In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features

Yiliang Ding, Yin Tang, Chun Kit Kwok, Yu Zhang, Philip C. Bevilacqua & Sarah M. Assmann

(2014)

Seminar RNA Bioinformatics

Introduction

- RNA high versatile molecule
- Ability to encode and to manipulate genetic information
- RNA has the aptitude to fold back on itself to form biologically functional structures
- RNA structure plays critical role in processes ranging

Introduction

- From ligand sensing to the regulation of translation, polyadenylation and splicing
- Structural data from RNA → How RNA structure regulates gene expression
- Most existing RNA structure mapping methods have been performed in vitro
- A method for genome-wide study of RNA structure in vivo has been lacking

Introduction

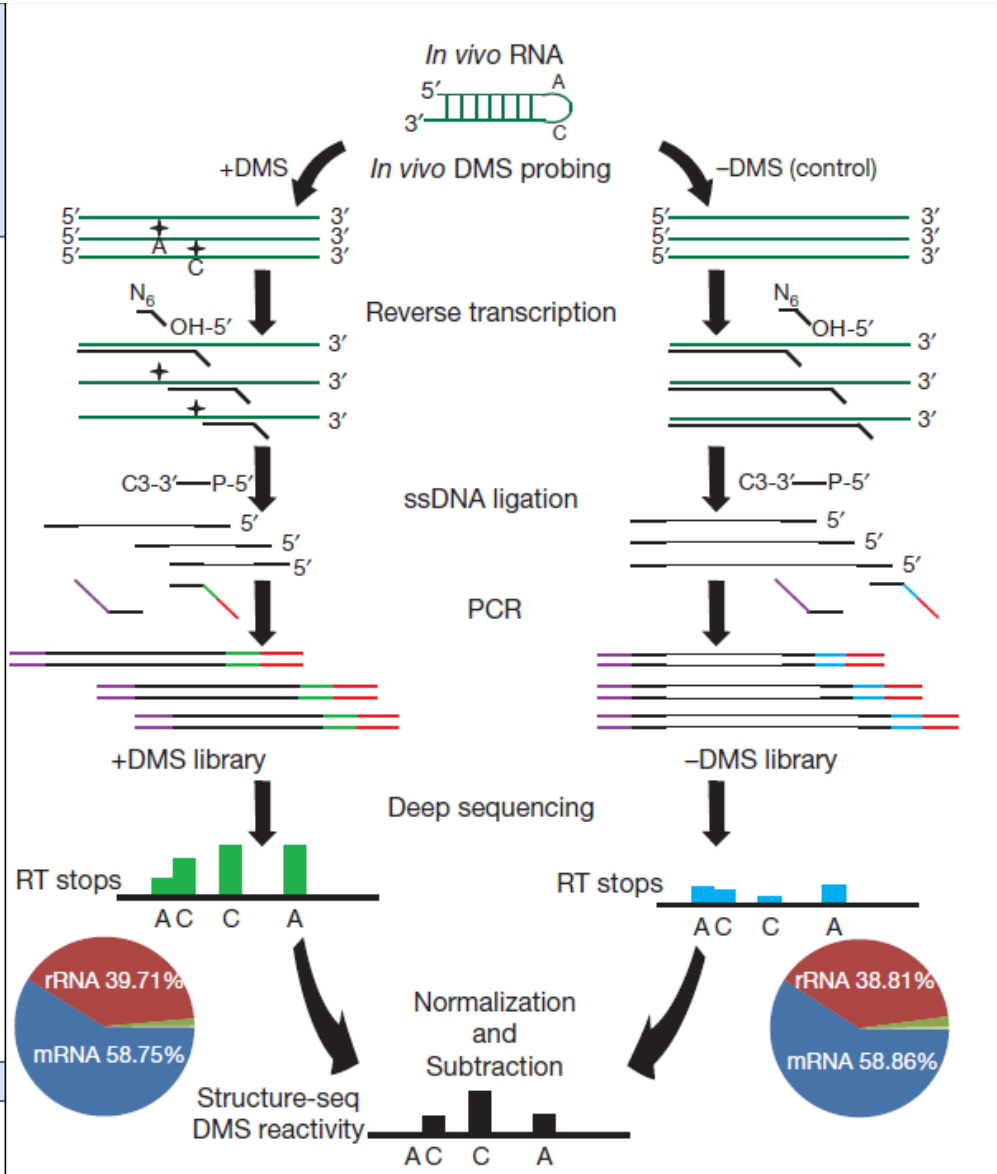
- Structure-Seq: High-throughput method for genome wide in vivo RNA structure probing
- In vivo quantitative measurement of genome-wide RNA secondary structure at nucleotide resolution
- Combine dimethyl sulphate (DMS) methylation method with next-generation sequencing

