Supplemental Material

A majority of m6A residues are in the last exons, allowing the potential for 3′ UTR regulation

Ke et. al. *Genes & Development* 2015

A. Supplemental Data:

Both the **processed data** an **raw data** have been deposited to GEO database (accession no. GSE71154, http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE71154).

The processed data includes the **precise locations** of all mapped m6A sites in mouse brain, mouse liver, human CD8 T cells and human A549 cells (30078, 8748, 19682 and 23479 precise m6A sites respectively) as well as the m6A peak region locations.

B. Supplemental Figures:

Sup. Fig. 1 to Sup. Fig. 7 (Page 1 to 12)

Supplemental Figure Legends

Supplemental Figure 1. Single nucleotide resolution achieved by m^{6}A-CLIP/IP, part A. (*A*) m^{6}A-CLIP pilot with two 25mer RNA oligos, one with an m^{6}A at site 13 and one with no m^{6}A. (*B*) UV induced insertion was identified at the +1 position (3' to the m^{6}A site) of m^{6}A in the pilot RNA oligo. Position 13 is the m^{6}A (marked in red). (*C*) UV induced substitution mutations are predominantly at +1 position of the two known m^{6}A sites in human rRNAs. Left panel: the A1832 site of 18S; Right panel: the A4190 site of 28S. (*D*) UV induced m^{6}A antibody covalent complex was evident in m^{6}A-CLIP but not in two negative controls. (*E*) Deletion CIMS maps m^{6}A sites at its -2 position. "Enrichment of RRACU or RAC" is the fold enrichment of RRACU (red) or RAC (blue) motif density at that position compared to the background motif density (horizontal grey dot line). Vertical black dot line represents the Peak Site 0 position. (*F*) Insertion CIMS maps m^{6}A sites at its -1 position. (*G*) Crosslink induced truncation site (CITS) maps m^{6}A sites at its -3 position.

Supplemental Figure 2. Single nucleotide resolution achieved by m^{6}A-CLIP/IP, Part B. (*A*) m^{6}A Peak Site maps m6A sites at its 0 position (human). The motif logo with its p value represents the de novo motif. (*B*) Substitution CIMS maps m^{6}A sites at its -1 position (human). (*C*) m^{6}A induced truncation sites (MITS) maps m^{6}A sites at its +1 position (human). (*D*) Summary of m^{6}A sites identified by different approaches (human). "All RACs" is the total number of RACs within m^{6}A peak regions. (*E*) m^{6}A sites identified by each type of analysis significantly overlaps with sites identified by other types of analyses. Left panel: mouse brain; Right panel: human. (*F*) Precisely mapped

m^6A sites are conserved in vertebrates regardless of the mapping approaches (human). "Control RACs" are the matched non-m^6A RACs within the same m^6A peak region as the m^6A sites. ** $p< 10^{-3}$, Wilcoxon rank-sum test. (*G*) m^6A-CLIP/IP verifies precise m^6A sites in MALAT1 RNA. "m^6A-IP enrich." is m^6A-IP reads normalized to its input. "m^6A-CLIP/IP" is by m^6A-CLIP/IP (the A marked in red); "SCARLET" is by Liu et al at (motif underlined). From left to right, these sites (Sites 2515, 2577, 2611, 2720) were located by 0 of substitution CIMS, +1 of MITS, -4 of CITS and -3 of CITS (see methods for precise mapping details).

Supplemental Figure 3. m^6A is enriched when entering last exons but not at stop codons. (*A*) Entering last exon, m^6A density increased sharply (lower panel) in contrast to its lagging increase when approaching stop codon (upper panel, data from human CD8 T cells and earlier publications). "m^6A peak density" was calculated as the number of m^6A peak region in a 10nt interval divided by the total number of mRNAs that contained this position. (*B*) m^6A is enriched in last exons but not around stop codons (data from human CD8 T cells and earlier publications). The upper panel is the m^6A peak region distribution around stop codons; the two lower panels are the distribution of m^6A peak regions and stop codons around last exon start. mRNAs were grouped according to their stop codon locations to last exon start. (*C*) m^6A is not enriched around stop codon when stop codon is not in last exon (data from human CD8 T cells and earlier publications). (*D*) Most exonic m^6As locate in last exon (human). The three pie graphs showed the relative proportions of m^6A peaks, RAC and RRACU motif in last exon and other exons with 100% representing all m^6A peaks/RAC/RRACU on mRNA.

Supplemental Figure 4. Short last exons have lower m^6A density . (*A*) The color-coded definition of last exon legnth. (*B*) Mouse brain data: The m^6A density in last exons were compared among long and short last exons. All genes considered are multi-exon coding genes with RPKM>=1 in mouse brain. "m^6A peak density" was calculated as the number of m^6A peak region in a 10nt interval divided by the total number of mRNAs that contained this position. (*C*) RAC density show little difference among long and short last exons expressed in mouse brain. (*D*) Human CD8 T cell data: All genes condisered are multi-exon coding genes with RPKM>=1 in human.

