

Additional file 1:

Uncovering the multi-layer cis-regulatory landscape of rice via integrative nascent RNA analysis

Marina Goliassse¹, Aurore Johary¹, Adrian E. Platts², Fabian Ortner¹, Patrick P. Edger²,
Jae Young Choi³, Michael D. Purugganan^{4,5*}, and Zoé Joly-Lopez^{1*}

¹ Département de Chimie, Université du Québec à Montréal, Montréal, Québec, Canada

² Department of Horticulture, Michigan State University, East Lansing, MI, USA

³ Department of Ecology and Evolutionary Biology, The University of Kansas State, Lawrence, KS, USA

⁴ Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY, USA

⁵ Center for Genomics and Systems Biology, NYU Abu Dhabi Research Institute, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, United Arab Emirates

This file includes:

Figs S1 to S39

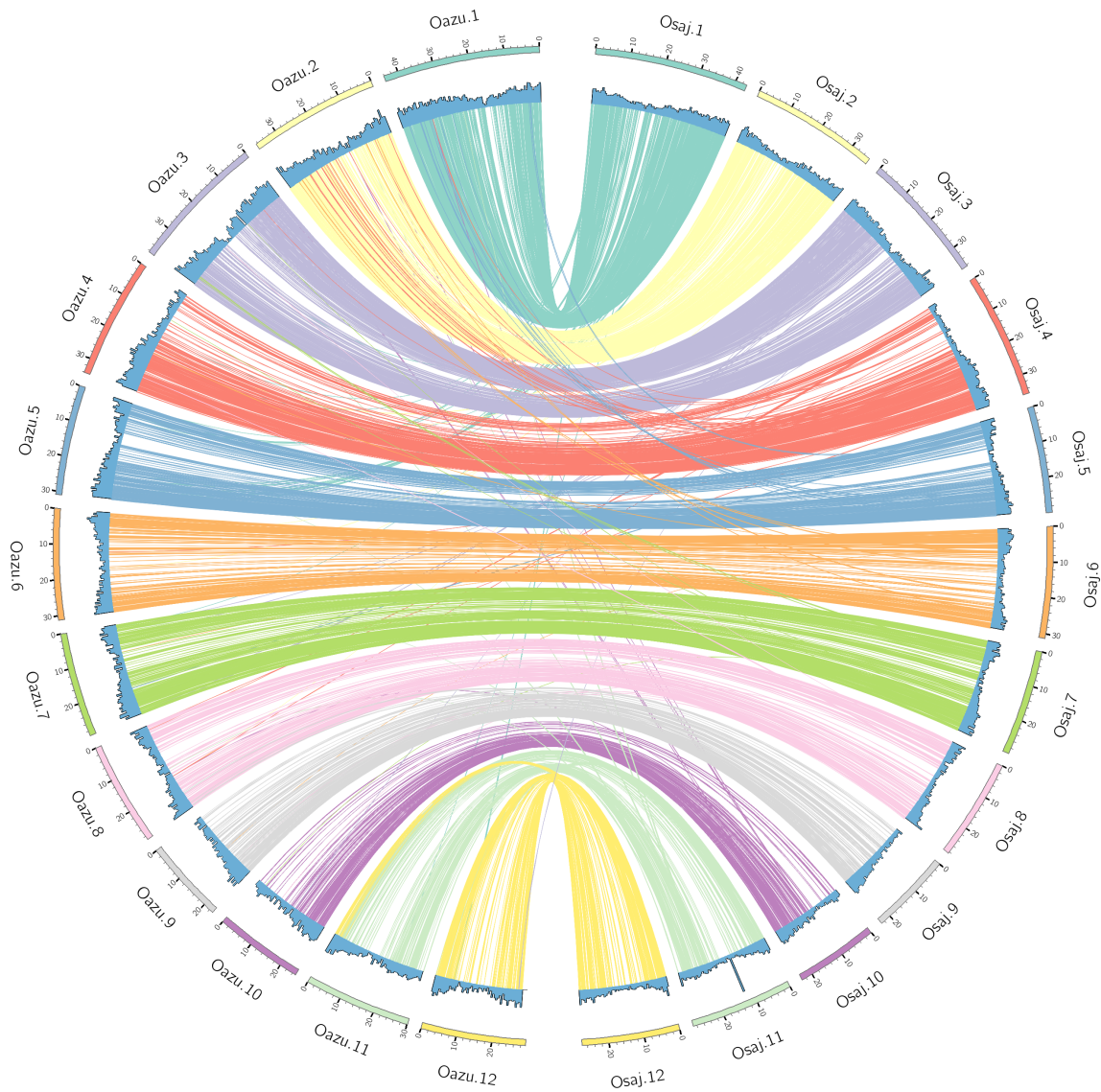


Fig S1. Circos plot for syntenic links and gene density. We note that the small regions that map between different chromosomes in Osaj12 and Oazu11 are likely a result of the grass tetraploidy event estimated about 70-90 million years ago.

Axis metrics are in nucleotides

y-axis organism: *Oryza sativa Japonica Group* (vOazu1)

