Multiple Testing Corrections

Contents at a glance

I. Introduction .................................................................2
II. Where to find Multiple Testing Corrections? .........................2
III. Technical details ..........................................................3
    A. Bonferroni correction ..................................................4
    B. Bonferroni Step-down (Holm) correction .........................4
    D. Benjamini & Hochberg False Discovery Rate ....................5
IV. Interpretation ...............................................................6
V. Recommendations ..........................................................7
VI. References .................................................................7
VII. Most commonly asked questions ....................................8
I. Introduction

When testing for a statistically significant difference between means, each gene is tested separately and a p-value is generated for each gene. If the p-value is set to 0.05, there is a 5% error margin for each single gene to pass the test. If 100 genes are tested, 5 genes could be found to be significant by chance, even though they are not. If testing a group of 10,000 genes, 500 might be found to be significant by chance. Therefore, it is important to correct the p-value of each gene when performing a statistical test on a group or genes. Multiple testing correction adjusts the individual p-value for each gene to keep the overall error rate to less than or equal to the user-specified p-cutoff value.

II. Where to find Multiple Testing Corrections?

The Multiple testing corrections can be found in two types of filters:

1) Filter on Confidence, found under the Filtering menu:

The multiple testing correction are available for a filter on the t-test p-value only.
2) The Statistical Analysis tool, found under the **Tools** menu.

The multiple testing corrections are available for all test types.

### III. Technical details

GeneSpring offers four types of multiple testing corrections:

- **A. Bonferroni**
- **B. Bonferroni Step-down (Holm)**
- **C. Westfall and Young Permutation**
- **D. Benjamini and Hochberg False Discovery Rate**

The methods are in order of their stringency, with the Bonferroni being the most stringent and the Benjamini and Hochberg FDR being the least stringent. Each method will be described in more details below. In the examples, a p-value cutoff of 0.05 and a gene list of 1000 genes are assumed.
A. Bonferroni correction
Take the p-value of each gene and multiply it by the number of genes in the gene list. If the corrected p-value is still below the cutoff, the gene will be significant:
Corrected P-value= p-value * n (number of genes in test) <0.05

B. Bonferroni Step-down (Holm) correction
This correction is very similar to the Bonferroni, yet less stringent:
1) Take the p-value of each gene and rank them from the smallest to the largest.
2) Multiply the first p-value by the number of genes present in the gene list: if the end value is less than 0.05, the gene is significant:
   Corrected P-value= p-value * n < 0.05
3) Multiply the second p-value by the number of genes less 1:
   Corrected P-value= p-value * n-1 < 0.05
4) Multiply the third p-value by the number of genes less 2:
   Corrected P-value= p-value * n-2 < 0.05
Follow that sequence until no gene is found to be significant.

Example:

Let n=1000, p-value cutoff=0.05

<table>
<thead>
<tr>
<th>Rank</th>
<th>Gene</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>0.00002</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>0.00004</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>0.00009</td>
</tr>
</tbody>
</table>

Correction:
Gene A p-value= 0.00002 * 1000=0.02 < 0.05 => gene is significant
Gene B p-value=0.00004*999=0.039 < 0.05 => gene is significant
Gene C p-value=0.00009*998=0.0898 > 0.05 => gene is not significant

The exact formula is:
\[ \tilde{p}_{(1)} = \min(np_{(1)}, 1) \]
\[ \tilde{p}_{(i)} = \min(\max(\tilde{p}_{(i-1)}, (n - i + 1)p_{(i)}), 1), i = 2, \ldots n \]

where \( p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(n)} \), \( n \) is the number of genes in the gene list.

C. Westfall and Young Permutation

The Westfall and Young permutation is a step-down procedure similar to the Holm method, combined with a bootstrapping method to compute the p-value distribution:

1) P-values are calculated for each gene based on the original data set and ranked.
2) The permutation method creates a pseudo-data set by dividing the data into artificial treatment and control groups.
3) P-values for all genes are computed on the pseudo-data set.
4) The successive minima of the new p-values are retained and compared to the original ones.
5) This process is repeated a large number of times, and the proportion of resampled data sets where the minimum pseudo-p-value is less than the original p-value is the adjusted p-value.

