Bioinformatics, 2016, 1–2
doi: 10.1093/bioinformatics/btw479
Advance Access Publication Date: 10 July 2016
Applications Note

OXFORD

Systems biology

MIMEAnTo: profiling functional RNA in mutational interference mapping experiments

Maureen R. Smith[1,*], Redmond P. Smyth[2], Roland Marquet[2] and Max von Kleist[1,*]

[1]Systems Pharmacology & Disease Control, Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany and [2]Architecture et Réactivité de l'ARN, Institut de Biologie Moléculaire et Cellulaire du Centre National de la Recherche Scientifique, Université de Strasbourg, Strasbourg, France

*To whom correspondence should be addressed.

Associate Editor: Bonnie Berger

Received on April 29, 2016; revised on June 22, 2016; accepted on July 6, 2016

Abstract

Summary: The mutational interference mapping experiment (MIME) is a powerful method that, coupled to a bioinformatics analysis pipeline, allows the identification of domains and structures in RNA that are important for its function. In MIME, target RNAs are randomly mutated, selected by function, physically separated and sequenced using next-generation sequencing (NGS). Quantitative effects of each mutation at each position in the RNA can be recovered with statistical certainty using the herein developed user-friendly, cross-platform software MIMEAnTo (MIME Analysis Tool).

Availability and implementation: MIMEAnTo is implemented in C++ using the boost library as well as Qt for the graphical user interface and is distributed under GPL (http://www.gnu.org/licences/gpl). The libraries are statically linked in a stand alone executable and are not required on the system. The plots are generated with gnuplot. Gnuplot-iostream (https://github.com/dstahlke/gnuplot-iostream) serves as gnuplot interface. Standalone executables including examples and source code can be downloaded from https://github.com/maureensmith/MIMEAnTo.

Contacts: msmith@zedat.fu-berlin.de or vkleist@zedat.fu-berlin.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

RNA has long been believed to mainly serve as a blueprint for proteins. However, a large diversity of so called noncoding RNAs have been discovered lately to regulate virtually all cellular processes (Wan *et al.*, 2011). Characterizing functional domains and structure–function–relationships in RNA is a major challenge as classical experimental approaches are time-consuming and require substantial investigator expertise. To overcome these obstacles we recently developed MIME (Smyth *et al.*, 2015), where RNA is randomly mutated, selected-by-function (e.g. binding to a protein), physically separated and sequenced using NGS. Consequently, quantitative effects of mutations at every nucleotide position can be recovered after a single experiment.

The mutation frequencies (recovered from NGS) in the functionally selected vs. non-selected pools contain information about the function and structural commitment of each nucleotide within the analyzed RNA: i.e. a mutation that does not affect selection is not required for the function of the RNA. Since NGS-reads R from the MIME experiment are confounded by errors introduced during library preparation and sequencing (see Fig. 1, upper panel), error correction and statistical assessment are essential. The following equation accommodates these considerations:

$$\frac{\mathrm{Kd}_{m,w}}{\mathrm{Kd}_{w,w}}(i,j) \approx \frac{\frac{R_{m,w}(i,j,\mathrm{unbound})}{R_{w,w}(i,j,\mathrm{unbound})} - \mathbb{E}_J(\kappa_{w\to m}(i))}{\frac{R_{m,w}(i,j,\mathrm{bound})}{R_{w,w}(i,j,\mathrm{bound})} - \mathbb{E}_J(\kappa_{w\to m}(i))},$$

allowing the relative dissociation constant of mutation m at nucleotide position i, $\mathrm{Kd}_m/\mathrm{Kd}_w(i)$, to be re-computed in a jackknife-like fashion for each position $j \in J\backslash i$. Above, $R_{m,w}(i,j)$ denotes the number of reads that carry a mutation m at position i and a wild