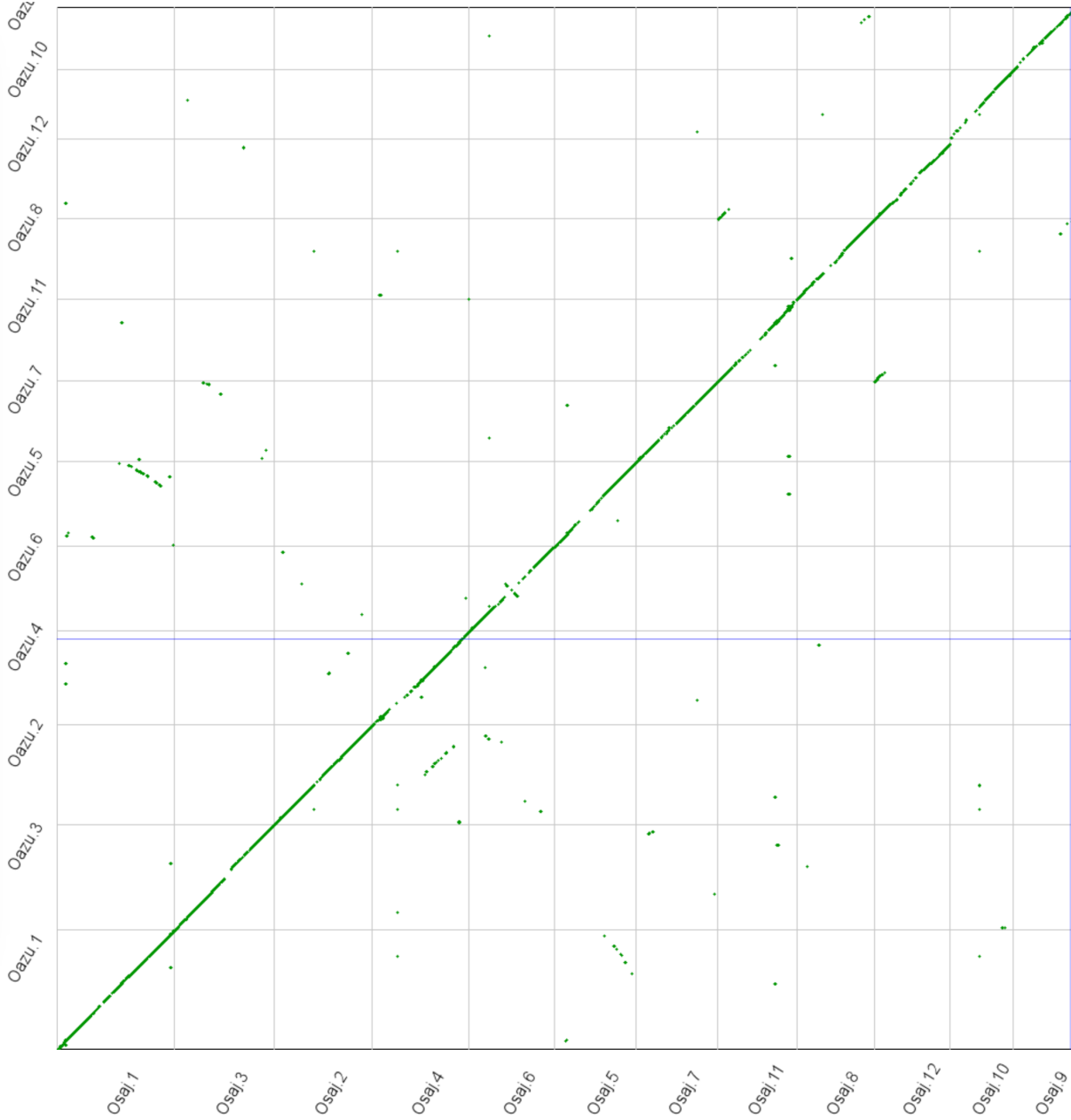


Fig S2. Synmap dotplot. Dotplot from which the synteny links were generated from in Figure S1. It shows again small region differences likely due to ancient polyploidy event.

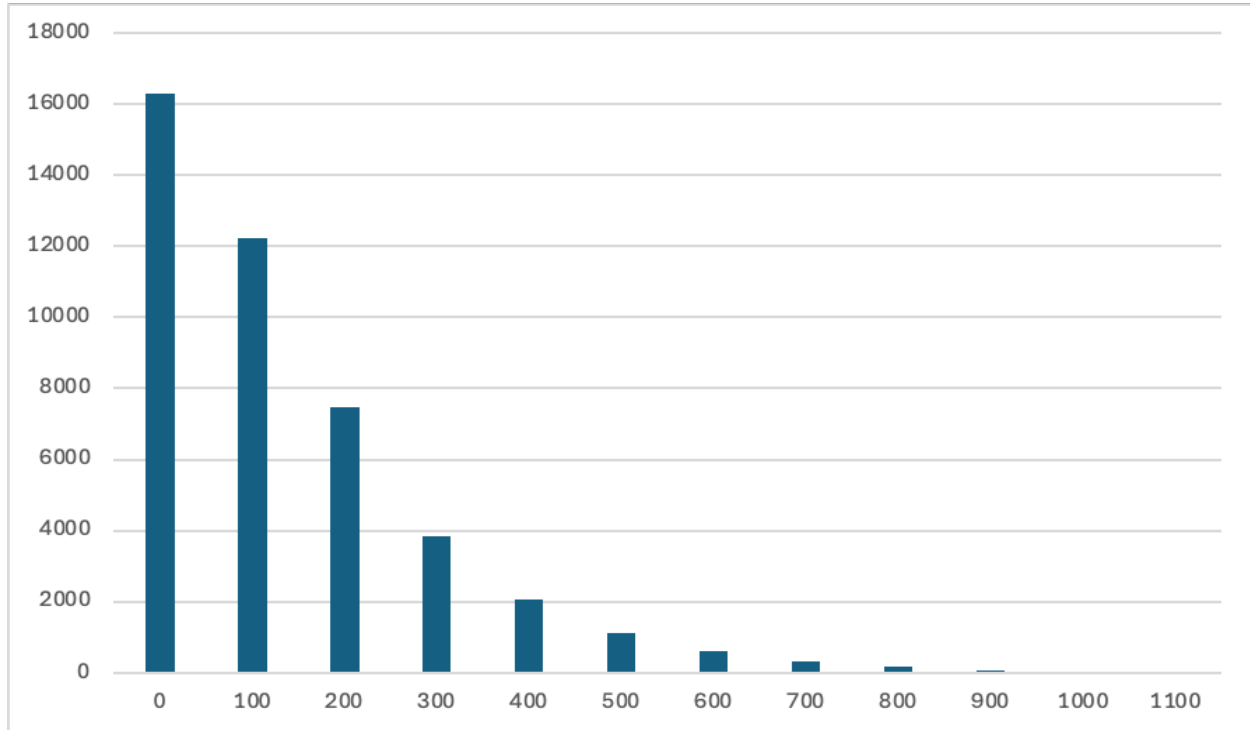
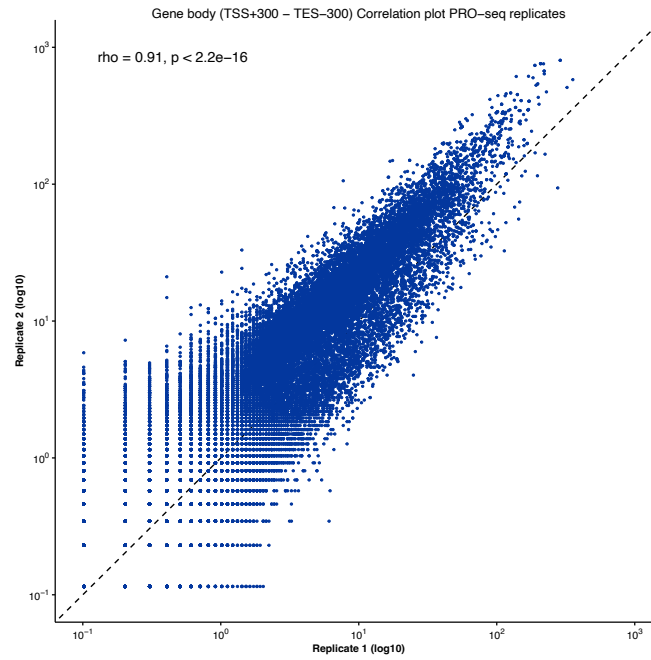


