Detection of Splice Junctions from Mid-to-Long RNA-seq Reads by Spaln

Chao Zeng[1], Hiroaki Iwata[2], Natsuhiro Ichinose[1], Tetsushi Yada[1], Osamu Gotoh[3]

[1] Laboratory of Computational Biology, Department of Intelligence Science and Technology, School of Informatics, Kyoto University, Yoshida-Honmachi 36-1, Sakyo-ku, Kyoto 606-8501, Japan
[2] Division of Bioinformatics, Medical Institute of Bioregulation, Kyushu University 3-1-1 Maidashi Higashi-ku, Fukuoka 812-8582, Japan
[3] Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), AIST Tokyo Waterfront Bio-IT Research Building, 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan

Keywords: alignment, splice junction, mapping, NGS, RNA-seq

Background:
High-throughput sequencing of mRNA (RNA-seq) provides new opportunities to study alternative splicing events. Advances in sequencing technologies yielded longer reads but it also means an increased rate of sequencing error. Moreover, longer reads are desirable to span one or more exons, that will significantly reduce fraction of continuously mapped reads. Most of currently popular pipelines are designed for only short reads with low rates of sequencing error and shows weak support for reads spanning exon-exon junctions.

Results:
Spaln is a space efficient and fast method for mapping and aligning cDNA/EST sequences onto genomic sequence. We expended Spaln to enable mapping mid-to-long RNA-seq reads against both exonic regions and splice junctions with high sensitivity and accuracy. When tested on H. sapiens RNA-seq simulated data (2,885,618 reads with length of 100bp, an error rate of 5%), 89.61% reads were mapped against the whole human genome and 675,346 splice events were predicted, using EST sequences as validation, of which 78.94% is proved to be trusted.

Conclusions:
The result of our experiment definitely indicates that the expended Spaln is promising to map and align mid-to-long RNA-seq reads onto genomic sequence, suitable for detection of novel as well as known splice junctions. Spaln is freely available to non-commercial users at: ***http://www.genome.ist.i.kyoto-u.ac.jp/~aln_user/spaln***.