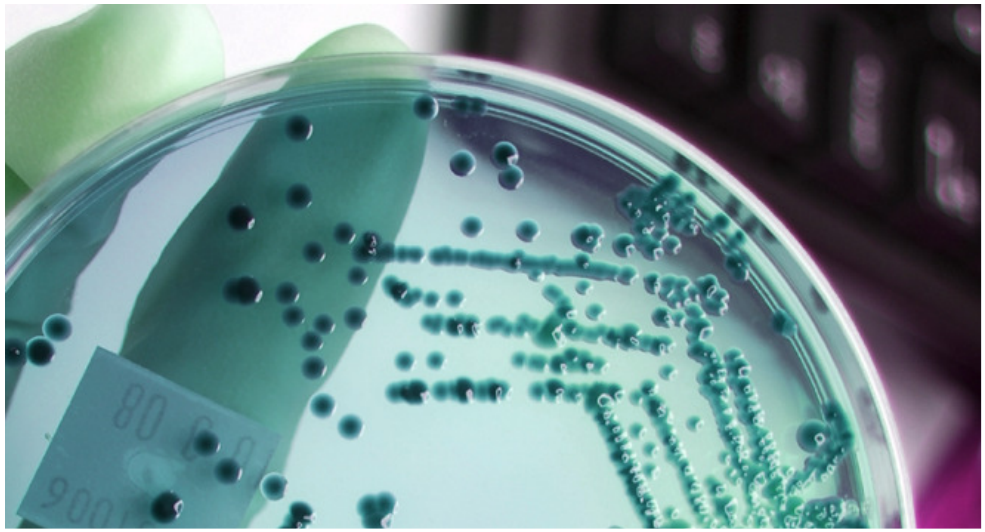
Era7 bioinformatics

RNA

www.era7bioinformatics.com



NATURE

The evolution of lncRNA repertoires and expression patterns in tetrapods

A large-scale evolutionary study of lncRNA repertoires and expression patterns, in 11 tetrapod species. We identify approximately 11,000 primate-specific lncRNAs and 2,500 highly conserved lncRNAs, including approximately 400 genes that are likely to have originated more than 300 millions years ago. We find that lncRNAs, in particular ancient ones, are in general actively regulated and may function predominantly in embryonic development. Most lncRNAs evolve rapidly in terms of sequence and expression levels, but tissue specificities are often conserved

Man is the only animal that stumbles twice in the same stone.....

And if the man is a scientist .....

Not twice.... "N" times

Although several highly conserved lncRNAs are known, lncRNAs generally have modest sequence conservation. Furthermore, in mouse liver, lncRNA transcription undergoes rapid evolutionary turnover.

These observations suggest that many lncRNAs may have no biological relevance.

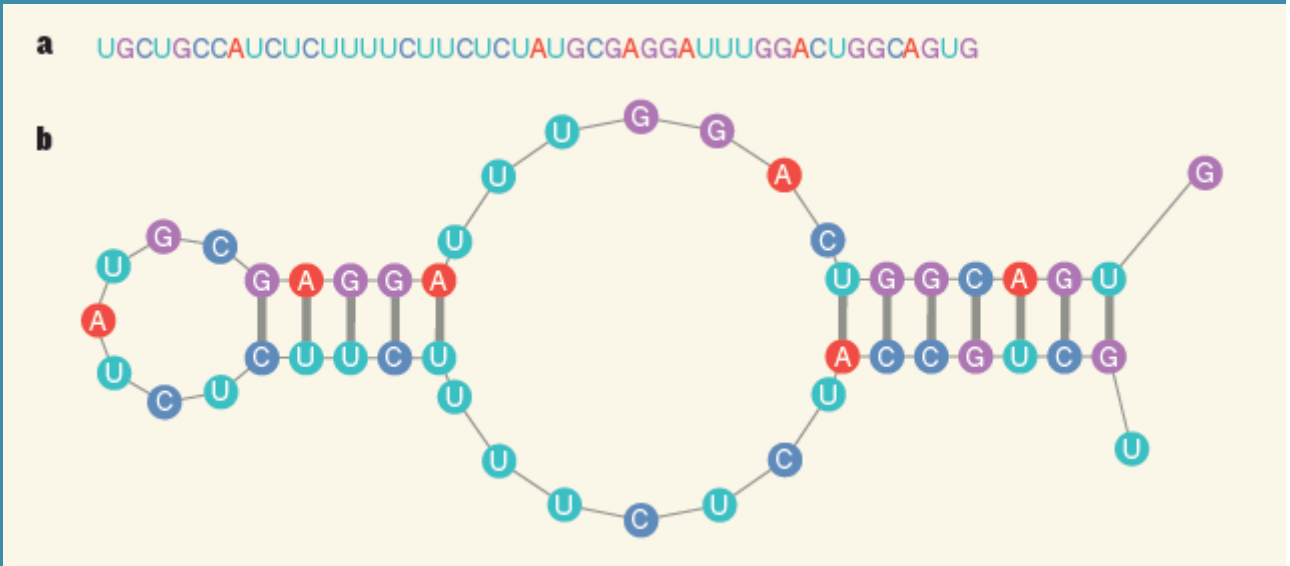
Detailed evolutionary analyses can clarify lncRNA functionality, but such analyses have been hampered by lack of annotations in non-model organisms