Fig S3. Distribution of CNS. We show CNS distribution content in genes within 2 kilo base pairs of gene coding start and gene coding end (including intronic CNS). The mean length is 106 nucleotides per gene, or about 3 CNSs per gene.

a



b

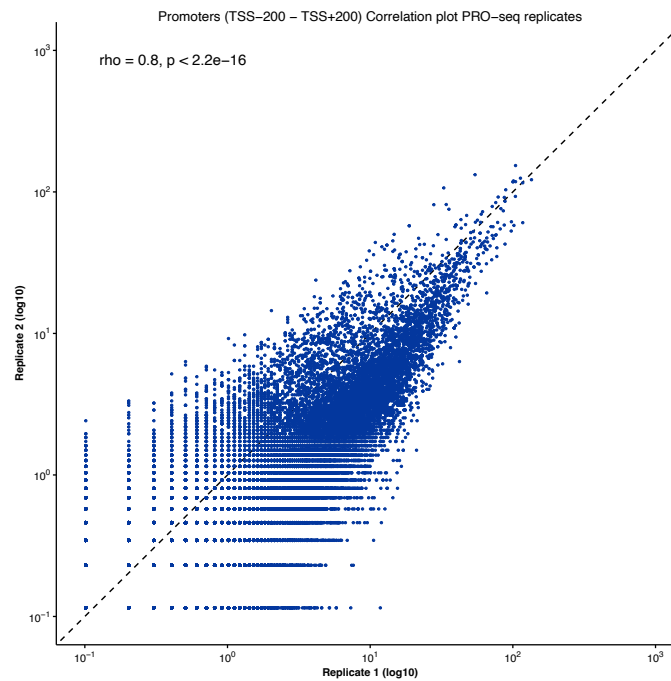


Fig S4. Correlation between PRO-seq replicates. Correlation of reads per million (RPM) in **(a)** gene body regions (TSS + 300 to TES-300) and **(b)** in promoter regions (TSS-200 to TSS+200) between PRO-seq replicates

