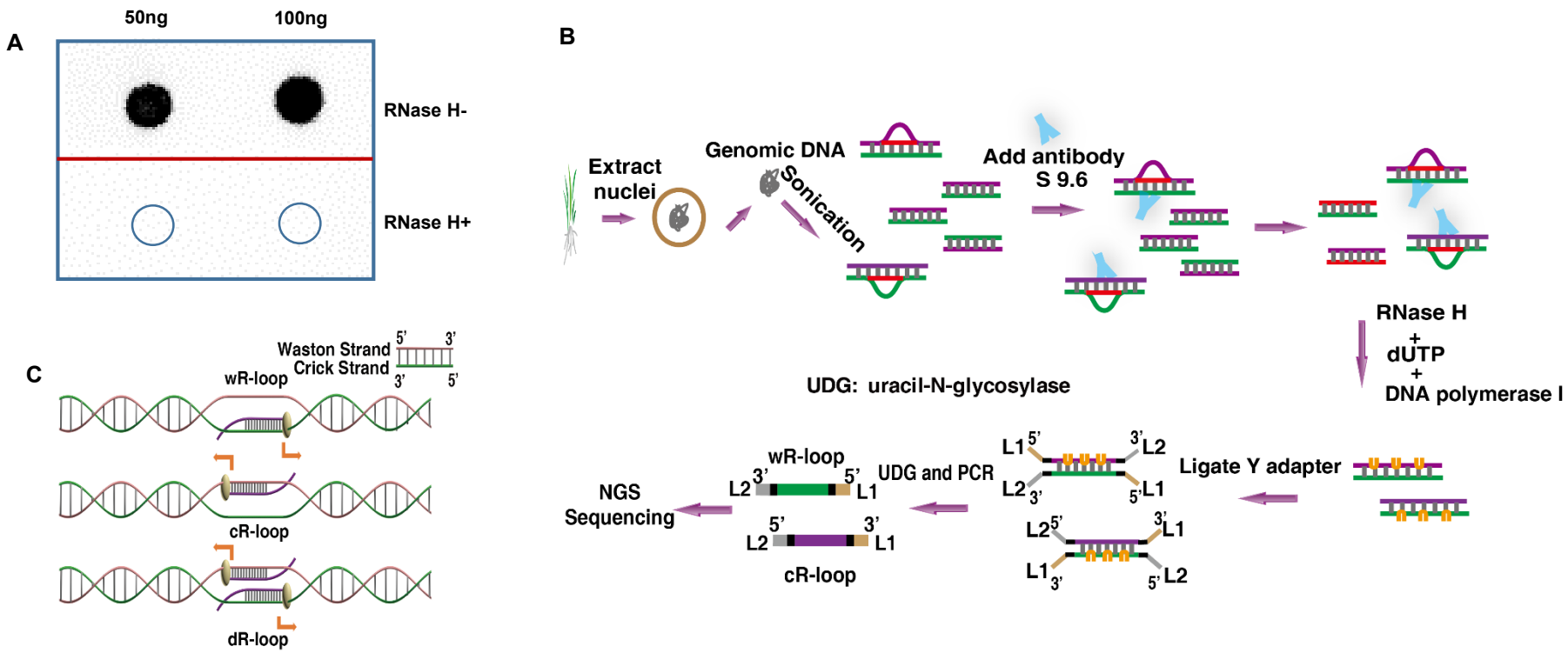


Figure S1. Adaptation of DRIP-seq in the rice genome. (A) Dot blotting assay. Genomic DNA (50 or 100 ng) treated with (RNase H+) or without (RNase H-) RNase H was loaded onto an Amersham Hybond-N^{+}-nylon membrane before it was probed with anti-S9.6 antibody (1:2500 dilution). (B) Schematic illustration of the primary procedures of the adapted DRIP-seq methodology. (C) Illustration of three types of R-loops: wR-loops represent R-loops formed with the Watson strand; and cR-loops represent R-loops formed with the Crick strand; dR-loops represent R-loops with a wR-loop and cR-loop in the same gene region.

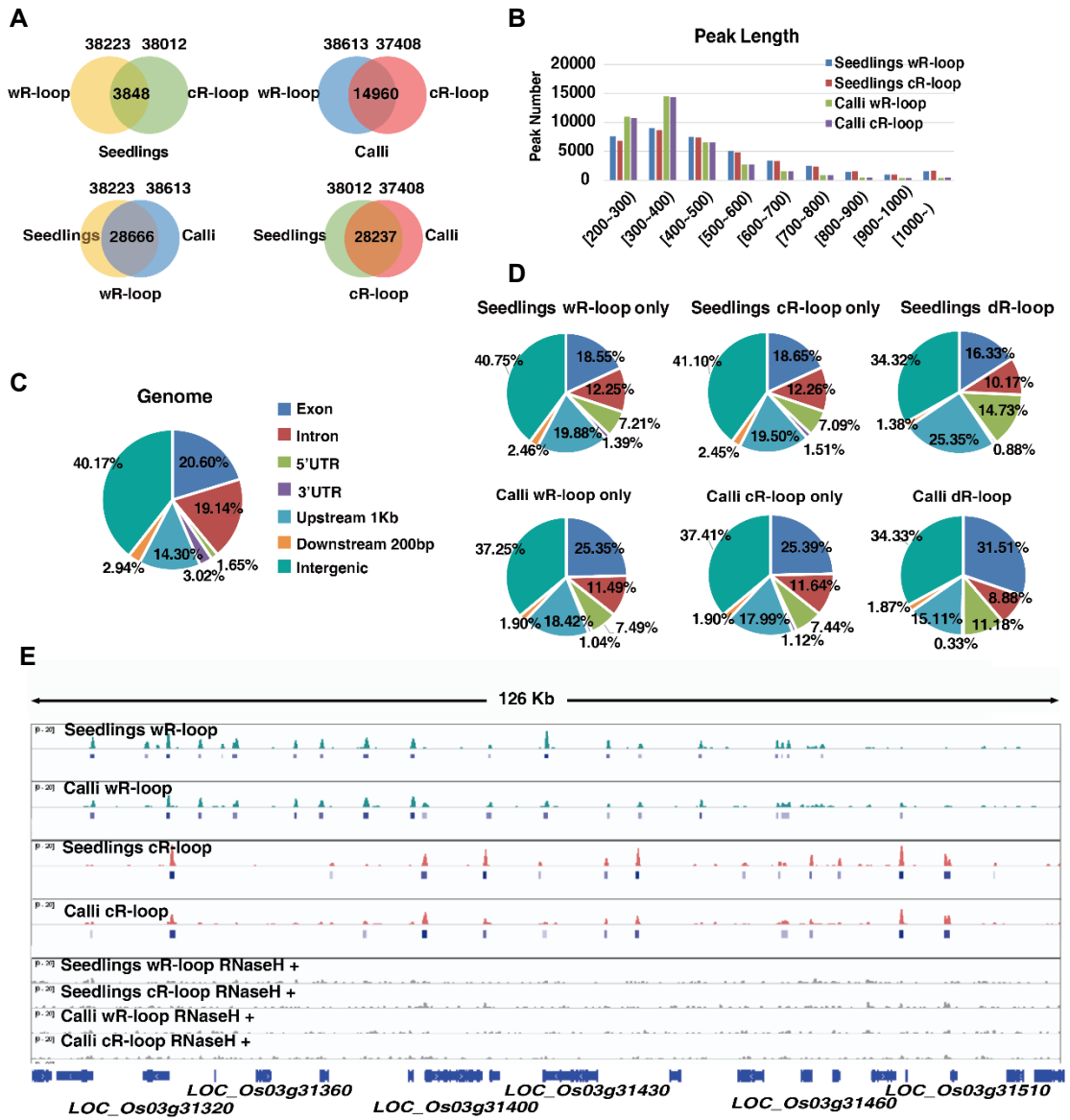

A
38223 38012
wR-loop
3848
cR-loop
Seedlings
38613 37408
wR-loop
14960
cR-loop
Calli
38223 38613
Seedlings
28666
Calli
wR-loop
38012 37408
Seedlings
28237
Calli
cR-loop
B
Peak Length
Peak Number
20000
15000
10000
5000
0
Seedlings wR-loop
Seedlings cR-loop
Calli wR-loop
Calli cR-loop
[200-300)
[300-400)
[400-500)
[500-600)
[600-700)
[700-800)
[800-900)
[900-1000)
[1000-)
C
Genome
40.17%
20.60%
19.14%
14.30%
1.65%
3.02%
2.94%
Exon
Intron
5'UTR
3'UTR
Upstream 1Kb
Downstream 200bp
Intergenic
D
Seedlings wR-loop only
40.75%
18.55%
12.25%
7.21%
19.88%
1.39%
2.46%
Seedlings cR-loop only
41.10%
18.65%
12.26%
7.09%
19.50%
1.51%
2.45%
Seedlings dR-loop
34.32%
16.33%
10.17%
14.73%
1.38%
25.35%
0.88%
Calli wR-loop only
37.25%
25.35%
11.49%
7.49%
18.42%
1.04%
1.90%
Calli cR-loop only
37.41%
25.39%
11.64%
7.44%
17.99%
1.12%
1.90%
Calli dR-loop
34.33%
31.51%
8.88%
11.18%
15.11%
0.33%
1.87%
E
126 Kb
Seedlings wR-loop
Calli wR-loop
Seedlings cR-loop
Calli cR-loop
Seedlings wR-loop RNaseH +
Seedlings cR-loop RNaseH +
Calli wR-loop RNaseH +
Calli cR-loop RNaseH +
LOC_Os03g31320
LOC_Os03g31360
LOC_Os03g31400
LOC_Os03g31430
LOC_Os03g31460
LOC_Os03g31510

Fig. S2

Figure S2. Genomic characterizations of R-loops. (A) Venn diagrams showing the numbers of wR-loops, cR-loops and dR-loops in seedlings and calli. dR-loops comprise wR-loops and cR-loops that overlapped at least 1bp. (B) Peak numbers of wR-loops and cR-loops in different bins with various peak lengths in seedlings and calli. (C) Percentages of the total lengths of different sub-genomic units in the rice genome, including: default promoters (1 kb upstream of TSSs), 5'UTRs, exons, introns, 3'UTRs, 200 bp downstream of the TESs, and intergenic regions. (D) Percentage of each of the seven sub-genomic regions associated with wR-loops, cR-loops and dR-loops. The color coding is the same as that used in (C). (E) An Integrative Genomics Viewers (IGV) snapshot spanning 126 kb window from the chromosome 3 illustrating R-loop peaks in the rice genome. Each significant R-loop identified was marked with a solid blue rectangular box below the peak.

