m^{6}A in Arabidopsis

Maurits Evers
12 May 2016

m^{6}A data

Luo et al.

From Yalin:

- Day 21 Arabidopsis. (above ground tissue harvested between 5th & 7th hour of the 16-hour light/8-hour dark light cycle)

- mRNA extraction by FastTrack MAG Maxi mRNA isolation kit. Fragment to ~200nt, incubate with m6A Ab, then do IP, eluate RNA

- Illumina HiSeq, 2x100 cycles Solexa PE sequencing

- Sequencing data analysis is based on PMID: 22608085

- m6A peak window 25nt. Also determined peak clusters

- m6A is enriched around stop codon, within 3'UTR, and also uniquely in plants-around start codon; highly enriched in chloroplast/plastid

Focus on 4,302 m^{6}A peaks

RNAModR analysis

- Build TAIR10-based reference genome

- Collapse gene isoforms:
    - Keep longest CDS
    - Keep longest adjoining 5'/3'UTRs

- Map sites to collapsed reference transriptome

- Generate null sites: All non-methylated adenosines in transcripts that contain at least one m^6A site

- All enrichment analyses test how likely it is so see an observed odds ratio of m^6A/null-m^6A as extreme or more extreme, **conditional on the odds ratio being equal to 1 (i.e. the null hypothesis)**.

- Validity of empirical null site distribution? Tricky question.
   E.g. Non-uniformity in transcript (section) coverage will affect the validity of the null.

Treat all meta-gene analyses with a (large) grain of salt.

RNAModR analysis: m^{6}A peak width distribution

RNAModR:
m^{6}A sites are the
"uncollapsed" peaks
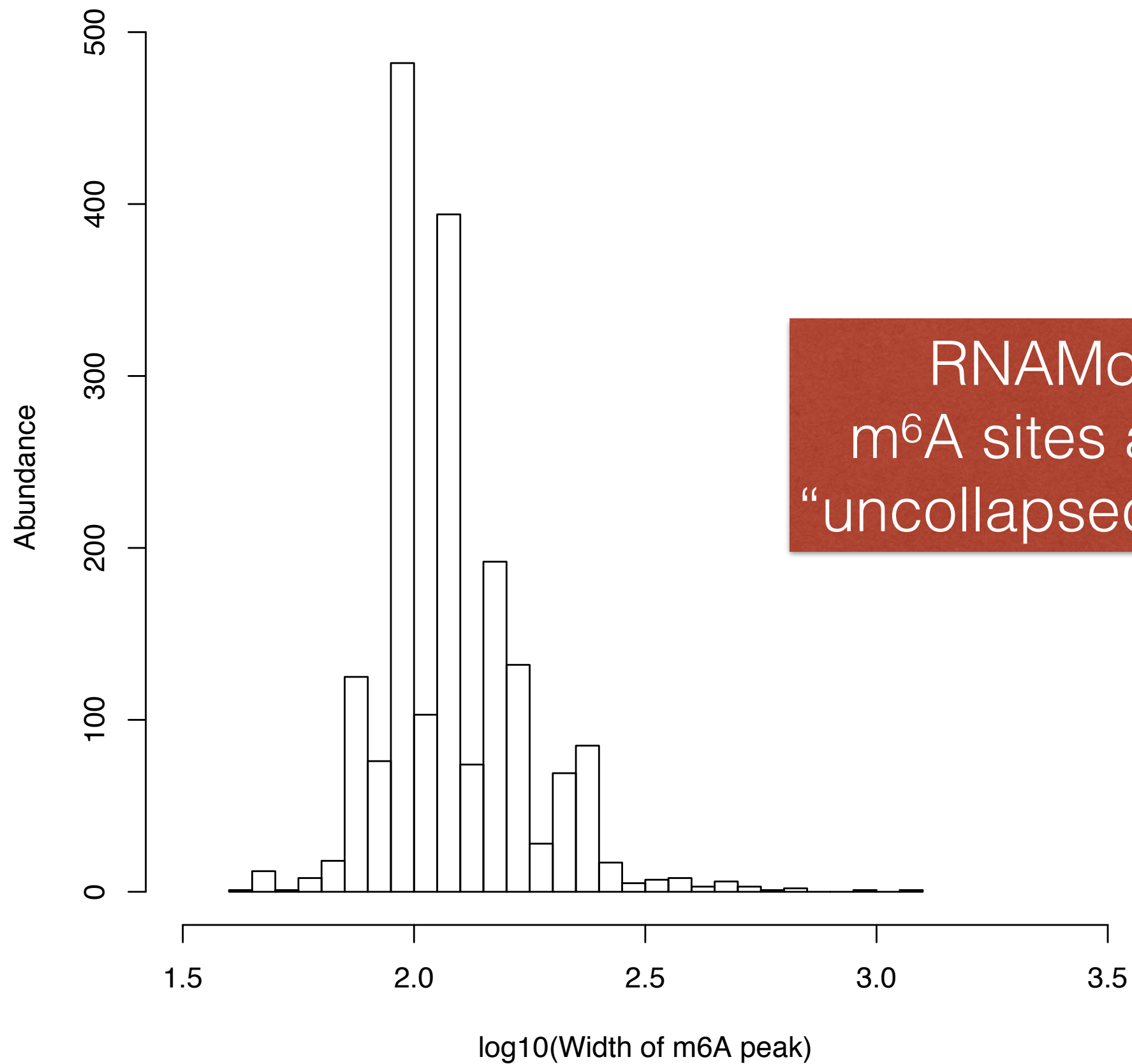
RNAModR analysis: m^{6}A distribution across transcript sections

Distribution of 1854 m6A sites across transcript sections

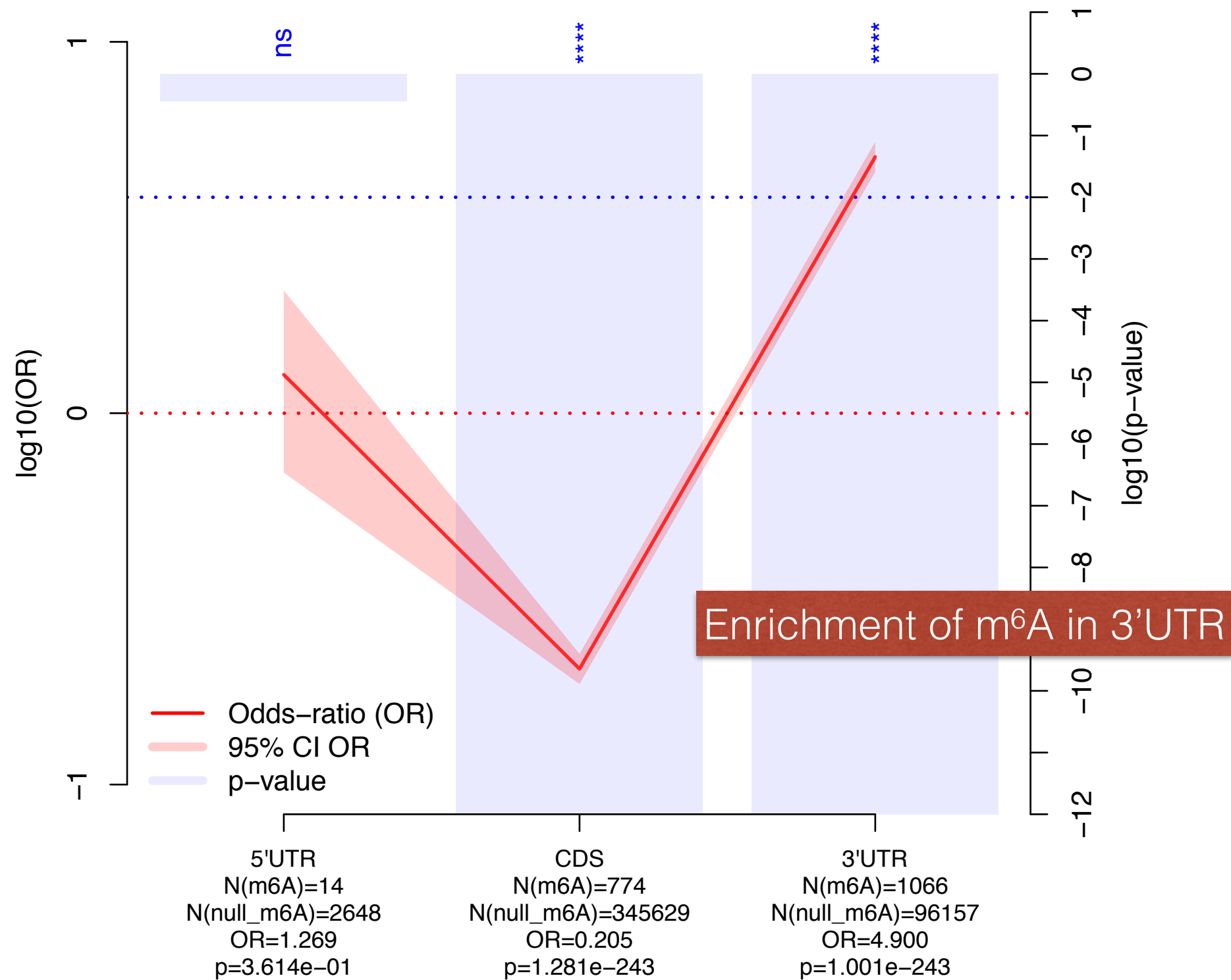

CDS 41.7% (774)

5'UTR 0.8% (14)

3'UTR 57.5% (1066)
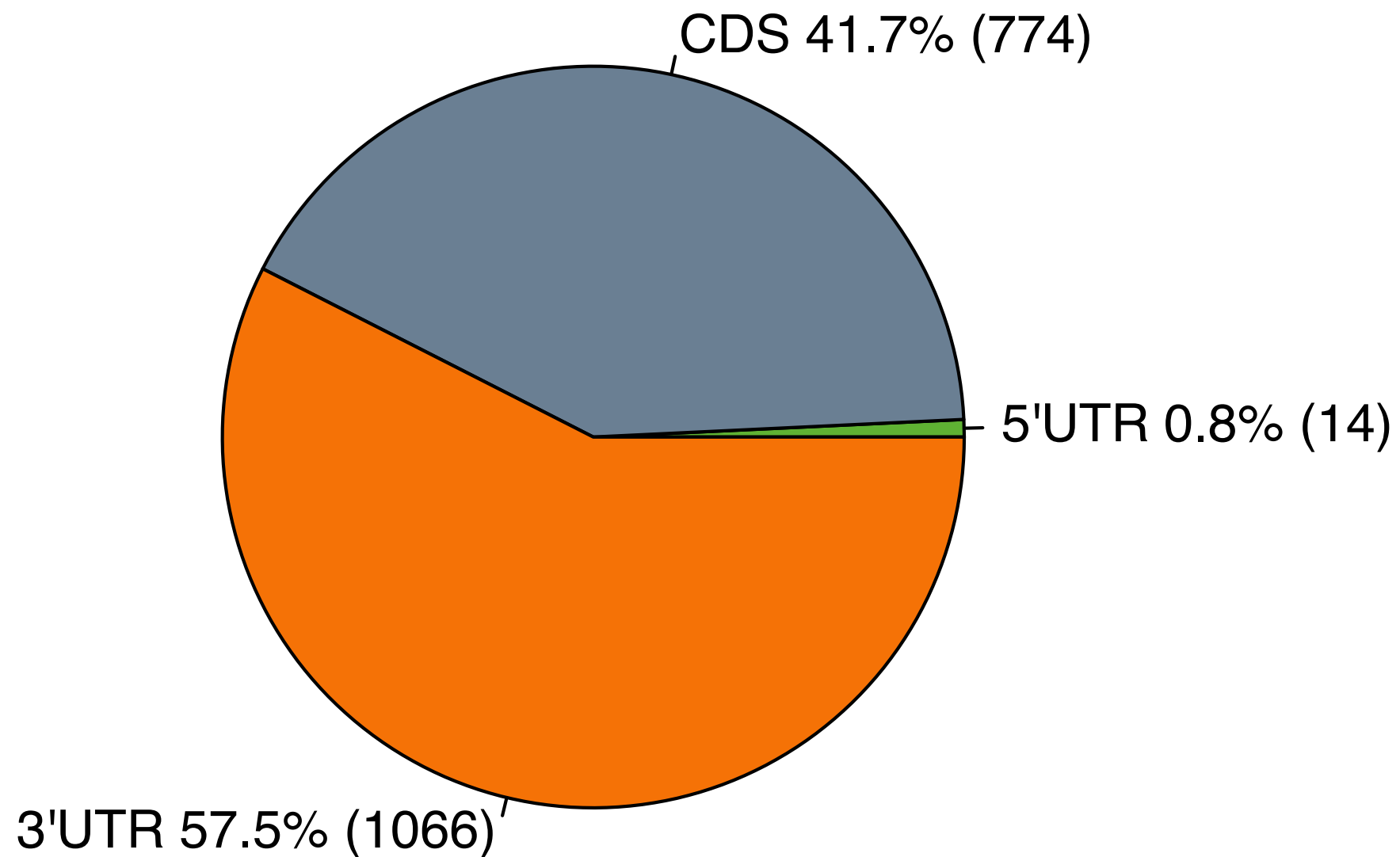
RNAModR analysis: m⁶A distribution across transcript sections