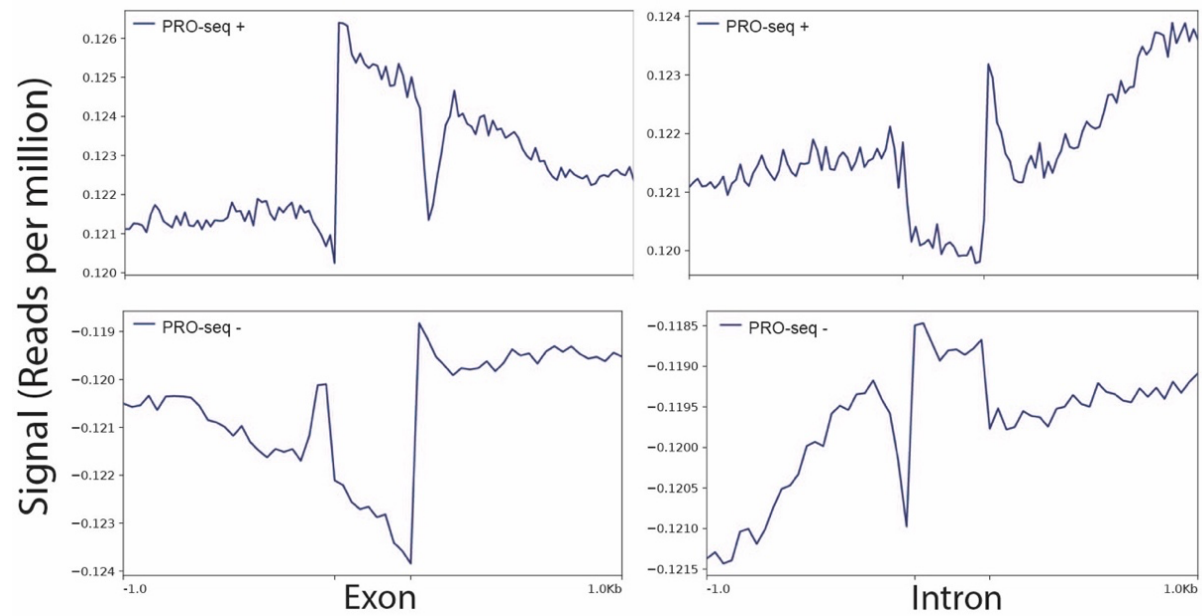
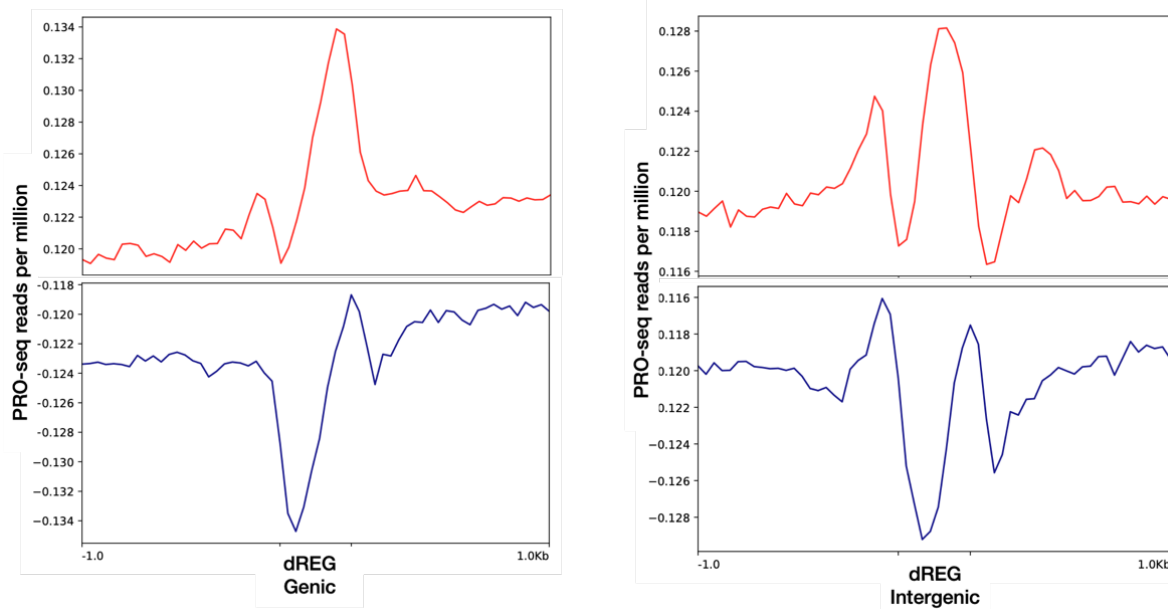
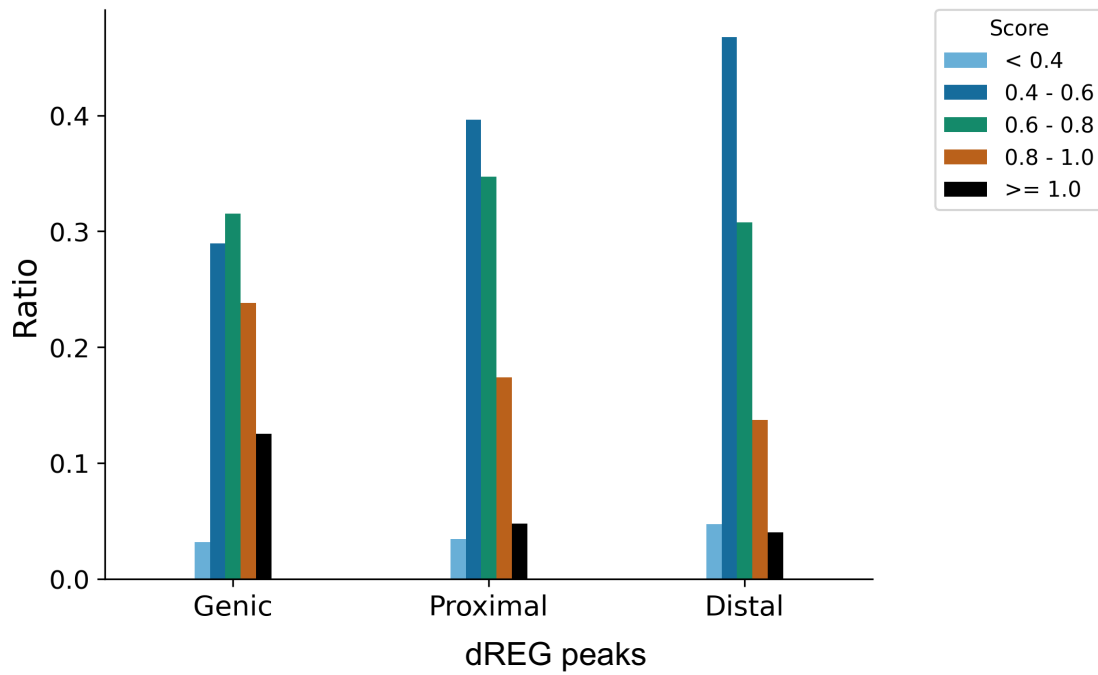
a**b**

Fig S5. PRO-seq signal within and around annotated regions. (a) PRO-seq signal (plus and minus) within exon and intron and +/- 1Kb around. **(b)** PRO-seq signal distribution within and +/- 1 kilo base pairs (kb) around dREG peaks annotated as genic (left) and intergenic (right).

a



b

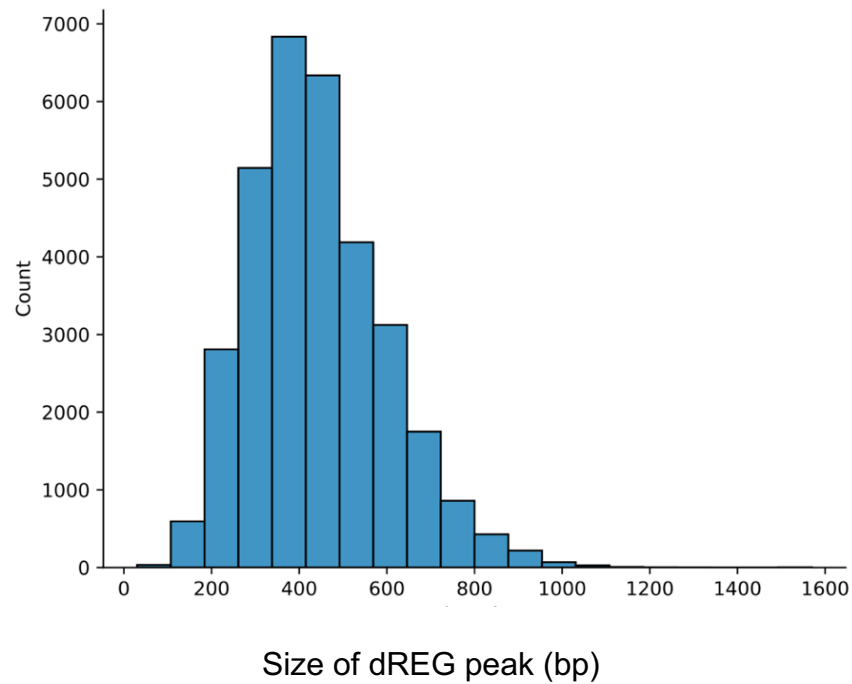


Fig S6. dREG characteristics. Distribution of genome-wide dREG scores **(a)** and dREG peak size **(b)**. Median dREG peak size is 420 base pairs (bp).

