RNA Framework: an all-in-one toolkit for the analysis of RNA structures and post-transcriptional modifications

Danny Incarnato[1,2,*], **Edoardo Morandi**[1,2], **Lisa Marie Simon**[1,2] **and Salvatore Oliviero**[1,2,*]

[1]Italian Institute for Genomic Medicine (IIGM), Via Nizza 52, 10126 Torino, Italy and [2]Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università di Torino, Via Accademia Albertina 13, Torino, Italy

Received March 17, 2018; Revised May 15, 2018; Editorial Decision May 17, 2018; Accepted May 23, 2018

ABSTRACT

RNA is emerging as a key regulator of a plethora of biological processes. While its study has remained elusive for decades, the recent advent of high-throughput sequencing technologies provided the unique opportunity to develop novel techniques for the study of RNA structure and post-transcriptional modifications. Nonetheless, most of the required downstream bioinformatics analyses steps are not easily reproducible, thus making the application of these techniques a prerogative of few laboratories. Here we introduce RNA Framework, an all-in-one toolkit for the analysis of most NGS-based RNA structure probing and post-transcriptional modification mapping experiments. To prove the extreme versatility of RNA Framework, we applied it to both an in-house generated DMS-MaPseq dataset, and to a series of literature available experiments. Notably, when starting from publicly available datasets, our software easily allows replicating authors' findings. Collectively, RNA Framework provides the most complete and versatile toolkit to date for a rapid and streamlined analysis of the *RNA epistructurome*. RNA Framework is available for download at: http://www.rnaframework.com.

INTRODUCTION

The advent of Next-Generation Sequencing (NGS) made possible the development of several novel techniques for the comprehensive characterization of the *RNA epistructurome* (1), a term we recently coined to collectively define both RNA structure and the set of post-transcriptional modifications of the transcriptome of an organism.

RNA footprinting experiments, aimed to interrogate the secondary/tertiary structure of RNA molecules, make use of either chemical reagents or nucleases that are able to specifically modify/cleave either single- or double-stranded RNA residues. Most high-throughput RNA footprinting methods, such as PARS, SHAPE-seq, DMS-seq, Structure-seq, CIRS-seq, icSHAPE and SPET-seq (2–9), are based on the detection of reverse transcriptase (RT) drop-off due to a nuclease cut (e.g. S1 Nuclease, cutting unpaired RNA residues) or to the presence of chemicals-mediated adducts on the free 2′-OH of the ribose moiety (e.g. SHAPE reagents, modifying structurally flexible residues), or due to modifications on the Watson-Crick face of nucleobases (e.g. Dimethyl sulfate, DMS, alkylating unpaired A/C residues). More recently, mutational profiling approaches (MaP) have emerged, such as SHAPE-MaP, RING-MaP and DMS-MaPseq (10–12). These methods are based on the use of special RT enzymes (e.g. TGIRT-III (10)) or reverse transcription conditions (11,12) enabling read-through at sites of adduct formation/modification, that are then recorded as mutations in the resulting cDNA molecule. Readout of these experiments is a reactivity profile that can be incorporated by RNA structure prediction software, such as ViennaRNA (13) or RNAstructure (14), in the form of *constraints* (or *restraints*) to guide *in silico* folding of the RNA molecule. These constraints are usually converted into pseudo-free energy terms that are then used to adjust the free energy contribution of each base-pair. Such incorporation of experimentally-derived footprinting data has been proven to greatly increase the accuracy of prediction algorithms (6,15–17).

RNA post-transcriptional mapping experiments either exploit specific chemical properties of modified nucleobases in order to achieve single-base resolution mapping, or otherwise make use of specific antibodies to selectively immunoprecipitate (IP) RNA fragments bearing the modified residues. Only a handful of the over 100 RNA post-transcriptional modifications known to date (18) have been mapped. Nucleotide methylations (and hydroxymethylations) such as m^6A, m^1A, m^5C and hm^5C can be read-

[*]To whom correspondence should be addressed. Tel: +39 011 670 9533; Email: salvatore.oliviero@unito.it
Correspondence may also be addressed to Danny Incarnato. Tel: +39 011 670 9531; Email: danny.incarnato@iigm.it

ily mapped by immunoprecipitation (19–23), an approach generically known as MeRIP-seq (methylated RNA immunoprecipitation). Beside IP, m^5C can also be detected by bisulfite sequencing (24,25), a single-base resolution approach largely used to map m^5C in genomic DNA, although it cannot discriminate hm^5C residues. To date, only a few other key modifications can be mapped by means of single-base resolution methods. Pseudouridine (Ψ) mapping by either Ψ-seq or Pseudo-seq (26,27), is based on the RT drop-off induced by the alkali-resistant adduct formed between *N*-cyclohexyl-*N*′-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate (CMCT) and the N3 of Ψ. 2′-*O*-Methylation (2′-OMe) mapping by 2OMe-seq is based on the differential processivity of the RT through 2′-OMe residues under limiting dNTP concentrations (28), whereas an alternative approach named RiboMeth-seq is based on the increased resistance of 2′-OMe residues toward alkaline hydrolysis (29,30).

Some attempts have been made to realize tools aimed at the analysis of the aforementioned experiments, although most of them are only able to handle individual protocols. StructureFold, Mod-seeker, Spats and ShapeMapper (31–34) have been designed to respectively analyze Structure-seq, Mod-seq, SHAPE-seq and SHAPE-MaP experiments. They all require a control sample to perform background subtraction, and thus are not suitable for the analysis of similar approaches lacking an untreated control, like DMS-seq and DMS-MaPseq. RME (35), although suffering of the same limitation of requiring a control sample, is more versatile as it can analyze structure-probing data from diverse RT drop-off methods, whereas it is not able to handle mutational profiling approaches. Much more complicated is the situation when looking at RNA post-transcriptional modification mapping experiments. Several tools have been specifically designed for the analysis of m^6A MeRIP-seq experiments, such as MeTCluster, DRME, MeRIP-PF, MeT-Diff and HEPeak (36–40), while to our knowledge no tool is currently available for the analysis of single-base resolution mapping experiments such as Ψ-seq, Pseudo-seq, 2OMe-seq and RiboMeth-seq.

We here present RNA Framework, the first comprehensive tool for the analysis of both RNA footprinting and RNA post-transcriptional modification mapping experiments. We show the ability of RNA Framework to deal with most *RNA epistructurome* NGS-based techniques by both analyzing an in-house produced DMS-MaPseq dataset and demonstrating its ability to replicate published analyses starting from raw data of several different experimental protocols.

MATERIALS AND METHODS

General RNA Framework structure

RNA Framework is written in Perl 5, and has multi-thread support. It consists of a suite of independent tools (Figure 1), built on top of a set of object-oriented modules enabling input/output of different file formats (FASTA, Vienna, CT etc.), process handling, nucleic acid sequence manipulation, SVG graphics generation, and more. A detailed API reference will be provided in future releases.

Reference index generation (rf-index)

The *rf-index* module enables the automatic generation of Bowtie v1 and v2 (41,42) reference indexes. This tool requires an internet connection as it relies on querying the UCSC Genome Browser database (43) to obtain transcript annotations and reference genome sequences. It requires the reference genome assembly (a complete list of UCSC available assemblies can be found at https://genome.ucsc.edu/FAQ/FAQreleases.html), and the name of the UCSC MySQL table containing the gene annotations (a complete list of available tables can be found at https://genome.ucsc.edu/cgi-bin/hgTables). The user has also the possibility to choose whether building a reference using either protein coding or non-coding transcripts only. Index generation relies on BEDTools (44) to extract transcript sequences, and on the *bowtie-build* (or *bowtie2-build*) utility shipped with the Bowtie package. Alternatively, pre-build indexes can be automatically retrieved from our webserver (available at http://www.rnaframework.com/indexes.html).

Reads mapping (rf-map)

The *rf-map* module can process any number of FastQ files, from both single-read or paired-end layout experiments. Mixed layout samples can be processed in parallel. It relies on Cutadapt (45) for sequencing adapters and low quality bases clipping, and on Bowtie v1 or v2 for reads mapping. Output alignments are automatically sorted and converted to BAM format (if required) using SAMtools (46).

Counting RT drop-off rates, mutations and coverage (rf-count)

The *rf-count* module constitutes the core component of the framework. It can process any number of SAM/BAM files to calculate per-base RT drop-off rates (or mutations) and coverage. Counts are reported using a proprietary binary format, named RNA Count (RC). RC files store transcript sequences, per-base RT-stop/mutation counts, per-base read coverage, and the total number of reads covering the transcript (Table 1).

Additionally, the RC end-of-file (EOF) stores the number of total mapped reads in the experiment (uint64_t), the RC version (uint16_t), and a seven-character long EOF marker. RC files can be indexed for fast random access. RC index files (RCI) are automatically generated by *rf-count*.

