Landscape and variation of RNA secondary structure across the human transcriptome

Yue Wan, Kun Qu, Qiangfeng Cliff Zhang, Ryan A. Flynn, Ohad Manor, Zhengqing Ouyang, Jiajing Zhang, Robert C. Spitale, Michael P. Snyder, Eran Segal & Howard Y. Chang

Nature 30 JANUARY 2014

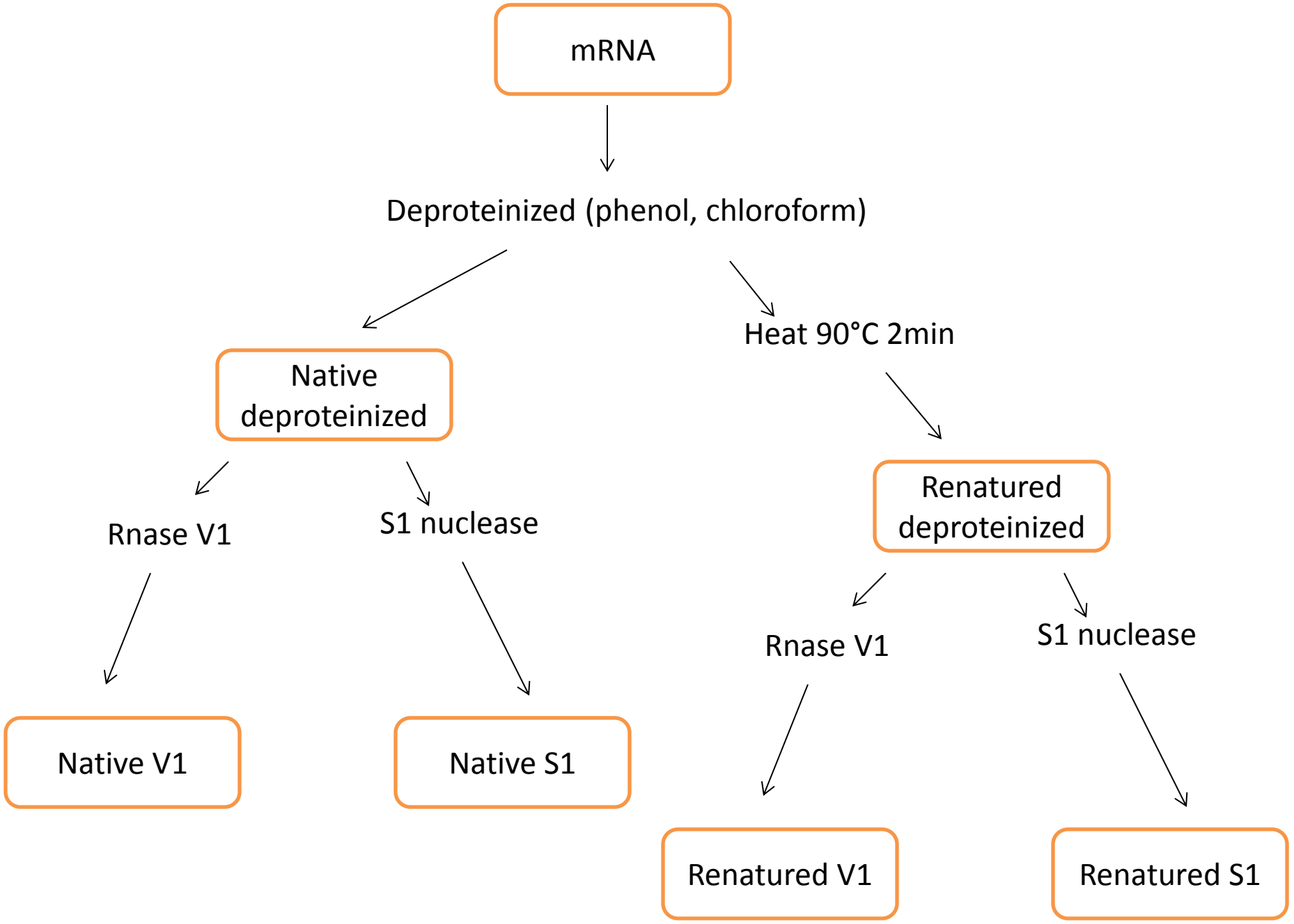
Mikk Puustusmaa

Introduction

- In parallel to the genetic code for protein synthesis, a **second layer of information** is embedded in all RNA transcripts in the form of RNA structure. RNA structure influences practically every step in the gene expression program

