SplitFusion: An Ultra-sensitive and Fast Method for Detecting Gene Fusions

Baifeng Zhang, Athena Chu, Siyu Bao, TSZ Yau, Wah Cheuk, Zongli Zheng

Ming Wai Lau Centre for Reparative Medicine, Karolinska Institutet, Hong Kong SAR

Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR

Email: baifeng.zhang@ki.se

Gene fusion is a hallmark of cancer and many fusions are effective therapeutic targets. Accurate and robust detection of gene fusion is challenging in both wet-lab and bioinformatics. We introduce SplitFusion, a clinic-grade bioinformatics solution for the gene fusion detection using split-reads from anchored multiplex PCR (AMP)-based next-generation sequencing data. SplitFusion leverages the fusion characteristics of BWA-mem secondary alignments and can efficiently identify fusion events in clinic data from low quality degraded samples. We compared the performance of SplitFusion with three other commonly used fusion detection software, namely Start-Fusion, Lumpy and EricScript. Using 11 ALK-EML4 and EWSR1-NFACT2 positive clinical tumor samples datasets sequenced to a high depth, with different down-sample sizes, SplitFusion showed the highest sensitivity and specificity, and consumed least computing resources with the least RAM footprint and an average time of 26 CPU seconds for processing 250K reads from FASTQ to inferred fusion breakpoints with 100% sensitivity. Further, SplitFusion can accurately infer in-frame and out-of-frame fusion of gene partners that is important to infer potential function of gene fusions.

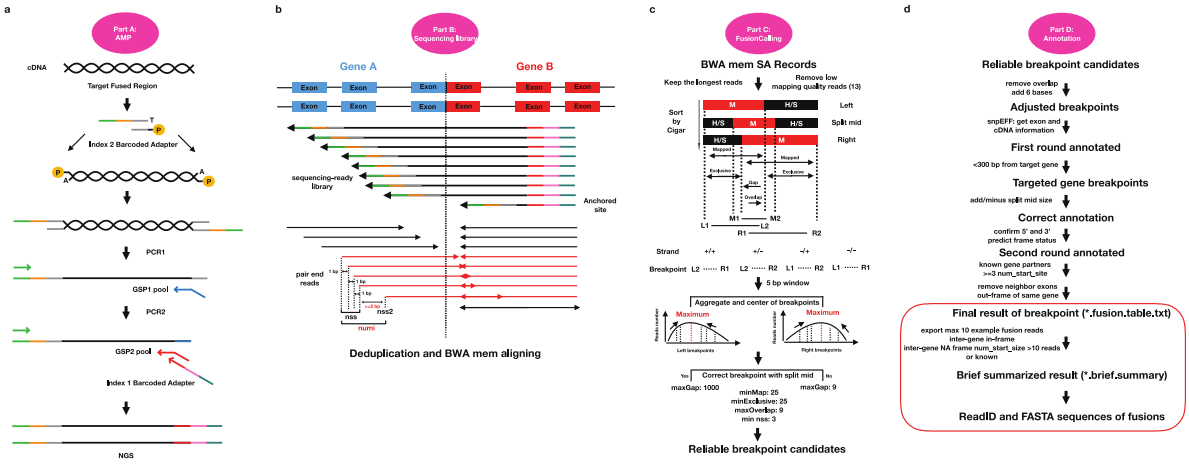

Figure 1. The workflow of SplitFusion

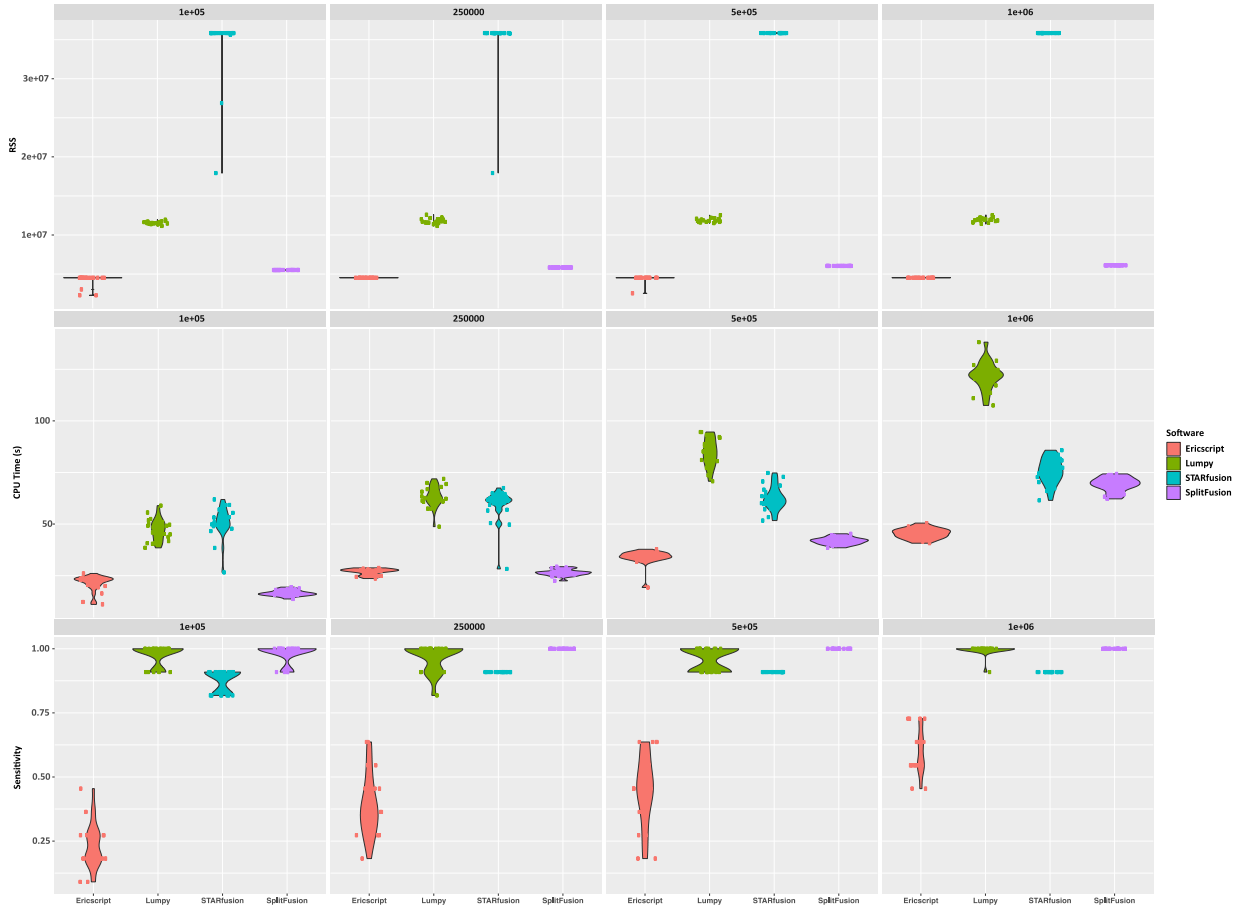

Figure 2. Benchmark test