



AlgoB

the seminar you can't miss

<http://algeb.univ-mlv.fr/seminar>

RNF: a general framework to evaluate NGS read mappers

Karel Břinda

LIGM Paris-Est

Tuesday, May 12, 2015 at 16:00 in 4B084

Laboratoire d'Informatique Gaspard-Monge

Université Paris-Est Marne-la-Vallée

5 bd Descartes, Champs-sur-Marne

Abstract: Read simulators combined with alignment evaluation tools provide the most straightforward way to evaluate and compare mappers. Simulation of reads is accompanied by information about their positions in the source genome. This information is then used to evaluate alignments produced by the mapper. Finally, reports containing statistics of successful read alignments are created. In default of standards for encoding read origins, every evaluation tool has to be made explicitly compatible with the simulator used to generate reads.

In order to solve this obstacle, we have created a generic format RNF (Read Naming Format) for assigning read names with encoded information about original positions. Furthermore, we have developed an associated software package RNFtools containing two principal components. MISH-mash applies one of popular read simulating tools (among DWGSIM, ART, MASON, CURESIM etc.) and transforms the generated reads into RNF format. LAVender evaluates then a given read mapper using simulated reads in RNF format. A special attention is payed to mapping qualities that serve for parametrization of ROC curves, and to evaluation of the effect of read sample contamination.