Cell hierarchy statistical tests

Maxwell W. Libbrecht

Terminology: Let K be the number of markers (9 in our data set). Define a population as a particular assignment to certain markers (i.e. A+B-C+). Let $P(p)$ be the proportion of cells in population p.

1 Expected proportions

One problem with using performing statistics using the proportion of cells in each population is that the proportions of different populations are dependent on one another. For example, if a perturbation (such as a gene KO) causes an increase in A+B+ cells, we will probably also see an increase in A+B+C+ and A+B+C-. So if we see an increase in A+B+C+, we have to ask, is that increase just a result of the increase in A+B+, or is there an additional increase in A+B+C+ on top of that?

We can get around this problem by accounting calculating the expected proportions in layer d given layers $1 \dots d-1$. We do this as follows.

Consider layer d. We would like to calculate the expected value of $P(A_{1:d})$. Assume that A_1 and A_2 are independent given $A_{3:d}$, and likewise for all other pairs of markers. Then,

$$P(A_1|A_{2:d}) = P(A_1|A_{3:d}) \tag{1}$$

$$\frac{P(A_{1:d})}{P(A_{2:d})} = \frac{P(A_1, A_{3:d})}{P(A_{3:d})} \tag{2}$$

$$P(A_{1:d}) = \frac{P(A_{2:d})P(A_1, A_{3:d})}{P(A_{3:d})} \tag{3}$$

Multiplying the above equation together for every i, j pair,

$$P(A_{1:d})^{\binom{d}{2}} = \prod_{i,j} \frac{P(A_{1:d\setminus i})P(A_{1:d\setminus j})}{P(A_{1:d\setminus\{i,j\}})} \tag{4}$$

$$= \frac{\prod_i P(A_{1:d\setminus i})^{d-1}}{\prod_{i,j} P(A_{1:d\setminus\{i,j\}})} \tag{5}$$

$$P(A_{1:d}) = \left(\frac{\prod_i P(A_{1:d\setminus i})^{d-1}}{\prod_{i,j} P(A_{1:d\setminus\{i,j\}})} \right)^{1/\binom{d}{2}} \tag{6}$$

One way to derive step (5) is to count factors: There are $2 * \binom{d}{2}$ total factors in the numerator, made up of d unique factors. Therefore, each unique factor has an exponent of $2 * \binom{d}{2}/d = d-1$.

Equation (6) gives the expected counts for a population in layer d given layers $1 \dots d-1$. We can evaluate the difference from expected using the log observed/expected enrichments, $\log P(A_{1:d})/\tilde{P}(A_{1:d})$.

1.1 Results

I wondered whether we would have more statistical power by comparing enrichments rather than raw population proportions. For each population, I performed a two-sided Wilcox rank-sum test to compare whether the enrichments in the 6 KO samples were uniformly higher or lower than enrichments in the 70 WT samples. I did the same using the raw proportions. As a null, I shuffled the KO/WT labels among the samples and performed the same test (using raw proportions).

Figures 1 and 2 show Q-Q plots of the distribution of p-values compared to a uniform distribution (uncorrected and corrected p-values respectively). A Q-Q plot against uniform is a way of displaying a 1D distribution, analogous to a histogram or eCDF. If the p-values followed a uniform distribution, they would follow a X=Y line. Normally, we hope to see two things in the Q-Q plot. First, we hope that our null test falls close to the X=Y line, indicating that the test is not overly liberal or conservative. The shuffled data falls close to the X=Y line, as we would hope.

Second, we hope that the smallest p-values from our real test fall well below the X=Y line, indicating that some tests are statistically significant. On the contrary, we see something strange in our real data. Almost all of the points lie below the X=Y line, indicating statistical differences in almost every population. However, this is not true of the smallest p-values; the smallest p-values are no smaller than we would expect by chance.

This is true for both the proportion p-values and the enrichment p-values. The enrichment p-values lie slightly above the proportion p-values indicating that, unfortunately, they give slightly less statistical power.

2 Population groups

Define a group of populations to be all populations at the same depth that involve the same markers. That is, A+B+, A+B-, A-B+ and A-B- are all part of the AB group.

There are $\sum_{d=1}^{K}\binom{K}{d} = 2^K$ groups.

2.1 If you know the proportion of one group member, you know the proportions of all group members

I realized that you can infer the proportions of every population given just one member of each group. We will prove this the simple case where we know $P(A+)$, $P(B+)$ and $P(A+B+)$. We can infer the rest of the depth-2 populations as follows:

$$\text{P(A+B-)} = \text{P(A+)} - \text{P(A+B+)}$$
$$\text{P(A-B+)} = \text{P(B+)} - \text{P(A+B+)}$$
$$\text{P(A-B-)} = 1 - \text{P(A+B+)} - \text{P(A+B-)} - \text{P(A-B+)} = 1 - \text{P(B+)} - \text{P(A+)} + \text{P(A+B+)}$$

This is significant because there are many fewer groups than there are populations. In particular, in our data set, at depth 7 there are 16,867 populations which make up 502 groups.