SAMPLE

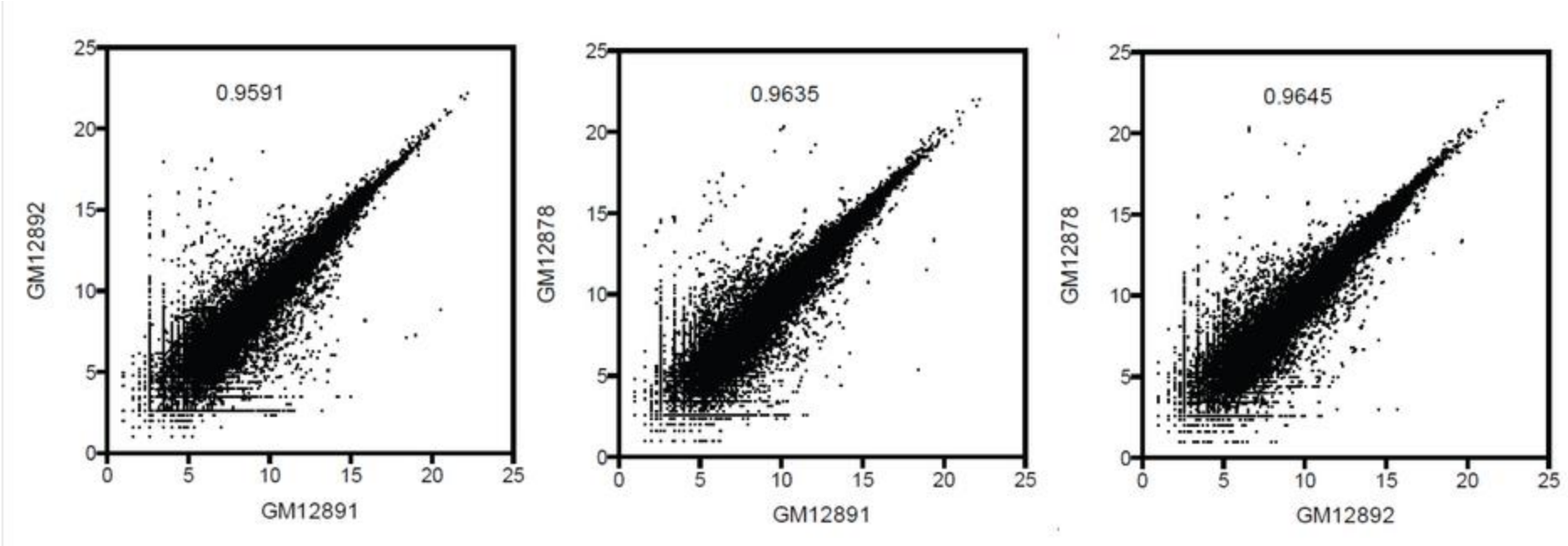
- Performed parallel analysis of RNA structure
  - Human lymphoblastoid cells (GM12878,GM12891,GM12892)
  - poly(A)+ RNA was obtained(mRNA)

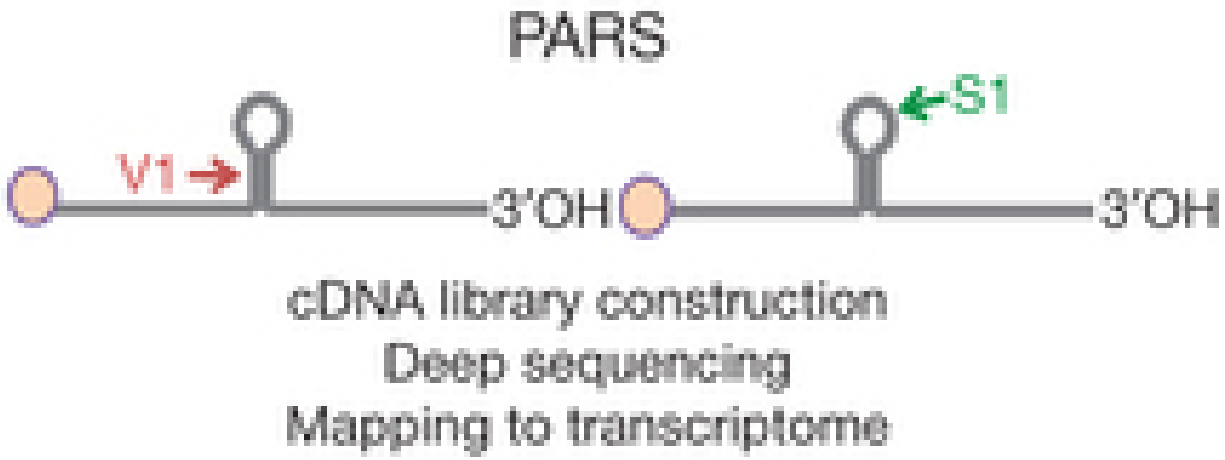

mRNA
Deproteinized (phenol, chloroform)
Native deproteinized
Heat 90°C 2min
Rnase V1
S1 nuclease
Renatured deproteinized
Native V1
Native S1
Rnase V1
S1 nuclease
Renatured V1
Renatured S1

Illumina sequencing and mapping

- Paired-End sequencing
- Raw reads were truncated to **50 bases** (51 bases from the 3’ end were trimmed)
- Mapped to human transcriptome with **Bowtie2**. Non-redundant transcripts from UCSC RefSeq and the Gencode v12 (hg19 assembly)
- They obtained **166- to 212-million mapped reads** for a Rnase v1 or S1 nuclease sample