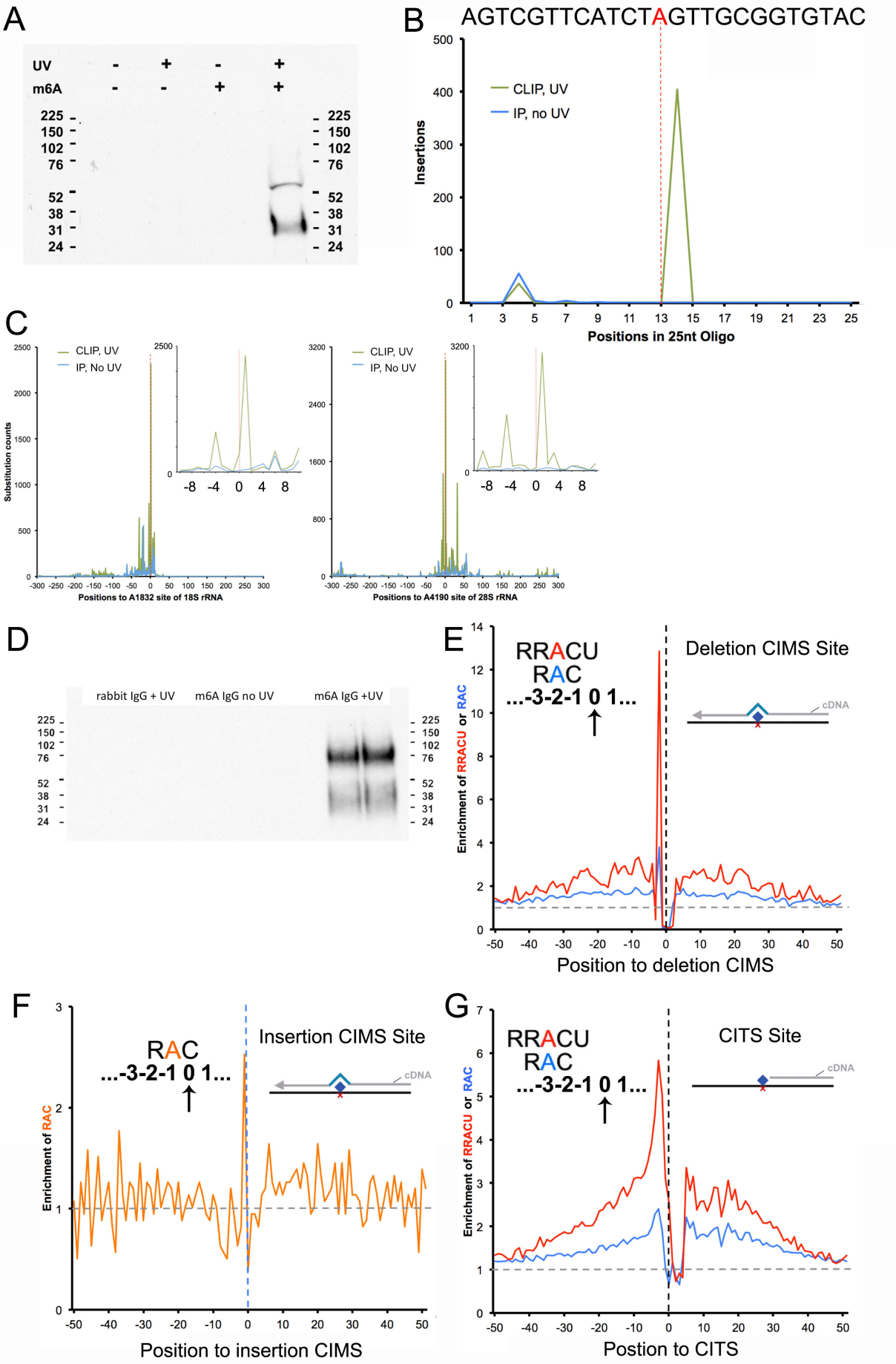
Supplemental Figure 5. Higher m^6A levels in last exons correlates with more distal polyA site usage. (*A*) m^6A is more abundant near proximal sites that are used less frequently (black line, left panel); when distal sites are used more, the m^6A level is higher until polyA site is reached (black line, right panel). The error bar is standard error of mean at each position. The shaded area highlights major area of difference between two groups, **, $p< 10^{-35}$, Fisher Exact Test. "used more" means a >=60% usage of all polyA sites and "used less" means a <=40% usage. Other cutoffs including 50% or 70% produce the same findings. The expression of mRNAs in regions between proximal and distal polyA sites is adequate for m^6A detection (RPKM>=1). Human data is presented here. (*B*) Distal alternative polyA sites in last exons are used more often in brains for co-expressed mRNAs. Alternative polyA site pairs with statistically significant differences in usage are highlighted in orange (higher usage of distal sites in brain) or blue (higher usage of distal sites in liver). FDR =5%, Fisher exact test. (*C*) Positional plot of m^6A