NATURE

Three reports describe analyses of all the mRNA molecules present in different populations of cells — transcriptome-wide analyses — using structure-probing techniques. These studies begin to reveal the extent of secondary structure in the transcriptomes of plants, humans and yeast

The nucleotides of RNA are free to interact with one another within each molecule, resulting in folding of the RNA chain into secondary structures

RNA is a single-stranded polymer, with nucleotide bases adenine (A), cytosine (C), guanine (G) and uracil (U). b, Unlike DNA, RNA molecules do not pair up to form helices. The bases of an individual molecule can pair with one another (G–C and A–U), causing the RNA to fold into secondary structures. G bases can also pair with U bases, forming a G–U wobble pair



The three new studies, each analysing different cell populations, use a combination of:

- Structure-probing techniques for determining RNA secondary
- Next Generation Sequencing

Each group reports that some of the RNA structures they observed in vitro were altered in vivo. In fact, Rouskin and colleagues found evidence in yeast that

RNA structures in the cell are actively unfolded by proteins.

Nonetheless, the papers show that

structural patterns are evolutionarily conserved at several functional sites within RNA molecules

Wan and co-workers probing in cell lines derived from a family trio (mother, father and child) found human inter-generational genetic variation on the transcriptome, and discovered more than 1,900 single-nucleotide mutations that alter RNA structure.

