BITS :: Call for Abstracts 2022 - Oral communication

Type	Oral communication
Session	Biological Databases
Title	REDIportal: the RNA editing comprehensive catalog
All Authors	Mansi L. (1), Lo Giudice C. (1), Pesole G.(1,2,3), Picardi E. (1,2,3)
Affiliation	

(1) Department of Biosciences, Biotechnologies and Biopharmaceutics (DBBB), University of Bari, Via Orabona 4, 70125 Bari, Italy.
(2) Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies (IBIOM), National Research Council, Via Amendola 122/O, 70126 Bari, Italy.
(3) National Institute of Biostructures and Biosystems (INBB), 00136 Roma, Italy.

Motivation

RNA editing is a relevant epitranscriptome phenomenon able to increase the transcriptome and proteome diversity of eukaryotic organisms. ADAR mediated RNA editing is widespread in humans in which millions of A-to-I changes modify thousands of primary transcripts. RNA editing has pivotal roles in the regulation of gene expression or modulation of the innate immune response or functioning of several neurotransmitter receptors. To support the study of A-to-I RNA editing we developed the specialized REDIportal database. In its last release we raised the content to about 16 millions of events detected in 9642 human RNAseq samples from the GTEx project by using a dedicated pipeline based on our REDItools software (PMID: 32838738).

Methods

GTEx RNAseq reads were aligned onto the human genome (hg19/GRCh37) using STAR providing known gene annotations from Gencode. DNAseq reads, instead, were aligned onto the human genome (hg19/GRCh37) using BWA.
The de novo detection of RNA editing events was performed by means of an optimized version of our REDItools package. Initially, REDItools was launched on individual aligned RNAseq reads with non-stringent parameters. DNAseq support was subsequently added if available. Each output table was then filtered according to our protocol described in Lo Giudice et al. 2020 (PMID: 31996844). Briefly, all detected positions were annotated using known SNP sites, repeated elements in RepeatMasker and known editing events stored in the first release of the REDIportal database. SNPs and sites not supported by at least 10 DNAseq reads (if available) were removed, and the remaining positions were grouped in "ALU" (residing in Alu elements), "REP NON ALU" (residing in repetitive non Alu elements) and "NON REP" (residing in non-repetitive regions) groups, according to RepeatMasker annotations. While "NON REP" and "REP NON ALU" variants underwent a second round of REDItools using stringent call criteria to exclude multi-mapping reads and PCR duplicates, changes in ALU regions were filtered only by coverage (at least 5 reads) and base quality (phred score of at least 30).
At the end, all filtered positions were collected returning the final list of RNA editing candidates per RNAseq.

Results

REDIportal provides RNA editing details for 9642 RNAseq samples from 549 individuals across 31 tissues and 54 body sites. According to our pipeline, each sample contains 60873 edited events on average. The highest number of events is detected in the cerebellar hemisphere and cerebellum. By contrast, skeletal muscle and heart contain the lowest number of candidates. The majority of A-to-I events reside in Alu elements, located mainly in intronic regions. A consistent fraction of sites is identified in intergenic and 3′ UTR regions, A-to-I events in exonic regions, instead, are very limited. The RNA editing activity, measured by the Alu Editing Index (PMID: 31636457), is largely body site specific as well as the activity at recoding sites.

Info

Rediportal is freely available at http://srv00.recas.ba.infn.it/atlas/index.html.

filename	-
Figure	

-

Availability	https://doi.org/10.1093/nar/gkaa916

Corresponding Author

Name, Surname	Luigi, Mansi
Email	luigi.mansi@uniba.it

Submitted on	28.04.2022