Material and methods

Dimethyl sulphate (DMS)

- Has been used to map structures of high-abundance RNAs in vivo in various organisms
- Methylates the base-pairing faces of adenine and cytosine of RNA whenever they are accessible to the methylation
- DMS can tell us which region of RNA is unpaired (unstructured), not how the structure is

Overview of Structure-seq



Material and methods

Plant materials and in vivo DMS chemical probing

- 5 days old Arabidopsis thaliana etiolated seedlings
- Added DMS and allowed to react for 15 min at room temperature (22°C)
- Reaction was quenched, the seedlings washed, frozen and submit to RNA extraction

Material and methods

Illumina library construction and mapping

- In vivo total RNA isolation was followed by one round of poly(A) selection
- The RNA was re-suspended in RNase-free water and subjected to reverse transcription
- The resultant first-strand cDNAs were then ligated at their 3’ ends to a ssDNA linker
- PCR amplification was performed on the ligated cDNA using Illumina TruSeq Primers

Material and methods

Illumina library construction and mapping

- Remove adapters and achieve a uniform size distribution of PCR products between 150 and 650 base pairs (bp)
- Subjected the DNA libraries to next-generation sequencing on Illumina HiSeq 2000
- An independent biological replicate was prepared in the same way and separately subjected to next-generation sequencing

Material and methods

Illumina library construction and mapping

- Illumina sequencing read were mapped to the Arabidopsis genome
- Mapping was performed using Bowtie
- High correlation between (+)DMS and (-)DMS libraries

Material and methods

- **Determination and normalization of DMS reactivity for each nucleotide on each transcript**:
  - ln[Pr(i)]: natural logarithm of the number of reverse transcriptase stops mapped to nucleotide position i
  - Divided by average of the ln of reverse transcriptase stops per position
  - This Normalise the number of reverse transcriptase stop for nucleotide i

$$P(i) = \frac{\ln[P_r(i)]}{(\sum_{i=0}^{l} \ln[P_r(i)])/l} \qquad M(i) = \frac{\ln[M_r(i)]}{(\sum_{i=0}^{l} \ln[M_r(i)])/l}$$

Material and methods

For each nucleotide calculate raw DMS reactivity:

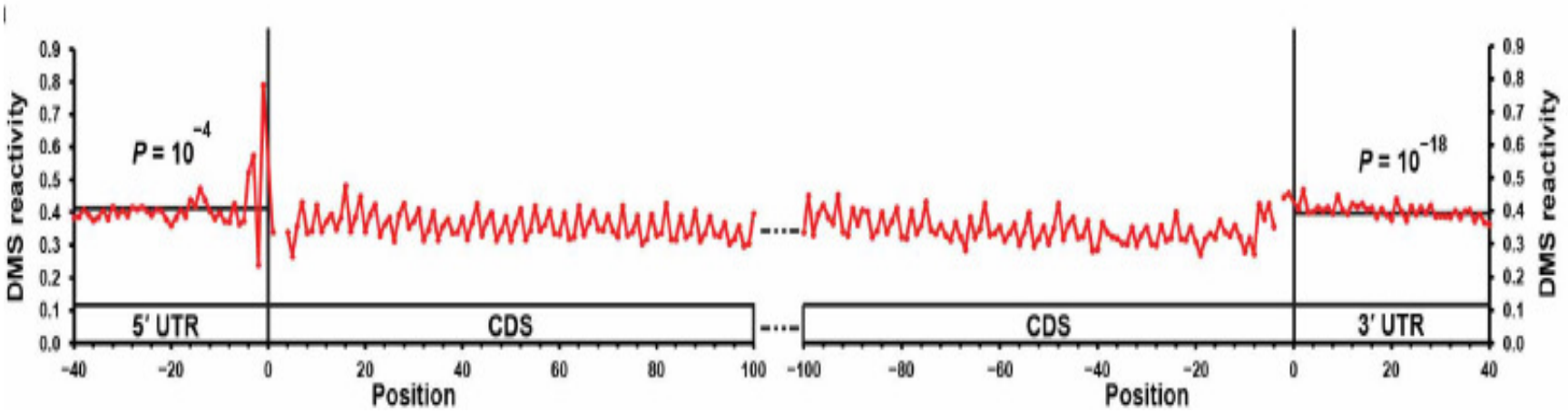
Substracting the normalized number of reverse transcriptase stop for the nucleotides between (+)DMS and (-)DMS libraries