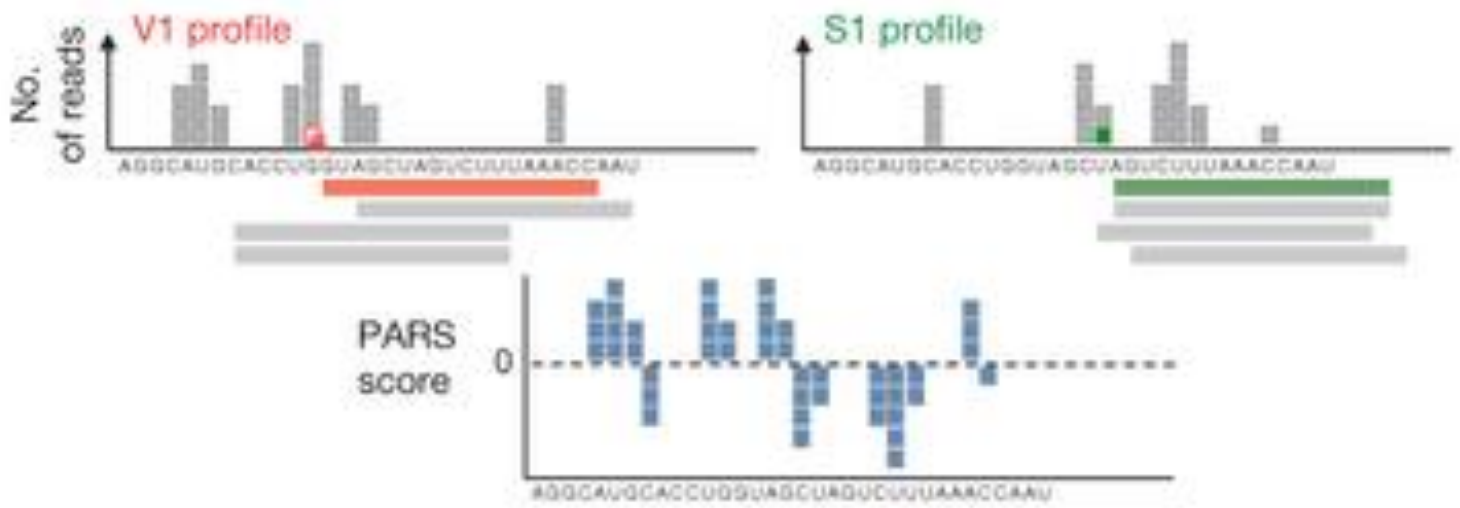
Transcript abundance




PARS
V1
S1
3'OH
3'OH
cDNA library construction
Deep sequencing
Mapping to transcriptome

PARS-score

- Parallel analysis of RNA structure (PARS)
- calculated the number of double-stranded reads and single-stranded reads that initiated on each base on an RNA.



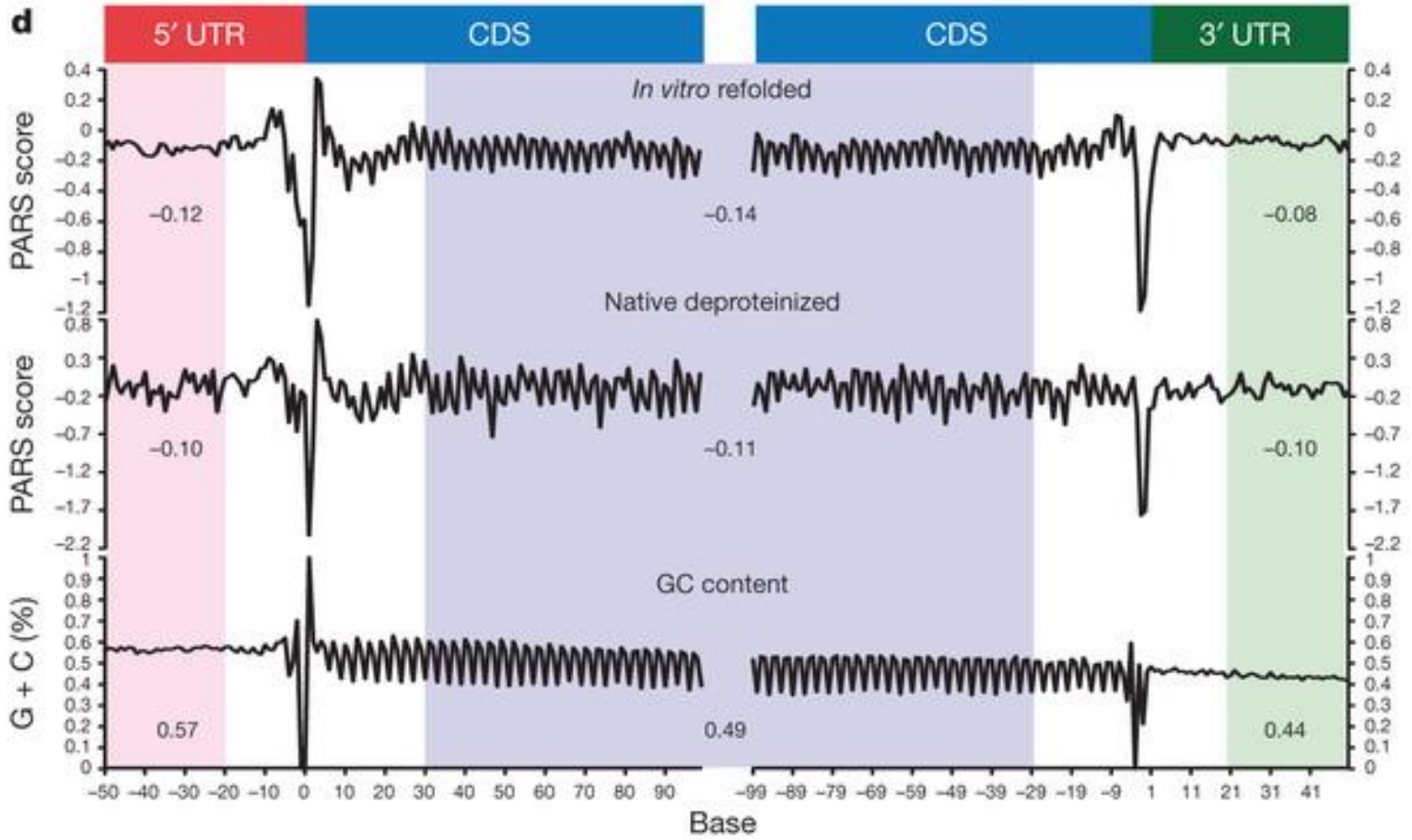
PARS-score

- The number of **double (V1)** and single stranded **reads (S1)** for each sequencing sample were then normalized by sequencing depth
- **Positive score** shows double stranded RNA
- **Negative score** shows single stranded RNA

$$\mathrm{PARS}_{i=1\ldots N} = \log_2(\mathrm{V1}_i + 5) - \log_2(\mathrm{S1}_i + 5)$$

