Bioinformatics, 33(18), 2017, 2954–2956
doi: 10.1093/bioinformatics/btx329
Advance Access Publication Date: 19 May 2017
Applications Note

OXFORD

Gene expression

chimeraviz: a tool for visualizing chimeric RNA

Stian Lågstad,[1,2,3] Sen Zhao,[1,3] Andreas M. Hoff,[1,3] Bjarne Johannessen,[1,3] Ole Christian Lingjærde[2,3] and Rolf I. Skotheim[1,2,3,*]

[1]Department of Molecular Oncology, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway, [2]Department of Informatics, Faculty of Mathematics and Natural Sciences and [3]Centre for Cancer Biomedicine, Faculty of Medicine, University of Oslo, Oslo, Norway

*To whom correspondence should be addressed.

Associate Editor: Bonnie Berger

Received on February 24, 2017; revised on May 3, 2017; editorial decision on May 16, 2017; accepted on May 18, 2017

Abstract

Summary: Advances in high-throughput RNA sequencing have enabled more efficient detection of fusion transcripts, but the technology and associated software used for fusion detection from sequencing data often yield a high false discovery rate. Good prioritization of the results is important, and this can be helped by a visualization framework that automatically integrates RNA data with known genomic features. Here we present *chimeraviz*, a Bioconductor package that automates the creation of chimeric RNA visualizations. The package supports input from nine different fusion-finder tools: deFuse, EricScript, InFusion, JAFFA, FusionCatcher, FusionMap, PRADA, SOAPfuse and STAR-FUSION.

Availability and implementation: *chimeraviz* is an R package available via Bioconductor (https://bioconductor.org/packages/release/bioc/html/chimeraviz.html) under Artistic-2.0. Source code and support is available at GitHub (https://github.com/stianlagstad/chimeraviz).

Contact: rolf.i.skotheim@rr-research.no

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Chimeric RNA molecules, or fusion transcripts, are composed of sequences from two or more genes. These may encode novel proteins and play roles in the development of cancer (Frenkel-Morgenstern *et al.*, 2012). Advances in RNA sequencing have enabled more efficient detection of fusion transcripts, but the technology and associated software used to detect fusions from sequencing data often yield a high false discovery rate. Also, there is a high degree of discordance between different fusion-finder tools (Kumar *et al.*, 2016). Good prioritization of the results is important. This can be enabled by a visualization framework that automatically integrates nucleotide level RNA data and known genomic features such as transcript annotation and exon structures.

Here, we present *chimeraviz*, a Bioconductor package that visualizes chimeric RNA. *chimeraviz* implements a unified format for representing fusion transcripts, and provides multiple tools for visualizing chimeric RNA molecules as well as functions for sorting and filtering candidates.

2 Features

Using *chimeraviz*, the user can import data from nine fusion-finders: deFuse (McPherson *et al.*, 2011), EricScript (Benelli *et al.*,2012), InFusion (Okonechnikov, 2016), JAFFA (Davidson *et al.*, 2015), FusionCatcher (Nicorici *et al.*, 2014), FusionMap (Ge *et al.*, 2011), PRADA (Torres-Garcia *et al.*, 2014), SOAPfuse (Jia *et al.* 2013) and STAR-FUSION (Haas *et al.*, 2015). With transcript annotation data from Ensembl and aligned RNA-sequencing data in a .BAM file, the user can create multiple visualizations of candidate fusion transcripts. The plots are useful for illustrative purposes and may indicate the biological consequence of the putative fusion transcript. The user can also sort and filter fusion results based on various criteria.

3 Implementation and demonstration

chimeraviz is an R package that can be obtained from the Bioconductor project (Gentleman *et al.*, 2004). The package

