SOFTWARE TOOL ARTICLE

DRETools: A tool-suite for differential RNA editing detection
[version 1; referees: awaiting peer review]

Tyler Weirick¹, Patrick Trainor²,3, Eric Rouchka⁴,5, Andrew DeFilippis²,3, Shizuka Uchida¹,3

¹Cardiovascular Innovation Institute, University of Louisville, Louisville, KY, 40202, USA
²Diabetes and Obesity Center, University of Louisville, Louisville, KY, 40202, USA
³Institute of Molecular Cardiology, University of Louisville, Louisville, KY, 40202, USA
⁴Kentucky Biomedical Research Infrastructure Network Bioinformatics Core, University of Louisville, Louisville, KY, 40202, USA
⁵Department of Computer Engineering and Computer Science, University of Louisville, Louisville, KY, 40202, USA

Abstract
Recent tools to detect RNA editing have expanded our understanding of epitranscriptomics, linking changes in RNA editing to both disease and normal cellular processes. However, the research community currently lacks tools for determining if change in RNA editing or "differential editing" has occurred. To meet this need, we present DRETools, a command-line tool-set for finding differential editing among samples, editing islands, and editing sites.

Keywords
epitranscriptomics, RNA-seq, RNA editing, differential RNA editing, editing-per-kilobase, EPK

Corresponding author: Shizuka Uchida (heart.lncrna@gmail.com)

Author roles: Weirick T: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Trainor P: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation; Rouchka E: Funding Acquisition, Writing – Review & Editing; DeFilippis A: Supervision, Writing – Review & Editing; Uchida S: Conceptualization, Funding Acquisition, Project Administration, Resources, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Funding provided by the V.V. Cooke Foundation (Kentucky, U.S.A.); University of Louisville 21st Century University Initiative on Big Data in Medicine (Z1762); National Institutes of Health (NIH; P20GM103436); and the startup funding from the Mansbach Family, the Gheens Foundation and other supporters at the University of Louisville. Its contents are solely the responsibility of the authors and do not represent the official views of the funding organization.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2018 Weirick T et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


Introduction

RNA editing is a class of epitranscriptomic post-transcriptional modification found throughout metazoan consisting of the abundant conversion of adenosine-to-inosine (A-to-I) by ADARs (adenosine deaminases acting on RNA) and rare conversion of cytosine-to-uridine (C-to-U) by APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like). RNA editing is particularly interesting as it is detectable as A-to-G and C-to-T mismatches to the reference genome within standard RNA-sequencing data via specialized computational pipelines. An increasing number of studies link changes in editing at specific sites or clusters-of-sites to diseases, such as epilepsy and atherosclerosis. Yet, no software for detecting differential editing is available. To meet this need, we present DRETools: 1) to calculate units that help reduce sample-bias, similar to FPKM for RNA expression; and 2) to find differentially edited sites and editing islands (i.e., clusters of editing sites). Further, we showcase two examples of finding differential editing and related tasks with DREtools.

Methods

Implementation

DRETools can be run via command-line by typing “dretools”, which will print the main help menu. The main help menu contains a list of operations that are available from dretools with short descriptions of each operation’s purpose. To run an operation, type dretools followed by the operation name. Further detail on each operation, including available command-line arguments and usage examples, can be found by running an operation with the --help argument. On the main help menu, operations are organized into sub-headings based on similar functions. Further detail of each sub-heading and corresponding operations can be found in the following sections.

Units

One fundamental problem of between groups of samples is a lack of standardized units for describing editing within samples, editing islands, and sites. To this end, DRETools implements Editing Per Kilobase (EPK) based on “overall editing” (OE). EPK builds upon OE by considering both A-to-G and C-to-T transitions, excludes editing sites with 100% edited bases as potential mutations, and scaling by 10^4 for readability (similar to FPKM). EPK is calculated by dividing the total number of “edited” bases by the total number of bases overlapping known editing sites and multiplying by 10^4. In addition to samples, DRETools can compute EPKs for editing islands and sites. Sample-wise editing can be computed with the “sample-epk” function and can be thought of as the global-editing-rate, whereas, the EPK of islands and sites can be computed with “region-epk” and “edsite-epk” respectively, and thought of as the “local-editing-intensity”.

Differential editing

Recently, a method was developed to find differentially edited sites between epileptic or control mouse hippocampi. However, methods capable of comparing different tissues are also needed. The problem is that unless the global-editing-rates are similar, we cannot determine if changes are due to differing global-editing-rates or other phenomena, such as competition with N6-methyladenosine (m6A). Furthermore, ADARs have been described to edit both specific sites in some cases and nonspecifically within small regions in other cases. Therefore, in addition to individual editing sites, looking at the clusters of editing is also of interest. DRETools addresses both these issues by allowing the normalization of both the global-editing-rate and site or island local-editing-intensity in EPK and testing for differential editing using a linear model (LM) with the formula: “logFeatureEPK ~ logSampleEPK + featureLength + averageReadDepth” (features can be sites or islands), which adjusts expectations for what constitutes differential editing.

