BITS :: Call for Abstracts 2021 - Poster

Type	Poster
Session	Gene regulation, transcriptomics and epigenomics
Title	Intron retention can bias accurate transposable elements quantification in RNA-seq data
All Authors	Gualandi N(1), Iperi C(1), Esposito M(1), Ansaloni F(1,2), Sanges R(1,2)
Affiliation	

(1) Computational Genomics Laboratory, Area of Neuroscience, Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste
(2) Central RNA Laboratory, Istituto Italiano di Tecnologia, Genova

Motivation

Intron retention (IR) is one of the 5 different alternative splicing (AS) events nature has evolved so far. It recently gain popularity due to its recognized role in gene expression regulation and its association with complex diseases [1]. IR occurs when introns, normally removed from the mature transcript by the splicing machinery, are maintained in the mature transcript [2]. Unlike previously expected, IR is a widespread and evolutionary conserved mechanism present in both health and disease [2], affecting many different genes with cell types specific patterns [3]. For example it was demonstrated that IR affects about the 17.8% (5462 introns) of expressed genes in human granulocytes [4].

Transposable elements (TEs) are repetitive sequences that comprise nearly half of the human genome [5]. Their main characteristic is the ability to change their location within the genome and they are virtually found in all eukaryotes [6,7]. TEs are key players in many different biological processes in health and disease [8,9,10], thus a reliable quantification of their expression as independent transcriptional units is crucial to distinguish TE expression from transcription of their sequences as part of canonical transcripts [11]. TE fragments are indeed embedded within most genes and in particular inside intronic sequences. The magnitude of this relation is so high that about 90% of all human RefSeq genes contain TEs in their introns [12]. This strict relation poses challenges in the quantification of intronic TE expression as reads coming from retained introns (RI) can bias the detection of overlapping TE expression resulting in an incorrect quantification that can lead to misleading conclusions.

We use RNA-seq data from lymphoblastoid cell lines to demonstrate the strong impact of IR on intronic TE quantification. Our work highlights a strong correlation between IR and TE expression that cannot be ignored when a unbiased quantification of TEs expression is required, in particular in samples derived from cell types in which IR is one of the most abundant AS event therefore playing a fundamental role in that specific biological context.

Methods

We use poly(A)+ RNA-seq data of lymphoblastoid cell lines from 462 healthy individuals from the Geuvadis project [13]. IR was quantified using IRFinder [14] while locus-specific TEs expression was assessed using SQuIRE [15] normalizing raw counts using the DESeq2 default median of ratios approach [16]. To understand the effect of IR on the measurements of intronic TEs quantification, correlations between IRratio and embedded TEs expression were performed using the Person correlation test. RI were defined as intron with an IRratio >= 0.1 and no overlap with other annotated features, while expressed TEs were defined as elements with normalized counts >= 5. Intersection were performed using bedtools intersect [17] and compared to 1000 randomization of the same size of input features.

Results

Our results show that IR is a widespread mechanism affecting about 6000 introns in lymphoblastoid cells in a physiological status. In this scenario we observe that IR affects the unbiased TE quantification mainly by increasing the number of reads mapped on TEs therefore leading to aberrant over-quantification. Using correlations analysis we show that more than 90% of intronic TE quantification positively correlates with the level of IR, while only the 0.28% have a negative correlation. Moreover, the number of expressed intronic TEs strongly correlates with the number of RI. Following randomization analysis using a set of highly confidence RI and expressed TEs shared by at least the 65% of samples, prove that RI are enriched to contain expressed TEs even if they result significantly depleted of TEs sequences with respect to a matched-size set of non-RI. These results confirm that IR affects TEs quantification and that the presence of intronic TEs sequences do not increase IR.

Our study demonstrates that IR can prevent an accurate and independent TEs quantification. This is particularly important when disease samples are compared with healthy ones because changes in IR, which appears to be a feature of several diseases, can lead to wrong interpretation of TEs differential expression and misleading biological conclusions.

Info

Bibliography

1. Monteuuis, G., Wong, J. J. L., Bailey, C. G., Schmitz, U. & Rasko, J. E. J. The changing paradigm of intron retention: regulation, ramifications and recipes. Nucleic Acids Research 47, 11497–11513 (2019).
2. Zheng, J.-T., Lin, C.-X., Fang, Z.-Y. & Li, H.-D. Intron Retention as a Mode for RNA-Seq Data Analysis. Front. Genet. 11, (2020).
3. Braunschweig, U. et al. Widespread intron retention in mammals functionally tunes transcriptomes. Genome Res 24, 1774–1786 (2014).
4. Schmitz, U. et al. Intron retention enhances gene regulatory complexity in vertebrates. Genome Biology 18, 216 (2017).
5. Lander, E. S. et al. Initial sequencing and analysis of the human genome. Nature 409, 860–921 (2001).
6. Wessler, S. R. Transposable elements and the evolution of eukaryotic genomes. PNAS 103, 17600 –17601 (2006).
7. Kidwell, M. G. Transposable elements and the evolution of genome size in eukaryotes. Genetica 115, 49 –63 (2002).
8. Payer, L. M. & Burns, K. H. Transposable elements in human genetic disease. Nature Reviews Genetics 20, 760–772 (2019).
9. Reilly, M. T., Faulkner, G. J., Dubnau, J., Ponomarev, I. & Gage, F. H. The Role of Transposable Elements in Health and Diseases of the Central Nervous System. J Neurosci 33, 17577–17586 (2013).
10. Saleh, A., Macia, A. & Muotri, A. R. Transposable Elements, Inflammation, and Neurological Disease. Front. Neurol. 10, (2019).
11. Lanciano, S. & Cristofari, G. Measuring and interpreting transposable element expression. Nature Reviews Genetics 21, 721–736 (2020).
12. Zhang, Y., Romanish, M. T. & Mager, D. L. Distributions of Transposable Elements Reveal Hazardous Zones in Mammalian Introns. PLOS Computational Biology 7, e1002046 (2011).
13. Lappalainen, T. et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 501, 506–511 (2013).
14. Middleton, R. et al. IRFinder: assessing the impact of intron retention on mammalian gene expression. Genome Biology 18, 51 (2017).
15. Yang, W. R., Ardeljan, D., Pacyna, C. N., Payer, L. M. & Burns, K. H. SQuIRE reveals locus-specific regulation of interspersed repeat expression. Nucleic Acids Res 47, e27–e27 (2019).
16. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, (2014).
17. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).

Figure

-

Availability -

Corresponding Author

Name, Surname	Remo, Sanges
Email	remo.sanges@gmail.com
Submitted on	08.05.2021