Distribution of 444434 null_m6A sites across transcript sections

CDS 77.8% (345629)

5'UTR 0.6% (2648)

3'UTR 21.6% (96157)
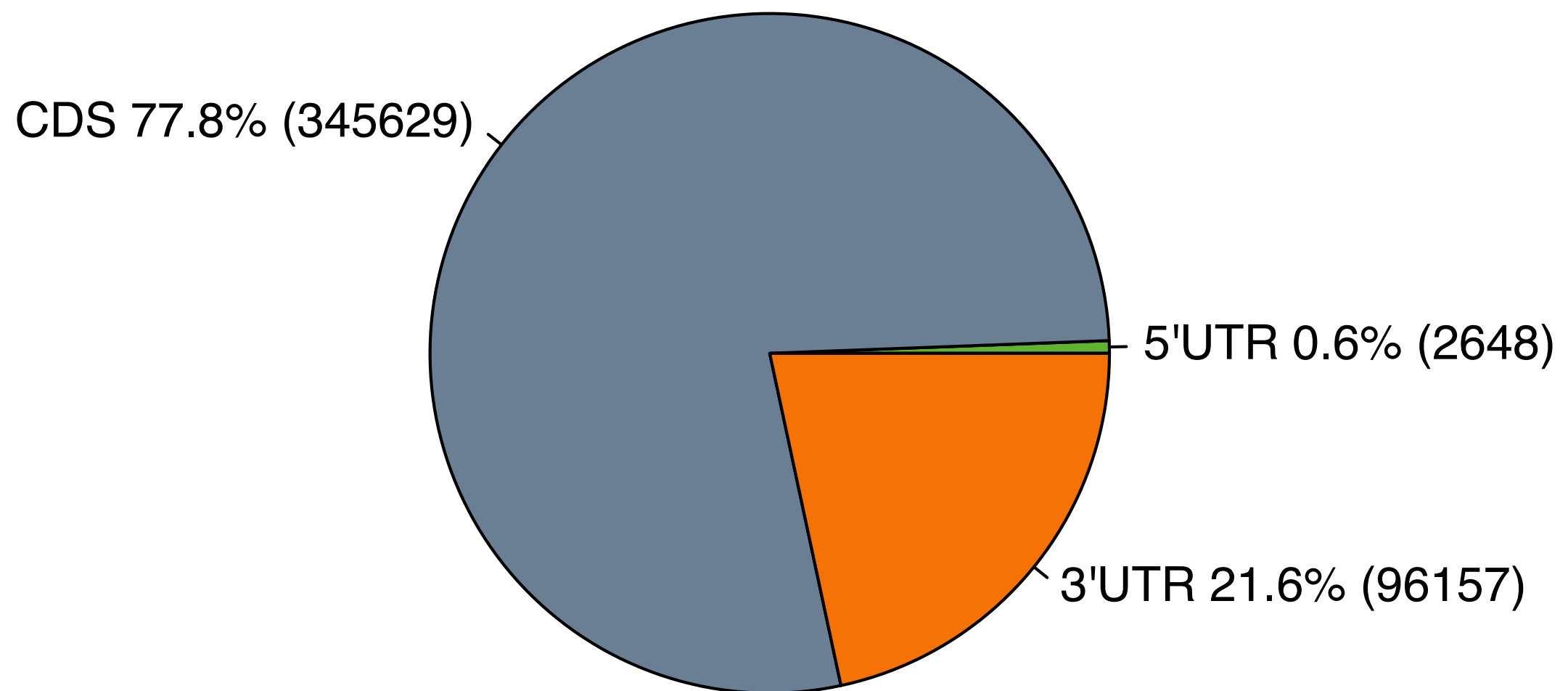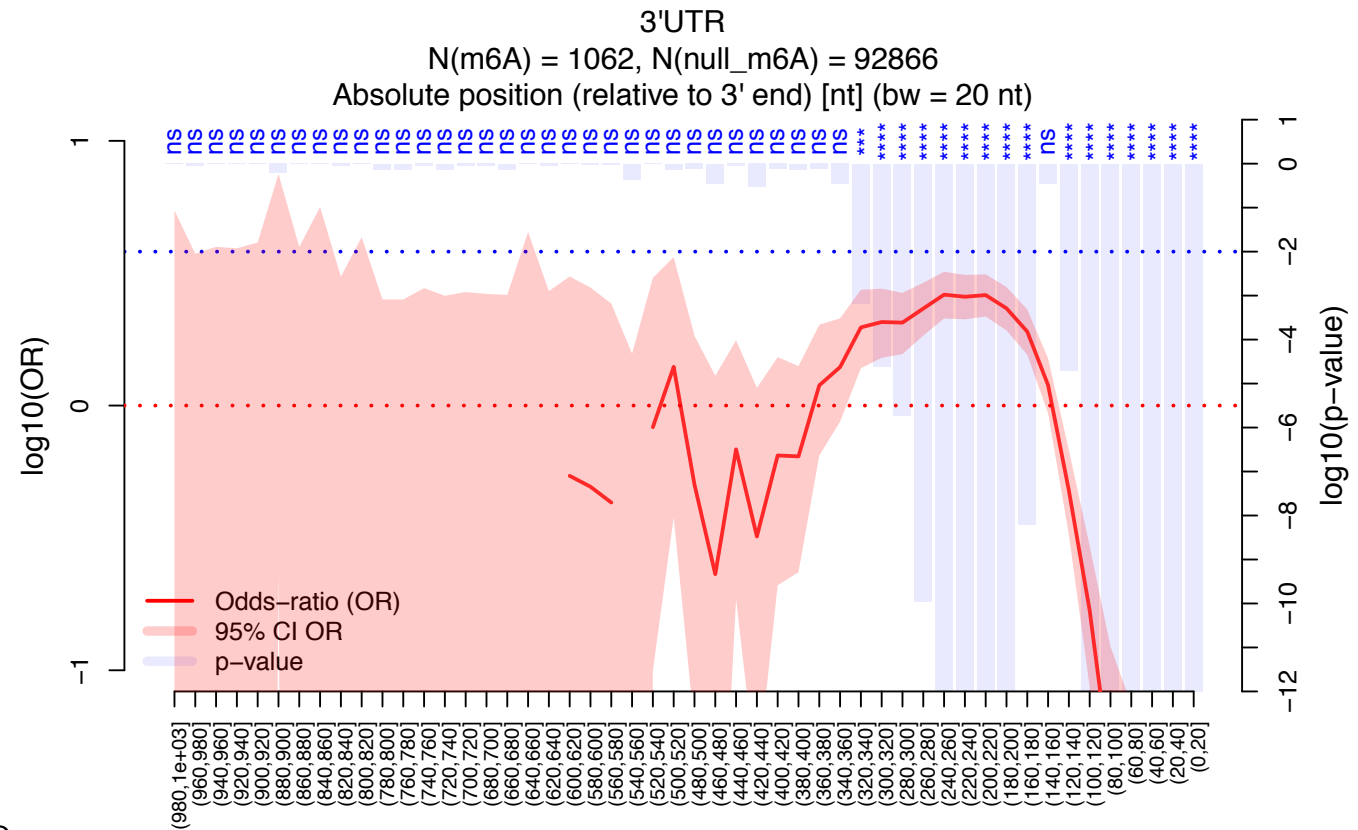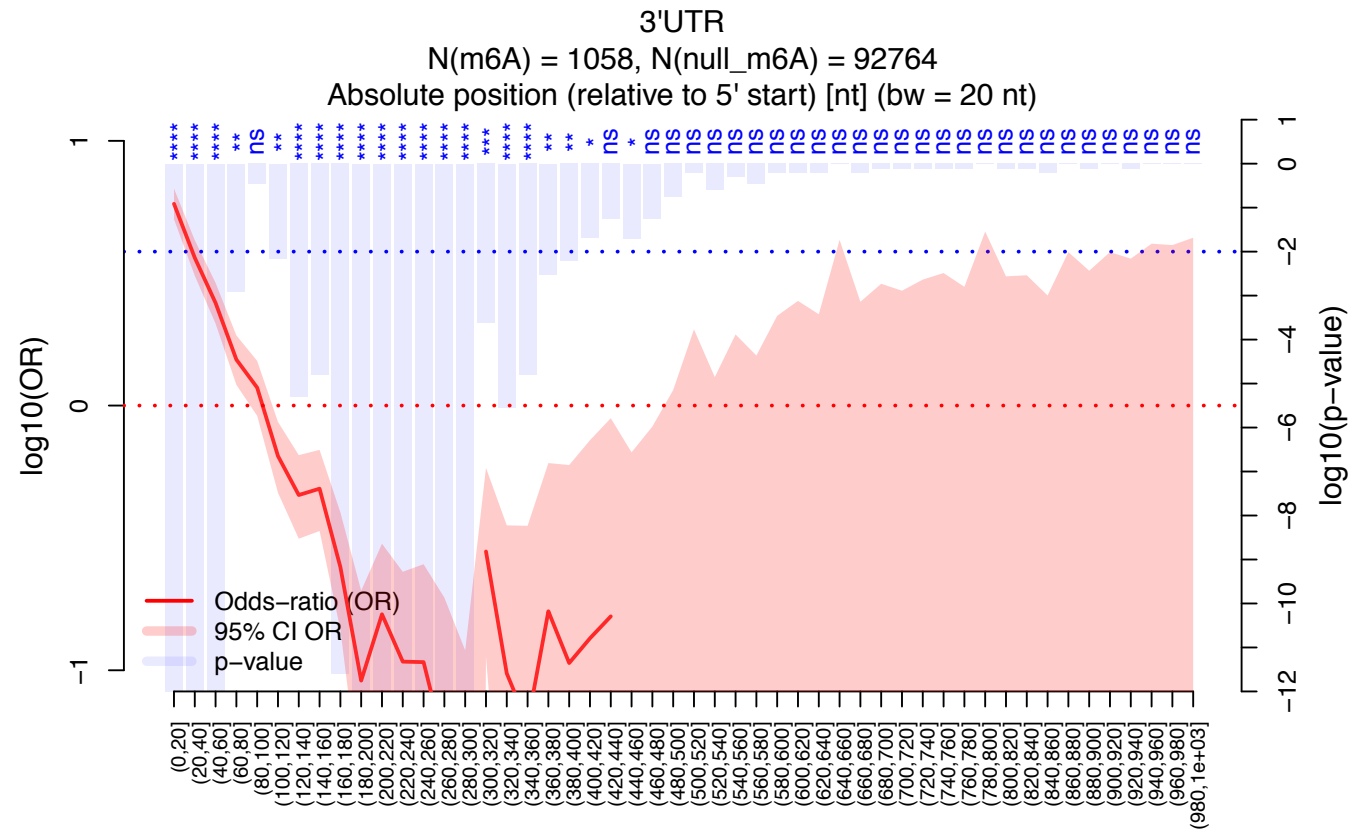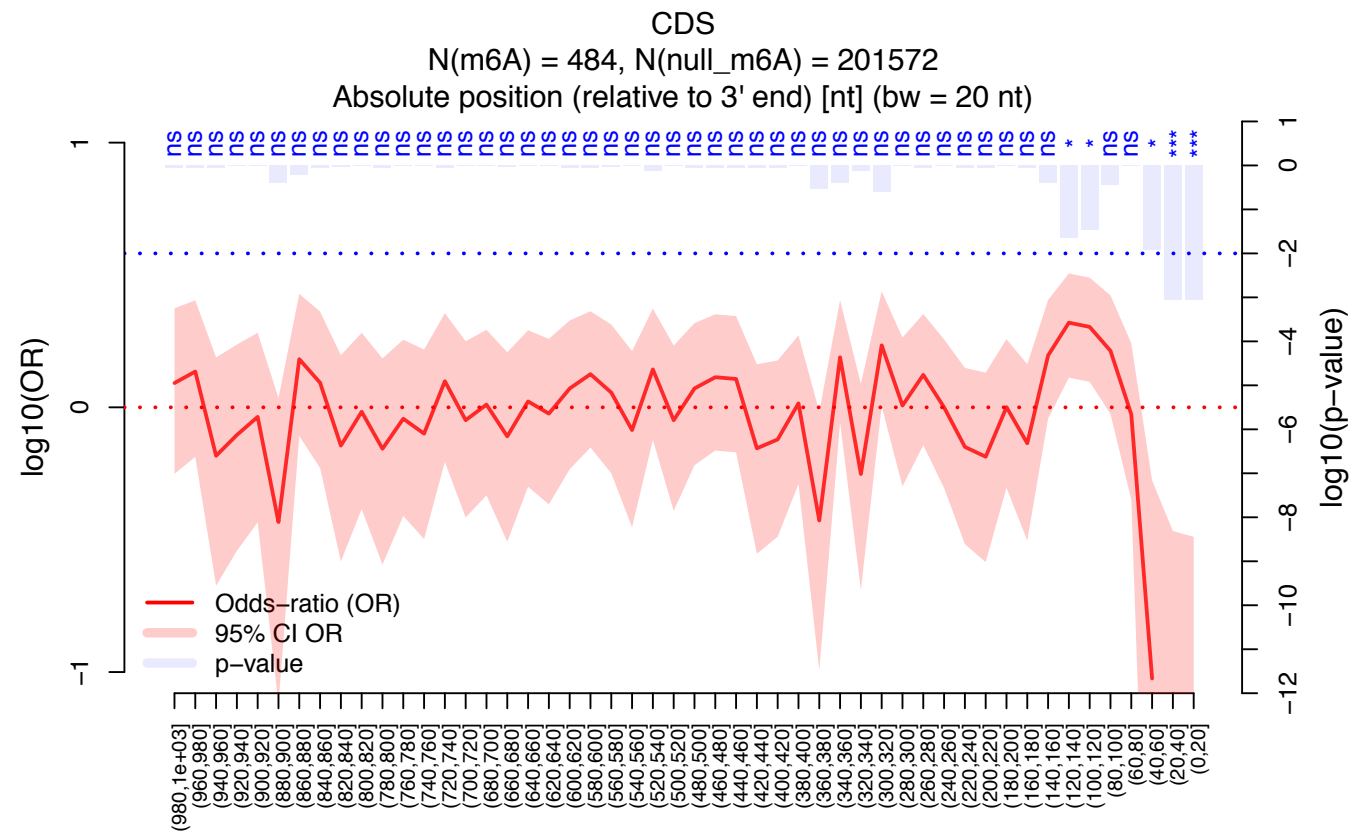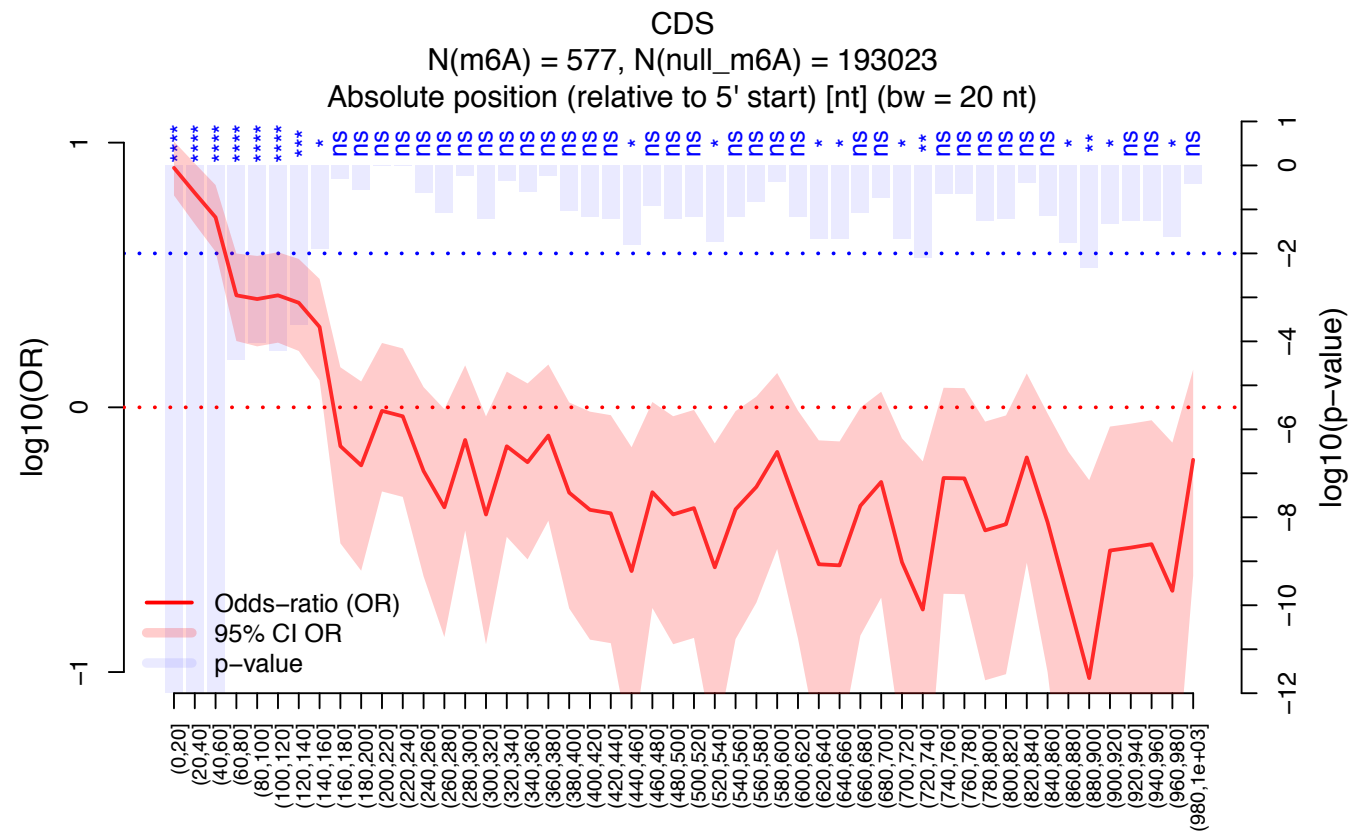
RNAModR analysis: m⁶A enrichment across transcript sections

