BIOINFORMATICS APPLICATIONS NOTE Vol. 29 no. 13 2013, pages 1700–1701
doi:10.1093/bioinformatics/btt252

Structural bioinformatics

Advance Access publication May 9, 2013

iFUSE: integrated fusion gene explorer

Saskia Hiltemann[1,*], Elizabeth A. McClellan[2], Jos van Nijnatten[2], Sebastiaan Horsman[2], Ivo Palli[2], Ines Teles Alves[1,3], Thomas Hartjes[1], Jan Trapman[3], Peter van der Spek[2], Guido Jenster[1] and Andrew Stubbs[2,*]

[1]Department of Urology, [2]Department of Bioinformatics and [3]Department of Pathology, Erasmus MC, 3015 GE, Rotterdam, The Netherlands

Associate Editor: Martin Bishop

ABSTRACT

Summary: We present iFUSE (integrated fusion gene explorer), an online visualization tool that provides a fast and informative view of structural variation data and prioritizes those breaks likely representing fusion genes. This application uses calculated break points to determine fusion genes based on the latest annotation for genomic sequence information, and where relevant the structural variation (SV) events are annotated with predicted RNA and protein sequences. iFUSE takes as input a Complete Genomics (CG) junction file, a FusionMap fusion detection report file or a file already analysed and annotated by the iFUSE application on a previous occasion.

Results: We demonstrate the use of iFUSE with case studies from tumour-normal SV detection derived from Complete Genomics whole-genome sequencing results.

Availability: iFUSE is available as a web service at http://ifuse.erasmusmc.nl.

Contact: s.hiltemann@erasmusmc.nl or a.stubbs@erasmusmc.nl

Received on January 7, 2013; revised on April 15, 2013; accepted on May 1, 2013

1 INTRODUCTION

Structural variation analysis is a common requirement in the study of cancer where many fusion genes have been implicated in the progression of cancer (Kumar-Sinha *et al.*, 2008; Mitelman *et al.*, 2007). The use of next-generation sequencing for fusion gene detection in cancer (Edgren *et al.*, 2011; Ge *et al.*, 2011; McPherson *et al.*, 2011), structural variation in non-cancerous diseases (Levy *et al.*, 2011; Sanders *et al.*, 2011) and in normal genomes (The 1000 Genomes Project Consortium, 2010) has expanded knowledge of the importance of these events. In a recent study, the use of *de novo* assembly in association with SV detection suggests that SVs account for more diversity between individuals than do single-nucleotide polymorphisms (Li *et al.*, 2011). Complete Genomics uses *de novo* assembly during SV, single-nucleotide variation and copy number variation determination (Carnevali *et al.*, 2012; Drmanac *et al.*, 2010; Reumers *et al.*, 2011).

Traditionally, for a given SV, a user can visualize the individual (sides of the) breaks in viewers, such as IGV (Robinson *et al.*, 2011; Thorvaldsdóttir *et al.*, 2012), but not the resulting event or fusion gene as a whole, and must manually retrieve the sequence of the resulting event based on the exons from both genes of the proposed fusion gene and determine the orientation and frame of the predicted transcript and encoded polypeptide sequences. Other applications also offer visualization, but with the caveat that the user must process their data within the application, e.g. inGAP-SV (Qi *et al.*, 2011), or that a bioinformatician is required to render the visualization using, e.g. Circos (Krzywinski *et al.*, 2009). Our aim is to deliver a web-based application that allows scientists to visualize all detected SV events and fusion genes determined in their results and to provide the concomitant candidate transcripts and polypeptide sequences associated with the detected fusion genes (Fig. 1). No other application exists at the moment to categorize and visualize the candidate fusion genes and to determine the proposed sequence for gDNA, RNA and polypeptides from standard SV files.

2 METHODS

iFUSE is a PHP-coded application running on an Apache web server with a mySQL database for user management and R for data analysis. Gene-based feature annotation is provided using the UCSC gene tables (hg18 and hg19). Documentation details, including full application configuration, are available from the website (http://ifuse.erasmusmc.nl).

iFUSE takes as input either a Complete Genomics (CG) junctions file or a fusion detection report as generated by the FusionMap tool (Ge *et al.*, 2011). To perform a comparison of two or more genomes, the Complete Genomics tool Junctions2Events can be used before visualization within iFUSE. The input file is validated and then analysed using R, after which a graphical representation for each event is generated. This representation displays the promotor, introns, exons and the junction site, and additional information, including DNA, RNA and protein sequences, is presented to the user. These events can be filtered and sorted by the user, either using general properties or by zooming in on a single event and filtering for nearby junctions or events with similar properties.

The input files uploaded to iFUSE are annotated in R using information retrieved from UCSC gene tables and the input files. The resulting output file can be retrieved from iFUSE for manual inspection and can also be used directly as input to iFUSE.