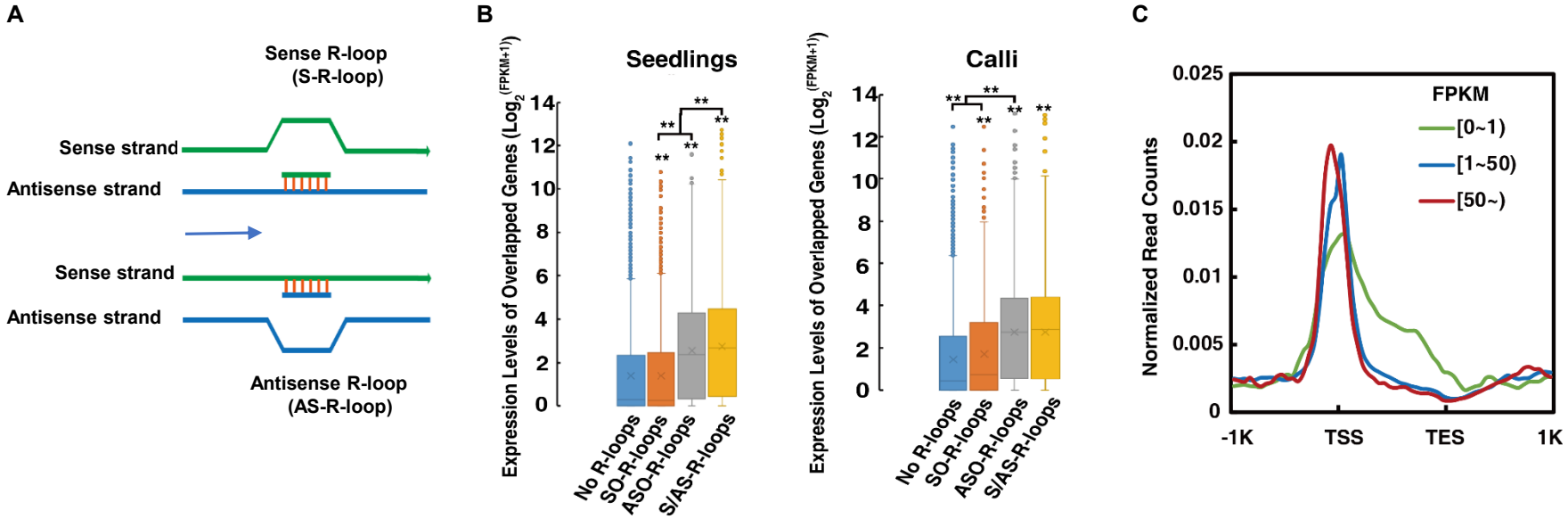
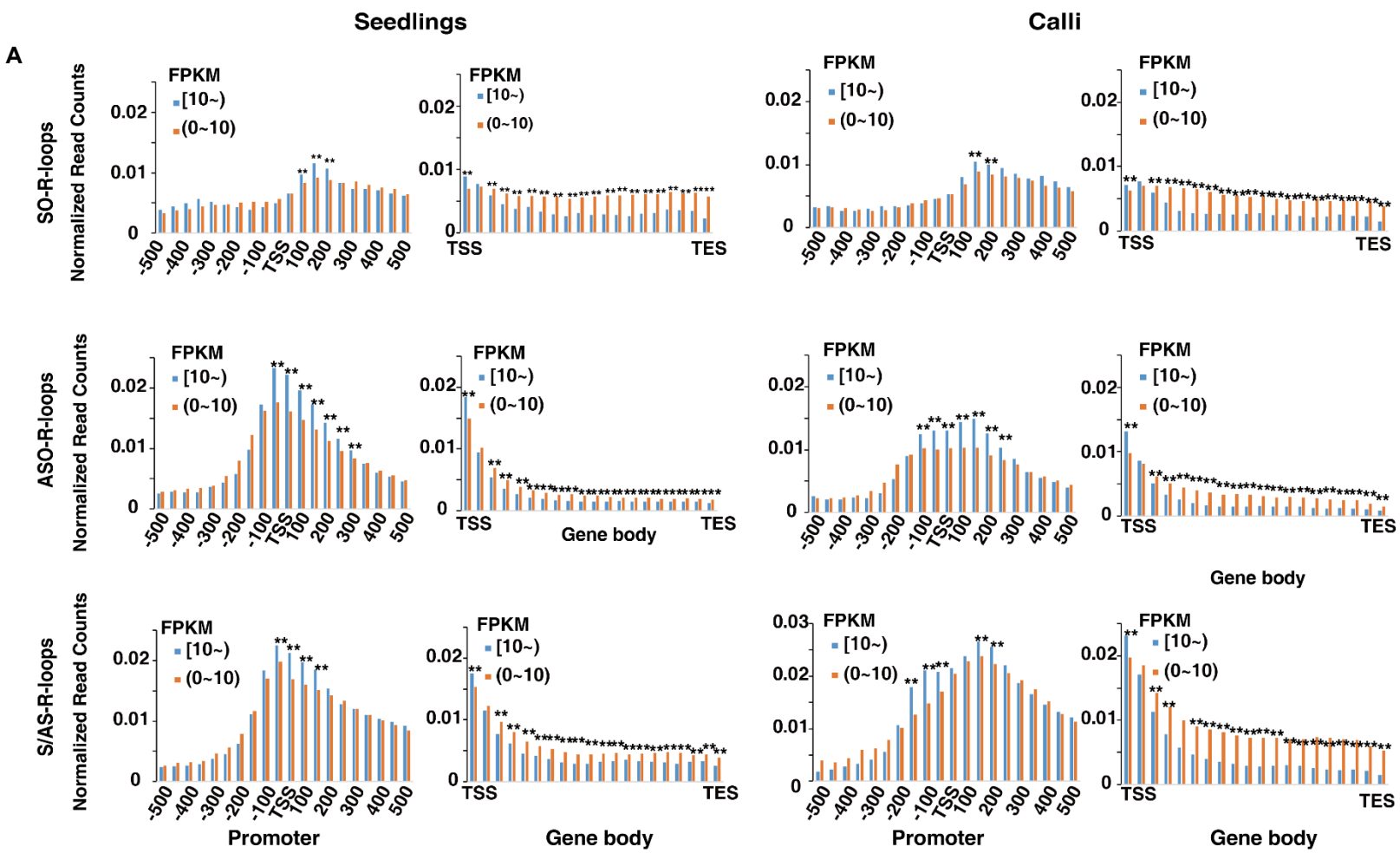


Figure S3. The relationships between the expression levels of R-loop overlapping genes and R-loop types. (A) Diagram illustrating the formation of sense and antisense R-loops in the rice genome. (B) Comparisons of the expression levels of non-TE genes without R-loops (control) to those with SO-R-loops, ASO-R-loops or S/AS-R-loops in seedling and callus tissues. (C) All genes with S/AS-R-loops located within ± 500 bp of the TSSs in calli were divided into three sub-classes according to FPKM values. R-loop read densities were then plotted for each of these sub-classes in this region. **denotes *p*-value < 0.01 was conducted using the Wilcoxon rank-sum test.


A
Seedlings
Calli
SO-R-loops
ASO-R-loops
S/AS-R-loops
Normalized Read Counts
FPKM
[10~)
(0~10)
TSS
TES
Gene body
Promoter

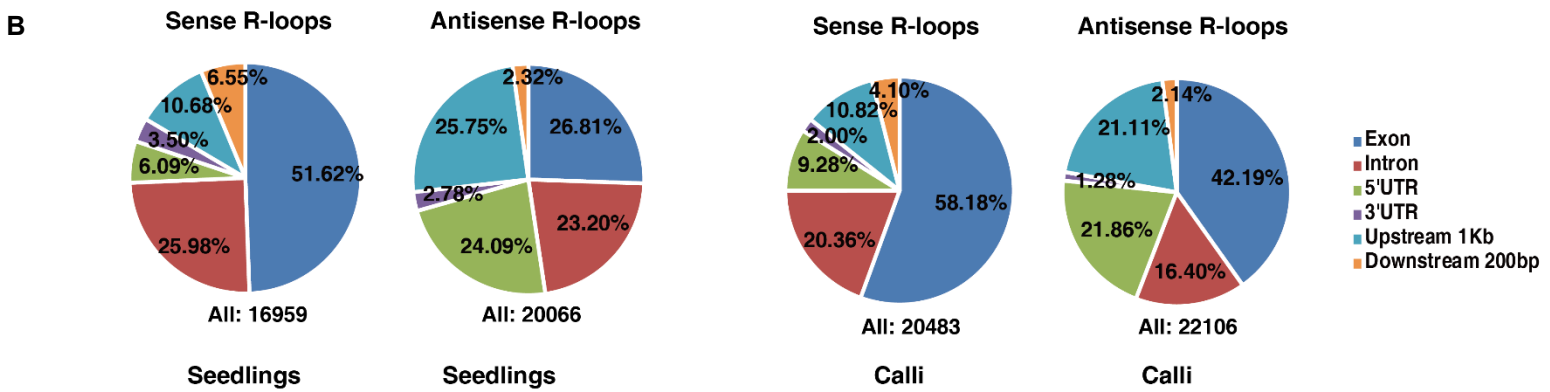

B
Sense R-loops
Antisense R-loops
Sense R-loops
Antisense R-loops
51.62%
25.98%
6.09%
3.50%
10.68%
6.55%
All: 16959
Seedlings
26.81%
23.20%
24.09%
2.78%
25.75%
2.32%
All: 20066
Seedlings
58.18%
20.36%
9.28%
2.00%
10.82%
4.10%
All: 20483
Calli
42.19%
16.40%
21.86%
1.28%
21.11%
2.14%
All: 22106
Calli
Exon
Intron
5'UTR
3'UTR
Upstream 1Kb
Downstream 200bp

Fig. S4

Figure S4. Genic distributions of sense and antisense R-loops and their relationship to gene expression. (A) Plots illustrating that normalized read counts (R-loop densities) vary with the type of R-loop, its genic position, and the level of gene expression. Gene expression is based on FPKM values, taken as low (FPKM 0-10) and high (FPKM $\geq$ 10) expression levels. (B) Genomic distribution of sense- and antisense- R-loops in seedlings (top) and calli (bottom). Normalized DRIP-seq read counts were calculated in step of 50 bp and ** denotes a *p*-value < 0.01 was conducted using the Wilcoxon rank-sum test.

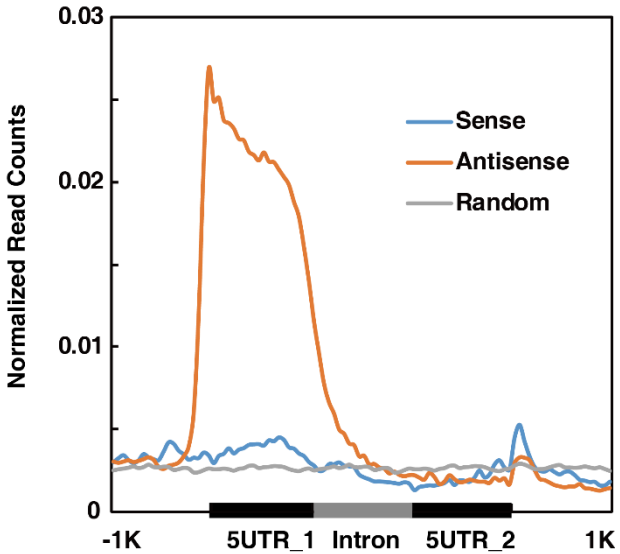


Figure S5. Distribution of sense and antisense R-loops in 5'UTRs contained introns. Normalized sense and antisense R-loops read counts were plotted across 5'UTRs that are split by introns and ± 1 kb region. The y-axis represents normalized read counts of R-loops, while the x-axis represents ± 1 kb of 5'UTRs; each 5'UTR is separated by an intron.