2.2 Aggregating p-values within each group

I tried using Fisher's p-value aggregation method to aggregate the p-values within each group. A positive answer from Fisher's method for a particular group indicates that at least one of the populations within a group is different between the KO and WT samples, without saying which population is different. Fisher's method assumes that the p-values from the different tests are independent, which is not true in this case, so we should expect it the test to be liberal. (Using the observed/expected score removes the dependence between layers of the hierarchy, but does not remove the dependence within a given layer.)

As expected, when we shuffle the KO/WT labels, we find a slight liberal bias. However, there is much more signal in the unshuffled data. If we could find a way of adjusting for the liberal bias due to correlated populations, this could be a good way to boost statistical power.

2.3 Variance within a group as a defining statistic

Another way to use groups is to use the following hypothesis: Because all the proportions in a group are dependent on each other (every group sums to 1), an abnormally large value in one population implies an abnormally low proportion in another population. In other words, the variance in proportions within a group is a summary statistic of the proportions within that group. If some perturbation (such as a gene KO) has an effect on one population in the group, it should influence the group's variance.

To test this, for each sample, I calculated the variance of proportions within each group (generating a 76×502) matrix. For each group, I used a two-sided Wilcox rank-sum test to compare whether the variances in the KO samples were uniformly higher or lower than the variances within the WT samples.

Unfortunately, this hypothesis doesn't seem to hold in the data. We don't see much separation between the real and shuffled p-values according to this test.