Example input files can be downloaded from the downloads section of the iFUSE website, and users can test the application without registration by selecting the option to start an anonymous session. Any files uploaded by the user will be deleted at the end of the anonymous session.

iFUSE accounts are password protected, the application has been tested for security risks, such as SQL injections, and our servers undergo periodical CERT vulnerability scans (http://www.cert.org). Furthermore,

*To whom correspondence should be addressed.

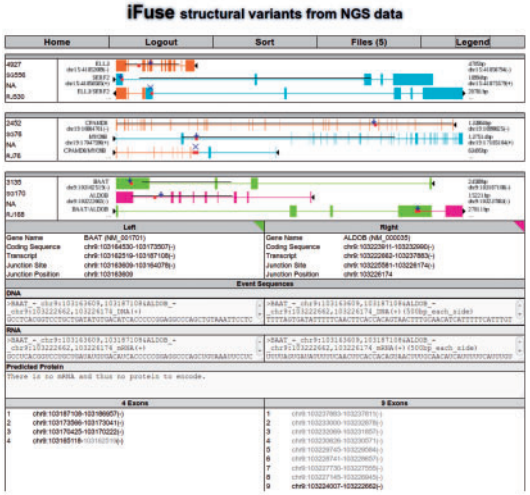

Fig. 1. Screenshot of iFUSE. SVs are visualized and where applicable, the predicted DNA, RNA and protein sequences are reported

Table 1. Results from iFUSE on public datasets

	S1 Tumour	S1 Normal	S2 Tumour	S2 Normal
HG18				
Junctions	1579	1594	1558	1387
Genes on both sides	32	15	23	4
Same orientation	21	14	16	1
HG19				
Junctions	1581	1592	1559	1390
Genes on both sides	21	12	31	7
Same orientation	10	6	17	3

Note: Sample 1 (S1): HCC1187, Sample 2 (S2): HCC2218, public datasets downloadable from Complete Genomics. (ftp://ftp2.completegenomics.com/Cancer_pairs/)

when a user deletes a file via the iFUSE web interface, it is purged completely from our systems.

3 DISCUSSION

Two public cancer genomes have been used to demonstrate the use of iFUSE for the prediction of fusion genes. Genomes were downloaded from Complete Genomics (ftp2.completegenomics.com). The results can be found in Table 1.

An event is labelled as a fusion gene if the break point has two different genes on either side. If the two sides also have the same orientation (are on the same strand, or in the case of an inversion on opposite strands) and are also in frame, the event is called a fusion protein.

4 CONCLUSION

iFUSE provides scientists with a convenient method to visualize, categorize and filter structural variation analysis output using Complete Genomics junction files or the FusionMap fusion detection report files as input to the application.

ACKNOWLEDGEMENTS

The authors thank Rick Tearle and Steve Lincoln from Complete Genomics whose valuable discussions on Complete Genomics analysis methods supported our application development.

Funding: Center for Translational Molecular Medicine. TraIT project (05T-401).

Conflict of Interest: none declared.

REFERENCES

Carnevali,P. *et al.* (2012) Computational techniques for human genome resequencing using mate gapped reads. *J. Comput. Biol.*, **19**, 279–292.

Drmanac,R. *et al.* (2010) Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science*, **327**, 78–81.

Edgren,H. *et al.* (2011) Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol.*, **12**, R6.

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

Krzywinski,M. *et al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.*, **19**, 1639–1645.

Kumar-Sinha,C. *et al.* (2008) Recurrent gene fusions in prostate cancer. *Nat. Rev.*, **8**, 497–511.

Levy,D. *et al.* (2011) Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron*, **70**, 886–897.

Li,Y. *et al.* (2011) Structural variation in two human genomes mapped at single-nucleotide resolution by whole genome de novo assembly. *Nat. Biotechnol.*, **29**, 723–730.

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

Mitelman,F. *et al.* (2007) The impact of translocations and gene fusions on cancer causation. *Nat. Rev.*, **7**, 233–245.

Qi,J. *et al.* (2011) inGAP-sv: a novel scheme to identify and visualize structural variation from paired end mapping data. *Nucleic Acids Res.*, **39**, 567–575.

Reumers,J. *et al.* (2011) Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. *Nat. Biotechnol.*, **30**, 61–68.

Robinson,J.T. *et al.* (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.

Sanders,S.J. *et al.* (2011) Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*, **70**, 863–885.

The 1000 Genomes Project Consortium. (2011) A map of human genome variation from population scale sequencing. *Nature*, **467**, 1061–1073.

Thorvaldsdóttir,H. *et al.* (2012) Integrative genomics viewer (IGV): high-performance genomics data visualisation and exploration. *Brief. Bioinform.*, **14**, 178–192.