N(m6A) = 1854, N(null_m6A) = 444434
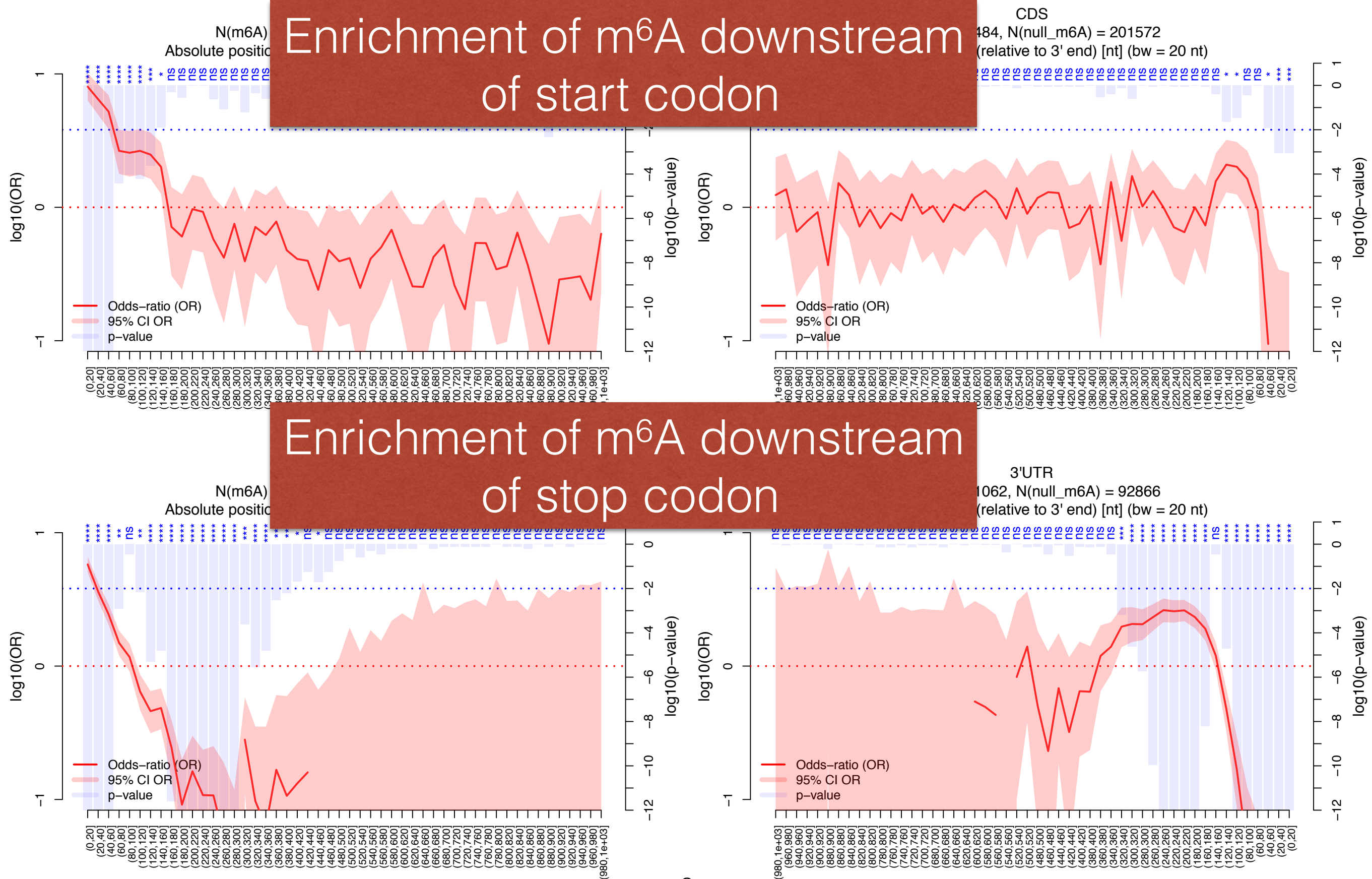
RNAModR analysis: m⁶A enrichment within transcript sections