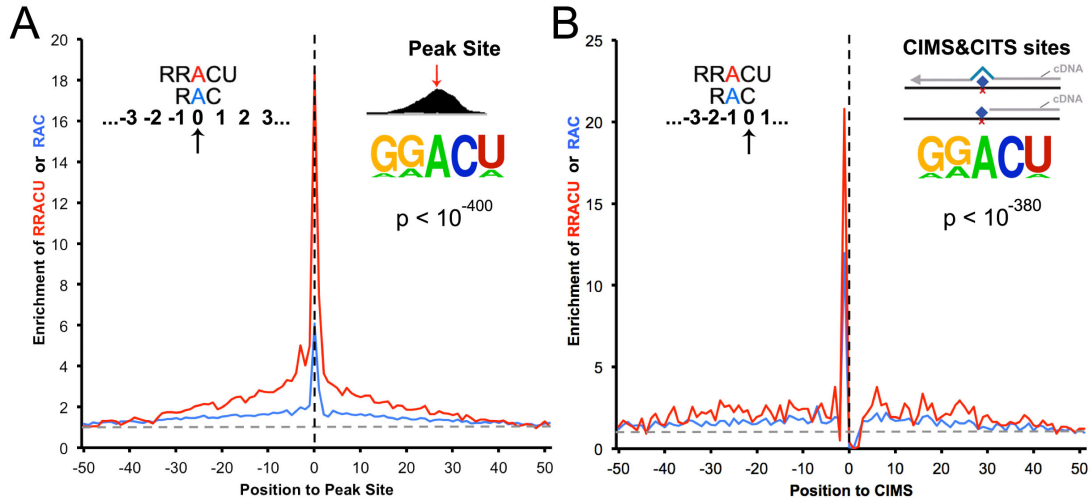
peaks around alternative polyA sites in last exons (FDR=5% for detecting changed alternative polyA sites). "m^6A peak density (brain - liver)" for each position was calculated as the number of m^6A peak regions that were higher in brain (orange points in Fig. 4C) in a 10nt interval divided by the total number of mRNAs that contained this position. Orange and grey lines represent more distal usage and no change in brain respectively. The expression of mRNAs in regions between proximal and distal polyA sites is adequate for m^6A detection in both tissues (RPKM>=1). Error bar: standard error of mean. The shaded area highlighted the major area of difference between two groups. **, $p< 10^{-20}$, Fisher exact test.

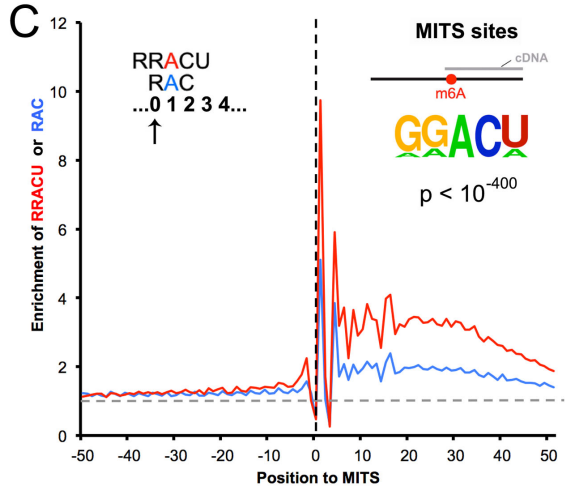
Supplemental Figure 6. Global limitation of m^6A changed alternative polyA sites in last exons. Alternative polyA site pairs with statistically significant differences in usage are highlighted in orange (higher usage of proximal sites in knockdown) or blue (higher usage of proximal sites in control). FDR=5%, Fisher exact test. The mRNAs using two alternative polyA sites in last exon and with m^6A presence at proximal polyA sites are summerized. The p value for directionality was caluclated by binomial test.

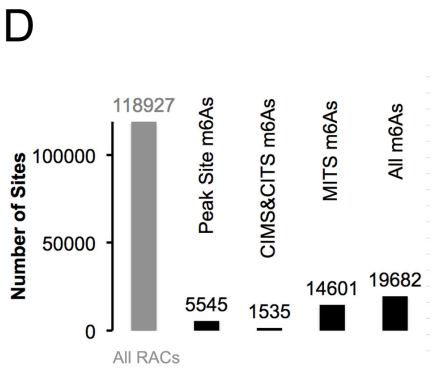
Supplemental Figure 7. Specific examples of m^6A loss and more usage of proximal polyA sites. (A) the DNAJB11 gene: m^6A-IP enrichment value is calculated by normalizing m^6A-IP reads to its input. "m^6A WT" (black) is the m^6A-IP enrichment value for WT (wild type, i.e. before knockdown); "m^6A KD" (grey) is the m^6A-IP enrichment value for KD (knockdown); "PolyA WT" (black) shows the polyA site usages for proximal polyA site (left) and distal polyA site (right) before knockdown, and "PolyA

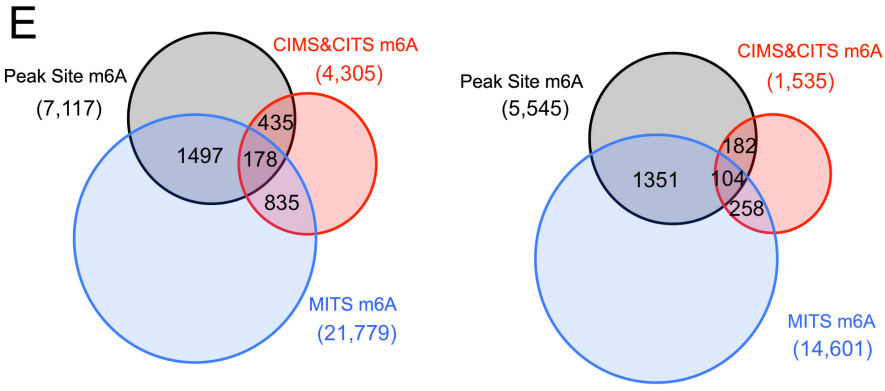
KD" (grey) shows the polyA site usages for proximal polyA site (left) and distal polyA site (right) after knockdown. The PolyA value for the major site between proximal and distal sites was set to 1 and the value for the minor site was adjusted proportionally. (*B*) the NDUFA6 gene. (*C*) the TMEM206 gene. (*D*) the ZNF142 gene. (*E*) the PTEN gene.