Merge and stats

DRETools also includes various helper functions. For example, the merge section contains functions to find editing islands and create consensus sets of editing sites by merging sites from multiple samples. Finally, the stats heading contains functions that calculate useful information about editing at the sample, gene, and site levels, such as the editable area or the number of editing sites falling in 3'UTR-untranslated regions, introns, or exons.

Operation

A standard laptop computer with the latest version of R and Python3 will handle most applications.

Results

To illustrate the utility of DRETools, we surveyed differential editing in human umbilical vein endothelial cells (HUVEC) transfected with either a siRNA against ADAR1 or against a random sequence (control) and the immortalized cell lines GM12787 and K562. First we surveyed sample-wise editing using the function “sample-epk.” (Figure 1A, B). Using EPK reduces variation within groups compared to the usage of number of editing sites. For example, the coefficient of variance drops from 0.21 to 0.05 for the silenced ADAR1 group and 0.52 to 0.01 for the control group. Similarly, when comparing the immortalized cell lines, the coefficient of variance is reduced from 0.57 to 0.25 and 0.46 to 0.11, respectively (Figure 1C, D).

Next, we compared the EPKs of editing islands within the immortalized cell lines using “epk-region”. Using EPK to represent editing islands as opposed to the number of edited bases reduces the coefficient of variance from 0.60 ± 0.21 to 0.31 ± 0.11 (p=2E-30). Finally, we tested for differential editing using the functions “region-diff” for islands and “site-diff” for editing sites (Figure 1E–H). Comparing silenced ADAR1 to the control, the LM yielded a uniform distribution of p-values. In contrast, when using t-test applied to the same data, the distribution of p-values is shifted to the left and exhibits greater skew. However, in the immortalized cell lines, p-values calculated by the LM are more leftward skewed while p-values from the t-test became more uniformly distributed. This provides evidence that the LM can effectively reduce type I errors when testing for differential editing. For example, the LM correctly recognizes that most of the differences between the silenced ADAR1 and control groups arise from the reduction of the global-editing-rate.
in the silenced samples. Whereas the t-test, which does not consider the global-editing-rates, finds many differentially edited sites and islands. Conversely, when comparing the immortalized cell lines, despite the large difference in EPK, many differentially edited sites and islands are detected. While deeper biological validation is needed to be certain, these could be instances of some other phenomena, such as m6A, affecting the editing in individual sites or islands.

**Conclusions**

DRETools is a command-line tool suite for finding differentially edited sites and islands. It allows users to calculate units that reduce sample-bias and find differentially edited sites and islands even when the global-editing-rate of groups being compared is different. Furthermore, it also includes a variety of other features for exploring RNA editing. These make DRETools a valuable tool for further investigating epitranscriptomics.

**Data availability**

All RNA-seq data are publically available and were downloaded from the NCBI SRA database. The HUVEC data sets were generated by Stellos et al., 2016 and the GM12787 and K562 cells by the ENCODE project. Lists of accession numbers, pipelines used to generate analyses, and intermediate files generated are archived on Zenodo.

**Software availability**

Source code available from: http://dretools.bitbucket.io/.

Data and analysis pipelines: https://zenodo.org/record/1400648.

Source code at time of publication: https://zenodo.org/record/1400005.

License: The software, and data and analysis pipelines are available under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

**Author contributions**

TW: Conception, Analysis, Investigation, Methodology, Project Administration, Software, Validation, Visualization, Writing - Original Draft Preparation, Writing - Review & Editing

PT: Conception, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation

AD: Supervision, Writing – Review & Editing

ER: Funding Acquisition, Writing – Review & Editing

Figure 1. (A) The number of editing sites in HUVEC control and silenced ADAR1 groups (p=0.77). NS, p>0.05. (B) HUVEC control and silenced ADAR1 (siADAR1) represented in EPK (p=0.78E-5). **p<0.0001. (C) The number of editing sites detected in GM12787 and K562 cells (p=1.2E-3). *p<0.05. (D) Editing in GM12787 and K562 cells represented in EPK (p=2.5E-6). **p<0.0001E-4. (E-H) Histograms detailing the distribution of p-values when testing for differential editing in a site- or island-wise manner. The site-wise comparison between: (E) siADAR1 and control; and (F) GM12787 and K562 cells. The island-wise comparison between: (G) siADAR1 and control; and (H) GM12787 and K562 cells.
Grant information
Funding provided by the V.V. Cooke Foundation (Kentucky, U.S.A.); University of Louisville 21st Century University Initiative on Big Data in Medicine (Z1762); National Institutes of Health (NIH; P20GM103436); and the startup funding from the Mansbach Family, the Gheens Foundation and other supporters at the University of Louisville. Its contents are solely the responsibility of the authors and do not represent the official views of the funding organization. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments
This work utilized the University of Louisville Cardinal Research Cluster.

References

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com