

## False discovery rate

By construction of hypothesis testing there is always a chance to wrongly reject the null hypothesis. For a single test, this is known as the type I error. When working with Omics data, the datasets are usually huge. That means as many chances to be wrong when hypothesis testing. That's why we introduce the False Discovery Rate (FDR) as the expectation of the false discovery proportion (FDP):

$$FDR = \mathbb{E} \left( \frac{\# \text{ false positives}}{\# \text{ Hypothesis}} \right) = \mathbb{E}(\text{FDP})$$

## Benjamini-Hochberg (BH) method

The BH procedure for controlling the FDR is one of the most widely used procedures for p-value adjustment :

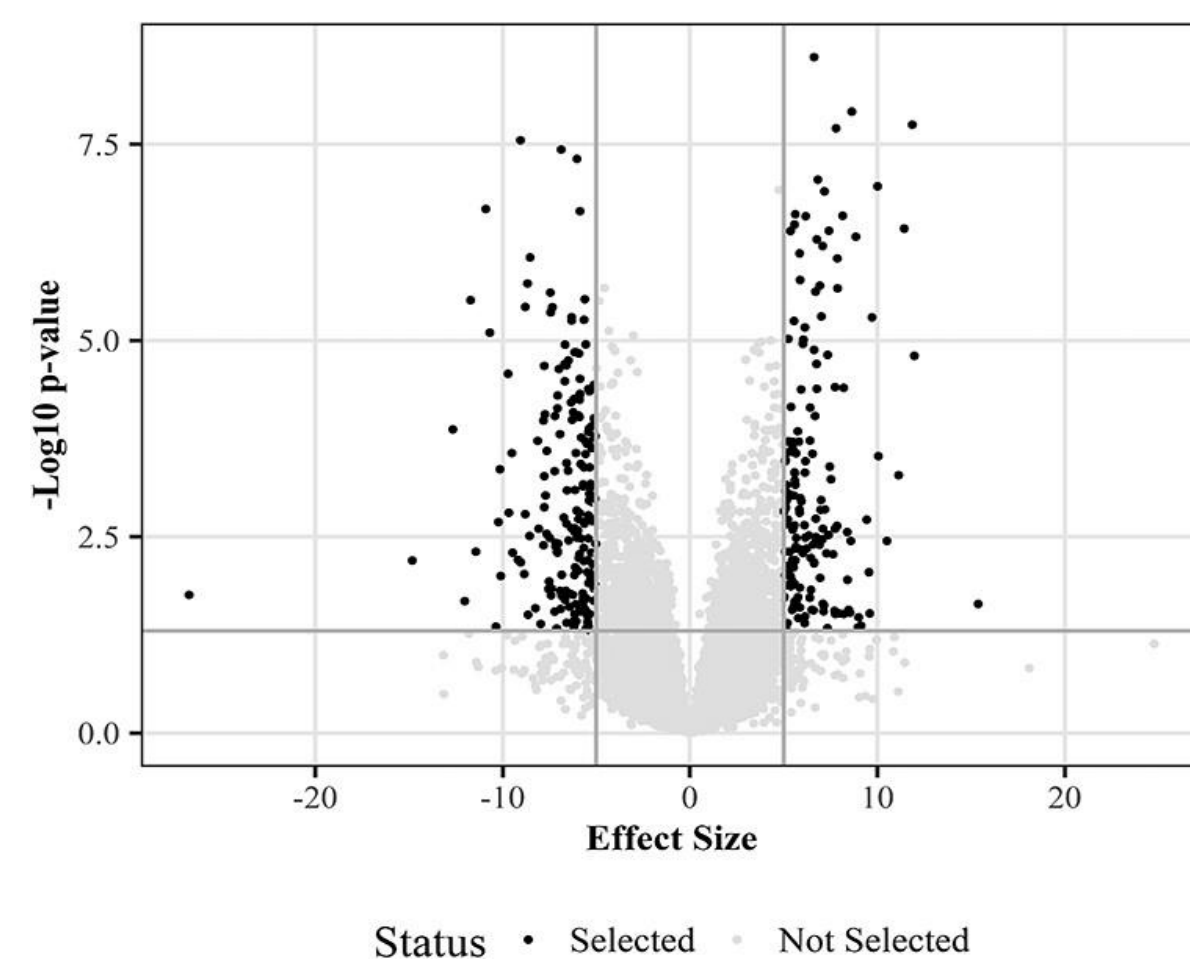
- Sort p-values :  $p_{(1)} \leq \dots \leq p_{(m)}$
- Define  $\hat{I} = \max\{k | p_{(k)} \leq \alpha \frac{k}{m}\}$
- Reject all  $i$  such that  $p_i \leq p_{\hat{I}}$



## Volcano plot

When too many discoveries remain after application of the BH procedure, the double filtering procedure can be applied, which keeps only the p-values that have a large effect size.