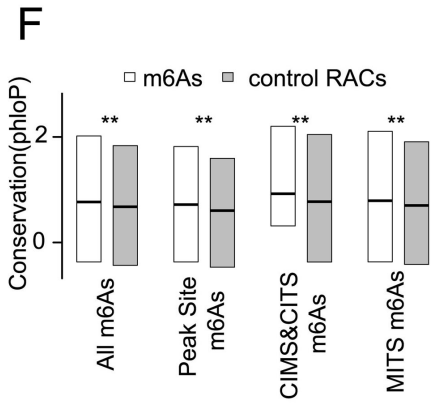

A
UV - + - +
m6A - - + +
225
150
102
76
52
38
31
24
B
AGTCGTTCATCTAGTTGCGGTGTAC
CLIP, UV
IP, no UV
Insertions
Positions in 25nt Oligo
C
CLIP, UV
IP, No UV
Substitution counts
Positions to A1832 site of 18S rRNA
Positions to A4190 site of 28S rRNA
D
rabbit IgG + UV
m6A IgG no UV
m6A IgG +UV
E
RRACU
RAC
...-3-2-1 0 1...
Deletion CIMS Site
cDNA
Enrichment of RRACU or RAC
Position to deletion CIMS
F
RAC
...-3-2-1 0 1...
Insertion CIMS Site
cDNA
Enrichment of RAC
Position to insertion CIMS
G
RRACU
RAC
...-3-2-1 0 1...
CITS Site
cDNA
Enrichment of RRACU or RAC
Postion to CITS

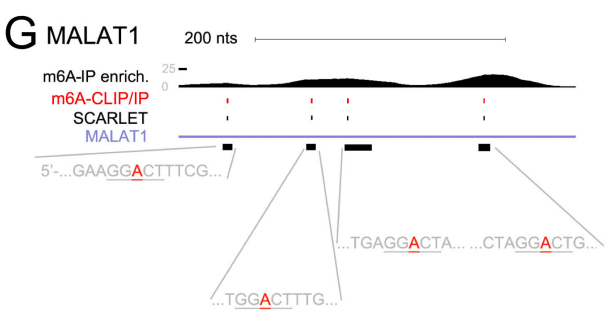

A
Peak Site
RRACU
RAC
...-3 -2 -1 0 1 2 3...
GGACU
p < 10-400
Enrichment of RRACU or RAC
Position to Peak Site
B
CIMS&CITS sites
cDNA
RRACU
RAC
...-3-2-1 0 1...
GGACU
p < 10-380
Enrichment of RRACU or RAC
Position to CIMS


C
MITS sites
cDNA
m6A
RRACU
RAC
...0 1 2 3 4...
GGACU
p < 10-400
Enrichment of RRACU or RAC
Position to MITS

