

Le problème des tests multiples : augmentation du risque...

Les tests sont basés sur des probabilités, il y a donc toujours un risque de tirer des conclusions erronées !

→ **Aucun test d'hypothèse n'est fiable à 100 %**

Effectuer des tests d'hypothèse :

Vous avez deux hypothèses :

- H_0 : Hypothèse nulle = l'hypothèse de référence : pas de différence
- H_1 : Hypothèse alternative : Il y a une différence



Vous rencontrez : Erreur de type I : α = Risque alpha

- $\alpha = 0,05$ C'est la probabilité de rejeter H_0 à tort ! (seuil de significativité)
- En d'autres termes, une chance acceptable de faux positif !!!

Abondance différentielle : Tests multiples !!

1 TEST :

$$P_{\text{False Positive}} = P_{\text{error}} = \underline{\alpha} = 0.05$$

Prob complémentaire

$$P_{\text{no_error}} = 1 - \underline{\alpha} = 0.95$$

2 TESTS sans faire d'erreur :

$$P_{\text{no_error in two tests}} = (1 - \underline{\alpha}) * (1 - \underline{\alpha}) = (1 - \underline{\alpha})^2$$

Prob complémentaire

$$P_{\text{at_least_ONE_error in two tests}} = 1 - (1 - \underline{\alpha})^2$$

Generalisation à n TESTS

$$P_{\text{at_least_ONE_error in n tests}} = 1 - (1 - \underline{\alpha})^n$$

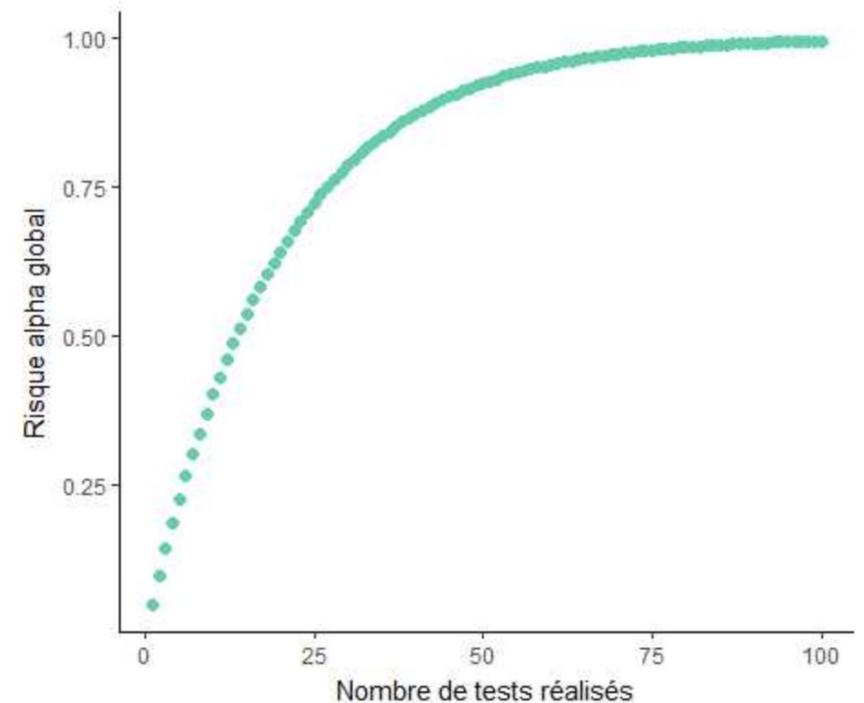
It's called the global $\underline{\alpha}$ risk

What does it means...

- You test **ONE** ASVs (n=1) for differential abundance: $1-(1-\alpha)^n = 1-(1-0.05)^1 = 0.05$
- You test **3** ASVs (n=3): $1-(1-0.05)^3 = 0.14$
- You test **100** ASVs (n=100): $1-(1-0.05)^{100} = 0.9941$

The global risk α reach **0.9941=99.41%!!!!**

→ **99% de rejet à tort H0 au moins une fois**



Il faut ajuster ce phénomène en utilisant la **p.value ajustée** !

FDR : False Discovery Rate : Benjamini-Hochker

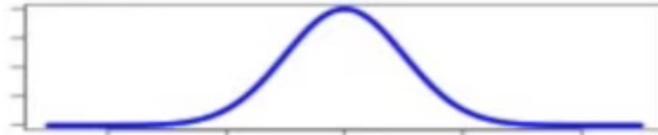
L'idée : Rejeter les mauvaises données qui ont l'air bonnes!!!