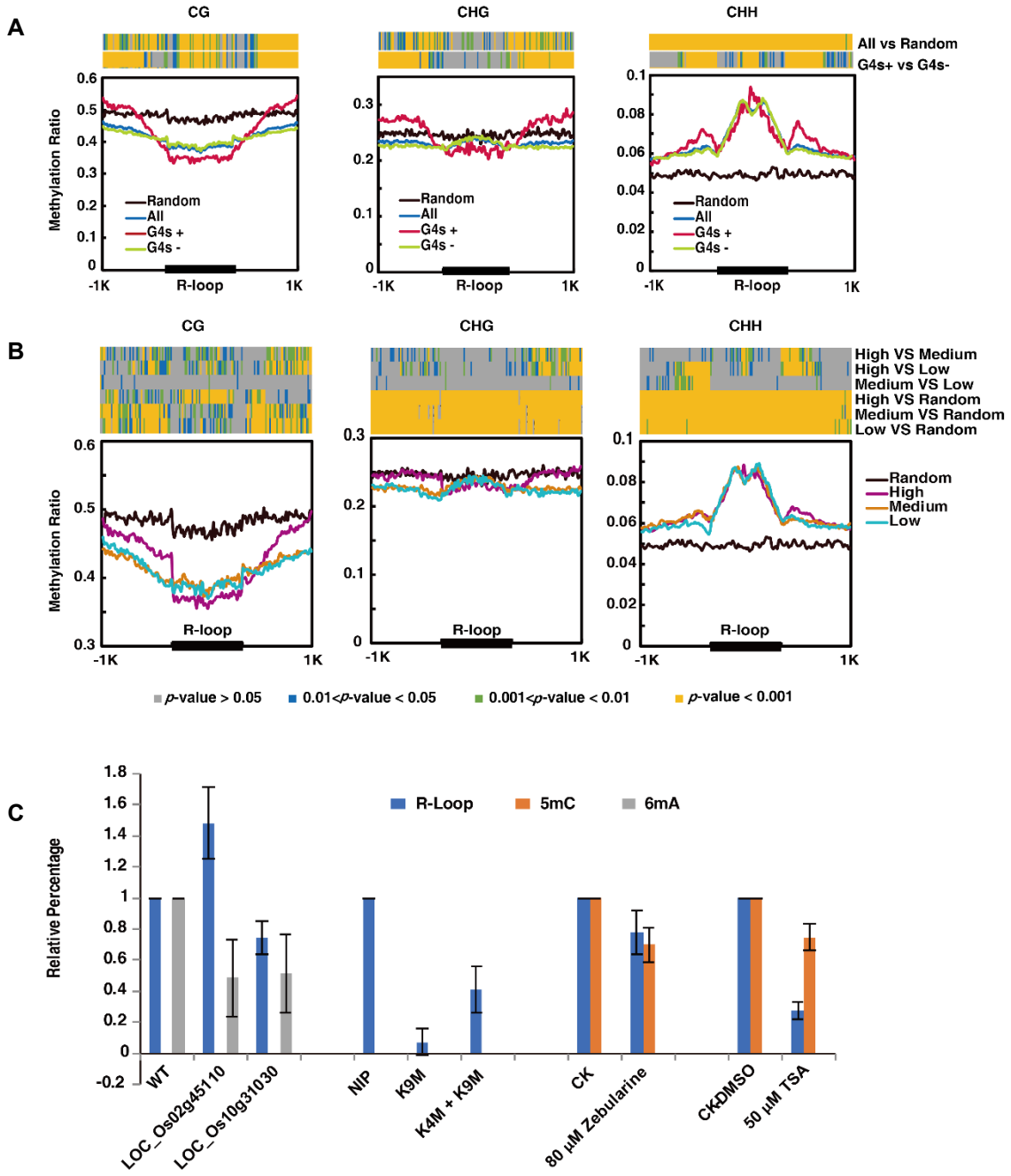

A
CG
CHG
CHH
All vs Random
G4s+ vs G4s-
Methylation Ratio
Random
All
G4s +
G4s -
-1K
R-loop
1K
B
High VS Medium
High VS Low
Medium VS Low
High VS Random
Medium VS Random
Low VS Random
Random
High
Medium
Low
p-value > 0.05
0.01<p-value < 0.05
0.001<p-value < 0.01
p-value < 0.001
C
R-Loop
5mC
6mA
Relative Percentage
WT
LOC_Os02g45110
LOC_Os10g31030
NIP
K9M
K4M + K9M
CK
80 μM Zebularine
CK-DMSO
50 μM TSA

Fig. S6

Figure S6. The relationship between DNA methylation levels within R-loop regions. (A) Average CG, CHG and CHH methylation ratios were calculated around ±1 kb of all R-loops, R-loops with G4s+, R-loops with G4s- and random controls. The window size in flanking regions was 20 bp, and all R-loop peaks were equally divided into 50 bins. (B) All R-loops were divided into three subgroups with different DNA 5mC levels (high, medium and low) according to reads counts. The average CG, CHG and CHH methylation ratios were calculated around ± 1 kb of each subgroup of R-loops. The window size in flanking regions was 20 bp, and all R-loop peaks were equally divided into 50 bins for calculation of DNA methylation rates. The colored strips located on the top of each plot reflect the *p*-values from the Wilcoxon rank-sum significance test; Grey bands represent *p*-values > 0.05, blue bands represent 0.05 < *p*-values < 0.01, green bands represent 0.001 < *p*-values < 0.01 and yellow bands represent *p*-values < 0.001. Each comparison is indicated with a specific label. The y-axis represents methylation ratio, while the x-axis represents the whole R-loop regions. (C) Statistical analyses of S9.6, 5mC and 6mA dot blotting signals. 50 ng DNA was used for dot blotting, and the results of two repeats were quantified.

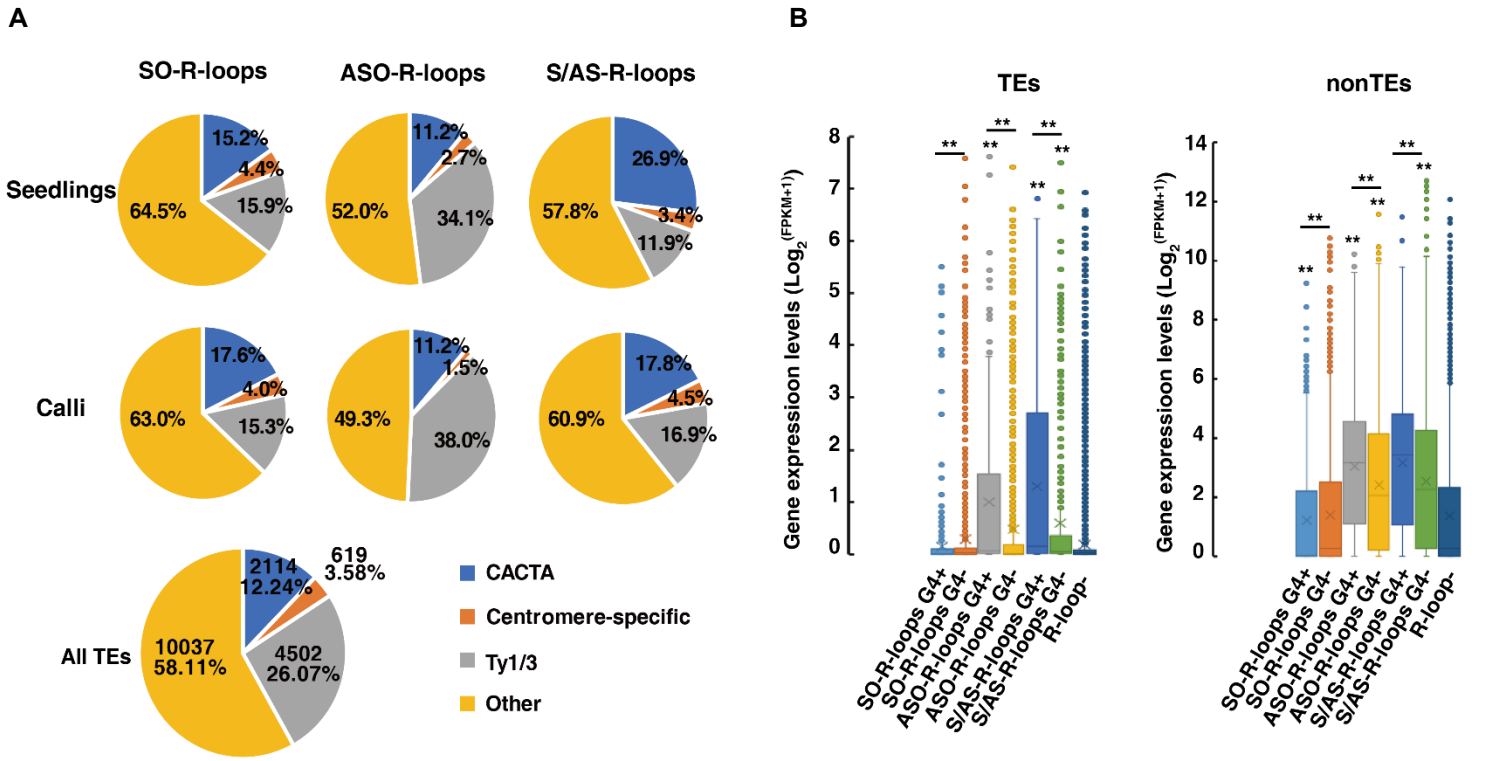


Figure S7. The percentages of four types of TEs and the expression levels of TEs and non-TEs with different types of R-loops that have G4+ or G4-. (A) Distribution of TEs with different types of R-loops in the rice genome. All TEs were classified into four categories, CACTA transposons, centromere-specific retrotransposons, Ty1/3 retrotransposons and others. (B) The expression levels of TEs (left) and non-TEs (right) overlapping SO-R-Loops, ASO-R-Loops, S/AS-R-Loops and non-R-loops (R-loops-). In each case, R-loops with G4s (G4+) or without G4s (G4-) were shown separately for comparison and the significance of difference was tested using Wilcoxon rank-sum test.

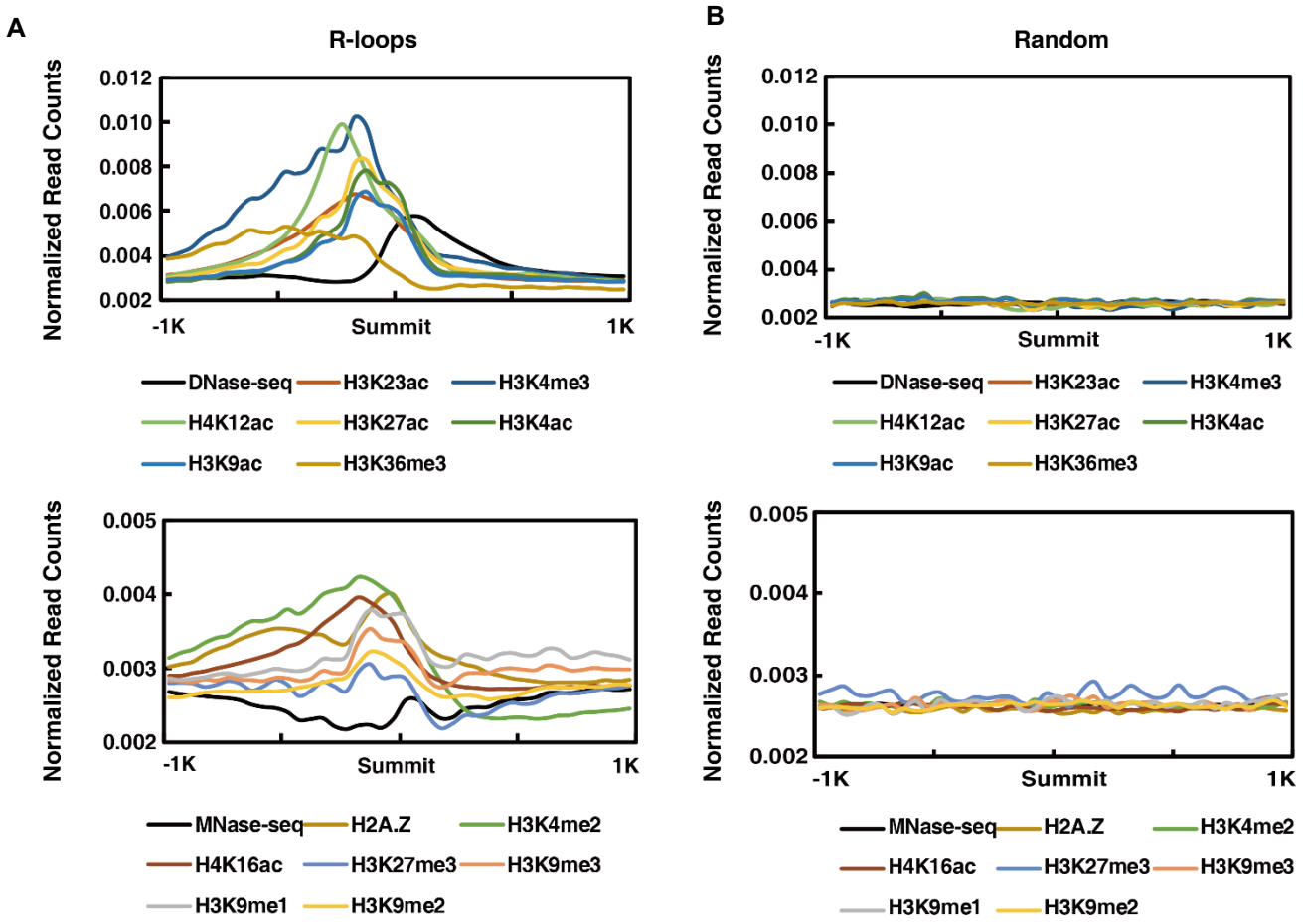


Figure S8. The enrichment of epigenomic marks in R-loops. (A) The normalized read counts from 16 epigenetic marks were plotted along ± 1 kb of the summits of the R-loop peaks. (B) 10,000 randomly selected 2 kb-regions were used as controls for comparison. A window size of 50 bp was used for calculating normalized read counts of each mark.