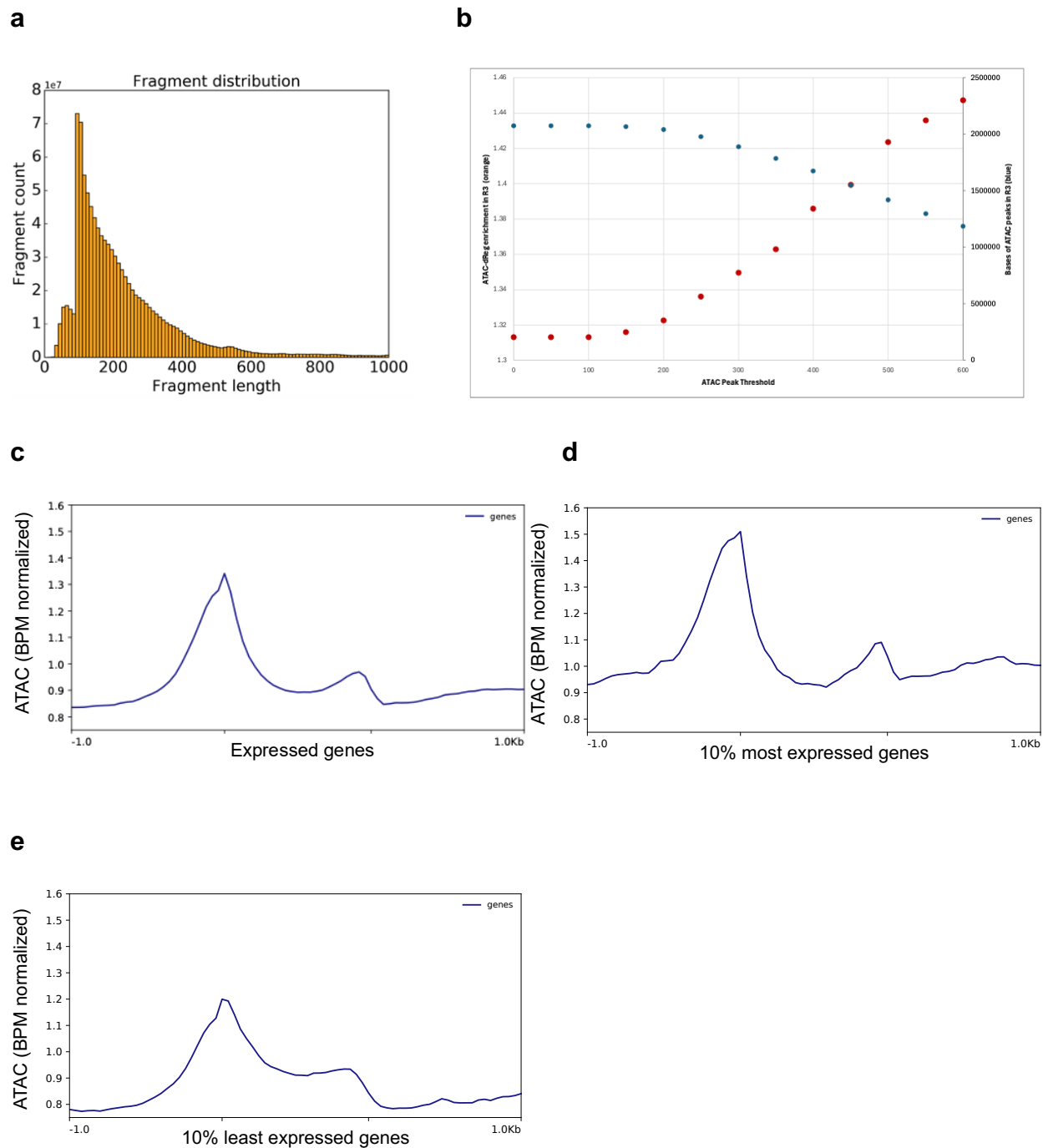


Fig S7. Quality control of ATAC peaks. (a) Fragment length distribution in ATAC peaks. (b) Selection of the optimal signal cutoff for ATAC peaks. (c) ATAC signal within expressed genes in rice lead tissues (based on RNA-seq data) as well as 1kb upstream and downstream of each expressed gene. In (d) the ATAC signal is plotted for the 10% most expressed genes and in (e) the 10% least expressed.

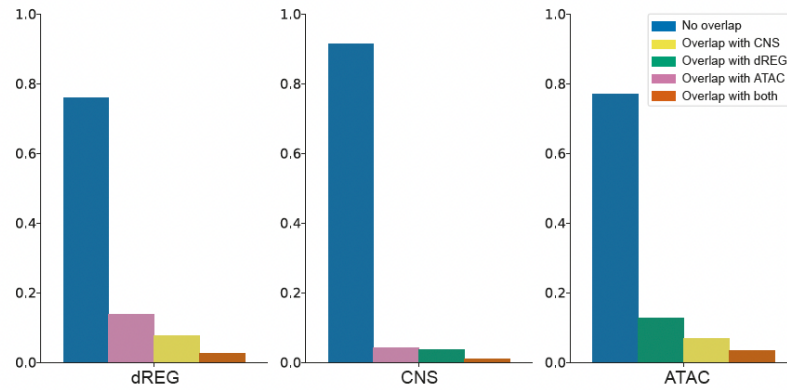
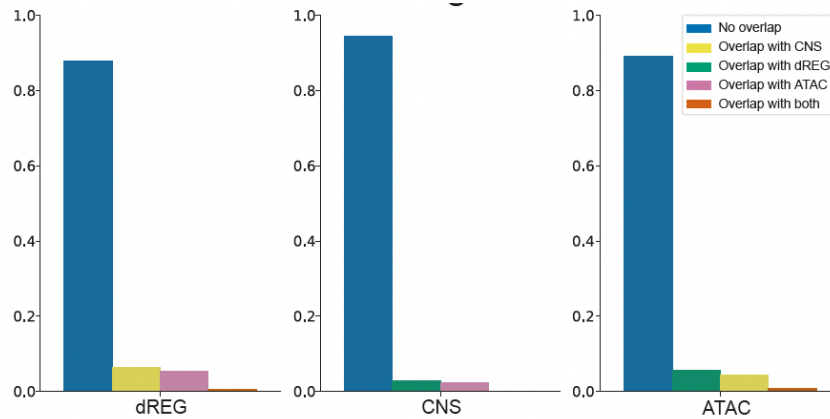
a**All regions****b****Intergenic**

Fig S8. Overlap analysis between dREG locations, ATAC peaks and CNSs. Overlap between the X-axis element and two other elements for all the genomic locations **(a)** and intergenic locations **(b)**. For each histogram, the overlap proportion is generated based on dREG locations (left), CNSs (middle) and ATAC peaks (right). For total, there were a small number of regions overlapping all three sequence features, corresponding to 832 CNSs (out of 96,570), 830 dREG (out of 32,424), and 921 ATAC (out of 27,772).

