

Statistical Analysis

- 1-way ANOVA test
- 2-way ANOVA test
- Parametric Test & Nonparametric Test
- Multiple Test Correction

Statistical Analysis (ANOVA)

- Gene Lists
- Experiments
 - Kacker
 - Stover
 - Wadler
 - XueKe
 - yi-chao
 - 8sam
 - Al
 - WYC
 - WYC

Choose Gene List >> all genes (22283)

Choose Experiment >> 8samples 021304, Default Interpretation [Mode: Log of ratio]

Cross-Gene Error Model is: Inactive

1-Way Tests | 2-Way Tests

You can expect a false discovery rate of about 5% of the genes identified.

Parameter to Test: Custom (2 groups) [Select Groups Manually...]

Test Type: Parametric test, assume variances equal

False Discovery Rate: 0.05

Multiple Testing Correction: Benjamini and Hochberg False Discovery Rate

Post Hoc Tests: None

Computation Preferences

Compute locally Compute on a GeNet RemoteServer

Progress: _____

Local run time estimate: Seconds

Start Close Help

One-Way ANOVA

- To filter out genes that don't vary significantly across different groups with multiple samples.
- This comparison is performed for each gene and genes with sufficiently small p-values are returned.
- Comparisons can be performed with parametric or non-parametric methods.

Test Options

Options	Specific test used (analyzing 2 groups)	Specific test used (analyzing more than 2 groups)
Parametric (variances equal)	Student's T-test	ANOVA
Parametric (variances not equal)	Welch t-test	Welch ANOVA
Parametric (use all available error estimate)	Welch t-test using error model variances	Welch ANOVA using error model variances
Nonparametric	Wilcoxon-Mann-Whitney test	Kruskal-Wallis test

Before Performing 1-way ANOVA - A Checklist

- Do you have replicates for the experimental groups that you are about to compare?
- Have you filtered out genes whose measurements are mostly unreliable?
- Have you defined one parameter in the **Experiment Parameters** window indicating which sample belongs to which group?
- If you plan to use a parametric test, have you changed the analysis mode to Log of Ratio in the **Experiment Parameters** window?

Overview of the 1-way ANOVA Tests Window

- **Choose Gene List:** Select the gene list containing the set of genes you would like to analyze. Statistical tests will be performed only on genes in the selected gene list. Again, it is recommended that the **all genes** gene list should not be used. Instead, use a list of genes that has been filtered to remove genes with measurements mostly in the noise range or mostly flagged Absent.
- **Choose Experiment:** Choose the experiment and its proper interpretation to analyze. If you are using parametric tests, then your experiment interpretation should be in log-of-ratio mode.
- **Parameter to Test:** Select the parameter and the underlying groups to compare. In the example shown above, the parameter, 'Drug Agent' was selected to compare the effect of different drug agents on Sprague-Dawley rats. If you would like to compare only selected conditions for this parameter, open the **Select Groups Manually** window, and uncheck the conditions that you would like to ignore. Only groups that are checked will be analyzed.
- **Test Type:** Select the appropriate 1-way ANOVA test type. If you are using a parametric test, make sure your data has been log-transformed (by selecting log-of-ratio mode in experiment interpretation window).
- **False Discovery Rate:** Indicates the overall rate of false positive. The wording for this option, and its final effect on the number of false positives, changes according to the multiple testing correction selected in the option below.
- **Multiple Testing Correction:** This test option is not required for analysis, but it will allow you to keep the overall rate of false positive low.
- **Post Hoc Tests:** This test option is also not required for analysis, but selecting this option will allow you to determine which pair(s) among the groups under study have expression means that are statistically different.

Multiple Testing Corrections (MTC)

- What is MTC?

MTC adjusts individual p value to account for multiple comparisons and keep error rate < user-defined p cut-off value.

- Without MTC

e.g. compare 5000 genes across various parameters. How many genes out of 5000 are differentially expressed between Drug A and B? If test N Genes with a p-value cut-off of p, then expect $\sim N \cdot P$ genes called significant by chance.

$$5000 \text{ genes} \times 0.05 \text{ p} = 250 \text{ genes}$$

- How MTC controls error rate?

How much error is controlled depends on the MTC chosen.

MTC

Types of Error

- Individual (genewise) error rate
None, lots of false positives.
- Family-wise error rate (FWER): very conservative
 - Bonferroni
 - Bonferroni step-down (Holm)
 - Westfall & Young permutation
- False discovery rate (FDR): balanced approach
Benjamini & Hochberg

Example

The recommended correction for multiple testing is Benjamini and Hochberg False Discovery Rate procedure. This procedure is the least stringent of all the methods mentioned above, but it provides a good balance between discovery of statistically significant differences in gene expression and protection against false positives (Type I error).