RNAModR analysis: m⁶A enrichment within transcript sections

Enrichment of m⁶A downstream of start codon

Enrichment of m⁶A downstream of stop codon

RNAModR analysis: Polyadenylation signal (PAS) motifs

PAS motifs

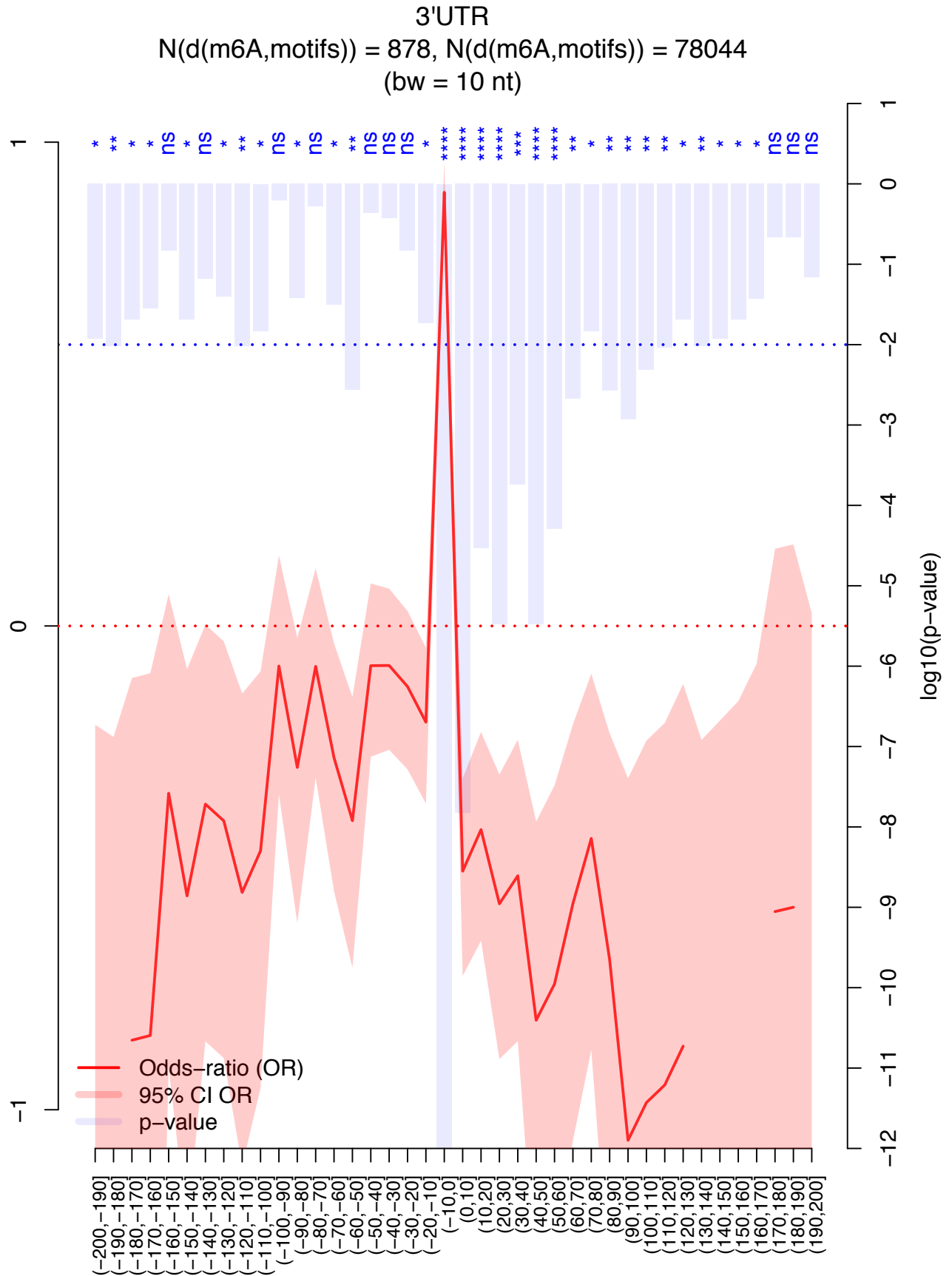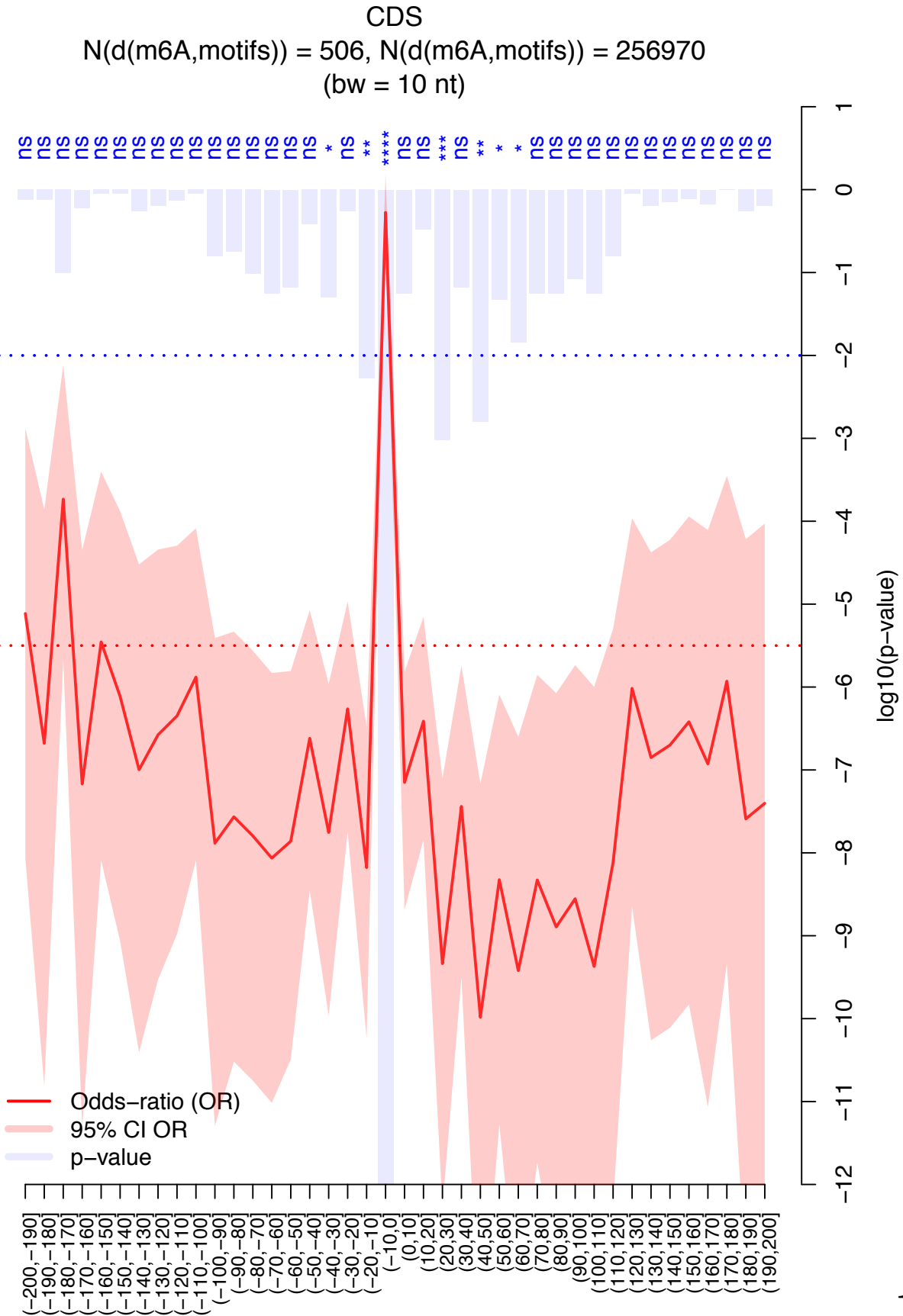canonical   AATAAA

non-canonical   ATTAAA

                              AGTAAA

                              TATAAA

                              AAGAAA

                              AATACA

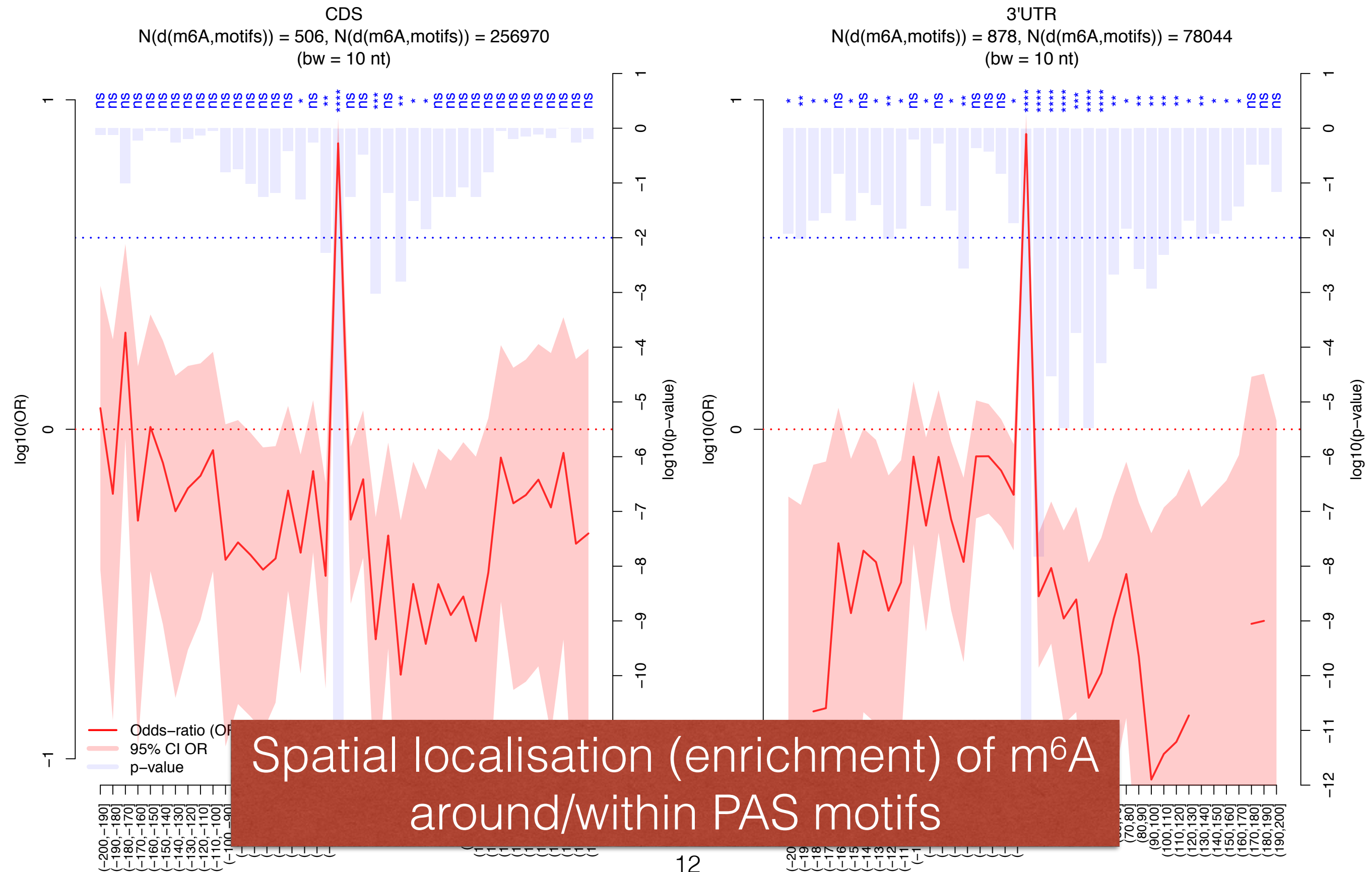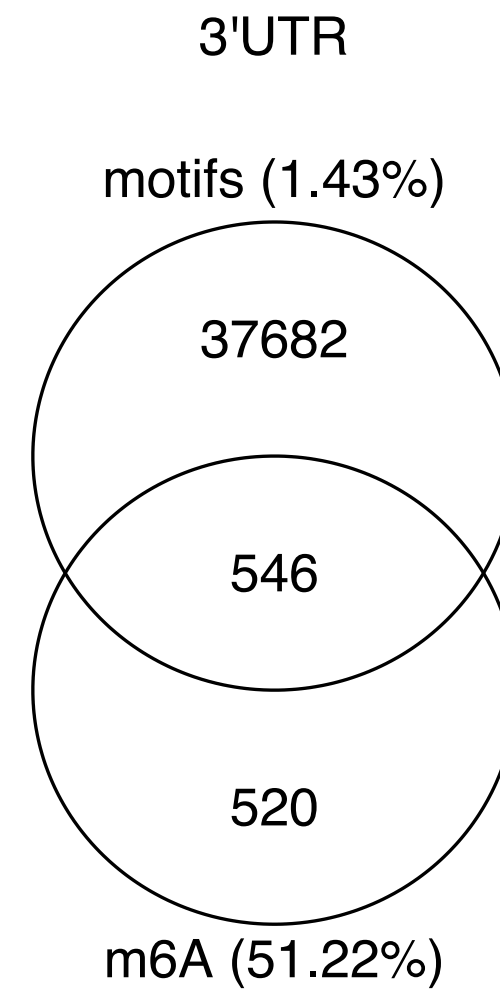   AATATA