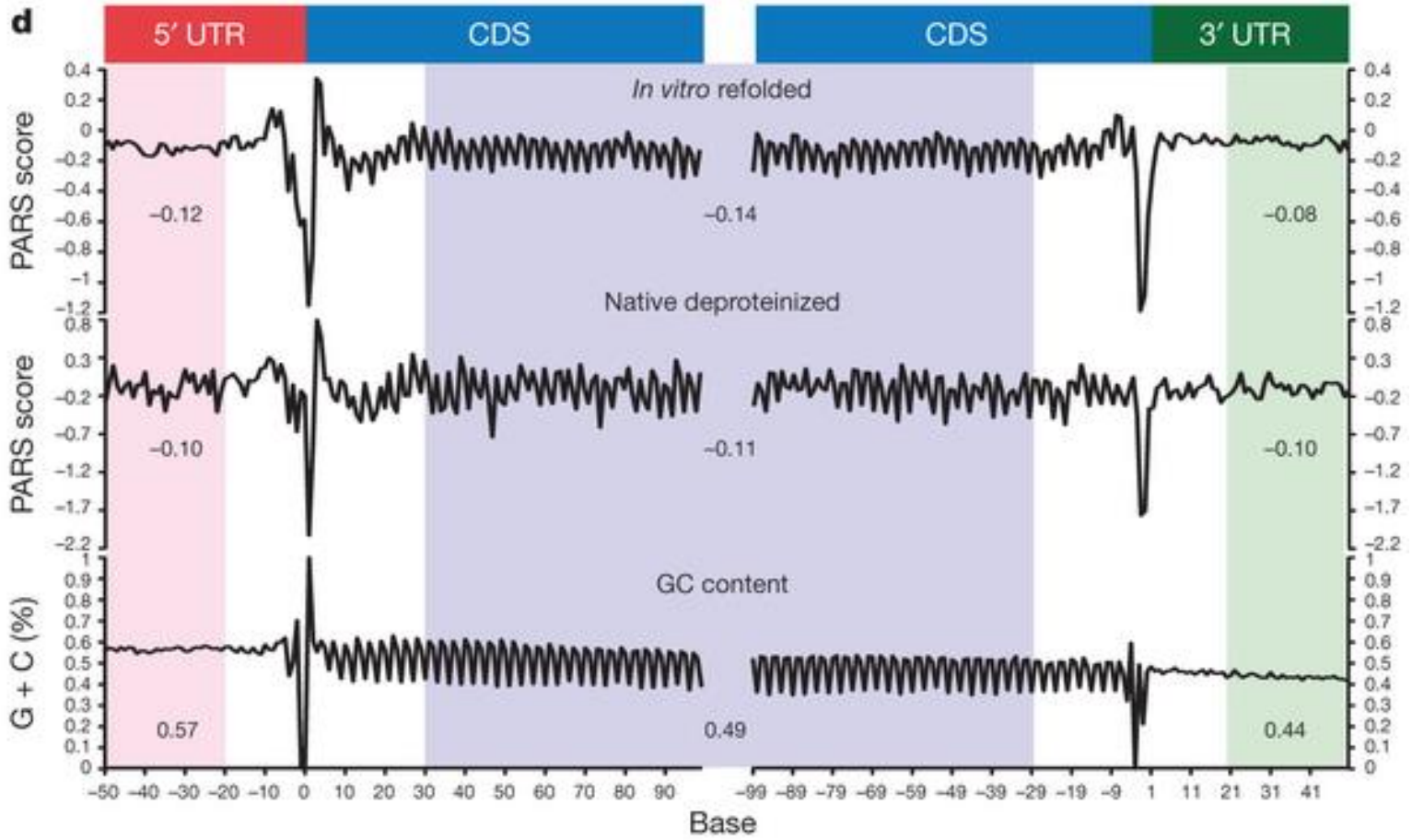
Result

- Both renatured and native mRNAs showed similar RSS features, suggesting that **RNA sequence is a strong determinant of RSS**



Result

- Metagene analysis show that, on average, CDS is more accessible near transational start site and stop codon



