

# Analyses de données de transcriptomique

**Cyprien Guérin**

CATI Sysmics

MaIAGE, INRA, Université Paris-Saclay, Jouy-en-Josas, France

Technologies d'acquisition du signal transcriptomique :

- Puces moyenne densité (*60k*) ;
- Puces tiling haute densité (*tiling array*) ;
- Séquençage haut-début de transcripts (*RNA-Seq*).

Types d'applications :

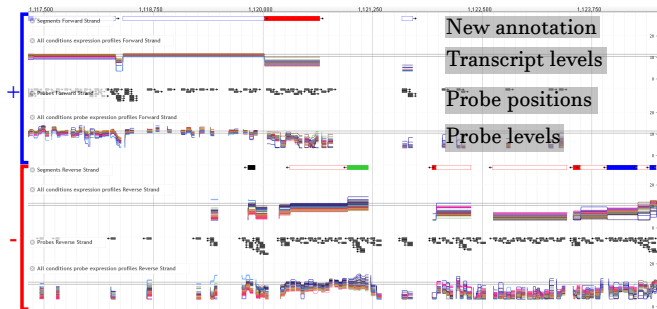
- Analyses d'expression différentielles (avec *a priori* d'annotation) ;
- Découverte du paysage transcriptionnel (sans *a priori* d'annotation) ;
- Analyse à façon des profils de transcription le long des génomes.

Organismes étudiés :

- Puces 60k : *F. psychrophilum* ;
- Puces tiling : *B. subtilis* et *A. aureus* ;
- RNA-Seq : les mêmes et bien d'autres.

Méthodologie d'analyse des données :

- Normalisation par *spike-in* ;
- Aggrégation des niveaux d'expression des sondes par régions annotées.



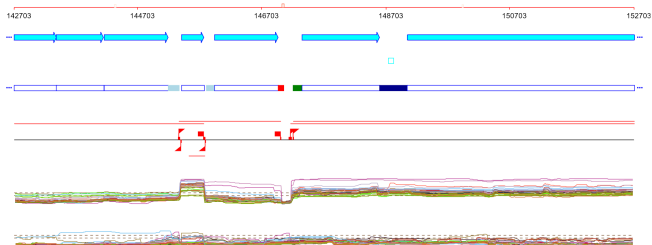
Visualisation type sous Jbrowse

Types d'applications :

- Analyses d'expression différentielle simple (*limma*, *Bioconductor*) ;
- Analyse à façon des départs de transcriptions.

Méthodologie d'analyse des données :

- Normalisation par *quantiles* ou *spike-in* ;
- Aggrégation des niveaux d'expression des sondes par régions annotées.



Visualisation type sous AEB

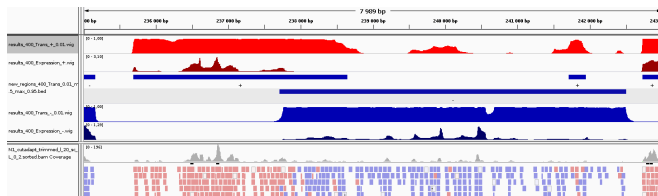
Types d'applications :

- Analyses d'expression différentielle simple (modèles linéaires avec contrôle du *FDR*) ;
- Enrichissement de l'annotation structurale.

# Séquençage haut débit de transcripts

Méthodologie d'analyse des données :

- Aggrégation des comptages par régions anotées ;
- Normalisation par taille de librairie de séquençage ;
- Calculs de *FPKM* (ou équivalent *FPKM* local) pour la visualisation.



Visualisation type sous IGV

Types d'applications :

- Analyses d'expression différentielle simple (*EdgeR* ou *DESeq2* sur comptages bruts avec contrôle du *FDR*) ;
- Enrichissement de l'annotation structurale par aggrégation de différents échantillons avant séquençage (*Parseq*, Mirauta et al., 2014).