Benjamini-hocherker **ajuste la p-values**
Pour limiter le nombre de **faux positifs**
qui sont signalés comme **significatifs** ($pvalue < 0.05$)

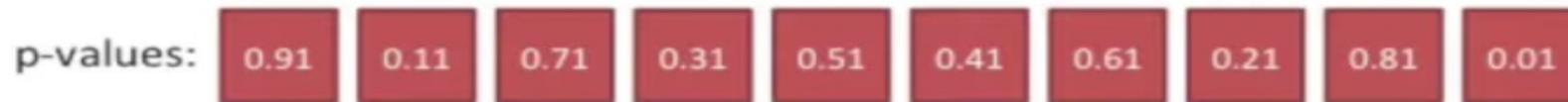
p-values ajustées
= augmente la valeur!

Using FDR cutoff < 0.05
signifie que moins de 5% des résultats significatifs seront des faux positifs

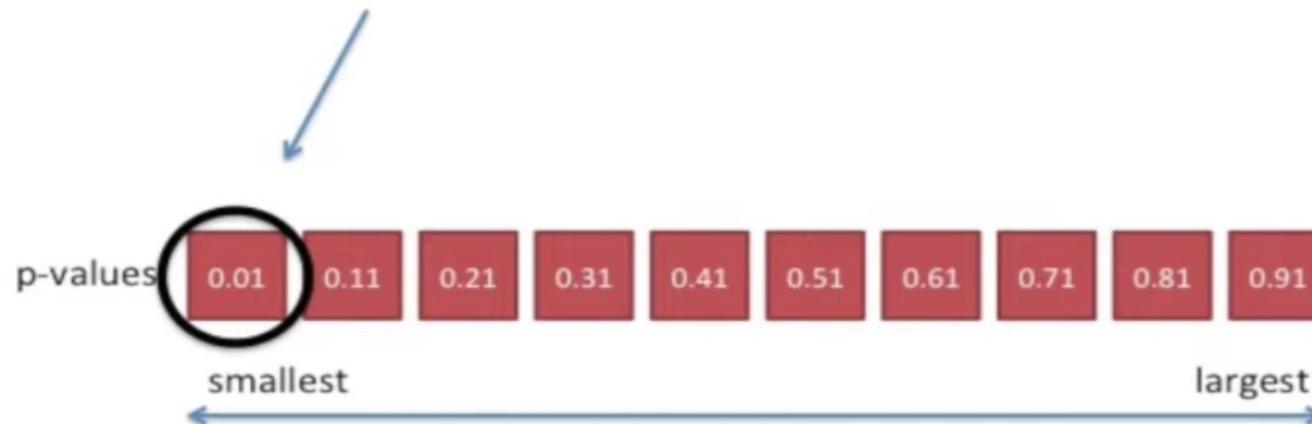
Mathematical approach FDR-Benjamini-Hochker



10 pairs of samples taken from the same distribution. (i.e. 10 genes that were not effected by the drug).