When analyzing MaP experiments, a substantial portion of the structural signal is encoded as deletions in the reverse transcribed cDNA molecules. Since when performing read mapping aligners often report a single possible alignment for a deletion, although multiple equally-scoring alignments are possible, *rf-count* also performs either the left realignment, or the removal, of unambiguously aligned deletions by using a previously described approach (47). Briefly, the portions of the sequence surrounding the deletion are concatenated, and the resulting sequence is stored. Then, the deletion is slid in one nucleotide steps along the sequence, and each time the surrounding sequences are extracted and concatenated. If the stored sequence matches any of the sequences generated during deletion sliding, then the deletion is marked as ambiguous (Figure 2A).

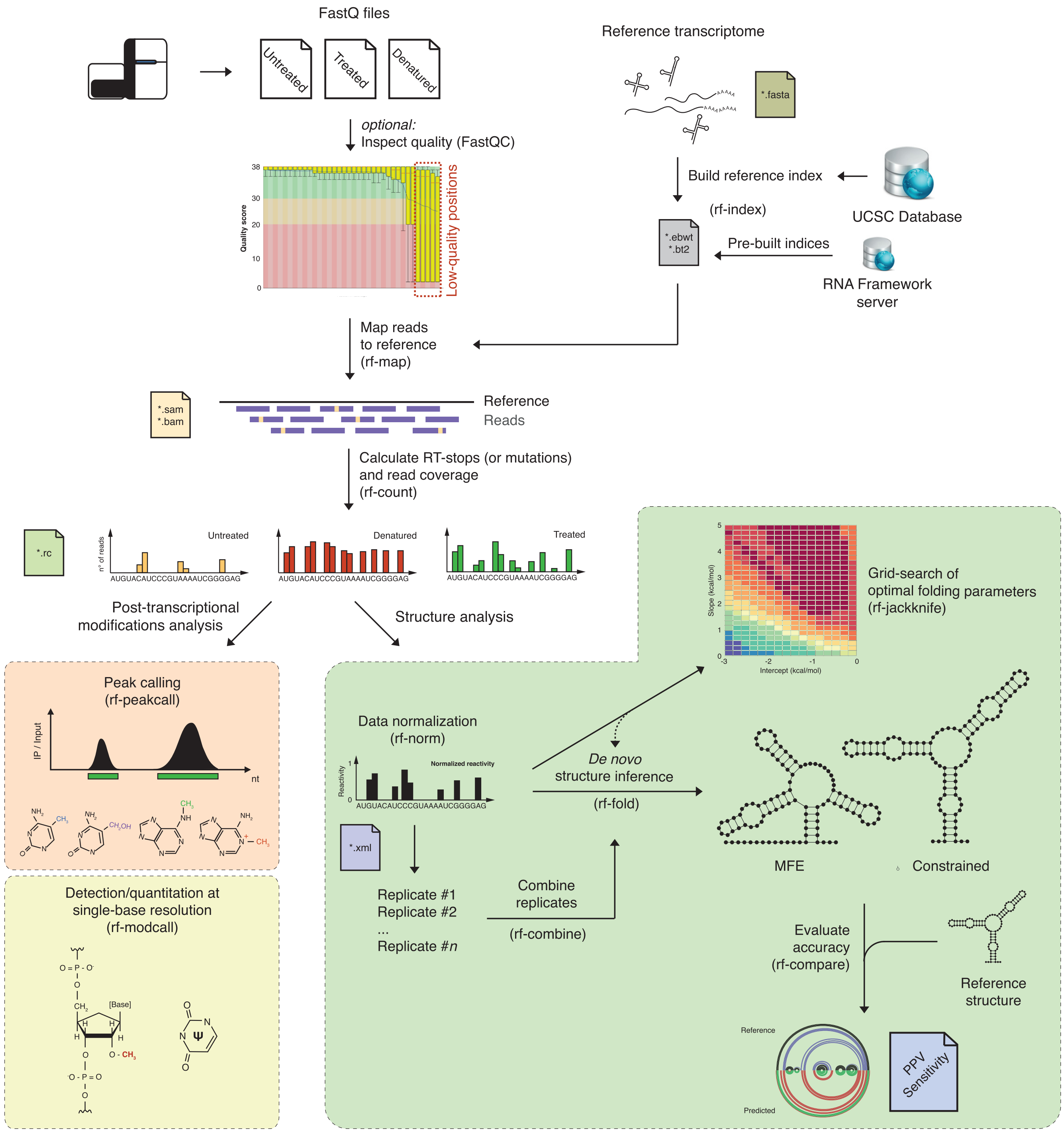

Figure 1. Overview of the RNA Framework suite. Schematics of RNA Framework step-by-step analysis of RNA structure probing and post-transcriptional modification mapping experiments starting from FastQ files.

Table 1. Structure of RC format file's transcript entry

Field	Description	Type
len_transcript_id	Length of the transcript ID (+1, including NULL)	uint32_t
transcript_id	Transcript ID (NULL terminated)	char[len_transcript_id]
len_seq	Length of sequence	uint32_t
seq	4-bit encoded sequence: 'ACGTN' → [0,4] (High nybble first)	uint8_t[(len_seq+1)/2]
counts	Transcript's per base RT-stops (or mutations)	uint32_t[len_seq]
coverage	Transcript's per base coverage	uint32_t[len_seq]
nt	Transcript's mapped reads	unint64_t

Normalization of structure probing data (rf-norm)

The *rf-norm* module processes RNA structure probing experiments' RC files to perform reactivity normalization. To ensure the maximal analysis flexibility, *rf-norm* currently implements four different scoring and three different normalization schemes. The following scoring schemes are available:

1. Ding *et al*. (5). Per-base signal is calculated as the ratio of the natural log(ln) of the raw count of RT-stops/nuclease cuts at a given position of a transcript, to the average of the ln of RT-stops/nuclease cuts along the whole transcript:

$$U_i = \frac{\ln\left(n_{Ui} + p\right)}{\left(\sum_{j=0}^{l} \frac{\ln\left(n_{Uj}+p\right)}{l}\right)}$$

$$T_i = \frac{\ln\left(n_{Ti} + p\right)}{\left(\sum_{j=0}^{l} \frac{\ln\left(n_{Tj}+p\right)}{l}\right)}$$

where n_{Ui} and n_{Ti} are respectively the raw read counts in the untreated and treated samples at position i of the transcript, l is the transcript's length, and p is a pseudocount added to deal with non-covered regions. U_i and T_i are respectively the normalized number of RT-stops at position