FIGURE 1:  
Volcano plot



The Volcano plot is a scatter plot of  $-\log_{10}$  of p-values among significant p-values after BH correction (y-axis) versus effect size (x-axis). The most interesting discoveries are on the right and left corners.

## Limitations & Objective

Nothing guarantees that a FDR-controlled subset is FDR-controlled itself. Intuitively, in the case of Volcano plots, we can see that the denominator of the FDP will be reduced because we discard features with lesser effect size but the numerator (i.e. the number of false positives) may not decrease at the same rate. Situations where it doesn't work include:

- Given the distribution of an estimate :  $\hat{\beta}_i \sim \mathcal{N}(\beta_i, \sigma_i^2)$ , we can see that  $\hat{\beta}_i$  will have a large effect size not only when its true effect  $\beta_i$  is large but also when its variance  $\sigma_i^2$  is large. When false positive features have a larger variance than true positive features, the FDP will inflate.
- There is generally more false positives than true features in a study, so even when the variances distributions are the same the biggest effect feature has a large probability to be a false positive.

Two approaches are proposed to overcome this weakness, the Focused BH and the Closed testing.

## Focused Benjamini-Hochberg (FBH)

The FBH method is a variant of the classical BH procedure that guarantees FDR-control over a subset of discoveries by applying a filter to the corrected p-values. The procedure works as follows:

- Input:** p-values  $\mathbf{p} = (p_1, \dots, p_m)$ , filter  $\mathcal{F}$
- 1 for  $r \in \{0, p_1, \dots, p_m\}$  do
  - 2 | Compute  $\widehat{FDP}(r) = \frac{m \times r}{\#\mathcal{F}(\{j: p_j \leq r\}, \mathbf{p})}$ ;
  - 3 end for
  - 4 Compute  $r^* = \max\{r \in \{0, p_1, \dots, p_m\} : \widehat{FDP}(r) \leq \alpha\}$ ;
  - 5 Compute the classical BH rejection set  $\mathcal{R}^* = \{j: p_j \leq r^*\}$ ;
  - 6 Compute  $\mathcal{U}^* = \mathcal{F}(\mathcal{R}^*, \mathbf{p})$ ;

**Result :** Filtered rejection set  $\mathcal{U}^*$

In the case of Volcano plots, we shall select  $\mathcal{F}$  such that it selects the  $k$  largest  $|\hat{\beta}_i|$  or all the  $|\hat{\beta}_i|$  above a certain threshold  $\tau$ .

## Closed testing

Instead of controlling directly the FDR, we can use a closed testing approach to estimate the FDP with a certain probability :

$$\mathbb{P}(FDP(D) \geq \widehat{FDP}(D) \text{ for all subset } D) \geq 1 - \alpha$$

Closed testing procedure is valid over all subsets of  $D$ . That insures us that the procedure works even with double filtration like Volcano plots. The procedure works as follows:

- Input:** p-values  $\mathbf{p} = (p_1, \dots, p_m)$ ,  $D = \{g: |\hat{\beta}_g| \geq \tau \text{ \& } p_g \leq p\}$  the Volcano plot subset
- 1 Let  $h_\alpha = \max\{i \in \{0, \dots, m\} : i p_{(m-i+j)} > j\alpha, \text{ for } j = 1, \dots, i\}$ ;
  - 2 Let  $\bar{t}(D) = \max_{1 \leq u \leq \#D} 1 - u + \{g \in D : h_\alpha p_g \leq u\alpha\}$ ;
  - 3 Compute  $\text{median}\widehat{FDP}(D) = 1 - \frac{\bar{t}(D)}{\#D}$  with  $\alpha = 0.5$ ;
  - 4 adjust  $\tau$  and  $p$  such that  $\text{median}\widehat{FDP}(D) \leq 0.1$ ;
- Result:**  $D$ , the FDP bounded Volcano plot set

## Results

In the real case study, the database contains 49 individuals and 58037 genes. Authors use a resampling method with 6 individuals for the test sample and 37 individuals for the validation sample. If among the 37 individuals there are significant genes for BH (adjusted P-value for BH  $< 0.05$ ) then they are considered differentially expressed.