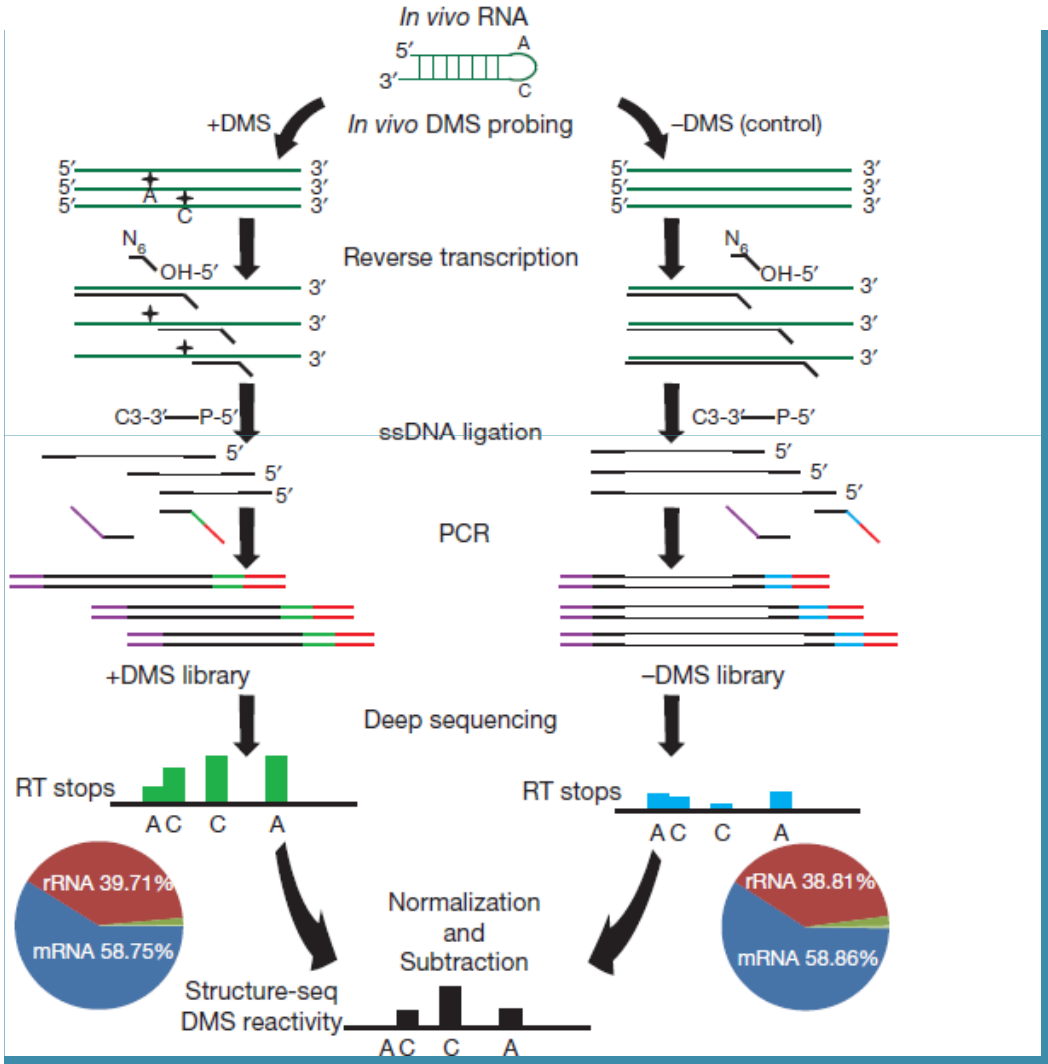
These experiments yielded thousands of new putative 'ribosnitches'— broadly defined as RNA sequences in which a specific single-nucleotide mutation alters structure. Ribosnitches are analogous to bacterial riboswitches, which change structure on binding of a small molecule and regulate transcription or translation.

Because RNA structure has the potential to influence post-transcriptional processes in the cell, a subset of the putative ribosnitches could be functional. Indeed, mutations that disrupt certain RNA secondary structural elements can cause human disease

The application of next-generation sequencing to the transcriptome has previously revealed the complexity of post-transcriptional regulatory networks. The structural dimension of this complexity is now accessible with the publication of these three papers

There are key differences in the specific features found by each approach. Such discrepancies may come from differences in experimental design, which can cause changes to the inherently dynamic structure of RNA. In this case, each study used different protocols for RNA extraction, library preparation and, in particular, determining levels of background noise. These experimental details must be taken into account when comparing structures discovered using the different approaches

Full characterization of transcriptome structure will require a concerted community effort, with an emphasis on standardization to allow quantitative comparisons of these data sets. Only then will it be possible to fully integrate these findings to determine the structural elements that are consequential in the transcriptome



