DRETools: A tool-suite for differential RNA editing detection [version 2; peer review: 1 approved, 1 approved with reservations, 1 not approved]

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

This article is included in the International Society for Computational Biology Community Journal gateway.
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 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Weirick T, Trainor P, Rouchka E et al. DRETools: A tool-suite for differential RNA editing detection [version 2; peer review: 1 approved, 1 approved with reservations, 1 not approved] F1000Research 2018, 7:1366
https://doi.org/10.12688/f1000research.16026.2
**Amendments from Version 1**

In this revised manuscript, a brief description of RNAEditor was added in the Methods section. Furthermore, the required hardware configuration for running DRETools, along with run times when analyzing each testing sample, were added.

See referee reports

---

**Introduction**

RNA editing is a class of epitranscriptomic post-transcriptional modification found throughout metazoa 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.

DRETools requires the output from RNA detection software. Here we used RNAEditor to detect editing sites in standard RNA-seq data. RNAEditor uses a specialized alignment and variant calling pipeline to find potential editing sites and then uses filters to remove false positives. In addition to RNAEditor, there are a number of other editing detection tools available. DRETools is usable with any of these tools that work in a similar manner and produces a VCF file containing editing sites and a BAM file containing aligned reads. Details regarding the usage of RNAEditor (e.g. analysis pipelines, configuration files, and non-downloadable reference files) can be found in the archived data.

**Units**

One fundamental problem 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 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. 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'/5'-untranslated regions, introns, or exons.

**Operation**

Minimum requirements for DRETools are 8 gigabytes (GB) of RAM, a 100 GB hard drive, and an operating system with a Bash command-line interface, R version 3.3+, and Python version 3.5+. The first two operations required on average (n=5) 4.2 minutes (min) and 449 megabytes of RAM memory (MB) for edssite-merge and 4.4 min and 192 MB for find-islands. The benchmarks of remaining operations are primarily dependent on the BAM files used for computation. The BAM files used
here ranged from 3.2-29 GB and 31-282 million reads. Performance was as follows: sample-epk (6-42 min, 40-54 MB), edsitedepk (6-41 min, 40-310 MB), region-epk (7-35 min, 40-323 MB), edsitediff (0.41-3.49 min and 534.09-2280 MB), and region-diff (0.05-0.23 min, 190.31-522 MB).

Results
To illustrate the utility of DRETools, we surveyed differential editing in human umbilical vein endothelial cells (HUVEC) transfected with either an 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.

![Figure 1](http://example.com/fig1.png)

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<0.00011E-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.
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
SU: Conceptualization, Funding Acquisition, Writing – Review & Editing
ER: Funding Acquisition, Writing – Original Draft Preparation, Writing – Review & Editing

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

F1000Research 2018, 7:1366 Last updated: 30 JUL 2020
Open Peer Review

Current Peer Review Status:  ❓ ✓ ✗

Version 2

Reviewer Report 04 October 2018

https://doi.org/10.5256/f1000research.17882.r38801

© 2018 Pesole G. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Graziano Pesole

Department of Biosciences, Biotechnology and Biopharmaceutics, University of Bari Aldo Moro (UBAM), Bari, Italy

This manuscript represents a valuable contribution addressing a timely and relevant problem in epitranscriptomics analysis. Indeed, several RNA editing analysis studies have been reported so far in the literature but reliable standard operative procedures for evaluating "differential editing" are still missing. However, the methodology presented raises some relevant concerns that need to be adequately addressed:

1) A basic question... how to label a site as edited? An editing level of 1% is enough? But this can be detected with a very high coverage. In other words different coverage levels in different samples may greatly affect the estimate of the number of editing sites.

2) The differential analysis is based on the EPK parameter which is related to the number of observed editing sites (regions). I have serious concerns on the metrics used in consideration that the extent of editing is variable and its detection somehow dependent on the coverage. For example, lets say we have 1000 shared editing sites, with editing level at 10% in one sample group and 80% in the other. If I understood correctly, the metric used considers only the number of editing sites and not the editing level. So, in this case any differential editing is detected. Please clarify.