- If differentially expressed gene in the validation and test sample  $\rightarrow$  true positive (TP);
- If gene differentially expressed in the validation sample and not differentially expressed in the test sample (or vice versa)  $\rightarrow$  false positive (FP);
- If gene not differentially expressed in validation and test sample  $\rightarrow$  true negative (TN).

17% of the genes selected by the Volcano plot in the test sample were not significant in the validation. Filtering the results may lead to FDP inflation and that a large effect size does not necessarily imply that the result is a true positive.

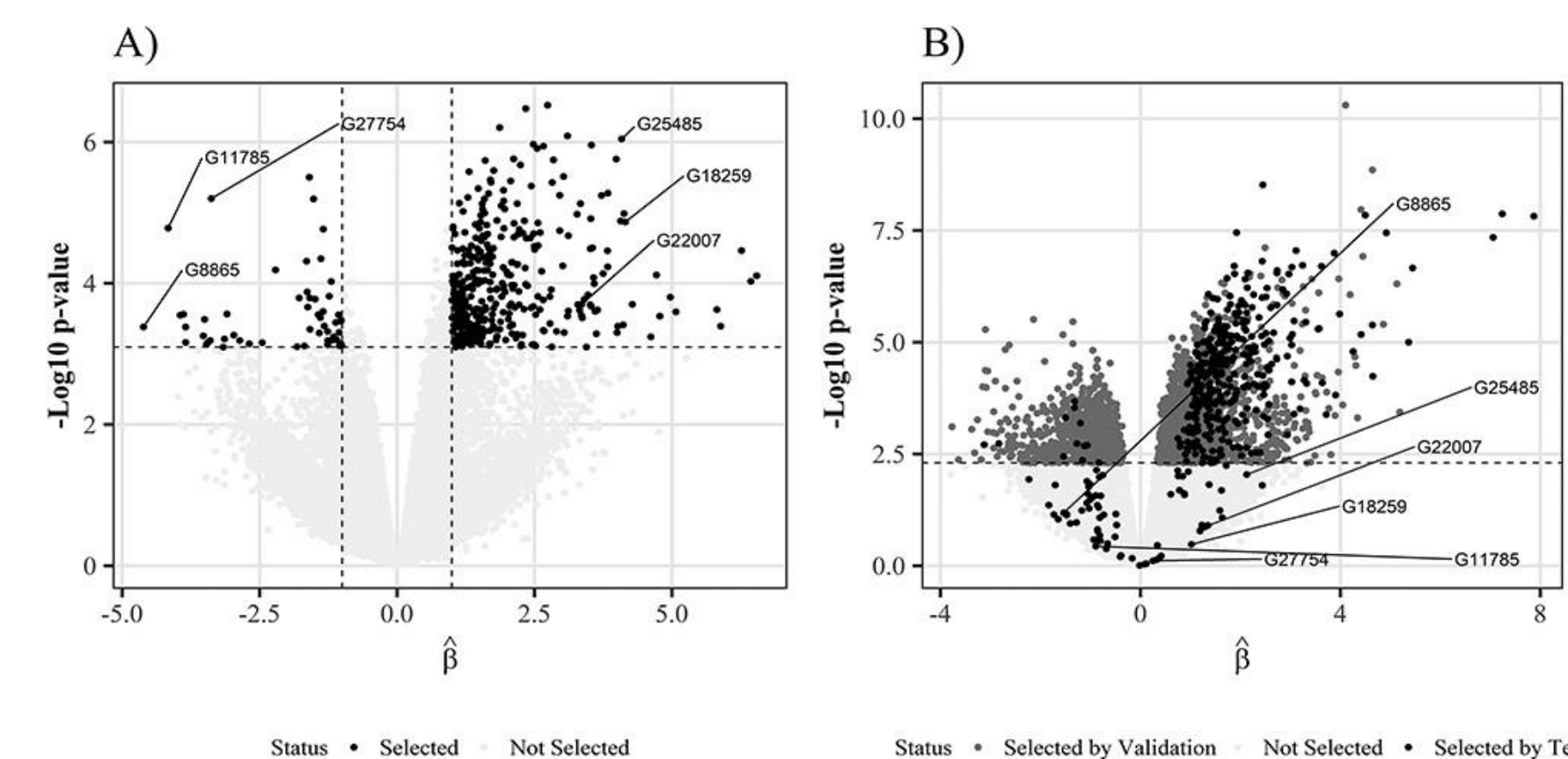
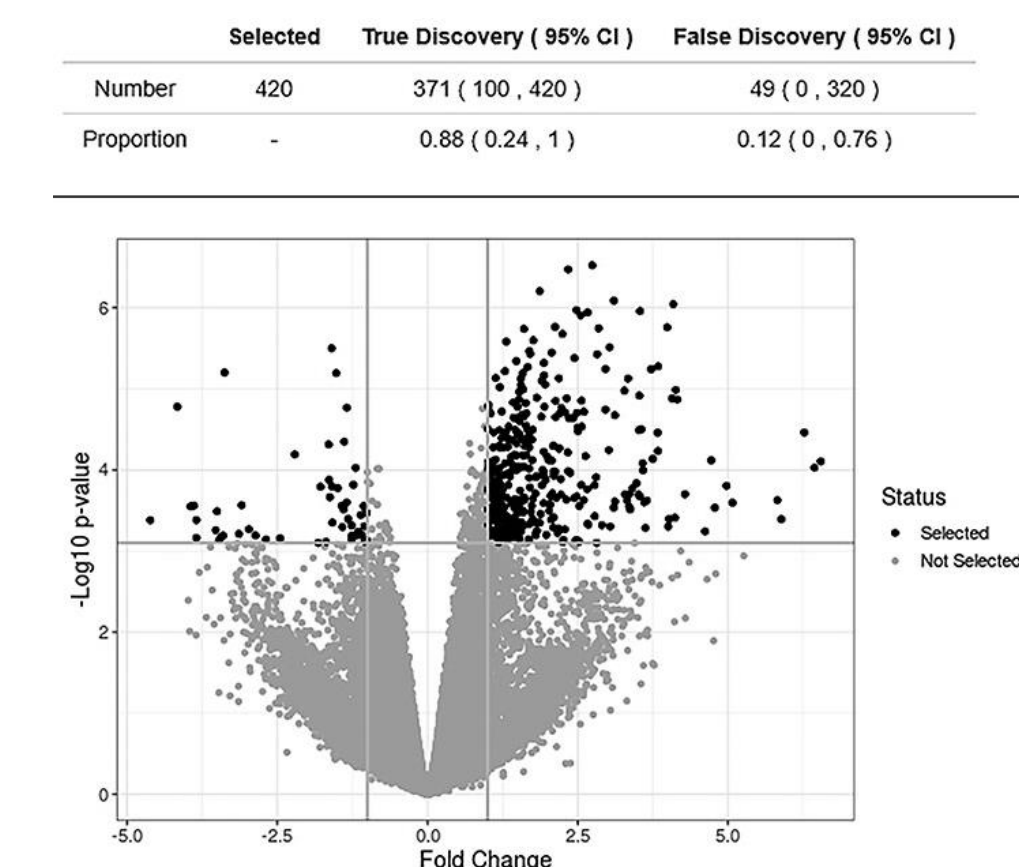


FIGURE 2: Volcano plot of the real case study. P-values are adjusted with BH method.

When we use the Closed testing method, we obtain an estimate medianFDP equals to 12%, that is better than the previous BH method (17%). It is also closer to the expected value because the medianFDP is conservative and it never underestimates the FDP value. Moreover, the confidence interval of the medianFDP includes the true value of FDP because it goes to 0% à 76%.

FIGURE 3: Volcano plot of the real case study with the Closed testing method



## Conclusion

FDR inflation occurs when FDR control procedures such as BH are combined with the Volcano Plots double filter procedure. FDR control on a set of discoveries does not imply FDR control on subsets of these discoveries. FDR inflation is high when the variance of differentially expressed genes is lower than the variance of non-differentially expressed genes. FDR inflation is less important when there are high correlations or a low proportion of null genes. Closed testing and FBH are two alternatives to the BH procedure that allow double screening while controlling the FDR.

## Bibliography

1. Ebrahimipoor (2021) Inflated false discovery rate due to Volcano plots: problem and solutions
2. Katsevich (2020) Filtering the rejection set while preserving false discovery rate control