                              CATAAA

                              AATGAA

                              GATAAA

                              ACTAAA

                              AATAGA

RNAModR analysis: Spatial enrichment of m⁶A relative to PAS motifs

CDS
N(d(m6A,motifs)) = 506, N(d(m6A,motifs)) = 256970
(bw = 10 nt)

3'UTR
N(d(m6A,motifs)) = 878, N(d(m6A,motifs)) = 78044
(bw = 10 nt)

RNAModR analysis: Spatial enrichment of m6A relative to PAS motifs

Spatial localisation (enrichment) of m6A around/within PAS motifs

RNAModR analysis: m⁶A overlap with PAS motifs

CDS

motifs (0.20%)

132827

267

507

m6A (34.50%)

3'UTR

motifs (1.43%)

37682

546

520

m6A (51.22%)
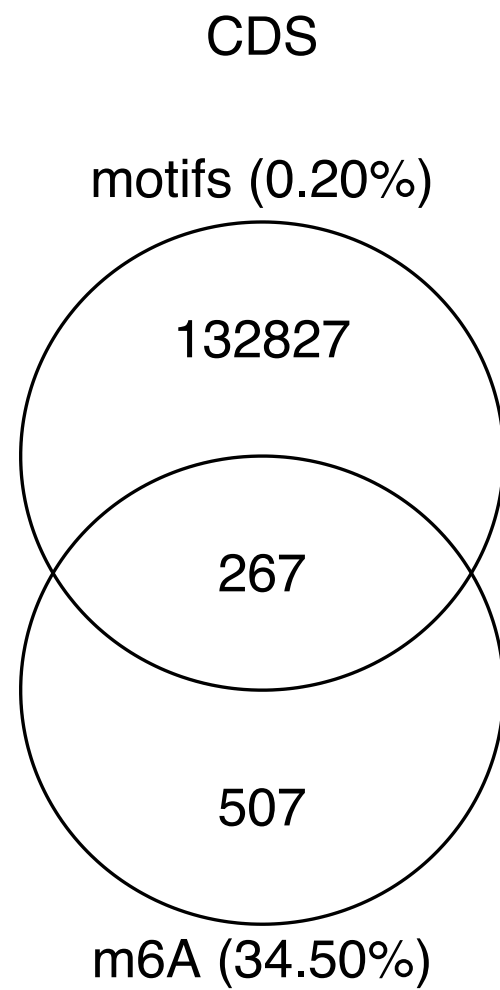
Polyadenylation site clusters (PAC)

Wu et al.

From Yalin:

- Material: Leaf 1: grow on soil, 8h daylight, unknown # of days; Leaf 2: sterile from germinating seed. Grow on filter paper with saturated nutrients, 16h daylight, 2wk; arabidopsis seed

- oligodT enrich polyA RNA, except for leaf 1

- 72 or 76 bp PE sequencing

- Internal priming removed. 71,000 polyA site clusters (PAC), with an average of 54 PATs per cluster. ~80% of PAC is in sense orientation, ~80% of these are in 3'UTR of known genes. There is also tissue specific APA, 73% of sense/anti-sense PACs are only in leaves (51%) or seeds (22%).
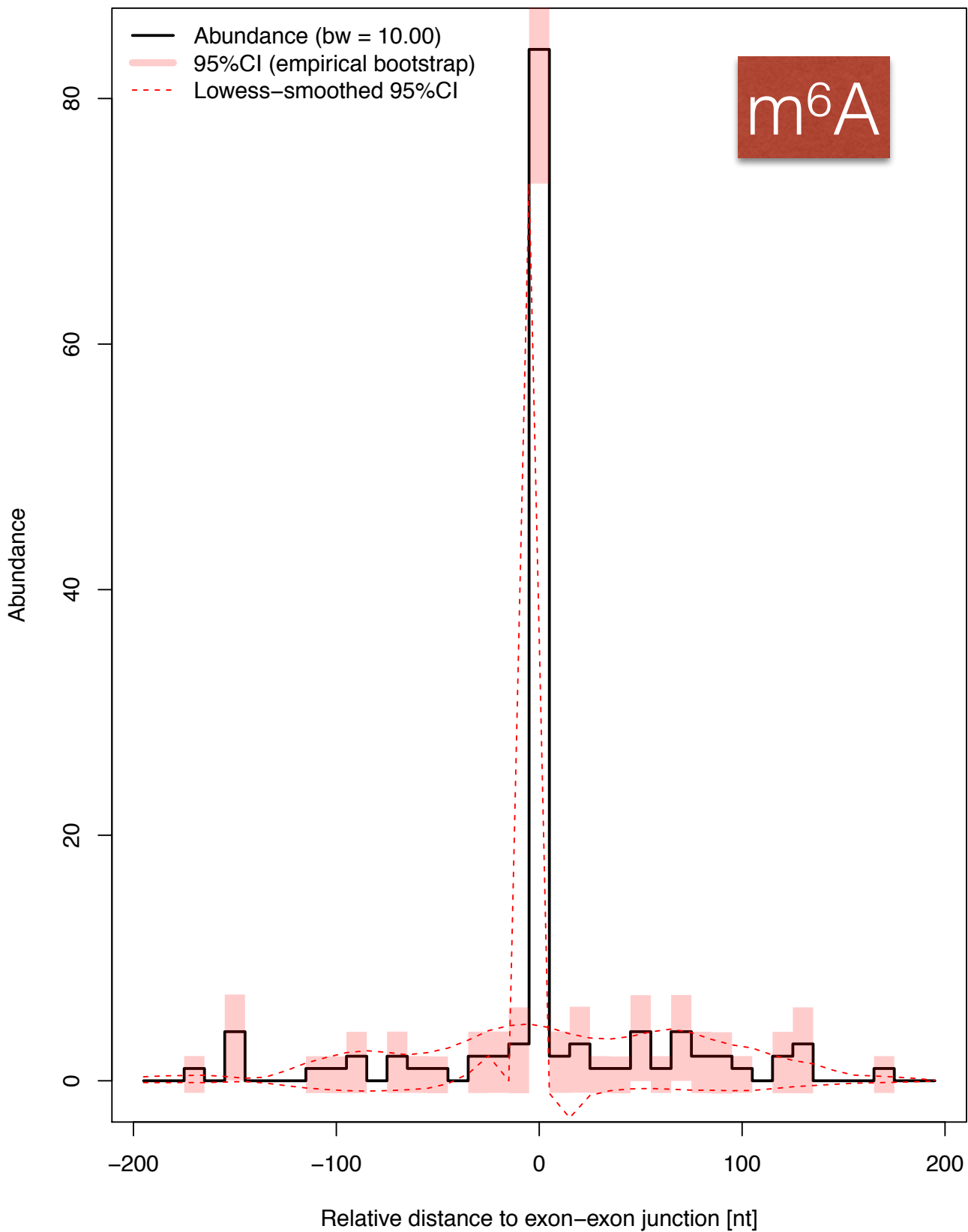
Focus on 57,473 sense PAC sites

RNAModR: PAC site distribution across transcript sections

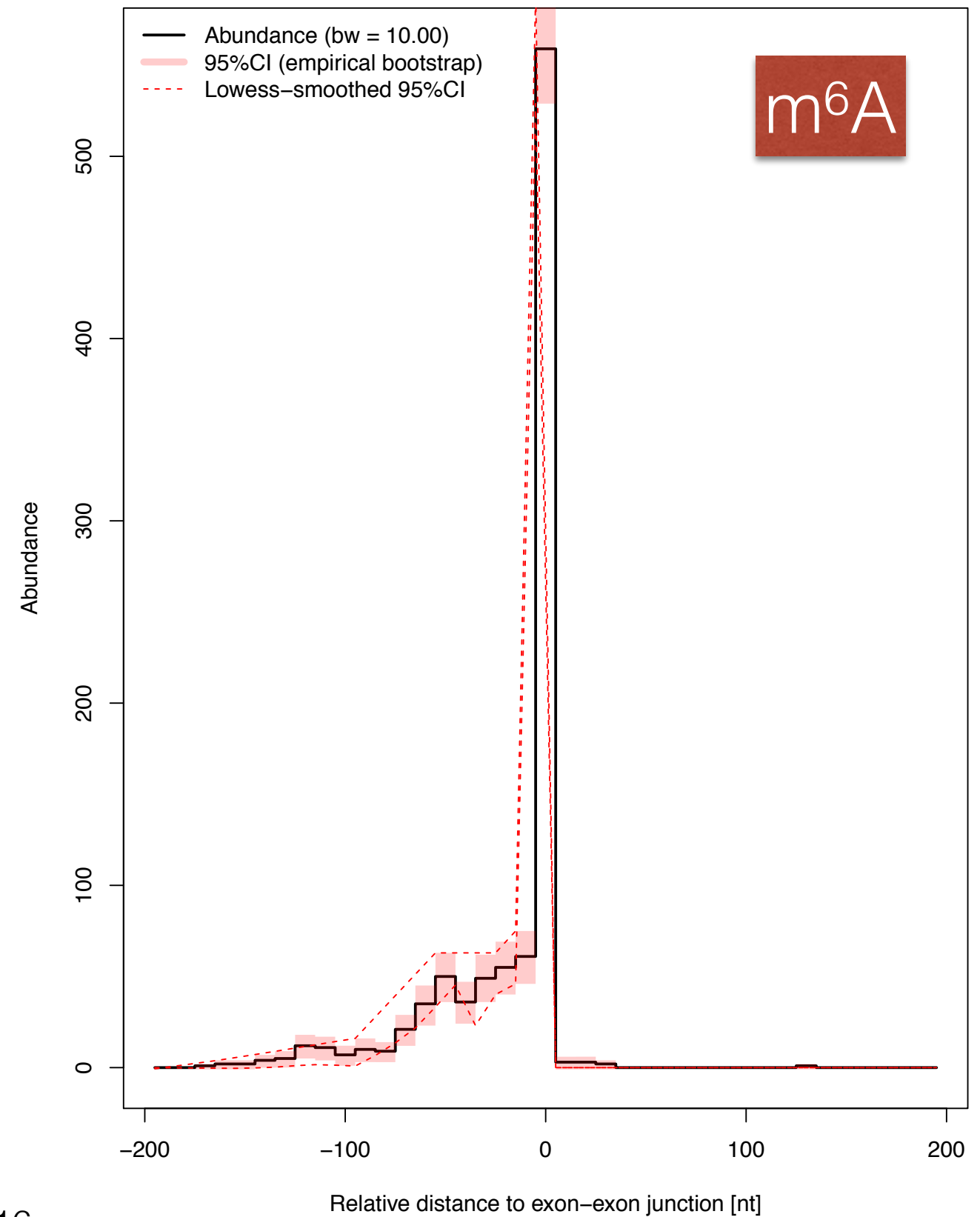## Distribution of 37845 PAC_sense sites across transcript sections