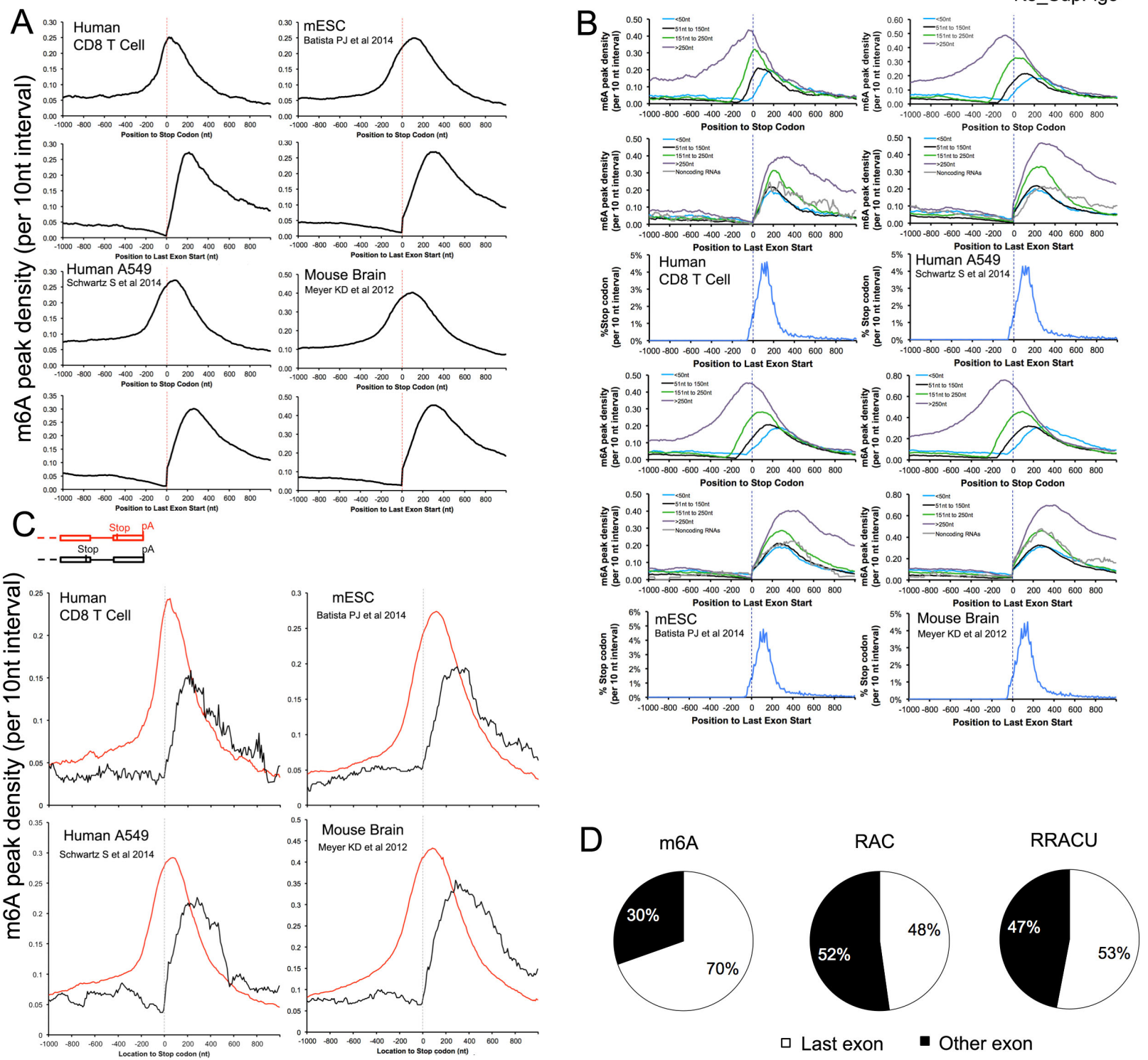

D
Number of Sites
118927
Peak Site m6As
CIMS&CITS m6As
MITS m6As
All m6As
5545
1535
14601
19682
All RACs

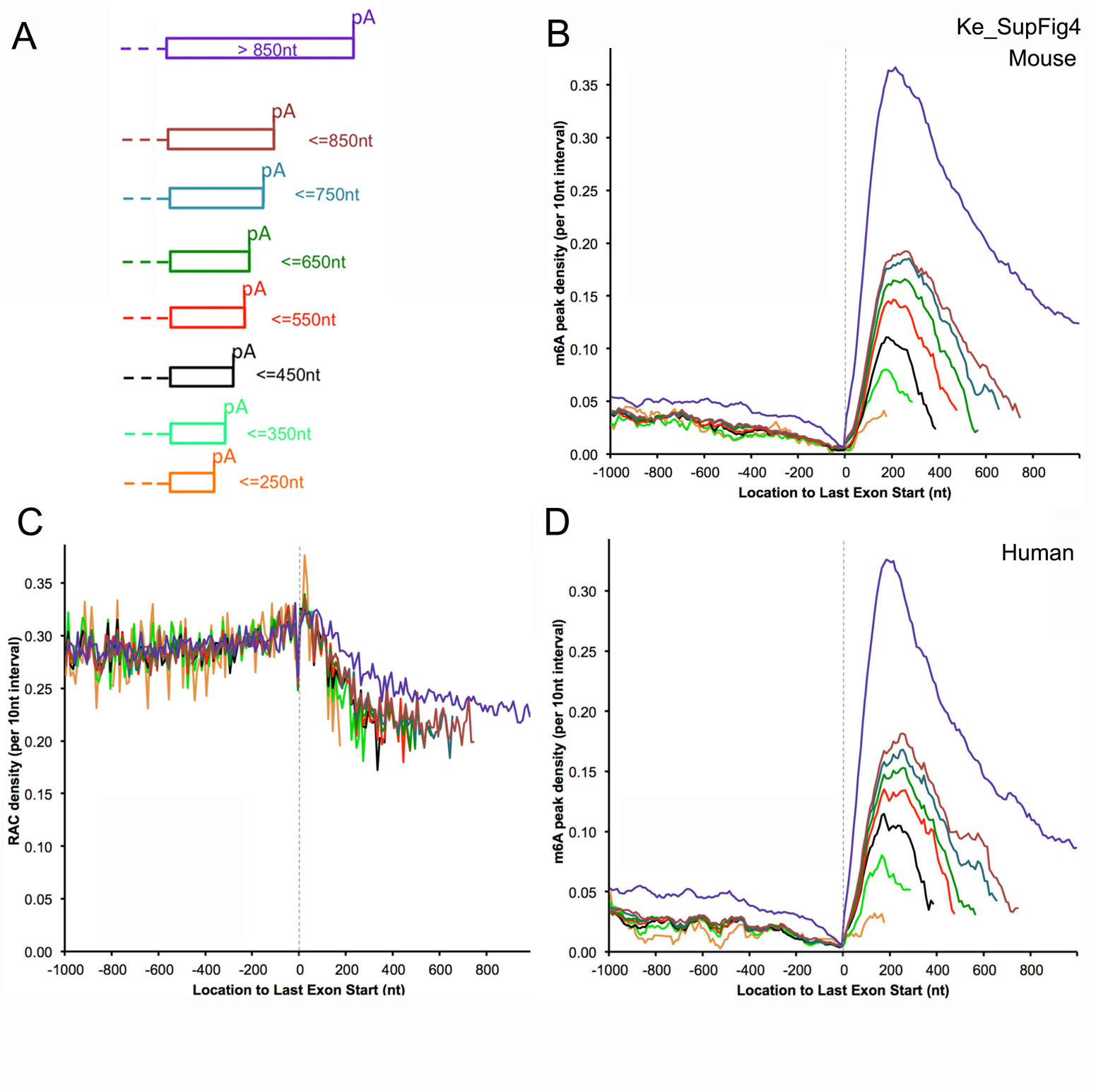

E
Peak Site m6A
(7,117)
CIMS&CITS m6A
(4,305)
MITS m6A
(21,779)
1497
435
178
835
Peak Site m6A
(5,545)
CIMS&CITS m6A
(1,535)
MITS m6A
(14,601)
1351
182
104
258