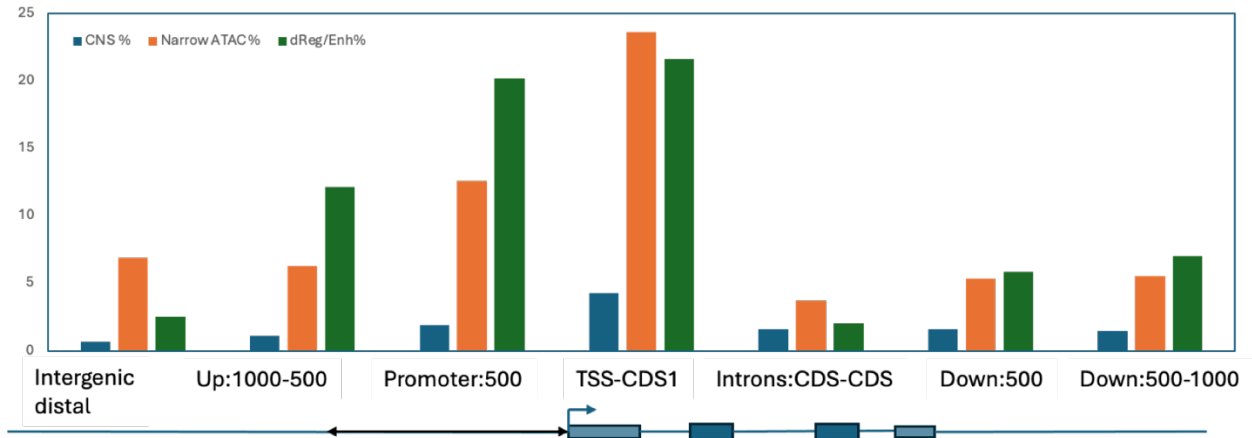


Fig S9. Percentage coverage of CNS, ATAC-seq and dREG peaks. The coverage of the different peaks within and around all annotated genes is shown in the rice genome. We counted only cases where the distance between the TSS and the first exon was less than 500 bp to exclude isoforms with potentially long 5' UTRs.

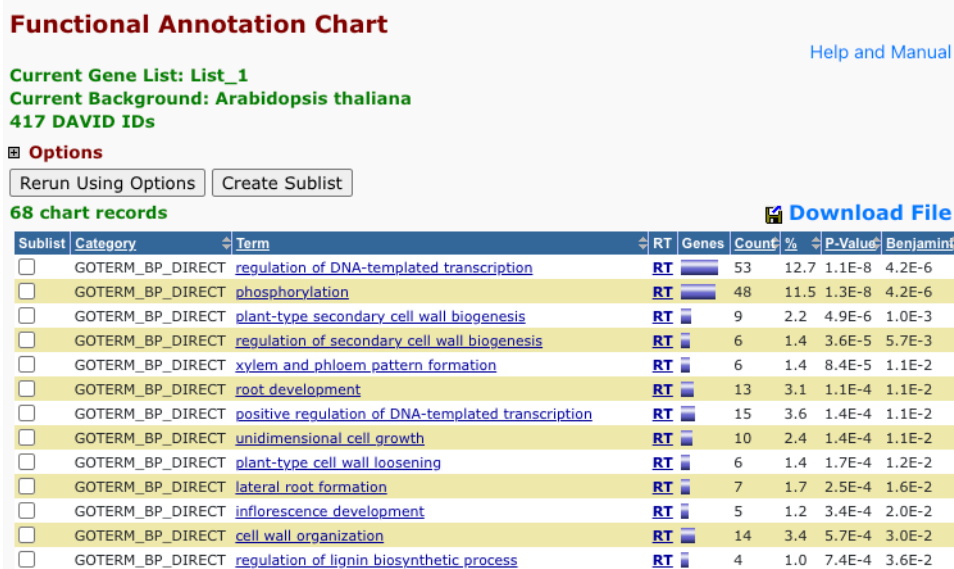
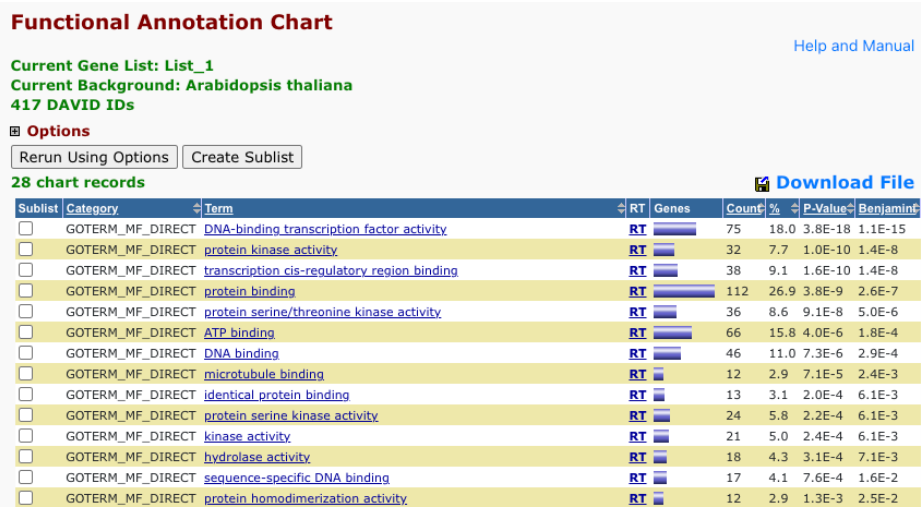


Fig S10. Gene ontology (GO) analysis for annotated rice genes with top 1,000 CNS content from translation start to 2 Kb upstream of the translation start upstream (coverage >233 nucleotides of CNS). Rice gene IDs were translated to *Arabidopsis thaliana* IDs to enable GO analysis with Database for Annotation, Visualization and Integrated Discovery (DAVID). Also see Additional file S4.