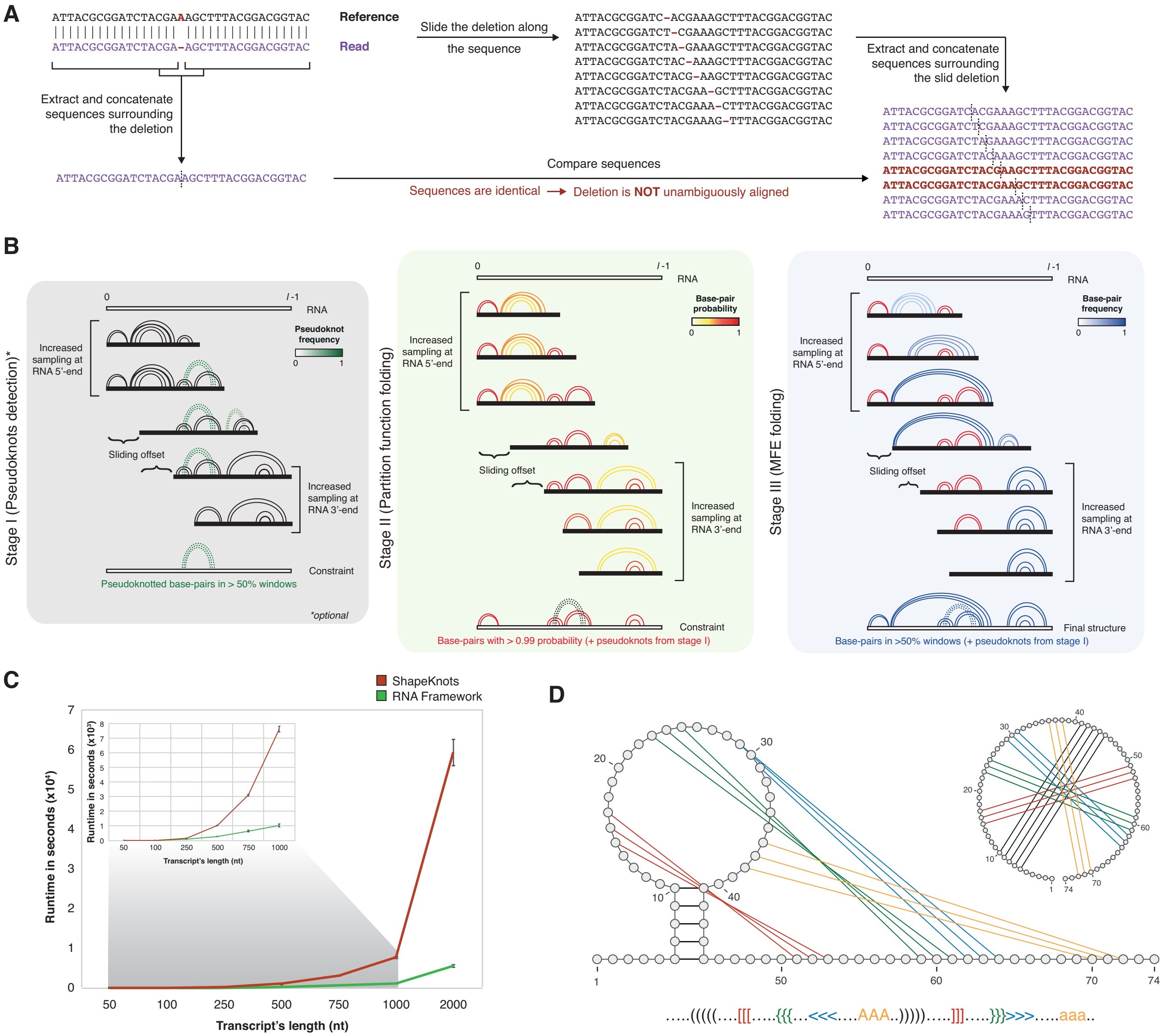
A
ATTACGCGGATCTACGAAAGCTTTACGGACGGTAC
ATTACGCGGATCTACGA-AGCTTTACGGACGGTAC
Reference
Read
Extract and concatenate sequences surrounding the deletion
ATTACGCGGATCTACGAAGCTTTACGGACGGTAC
Slide the deletion along the sequence
ATTACGCGGATC-ACGAAAGCTTTACGGACGGTAC
ATTACGCGGATCT-CGAAAGCTTTACGGACGGTAC
ATTACGCGGATCTA-GAAAGCTTTACGGACGGTAC
ATTACGCGGATCTAC-AAAGCTTTACGGACGGTAC
ATTACGCGGATCTACG-AAGCTTTACGGACGGTAC
ATTACGCGGATCTACGAA-GCTTTACGGACGGTAC
ATTACGCGGATCTACGAAA-CTTTACGGACGGTAC
ATTACGCGGATCTACGAAAG-TTTACGGACGGTAC
Extract and concatenate sequences surrounding the slid deletion
ATTACGCGGATCACGAAAGCTTTACGGACGGTAC
ATTACGCGGATCTCGAAAGCTTTACGGACGGTAC
ATTACGCGGATCTAGAAAGCTTTACGGACGGTAC
ATTACGCGGATCTACAAAGCTTTACGGACGGTAC
ATTACGCGGATCTACGAAGCTTTACGGACGGTAC
ATTACGCGGATCTACGAAGCTTTACGGACGGTAC
ATTACGCGGATCTACGAAACTTTACGGACGGTAC
ATTACGCGGATCTACGAAAGTTTACGGACGGTAC
Compare sequences
Sequences are identical → Deletion is NOT unambiguously aligned
B
Stage I (Pseudoknots detection)*
0
l -1
RNA
Pseudoknot frequency
0
1
Increased sampling at RNA 5'-end
Sliding offset
Increased sampling at RNA 3'-end
Constraint
Pseudoknotted base-pairs in > 50% windows
*optional
Stage II (Partition function folding)
0
l -1
RNA
Base-pair probability
0
1
Increased sampling at RNA 5'-end
Sliding offset
Increased sampling at RNA 3'-end
Constraint
Base-pairs with > 0.99 probability (+ pseudoknots from stage I)
Stage III (MFE folding)
0
l -1
RNA
Base-pair frequency
0
1
Increased sampling at RNA 5'-end
Sliding offset
Increased sampling at RNA 3'-end
Final structure
Base-pairs in >50% windows (+ pseudoknots from stage I)
C
ShapeKnots
RNA Framework
Runtime in seconds (x10⁴)
0
1
2
3
4
5
6
7
Runtime in seconds (x10³)
0
1
2
3
4
5
6
7
8
50
100
250
500
750
1000
Transcript's length (nt)
50
100
250
500
750
1000
2000
Transcript's length (nt)
D
1
10
20
30
40
50
60
70
74
1
10
20
30
40
50
60
70
74
.....(((((....[[[.....{{{...<<<....AAA..)))))).....]]].....}}}>>>.....aaa..

Figure 2. RNA Framework features. (**A**) Ambiguously mapped deletions removal by RNA Framework's *rf-count* for the accurate analysis of mutational profiling (MaP) experiments. (**B**) Schematics of the *rf-fold* windowed RNA folding approach. (**C**) Runtimes comparison for ShapeKnots and RNA Framework pseudoknots detection algorithms on a set of 21 randomly generated RNA sequences (3 × 50–100–250–500–750–1000–2000 nucleotides). (**D**) Sample of the expanded dot-bracket alphabet implemented by RNA Framework to deal with complex topology pseudoknotted RNAs.

i in the untreated and treated samples. Score at position i is then calculated as:

$$S_i = \max(0, (T_i - U_i))$$

2. Rouskin *et al.* (4). Oppositely to the previous scoring scheme, this method does not need an untreated control sample. Raw per-base RT-stop counts are used as a direct measure of the reactivity signal.
3. Siegfried *et al.* (12). This method takes into account both an untreated control sample, and (optionally) a denatured control sample. Per-base raw signal is calculated as:

$$S_i = \frac{\frac{n_{Ti}}{c_{Ti}} - \frac{n_{Ui}}{c_{Ui}}}{\frac{n_{Di}}{c_{Di}}}$$

where n_{Ti}, n_{Ui}, and n_{Di} are respectively the mutation counts in the treated, untreated, and denatured samples at position i of the transcript, while c_{Ti}, c_{Ui}, and c_{Di} are respectively the reads covering position i of the transcript in the treated, untreated, and denatured samples. When no denatured control sample is provided, raw signal is simply calculated as:

$$S_i = \frac{n_{Ti}}{c_{Ti}} - \frac{n_{Ui}}{c_{Ui}}$$

4. Zubradt *et al.* (10). Oppositely to the previous scoring scheme, this method does not need an untreated control sample. Per-base raw signal is calculated as:

$$S_i = \frac{n_{Ti}}{c_{Ti}}$$

where n_{Ti} and c_{Ti} are respectively the mutations count and the read coverage at position i of the transcript.

Following score calculation, raw reactivities can be normalized using one of the following normalization schemes:

1. 2–8% normalization. From the top 10% of raw reactivity values, the top 2% is discarded, then each reactivity value is divided by the average of the remaining 8%. This generally yields reactivities ranging from 0 to ~2.
2. 90% Winsorizing. Raw reactivity values above the 95th percentile are set to the 95th percentile and raw reactivity values below the 5th percentile are set to the 5th percentile, then each reactivity value is divided by the 95th percentile. Yields reactivities ranging from 0 to 1.
3. Box-plot normalization. Raw reactivity values more than 1.5 times greater than the interquartile range (the numerical distance between the 25th and 75th percentiles) above the 75th percentile are removed. After excluding these outliers, the next 10% of remaining reactivity values are averaged, and all reactivities (including outliers) are divided by this value. This generally yields reactivities ranging from 0 to ~1.5.

While 90% Winsorizing yields normalized reactivities ranging from 0 to 1 (respectively indicating residues with a low or high propensity toward single-strandedness), 2–8% normalization and box-plot normalization yield reactivities ranging from 0 to >1. These values can be further remapped to values ranging from 0 to 1 by applying the approach previously proposed by Zarringhalam *et al.* (48). Briefly, values <0.25 are linearly mapped to values in the range 0-0.35, values ≥0.25 and <0.3 are linearly mapped to values in the range 0.35–0.55, values ≥0.3 and <0.7 are linearly mapped to values in the range 0.55–0.85, and values ≥0.7 are linearly mapped to values in the range 0.85–1. Normalization can be performed either in a single step on the whole transcript or using a sliding window approach. The user can moreover specify which bases should be considered during normalization (e.g. A/C for DMS treatment, G/U for CMCT treatment etc.). For each transcript being analyzed, the *rf-norm* module outputs an XML file reporting the used scoring/normalization parameters, the transcript's sequence, and the per-base normalized reactivities.

RNA secondary structure prediction (rf-fold)

The *rf-fold* module is designed to allow transcriptome-wide reconstruction of RNA structures, starting from XML files generated by the *rf-norm* tool. Predictions can be performed using either the ViennaRNA package (13) or RNAstructure (14). Folding can be either computed in a single step on the whole transcript or using a previously described sliding window approach (12). This windowed folding approach consists of 3 stages (Figure 2B). At each stage, normalized reactivities are included in the form of soft constraints. In stage I (optional), a window is slid along the transcript, and potential pseudoknots are detected using the ShapeKnots algorithm (16). The user can choose between the original algorithm, built on top of RNAstructure, or an RNA Framework-specific implementation, built on top of the ViennaRNA package. The latter has comparable performances with transcripts up to 250 nucleotides in length but becomes exponentially faster as the transcript's length increases (Figure 2C). The *rf-fold* module can support even very complex pseudoknotted RNA topologies thanks to the use of an expanded dot-bracket notation's alphabet (Figure 2D), similar to the one employed by the Stockholm format of Rfam (49). Predicted pseudoknotted base-pairs are retained if they appear in >50% of the analyzed windows, and if their average reactivity is below a user-defined threshold. In stage II, a window is slid along the transcript, and the partition function is calculated. If stage I has been performed, pseudoknotted base-pairs are hard-constrained to be single-stranded. Predicted base-pair probabilities are then averaged across all windows in which they have appeared, and base-pairs occurring with >99% probability are retained, and hard-constrained to be base-paired in stage III. In stage III, a window is slid along the transcript, and minimum free energy (MFE) folding is performed, including hard-constraints derived from stages I and II. Predicted base-pairs are retained if they appeared in >50% of the analyzed windows. At this stage, if step I has been performed, pseudoknotted base-pairs are added back to the structure, and the free energy is computed. It is worth noting that, at all stages, increased sampling is performed at both 5′ and 3′ ends to avoid end biases. Predicted structures can be reported either in Vienna format (dot-bracket notation), or in connectivity table (CT) format. The module also produces Wiggle track files containing per-base Shannon entropies, calculated as:

$$H_i = -p_i \, log_{10} \, p_i$$

where p_i is the probability of base i of being base-paired (12), Furthermore, dot-plot files of base-pairing probabilities are produced, along with vector graphical reports in SVG format, reporting a bar-plot of per-base reactivities, the predicted structure, a chart of per-base Shannon entropies, and base-pairing probabilities.