F
m6As
control RACs
Conservation(phloP)
All m6As
Peak Site m6As
CIMS&CITS m6As
MITS m6As

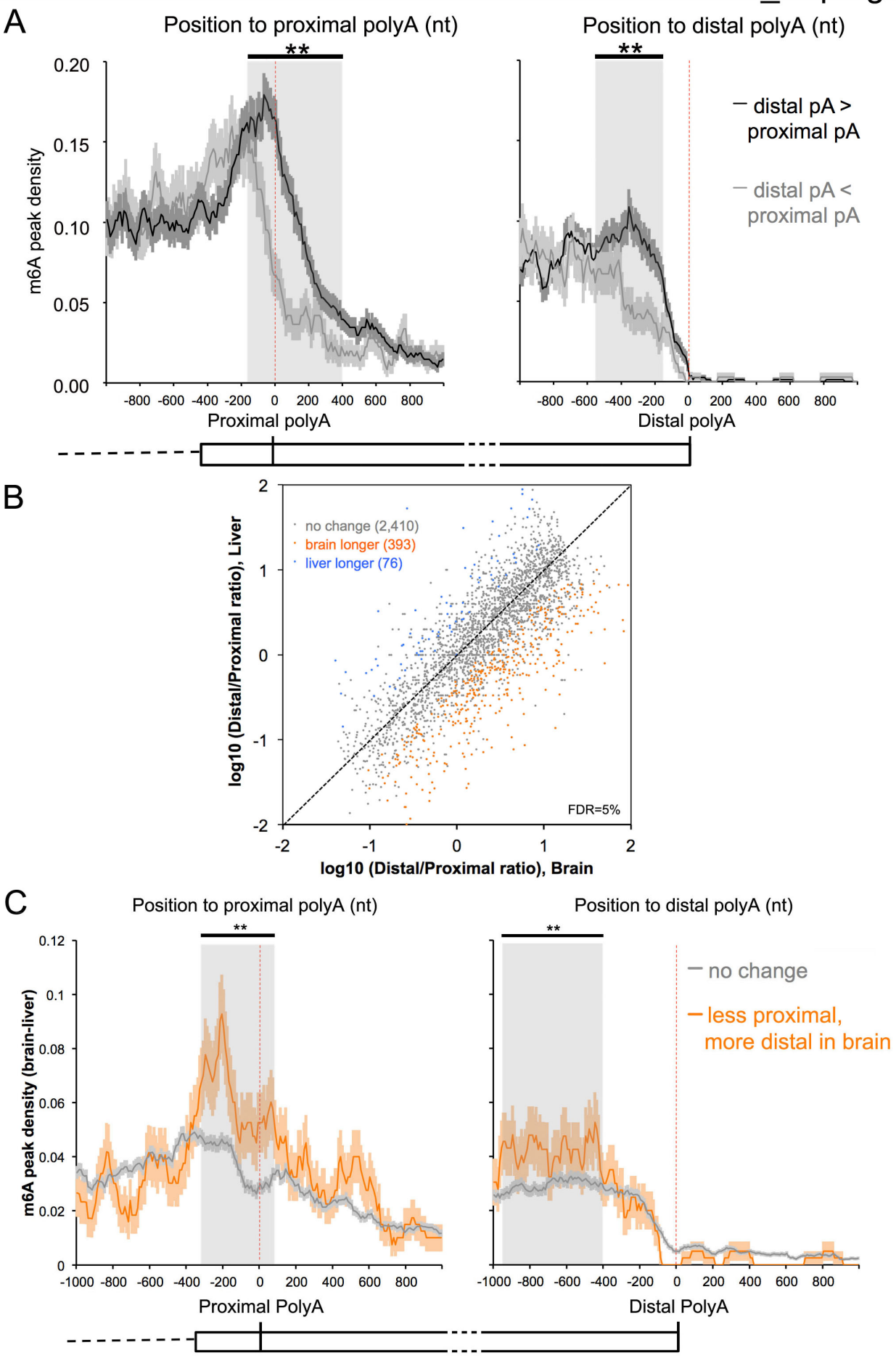

G MALAT1
200 nts
m6A-IP enrich.
m6A-CLIP/IP
SCARLET
MALAT1
5'-...GAAGGACTTTCG...
...TGAGGACTA...
...CTAGGACTG...
...TGGACTTTG...