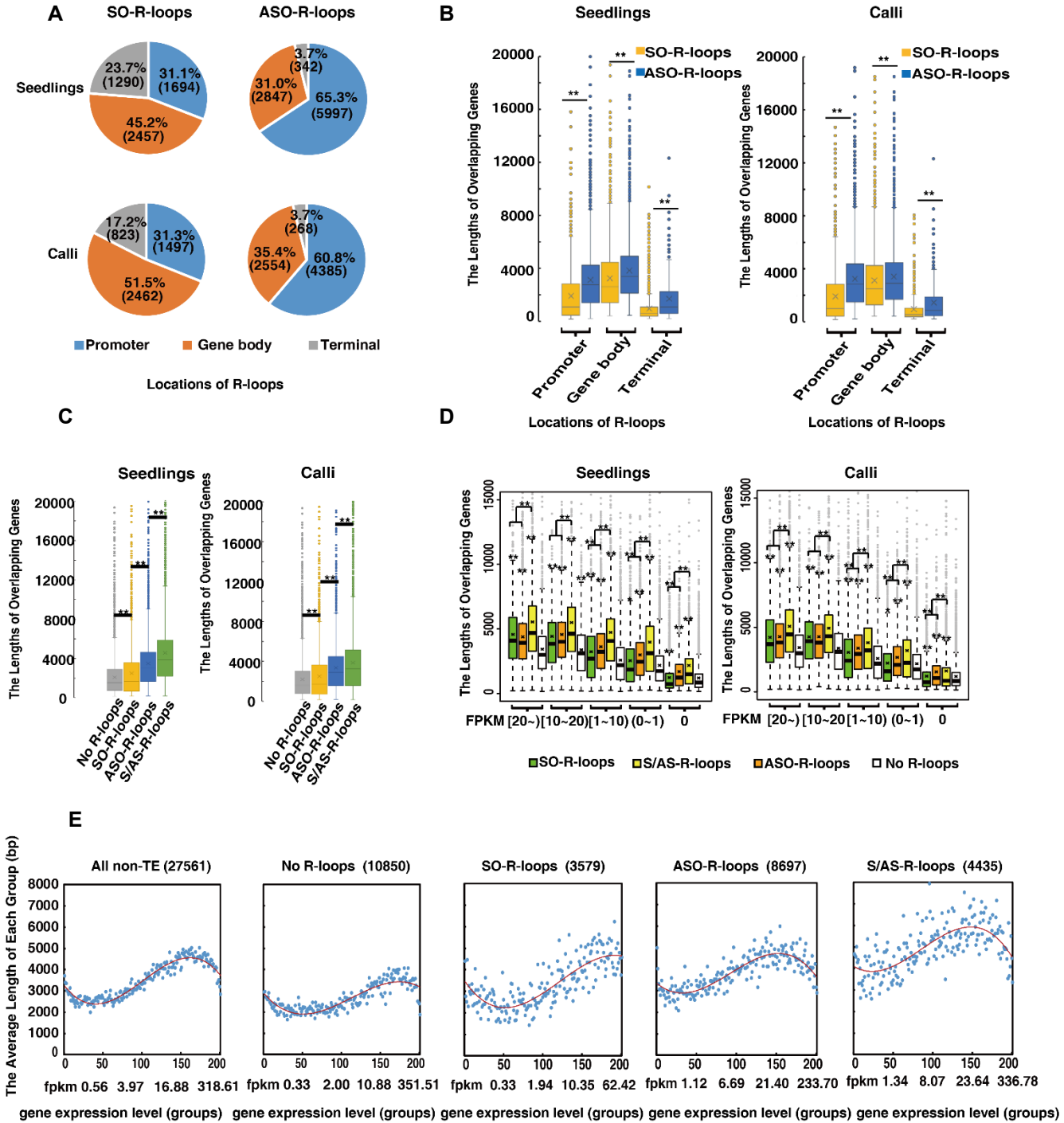
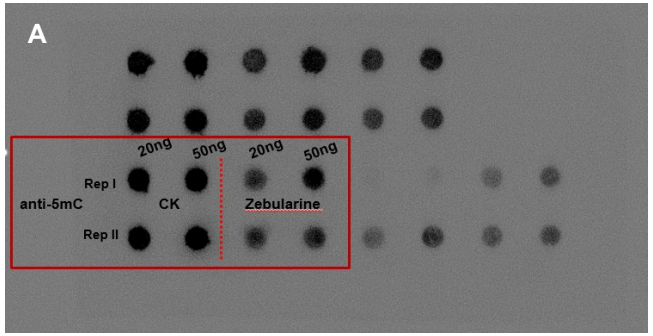
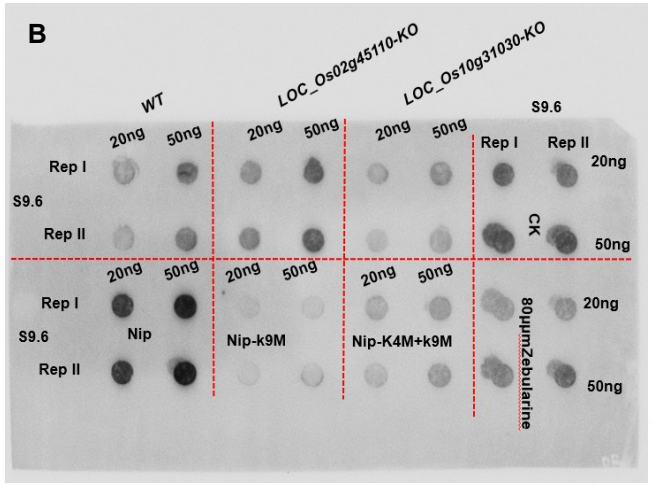
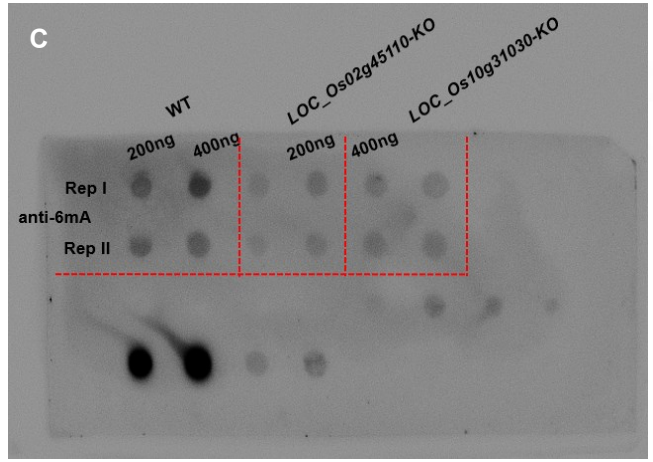

A
SO-R-loops
ASO-R-loops
Seedlings
23.7% (1290)
31.1% (1694)
45.2% (2457)
3.7% (342)
31.0% (2847)
65.3% (5997)
Calli
17.2% (823)
31.3% (1497)
51.5% (2462)
3.7% (268)
35.4% (2554)
60.8% (4385)
Promoter
Gene body
Terminal
Locations of R-loops
B
Seedlings
Calli
SO-R-loops
ASO-R-loops
The Lengths of Overlapping Genes
Promoter
Gene body
Terminal
Locations of R-loops
C
Seedlings
Calli
The Lengths of Overlapping Genes
No R-loops
SO-R-loops
ASO-R-loops
S/AS-R-loops
D
Seedlings
Calli
The Lengths of Overlapping Genes
FPKM [20~) [10~20) [1~10) (0~1) 0
SO-R-loops
S/AS-R-loops
ASO-R-loops
No R-loops
E
All non-TE (27561)
No R-loops (10850)
SO-R-loops (3579)
ASO-R-loops (8697)
S/AS-R-loops (4435)
The Average Length of Each Group (bp)
fpkm 0.56 3.97 16.88 318.61
fpkm 0.33 2.00 10.88 351.51
fpkm 0.33 1.94 10.35 62.42
fpkm 1.12 6.69 21.40 233.70
fpkm 1.34 8.07 23.64 336.78
gene expression level (groups)

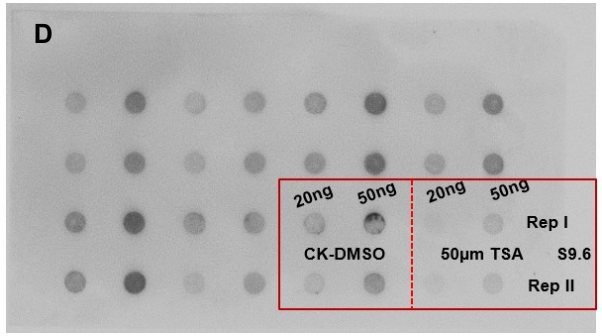
Fig. S9

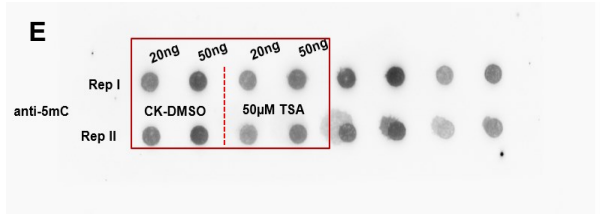
Figure S9. Association of R-loops with gene lengths. (A) Percentages of genes having one SO-R-loop or ASO-R-loop located in promoters (-1 kb ~ 200 bp of TSSs), gene bodies (200 bp of TSSs ~ -200 bp of TESs) or terminal regions (-200 bp ~ 1 kb of TESs). (B) Lengths of non-TE genes having only one SO-R-loop or ASO-R-loop located in each specific region. (C) Lengths of all non-TE genes overlapping SO-, ASO- or S/AS-R-loops. (D) Lengths of genes overlapped different types of R-loops with respect to the different expression levels (FPKM). (E) Correlation between the expression levels of R-loop overlapping genes and their lengths in seedlings. All 27,561 expressed non-TE genes (FPKM>0) were divided into 200 bins from the lowest to the highest expression levels in seedlings. (F) Correlation between expression levels of R-loop overlapping genes and their lengths in seedlings. Similar analysis was conducted in Supplemental Figure S4E using expressed genes (FPKM>0) with different types of R-loops and control genes without R-loops. ***p*-value < 0.01 was determined by Wilcoxon rank-sum test.

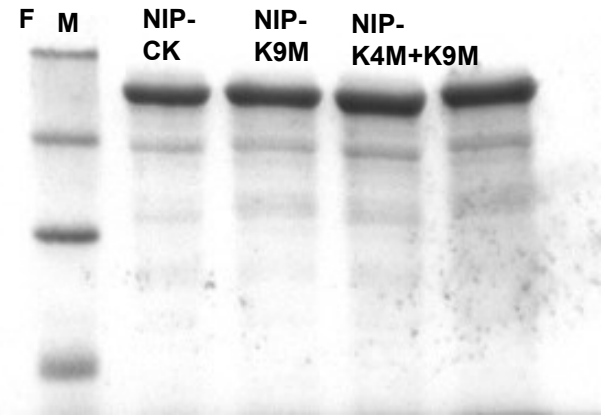

A
20ng
50ng
20ng
50ng
Rep I
anti-5mC
CK
Zebularine
Rep II

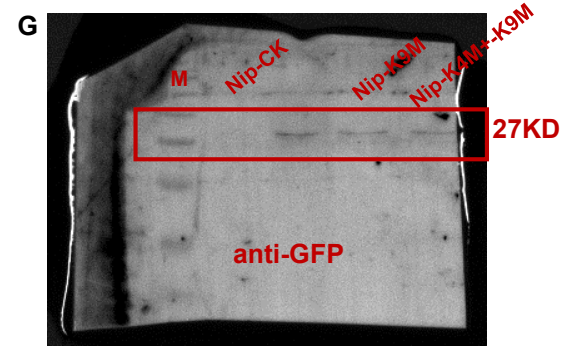

B
WT
LOC_Os02g45110-KO
LOC_Os10g31030-KO
S9.6
20ng
50ng
Rep I
Rep II
S9.6
CK
50ng
Nip
Nip-k9M
Nip-K4M+k9M
80μmZebularine