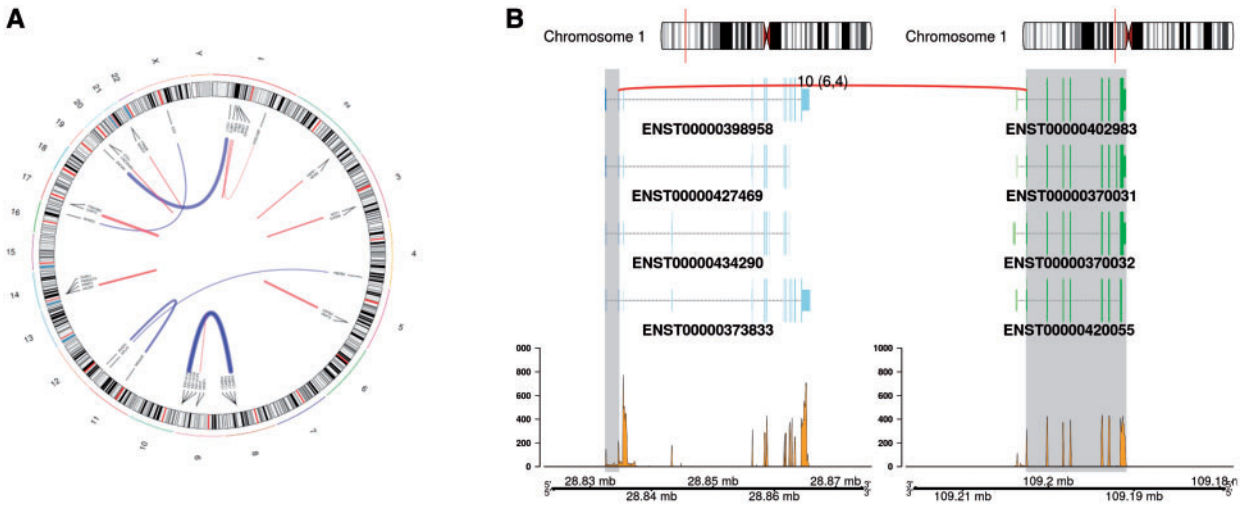

Fig. 1. Fusion landscape. (**A**) The circular plot shows an overview of fusion events between locations in chromosomes. Red and blue links indicate intra and inter-chromosomal fusions. (**B**) Example of fusion *RCC1-HENMT1*. The genomic view of the fusion event is from the top showing the gene loci within the chromosomes, the number of discordant (split and spanning) reads supporting breakpoint (curved red line), annotated exons of known transcripts of fusion partner genes, and plots of the RNA read counts along the genomic coordinates for the fusion partners in mega basepairs from p-telomere of chromosome 1

includes tutorials in the form of R vignettes, and test data required to produce the various visualizations available in *chimeraviz*. These data include RNA-sequencing data from an embryonal carcinoma cell line, 833Ke, recently used to detect and characterize novel fusion genes in testicular germ cell tumors (Hoff *et al.*, 2016).

Among the visualization types *chimeraviz* can produce, the overview plot and the fusion plot are demonstrated here (Fig. 1). Demonstration of the other plots, and filtering and sorting options in *chimeraviz*, can be seen in the Supplementary Material (vignette at https://bioconductor.org/packages/release/bioc/vignettes/chimeraviz/inst/doc/chimeraviz-vignette.html).

3.1 Overview plot

The overview plot gives a genome-scale summary of a sample's fusion landscape (Fig. 1A). The plotCircle() function in *chimeraviz* produces a circos plot with links indicating fusion transcripts, using the R package RCircos (Zhang *et al.*, 2013). The blue links indicate inter-chromosomal fusions whereas red links indicate intra-chromosomal fusions. The width of each link indicates the number of reads that the fusion-finder found to support the fusion junction.

3.2 Fusion plot

The fusion plot is a gene-pair centric and comprehensive visualization, which is useful in the evaluation of whether a fusion event is a true positive (Fig. 1B).

The genomic coordinates of both partner genes are shown with chromosome ideograms in the upper part of the plot and, more precisely, with basepair resolution on the x-axis in the bottom part.

Known transcript structures for the partner genes are shown in the middle section. Exons that are likely part of the fusion transcript are drawn in darker colors. These exons are also highlighted with gray rectangles. Untranslated regions are drawn as slightly thinner boxes.

A red link connects the partner genes at the breakpoint sites. The number of reads found to support the fusion junction is indicated on top of this link. The width of the link also represents the number of supporting reads.

RNA-seq coverage is shown in histograms below the transcripts. This visualization enables evaluation of whether exons included in the fusion transcripts have higher expression than other exons of the fusion partner genes.

4 Conclusion

We have developed an R package which can take input from multiple fusion-finder tools, create visualizations to illustrate the fusion transcripts, and apply functions for filtering and sorting lists of candidate fusion transcripts. These functionalities will facilitate the prioritization of true positive and important fusion transcripts.

Acknowledgements

We would like to thank Matteo Benelli, author of EricScript, and Roeland Verhaak, author of PRADA, for providing example output data to be included in chimeraviz.

Funding

This work has been supported by grants from the Norwegian Cancer Society (PR-2007-0166), Notur (project NN9313K), and the Research Council of Norway through its Centers of Excellence funding scheme (project number 179571).

Conflict of Interest: none declared.

References

Benelli,M. *et al.* (2012) Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. *Bioinformatics*, **28**, 3232–3239.

Davidson,N. *et al.* (2015) JAFFA: high sensitivity transcriptome-focused fusion gene detection. *Genome Med.*, **7**, 43.

Frenkel-Morgenstern,M. *et al.* (2012) Chimeras taking shape: potential functions of proteins encoded by chimeric RNA transcripts. *Genome Res.*, **22**, 1231–1242.

Ge,H. *et al.* (2011) FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. *Bioinformatics*, **27**, 1922–1928.

Gentleman,R. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.

Haas,B. *et al*. (2015) STAR-FUSION, https://github.com/STAR-Fusion/STAR-Fusion (13 April 2016, date last accessed).

Hoff,A.M. *et al*. (2016) Identification of novel fusion genes in testicular germ cell tumors. *Cancer Res*., **76**, 108–116.

Jia,W. *et al*. (2013) SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. *Genome Biol*., **14**, R12.

Kumar,S. *et al*. (2016) Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data. *Sci. Rep*., **6**, 21597.

McPherson,A. *et al*. (2011) deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput. Biol*., **7**, e1001138.

Nicorici,D. *et al*. (2014) FusionCatcher - a tool for finding somatic fusion genes in paired-end RNA-sequencing data. doi: https://doi.org/10.1101/011650.

Okonechnikov,K. (2016) InFusion, https://bitbucket.org/kokonech/infusion (21 April 2016, date last accessed).

Torres-Garcia,W. *et al*. (2014) PRADA: pipeline for RNA sequencing data analysis. *Bioinformatics*, **30**, 2224–2226.

Zhang,H. *et al*. (2013) RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics*, **14**, 244.