Bioinformatics Consortium of Taiwan

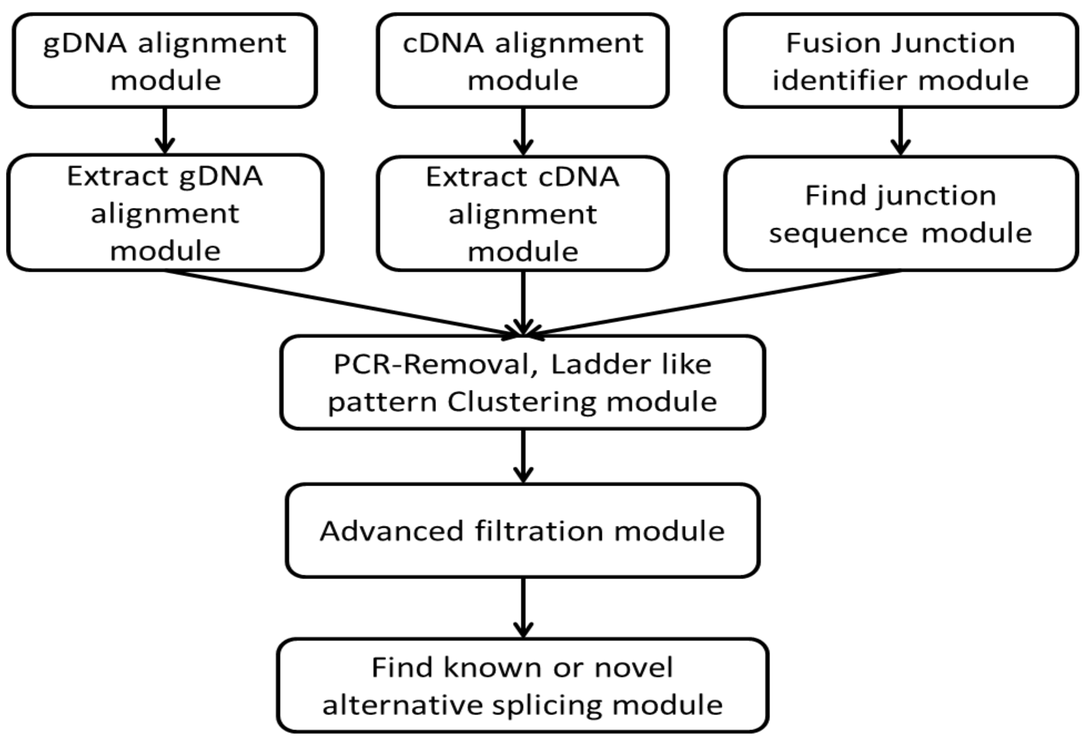
http://bct.binfo.org.tw

National Core Facility Program for Biotechnology- Bioinformatics Consortium of Taiwan

FusionDetector: a flexible pipeline for finding aberrant fusion genes using RNA-seq data