Ke_SupFig3

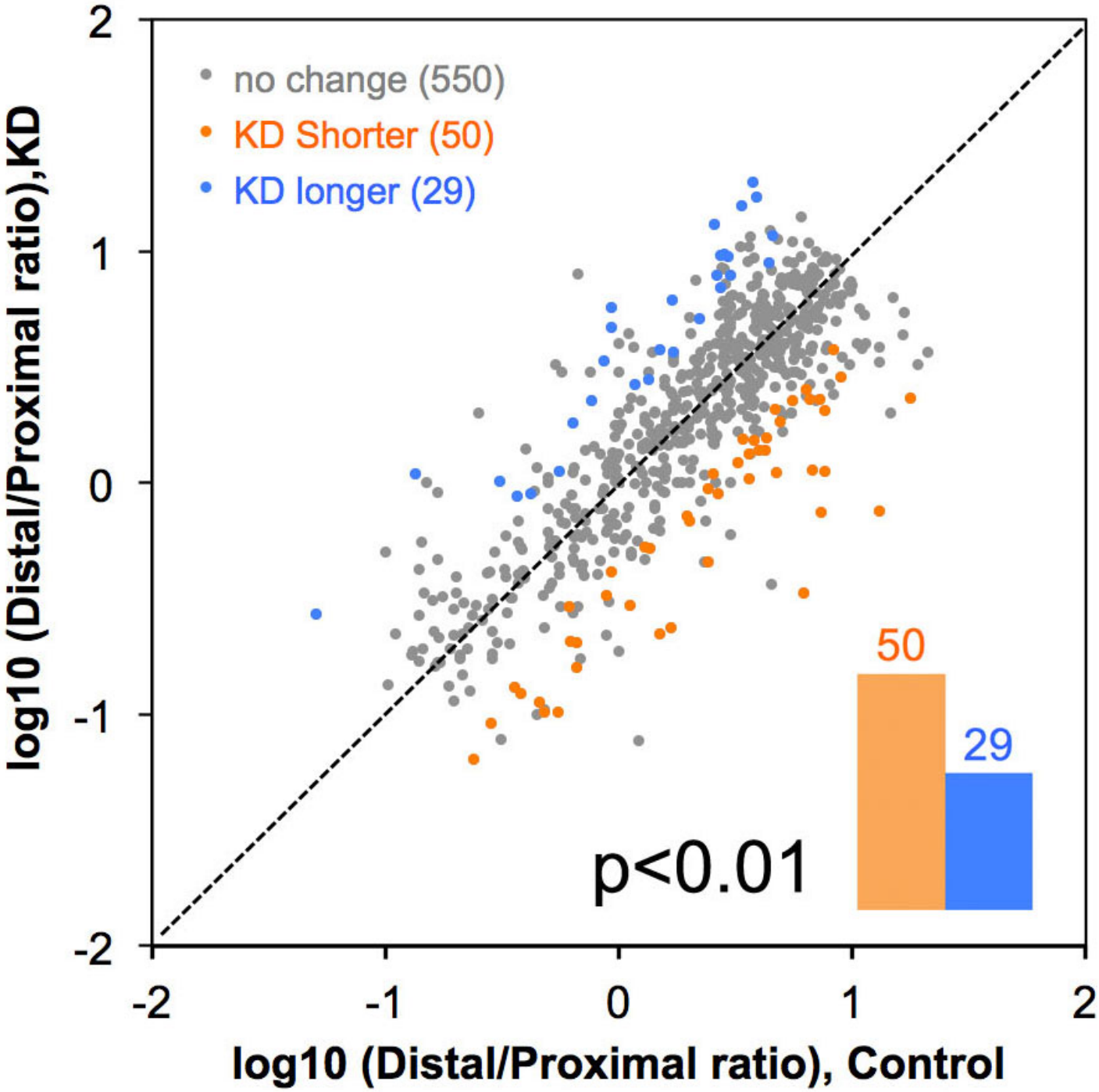

A
m6A peak density (per 10nt interval)
Human CD8 T Cell
mESC
Batista PJ et al 2014
Human A549
Schwartz S et al 2014
Mouse Brain
Meyer KD et al 2012
Position to Stop Codon (nt)
Position to Last Exon Start (nt)
B
m6A peak density (per 10 nt interval)
% Stop codon (per 10 nt interval)
<50nt
51nt to 150nt
151nt to 250nt
>250nt
Noncoding RNAs
Position to Stop Codon
Position to Last Exon Start
C
Stop
pA
Location to Stop codon (nt)
D
m6A
RAC
RRACU
30%
70%
52%
48%
47%
53%
Last exon
Other exon


A
pA
> 850nt
pA
<=850nt
pA
<=750nt
pA
<=650nt
pA
<=550nt
pA
<=450nt
pA
<=350nt
pA
<=250nt
B
Ke_SupFig4
Mouse
m6A peak density (per 10nt interval)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-1000 -800 -600 -400 -200 0 200 400 600 800
Location to Last Exon Start (nt)
C
RAC density (per 10nt interval)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-1000 -800 -600 -400 -200 0 200 400 600 800
Location to Last Exon Start (nt)
D
Human
m6A peak density (per 10nt interval)
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-1000 -800 -600 -400 -200 0 200 400 600 800
Location to Last Exon Start (nt)