The stringency of MTC procedures mentioned increases as the number of genes being tested (genes on selected gene list) increases. The following example illustrates this situation:

If:

- number of genes on gene list = 10,000
- p-value cutoff = 0.05
- p-value for Gene A **without** MTC equals 0.000006

If the Bonferroni multiple testing correction was applied to this analysis, then the p-value for Gene A **with** MTC equals 0.06:

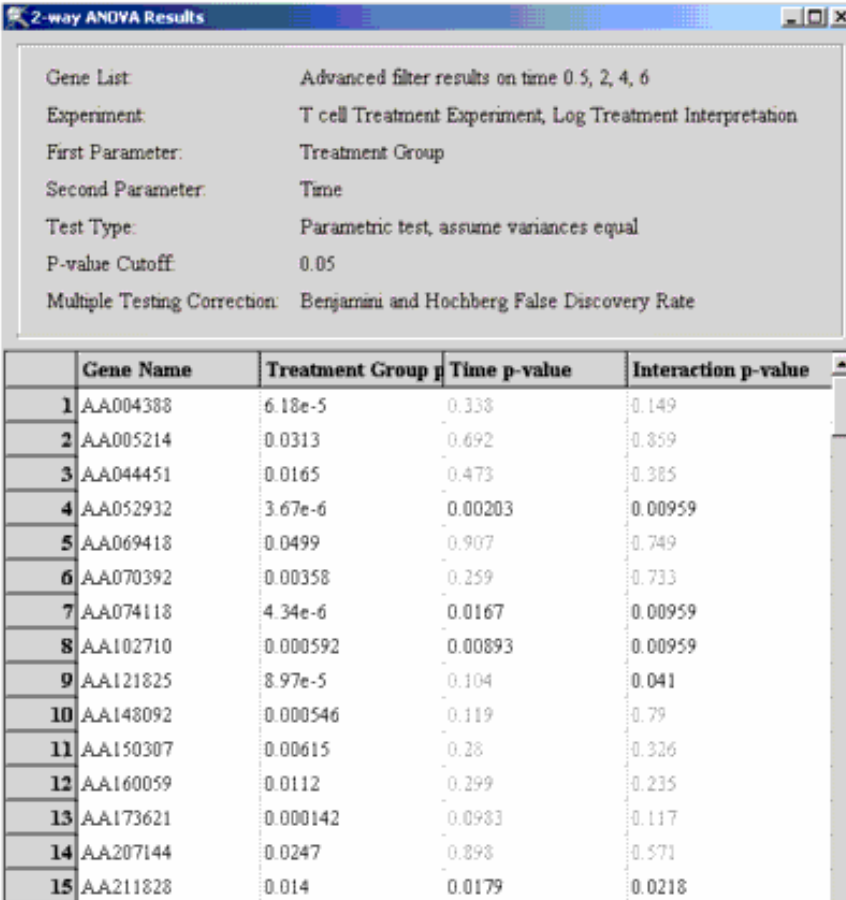
$$P\text{-value with MTC} = 10,000 \times 0.000006$$

It is therefore recommended that you perform statistical analysis on a list of genes that have been filtered for unreliable genes since the multiple testing corrections are directly affected by the number of genes on your gene list.

2-Way ANOVA

For significant differences across groups by 2 parameters.

such as parameters for time point and treatment



2-way ANOVA Results

Gene List: Advanced filter results on time 0.5, 2, 4, 6
Experiment: T cell Treatment Experiment, Log Treatment Interpretation
First Parameter: Treatment Group
Second Parameter: Time
Test Type: Parametric test, assume variances equal
P-value Cutoff: 0.05
Multiple Testing Correction: Benjamini and Hochberg False Discovery Rate

	Gene Name	Treatment Group p	Time p-value	Interaction p-value
1	AA004388	6.18e-5	0.338	0.149
2	AA005214	0.0313	0.692	0.359
3	AA044451	0.0165	0.473	0.385
4	AA052932	3.67e-6	0.00203	0.00959
5	AA069418	0.0499	0.907	0.749
6	AA070392	0.00358	0.259	0.733
7	AA074118	4.34e-6	0.0167	0.00959
8	AA102710	0.000592	0.00893	0.00959
9	AA121825	8.97e-5	0.104	0.041
10	AA148092	0.000546	0.119	0.79
11	AA150307	0.00615	0.28	0.326
12	AA160059	0.0112	0.299	0.235
13	AA173621	0.000142	0.0983	0.117
14	AA207144	0.0247	0.898	0.571
15	AA211828	0.014	0.0179	0.0218