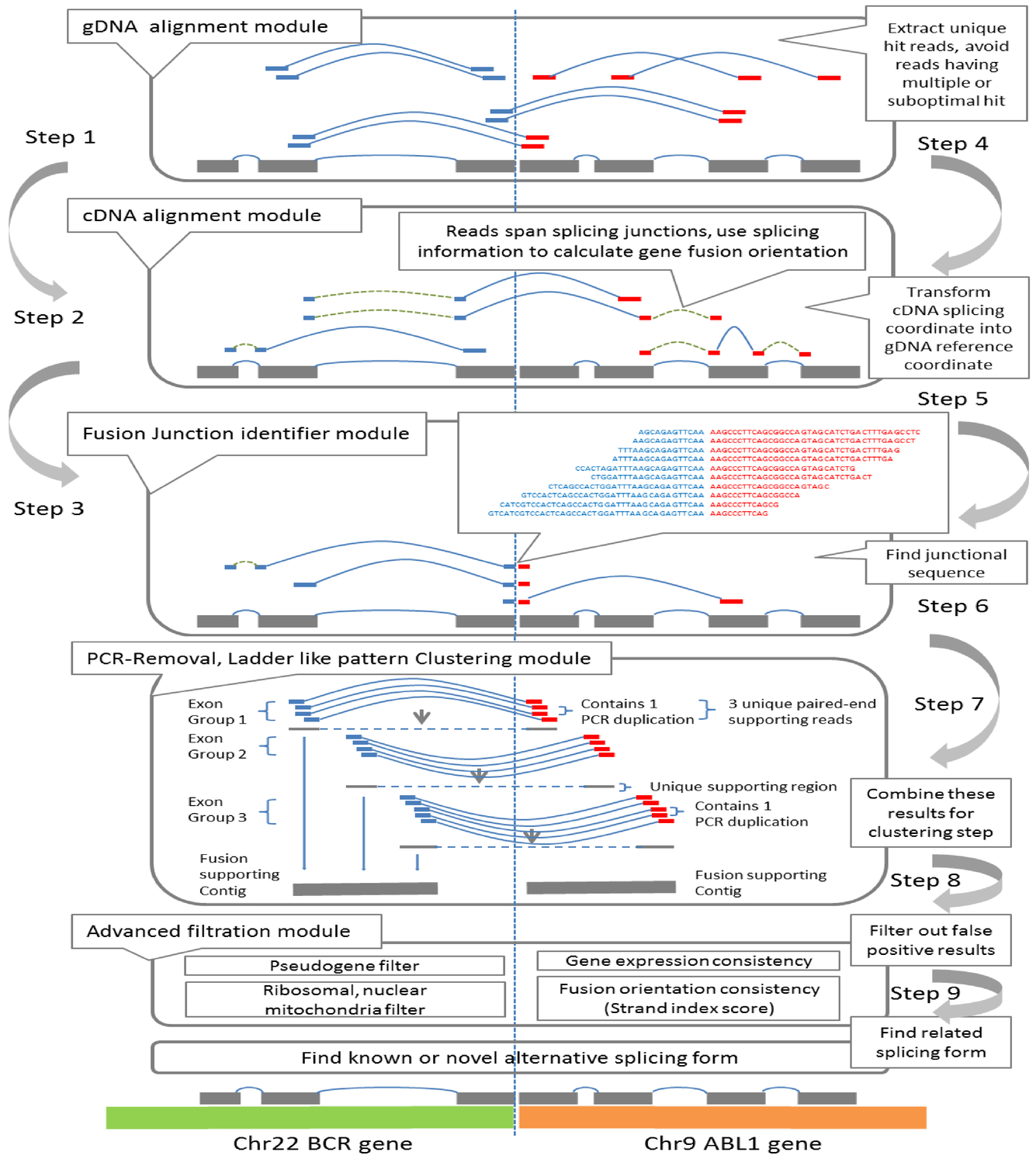
Introduction

Aberrant fusion genes are suggestive to play roles in carcinogenesis. However, finding them systematically has not been practical until the development of RNA-seq technology. To facilitate the decoding of RNA-seq results, a number of alignment algorithms as well as pipelines have been developed. However, available RNA-seq tools were generally not designed to give comprehensive information to facilitate the assessment of the quality of each candidate fusion gene.



Results

FusionDetector, a fusion-gene discovery pipeline, is designed to take advantage of available public-domain databases and tools, as well as data-exchange standards, to extract information that might be useful to determine aberrant fusion genes. FusionDetector takes the alignments generated by available alignment algorithms (1,2) as the start point to extract aberrant fusion-gene information. FusionDetector can integrate two types of reads-*versus*-references alignments, namely, the alignment of RNA-seq reads *versus* a reference genome and the alignment of RNA-seq reads *versus* known transcripts. FusionDetector annotates putative fusion genes based on Ensembl transcripts. To help biologists remove false positives, FusionDetector is implemented with three filters – a HGNC-based mis-alignment filter, a blacklist-based filter, and an expression-balance filter. Other supporting information like sequencing coverage, orientation of paired reads, and splice-junction patterns are also given in a tabulated format.



Conclusion

FusionDetector is an easy-to-use fusion-gene detection pipeline. A simple multi-core desktop computer is quite sufficient to run FusionDetector. Using public-domain RNA-seq data sets, we concluded that the performance of FusionDetector is quite comparable to that of other fusion-gene finding tools, such as FusionMap (3) and FusionSeq (4). With a simple multi-core desktop computer equipped with 8G RAM, the analysis of 0.25 giga reads could be finished within a week. FusionDetector can be easily adapted to other new RNA-seq aligning tools.

Reference

1.Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754-1760.

2.Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, **10**, R25.

3. Ge, H., Liu, K., Juan, T., Fang, F., Newman, M. and Hoeck, W. (2011) FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. *Bioinformatics*.

4.Sboner, A., Habegger, L., Pflueger, D., Terry, S., Chen, D.Z., Rozowsky, J.S., Tewari, A.K., Kitabayashi, N., Moss, B.J., Chee, M.S. *et al.* (2010) FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. *Genome Biol*, **11**, R104.

Welcome to contact with BCT E-mail: bct@ym.edu.tw Tel: 02-28267359 http://bct.binfo.org.tw

NCFPB 生技類核心設施平台維運計畫 National Core Facility Program for Biotechnology