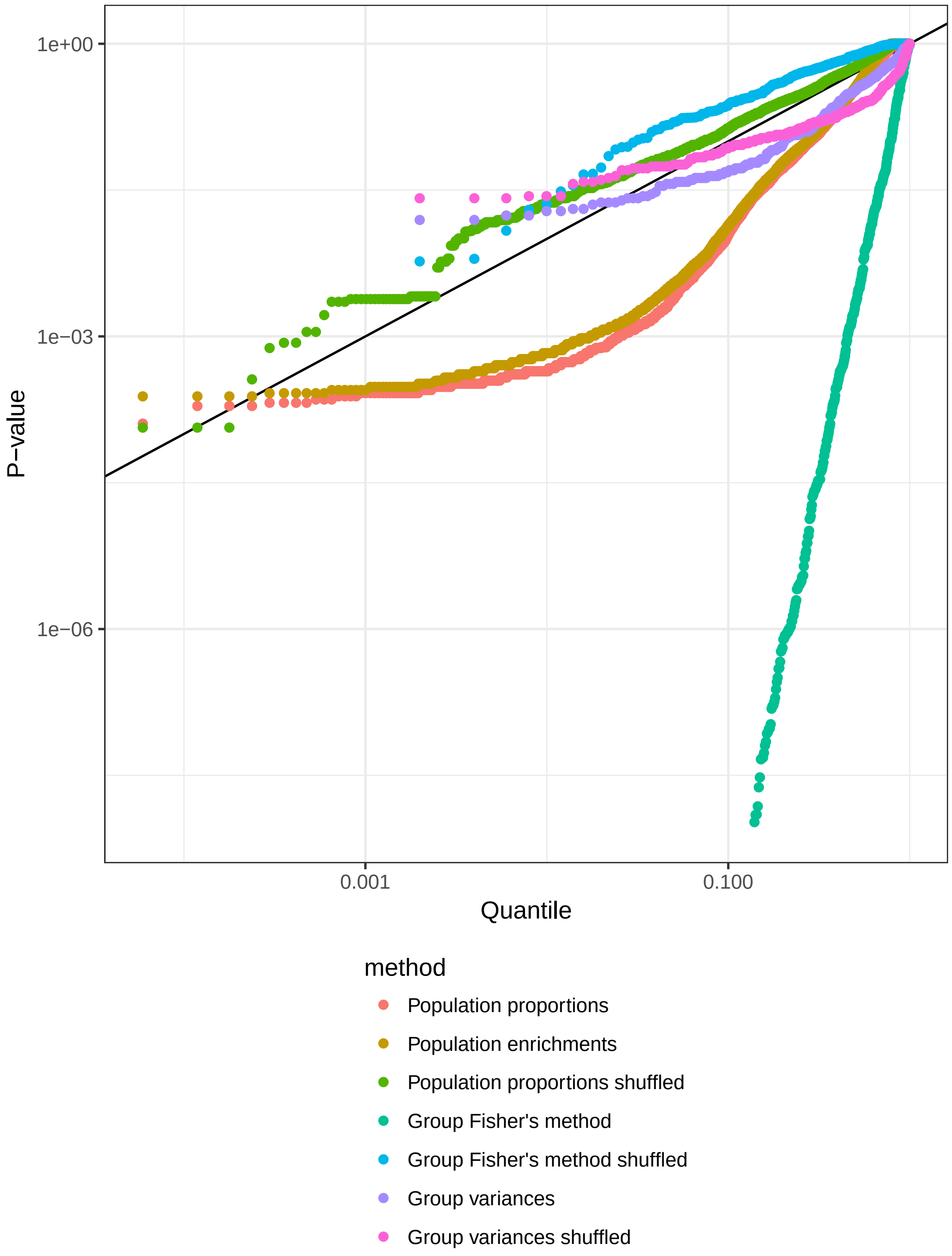

Figure 1: Q-Q plot of p-values generated in different ways. Vertical axis is the uncorrected p-value. Horizontal axis is the p-value quantile—that is, the highest p-value has quantile 1 and the lowest p-value has quantile close to 0. Black line indicates X=Y. Vertical and horizontal axes are shown with a log transformation.

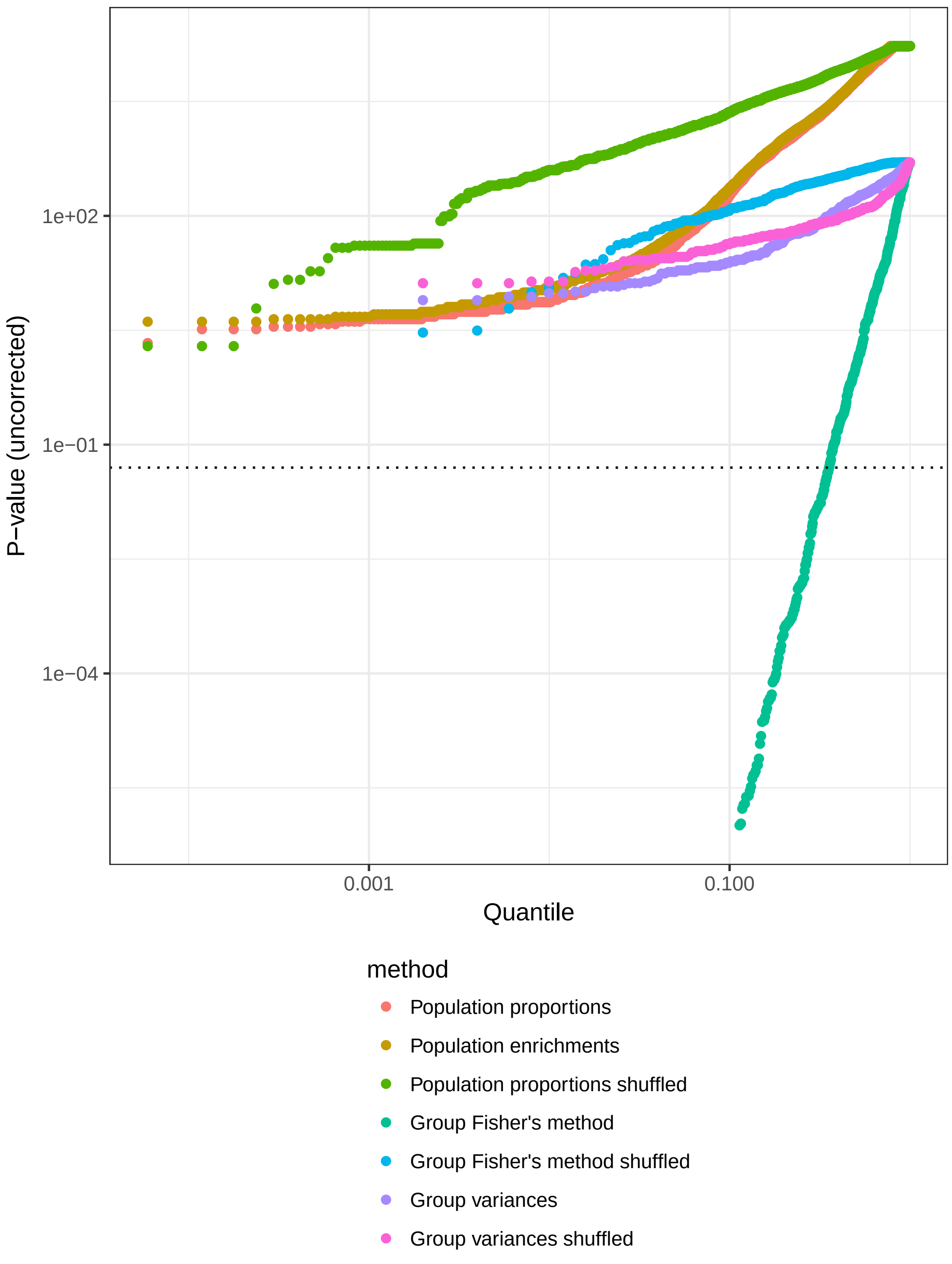

Figure 2: Same as Figure 1, but with Bonferroni corrected p-values. Horizontal dotted line indicates a 0.05 significance threshold.