Optimization of RNA folding parameters (rf-jackknife)

Incorporation of RNA footprinting-derived constraints into RNA folding algorithms is usually performed through the formula (15):

$$pseudo\Delta G\ (i) = m \ln(Reactivity(i) + 1) + b$$

where m is the slope, b is the intercept, and the $pseudo\Delta G$ term represents the pseudo-free energy contribution of nucleotide i. Tuning of slope and intercept on a set of reference RNA structures allows determining the values yielding the prediction that better approximates the true reference structure, an operation commonly known as *grid search*, or *jackknifing*. Although this represents a key step in RNA structure inference experiments, as the optimal slope/intercept pair can vary in an experiment-to-experiment fashion, no tool exists to date (to our knowledge) which allows performing automatic grid search. The *rf-jackknife* module takes a set of XML reactivity files generated by *rf-norm*, and a set of reference RNA structures, and iteratively calls the *rf-fold* module by tuning the slope and intercept parameters. For each slope/intercept pair, the structure is predicted and compared to the reference using two metrics: the Positive Predictive Value (PPV), measured as the fraction of base-pairs present in the predicted structure that are also present in the reference structure, and the sensitivity, measured as the fraction of base-pairs in the reference structure that are also present in the predicted structure. The module outputs the optimal slope/intercept pair, as well as three CSV tables respectively reporting the PPV, the sensitivity, and the geometric mean of the 2 metrics for each slope/intercept pair.

Comparison of predicted structures to reference (rf-compare)

The *rf-compare* module allows comparing a set of *rf-fold* inferred structures to a set of reference structures, thus allowing assessing the overall accuracy of the experiment. For each structure pair the module returns the PPV and sensitivity. Moreover, for each comparison it generates an SVG graph, reporting specular arc plots for the reference and compared structures, with base-pairs colored according to their presence in both structures.

Analysis of RNA immunoprecipitation experiments (rf-peakcall)

The *rf-peakcall* module allows performing peak calling of RNA immunoprecipitation (IP) experiments starting from RC files generated by the *rf-count* module, both in the presence or in the absence of a control (or input) sample. Analysis is performed by sliding a window along the transcript, and calculating the signal enrichment in the IP sample versus the control (or input) sample as:

$$E_{(i\ ..\ i+w)} = \log_2\left(\frac{(\mu_{IP(i\ ..\ i+w)} + p}{(Md_{IP} + p)}\right) - \log_2\left(\frac{(\mu_{Ctrl(i\ ..\ i+w)} + p}{(Md_{Ctrl} + p)}\right)$$

where w is the window's length, i and i+w are the start and end position of the window, $\mu_{IP(i..i+w)}$ and $\mu_{Ctrl(i..i+w)}$ are respectively the mean coverage within the analyzed window in the IP and in the control samples, Md_{IP} and Md_{Ctrl} are respectively the median transcript's coverage in the IP and control samples, and p is a pseudocount added to deal with non-covered regions.

If no control sample is provided, the signal enrichment is simply calculated as:

$$E_{(i\ ..\ i+w)} = \log_2\left(\frac{(\mu_{IP(i\ ..\ i+w)} + p}{(Md_{IP} + p)}\right)$$

A P-value is then calculated for each window with an enrichment above a user-defined threshold using a Fisher's exact test. The following 2 × 2 contingency matrix is defined for each threshold-passing window:

	n_{11}	n_{12}
n_{21}	$\mu_{IP(i..i+w)}$	Md_{IP}
n_{22}	$\mu_{Ctrl(i..i+w)}$	Md_{Ctrl}

If no control sample is provided, the contingency matrix is instead defined as:

	n_{11}	n_{12}
n_{21}	$\mu_{IP(i..i+w)}$	Md_{IP}
n_{22}	$\mu_{IP\ windows}$	Md_{IP}

where $\mu_{IP\ windows}$ is the average of each possible window in the IP sample. P-values are Benjamini-Hochberg corrected. Consecutive significantly enriched windows are then merged together, and P-values are combined by Stouffer's method.

Analysis of single-base resolution post-transcriptional modification mapping experiments (rf-modcall)

The *rf-modcall* module allows performing analysis of single-base resolution RNA post-transcriptional modification mapping experiments such as Ψ-seq/Pseudo-seq, 2OMe-seq and RiboMeth-seq. For each position of a transcript 2 measures are computed: the score, a measure of the modification's enrichment, and the ratio, a measure of the modification stoichiometry. For Ψ-seq/Pseudo-seq and 2OMe-seq experiments the score and the ratio are computed as previously described (26–28):

$$S_i = w\,\frac{n_{Ti} - n_{Ui}}{\sum_{j=i-\frac{w}{2}}^{i+\frac{w}{2}}\left(n_{Tj} + n_{Uj}\right) - n_{Ti} - n_{Ui}}$$

$$R_i = \frac{n_{Ti}}{c_{Ti}}$$

where S_i and R_i are respectively the score and the ratio at position i, w is the size of a window centered on position

i, n_{Ti} and n_{Ui} are respectively the number of RT-stops in the CMCT-treated (for Ψ-seq/Pseudo-seq) or low dNTP (for 2OMe-seq) sample and in the CMCT-untreated (for Ψ-seq/Pseudo-seq) or high dNTP (for 2OMe-seq) sample, and c_{Ti} is the read coverage at position i in the CMCT-treated (for Ψ-seq/Pseudo-seq) or low dNTP (for 2OMe-seq) sample.

For RiboMeth-seq experiments the score and the ratio are instead calculated as:

$$S_i = \frac{\left| n_i - 0.5\left(\frac{\sum_{j=i-6}^{i-1} \omega_j n_j}{\sum_{j=i-6}^{i-1} \omega_j} + \frac{\sum_{j=i+1}^{i+6} \omega_j n_j}{\sum_{j=i+1}^{i+6} \omega_j} \right) \right|}{n_i + 1}$$

$$R_i = \max \begin{cases} 1 - \dfrac{n_i}{0.5\left(\frac{\sum_{j=i-6}^{i-1} \omega_j n_j}{\sum_{j=i-6}^{i-1} \omega_j} + \frac{\sum_{j=i+1}^{i+6} \omega_j n_j}{\sum_{j=i+1}^{i+6} \omega_j} \right)} \\ 0 \end{cases}$$

where n_i is the number of RT drop-off at position i, and ω is a weighting parameter linearly varying in 0.1 increments with respect to distance (d) from position i, so that $\omega = 0.5$ for $d = \pm 6$, and $\omega = 1$ for $d = \pm 1$. For each transcript being analyzed, the *rf-modcall* module outputs a XML file reporting the transcript's sequence, and the per-base score and ratio.

Additional utilities

In addition to the aforementioned modules, two additional utilities are shipped with the RNA Framework package. The *rf-combine* module allows combining XML files from multiple experiments (generated either by *rf-norm* or *rf-modcall*) into a single profile (e.g. multiple replicates, or two complementary experiments such as DMS and CMCT probings in CIRS-seq experiments). When replicates are combined, the resulting XML files will contain an optional *error* tag for each combined measure (*reactivity-error* for experiments analyzed using *rf-norm*, and *score-error*/*ratio-error* for experiments analyzed using *rf-modcall*). The *rf-wiggle* module allows processing both RC or XML files into Wiggle tracks for visualization in IGV (50) or other genomics data visualization browsers.

Software help and support

A detailed RNA Framework documentation can be found at the address: http://rnaframework.readthedocs.io/en/latest/. Users can post support requests or report bugs/issues using either the GitHub interface (https://github.com/dincarnato/RNAFramework/issues) or the forum (https://groups.google.com/forum/#!forum/rnaframework). News and updates are available through our blog at the address: http://www.rnaframework.com/blog.

