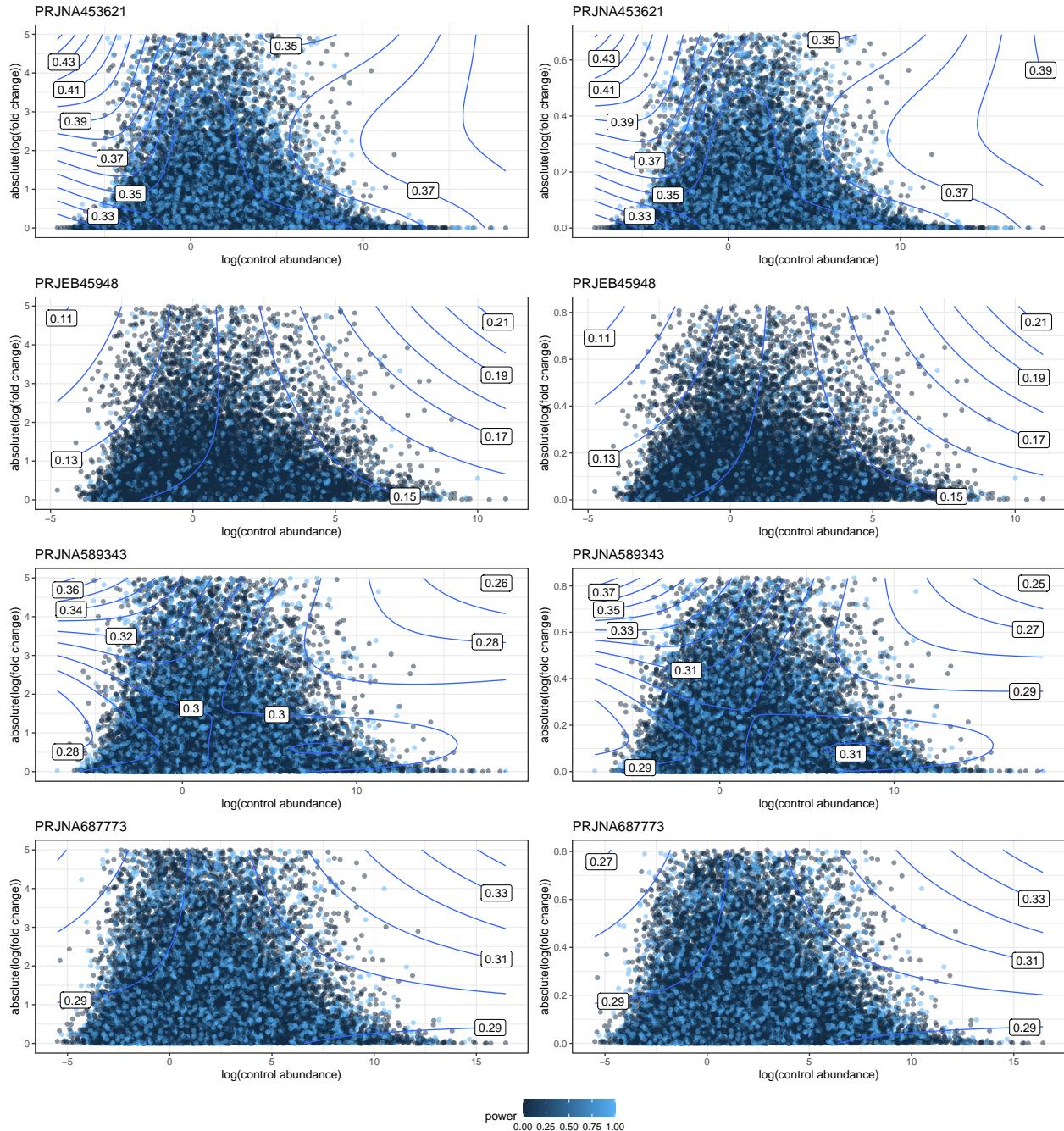


Figure 6: Heatmap showing the power and total number of taxa in grids of control and effect sizes