Multiple Testing in Genetic Research

Maia Czerwonka

December 10, 2025

1 Introduction

Multiple testing occurs when researchers have a large number of hypothesis tests that they must run in an experiment. Corrections for multiple testing are important to implement as they ensure more reliable conclusions to be made.

Foundations of hypothesis testing build upon the following key concepts: type 1 error, probability of a type 1 error occuring, and power. A type 1 error occurs when we reject a null hypothesis if it is true. For example, if there is a null hypothesis stating that a gene is not related to hair color, a type 1 error would occur if the gene was truly not related to hair color, but the null is rejected. By definition, valid tests satisfy $P(\text{T1 Error}) \leq \alpha$, meaning that α is the maximum probability one is willing to accept for a type 1 error. The power of a test is the probability of correctly rejecting a false null hypothesis.

Suppose a researcher is completing 12,000 hypothesis tests; one for each gene they are testing to see if it is related to hair color or not.

$$\begin{aligned} H_0^1 &= \text{Gene 1 not related to hair color} & H_1^1 &= \text{Gene 1 is related to hair color} \\ H_0^2 &= \text{Gene 2 not related to hair color} & H_1^2 &= \text{Gene 2 is related to hair color} \\ &\vdots & &\vdots \\ H_0^{12000} &= \text{Gene 12000 not related to hair color} & H_1^{12000} &= \text{Gene 12000 is related to hair color} \end{aligned}$$

Hypothetically, if none of the 12,000 genes are related to hair color, we are almost guaranteed to commit many type 1 errors. The probability of committing a type 1 error for any single hypothesis test is $\leq \alpha$, however, the probability of committing a type 1 error for all 12,000 tests is approximately 100%, because there are so many tests. Therefore, conclusions drawn from the hypothesis tests may be incorrect.

If we present all test results together, we lose the ability to say anything meaningful about any single test independently. We cannot say if we are actually rejecting a null or if we are just rejecting because we ran so many tests. To solve this problem, we use corrections for multiple testing.

Corrections for multiple testing are a group of processes with which the results of many hypothesis tests can be presented in a meaningful way.

2 False Discovery Rate

False Discovery Rate (FDR) is the proportion of false discoveries (type 1 errors) among all discoveries (any rejection of a null). Controlling the FDR at a rate of 0.2 means that on average no more than 20 out of 100 significant results will be false positives.

One way to implement FDR is through the Benjamini-Hochberg procedure, which ensures that the proportion of false positives is less than or equal to a set FDR. This procedure orders the p-values, finds the largest one meeting its FDR-based threshold, and declares all p-values up to that point significant.

3 Breast Cancer Genetic Predictor Study

3.1 Background

One study that uses multiple testing investigates the relationship between genes and survival time for patients with breast cancer. One of the gene datasets that were studied had 12,649 genes. Researchers tested associations between genes and survival time using the Cox proportional hazards regression model. They then controlled FDR using Benjamini-Hochberg at a rate of 0.2.

3.2 Results

Of the 12,649 genes, 3,246 were found to be significant after controlling for FDR. This means that roughly 650 may be false positives. After focusing our attention to all genes that are seemingly related to breast cancer survival time, there is a 20% chance that any given gene is not actually related.

Although 0.2 is a relatively high percentage of false positives, the goal of the study was not to make clinical decisions-it was to identify potential genes that may be linked to breast cancer survival time so that they may be included in follow-up studies.

4 Conclusions

Simultaneously testing tens of thousands of genes increases the risk of false positives. FDR allows us to control the expected proportion of false positives in a way that is less conservative than other methods, and balances discovery with high reliability.

5 References

Miecznikowski, J.C., Wang, D., Liu, S. et al. Comparative survival analysis of breast cancer microarray studies identifies important prognostic genetic pathways. BMC Cancer 10, 573 (2010). https://doi.org/10.1186/1471-2407-10-573