Comparison with ShapeMapper 2

Paired-end FastQ files for *E. coli* TPP, 5S, 16S, 23S, and tRNA^{Phe}, *T. thermophila* Group I intron and *O. iheyensis* Group II intron SHAPE-MaP data (untreated, 1M7-treated denatured and 1M7-treated *in vitro* folded) from Siegfried *et al.* (12) have been analyzed using ShapeMapper 2 with parameters: *–min-depth 1000 –min-mapq 10 –min-qual-to-count 20 –indiv-norm*. For analysis with RNA Framework, reads were mapped using *rf-map* with parameters: *-b2 -cqo -cq5 20 -mp '–local –sensitive-local –mp 3,1 –rdg 5,1 –rfg 5,1 –dpad 30 –maxins 800 –ignore-quals –no-unal –dovetail' -mo*. Mapping parameters were selected to reflect those used by ShapeMapper 2. Count of mutations was performed using *rf-count* with parameters: *-r -nm -m -es -cc*. Reactivity normalization was performed using *rf-norm* with parameters: *-sm 3 -nm 3 -n 1000*. Folding was performed using either the ViennaRNA package (13) or RNAstructure (14), with default slope and intercept values (slope: 1.8 kcal/mol, intercept: -0.6 kcal/mol).

DMS treatment

A single colony of *Escherichia coli* strain DH10B was inoculated into 250 ml of LB medium without antibiotics and grown at 37°C with shaking (150 RPM) for ~4 h, until OD_{600} was ~0.3 (log phase). 25 ml aliquots of bacteria were then pelleted by centrifugation at 1800*g* for 15 minutes (4°C). After centrifugation, medium was decanted, and cells from 25 ml of culture each were resuspended in 1 ml of structure probing buffer [50 mM HEPES–KOH pH 7.9; 100 mM NaCl; 3 mM KCl]. DMS was diluted 1:6 in 100% ethanol to a final concentration of 1.76 M. Diluted DMS was added to bacteria to a final concentration of ~105 mM. Samples were incubated with moderate shaking (800 RPM) at 25°C for 2 min, after which reactions were immediately transferred on ice. DTT was added to a final concentration of 0.7 M to quench DMS, and samples were vigorously vortexed for 10 s. Bacteria were then pelleted by centrifugation at 10 000*g* for 30 s (4°C), and the supernatant was decanted. Pellets were then washed once with 1 ml Isoamyl alcohol (Sigma Aldrich, cat. W205702) to remove traces of water-insoluble DMS. Bacteria were then pelleted by additional centrifugation at 10 000*g* for 30 s (4°C), supernatant was decanted, and samples were snap-frozen in liquid nitrogen.

DMS-MaPseq library preparation

1 μg of DMS-treated RNA, either total, or rRNA-depleted using Ribo-Zero rRNA Removal Kit (Illumina), was fragmented in the presence of 10 mM $MgCl_2$ at 94°C for 8 min. RNA was end-repaired by treatment with 1 U rSAP (NEB) at 37°C for 30 min. After heat-inactivating the enzyme at 65°C for 5 minutes, RNA fragments were phosphorylated by treatment with 10 U T4 Polynucleotide kinase (NEB) in the presence of 1 mM final ATP at 37°C for 1 h. After reaction cleanup on RNA Clean & Concentrator™-5 columns (Zymo Research), 3′ and 5′ adapters were ligated to RNA fragments using the NEBNext® Small RNA Library Prep Set (NEB). RNA was then heat-denatured at 70°C for 5 min and incubated at room temperature for 5 min. Reverse transcription was carried out in a final volume of 10 μl, in the presence of 1 mM dNTPs, 10 pmol of SR RT primer (NEBNext® Small RNA Library Prep Set kit), 20 U RNaseOUT™ Recombinant Ribonuclease Inhibitor, and 100 U TGIRT™-III Reverse Transcriptase (InGex), by incubating at 50°C for 5 min, followed by 2 h at 57°C. Template

RNA was degraded by adding 1 μl of 5 M NaOH and incubating at 95°C for 3 min. After reaction cleanup, cDNA was eluted in 20 μl nuclease-free water, and barcodes were introduced by 15 cycles of PCR in the presence of 25 pmol of each primer, and 25 μl NEBNext® High-Fidelity 2× PCR Master Mix.

Deposited data

Sequencing data have been deposited on the NCBI Gene Expression Omnibus (GEO) under the accession GSE111962.

RESULTS AND DISCUSSION

To demonstrate the features, versatility and easiness of use of RNA Framework, and its applicability to a wide range of cases, we here analyze an in-house produced DMS-MaPseq dataset, and re-analyze a set of previously published experiments, showing the ability of RNA Framework to replicate authors' analyses and findings.

RNA structure probing experiments data analysis

In order to demonstrate some of the RNA Framework features, we here produced a small DMS-MaPseq mutational profiling dataset by probing exponential phase *E. coli* cells with dimethyl sulfate (DMS, Supplementary Note 1). DMS is a cell-permeable reagent that modifies unpaired A and C residues. We produced libraries from both total RNA and ribo-depleted RNA (Figure 3A), that were mapped by *rf-map* using Bowtie v2. The resulting BAM files were combined and passed to *rf-count* to perform mutation count, revealing a mutational signature enriched on A and C residues (48.3% and 32.4% respectively), typical of DMS-MaPseq experiments. To determine optimal folding parameters (slope and intercept), we applied the Zubradt *et al.*, 2016 scoring scheme (10) and box-plot normalization to 16S and 23S rRNA data through the *rf-norm* module and passed the resulting normalized XML files to the *rf-jackknife* module, along with the phylogenetically-inferred reference rRNA structures (51) to perform grid search (Figure 3B). This analysis revealed optimal slope and intercept values to be respectively 1.2 kcal/mol and –0.8 kcal/mol. We then applied the same normalization scheme to all other transcripts (Figure 3C), and performed constrained folding using the *rf-fold* module. Application of constraints derived from DMS probing data increased the overall prediction accuracy for both pseudoknotted (Figure 3D) and non-pseudoknotted (Figure 3E) structures.

Beside DMS, SHAPE reagents are widely used as probes of RNA structural flexibility as they can form adducts with the free 2′-OH of the ribose moiety of structurally flexible residues. 1-methyl-7-nitroisatoic anhydride (1M7) is one of the best-known SHAPE reagents and can be used to readily probe RNA structures both *in vitro* and *in vivo* (although its suitability for *in vivo* probing is controversial (52)) on a small time-scale (half-life: ∼17 s at 37°C) (53). In a previous work Weeks *et al.* have applied SHAPE-MaP to probe the *ex virio* deproteinized structure of the HIV-1 genome (12). The original dataset was composed of three samples: an untreated control, a denatured control obtained by treating HIV-1 RNA with 1M7 under denaturing conditions (high temperature in the presence of formamide), and a 1M7-treated sample under native conditions (Supplementary Note 2). Samples were mapped to the HIV-1 genome (GenBank: M19921.2) using *rf-map* and Bowtie v2, then mutations were counted using *rf-count*. Raw reactivities were computed using the Siegfried *et al.* (2014) scoring scheme, and normalized by box-plot normalization using *rf-norm* (Figure 4A and B). The reconstructed SHAPE reactivity profile well resembled the original profile from Siegfried *et al.* (2014) (PCC = 0.91, Supplementary Figure S1A and B). Normalized data was then provided to *rf-fold*, and the structure was inferred using the windowed folding approach, using the original authors-defined parameters. Besides predicting the minimum expected accuracy (MEA) structure (Figure 4B), the algorithm also computed base-pairing probabilities and the Shannon entropy (Figure 4C). According to the authors, regions of low median SHAPE reactivity (Figure 4A) and low Shannon entropy (Figure 4C) correspond to stable secondary structure elements. In agreement with this observation, our algorithm successfully inferred key structural elements of the HIV-1 genome in these regions (Figure 4D).

We further sought to compare the performances of RNA Framework to those of ShapeMapper 2 (33), the only other tool available to date for the analysis of SHAPE-MaP data, using a previously published dataset of *in vitro* folded transcripts subjected to SHAPE-MaP analysis (12). Pseudoknot-free structures were predicted using either the ViennaRNA package (13) or RNAstructure (14). Notably, while analysis conducted using RNAstructure with soft constraints derived by either tools yielded comparable results in terms of both PPV and sensitivity (median PPV/sensitivity: 0.87/0.89 for both tools, Supplementary Figure S2A), analysis conducted using the ViennaRNA package gave markedly better results with soft constraints inferred by RNA Framework (RNA Framework median PPV/sensitivity: 0.92/0.89, ShapeMapper 2 median PPV/sensitivity: 0.76/0.78, Supplementary Figure S2B). Although we have not further looked into the reasons of these differences, this data suggests that RNAstructure is less sensitive to small SHAPE reactivity variations than the ViennaRNA package and confirms the robustness of RNA Framework despite of the employed RNA structure prediction software.

