



Université Lille 2
Droit et Santé



metaRNASeq: un package pour la méta-analyse de données RNA-seq

Guillemette Marot, Florence Jaffrézic, Andrea Rau

28/06/13

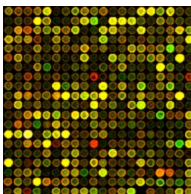
Overview

- 1 Introduction
- 2 Analyse statistique d'une seule étude
- 3 Méta-analyse
- 4 Utilisation du package
- 5 Conclusion

Introduction

Technologies qui mesurent l'expression de **milliers de gènes** simultanément.

Puces à ADN



source : bio.cse.ohio-state.edu/MicroarrayDesigner/

Séquençage à haut débit



source : Haas and Zody, Nature Biotechnology (2010)

Application

Recherche de gènes **différentiellement exprimés** entre deux conditions (malade/sain par exemple).

Analyse statistique d'une seule étude

Recherche de gènes différentiellement exprimés

- Normalisation
- Utilisation d'une statistique de test adaptée pour l'analyse d'expression différentielle.
- Correction pour les tests multiples

	Puces à ADN	RNA-Seq
Information	Mesures d'intensités lumineuses	Comptages de lectures
Modélisation	Loi normale	Loi de Poisson, loi binomiale négative
Tests	T-tests modérés	Tests de rapports de vraisemblances

Adaptation des statistiques de test usuelles pour trouver un compromis entre une analyse "gène-à-gène" et une analyse "paramètre commun"

Méta-analyse

Contexte particulier :

- Plusieurs études disponibles mais **comparaison directe impossible**
- Peu d'individus dans chaque étude individuelle, beaucoup de gènes ou transcrits

Objectifs de la méta-analyse

- Accroître la sensibilité, la proportion de gènes déclarés à raison différentiellement exprimés
- Eliminer des faux positifs, gènes déclarés à tort différentiellement exprimés

Méta-analyse de données RNA-seq

Analyse statistique d'une étude

Package **DESeq** (Anders and Huber, 2010, Genome Biology)

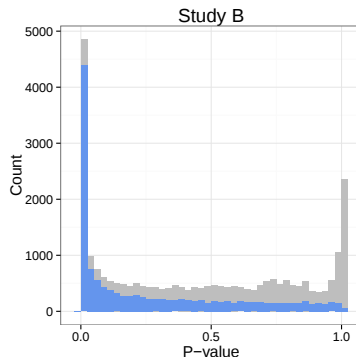
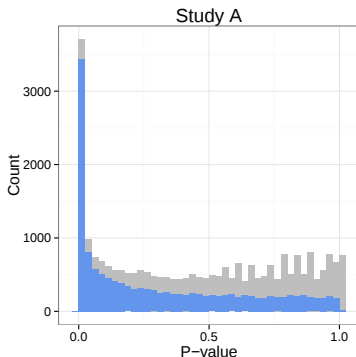
⇒ utilisation de distributions **binomiales négatives** et rétrécissement des paramètres de dispersion.

Méta-analyse

- Inclure l'effet étude dans DESeq
 - Combiner les résultats (p-values) des analyses individuelles
- Avantage : possibilité d'utiliser une autre statistique de test que celle de DESeq
- Hypothèse : la distribution des p-values doit être uniforme sous l'hypothèse nulle

Méta-analyse de données RNA-seq

Recommandation : utiliser HTSFilter (Rau, 2013) pour filtrer les gènes avec de faibles comptages.



Exemple : Histogrammes de p-values brutes de deux études comparant l'expression des gènes dans des lignées cellulaires du mélanome qui expriment ou pas MiTF.

Méta-analyse de données RNA-seq

Packages pour la méta-analyse de données transcriptomiques :

- metaMA (2009) : puces à ADN
- metaRNASeq (2013) : séquençage à haut débit

Différences de metaRNASeq par rapport à metaMA

- Combinaison des tailles d'effet absente à cause du problème de surdispersion qui rend difficile leur calcul
- Séparation des gènes sur- ou sous-exprimés compliquée dans les techniques de combinaison de p-values

Retour sur la méta-analyse de données de puces à ADN

Combinaison inverse normale

$$N_g = \sum_{s=1}^S w_s \Phi^{-1}(1 - p_{gs})$$

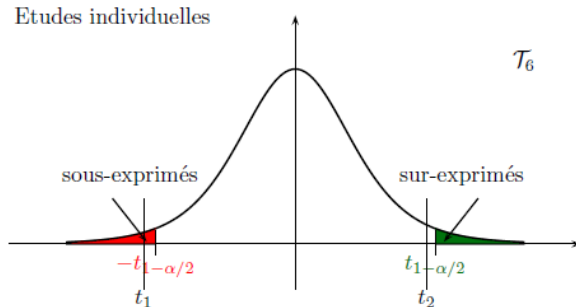
$$w_s = \sqrt{\frac{\sum_c R_{cs}}{\sum_\ell \sum_c R_{c\ell}}}$$

$$N_g \sim \mathcal{N}(0, 1)$$

Pour les puces à ADN, il était conseillé d'utiliser des p-values unilatérales pour éviter les conflits.

