Bioinformatics, 31(15), 2015, 2568–2570
doi: 10.1093/bioinformatics/btv187
Advance Access Publication Date: 2 April 2015
Applications Note

OXFORD

Genome analysis

DREAM: a webserver for the identification of editing sites in mature miRNAs using deep sequencing data

Shahar Alon[1,2], Muhammad Erew[3] and Eli Eisenberg[2,3,*]

[1]Department of Neurobiology, George S. Wise Faculty of Life Sciences, [2]Sagol School of Neuroscience and [3]Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel

*To whom correspondence should be addressed.

Associate Editor: Ivo Hofacker

Received on November 18, 2014; revised on March 15, 2015; accepted on March 29, 2015

Abstract

Summary: DREAM: **d**etecting **R**NA **e**diting **a**ssociated with **m**icroRNAs, is a webserver for the identification of mature microRNA editing events using deep sequencing data. Raw microRNA sequencing reads can be provided as input, the reads are aligned against the genome and custom scripts process the data, search for potential editing sites and assess the statistical significance of the findings. The output is a text file with the location and the statistical description of all the putative editing sites detected.

Availability and implementation: DREAM is freely available on the web at http://www.cs.tau.ac.il/~mirnaed/.

Contact: elieis@post.tau.ac.il

1 Introduction

Deep sequencing has many possible applications; one of them is the identification and quantification of RNA editing sites. The most common type of RNA editing is adenosine to inosine (A-to-I) editing, catalyzed by enzymes of the adenosine deaminase that act on RNA family (Nishikura, 2010). A prerequisite for this editing process is a double-stranded RNA (dsRNA) structure. Such dsRNAs are formed as part of the microRNA (miRNA) maturation process, and it is therefore expected that miRNAs are affected by A-to-I editing. Indeed, dozens of editing sites were found in miRNAs: A-to-I editing of primary miRNA (pri-miRNA) was shown to affect the processing of the pri-miRNA to precursor miRNA (pre-miRNA) or the processing of the pre-miRNA to mature miRNA (Chawla and Sokol, 2014; Yang *et al.*, 2006). Pri-miRNA and pre-miRNA editing events may lead to an edited version of the mature miRNAs. Mature miRNAs can bind to partially complementary targets located in the 3′ untranslated region of specific mRNAs and block the translation or trigger degradation of the target mRNAs (Bartel, 2004). Editing at the 'seed' region, bases 2–8 at the 5′ end of the mature miRNA that are critical for target recognition, is expected to change the set of target genes. A striking example is mouse miR-376 in which editing in the recognition site alters the target specificity of the miRNA and profoundly affects cellular processes (Kawahara *et al.*, 2007). Deep sequencing data analysis may unravel the full extent of A-to-I editing in miRNAs as well as other types of RNA editing. However, technical problems in the analysis of deep sequencing datasets may result in a false detection of miRNA modification events (de Hoon *et al.*, 2010). We have previously described a protocol which utilizes deep sequencing data to characterize RNA editing in mature miRNAs from human and mouse brain while avoiding technical data analysis problems (Alon and Eisenberg, 2013; Alon *et al.*, 2012). Other groups have also detected miRNA editing events in the mouse brain during development, in *Drosophila* and in *C. elegans*, demonstrating that editing can add complexity to the miRNA gene regulation pathway (Berezikov *et al.*, 2011; Ekdahl *et al.*, 2012; Vesely *et al.*, 2012; Warf *et al.*, 2012). However, so far there is no tool allowing researchers to detect mature miRNA editing sites without extensive use of data analysis procedures. The DREAM webserver is expected to make the editing detection procedure accessible to the large

community of researchers studying miRNA and RNA modifications processes.

2 Functionality

DREAM has been developed in Perl programming language and using hypertext preprocessor (PHP) scripting. The required input is sequencing reads of mature miRNAs in a standard Fastq format. The webserver supports only human and mouse data obtained using Illumina platform. Careful reads alignment is crucial for correct identification of editing sites (Alon *et al.*, 2012; de Hoon *et al.*, 2010), and thus the webserver does not support previously aligned sequencing reads in SAM or BAM format. The preferred method to supply the sequencing data to the webserver is by uploading it to a public directory in a Dropbox account and providing the direct link to the webserver. Using the direct link, the following compressed file formats are allowed: zip, tar, tar.gz, tar.bz2 and rar. Fastq file, either compressed or not, can be also uploaded directly to the webserver, but this option is not recommended as upload times can be quite long for large files (typically ~5 GB/h) and network interruptions may prevent the completion of the upload. The user may follow the upload progress using a progress bar at the bottom of the webserver page, and receives a notification email upon upload completion, including an estimate on the process time, based on the size of the data and the files in queue. Importantly, the webserver supports uploading and processing of several runs in parallel. Therefore, the user must use a unique file name for every uploaded file to avoid conflicting runs. To get a grasp of the usability of the webserver and the expected results, a test dataset can also be used. The test dataset is of mature miRNAs reads from mouse cerebellum (Short Reads Archive accession number SRR346417). For the test dataset the webserver parameters are automatically adjusted.

The Fastq file may include either unfiltered raw sequencing reads, likely to contain parts of the adapter sequences, or reads that were already filtered using Illumina software tools. In the former case, the webserver can trim the adapters if the user supplies their sequences. As the expected length of mature miRNAs is ~21 bases, reads longer than 28 bases or shorter than 15 bases (after trimming) are discarded (Alon *et al.*, 2012). Low-quality reads (as defined by the read quality score) are unlikely to be informative. Therefore reads with low sequencing quality (Phred-score <20) in more than three positions are discarded (Alon *et al.*, 2012). This pre-process step takes place only if the user chooses the pre-processing option.