3) In order to carry pout a reliable comparison only inter-samples common sites should be considered... If different set of editing sites are compared in different samples I believe that this may introduce a serious bias affecting the results. The Authors should address this issue.

4) I am not convinced that 100% edited sites are to be excluded. Many of these are validated through comparison with the relevant genomic sequences.

5) The description of metrics for the island/cluster differential editing is not clear.

Is the rationale for developing the new software tool clearly explained?
Partly
Is the description of the software tool technically sound?
Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
No

**Competing Interests:** No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

---

**Version 1**

Reviewer Report 13 September 2018

https://doi.org/10.5256/f1000research.17503.r37750

© 2018 Zhao Y. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Yicheng Zhao
Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston (UTHealth), Houston, TX, USA

Currently, more research has focused on human RNA editing. This software is very useful in detecting human differentially edited sites and islands. However, I suggest the authors to add some brief description about RNAEditor in the method section, which will be helpful for understanding how to handle and analyse RNA editing via RNA seq data. Besides, the authors should provide the required hardware configuration for running DREtools, and the run times when analysing each testing sample.

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Non coding RNA function, RNA editing and related bioinformatics tech

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 17 Sep 2018

Shizuka Uchida, University of Louisville, Louisville, USA

Thank you very much for your valuable comments. We have now added a brief description of RNAEditor in the Method section. Furthermore, we have included the required hardware configuration for running DRETools along with run times when analyzing each testing sample.

Competing Interests: None

Reviewer Report 13 September 2018

https://doi.org/10.5256/f1000research.17503.r37753

© 2018 Picardi E. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ernesto Picardi

1 University of Bari, Bari, Italy
2 Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, National Research Council (CNR), Bari, Italy

The manuscript by Weirick et al. introduces a tool-suite to calculate differential RNA editing.

The interest towards RNA editing is rapidly growing and, thus, similar tools to improve the investigation of
RNA editing in different experimental conditions are demanding.

DRETools include some functions to mainly post-process results from RNAEditor, developed in the same research group. Even though they can be applied to results from other tools after an ad hoc parsing.

Calculations are based on the definition of EPK (editing per kilobase), in turn, based on the overall editing concept introduced by Tan et al. 2017. The overall editing is simply calculated as the total number of reads with G at all known editing positions as compared to all reads covering the position. In other terms, this metric is a global editing frequency per sample.

Authors multiply the overall editing by 1000 in order to improve the readability because in same cases very low numbers may appear. Although authors show that EPK values are useful over the raw count of editing sites, the properties of EPK are not well investigated. The number of As andGs is dependent on filters used to detect editing and the quantity of reads generated by sequencing. Base quality is also an additional factor to consider. I'm not completely sure if EPK can take into account the number of reads per sample. I suggest to perform further investigations calculating global EPK in samples belonging to the same tissue. For example, authors could use GTEx RNAseq from three or four tissues and at least 10 experiments per tissue.

Other authors proposed similar indices to detect editing activity in a sample. For example, Paz-Yaacov introduced the Alu editing index, a robust measure that can be calculated also on other additional genomic properties (recoding sites, conserved sites and so on). This index has been successfully used in several cases and authors need to perform appropriate comparisons.

Regarding statistical tests used to detect differential editing, authors implement a linear model and the t-test. In figure 1 (panels E to H), p-values distributions are shown and great differences seem to appear. Authors should discuss the reason why of these observed discrepancies. I suggest authors to check the use of non parametric tests since they could be robust in case of small samples or when users cannot easily establish if normality and other assumptions are encountered.

Additionally, the tool does not take into account the correction for multiple testing. So it needs to be implemented.

In humans, RNA editing has different properties depending on affected genomic regions. For example, Alu editing is different from recoding editing. Is there a way to take such properties into account?

Finally, DRETools features should better described in the manuscript and details about input and output files should be included in the wiki pages.

Some experimental validations are required to corroborate tool results.

In my opinion DRETools are useful but major improvement is deeply needed.

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Partly
Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Partly

**Competing Interests:** No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

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