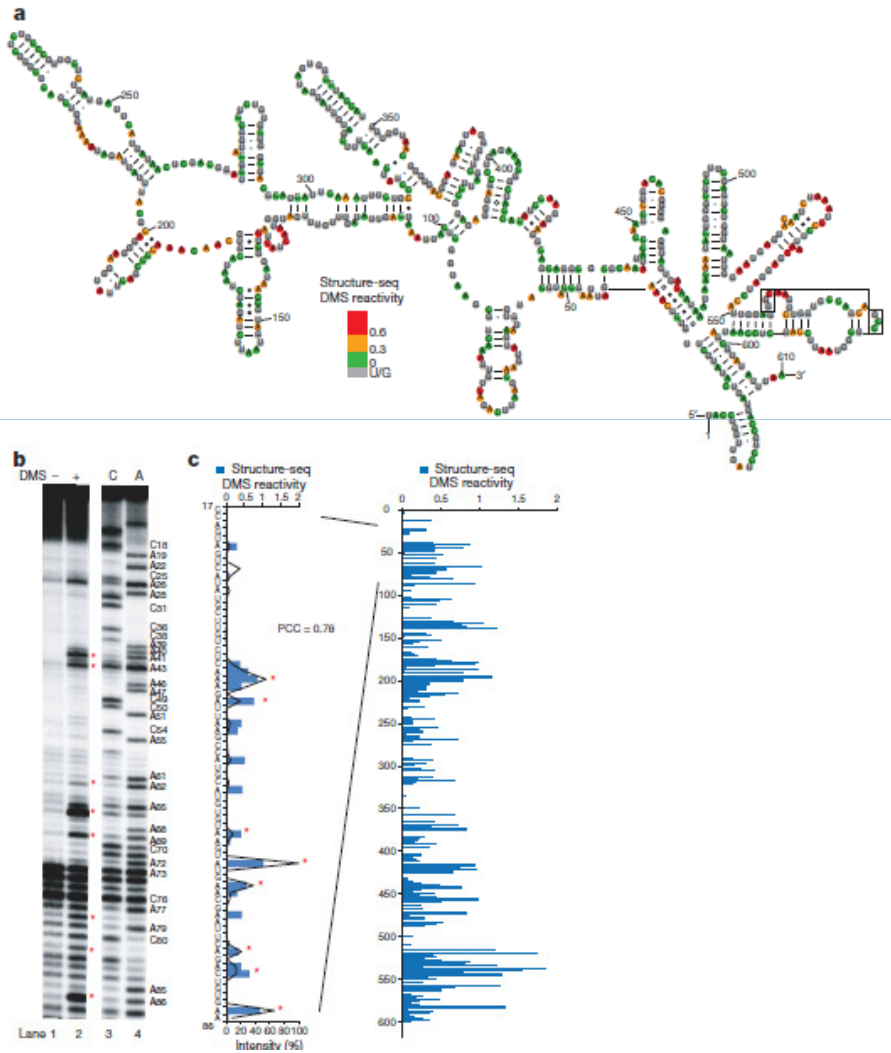
In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features.

Arabidopsis seedlings (brotes)

RNA structure probing reagents, dimethyl sulphate (DMS) can penetrate cells and has been used to map structures of high-abundance RNAs in vivo in various organisms. DMS methylates the base-pairing faces of A and C of RNA in loops, bulges, mismatches and joining regions

To validate in vivo structure-seq, we mapped DMS reactivities of 18S rRNA . Overall, the reactivities are consistent with structure mapping of 30 subunit-bound 16S rRNA. Conventional gel-based in vivo structure probing yielded strong agreement for all regions of 18S rRNA tested

RNA structures in vivo on a genome-wide basis. Importantly, complete coverage can be provided in a single experiment even for long transcripts, which is not the case for conventional gel-based methods