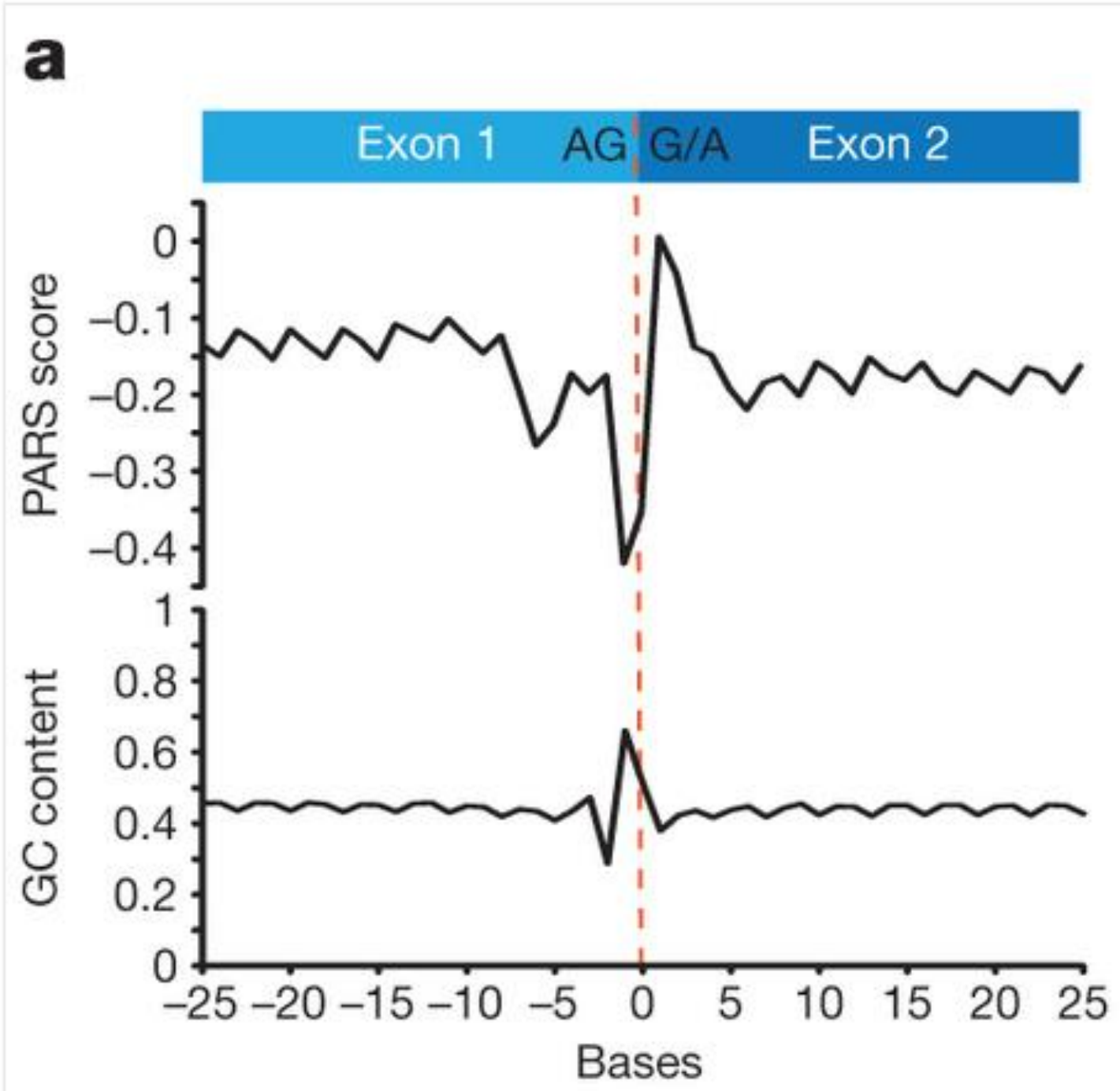
Result

- They also identified 583 (5.7%) consistently different regions between native deproteinized and renatured structure profiles, **providing candidate sites for regulation of RNA structure** *in vivo*

- They noted that 3.7% of bases (residing in 9.7% of transcripts) have both strong V1 and S1 reads, indicating the existence of **multiple mRNA conformations**