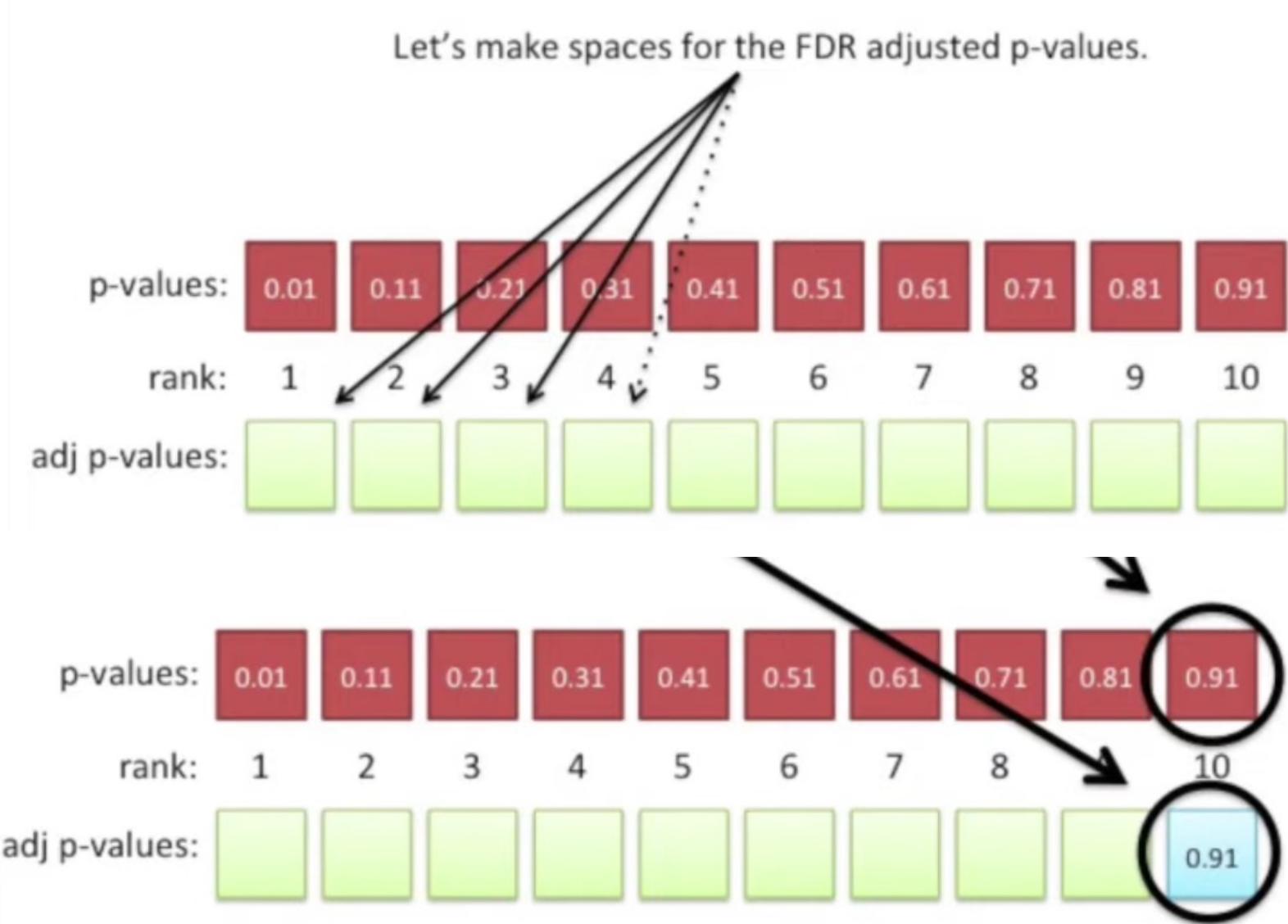


Notice that one of the p-values is a false positive (that is to say, less than 0.05)