CDS 17.6% (6667)

3'UTR 82.4% (31178)
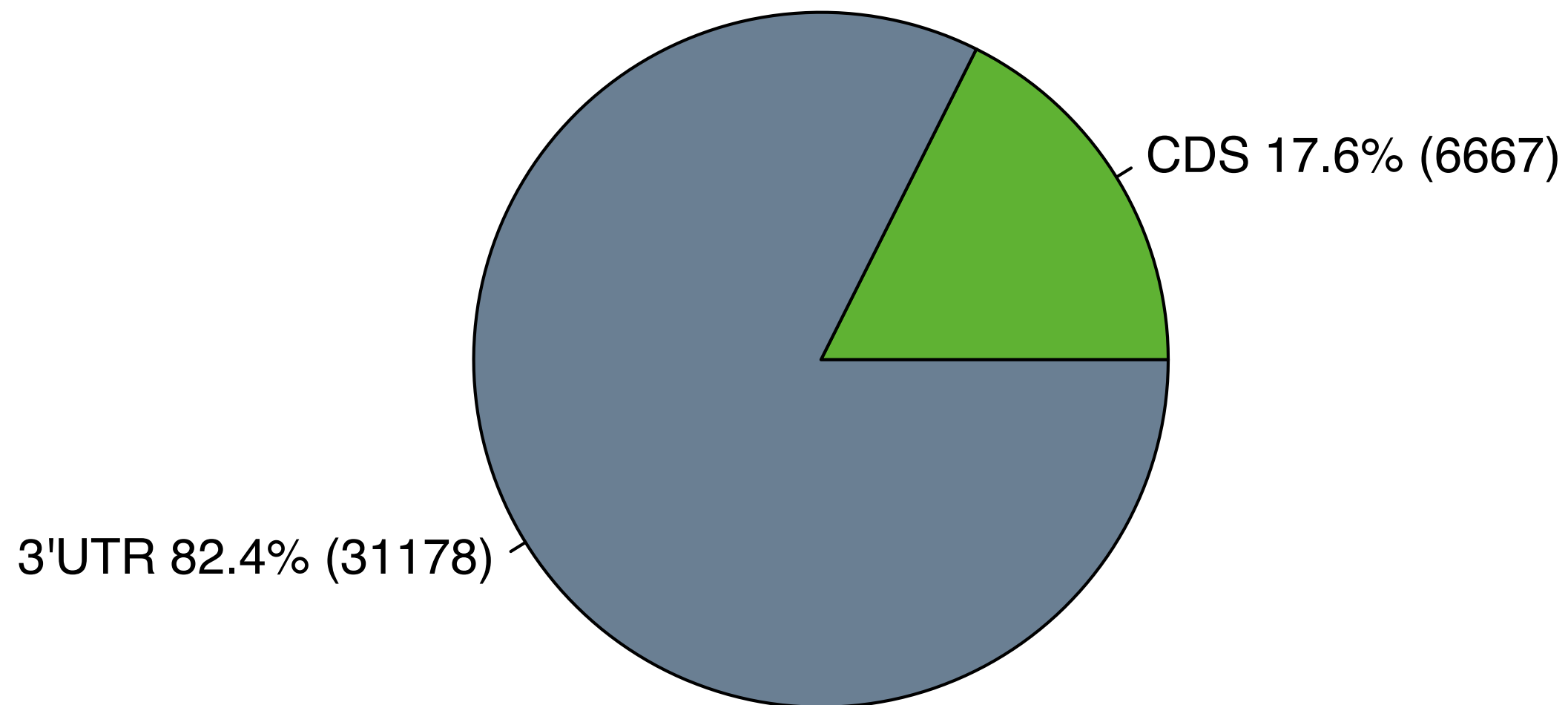
RNAModR analysis: Spatial distribution of m⁶A relative to PAC
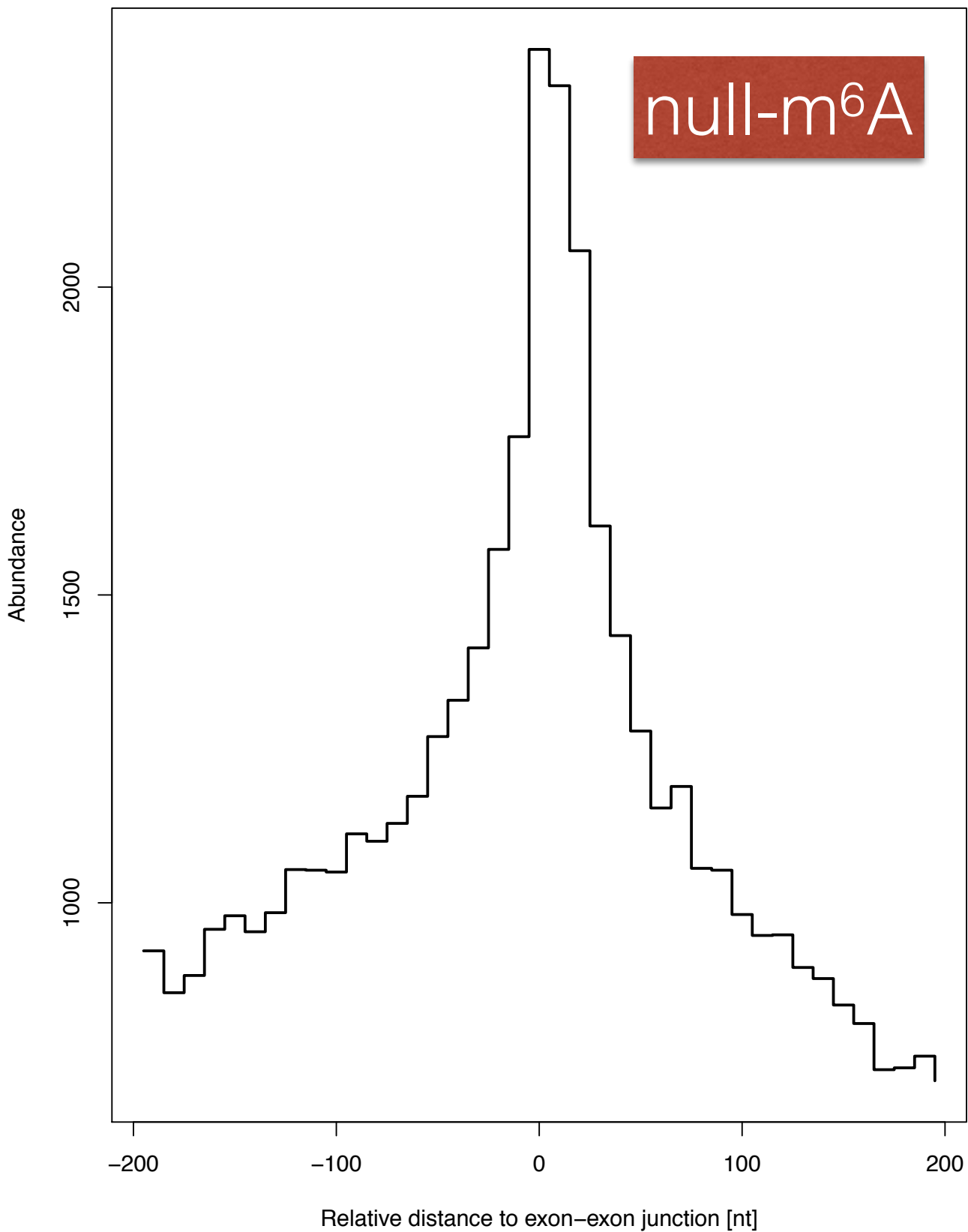
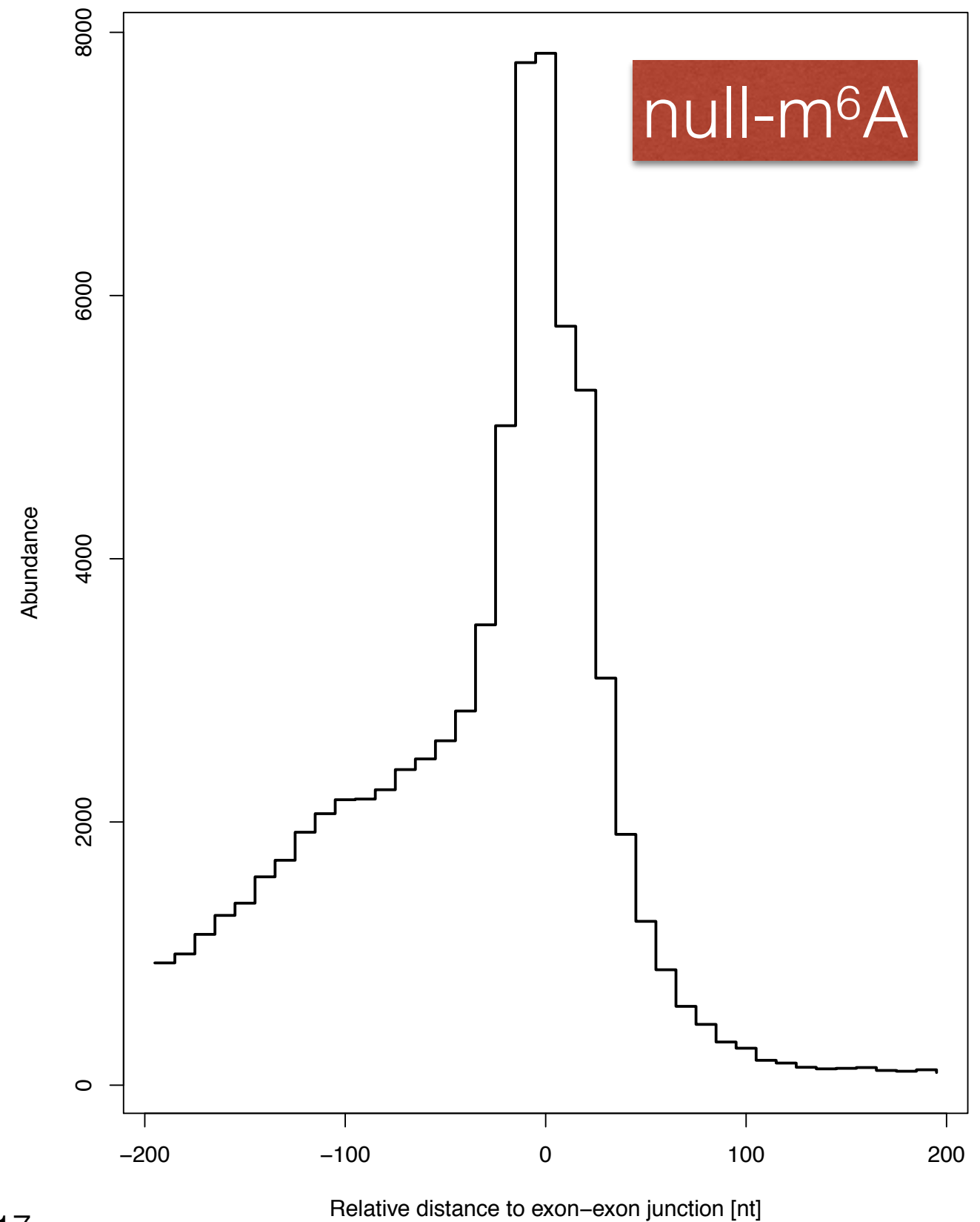d(m6A,PAC_sense) in CDS (N=131)

d(m6A,PAC_sense) in 3'UTR (N=938)