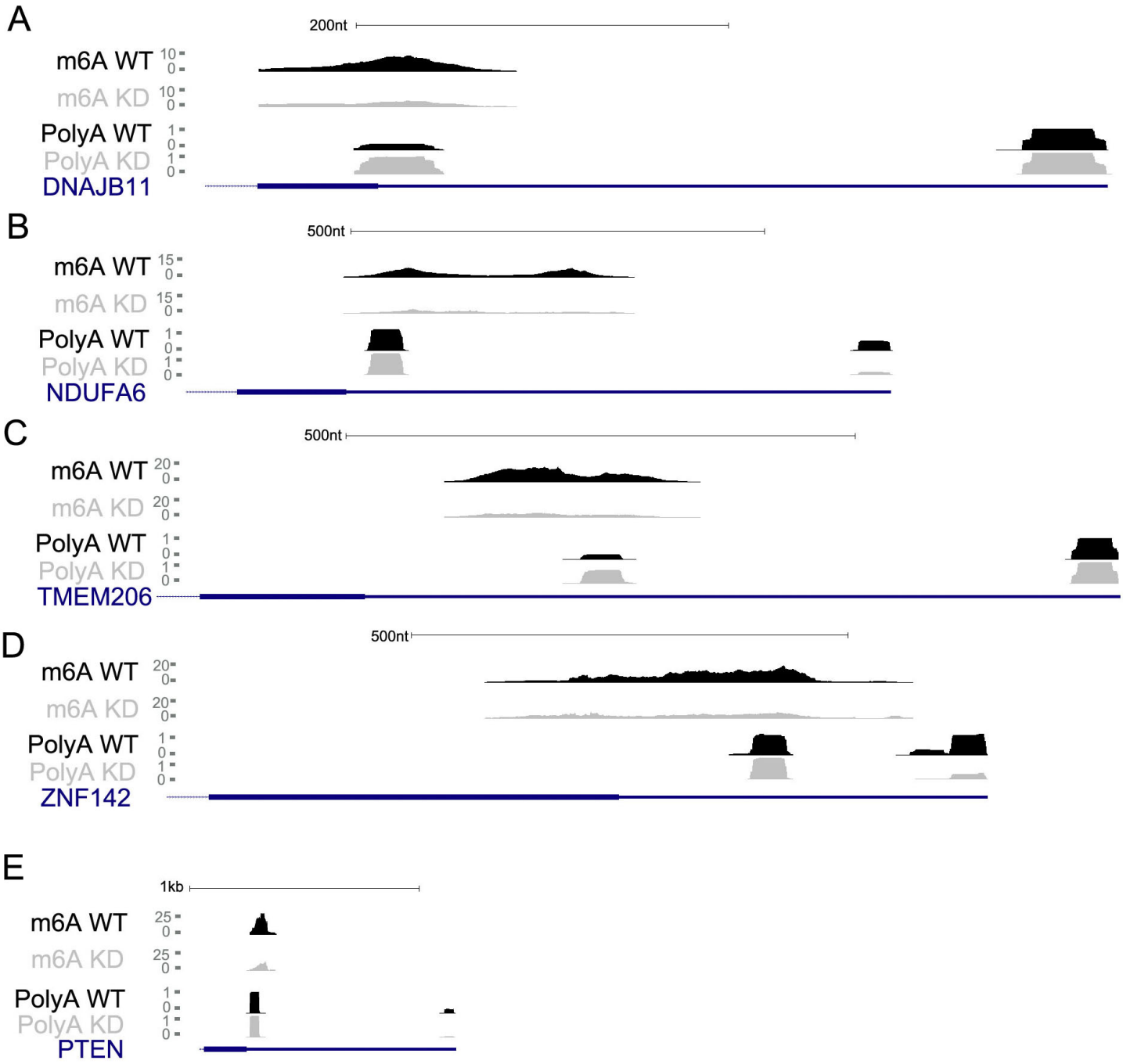
Ke_SupFig5


A
Position to proximal polyA (nt)
Position to distal polyA (nt)
**
**
m6A peak density
0.20
0.15
0.10
0.05
0.00
-800 -600 -400 -200 0 200 400 600 800
Proximal polyA
Distal polyA
distal pA > proximal pA
distal pA < proximal pA
B
log10 (Distal/Proximal ratio), Liver
log10 (Distal/Proximal ratio), Brain
no change (2,410)
brain longer (393)
liver longer (76)
FDR=5%
2
1
0
-1
-2
C
Position to proximal polyA (nt)
Position to distal polyA (nt)
**
**
m6A peak density (brain-liver)
0.12
0.1
0.08
0.06
0.04
0.02
0
-1000 -800 -600 -400 -200 0 200 400 600 800
Proximal PolyA
Distal PolyA
no change
less proximal, more distal in brain

Ke_SupFig6


no change (550)
KD Shorter (50)
KD longer (29)
50
29
p<0.01
log10 (Distal/Proximal ratio),KD
log10 (Distal/Proximal ratio), Control
2
1
0
-1
-2
-2
-1
0
1
2

Ke_SupFig7


A
200nt
m6A WT
m6A KD
PolyA WT
PolyA KD
DNAJB11
B
500nt
m6A WT
m6A KD
PolyA WT
PolyA KD
NDUFA6
C
500nt
m6A WT
m6A KD
PolyA WT
PolyA KD
TMEM206
D
500nt
m6A WT
m6A KD
PolyA WT
PolyA KD
ZNF142
E
1kb
m6A WT
m6A KD
PolyA WT
PolyA KD
PTEN