[Talk / CBRC2012:The Impact of Next Generation Sequencers]

A graphical exon model for splice site sensitive RNA-seq mapping with integration of variable transcriptional evidences

Toutai Mitsuyama
RNA Informatics Team, Leader, CBRC, AIST

Splice junction sensitive mapping problem has been one of the most crucial problems for expression data analysis. Recent proliferation of next-generation sequencing technology drives demands of precise genome mapping. Although this problem has been a major problem in the field of bioinformatics since a query sequence was an expressed sequence tag (EST), it requires different approach to solve the same problem with shorter and far greater number of query sequences.
The existing sophisticated methods: Tophat (http://tophat.cbcb.umd.edu/) and Scripture (http://www.broadinstitute.org/software/scripture/) take full advantage of large-scale sequencing mapping. They identify the most probable splice junctions by sorting out most frequently supported junctions in a large set of queries. They can also take into account previously identified or human-curated splice junctions from supplementary files.
Their approach is sound and valid for identifying novel splice junctions where there is less existing junction data available. However, these days, there are a large number of sequencing data acquired in a database which can be utilized for identification of splice junctions. Not only splice junctions but also single nucleotide polymorphisms (SNPs), DNA methylation, RNA editing, etc., which should be taken into account for more sensitive and precise genome mapping.
Our original method is able to integrate multiple types of data associated to a genomic sequence named above to realize more sensitive and precise genome mapping of RNA-seq data than the existing methods. This poster will present the basic idea of our method and early benchmark results that demonstrate the performance of our method.