

Bologna International Master in Bioinformatics (September 22-26, 2014)

Special Advanced Course in NGS data analysis:

Nucleotide Sequences Homology Modeling in the Next Gen Sequencing Era

Lecturer: Cédric Notredame

Centro de Regulacio Genomica - Barcelona

In this course, I will start by introducing mainstream next generation sequencing methods. I will then discuss how these methods can be used to sequence, or re-sequence large genomes. I will then introduce RNA-Seq and the systematic sequencing of cell transcriptomes, along with the many challenges it entails, such as gene modeling and isoform quantification. I will then introduce the notion of multiple genome alignment and review existing methods, including some developed in our group. The course will then be focused on Long Non Coding RNAs. I will introduce the latest ENCODE results on this new class of transcripts and present the challenges of homology based annotation for Long Non Coding RNA. Some methods available for this task will be described, including the pipeline we developed for the ENCODE companion paper on Long Non Coding RNAs. We will see how this pipeline, and similar tools, can be deployed to produce homology-based annotation of newly sequenced genomes. The last part of the course will be dedicated to the computation of RNA multiple sequence alignment methods. I will first introduce the standard multiple sequence alignment method and show how these methods can be applied to nucleotide sequences. I will finish this series of lectures by showing how multiple sequence alignment methods can be adapted to deal with RNA secondary structure analysis, and how they can possibly be used to combine structural and sequence information.