D. Benjamini & Hochberg False Discovery Rate

This correction is the least stringent of all 4 options. Here is how it works:

1) Rank the p-value of each gene in order from the smallest to the largest.
2) Multiply the largest p-value by the number of genes in test.
3) Take the second largest p-value and multiply it by the total number of genes in gene list divided by its rank. If less than 0.05, it is significant.
   Corrected p-value = \( p\)-value* \( \left(\frac{n}{n-1}\right) < 0.05 \), if so, gene is significant.
4) Take the third p-value and proceed as in step 3:
   Corrected p-value = \( p\)-value* \( \left(\frac{n}{n-2}\right) < 0.05 \), if so, gene is significant.
And so on.

Example:

Let \( n=1000 \), p-value cutoff=0.05
<table>
<thead>
<tr>
<th>Rank</th>
<th>Gene</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>A</td>
<td>0.1</td>
</tr>
<tr>
<td>999</td>
<td>B</td>
<td>0.04</td>
</tr>
<tr>
<td>998</td>
<td>C</td>
<td>0.01</td>
</tr>
</tbody>
</table>

**Correction:**

P-value (A) > 0.05 => gene D is not significant

P-value (B) = 1000/999*0.04 = 0.04004 < 0.05, gene B is significant.

P-value (C) = 1000/998*0.01 = 0.01002 < 0.05, gene C is significant.

The exact formula is:

\[ \tilde{p}(n) = p(n) \]

\[ \tilde{p}(i) = \min(\tilde{p}(i+1), \frac{n}{i} p(i)), i = (n-1), \ldots, 1 \]

where \( p(1) \leq p(2) \leq \ldots \leq p(n) \) is the number of genes in the gene list.

### IV. Interpretation

The number of false-positive genes after multiple testing correction can be interpreted as shown in the table below.

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Type of Error control</th>
<th>Genes identified by chance after correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonferroni</td>
<td>Family-wise error rate</td>
<td>If testing 10,000 genes with p-cutoff equals 0.05, expects 0.05 genes to be significant by chance</td>
</tr>
<tr>
<td>Bonferroni Step-down</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Westfall and Young</td>
<td>False Discovery Rate</td>
<td>If tested 10,000 genes with p-cutoff equals 0.05, genes identified by chance is 5% of genes that are considered statistically significant</td>
</tr>
<tr>
<td>permutation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table represents the multiple testing correction options and the type of error that they control.
V. Recommendations

1. The default multiple testing correction is the Benjamini & Hochberg False Discovery Rate. It is the least stringent of all corrections and provides a good balance between discovery of statistically significant genes and limitation of false positive occurrences.

2. The Bonferroni correction is the most stringent test of all, but offers the most conservative approach to control for false positives.

3. As the multiple testing correction algorithm depends on the number of tests performed, or number of genes tested, it is recommended to select a prefiltered gene list in the Filter on Confidence or the Statistical Analysis tool. Example:

   If there are 10,000 genes in the gene list, assuming a p-value cutoff of 0.05, the individual p-value for a gene must be at least 0.05/10,000=0.000005 to pass a statistical test after a Bonferroni Multiple Testing Correction.

   If there are 4000 genes in the gene list, the minimal individual p-value must be only 0.0000125.

VI. References


VII. Most commonly asked questions

Q. Why do I get more genes with a smaller gene list than with all genes list when I perform a one-way ANOVA using a Multiple Testing Correction?

A. As multiple testing corrections depend on how many genes are tested, the larger the gene list, the more stringent the correction will be. For instance, the Bonferroni correction will multiply the p-values of each gene by the number of tests performed. The more tests (or the more genes, since there is one test per gene), the smaller the p-value must be to pass the restriction.

Q. Why should I use a Multiple Testing Correction? If I select one, no genes pass the restriction.

A. Even though no genes pass the statistical restriction, it is important to keep in mind that the genes that will pass a restriction without multiple testing correction might all be false positives, thus not significant at all. If you have 10,000 genes in your genome, and perform a statistical analysis, a p-value cutoff of 0.05 allows a 5% chance of error. That means that 500 genes out of 10,000 could be found to be significant by chance alone.

Q. What should I do if no genes pass the statistical test when I apply the multiple testing correction?

A. To improve your statistical results, try one or more of the following suggestions:

1) Increase the p-value cutoff.

2) Increase the number of replicates in your experiment.

3) Select a smaller list of genes to use with your analysis. The smaller the list, the less stringent the multiple testing correction will be (see first question).

4) Select to use a less stringent or no multiple testing correction. If you choose to apply no multiple testing correction, rank the genes by their p-values to inspect them manually. Genes with the smallest p-values will be the most reliable.