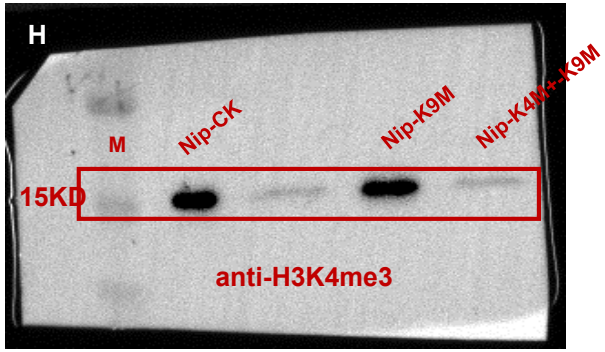

C
WT
LOC_Os02g45110-KO
LOC_Os10g31030-KO
200ng
400ng
Rep I
anti-6mA
Rep II

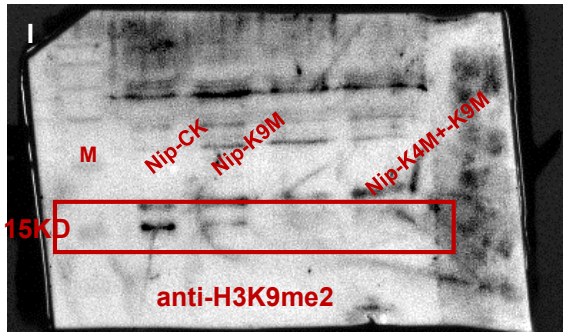

D
20ng
50ng
Rep I
CK-DMSO
50μm TSA
S9.6
Rep II

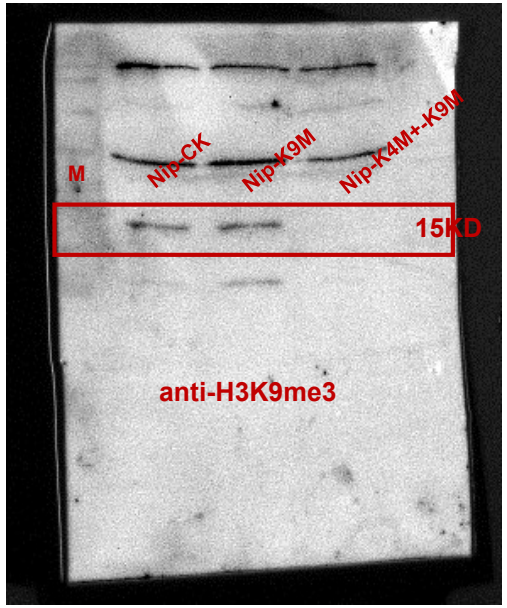

E
20ng
50ng
Rep I
anti-5mC
CK-DMSO
50μM TSA
Rep II


F
M
NIP-CK
NIP-K9M
NIP-K4M+K9M


G
M
Nip-CK
Nip-K9M
Nip-K4M+-K9M
27KD
anti-GFP


H
M
Nip-CK
Nip-K9M
Nip-K4M+-K9M
15KD
anti-H3K4me3


I
M
Nip-CK
Nip-K9M
Nip-K4M+-K9M
15KD
anti-H3K9me2


J
M
Nip-CK
Nip-K9M
Nip-K4M+-K9M
15KD
anti-H3K9me3

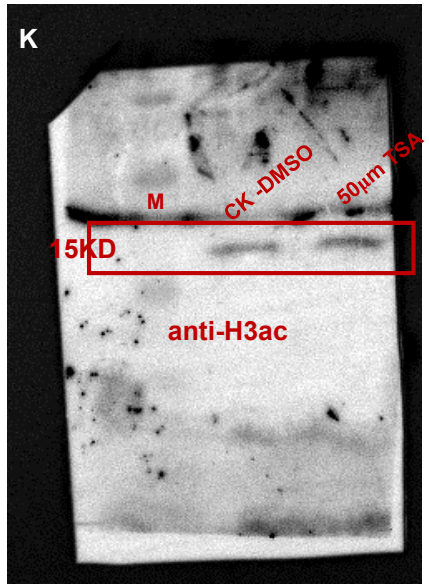

K
M
CK -DMSO
50μm TSA
15KD
anti-H3ac

Fig. S10

Figure S10. Original images for DNA and Histone dot blotting assays. (A) Original image for anti-5mC dot blotting assay, which was corresponding to Fig. 3D with anti-5mC antibody. (B) Original image for the dot blotting assay with anti-S9.6 antibody, which was corresponding to Fig. 3D, Fig. 4F and Fig. 5B with anti-S9.6 antibody, respectively. (C) Original image for the dot blotting assay with anti-6mA antibody, which was corresponding to Fig. 4F with anti-6mA antibody. (D) Original image for the dot blotting assay with anti-S9.6 antibody, which was corresponding to Fig. 5C with anti-S9.6 antibody. (E) Original image for the dot blotting assay with anti-5mC antibody, which was corresponding to Fig. 5C with anti-5mC antibody. (F) Original gel image with Commassie blue staining for indicating loading control of total protein, which was corresponding to Fig. 5B loading control. (G) Original western blot with anti-GFP antibody, which was corresponding to Fig. 5B with anti-GFP antibody. (H) Original western blot with anti-H3K4me3 antibody, which was corresponding to Fig. 5B with anti-H3K4me3 antibody. (I) Original western blot with anti-H3K9me2 antibody, which was corresponding to Fig. 5B with anti-H3K9me2 antibody. (J) Original western blot with anti-H3K9me3 antibody, which was corresponding to Fig. 5B with anti-H3K9me3 antibody. (K) Original western blot with anti-H3ac antibody, which was corresponding to Fig. 5C with anti-H3ac antibody. Dots/western bands showing in the main results were cut from each corresponding original image as indicated or marked with the red box. The specific western band was determined as the estimated molecular weight indicated in each original blot. All other unmarked dot signals or whole lane western bands were unrelated to this study.

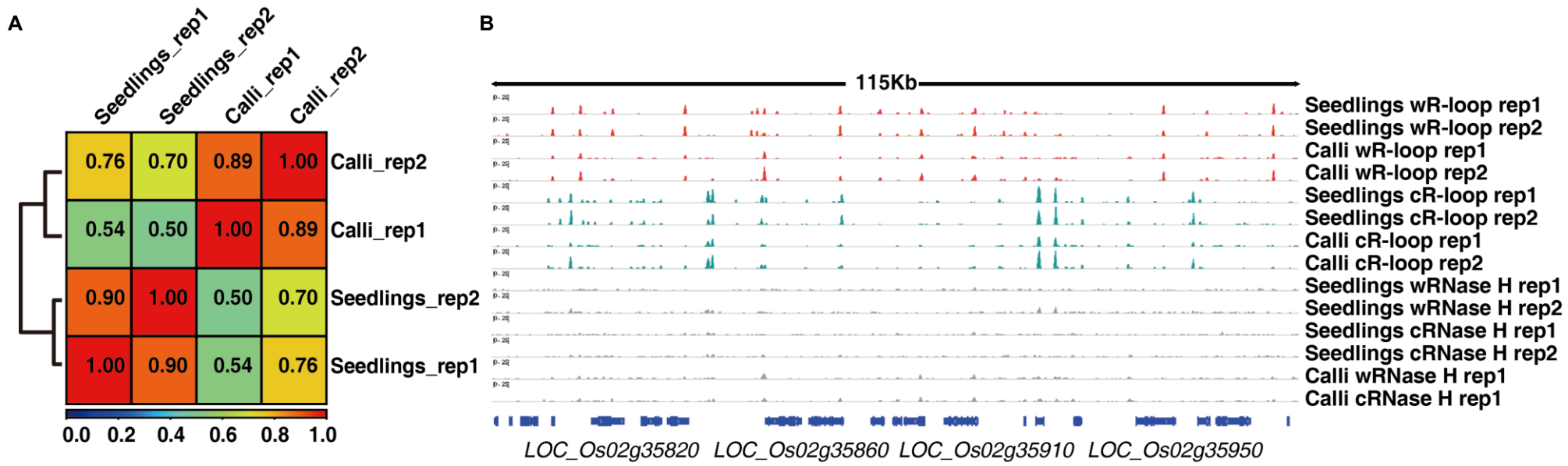


Figure S11. The correlation between two independent biological replicates from seedlings and calli. (A) Spearman's correlation of read densities (the bin size as 10 bp) was calculated between two independent biological replicates using the deepTools software. (B) Comparison of R-loop landscapes (wR-loops and cR-loops) between two biological replicates of the same tissue and also between two tissues with/without RNase H treatment. All snapshots of Integrative Genomics Viewer (IGV) span the same region of 115 kb in the rice genome that harbors 20 genes (only four genes are labeled).

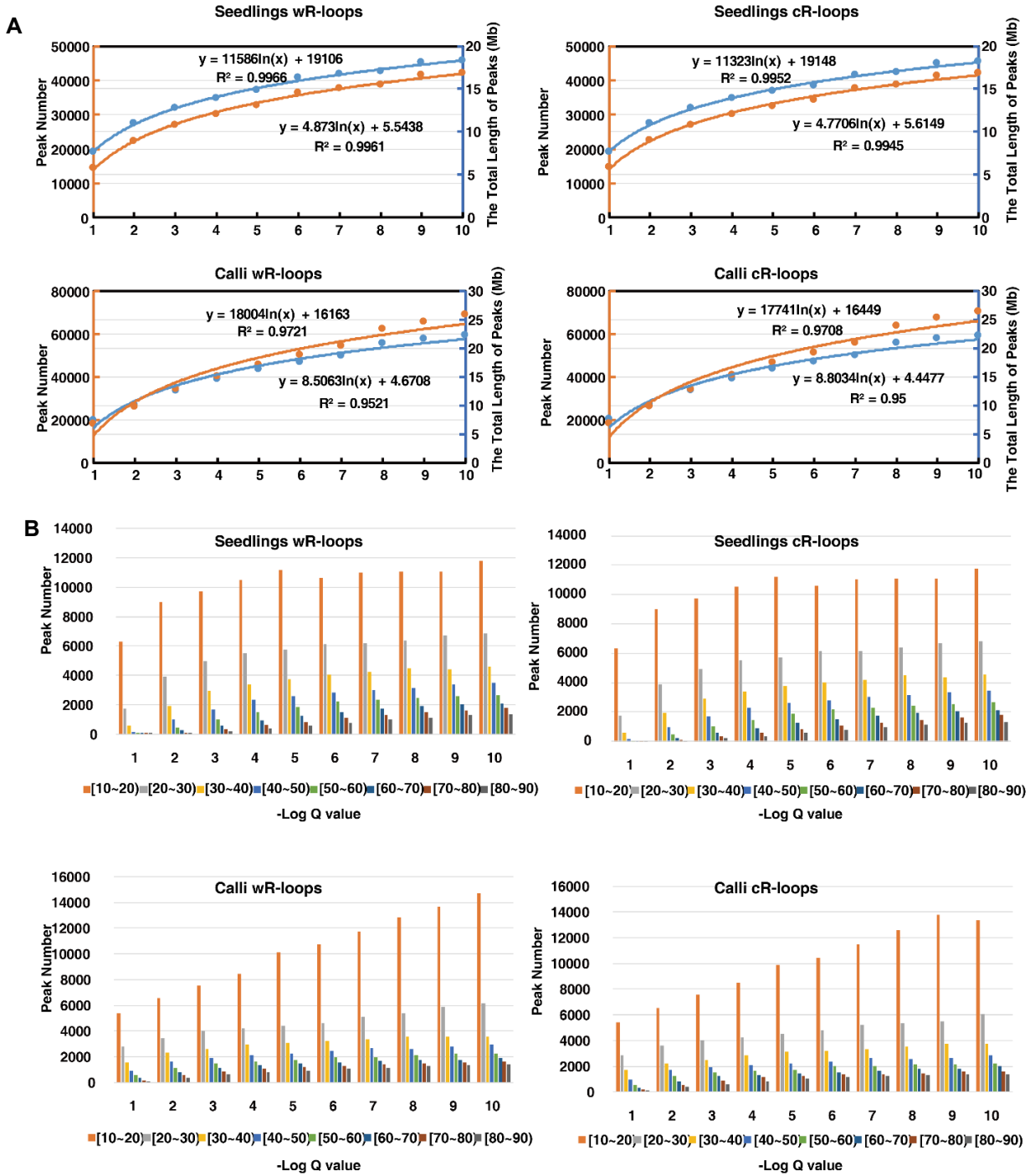

A
Seedlings wR-loops
y = 11586ln(x) + 19106
R² = 0.9966
y = 4.873ln(x) + 5.5438
R² = 0.9961
Peak Number
The Total Length of Peaks (Mb)
Seedlings cR-loops
y = 11323ln(x) + 19148
R² = 0.9952
y = 4.7706ln(x) + 5.6149
R² = 0.9945
Calli wR-loops
y = 18004ln(x) + 16163
R² = 0.9721
y = 8.5063ln(x) + 4.6708
R² = 0.9521
Calli cR-loops
y = 17741ln(x) + 16449
R² = 0.9708
y = 8.8034ln(x) + 4.4477
R² = 0.95
B
Seedlings wR-loops
Seedlings cR-loops
Calli wR-loops
Calli cR-loops
Peak Number
[10~20) [20~30) [30~40) [40~50) [50~60) [60~70) [70~80) [80~90)
-Log Q value

Fig. S12

Figure S12. The simulated relationships between peak numbers of two types of R-loops (wR- and cR-loops) and sequence depths. (A) Peak number and total peak length of w-R-loops or cR-loops in each tissue exhibited a logarithmic growth as the sequence depth increases. The total sequencing data was first divided into 10 equal parts, a bin i^{th} was generated by merging the first i^{th} parts into the i^{th} bin, where $i = 1, 2, 3 \ldots, 10$. There are two y-axes: one represents the peak number (left) and the other represents the total peak length (right), while the x-axis represents the bins of 1, 2, …,10 with increasing sequence depths. (B) Peak numbers for the different ranges of Q values. A higher negative logarithmic Q value represents a more significant peak level. The y-axis represents the peak number, while the x-axis represents the bins of 1, 2, …,10 with increasing sequence depths. The logarithmic Q value range was indicated using different colors.

