Genome wide quantification of A-to-I RNA editing activity

CURRENT STATUS: POSTED

Shalom Hillel Roth
The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan 5290002, Israel

ORCiD: https://orcid.org/0000-0003-3773-0798

Erez Y. Levanon
The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan 5290002, Israel

ORCiD: https://orcid.org/0000-0002-3641-4198

Eli Eisenberg
Raymond and Beverly Sackler School of Physics and Astronomy and 3Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

elieis@post.tau.ac.il*Corresponding Author*
ORCiD: https://orcid.org/0000-0001-8681-3202

DOI:
10.21203/rs.2.12827/v1

SUBJECT AREAS
Computational biology and bioinformatics *Genetics*

KEYWORDS
ADAR, Editing, RNA Editing, A-to-I RNA Editing, Global Editing, Computational tool

Abstract

Adenosine to inosine (A-to-I) RNA editing by the ADAR enzymes is a common RNA modification, preventing false activation of the innate immune system by endogenous dsRNAs. Methods for quantification of ADAR activity are sought after, due to an increasing interest in the role of ADARs in cancer and auto-immune disorders, as well as attempts to harness the ADAR enzymes for RNA engineering. Here we present the *Alu* Editing Index (AEI), a robust and simple-to-use computational tool devised for this purpose that produces a single number representing the global editing level from BAM files. The AEI tool is available at https://github.com/a2iEditing/RNAEditingIndexer

Introduction

Adenosine to inosine (A-to-I) RNA editing is a common and critical RNA modification, affecting millions of sites in the human genome. The vast majority of RNA editing activity, mediated primarily by ADAR (ADAR1), occurs in noncoding parts of the transcriptome[1]. The typical target consists of a long noncoding dsRNA formed by pairing of two inverted copies of repetitive elements[2]. In particular, the human transcriptome harbors more than a million copies of the ~300bp-long *Alu* repeat. Due to their abundance, pairing of two neighboring *Alu* elements is common, resulting in long dsRNAs. Millions of editing sites have been identified within human *Alu* repeats[2–8], accounting for almost all editing activity in the human transcriptome.

Comparative analysis of ADAR activity across multiple samples is as yet an unmet challenge. For this purpose, one desires a single quantitative measure representing the global editing behavior in each sample. Since virtually all human editing events take place in *Alu* repeats[8], we define and present here the *Alu* Editing Index (AEI), a robust measure that retains the full *Alu* editing signal. We provide a software package, enabling a straight-forward calculation of the AEI for multiple samples from raw RNA-sequencing data.

For any set of pre-determined genomic regions, the editing index is defined as the weighted average of the observed editing levels over all adenosines, where the weights are the total coverage of each site[9]. It avoids the need to quantify the editing level per-site and accounts for the editing activity, including editing events residing in low-coverage regions, while maintaining a very low false discovery

rate including lowly-covered regions. For convenience, we multiply this by 100, so that the index describes the per-cent level of editing in the set of regions. In human, virtually all ADAR activity occurs within *Alu* elements[5–8], spanning ~11% of the human genome. Thus, we study hereafter the *Alu* Editing Index (AEI), which is the weighted average over tens of millions of genomic adenosine sites located within *Alu* sequences. Note that the same approach can be used for other sets of genomic regions in any organism, as long as editing therein is strong enough to allow for sufficiently large signal-to-noise ratio (SNR).

To estimate the noise, mismatches due to technical and biological source rather than *bona fide* A-to-I editing, we calculate "control indexes", measuring the abundance of other types of mismatches (many times the highest of these is the C-to-T index, as it represents both a common genomic mutation and C-to-U RNA editing).

Equipment

A computer (or cluster) capable of running SAMtools and BEDTools.

For installation instructions, please see the installation guide at

https://github.com/a2iEditing/RNAEditingIndexer

Procedure

Step A – Alignment

First, remove duplicates and align the reads to the reference genome using your aligner of choice. Retain uniquely mapped reads only. Sorting and indexing your BAM files can save run-time later, but is not necessary.

Step B – Running RNAEditingIndexer

The BAM files and the corresponding reference genome are to be used as the input to the RNAEditingIndexer. We recommend using either a virtual environment such as conda or to use the provided Docker image).

For human, the RNAEditingIndexer averages (by default) on Alu repeats. For mouse, the default regions are B1 and B2 repeats. If your genome doesn’t have built-in support you will need to prepare the needed regions, annotations and SNPs files as described in the documentation.

This step uses all the programs described in the documentation (BamUtils, SAMtools, and BEDTools). It is recommended to check both the logs and the output to the screen for errors as some errors will not abort the execution

Anticipated Results

This section describes the most common (and thus the default) output. More complex output files can be generated, please check the full documentation at https://github.com/a2iEditing/RNAEditingIndexer/tree/master/Docs.

A successful run of the RNAEditingIndexer should produce a CSV output file containing the sample metadata (e.g. name, path, etc.) and the “editing” index (i.e. the weighted-average mismatch rate over the regions of interest) for each mismatch type selected (by default, all 6 strand-insensitive mismatch types). The A-to-I editing level is the A2GEditingIndex. The noise level may be estimated (conservatively) by the C2TEditingIndex. A signal to noise ratio below 2 or 3 is a signature of unreliable results. You can then use the values per sample as is, with or without baseline subtraction. No further normalization is required.

Please note that it is extremely unlikely to receive exactly zero as the index values for all mismatch types. If this appears in the output, the most likely scenario is that the calculation failed for this sample.

References

1. Tan, M. H. *et al.* Dynamic landscape and regulation of RNA editing in mammals. *Nature* **550**, 249–254 (2017).
2. Porath, H. T., Knisbacher, B. A., Eisenberg, E. & Levanon, E. Y. Massive A-to-I RNA editing is common across the Metazoa and correlates with dsRNA abundancee. *Genome Biol.* **18**, 185 (2017).
3. Bazak, L. *et al.* A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res.* **24**, 365–376 (2014).
4. Ramaswami, G. & Li, J. B. RADAR: A rigorously annotated database of A-to-I RNA editing.

Nucleic Acids Res. **42**, D109-13 (2014).

5. Kim, D. D. Y. *et al.* Widespread RNA editing of embedded Alu elements in the human transcriptome. *Genome Res.* **14**, 1719–1725 (2004).

6. Blow, M., Futreal, A. P., Wooster, R. & Stratton, M. R. A survey of RNA editing in human brain. *Genome Res.* **14**, 2379–2387 (2004).

7. Athanasiadis, A., Rich, A. & Maas, S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.* **2**, e391 (2004).

8. Levanon, E. Y. *et al.* Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat. Biotechnol.* **22**, 1001–1005 (2004).

9. Bazak, L., Levanon, E. Y. & Eisenberg, E. Genome-wide analysis of Alu editability. *Nucleic Acids Res.* **42**, 6876–6884 (2014).

Acknowledgements

We thank Orshay Gabay, Ilana Buchumenski, Ariel Fieglin, Yishay Pinto and other lab members for assistance in data analysis and many helpful comments. EYL was supported by JDRF Innovative Grant (1-INO-2018-639-A-N) and the Israel Science Foundation (1380/14). EE was supported by the Israel Science Foundation (2673/17 and 1945/18).

Genome-wide quantification of ADAR adenosine-to-inosine RNA editing activity

by Shalom Hillel Roth, Erez Y. Levanon, Eli Eisenberg

Nature Methods (21 October, 2019)

"Genome wide quantification of A-to-I RNA editing activity"; Shalom Hillel Roth, Erez Y. Levanon, Eli Eisenberg; In revision in Nature Methods