The filtered reads are aligned against the genome of interest using Bowtie (Langmead *et al.*, 2009). The UCSC hg19 and mm9 versions of the genome are used for human and mouse, respectively. The webserver require unique best alignment, i.e. reads that do not align to other locations in the genome with the same number of mismatches. Only alignments with up to one mismatch are used. These steps are taken to solve the cross mapping problem that significantly hinders identification of true editing sites in mature miRNAs (Alon *et al.*, 2012; de Hoon *et al.*, 2010). Note that this procedure might fail to detect some mature miRNAs with many edited positions. The last two bases (the 3′ end) of mature miRNA undergo extensive adenylation and uridylation (Burroughs *et al.*, 2010), and are therefore excluded (thus, editing events in these two bases cannot be detected). This measure is necessary to ensure that sufficient number of reads map with a single mismatch at most. Next, the reads aligned to the genome are converted to counts of each of the four possible nucleotides at each position along the pre-miRNA sequence (miRBase, release 21; Kozomara and Griffiths-Jones, 2011; http://www.mirbase.org/), for all the pre-miRNAs. Performing this transformation allows focusing the analysis on *bona fide* miRNA only. In the following step, binomial statistics is applied to distinguish significant modifications in these regions from sequencing errors. The binomial statistics requires the number of mismatches of a given type in a given position, the number of total reads in the given position, and the a priori sequencing error probability. As only mismatches with Phred score of 30 are allowed (Alon *et al.*, 2012), we use 0.1% as the expected base call error rate. Importantly, binomial statistics do not require any arbitrary expression level filter. The search for possible RNA editing sites is performed for every position in every mature miRNA separately. As multiple tests are performed, the resulting *P*-value for each position must be corrected accordingly. The type of the multiple testing correction (either 'Bonferroni' or 'Benjamini–Hochberg') and the required *P*-value (or False Discovery Rate) cutoff is chosen by the user. Upon completion, a second email is sent to the user, containing the results in text format. Spreadsheet software, like Microsoft Excel, can be used to open the text file. For each significant modifications detected, the output text file gives its type and the location in the genome and in the miRNA (both inside the pre-miRNA sequence and the mature miRNA sequence) according to miRBase definitions (release 21). In addition, descriptive information is given: the number of reads which show the modified base and the total number of reads in this position. Lastly, the *P*-values calculated using the binomial test are given, as well as multiple testing adjusted *P*-values. Most of the detected modifications are expected to be A-to-G (Alon *et al.*, 2012), which can be a result of A-to-I editing as the sequencing process identifies inosines as guanosines. Known SNPs are automatically removed during the analysis (dbSNP, builds 138 and 142 for mouse and human, respectively). However, the detected modifications can also be due to previously undetected or rare SNPs, somatic mutations, other RNA modifications processes or technical artifacts (Li *et al.*, 2011; Lin *et al.*, 2012), and therefore post-processing is recommended (Alon and Eisenberg, 2015).

Acknowledgements

The authors thank Dr. Gianni Nigita for helpful suggestions.

Funding

This work was partially supported by the Israel Science Foundation [379/12] and the United States–Israel Binational Science Foundation (2009/290), Jerusalem, Israel, to E.E. S.A. was supported by a Clore Fellowship.

Conflict of Interest: none declared.

References

Alon,S. and Eisenberg,E. (2013) Identifying RNA editing sites in miRNAs by deep sequencing. *Methods Mol. Biol.,* **1038**, 159–170.

Alon,S. and Eisenberg,E. (2015) Using deep sequencing data for identification of editing sites in mature miRNAs. *Methods Mol. Biol.,* **1269**, 231–242

Alon,S. *et al.* (2012) Systematic identification of edited microRNAs in the human brain. *Genome Res.*, **22**, 1533–1540.

Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell,* **116**, 281–297.

Berezikov,E. *et al.* (2011) Deep annotation of Drosophila melanogaster microRNAs yields insights into their processing, modification, and emergence. *Genome Res.*, **21**, 203–215.

Burroughs,A.M. *et al.* (2010) A comprehensive survey of 3′ animal miRNA modification events and a possible role for 3′ adenylation in modulating miRNA targeting effectiveness. *Genome Res.*, **20**, 1398–1410.

Chawla,G. and Sokol,N.S. (2014) ADAR mediates differential expression of polycistronic microRNAs. *Nucleic Acids Res.,* **42**, 5245–5255.

de Hoon,M.J.L. *et al.* (2010) Cross-mapping and the identification of editing sites in mature microRNAs in high-throughput sequencing libraries. *Genome Res.,* **20**, 257–264.

Ekdahl,Y. *et al.* (2012) A-to-I editing of microRNAs in the mammalian brain increases during development. *Genome Res.,* **22**, 1477–1487.

Kawahara,Y. *et al.* (2007) Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science,* **315**, 1137–1140.

Kozomara,A. and Griffiths-Jones,S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *NAR,* **39**, D152–D157.

Langmead,B. *et al.* (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.,* **10**, R25.

Li,M. *et al.* (2011) Widespread RNA and DNA sequence differences in the human transcriptome. *Science,* **333**, 53–58.

Lin,W. *et al.* (2012) Comment on "Widespread RNA and DNA sequence differences in the human transcriptome." *Science,* **335**, 1302.

Nishikura,K. (2010) Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.,* **79**, 321–349.

Vesely,C. *et al.* (2012) Adenosine deaminases that act on RNA induce reproducible changes in abundance and sequence of embryonic miRNAs. *Genome Res.,* **22**, 1468–1476.

Warf,M.B. *et al.* (2012) Effects of ADARs on small RNA processing pathways in *C. elegans Genome Res.,* **22**, 1488–1498.

Yang,W. *et al.* (2006) Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat. Struct. Mol. Biol.,* **13**, 13–21.