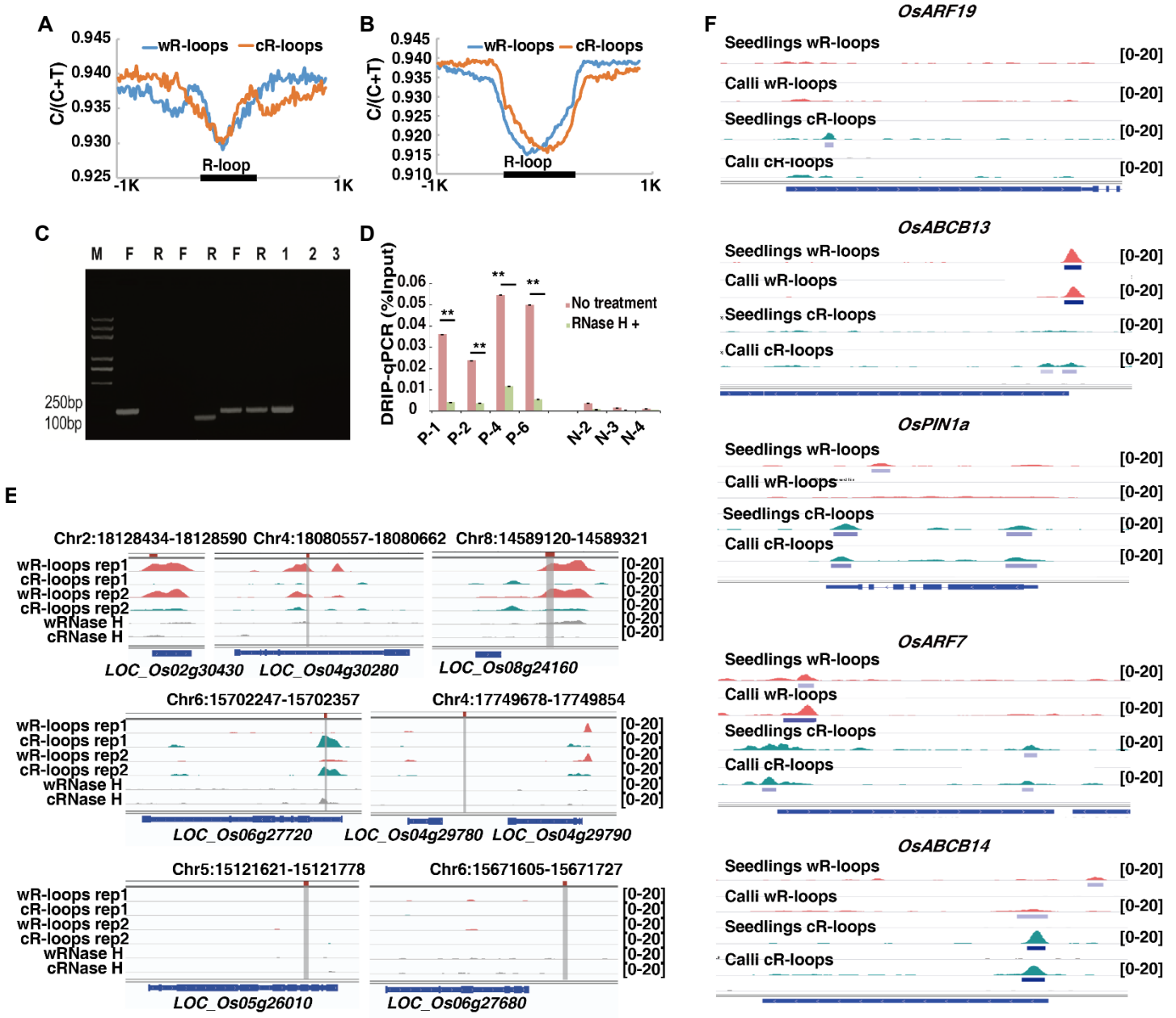

A
wR-loops
cR-loops
C/(C+T)
R-loop
-1K
1K
B
wR-loops
cR-loops
C/(C+T)
R-loop
-1K
1K
C
M F R F R F R 1 2 3
250bp
100bp
D
DRIP-qPCR (%Input)
No treatment
RNase H +
P-1 P-2 P-4 P-6 N-2 N-3 N-4
E
Chr2:18128434-18128590 Chr4:18080557-18080662 Chr8:14589120-14589321
wR-loops rep1
cR-loops rep1
wR-loops rep2
cR-loops rep2
wRNase H
cRNase H
[0-20]
LOC_Os02g30430 LOC_Os04g30280 LOC_Os08g24160
Chr6:15702247-15702357 Chr4:17749678-17749854
LOC_Os06g27720 LOC_Os04g29780 LOC_Os04g29790
Chr5:15121621-15121778 Chr6:15671605-15671727
LOC_Os05g26010 LOC_Os06g27680
F
OsARF19
Seedlings wR-loops
Calli wR-loops
Seedlings cR-loops
Calli cR-loops
[0-20]
OsABCB13
OsPIN1a
OsARF7
OsABCB14

Fig. S13

Figure S13. Validation of the R-loops identified by the DRIP-seq. NBS-seq and RT-PCR assays were used to validate the R-loops identified by DRIP-seq. (A) and (B): The ratios of unconverted cytosine in R-loop regions in seedlings (A) and calli (B). The higher conversion rates of cytosine to thymine, which corresponds to a lower relative ratio of C/(C+T), occurred within R-loops when compared to the flanking regions for both tissues, indicating the existence of single-stranded DNAs within in R-loop regions. (C) Strand-specific RT-PCR (ssRT-PCR) assay using the DRIPed DNA from seedlings. 104 R-loop loci were randomly selected for this assay. Final PCR products are shown. Bands in lanes marked “F” indicate that transcription occurred at the Watson DNA strand; bands in lanes marked “R” indicate that transcription occurred at the Crick DNA strand. The presence of DNA bands in two adjacent F and R lanes indicate that bidirectional transcription occurred at the R-loop region. Lane 1, total RNA without DNase I treatment; Lane 2, total RNA with DNase I treatment; Lane 3, negative control with H_2O as the PCR template. Lane M, DNA marker. (D) Four positive controls (P-1, P-2, P-4, and P-6) with R-loops, and three negative controls (N-2, N-3, and N-4) lacking R-loops were selected for the DRIP-qPCR assay. Primer sequences are listed in Supplemental Table S9. (E) The snapshots of Integrative Genomics Viewer (IGV) spanning seven regions for DRIP-qPCR assay in the rice genome. (F) Five previously validated auxin response genes (OsARF19, OsARF7, OsABCB13, 788 OsABCB14, OsPIN1a) with R-loop in *Oryza sativa* were re-examined in our study.

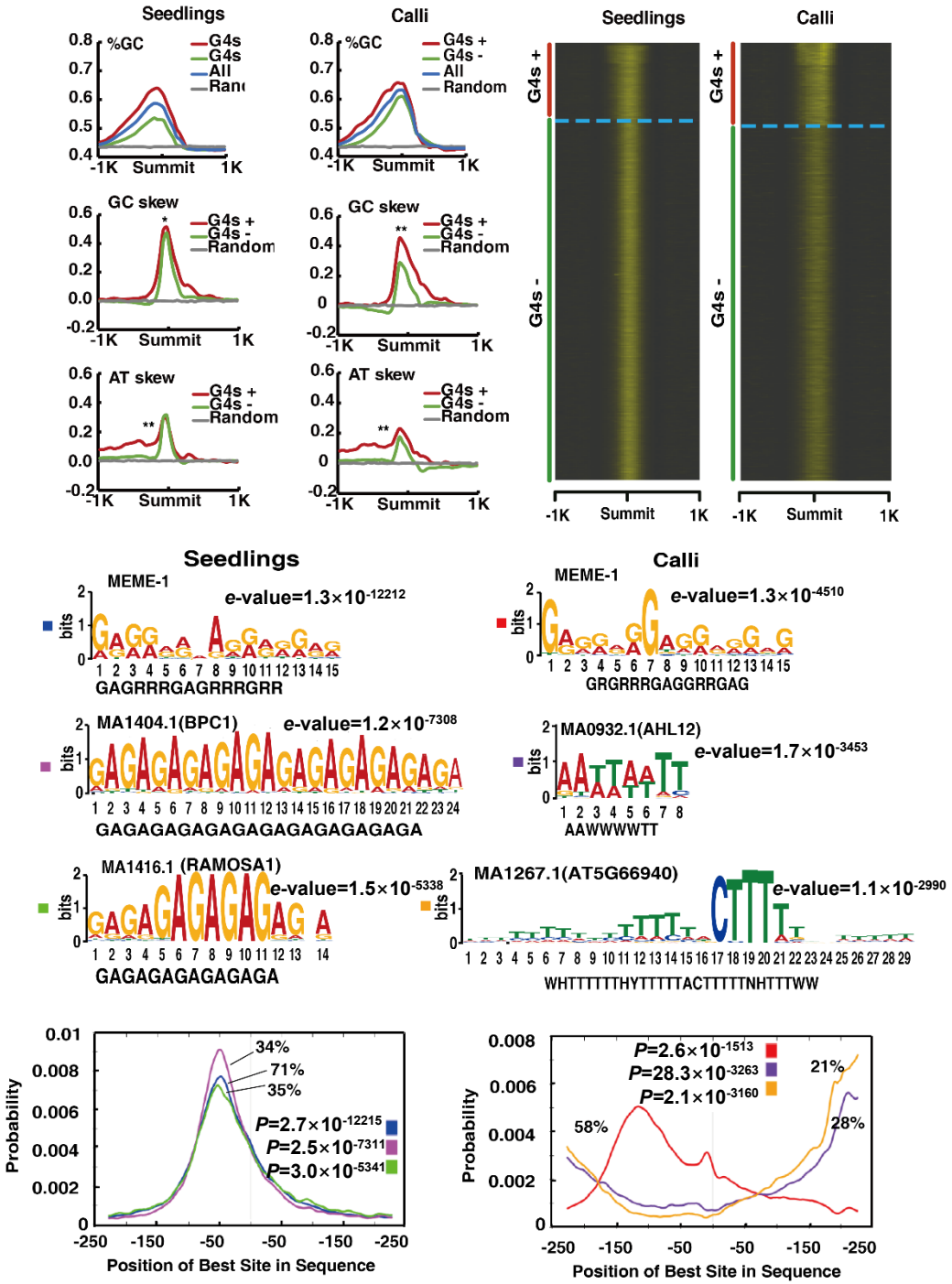

Seedlings
Calli
%GC
G4s
G4s +
G4s -
All
Random
GC skew
AT skew
Summit
-1K
1K
MEME-1
e-value=1.3×10-12212
GAGRRRGAGRRRGRR
e-value=1.3×10-4510
GRGRRRGAGGRRGAG
MA1404.1(BPC1)
e-value=1.2×10-7308
GAGAGAGAGAGAGAGAGAGAGAGA
MA0932.1(AHL12)
e-value=1.7×10-3453
AAWWWWTT
MA1416.1 (RAMOSA1)
e-value=1.5×10-5338
GAGAGAGAGAGAGA
MA1267.1(AT5G66940)
e-value=1.1×10-2990
WHTTTTTHYTTTTACTTTTTNHTTTWW
Probability
34%
71%
35%
P=2.7×10-12215
P=2.5×10-7311
P=3.0×10-5341
Position of Best Site in Sequence
P=2.6×10-1513
P=28.3×10-3263
P=2.1×10-3160
58%
21%
28%

Fig. S14

Figure S14. DNA sequence features associated with R-loops. (A) GC contents distributed around the summits of all R-loop peaks (All, blue line), R-loops with G-quadruplex sequences (G4s+: red line), R-loops without G4s (G4s-: green line), and the controls of 10,000 randomly selected regions (random: grey line). Each GC content was calculated for a sliding window of 20 bp to ± 1 kb from the summit. (B) Distribution of GC and AT skews in R-loops, which were calculated with the same window size of 20 bp. The GC-skews and AT-skews were calculated using the following formulae: (G%-C%)/(G%+C%) for GC-skew and (A%-T%)/(A%+T%) for AT-skew, while *p*-values were determined by Mann-Whitney U test. (C) Heatmap showing the read counts of R-loops distributed around ± 1 kb of the summit of the R-loop peaks. The different colors represent signal intensities of R-loops ranging from low (black) to high (yellow). The red or green sidebars represent G4s+ or G4s- R-loops, respectively, which are separated by the blue dashed line. (D) The top three significantly enriched DNA motifs in seedlings and calli. The e-value was listed in upside of each motif logo. (E) Distinct profiles of representative motifs in seedlings and calli. The percentages given are R-loop peaks containing a particular DNA motif sequences as a percent of all R-loop peaks.

