Investigating the transcriptomic repertoire based on High Throughput Sequencing data

Eric Rivals

Investigating the transcriptomic repertoire based on High Throughput Sequencing data.

Eric Rivals - CNRS, LIRMM, Montpellier, France

Keynote speech - Colloque 2009 du GdR Bioinformatique Moléculaire

Ultra-high throughput sequencing (HTS) is used to analyse the transcriptome or interactome at unprecedented depth on a genome-wide scale. These techniques yield short sequence reads that are then mapped on a genome sequence to predict putatively transcribed or protein-interacting regions. We argue that factors such as background distribution, sequence errors, and read length impact on the prediction capacity of sequence census experiments. Here we suggest a computational approach to measure these factors and analyse their influence on both transcriptomic and epigenomic assays. We developped and tuned a bioinformatic pipeline to assess the expression level of known mRNAs and predict novel splicing variants based on the transcript signatures (reads) obtained by Digital Gene Expression (DGE). However, almost 30% of the signatures map to non coding regions, suggesting the existence of unknown transcripts. To cross validate in silico those novel RNAs, we take advantage of RNA-seq, as well as other publicly available DGE data, and visualise all data in the genomic context.

Related publications :

1. Using reads to annotate the genome : influence of length, background distribution, and sequence errors on prediction capacity N. Philippe*, A. Boureux*, L. Bréhèlin, J. Tarhio, T. Commes, E. Rivals Nucleic Acids Research (NAR) doi:10.1093/nar/gkp492 ; 2009.
2. MPSCAN : fast localisation of multiple reads in genomes E. Rivals, L. Salmela, P. Kiiskinen, P. Kalsi, J. Tarhio Proc. 9th Workshop on Algorithms in Bioinformatics Lecture Notes in BioInformatics (LNBI), Springer-Verlag, Vol. 5724, p. 246-260, 2009.