BITS :: Call for Abstracts 2023 - Oral communication

Type	Oral communication
Session	Artificial Intelligence for Bioinformatics
Title	Detecting A-to-I RNA editing by Convolutional Neural Network
All Authors	Pietro Luca Mazzacuva2, 5, Adriano Fonzino1, Graziano Pesole1,2,3, Ernesto Picardi1,2,4

Affiliation

1. Department of Biosciences, Biotechnology and Environment, University of Bari, Via Orabona 4,
70126 Bari, Italy
2. Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies (IBIOM), National Research Council, Via Amendola 122/O, 70126 Bari, Italy
3. Consorzio Interuniversitario Biotecnologie, via Flavia 23/1, 34148 Trieste, Italy
4. National Institute of Biostructures and Biosystems (INBB), 00136 Roma, Italy
5 Department of Engineering, University Campus Bio-Medico di Roma , Via Alvaro Del Portillo 21,
00128 Roma, Italy

Motivation

In mammals, the main non-transient RNA modification involves the conversion of Adenosine (A) into Inosine (I), referred to as A-to-I RNA editing. It is carried out by ADAR enzymes and plays relevant physiological roles, some of them not yet fully explored. Additionally, the deregulation of A-to-I editing has been associated with several diseases. Genuine RNA editing events can be detected using transcriptome reads from second and third generation sequencing technologies. Several attempts have also been done employing machine learning algorithms and known RNA editing sites from public databases. Recently, we profiled A-to-I editing in about 10,000 RNAseq experiments from the GTEx project, unveiling more than 15M events that have been collected in the specialized REDIportal database. This huge amount of editing sites provided us the unique opportunity to develop an ad hoc machine learning algorithm to accurately detect A-to-I events in RNAseq data.

Methods

To identify A-to-I editing sites from RNAseq data, we have developed a novel Convolutional Neural Networks (CNN) based Machine Learning method, in which sequencing data have been processed and treated like audio data. In particular, known RNA editing events as well as RNA variants not due to ADAR enzymes have been obtained from REDItools tables used to populate the REDIportal database. For each site, we extracted an interval of 101 bases, centered at the putative editing position, and collected the corresponding RNAseq and WGS/WES frequencies (if any). After the encoding and several preprocessing standardization steps, the CNN was trained via tenfold cross-validation and the performances were measured in terms of specificity, sensitivity, precision, F1 score and accuracy

Results

We tested our new CNN machine learning classifier in ten datasets with a growing percentage of known RNA editing events (from 20% to 100%) and found a mean detection accuracy of 0.99. Next, we applied our classifier to RNA sites from a wild type and ADAR knockout human kidney cell line and reached an average accuracy of about 0.98. Our preliminary results clearly demonstrate the predicting power of our machine learning approach for unveiling transcriptome-wide A-to-I editing events.

Info

-

filename	-

Figure

-

Availability	-

Corresponding Author

Name, Surname	Pietro Luca, Mazzacuva
Email	pietroluca.mazzacuva@unicampus.it

Submitted on	02.05.2023