RNA post-transcriptional modification mapping data analysis

To demonstrate data analysis of RNA post-transcriptional modification mapping experiments we re-analyzed three previously published datasets. The first two are datasets of m^6A-seq and m^1A-seq in HepG2 cells (20,21), while the second is a dataset of single-base resolution Ψ mapping by Pseudo-seq in exponential phase yeast (26). m^6A has been reported by several authors to occur in highly conserved regions within the consensus sequence RRACH (mostly GGACH), in strict proximity to stop codons (20,23,54), while m^1A has been reported to be mainly enriched in thermodynamically stable 5′-UTRs and proposed to occur in

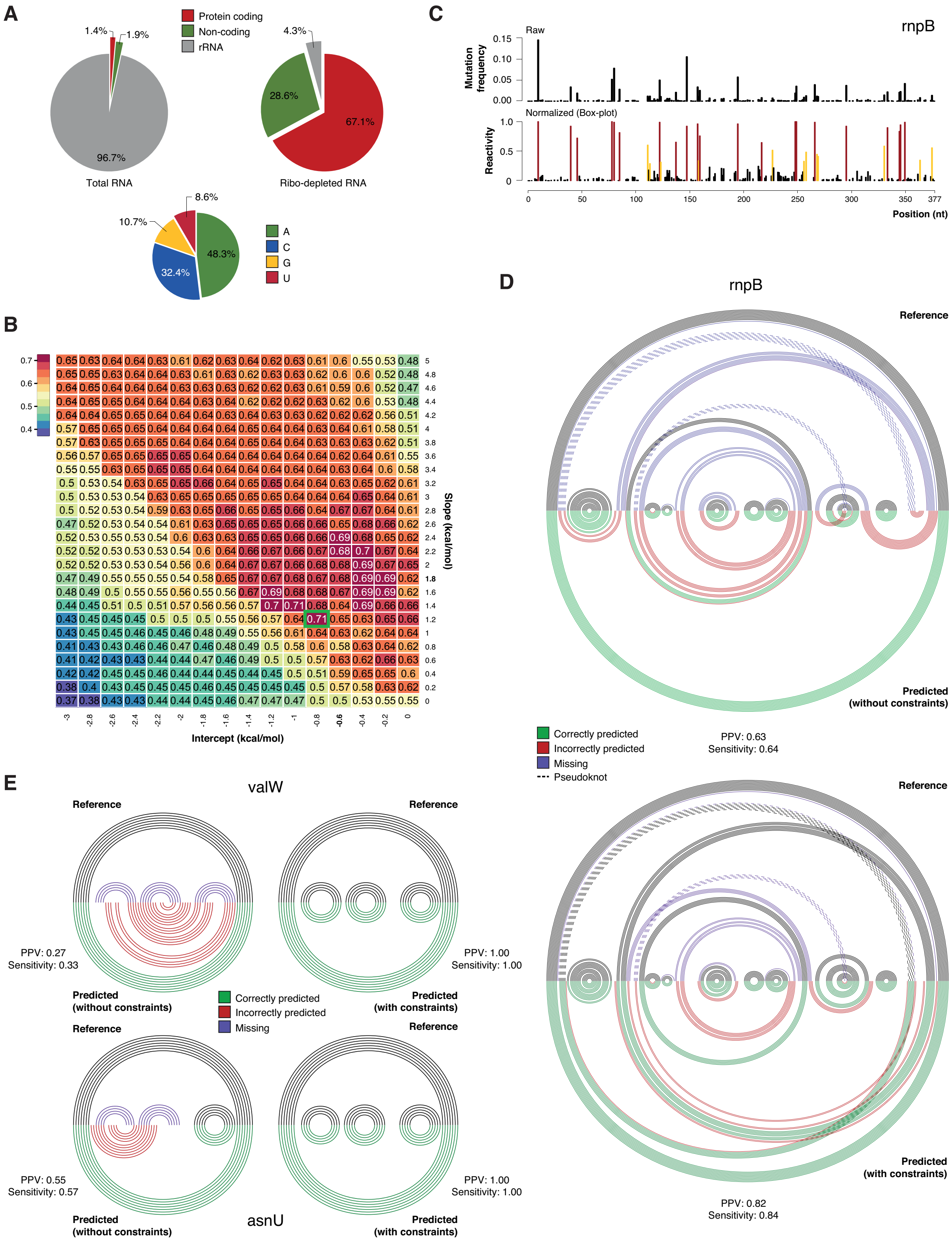

Figure 3. DMS-MaPseq data analysis. (**A**) Distributions of read mappings on different classes of transcripts in the total RNA and Ribo-depleted DMS-MaPseq libraries (*top*), and combined base mutation frequencies (*bottom*). (**B**) Matrix of PPV/sensitivity geometric means calculated by *rf-jackknife* by tuning slope and intercept parameters on 16S and 23S rRNA reference structures. The optimal slope/intercept pair is marked by a green rectangle (slope: 1.2; intercept: –0.8). (**C**) Example of raw DMS-MaPseq data normalization by *rf-norm* using the box-plot normalization method on RNase P (rnpB). (**D**) Arc plot comparison of RNase P (rnpB) reference structure, and structure inferred by *rf-fold* both in the absence (*top*) and presence (*bottom*) of DMS-MaPseq soft constraints. Image was automatically generated by the *rf-compare* utility. (**E**) Arc plot comparison of tRNA Val and Asn (valW and asnU) reference structures, and structures inferred by *rf-fold* both in the absence (*top*) and presence (*bottom*) of DMS-MaPseq soft constraints. Images were automatically generated by the *rf-compare* utility.

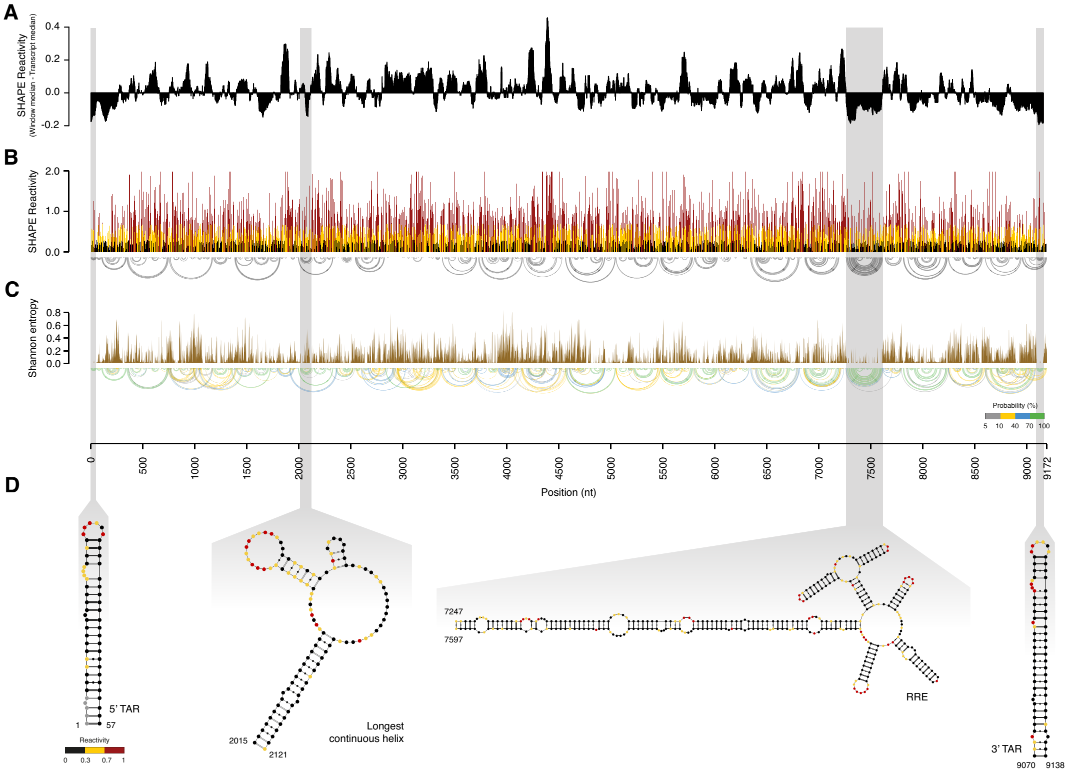

Figure 4. SHAPE-MaP data analysis. (**A**) Bar-plot of median SHAPE reactivity in 55 nt windows compared to median SHAPE reactivity along the whole HIV-1 genome. Low SHAPE reactivity regions correspond to stable secondary structure elements. (**B**) Bar-plot of base-level SHAPE reactivity (*top*) and arc plot of HIV-1 MEA structure (*bottom*). Reactivity was capped to a value of 2 for graphical reasons. Plots were automatically generated by the *rf-fold* module. (**C**) Plot of base-level Shannon entropy (*top*) and arc plot of base-pairing probabilities. Plots were automatically generated by the *rf-fold* module. (**D**) Secondary structure models of key HIV-1 structural motifs. Models were generated using VARNA (55). Bases were colored according to their SHAPE reactivity.

purine-rich contexts (21,22). Samples were mapped to the human transcriptome using *rf-map* and Bowtie v1, coverage was calculated using *rf-count*, and peaks were called using *rf-peakcall* (Supplementary Note 3, 4). In agreement with previous reports, m^6A-seq peaks were enriched around stop codons, and motif discovery analysis revealed a significant enrichment of the core m^6A consensus GGAC ($P = 5.6$e–14, Wilcoxon Rank Sum test, Figure 5A). As previously described, m^6A peaks are more conserved than expected by chance ($P = 1.9$e–5, Wilcoxon Rank Sum test, Figure 5B). Conversely, m^1A peaks are enriched toward TSSs, and motif enrichment analysis revealed the presence of the expected purine-rich motif ($P = 9.0$e–16, Figure 5C), beside an enrichment for CG-rich sequences (*data not shown*) in agreement with previous reports showing preferential positioning of m^1A residue within CG-rich regions (21). Since m^1A positive 5′-UTRs have been shown to be more thermodynamically stable than unmethylated 5′-UTRs, we further applied RNA Framework to the analysis of PARS data from HepG2 cells and compared length-adjusted MFE values for m^1A positive versus m^1A negative 5′-UTRs. As previously described, m^1A positive UTRs form significantly more thermodynamically stable secondary structures (median: –57.8 kcal/mol) than their unmethylated counterparts (median: −49.2 kcal/mol, $P < 2.2$e–16, Figure 5D).