16

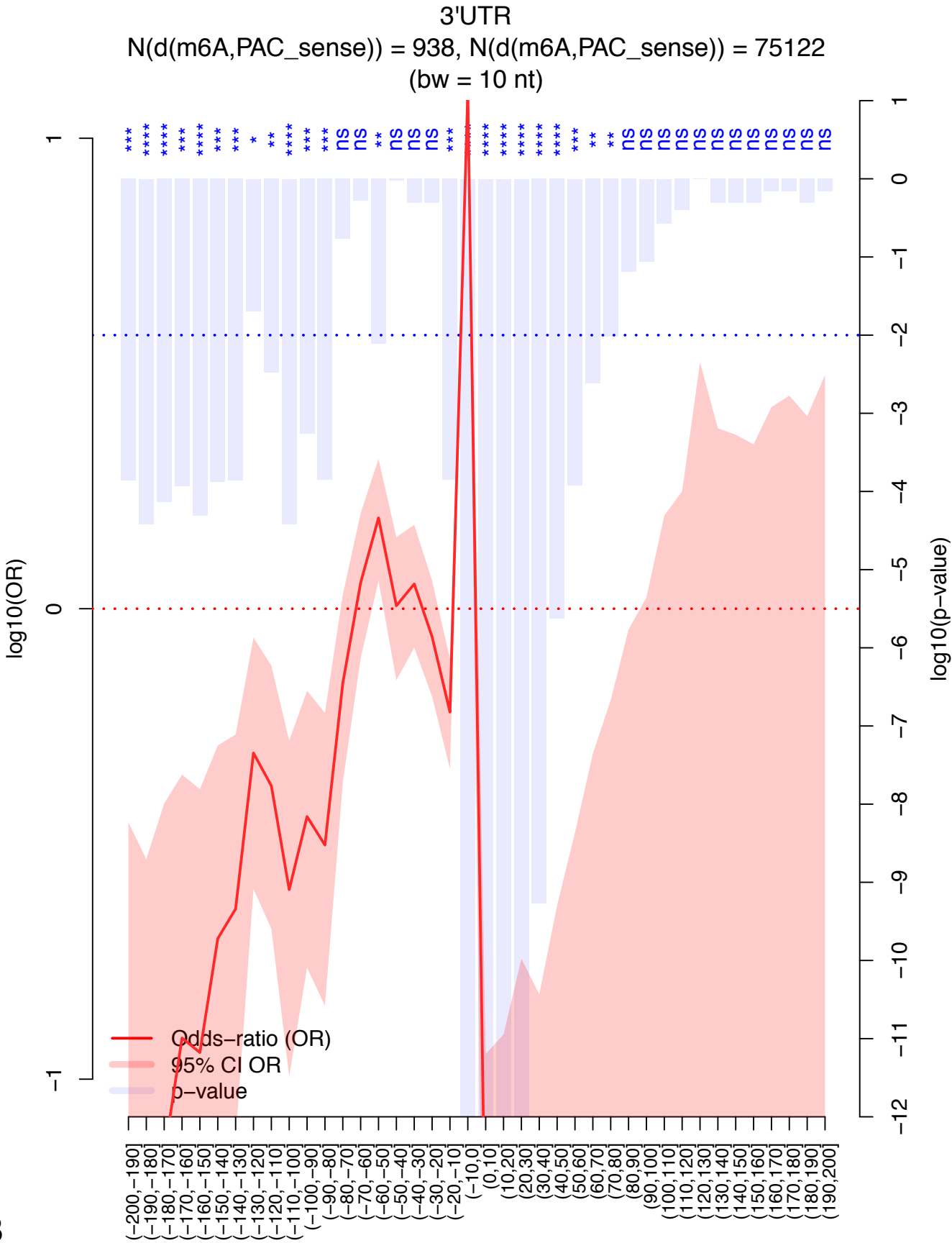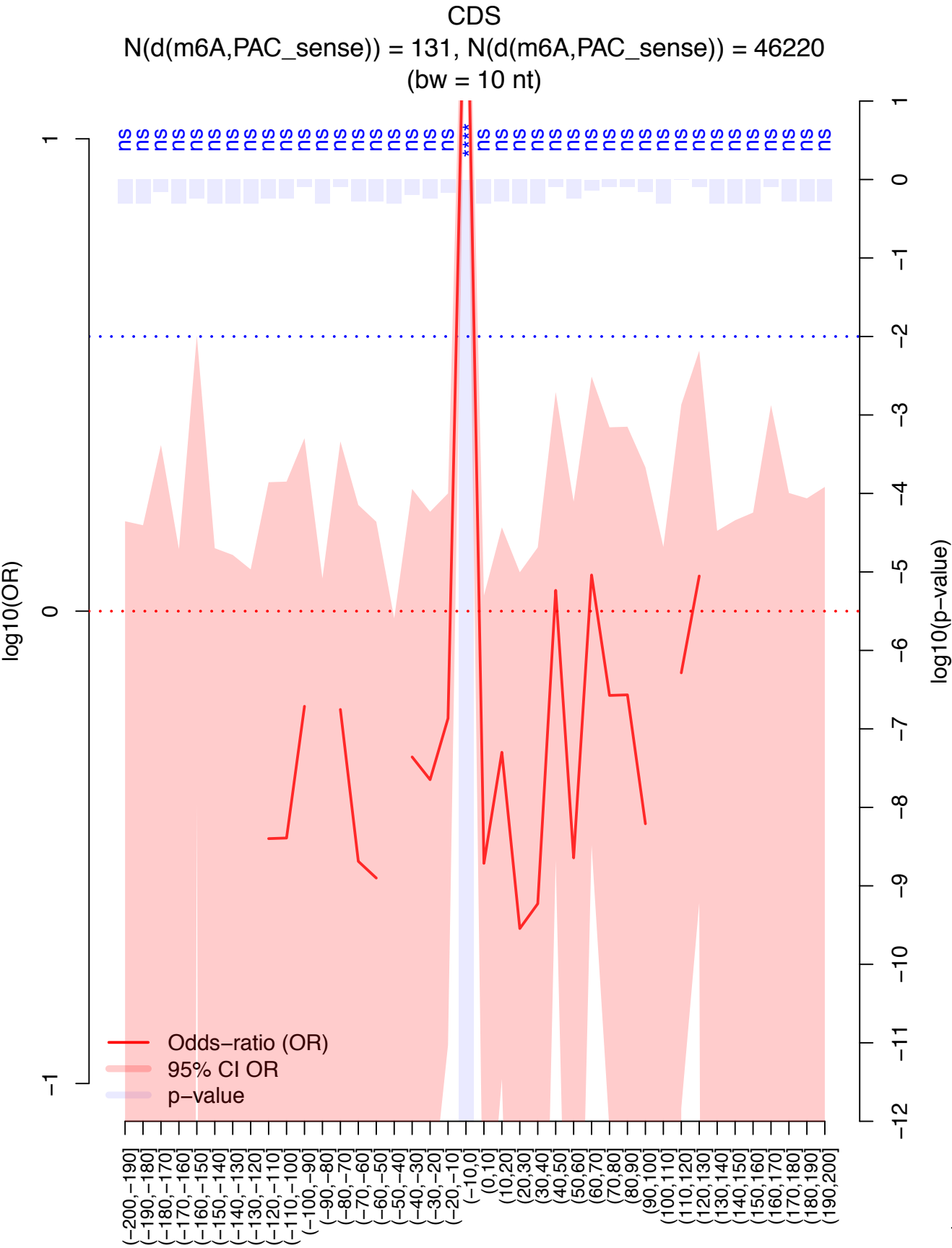# RNAModR analysis: Spatial distribution of m⁶A relative to PAC

d(null_m6A,PAC_sense) in CDS (N=46304)

d(null_m6A,PAC_sense) in 3'UTR (N=75216)

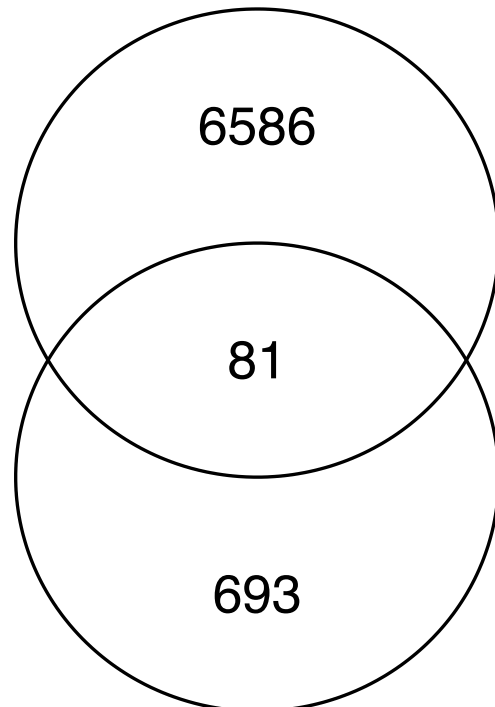RNAModR analysis: Spatial enrichment of m⁶A relative to PAC

CDS
N(d(m6A,PAC_sense)) = 131, N(d(m6A,PAC_sense)) = 46220
(bw = 10 nt)

3'UTR
N(d(m6A,PAC_sense)) = 938, N(d(m6A,PAC_sense)) = 75122
(bw = 10 nt)

RNAModR analysis: Spatial enrichment of m^{6}A relative to PAC

CDS
N(d(m6A,PAC_sense)) = 131, N(d(m6A,PAC_sense)) = 46220
(bw = 10 nt)

3'UTR
N(d(m6A,PAC_sense)) = 938, N(d(m6A,PAC_sense)) = 75122
(bw = 10 nt)

