

---

# High-Dimensional Linear Model Inference with Conditional Resampling for Gene Discovery

Daniela Corbetta\*<sup>1</sup>, Pierre Neuvial<sup>2</sup>, Paolo Della Penna<sup>1</sup>, and Livio Finos<sup>1</sup>

<sup>1</sup>Department of Statistical Sciences, University of Padova – Italy

<sup>2</sup>Institut de Mathématiques de Toulouse – Université de Toulouse – France

## Abstract

A central challenge in genomics is to identify the subset of genes that are active in modulating an outcome of interest, such as the insurgence of a disease or a clinical biomarker. The standard approach involves fitting one model for each gene, using the gene as response and the outcome of interest as covariate, and then performing multiplicity correction to control the False Discovery Rate (FDR). Although this strategy is widely adopted, it presents two important limitations. First, it targets marginal rather than conditional associations, potentially overlooking the joint effect of correlated genes. Second, FDR control is restrictive in exploratory analyses, where researchers may wish to adaptively refine the set of reported discoveries based on additional biological knowledge. In such cases, it is more natural to rely on post-hoc inference procedures that provide valid upper bounds on the False Discovery Proportion (FDP) for any selected subset of genes.

To address both challenges, we build upon the flipscore test, a recently proposed nonparametric test for generalized linear models, that approximates the null distribution of the test statistic through random sign-flipping of the individual score contributions. This test is particularly appealing for genomic applications, as it is robust to variance misspecification and directly considers dependencies among test statistics.

However, the classical flipscore test cannot be directly applied in high-dimensional settings where the number of variables exceeds the sample size, as it requires the inversion of a non-full-rank matrix. To overcome this limitation, we propose an extension of the flipscore test that remains valid in high dimensions by introducing a preliminary variable selection step. We establish theoretical results ensuring the validity of the resulting procedure and empirically show its promising performance comparing it with state-of-the-alternatives, including ridge projection and debiased lasso.

Finally, we show that the resampling-based structure of the flipscore test allows integration with post-hoc inference methods that adapt to the dependence structure of test statistics. This results in FDP upper bounds that are typically less conservative than those obtained under standard independence assumptions, offering a powerful and flexible tool for high-dimensional genomic analyses.

**Keywords:** Flipscore test, permutation test, high dimensional inference, post hoc inference

---

\*Speaker