Negative values are taken as 0

$$\theta(i) = \max((\mathrm{P}(i) - \mathrm{M}(i)), 0)$$

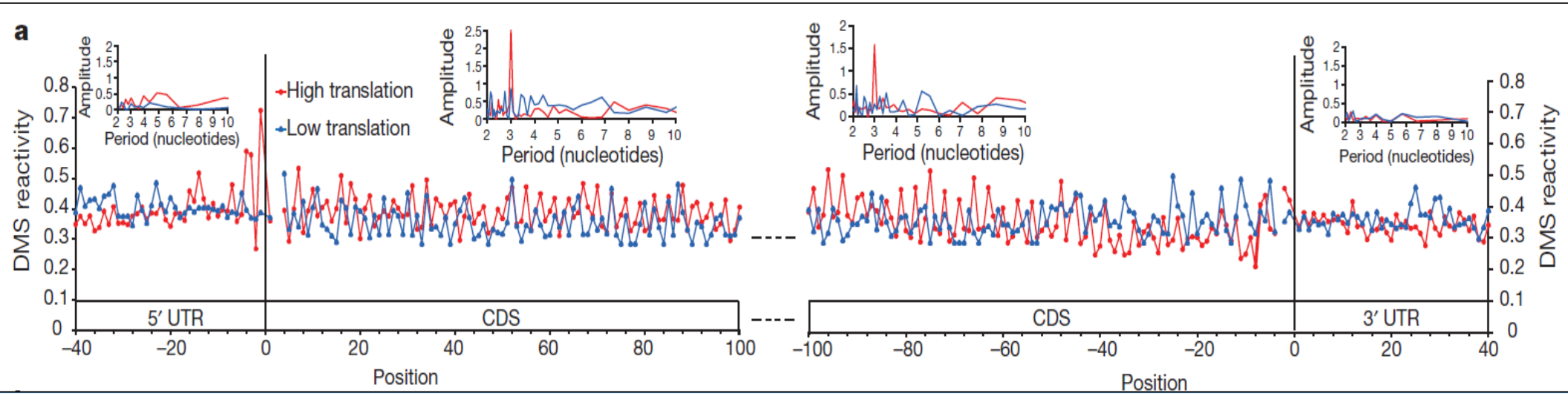
Results

- The average DMS reactivity of untranslated regions (UTRs) is significantly higher than that of coding sequences (CDS)