Qu'en est-il pour le séquençage à haut débit ?

Retour sur la méta-analyse de données de puces à ADN



Gène pour lequel on observe un conflit d'expression

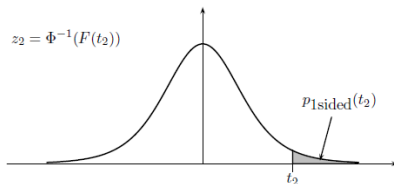
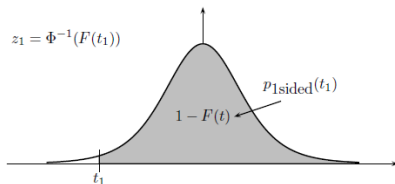
$t_1 < 0$ valeur observée de la statistique de test pour l'étude 1

$t_2 > 0$ valeur observée de la statistique de test pour l'étude 2

Retour sur la méta-analyse de données de puces à ADN

scores à partir de p-values unilatérales

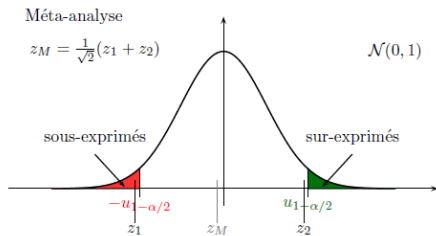
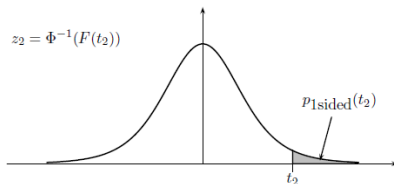
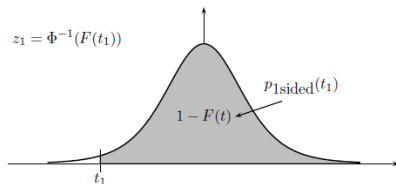
$$z = \Phi^{-1}(1 - p_{1\text{sided}}) = \Phi^{-1}(F(t))$$



Retour sur la méta-analyse de données de puces à ADN

scores à partir de p-values unilatérales

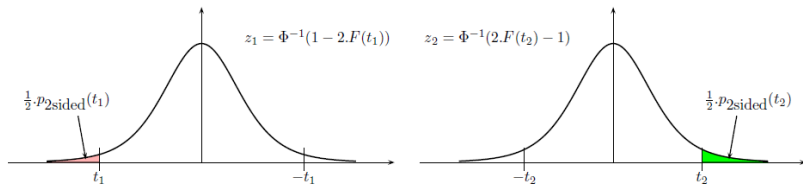
$$z = \Phi^{-1}(1 - p_{1\text{sided}}) = \Phi^{-1}(F(t))$$



Retour sur la méta-analyse de données de puces à ADN

scores à partir de p-values bilatérales

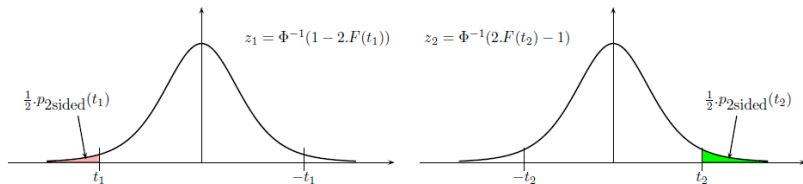
$$z = \Phi^{-1}(1 - p_{2\text{sided}}) = \Phi^{-1}(1 - 2.(1 - F(|t|)))$$



Retour sur la méta-analyse de données de puces à ADN

scores à partir de p-values bilatérales

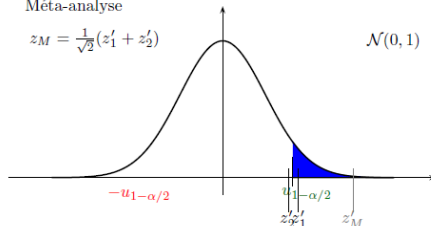
$$z = \Phi^{-1}(1 - p_{2\text{sided}}) = \Phi^{-1}(1 - 2.(1 - F(|t|)))$$