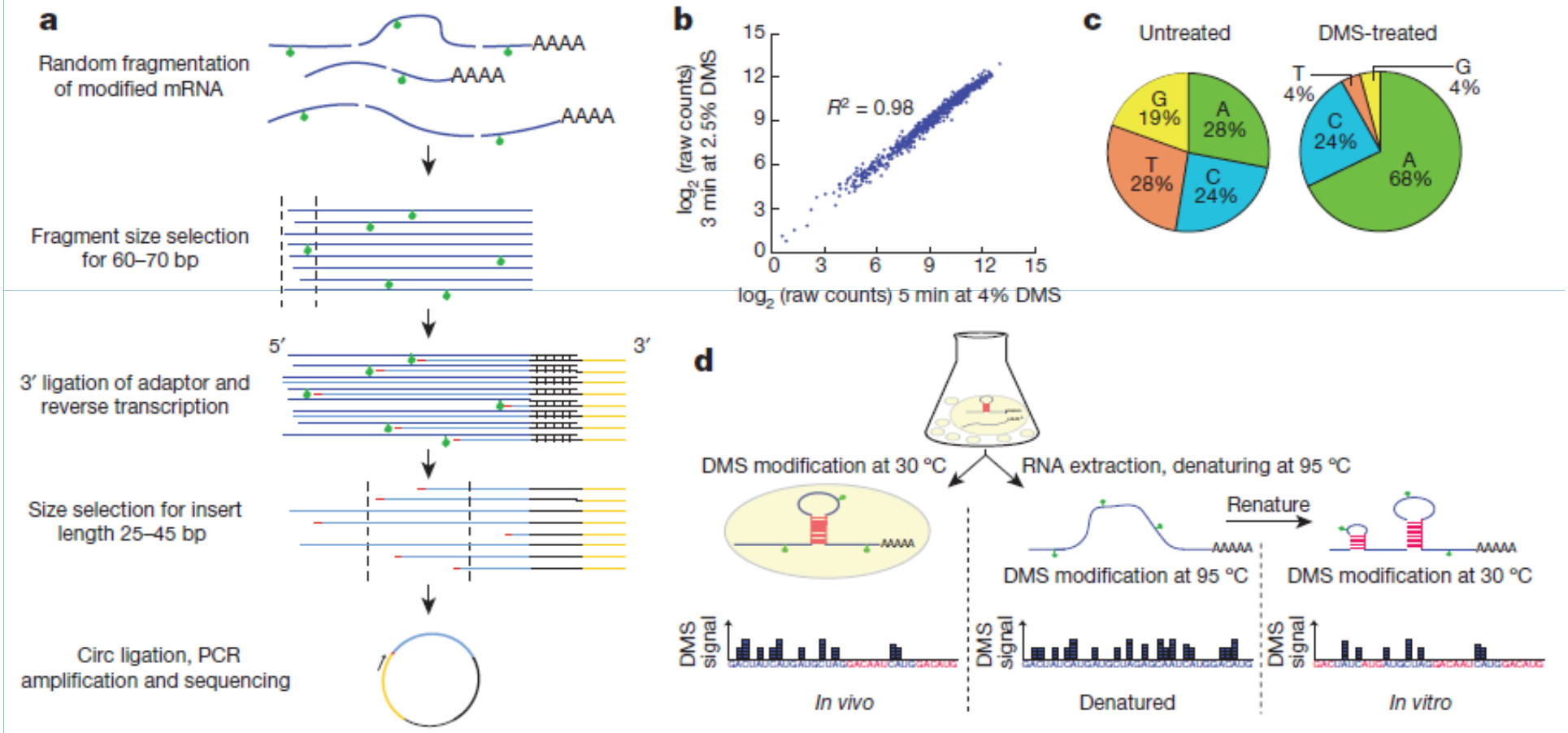
Genome-wide in vivo RNA structural properties of Arabidopsis mRNAs. We found that the average DMS reactivity of untranslated regions (UTRs) is significantly higher than that of coding sequences (CDS). The 5 nucleotides (nt) immediately upstream of the start codon show particularly high DMS reactivity, which indicates less structure