Odds−ratio (OR)
95% CI OR
p−value

Spatial localisation (enrichment) of m^{6}A around/within PAC's
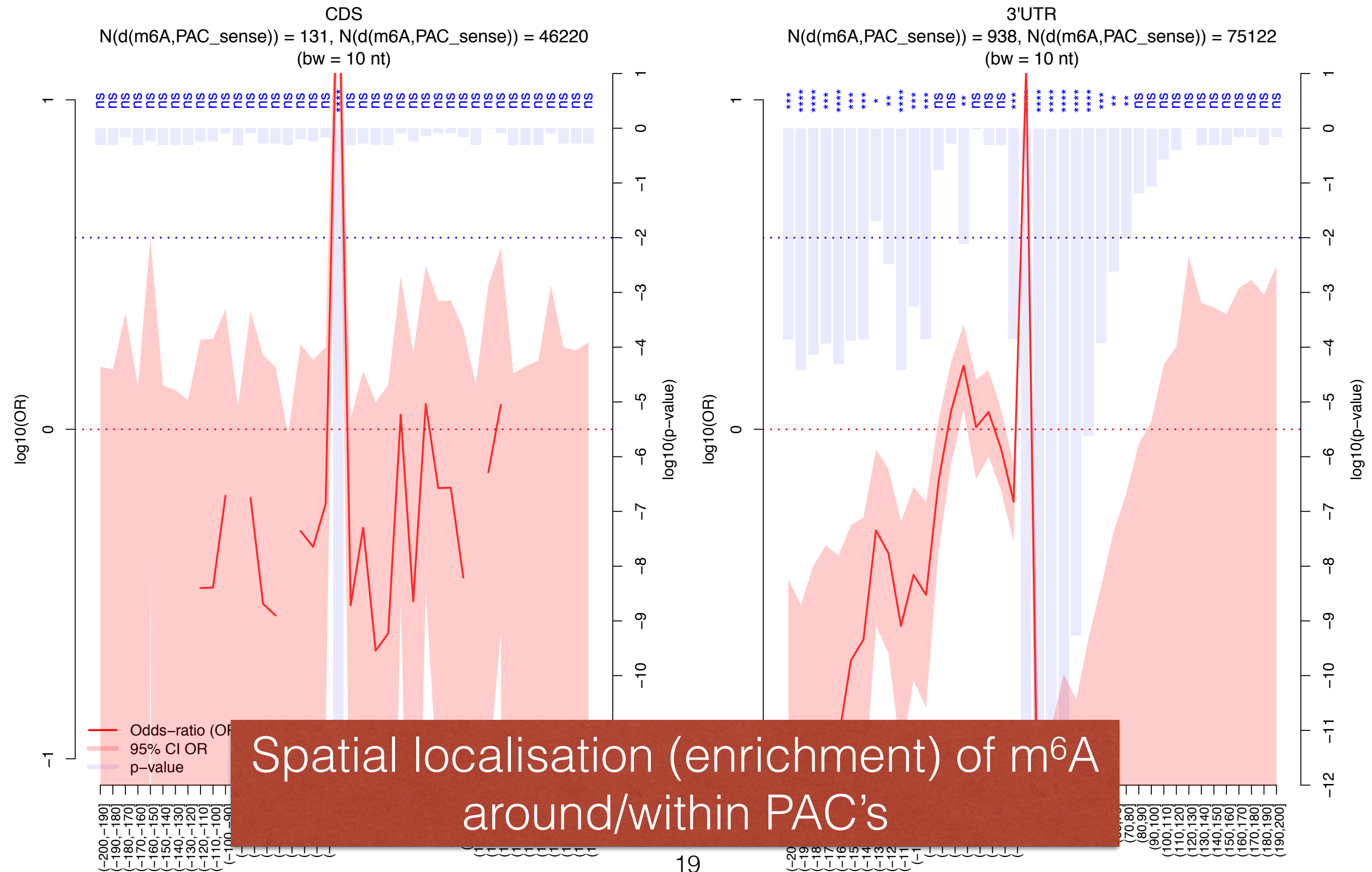
RNAModR analysis: m⁶A overlap with PAC's

CDS

PAC_sense (1.21%)

6586

81

693
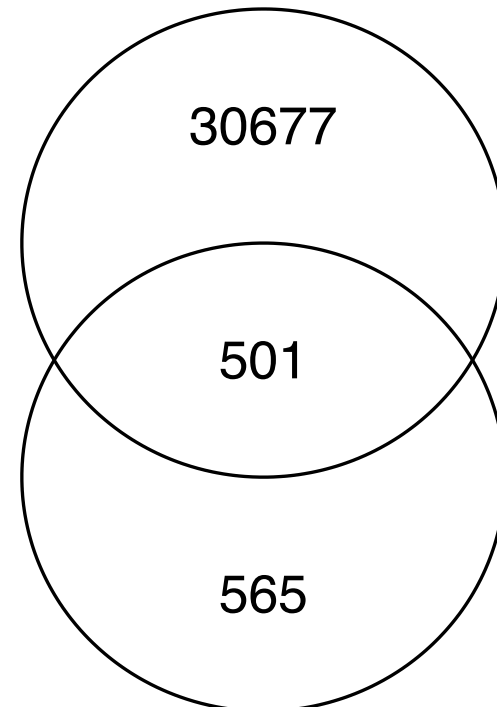
m6A (10.47%)

3'UTR

PAC_sense (1.61%)

30677

501

565

m6A (47.00%)

Cleavage sites

Sherstnev et al.

From Yalin:

- Arabidopsis seedling. 2 days at 4 degree C in dark, then harvest 14 days after transfer to 24 degree C.

- Direct RNA Sequencing (DRS) technology (PMID:21145465) to reveal 3'end formation.

Focus on 49,916 protein coding gene-associated cleavage sites

RNAModR: Cleavage site distribution across transcript sections

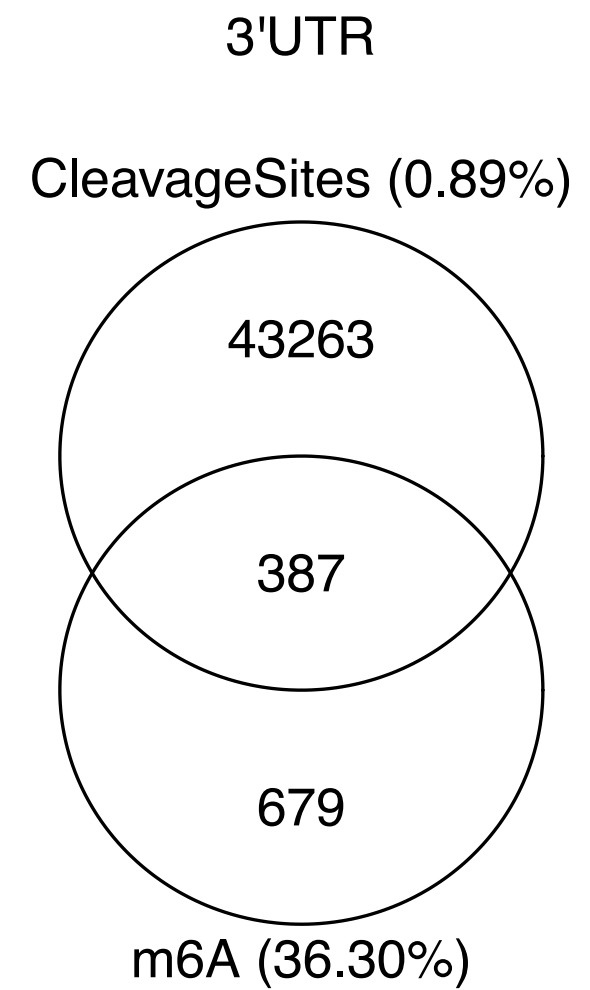Distribution of 43722 CleavageSites sites across transcript sections

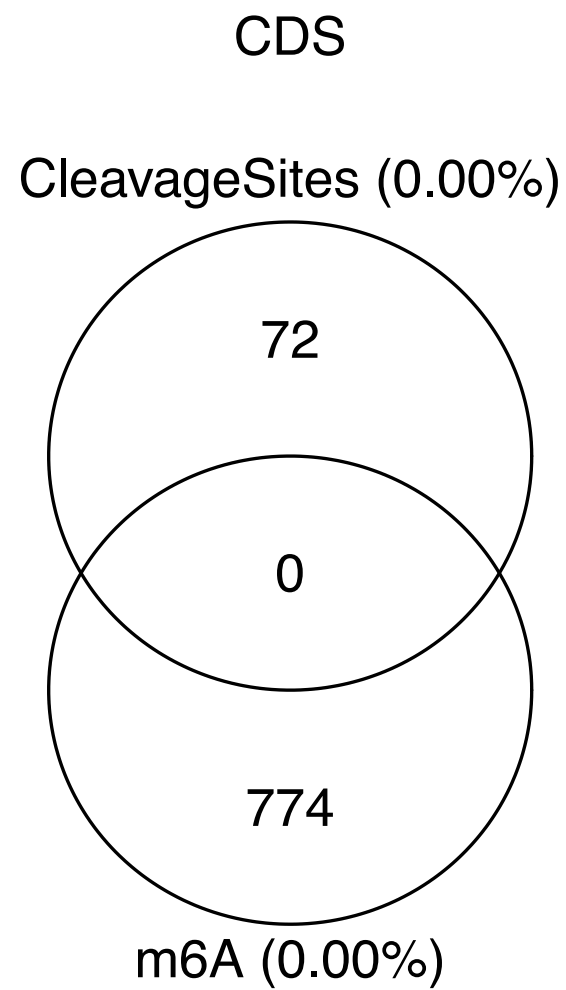3'UTR 99.8% (43650) — CDS 0.2% (72)

RNAModR analysis: Spatial enrichment of m⁶A relative to cleavage sites

RNAModR analysis: Spatial enrichment of m⁶A relative to cleavage sites

Spatial localisation (enrichment) of m⁶A around/within cleavage sites in 3'UTR
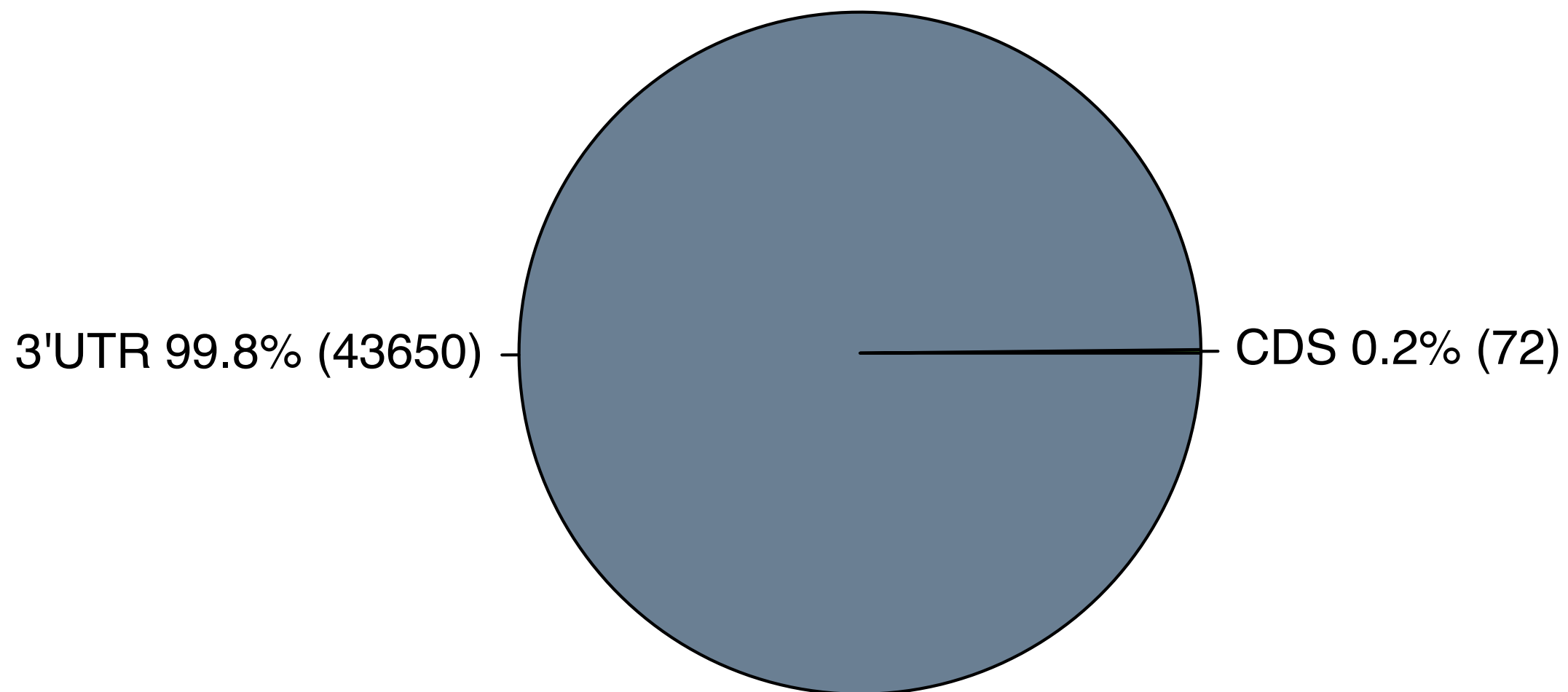
RNAModR analysis: m⁶A overlap with cleavage sites

CDS

CleavageSites (0.00%)

72

0

774

m6A (0.00%)

3'UTR

CleavageSites (0.89%)

43263

387

679

m6A (36.30%)

RNAModR analysis

Source data and files on github:

https://github.com/mevers/m6A_Arabidopsis