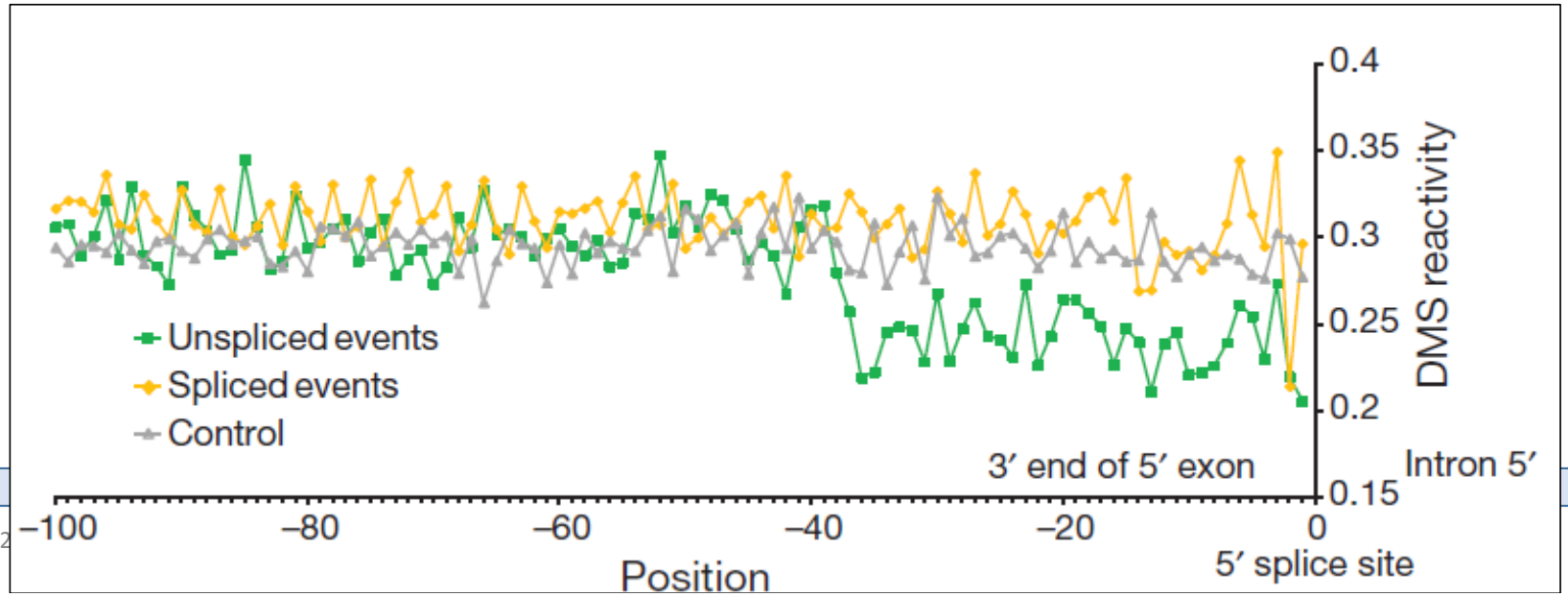
Results

- Averaging DMS reactivity along the CDS across mRNAs in this data set reveals a periodic trend
- Discrete Fourier transformation applied to the CDS yielded a period of 3
- Periodicity was absent in UTR regions



Results

- RNA secondary structure → Alternative splicing
- Considering a previous compilation of alternative splicing events in Arabidopsis seedlings
- Identified for each mRNA in data set whether alternative splicing occurred

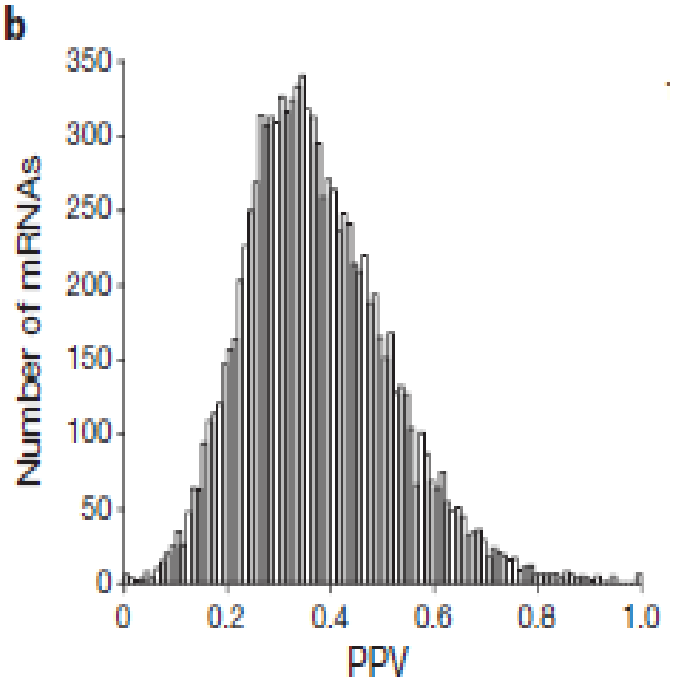


Results

- Using a current in silico structure prediction (RNA structure), a set of probable RNA structures was estimated
- For each mRNAs the positive predictive value (PPV) was calculated
- PPV indicates the proportion of base pairs in the in vivo DMS-RNA structure that also appears in the in silico predicted RNA structure (The number of true positives divided by the total number of positives)

Results

- Higher PPV value indicates less difference
- Most mRNAs did not fold in vivo according to in silico-predicted structures
- The poor correlation could be explained by mRNA association with proteins that block DMS reactivity in vivo



Discussion

- Structure-seq:

  - provides a broadly applicable method for the investigation of RNA structure–function relationships in living systems

  - Apart of this the experimental data can help prediction algorithms in different ways

Discussion

- For energy based secondary structure prediction methods they can improve the energy parameter

- They can be incorporated into dynamic programming algorithms as cells that do not have to be computed in the Dynamic Programming matrix. This decreases the runtime of the algorithms!

Source

[1] Advances in RNA structure analysis by chemical probing,
Kevin M Weeks

Thank you

For your attention