ID de Contribution: **15** Type: **Non spécifié**

Clustering of co-expressed genes

vendredi 22 juin 2018 11:30 (1 heure)

Complex studies of transcriptome dynamics are now routinely carried out using RNA sequencing (RNA-seq). A common goal in such studies is to identify groups of co-expressed genes that share similar expression profiles across several treatment conditions, time points, or tissues. These co-expression analyses can in fact serve a double purpose: (1) as an exploratory tool to visualize cluster-specific profile trajectories; and (2) as a hypothesis-generating tool for poorly annotated genes, as co-expression clusters may correspond to genes involved in similar biological processes or that are candidates for co-regulation.

Although a large number of clustering algorithms have been proposed in the past to identify groups of co-expressed genes from microarray data, the question of if and how such methods may be applied to RNA-seq data has only recently been addressed. During the MixStatSeq project, we have proposed several methods to solve this gene clustering problem. After a first procedure based on a Poisson mixture model (Rau et al, 2015) on the raw counts of sequenced reads for each gene, the problem was reformulated as the clustering of normalized expression profiles, which represent compositional data. Data transformations in conjunction with Gaussian mixture models were considered as an effective strategy to identify RNA-seq co-expression clusters in Rau and Maugis-Rabusseau (2017). Some related strategies were investigated in Godichon-Baggioni et al. (2018) using K-means. All of these procedures are implemented in the R/Bioconductor package coseq.

Auteur principal: Dr MAUGIS-RABUSSEAU, Cathy (IMT / INSA Toulouse)

Co-auteurs: RAU, Andréa (INRA Jouy-en-Josas); GODICHON-BAGGIONI, Antoine (INSA Rouen)

Orateur: Dr MAUGIS-RABUSSEAU, Cathy (IMT / INSA Toulouse)

Classification de Session: Biological and medical applications