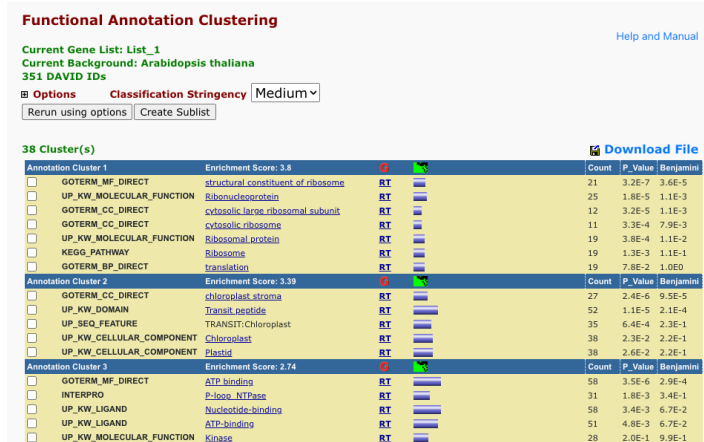
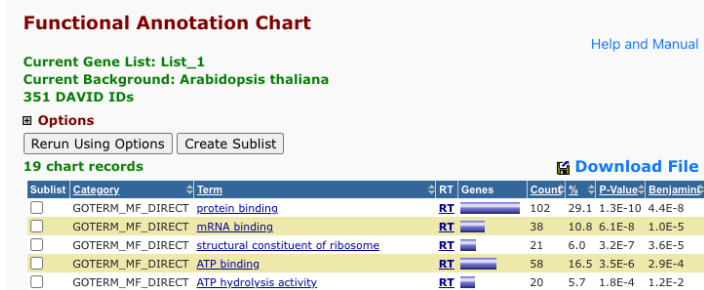
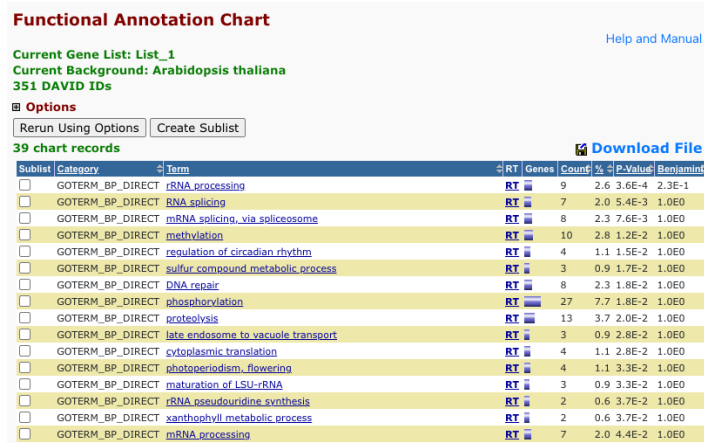


Fig S11. Gene ontology (GO) analysis for annotated rice genes with top 1,000 ATAC content from translation start to 2 Kb upstream of the translation start upstream (coverage >1274 nucleotides of ATAC). Rice gene IDs were translated to *Arabidopsis thaliana* IDs to enable GO analysis with Database for Annotation, Visualization and Integrated Discovery (DAVID). Also see Additional file S5.

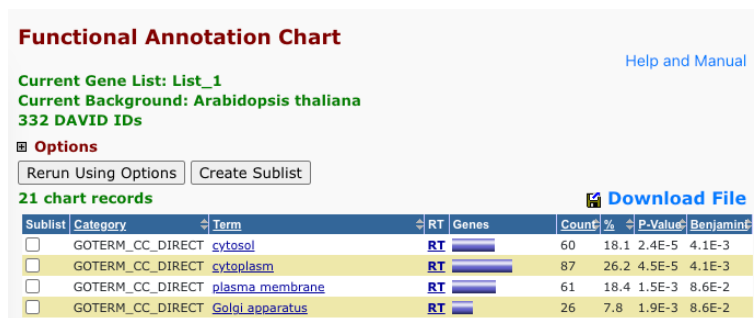
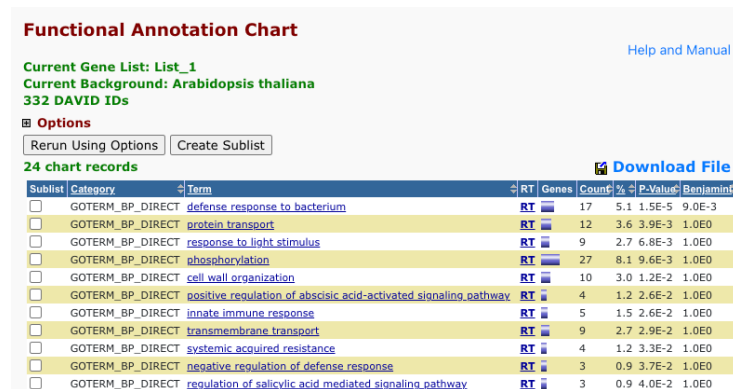


Fig S12. Gene ontology (GO) analysis for annotated rice genes with top 709 dREG content from translation start to 2 Kb upstream of the translation start upstream (coverage ≥ 1 nucleotides of dREG). Rice gene IDs were translated to *Arabidopsis thaliana* IDs to enable GO analysis with Database for Annotation, Visualization and Integrated Discovery (DAVID). Also see Additional file S6.