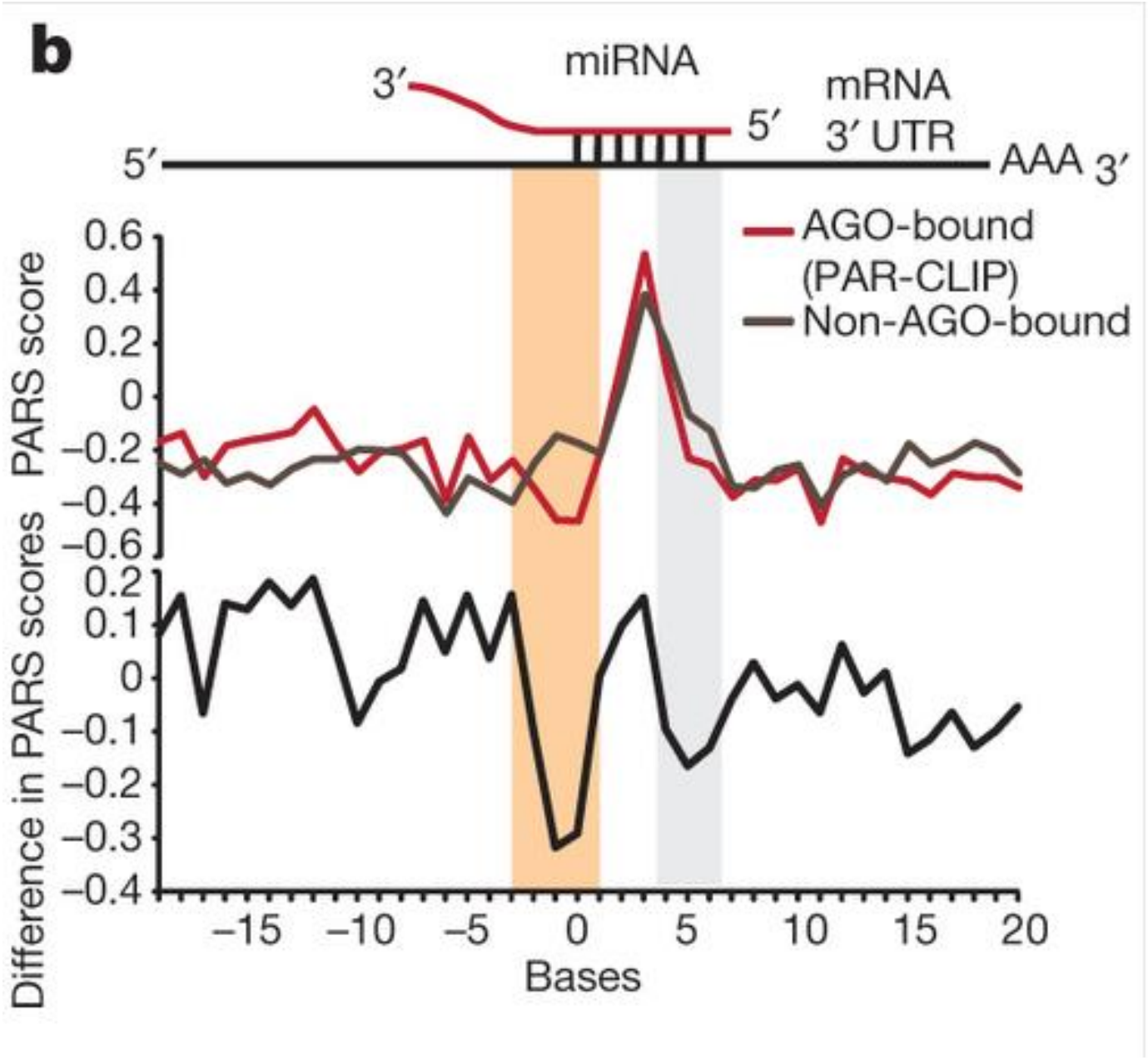
Result

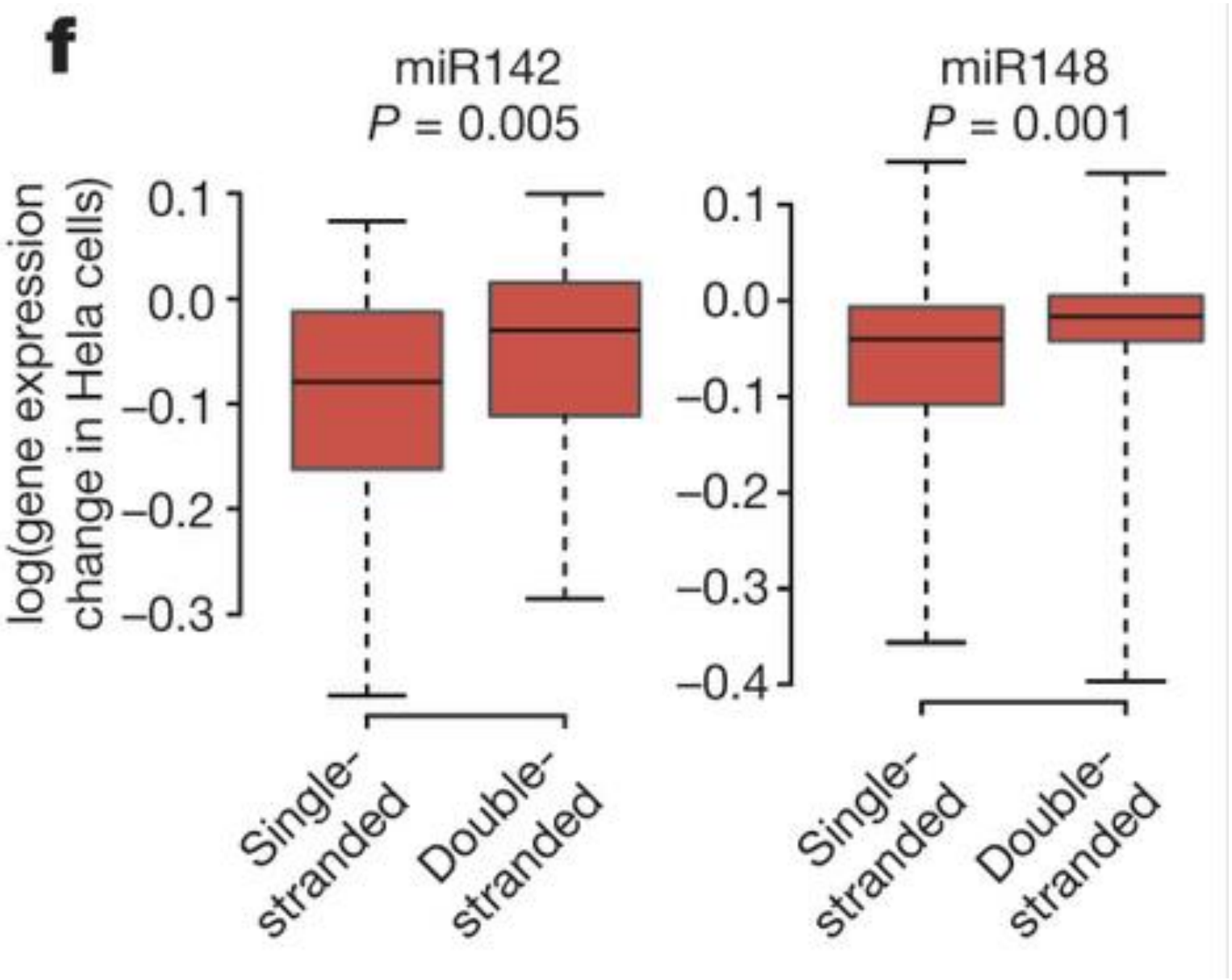
- They observed a unique asymmetric RSS signature at the **exon–exon junction** in both renatured and native deproteinized transcripts that is not simply explained by GC content. The terminal **AG dinucleotide** at the end of the 5′ exon tends to be **more accessible**, whereas the first nucleotides of **the 3′ exon are more structured**


a
Exon 1
AG
G/A
Exon 2
PARS score
0
−0.1
−0.2
−0.3
−0.4
GC content
1
0.8
0.6
0.4
0.2
0
−25 −20 −15 −10 −5 0 5 10 15 20 25
Bases