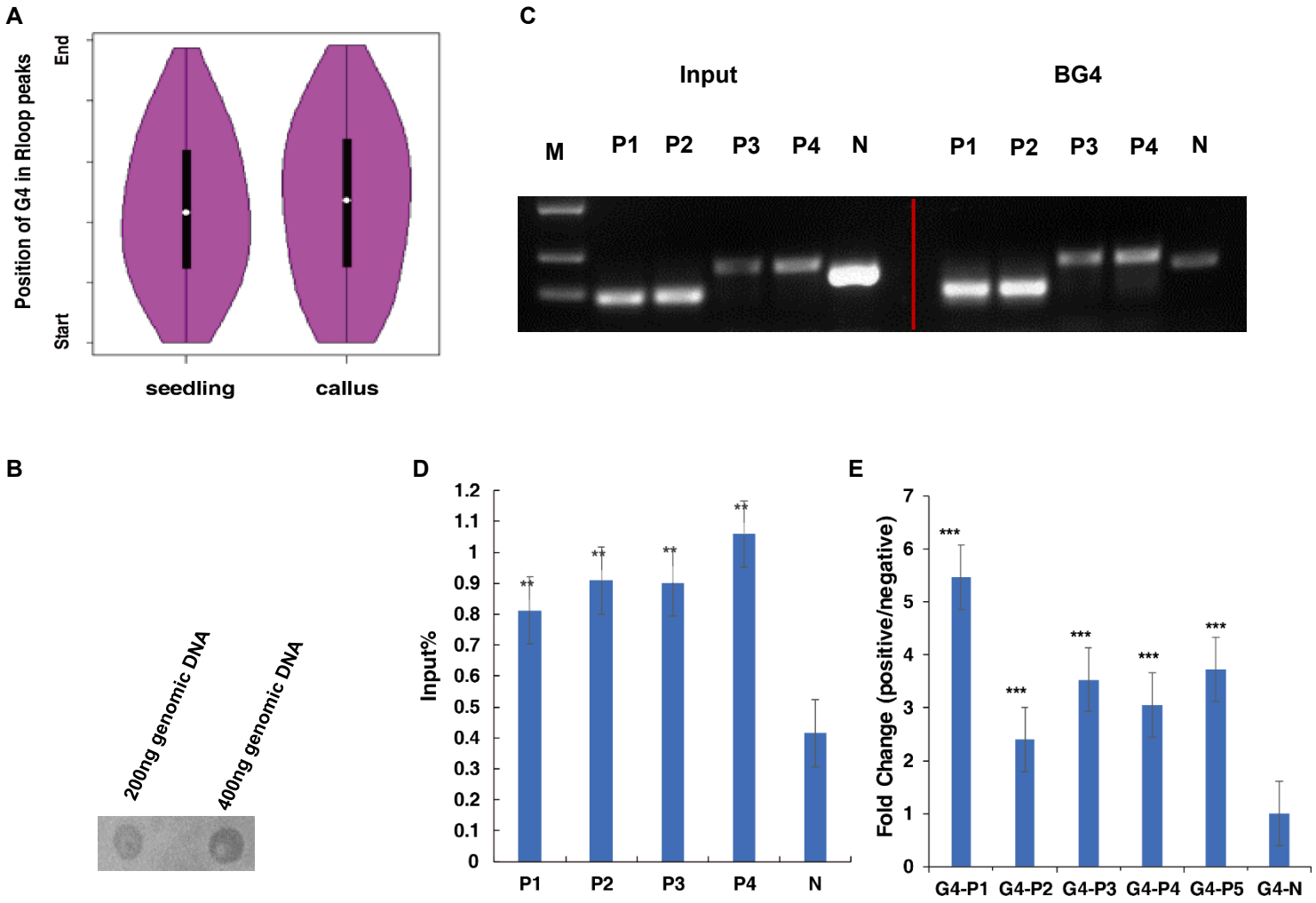

A
Position of G4 in Rloop peaks
End
Start
seedling
callus
C
Input
BG4
M
P1
P2
P3
P4
N
P1
P2
P3
P4
N
B
200ng genomic DNA
400ng genomic DNA
D
Input%
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
P1
P2
P3
P4
N
E
Fold Change (positive/negative)
7
6
5
4
3
2
1
0
G4-P1
G4-P2
G4-P3
G4-P4
G4-P5
G4-N

Fig. S15

Figure S15. Characterization of G-Quadruplex sequences in R-loops. (A) Violin plot showing that the G-quadruplex sequences (G4s) are primarily present in the approximate centers of R-loop peaks in both tissues. (B) Dot blotting assay. Genomic DNA (200 or 400ng) were loaded onto an Amersham Hybond-N+-nylon membrane followed by probing with recombinant BG4 fused with FLAG (1:2,500 dilution), and then detected with anti-FLAG antibody. The dot immuno-signal was detectable in 200 ng and 400 ng DNA. (C) G4-ChIP-PCR assay. Four positive R-loop loci with predicted G4s (P1-P4) and one negative control without G4s (N) were randomly selected for G4-ChIP-PCR assay. The PCR final products were detected by running 1.5% agarose gel. The left and right panels represent DNA products were produced using input and G4-ChIPed DNA, respectively. (D) Statistical analysis of enrichment at each locus relative to % of input. (E) G4-ChIP-qPCR assay. Five positive R-loop loci with predicted G4s (G4s-P1-P5) and one negative control without G4 (G4-N) were randomly selected for G4-ChIP-qPCR assay. Fold change was used to indicate the enrichment level of the positive locus relative to the negative locus. Primer sequences for G4-ChIP-PCR or -qPCR are listed in Supplemental Table S11. Significance in (D) and (E) was determined using variance analysis. *** and ** stand for extremely significance and highly significance, respectively.

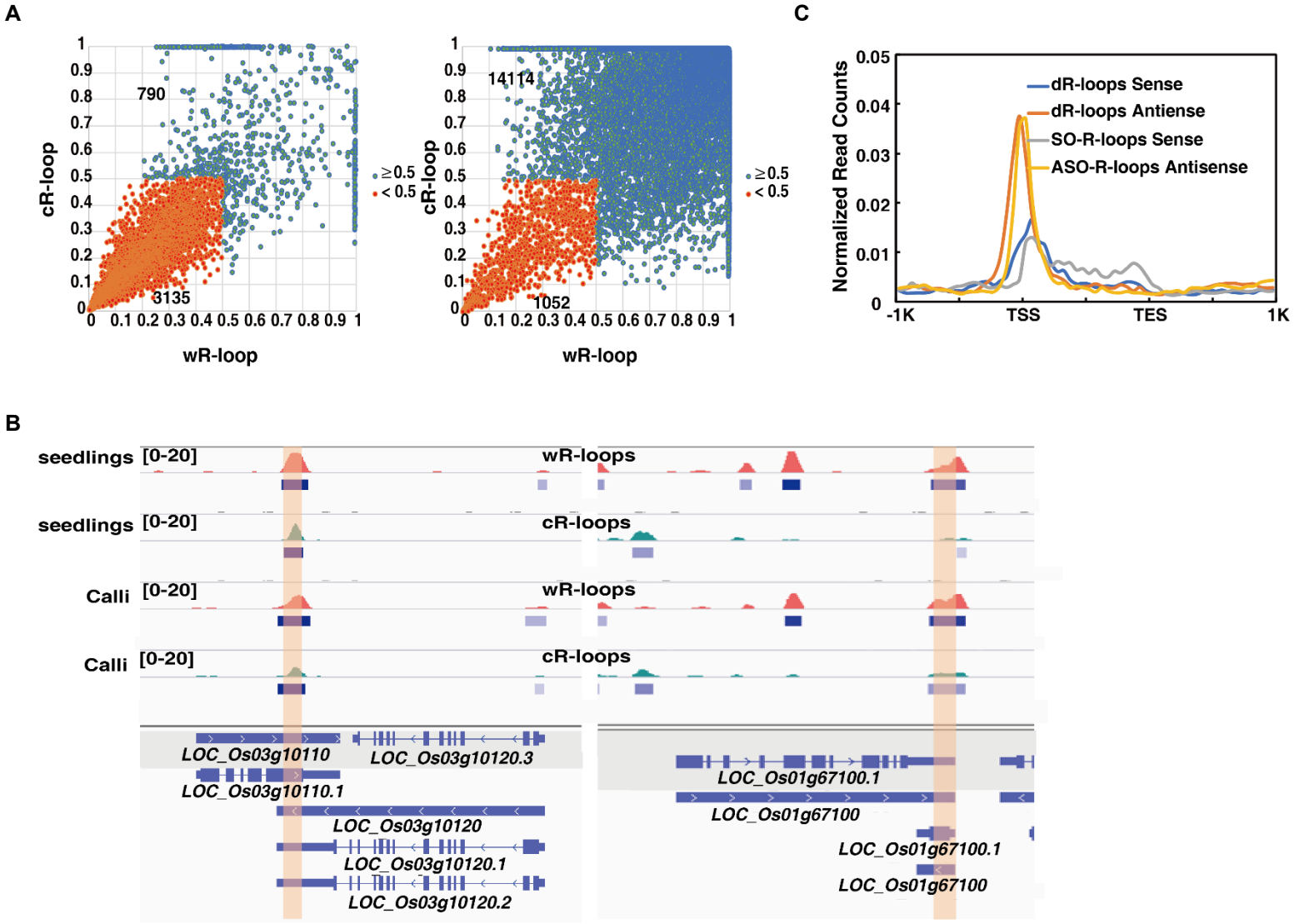


Figure S16. (A) Number of dR-loops with overlapping regions between cR-loops and wR-loops greater (green) or less (red) than 50% in seedlings (left) and calli (right). (B) Integrative Genomics Viewer (IGV) snapshots illustrating the existence of dR-loops in each tissue. (C) Comparisons of normalized read counts of sense or antisense R-loops associated with dR-loops with those of SO-R-loops and ASO-R-loops, respectively. The y-axis represents the normalized R-loop read counts, while the x-axis represents $\pm$ 1 kb of R-loop overlapping genes.

Supplemental Table S1. Classification of non-TE/TE genes associated with different types of R-loops in the rice genome.

Type of R-loops	Seedlings (non-TE/TE)		Calli (non-TE/TE)	
	the number of genes	percentage (%)	the number of genes	percentage (%)
sense only（SO-Rloop)	6,006/3,120	15.46/18.42	5,132/2,321	13.21/13.70
antisense only (ASO-Rloop)	10,278/1,539	26.45/9.08	7,720/1,485	19.87/8.77
sense and antisense overlapped (S/AS-Rloop)	5,125/772	13.19/4.56	9,177/1,773	23.62/10.47
R-loops -	17,451/11,510	44.91/67.94	16,831/11,362	43.31/67.07
Total	38,860/16,941	100/100	38,860/16,941	100/100

Note: R-loops- representing non-TE genes without R-loops.

Supplemental Table S2. Summary of transcripts containing spliced 5'UTRs.

Exon numbers of 5'UTRs	Numbers of transcripts	Percentage (%)
1	24,586	74.83
2	6,020	18.32
3	1,445	4.40
4	639	1.95
5	56	0.17
> 5	108	0.33

Note: The numbers with red were used for further analysis in the study.

Supplemental Table S4. Summary of primer information used for preparation of rice transgenic vectors.

ID	F primer sequences	R primer sequences
H3.3	TTGGAGAGGACAGCCGTCGACATGGCCCGTACGAAGCAGAC	GCCCTTGCTCACCATCCCGGGCGCGCGCTCGCCGCGGA
H3.3-K9M	GAGGACAGCCGTCGACATGGCCCGTACGAAGCAGACCGCCCGCATGTCCA	GCCCTTGCTCACCATCCCGGGCGCGCGCTCGCCGCGGA
H3.3-K4 and K9M	GAGGACAGCCGTCGACATGGCCCGTACGATGCAGACCGCCCGCATGTCCA	CCAGCTCGCCCGCCGCATCCGCGGCGAGCGCGCGTAACCCGGG
LOC_Os02g45110	GP2938-2011-F:CACAAACAAGGCCGTATCGC	GP2938-2011-R:GAGCTCCAGGAGGCACAC
LOC_Os1031030	GP2940-2013-F:TGTCCAGTTGTGCTGCGTAT	GP2940-2013-R:CCGATCAGAGGACCTGCTTT

Supplemental Table S5. Summary of sequencing data used in the study.