To further show the ability of RNA Framework to deal with the analysis of post-transcriptional modification mapping experiments, we also re-analyzed a dataset of Ψ mapping by Pseudo-seq. Ψ is the most abundant RNA post-transcriptional modification and occurs on several rRNA sites. After mapping the Pseudo-seq dataset, composed of a CMC-treated (CMC+) and a CMC-untreated (CMC-) sample, on 18S and 25S sequences from *Saccharomyces cerevisiae* using *rf-map* and Bowtie v1, we counted RT drop-off rates using *rf-count* and proceeded to the analysis using *rf-modcall* (Supplementary Note 5). Analysis successfully identified known rRNA Ψ residues using the score metric (26), and their relative stoichiometry using the ratio metric (27) (Figure 5E, see Material and Methods).

CONCLUSION

The increasing interest RNA is receiving from the scientific community is rapidly leading to the generation of large amounts of NGS data. Despite this growing interest, efficient tools enabling fast and streamlined data analysis are still missing. We here presented RNA Framework, the to date most complete toolkit for the comprehensive analysis of NGS-based RNA structure and post-transcriptional modification mapping experiments. We demonstrated through different application cases that our software enables the rapid analysis of most NGS ap-

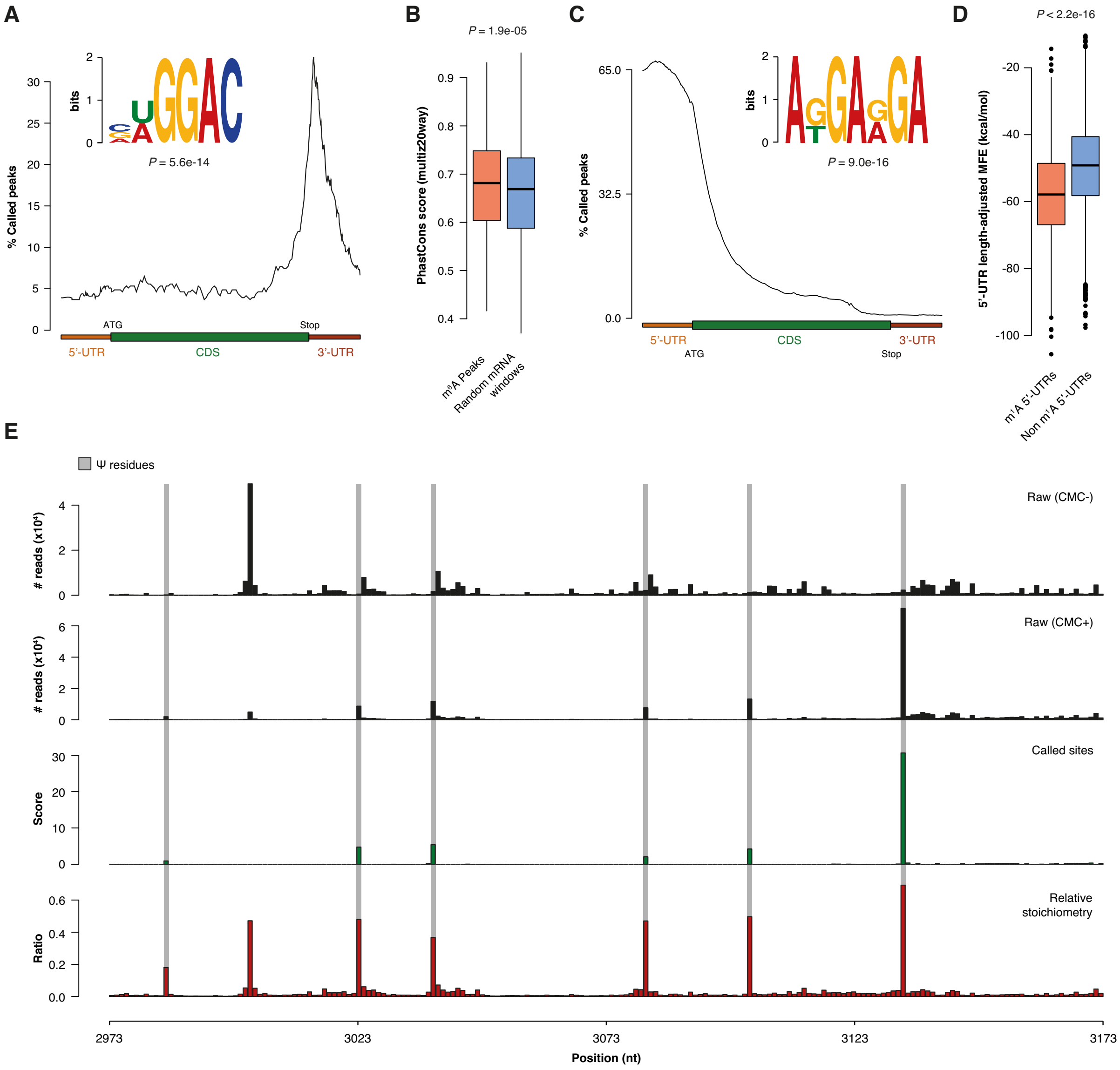

Figure 5. RNA post-transcriptional modification mapping experiments data analysis. (**A**) Meta-gene plot of m^6A peaks distribution and top enriched motif as detected by DREME (56). (**B**) Box-plot of PhastCons scores from multiz20way alignment for m^6A peaks (scaled to 200 nt with respect to the peak's center) compared to random mRNA windows of matching size. (**C**) Meta-gene plot of m^1A peaks distribution and top A-containing enriched motif as detected by DREME. (**D**) Box-plot of 5′-UTR length-adjusted MFE values for m^1A 5′-UTRs compared to non m^1A 5′-UTRs. (**E**) Bar-plots of raw RT drop-off signal in CMC– and CMC+ samples, score, and ratio in a 200 nt region of yeast 25S rRNA containing 6 known Ψ residues (highlighted in gray).

proaches developed to date, thus providing a cornerstone for the study of the *RNA epistructurome*.

DATA AVAILABILITY

RNA Framework is available at http://www.rnaframework.com. For support request or to report bugs a user group is available at https://groups.google.com/forum/#!forum/rnaframework. For news and walkthrough examples a blog is available at http://www.rnaframework.com/blog. Sequencing data have been deposited on the NCBI Gene Expression Omnibus (GEO) under the accession GSE111962.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Prof. Weeks (UNC Chapel Hill) for kindly sharing SHAPE-MaP data from Siegfried *et al.* (2014).

FUNDING

Associazione Italiana Ricerca sul Cancro (AIRC) [IG 20240 to S.O.]; AIRC TRansforming IDEas in Oncological research (TRIDEO) [17182 to D.I.]. Funding for open access charge: Hosting institution's funds.

Conflict of interest statement. None declared.