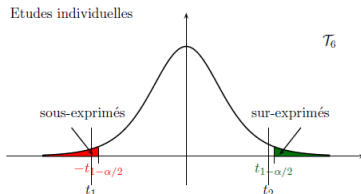
Méta-analyse

$$z_M = \frac{1}{\sqrt{2}}(z'_1 + z'_2)$$

$\mathcal{N}(0, 1)$



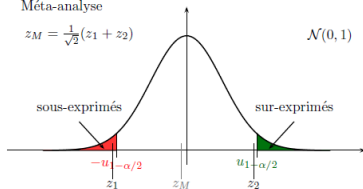
Retour sur la méta-analyse de données de puces à ADN



Construction des scores individuels à partir des p-values

Méta-analyse

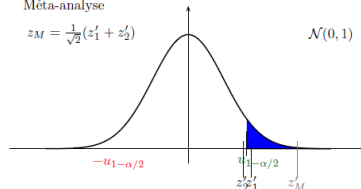
$$z_M = \frac{1}{\sqrt{2}}(z_1 + z_2)$$



unilatérales
puis test bilatéral

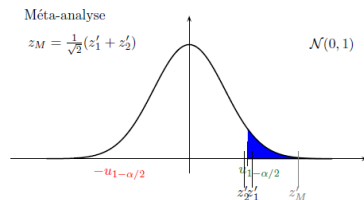
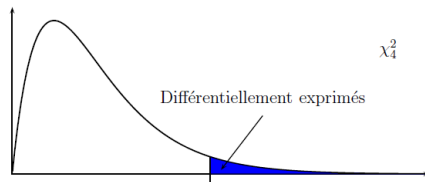
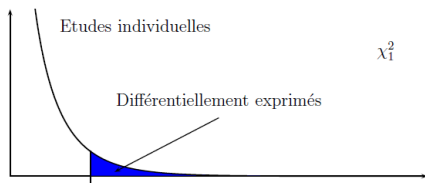
Méta-analyse

$$z_M = \frac{1}{\sqrt{2}}(z'_1 + z'_2)$$



bilatérales
puis test unilatéral

Méta-analyse de données RNA-seq



Méta-analyse de données RNA-seq

Combinaison inverse normale (Liptak, 1958)

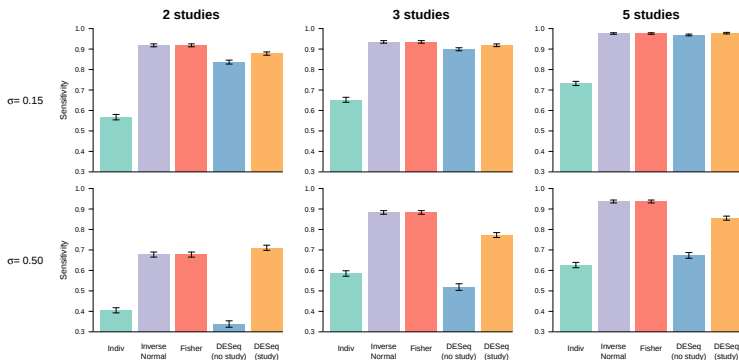
$$N_g = \sum_{s=1}^S w_s \Phi^{-1}(1 - p_{gs})$$
$$N_g \sim \mathcal{N}(0, 1)$$

Méthode de Fisher (1932)

$$F_g = -2 \sum_{s=1}^S \ln(p_{gs})$$
$$F_g \sim \chi_{2S}^2$$

Méta-analyse

Sensibilité : $E\left(\frac{VP}{VP+FN}\right)$



En pratique...

Preprint disponible sur HAL :

(Rau et al., 2013) Differential meta-analysis of RNA-seq data from multiple studies.

```
library(metaRNASeq)  
vignette("metaRNASeq")
```

- simulation de données avec `sim.function`
- analyses individuelles avec `HTSFilter` et `DESeq`
- combinaison de p-values avec `fishercomb` et `invnorm`
- gestion des conflits par extraction des log-ratios

Conclusions

Conclusions

- Méta-analyse à partir des p-values plus difficile dans les données de RNA-seq que de puces à ADN
- Performances identiques pour les deux méthodes de combinaisons de p-values
- Faible variabilité inter-études $\Rightarrow$ les résultats sont très proches d'une analyse globale GLM incluant un effet étude
- Augmentation de la variabilité inter-études $\Rightarrow$ gains en terme de sensibilité, AUC (particulièrement en présence de plus de 3 études)
- Package metaRNASeq et sa vignette déjà disponibles sur la forge R, bientôt sur le CRAN.

Merci de votre attention !