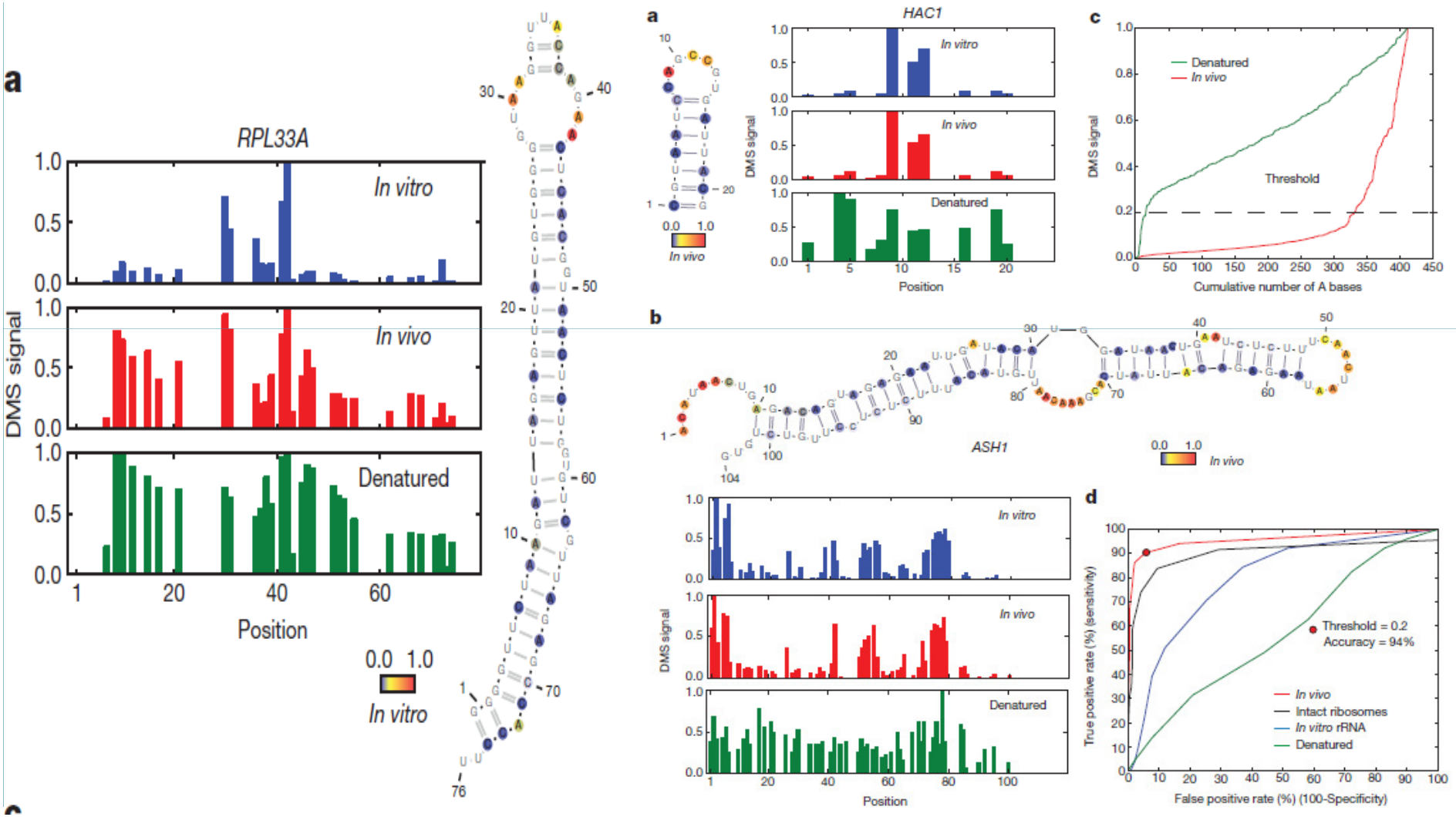
The unstructured region upstream of the start codon was enriched in high translation efficiency mRNAs and was absent in low translation efficiency mRNAs

Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo

Yeast

DMS is highly reactive with solvent-accessible, unpaired residues but reliably unreactive with bases engaged in Watson–Crick interactions, thus nucleotides that are strongly protected or reactive to DMS can be inferred to be base-paired or unpaired, respectively


a
Random fragmentation of modified mRNA
AAAA
AAAA
AAAA
Fragment size selection for 60–70 bp
5′
3′
3′ ligation of adaptor and reverse transcription
Size selection for insert length 25–45 bp
Circ ligation, PCR amplification and sequencing
b
log2 (raw counts) 3 min at 2.5% DMS
R2 = 0.98
log2 (raw counts) 5 min at 4% DMS
c
Untreated
DMS-treated
G 19%
A 28%
T 28%
C 24%
T 4%
G 4%
C 24%
A 68%
d
DMS modification at 30 °C
RNA extraction, denaturing at 95 °C
Renature
DMS modification at 95 °C
DMS modification at 30 °C
DMS signal
In vivo
Denatured
In vitro


a
RPL33A
In vitro
In vivo
Denatured
DMS signal
Position
0.0 1.0
In vitro
HAC1
In vitro
In vivo
Denatured
DMS signal
Position
In vivo
c
Denatured
In vivo
Threshold
DMS signal
Cumulative number of A bases
b
ASH1
In vivo
In vitro
In vivo
Denatured
DMS signal
Position
d
Threshold = 0.2
Accuracy = 94%
In vivo
Intact ribosomes
In vitro rRNA
Denatured
True positive rate (%) (sensitivity)
False positive rate (%) (100-Specificity)

Era7 bioinformatics

Thanks
for your attention!

Eduardo Pareja

epareja@era7.com

www.era7bioinformatics.com