Rsults

- Analysis of RSS from renatured RNA around predicted miRNA targets revealed that true Argonaute **(AGO)-bound target sites** show strong structural accessibility from −1 to 3 nucleotides upstream of the miRNA-target site compared to predicted targets not bound by AGO


b
miRNA
3′
5′
mRNA
3′ UTR
5′
AAA 3′
AGO-bound
(PAR-CLIP)
Non-AGO-bound
PARS score
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
Difference in PARS scores
0.2
0.1
0
−0.1
−0.2
−0.3
−0.4
−15
−10
−5
0
5
10
15
20
Bases


f
miR142
P = 0.005
miR148
P = 0.001
log(gene expression change in Hela cells)
0.1
0.0
−0.1
−0.2
−0.3
−0.4
Single-stranded
Double-stranded
Single-stranded
Double-stranded

Part 2 riboSNitches

- Comparison of RNA structural landscapes between individuals revealed the **impact of diverse sequence variants on RNA structure**.

- SNVs that alter RNA structure, known as **‘riboSNitches’**

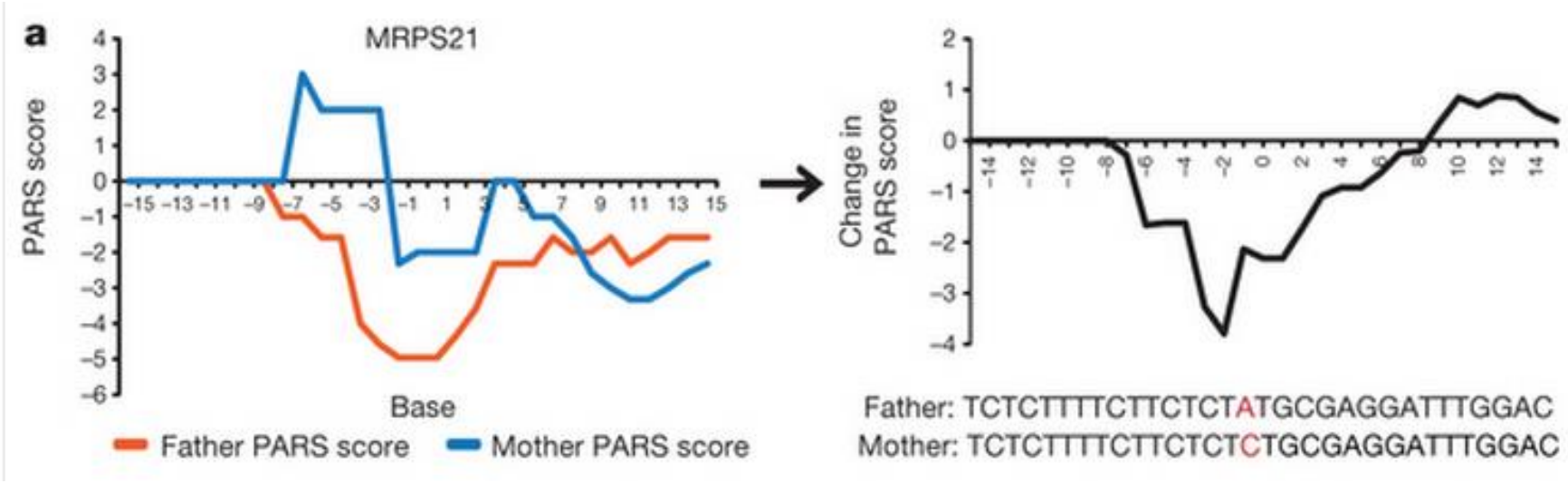
Part 2 riboSNitches

- Furthermore, **78.2%** of all structure changing bases lie in transcripts that contain either SNVs or indels, suggesting that sequence variation is important in shaping RSS variation in the human transcriptome

Part 2 riboSNitches

- To **pinpoint riboSNitches**, we calculated structure changes between each pair of individuals and selected SNVs that had large **PARS score differences, low false discovery rate** (FDR), **significant *P* value**, **and high local read coverage**

Part 2 riboSNitches



Result

- We found that **1,907** out of 12,233 (15%) SNVs switched RNA structure in the trio.
- They also **experimentally validated** nine riboSNitches using independent structure probing methods such as nucleases, selective 2′ hydroxyl acylation and primer extension (SHAPE) or dimethyl sulphate (DMS), and **confirmed the ability of PARS to discover riboSNitches**

Result

- Out of **172** parental homozygous riboSNitches, **117** (68%) were validated by allele-specific mapping in the child. **As only reads upstream of the riboSNitch can be uniquely mapped and detected, this is likely to be an underestimate**. We also observed a validation rate of 61% in native deproteinized samples of the child, indicating that the structural changes are biologically relevant *in vivo*