REFERENCES

1. Incarnato,D. and Oliviero,S. (2017) The RNA Epistructurome: Uncovering RNA function by studying structure and Post-Transcriptional modifications. *Trends Biotechnol.*, **35**, 318–333.
2. Loughrey,D., Watters,K.E., Settle,A.H. and Lucks,J.B. (2014) SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res.*, **42**, e165.
3. Kertesz,M., Wan,Y., Mazor,E., Rinn,J.L., Nutter,R.C., Chang,H.Y. and Segal,E. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*, **467**, 103–107.
4. Rouskin,S., Zubradt,M., Washietl,S., Kellis,M. and Weissman,J.S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, **505**, 701–705.
5. Ding,Y., Tang,Y., Kwok,C.K., Zhang,Y., Bevilacqua,P.C. and Assmann,S.M. (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, **505**, 696–700.
6. Incarnato,D., Neri,F., Anselmi,F. and Oliviero,S. (2014) Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol.*, **15**, 491.
7. Incarnato,D., Morandi,E., Anselmi,F., Simon,L.M., Basile,G. and Oliviero,S. (2017) In vivo probing of nascent RNA structures reveals principles of cotranscriptional folding. *Nucleic Acids Res.*, **45**, 9716–9725.
8. Lucks,J.B., Mortimer,S.A., Trapnell,C., Luo,S., Aviran,S., Schroth,G.P., Pachter,L., Doudna,J.A. and Arkin,A.P. (2011) Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11063–11068.
9. Spitale,R.C., Flynn,R.A., Zhang,Q.C., Crisalli,P., Lee,B., Jung,J.-W., Kuchelmeister,H.Y., Batista,P.J., Torre,E.A., Kool,E.T. *et al.* (2015) Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*, **519**, 486–490.
10. Zubradt,M., Gupta,P., Persad,S., Lambowitz,A.M., Weissman,J.S. and Rouskin,S. (2016) DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods*, **14**, 75–82.
11. Homan,P.J., Favorov,O.V., Lavender,C.A., Kursun,O., Ge,X., Busan,S., Dokholyan,N.V. and Weeks,K.M. (2014) Single-molecule correlated chemical probing of RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 13858–13863.
12. Siegfried,N.A., Busan,S., Rice,G.M., Nelson,J.A.E. and Weeks,K.M. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods*, **11**, 959–965.
13. Lorenz,R., Bernhart,S.H., Höner Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol Biol*, **6**, 26.
14. Reuter,J.S. and Mathews,D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, **11**, 129.
15. Deigan,K.E., Li,T.W., Mathews,D.H. and Weeks,K.M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 97–102.
16. Hajdin,C.E., Bellaousov,S., Huggins,W., Leonard,C.W., Mathews,D.H. and Weeks,K.M. (2013) Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 5498–5503.
17. Underwood,J.G., Uzilov,A.V., Katzman,S., Onodera,C.S., Mainzer,J.E., Mathews,D.H., Lowe,T.M., Salama,S.R. and Haussler,D. (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods*, **7**, 995–1001.
18. Behm-Ansmant,I., Helm,M. and Motorin,Y. (2011) Use of specific chemical reagents for detection of modified nucleotides in RNA. *J. Nucleic Acids*, **2011**, 1–17.
19. Delatte,B., Wang,F., Ngoc,L.V., Collignon,E., Bonvin,E., Deplus,R., Calonne,E., Hassabi,B., Putmans,P., Awe,S. *et al.* (2016) RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science*, **351**, 282–285.
20. Dominissini,D., Moshitch-Moshkovitz,S., Schwartz,S., Salmon-Divon,M., Ungar,L., Osenberg,S., Cesarkas,K., Jacob-Hirsch,J., Amariglio,N., Kupiec,M. *et al.* (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*, **485**, 201–206.
21. Dominissini,D., Nachtergaele,S., Moshitch-Moshkovitz,S., Peer,E., Kol,N., Ben-Haim,M.S., Dai,Q., Di Segni,A., Salmon-Divon,M., Clark,W.C. *et al.* (2016) The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature*, **530**, 441–446.
22. Li,X., Xiong,X., Wang,K., Wang,L., Shu,X., Ma,S. and Yi,C. (2016) Transcriptome-wide mapping reveals reversible and dynamic N^1-methyladenosine methylome. *Nat. Chem. Biol.*, **12**, 311–316.
23. Meyer,K.D., Saletore,Y., Zumbo,P., Elemento,O., Mason,C.E. and Jaffrey,S.R. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell*, **149**, 1635–1646.
24. Edelheit,S., Schwartz,S., Mumbach,M.R., Wurtzel,O. and Sorek,R. (2013) Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet.*, **9**, e1003602.
25. Squires,J.E., Patel,H.R., Nousch,M., Sibbritt,T., Humphreys,D.T., Parker,B.J., Suter,C.M. and Preiss,T. (2012) Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.*, **40**, 5023–5033.
26. Carlile,T.M., Rojas-Duran,M.F., Zinshteyn,B., Shin,H., Bartoli,K.M. and Gilbert,W.V. (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature*, **515**, 143–146.
27. Schwartz,S., Bernstein,D.A., Mumbach,M.R., Jovanovic,M., Herbst,R.H., León-Ricardo,B.X., Engreitz,J.M., Guttman,M., Satija,R., Lander,E.S. *et al.* (2014) Transcriptome-wide mapping reveals widespread Dynamic-Regulated pseudouridylation of ncRNA and mRNA. *Cell*, **159**, 148–162.
28. Incarnato,D., Anselmi,F., Morandi,E., Neri,F., Maldotti,M., Rapelli,S., Parlato,C., Basile,G. and Oliviero,S. (2016) High-throughput single-base resolution mapping of RNA 2′-O-methylated residues. *Nucleic Acids Res.*, **45**, 1433–1441.
29. Birkedal,U., Christensen-Dalsgaard,M., Krogh,N., Sabarinathan,R., Gorodkin,J. and Nielsen,H. (2015) Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew. Chem. Int. Ed. Engl.*, **54**, 451–455.
30. Marchand,V., Blanloeil-Oillo,F., Helm,M. and Motorin,Y. (2016) Illumina-based RiboMethSeq approach for mapping of 2′-O-Me residues in RNA. *Nucleic Acids Res.*, **44**, e135.
31. Tack,D.C., Tang,Y., Ritchey,L.E., Assmann,S.M. and Bevilacqua,P.C. (2018) StructureFold2: bringing chemical probing data into the computational fold of RNA structural analysis. *Methods*, doi:10.1016/j.ymeth.2018.01.018.
32. Talkish,J., May,G., Lin,Y., Woolford,J.L. and McManus,C.J. (2014) Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA*, **20**, 713–720.
33. Busan,S. and Weeks,K.M. (2018) Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA*, **24**, 143–148.
34. Aviran,S., Trapnell,C., Lucks,J.B., Mortimer,S.A., Luo,S., Schroth,G.P., Doudna,J.A., Arkin,A.P. and Pachter,L. (2011) Modeling and automation of sequencing-based characterization of RNA structure. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11069–11074.
35. Wu,Y., Shi,B., Ding,X., Liu,T., Hu,X., Yip,K.Y., Yang,Z.R., Mathews,D.H. and Lu,Z.J. (2015) Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data. *Nucleic Acids Res.*, **43**, 7247–7259.
36. Cui,X., Meng,J., Zhang,S., Rao,M.K., Chen,Y. and Huang,Y. (2016) A hierarchical model for clustering m(6)A methylation peaks in MeRIP-seq data. *BMC Genomics*, **17**(Suppl. 7), 520.
37. Liu,L., Zhang,S.-W., Gao,F., Zhang,Y., Huang,Y., Chen,R. and Meng,J. (2016) DRME: Count-based differential RNA methylation analysis at small sample size scenario. *Anal. Biochem.*, **499**, 15–23.
38. Li,Y., Song,S., Li,C. and Yu,J. (2013) MeRIP-PF: an easy-to-use pipeline for high-resolution peak-finding in MeRIP-Seq data. *Genomics Proteomics Bioinformatics*, **11**, 72–75.
39. Cui,X., Zhang,L., Meng,J., Rao,M., Chen,Y. and Huang,Y. (2018) MeTDiff: a novel differential RNA methylation analysis for MeRIP-Seq Data. *IEEE/ACM Trans. Comput. Biol. Bioinf.*, **15**, 526–534.

40. Cui,X., Meng,J., Rao,M.K., Chen,Y. and Huang,Y. (2015) HEPeak: an HMM-based exome peak-finding package for RNA epigenome sequencing data. *BMC Genomics*, **16**(Suppl. 4), S2.
41. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
42. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
43. Casper,J., Zweig,A.S., Villarreal,C., Tyner,C., Speir,M.L., Rosenbloom,K.R., Raney,B.J., Lee,C.M., Lee,B.T., Karolchik,D. *et al.* (2018) The UCSC genome browser database: 2018 update. *Nucleic Acids Res.*, **46**, D762–D769.
44. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
45. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10–12.
46. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
47. Smola,M.J., Rice,G.M., Busan,S., Siegfried,N.A. and Weeks,K.M. (2015) Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.*, **10**, 1643–1669.
48. Zarringhalam,K., Meyer,M.M., Dotu,I., Chuang,J.H. and Clote,P. (2012) Integrating chemical footprinting data into RNA secondary structure prediction. *PLoS ONE*, **7**, e45160.
49. Kalvari,I., Argasinska,J., Quinones-Olvera,N., Nawrocki,E.P., Rivas,E., Eddy,S.R., Bateman,A., Finn,R.D. and Petrov,A.I. (2018) Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.*, **46**, D335–D342.
50. Thorvaldsdóttir,H., Robinson,J.T. and Mesirov,J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics*, **14**, 178–192.
51. Cannone,J.J., Subramanian,S., Schnare,M.N., Collett,J.R., D'Souza,L.M., Du,Y., Feng,B., Lin,N., Madabusi,L.V., Müller,K.M. *et al.* (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, **3**, 2.
52. Lee,B., Flynn,R.A., Kadina,A., Guo,J.K., Kool,E.T. and Chang,H.Y. (2017) Comparison of SHAPE reagents for mapping RNA structures inside living cells. *RNA*, **23**, 169–174.
53. Smola,M.J., Rice,G.M., Busan,S., Siegfried,N.A. and Weeks,K.M. (2015) Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.*, **10**, 1643–1669.
54. Batista,P.J., Molinie,B., Wang,J., Qu,K., Zhang,J., Li,L., Bouley,D.M., Lujan,E., Haddad,B., Daneshvar,K. *et al.* (2014) m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell*, **15**, 707–719.
55. Darty,K., Denise,A. and Ponty,Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.
56. Bailey,T.L. (2011) DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics*, **27**, 1653–1659.