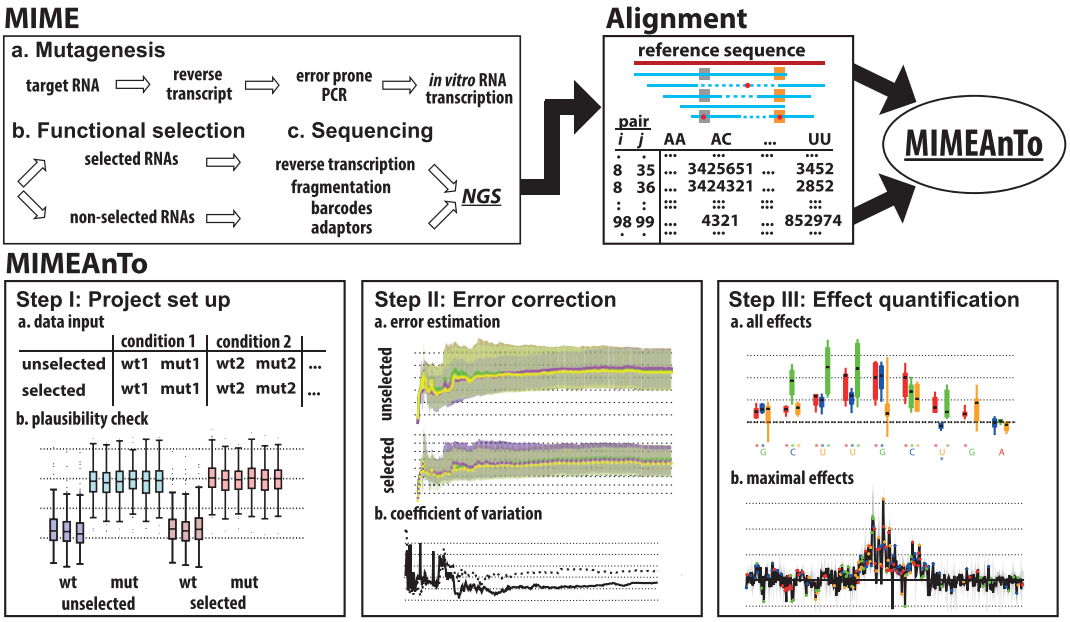

Fig. 1. The MIMEAnTo software identifies functional domains in RNA from MIME data. **Top,** left: In MIME, a target RNA is (**a**) randomly mutated, (**b**) subjected to selection and physically separated into selected and non-selected fractions. (**c**) Both pools are sequenced using NGS. Right: Sequence reads are aligned to the reference RNA. Base counts for each pair of positions (*i, j*) serve as input for MIMEAnTo. **Bottom,** left: Raw mutation frequencies are depicted allowing for plausibility checks. Middle: Inter-experimental differences in position- and mutation specific error estimates are explored. Right: Quality criteria can be entered and quantification of effects, including statistical ascertainment is provided, as well as output functions

type residue at position $j \neq i$ and $\mathbb{E}_J(\kappa_{w\to m}(i))$ denotes the expected probability of falsely detecting a wild type at position i as a mutant m, which is used for error-correction.

In this note, we present the user-friendly cross-platform software MIMEAnTo, which evaluates MIME data, following error-correction, statistical ascertainment and quality assessment. Mathematical foundations and a user-manual are provided as Supplementary Texts.

MIMEAnTo has a wizard-like graphical user interface, guiding through the data analysis procedure in three steps: (i) data input and assessment, (ii) sequencing and reverse transcription error correction and (iii) quantification of raw effects and quality filtering. Note because RNA is structurally flexible, our method is unique as it allows to characterize *functionally relevant* structure(s) of the RNA, which can inform fundamental research, as well as drug design.

2 Usage

Input: MIMEAnTo takes plain text files with absolute nucleotide occurrences in the selected- and non-selected samples, captured with different experimental conditions, as input. Two types of files are required: (i) simple input tables count the number of A, C, G and U at each position and are used for plausibility checks (Fig. 1, bottom left). (ii) extended input tables count all non-redundant combinations of two positions (*i, j*) filling an array containing the number of dinucleotide occurrences for each pair of positions (number of AA, AC,. . .UG and UU). The latter is used for error correction, for inferring the effect of mutations on protein binding and statistical ascertainment. A program for parsing SAM files is provided on the project webpage. A first plausibility check of the data (step I) showing the gross mutation rates of each sample provides the opportunity to interactively validate the data (Fig. 1, bottom left).

Error correction: Sequencing and reverse transcription can produce errors. In particular, the actual mutation frequencies in the samples need to exceed these errors (signal-to-noise ratio). The MIME workflow incorporates control experiments with wild type libraries allowing the quantification of this error for each position and mutation. The errors are explored in step II of the pipeline (Fig. 1, bottom center).

Quantification, filtering & output: In step III raw error-corrected effects can be quantified using eq. (1). Simultaneously, for each re-sampled effect at nucleotide positions (*i, j*), a signal-to-noise ratio $D_{m,w}(i,j)$ is derived. This step is only performed once (runtime $\mathcal{O}(L^2)$), where L denotes the length of the functional RNA. Next, the user can provide several quality criteria for filtering the results: (i) a minimum signal-to-noise ratio $D_{m,w}(i,j)$ and (ii) two criteria ensuring a meaningful re-sampling. An absolute minimum coverage and a threshold that ensures that the coverage for each pair of positions (*i, j*) exceeds a given percentage of the total coverage at position *i*. Guidelines for choosing the exact values are provided in the Supplementary Text.

The results are then filtered and publication-ready graphics generated showing (i) the $\frac{\mathrm{Kd}_{m,w}}{\mathrm{Kd}_{w,w}}(i,*)$ distribution of each individual mutation m at each position i and (ii) the $\frac{\mathrm{Kd}_{m_{\max},w}}{\mathrm{Kd}_{w,w}}(i,*)$ distribution for only the maximum effect mutation $m_{\max}$ with significant effects highlighted (Fig. 1, bottom right). Results and graphics can be exported in csv or eps and pdf format respectively.

Funding

MvK received funding through BMBF grant 031A307 and RM through ANR grant ANR-15-CE11-0014-01 'PSiFlu'.

Conflict of Interest: none declared.

References

Smyth,R.P. *et al.* (2015) Mutational interference mapping experiment (MIME) for studying RNA structure and function. *Nat. Methods*, **12**, 866–872.

Wan,Y. *et al.* (2011) Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.*, **12**, 641–655.