Result

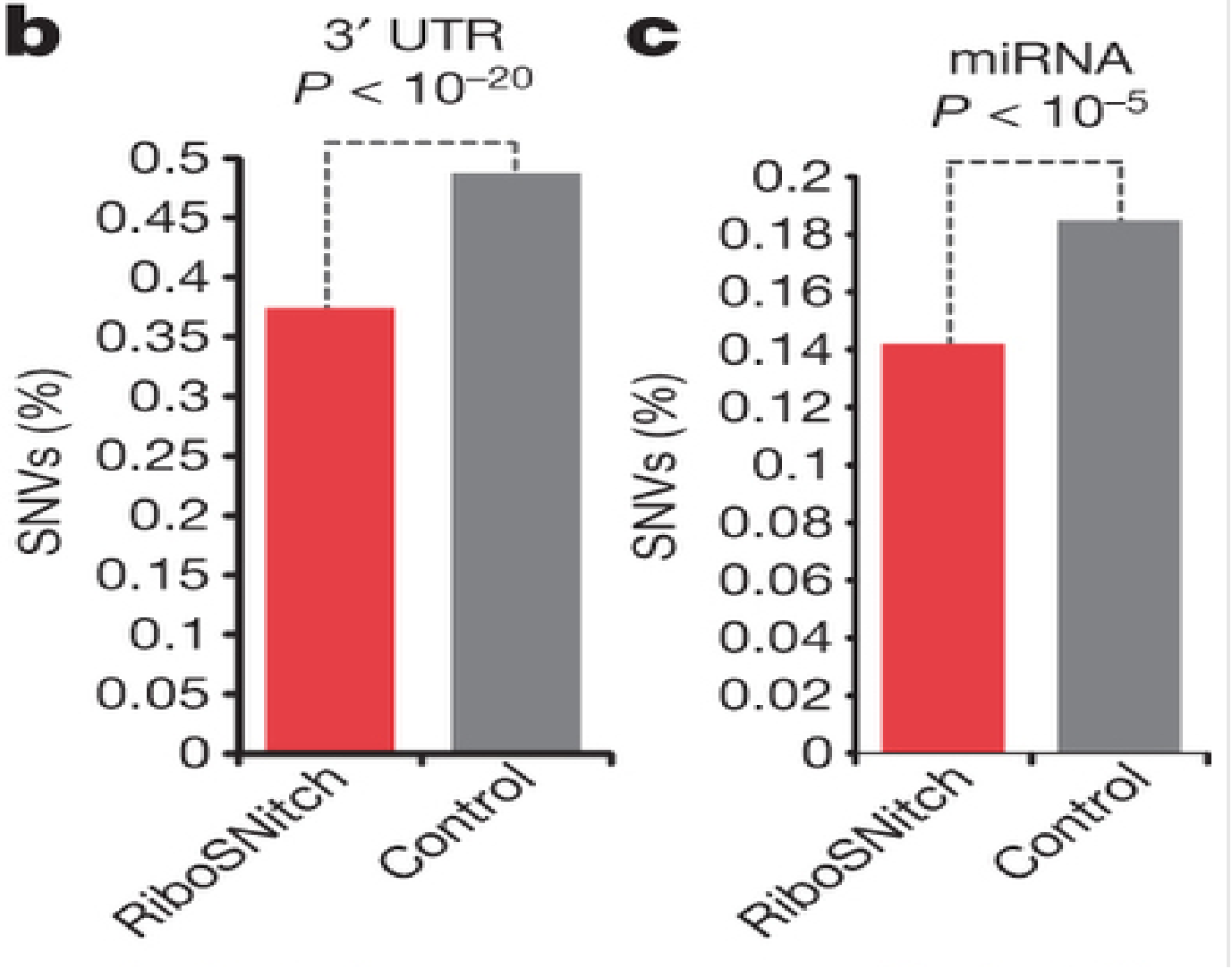
- Intersection with expression quantitative trait loci (eQTL) identified **211** riboSNitches that are associated with changes in gene expression.

sequence and context rules in riboSNitches

- **First**, riboSNitches that lie in double- or single-stranded regions tend to become more single- or double-stranded, respectively, after nucleotide change.
- **Second**, the nucleotide content of the riboSNitch is instructive of the direction of RSS change. Bases that undergo G/C to A/T changes tend to become more single-stranded, whereas bases that change from A/T to G/C tend to become more paired

sequence and context rules in riboSNitches

- **Third**, the structural context flanking SNVs influence their transition to become more single- or double-stranded
- **Fourth**, riboSNitches have fewer SNVs around them as compared to non-structure changing SNVs, suggesting that co-variation of some SNVs may help to maintain functional RNA structures



- riboSNitches are also significantly depleted at 3’ UTR compared to control (structurally synonymous SNV) and also around predicted miRNA target sites

RiboSNitches

- RiboSNitches may also influence gene regulation through splicing. Indeed, riboSNitches near splice junctions are associated with greater alternative splicing changes ,suggesting that **RNA structures could regulate splicing.**

Conclusion

- They identified unique RSS signatures that demarcate **open reading frames** and **splicing junctions**, **and define authentic microRNA-binding sites**.
- Majority of the RSS information is encoded within RNA sequence.
- Over **1,900** transcribed single nucleotide variants (approximately 15% of all transcribed single nucleotide variants) **alter local RNA** structure.
- They discovered **simple sequence and spacing rules** that determine the ability of point mutations to impact RSSs.
- Selective depletion of **'riboSNitches**' versus structurally synonymous variants at precise locations suggests selection for specific RNA shapes at thousands of sites, including 3'UTR, binding sites of microRNAs and RNA-binding proteins genome-wide.

Conclusion

- These results highlight the potentially broad contribution of RNA structure and its variation to gene regulation.

Tänan kuulamast!