Detecting Chimeric Transcripts in RNA-seq Data

Lucas Swanson[1,2], Karen Mungall[1], Gordon Robertson[1], Readman Chiu[1], Shaun D Jackman[1], Jenny Q Qian[1], Sam Lee[3], Deniz Yorukoglu[2], Rong She[1], Yongjun Zhao[1], Richard Moore[1], Marco A Marra[1], Steven JM Jones[1], Aly Karsan[1], Pamela A Hoodless[3], S Cenk Sahinalp[2], Inanc Birol[1]

[1] Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, B.C., Canada
[2] School of Computing Science, Simon Fraser University, Burnaby, B.C., Canada
[3] Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, B.C., Canada

Corresponding Authors:
Lucas Swanson: lucas.a.swanson@gmail.com
Cenk Sahinalp: cenk@cs.sfu.ca
Inanc Birol: ibirol@bcgsc.ca

Though its functional importance remains controversial, chimeric transcripts due to trans-splicing appear to occur not only in nematodes and kinetoplastids, but also in higher animals, including mammals.

We have developed a suite of analysis tools for identifying a range of types of chimeric transcripts using RNA-seq data. Based on the Trans-ABySS transcriptome assembly platform, these tools integrate event-support evidence from transcriptome contigs, their alignments to a reference genome, read alignments to assembled contigs and the genome, and gene and repeat annotations. Applied to a set of libraries, the toolset can also identify recurrent events, and can prioritize events for manual review.

Using these methods, we have characterized transcriptomes of seven embryonic and adult normal mouse tissues to identify fusions involving 22 genes, partial tandem duplications (PTDs) involving six genes, and internal tandem duplications (ITDs) involving 35 genes. Of these genes, fusions recurred across libraries in eight genes and ITDs in three genes. A subset of the fusions, PTDs, and ITDs appear to occur in homologous genes in humans, in public EST databases. Finally, we estimate whether the trans-splicing that we detected is likely to be functionally important by calculating the relative expression level of chimera and wild type transcripts.