

Les défauts du calcul du FDR (False Discovery Rate)

Thomas Burger

iRTSV (FR3425) / BGE (U1038), CNRS/CEA/UJF/INSERM, Grenoble, France

L'analyse haut débit de mélanges biologiques complexes par des dispositifs de type LC-MS/MS permet d'identifier un nombre toujours plus important de protéines. Néanmoins, certaines de ces identifications peuvent être erronées (ce sont des faux positifs) ; afin d'en limiter le nombre dans la publication de listes de protéines toujours plus longues, il est désormais demandé de contrôler la proportion estimée de ces faux positifs, via le calcul d'un FDR (False Discovery Rate).

La présentation commencera par un bref rappel des notions de test d'hypothèse simple, de test d'hypothèses multiples, de proportions de faux positifs et de fausses découvertes, de FPR (False Positive Rate) puis, enfin de FDR (False Discovery Rate) ; Nous expliquerons pourquoi le FDR est important en protéomique et nous résumerons rapidement les différentes utilisations de base Decoy pour le calcul du FDR

Ensuite, nous détaillerons un certain nombre de limitations associées au calcul du FDR (flou de la définition, biais de l'estimation, etc.) ; Nous montrerons que son intérêt réel est largement surestimé, et que dans certains cas, cet intérêt est même discutable... Enfin, nous ouvrirons une discussion sur les possibles méthodes alternatives permettant de contrôler la qualité des listes de protéines sans passer par le calcul d'un FDR.