Data	Accession number	Database	Reference	Tissue
H2A.Z	PRJNA326065	NCBI	(Zhang et al. 2017)	Seedlings
H3K4me3	GSM489075	NCBI	(He et al. 2010)	Seedlings
H3K4me2	GSM658110	NCBI	(Zhang et al. 2012)	Seedlings
H3K36me3	GSM658111	NCBI	(Zhang et al. 2012)	Seedlings
H4K12ac	GSM658112	NCBI	(Zhang et al. 2012)	Seedlings
DNase-seq	GSM655033	NCBI	(Zhang et al. 2012)	Seedlings
RNA-seq	GSE33265	NCBI	(Wu et al. 2011)	Seedlings/Calli
MNase-seq	SRP045236	NCBI	(Zhang et al. 2015)	Seedlings
H3K9me1	GSE79033	NCBI	(Fang et al. 2016)	Seedlings
H3K4ac	GSE79033	NCBI	(Fang et al. 2016)	Seedlings
H3K27me3	GSE79033	NCBI	(Fang et al., 2016)	Seedlings
H3K27ac	GSE79033	NCBI	(Fang et al. 2016)	Seedlings
H3K9ac	GSE79033	NCBI	(Fang et al. 2016)	Seedlings
H3K23ac	GSE69426	NCBI	(Lu et al. 2015)	Seedlings
H4K16ac	GSE69426	NCBI	(Lu et al. 2015)	Seedlings
H3K9me3	GSE79033	NCBI	(Fang et al. 2016)	Seedling
H3K9me2	GSM2152477	NCBI	(Tan et al. 2016)	Seedlings
DRIP-seq	GSE111944	NCBI	New	Seedlings/Calli

Supplemental Table S6. Summary of sequencing data sets.

Samples	Biological replicates	Clean read	Mappable read	Mappable rate	Unique mappable read	Unique mappable ratio
DRIP-seq_seedlings_rep1	1	11,533,822	9,929,403	86.19%	5,764,252	49.98%
DRIP-seq_seedlings_rep2	2	14,816,281	11,882,548	80.20%	8,594,178	58.00%
DRIP-seq_RNaseH+_seedling_rep1	1	6,200,836	3,847,428	62.05%	2,905,571	46.86%
DRIP-seq_RNaseH+_seedlings_rep2	2	12,024,817	7,259,539	61.87%	4,008,149	33.33%
DRIP-seq_calli_rep1	1	36,049,161	27,021,995	80.52%	22,217,046	61.63%
DRIP-seq_calli_rep2	2	27,113,050	23,143,427	78.30%	15,386,009	56.75%
DRIP-seq_RNaseH+_calli	1	16,561,703	11,226,516	69.06%	7,575,653	45.74%
NBS-seq_seedlings	1	30,026,944	28,214,828	93.97%	20,254,170	67.45%
NBS-seq_calli	1	80,891,014	76,039,843	94.00%	55,133,140	68.16%

Supplemental Table S8. Genomic information for seven genomic loci used for DRIP-qPCR assay.

ID	Region	F primer sequences	R primer sequences
P-1	Chr2:18128434-18128590	AGAAGGTCTCCGGAATCACC	CCTCCGTCGGAATCCACC
P-2	Chr4:18080557-18080662	GGGAAGAAGGAACATTTTGTGC	ACTTTGCTTATTTGGCCCTTCT
P-4	Chr8:14589120-14589321	GATGAATCCGGCTTGTAGACCT	CTCCTCCTGCCTCATCACATAG
P-6	Chr6:15702247-15702357	TCGAATCCTACCTCTCTCTCCT	AGGGAACGTGGGTGCTTTATAT
N-2	Chr4:17749678-17749854	AAGAGGCCATGAGAAGCCAT	AACTTTTGCCACAAGTCGCA
N-3	Chr5:15121621-15121778	CACTGAGGAGTCGTACCAGG	GGGTACTCACAACACAAGCG
N-4	Chr6:15671605-15671727	CGAAGAGAAGCAACAGAGAAGG	TTGGAATGGTCAAAATTGCCCT

Supplemental Table S9. Summary of sequence features associated with all R-loops or R-loops in promoters and introns.

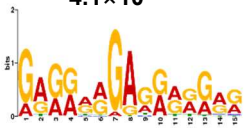
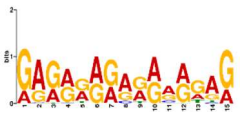
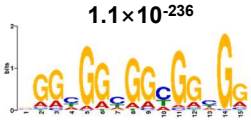
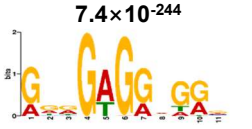
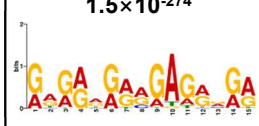
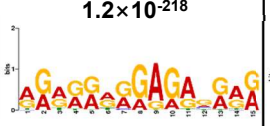
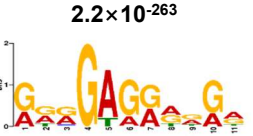
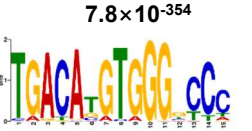
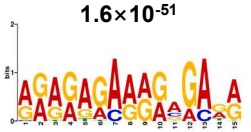
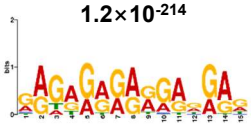
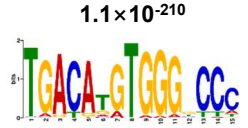
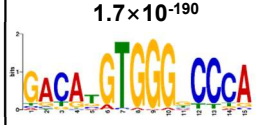
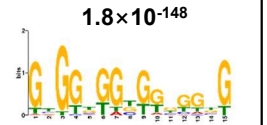
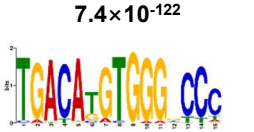
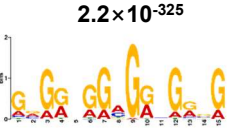
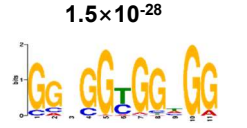
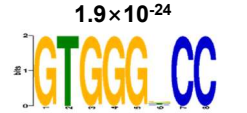
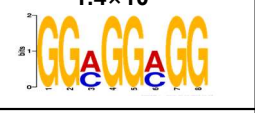
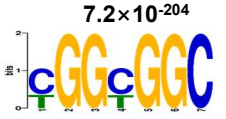
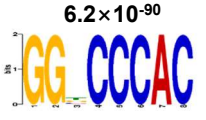
	R-loops	peak number	%GC	SD (±)	GC-skew (%)	SD (±) 2	%AT	SD (±) 3	AT-skew (%)	SD (±) 4
All R-loops	All	76,235 (100%)	53.21	0.1060	27.05	0.1431	46.79	0.1060	14.39	0.1631
	R-loop with DHSs	16,196 (21.24%)	56.72	0.0873	25.27	0.1375	43.28	0.0873	11.77	0.1596
	R-loop with CNSs	5,426 (7.12%)	54.27	0.0920	2586	0.1404	45.73	0.0920	10.36	0.1827
R-loops in Promoter	All	15,620 (100%)	53.25	0.0964	26.33	0.1495	46.75	0.0964	11.87	0.1664
	R-loop with DHSs	5,241 (33.55%)	55.77	0.0819	25.82	0.1471	44.23	0.0819	8.57	0.1685
	R-loop with CNSs	1,924 (12.33%)	54.80	0.0852	28.37	0.1487	45.20	0.0852	6.40	0.1780
R-loops in Intron	All	8,760 (100%)	50.70	0.1013	27.05	0.1432	49.30	0.1014	15.00	0.1721
	R-loop with DHSs	977(11.15%)	54.11	0.0804	22.94	0.1359	45.89	0.0804	16.03	0.1552
	R-loop with CNSs	336 (3.84%)	48.72	0.0867	24.73	0.1606	51.28	0.0868	12.36	0.2566

DHS: DNase I hypersensitive sites; CNS: conserved non-coding sequences (CNSs)

Supplemental Table S10. Genomic information for genomic loci used for G4-ChIP-PCR and -qPCR assays.

	ID	Region	F primer sequences	R primer sequences
G4-ChIP-PCR	P1	Chr7:00001072-00001859	GAAGACGACGGCGAACG	TCTAATCCCTCCCATCTTGTT
	P2	Chr5:11499044-11499882	TAGGGGAGAGGTTTCCGAAC	CCTCCCTCTCTCTCCTCTCC
	P3	Chr6:05762558-05763263	GAGGGAGGGAGGGGTATAGA	GGCTTCCTAGTGCGAGCTTA
	P4	Chr6:08593945-08594825	ACGGGGAGGGTTTAAATAGG	GAATCATTCCCCGTGATCC
	N	Chr4:18080557-18080662	GGGAAGAAGGAACATTTTGTGC	ACTTTGCTTATTTGGCCCTTCT
G4-ChIP-qPCR	G4-P1	Chr1:7408935- 7409496	ACCACCACCTCAGACAAACC	CCCCAATTCCACAGAGAAAA
	G4-P2	Chr2:5630922-5631327	CGTTTCGCCTGTAAATGCTT	GCTCACCTCCGAGTCCATAG
	G4-P3	Chr6 :2274746 2275056	GAGGAGGAAGAGAGGGAAGG	GCAGCAGGAATCACTCATCA
	G4-P4	Chr2:23780501-23781520	GGTGACGAGCACCAACCTAT	CAGTTGCAAGAAGACGGACA
	G4-P5	Chr8:3392736-3393108	TCCAGGTATGGAGAGCTTGG	GGGGAAACCCCAGTAGTGAT
	G4-N	Chr1:35835852-35836173	GGACGCATAGGAAGAACAGG	GCGATACCTTGGACGTTGTT

Supplemental Table S11. Summary of novel motifs identified in R-loops for seedlings.

Motif	All R-loop Peaks	R-loop Peaks in Promoter			R-loop Peaks in Intron		
		all	Overlapped with DHs	Overlapped with CNS	all	Overlapped with DHs	Overlapped with CNS
MEME-1	4.1×10^{-1265} 	5.4×10^{-1241} 	5.3×10^{-1346} 	4.6×10^{-1305} 	7.1×10^{-1219} 	3.7×10^{-1310} 	2.2×10^{-354} 
MEME-2	1.1×10^{-236} 	7.4×10^{-244} 	1.5×10^{-274} 	1.2×10^{-218} 	2.2×10^{-263} 	7.8×10^{-354} 	1.6×10^{-51} 
MEME-3	1.2×10^{-214} 	1.1×10^{-210} 	1.7×10^{-190} 	1.8×10^{-148} 	7.4×10^{-122} 	2.2×10^{-325} 	1.5×10^{-28} 
DREME-1	1.6×10^{-1511} 	2.9×10^{-274} 	4.1×10^{-121} 	2.3×10^{-31} 	1.0×10^{-106} 	1.9×10^{-24} 	3.4×10^{-3} 
DREME-2	2.2×10^{-1238} 	1.8×10^{-237} 	4.3×10^{-112} 	3.9×10^{-31} 	2.3×10^{-101} 	1.4×10^{-19} 	
DREME-3	2.3×10^{-827} 	7.2×10^{-204} 	7.8×10^{-73} 	2.2×10^{-21} 	6.2×10^{-90} 	1.8×10^{-14} 	

Additional Supplemental Table legends

Table S3. Summary of the KEGG enrichment analysis for genes associated with each cluster within each R-loop type. All genes associated with each cluster within each R-loop type were used for KEGG and GO enrichment analysis, and only a subset of genes with significant GOTerm enrichment were listed in the table.

Table S7. Summary of PCR primer sequence information used for ssRT-PCR assay. Genomic loci associated with R-loops were randomly chose for designing PCR primers for ssRT-PCR assay. The transcription direction of each R-loop examined was determined according to the presence or absence of PCR products. "+" represent the sense R-loop; "-" represent the antisense R-loop; "Null" represent the R-loop was not detected.