This paper introduces the concept of False Discovery Rate (FDR), which is a method to account for type I errors when conducting multiple comparisons. The author proposes a procedure to control FDR: ranking the p-values of each test, setting different thresholds for these p-values, and deciding which tests meet the threshold for significance. Experimental results show that the FDR controlling methods have larger power in correctly rejecting false hypotheses. This advantage increases when the size of the dataset is larger.

Understanding the control of type I errors is critical for data scientists as they may result in misleading conclusions and negative real-world impact, such as the adoption of ineffective treatment. The FDR methods described in this paper are discussed in our bioinformatics class (BMI776) and are significant statistical methodologies for hypothesis testing, especially in the era of big data.