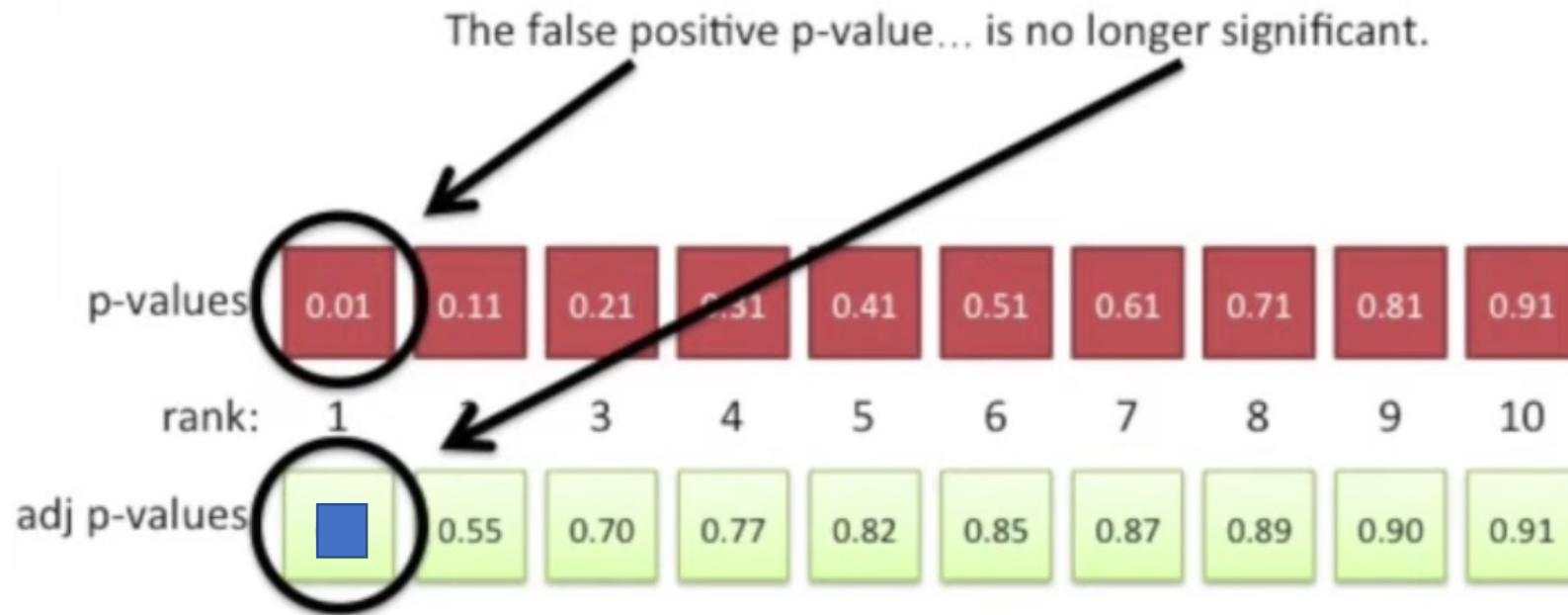
1- Ranking pvalue

Prepare space for adjusted p-value

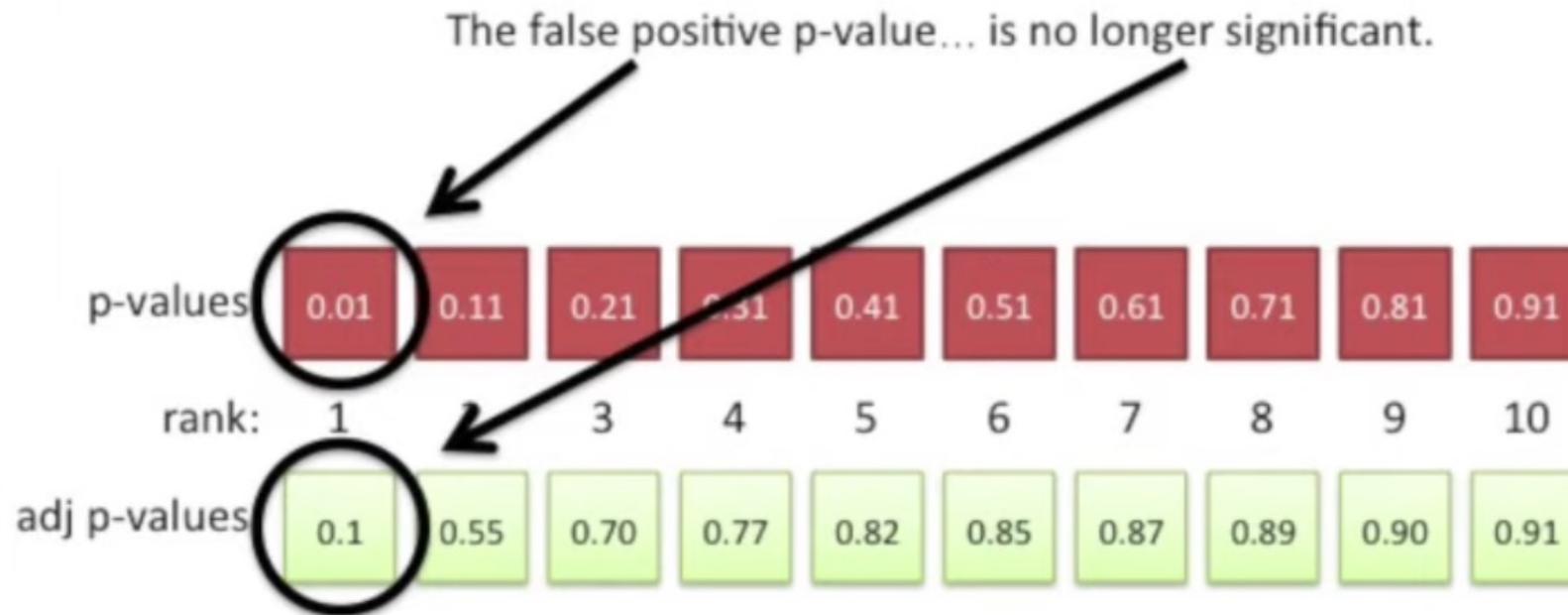


2- Largest adjusted pvalue and larger pvalue are same

Finalemment...



Finalemment...



Le faux positif ne l'est plus grâce à cette méthode d'ajustement !!