Enriched motifs in Proximal-promoter CNSs

Rank	Motif	P-value	log P-value	% of Targets	% of Background	STD(Bg STD)	Best Match/Details	Motif File
1		1e-157	-3.625e+02	35.10%	21.06%	18.5bp (18.5bp)	PHD1/MA0355.1/jaspar(0.930) More Information Similar Motifs Found	motif file (matrix)
2		1e-99	-2.303e+02	21.78%	12.51%	18.6bp (18.6bp)	PFP1/RRM1/Drosophila_melanogaster-PFP1-Seq(GSE35062) (Homer0.951) More Information Similar Motifs Found	motif file (matrix)
3		1e-87	-2.012e+02	33.27%	22.79%	18.6bp (18.6bp)	JRNKRP/RRM1/Gallus_gallus-RNCMP00286-PBM/HughesRNA(0.860) More Information Similar Motifs Found	motif file (matrix)
4		1e-89	-1.863e+02	15.80%	8.67%	18.6bp (18.6bp)	phodungp/Bergmann/Rv(0.782) More Information Similar Motifs Found	motif file (matrix)
5		1e-76	-1.760e+02	26.10%	17.17%	18.7bp (19.0bp)	BHLHE22/ncv2/MA1435.1/jaspar(0.864) More Information Similar Motifs Found	motif file (matrix)
6		1e-41	-1.011e+02	7.56%	3.86%	18.6bp (19.0bp)	WKKY2/MA1091.1/jaspar(0.870) More Information Similar Motifs Found	motif file (matrix)
7		1e-33	-7.640e+01	8.84%	5.26%	17.9bp (18.6bp)	CRP2/MA0975.1/jaspar(0.834) More Information Similar Motifs Found	motif file (matrix)
8		1e-31	-7.353e+01	4.33%	1.99%	19.1bp (19.1bp)	Iron-Alpha1/HLH/NCCT-Ara-CHP-Seq(La.../Homer(0.871)) More Information Similar Motifs Found	motif file (matrix)
9		1e-25	-5.780e+01	1.74%	0.54%	16.7bp (17.7bp)	REMS/RRM1/Homo_sapiens-RNCMP00173-PBM/HughesRNA(0.890) More Information Similar Motifs Found	motif file (matrix)
10		1e-25	-5.760e+01	2.24%	0.82%	19.8bp (18.3bp)	RFX1/MA1554.1/jaspar(0.767) More Information Similar Motifs Found	motif file (matrix)
11		1e-22	-5.212e+01	2.92%	1.31%	17.7bp (19.0bp)	CG1471/RRM1/Drosophila_melanogaster-RNCMP00066-PBM/HughesRNA(0.905) More Information Similar Motifs Found	motif file (matrix)
12		1e-22	-5.103e+01	2.33%	0.94%	18.6bp (17.7bp)	TF3/AC2H2/col-TF3A-DAP-Seq(GSE60143)/Homer(0.916) More Information Similar Motifs Found	motif file (matrix)
13		1e-21	-5.049e+01	6.07%	3.65%	19.2bp (17.7bp)	URF1/RRM1/Homo_sapiens-RNCMP00090-PBM/HughesRNA(0.870) More Information Similar Motifs Found	motif file (matrix)
14		1e-16	-3.806e+01	4.00%	2.31%	17.6bp (18.2bp)	Met1/Homer03/Mat-Cdb-Mat1-CHP-Seq(GSE44085) (Homer0.955) More Information Similar Motifs Found	motif file (matrix)
15		1e-16	-3.756e+01	9.68%	6.13%	18.8bp (20.1bp)	CUX1/Homer03/KG-CUX1-CHP-Seq(GSE92882)/Homer(0.873) More Information Similar Motifs Found	motif file (matrix)
16		1e-15	-3.595e+01	1.25%	0.43%	17.3bp (18.2bp)	TRP2/MT/Related/colamp-TRP2-DAP-Seq(GSE60143)/Homer(0.965) More Information Similar Motifs Found	motif file (matrix)
17		1e-12	-2.850e+01	1.16%	0.45%	18.6bp (19.1bp)	MAC1/Literature/Harrison/Yeast(0.857) More Information Similar Motifs Found	motif file (matrix)

Enriched motifs in Intergenic CNSs

Rank	Motif	P-value	log P-value	% of Targets	% of Background	STD(Bg STD)	Best Match/Details	Motif File
1		1e-1333	-3.071e+03	64.60%	35.72%	19.8bp (18.5bp)	A2BP1/RRM1/Drosophila_melanogaster-RNCMP00123-PBM/HughesRNA(0.940) More Information Similar Motifs Found	motif file (matrix)
2		1e-858	-1.978e+03	62.22%	38.90%	20.2bp (18.6bp)	GATA15/MA1016.1/jaspar(0.961) More Information Similar Motifs Found	motif file (matrix)
3		1e-660	-1.521e+03	64.20%	43.80%	20.3bp (19.1bp)	CDMA0015.1/jaspar(0.819) More Information Similar Motifs Found	motif file (matrix)
4		1e-641	-1.476e+03	33.24%	16.96%	20.1bp (18.5bp)	WKKY2/MA1091.1/jaspar(0.842) More Information Similar Motifs Found	motif file (matrix)
5		1e-572	-1.319e+03	38.11%	20.86%	20.2bp (19.2bp)	phodungp/Bergmann/Rv(0.991) More Information Similar Motifs Found	motif file (matrix)
6		1e-448	-1.034e+03	28.85%	15.80%	20.3bp (18.6bp)	PBF/MA0604.1/jaspar(0.856) More Information Similar Motifs Found	motif file (matrix)
7		1e-448	-1.034e+03	22.68%	10.35%	20.3bp (19.2bp)	via/MA0352.1/jaspar(0.941) More Information Similar Motifs Found	motif file (matrix)
8		1e-367	-8.472e+02	28.15%	15.44%	20.0bp (18.5bp)	ATHB92/Homer03/col-ATHB9-DAP-Seq(GSE60143)/Homer(0.972) More Information Similar Motifs Found	motif file (matrix)
9		1e-310	-7.157e+02	13.00%	5.13%	20.7bp (19.0bp)	Brc1/PG1/Homer03/NPC-Brc1-CHP-Seq(GSE35496) (Homer0.772) More Information Similar Motifs Found	motif file (matrix)
10		1e-195	-4.486e+02	11.12%	5.16%	19.9bp (18.3bp)	SFR/MA1322.1/jaspar(0.892) More Information Similar Motifs Found	motif file (matrix)
11		1e-185	-4.279e+02	16.10%	9.00%	19.8bp (19.4bp)	LIN54/MA0619.1/jaspar(0.794) More Information Similar Motifs Found	motif file (matrix)
12		1e-174	-4.008e+02	10.26%	4.82%	19.7bp (17.9bp)	PF11695/RRM1/Plasmodium_falciparum-RNCMP00203-PBM/HughesRNA(0.804) More Information Similar Motifs Found	motif file (matrix)
13		1e-128	-2.951e+02	5.55%	2.18%	20.1bp (19.4bp)	bhm/MA0226.1/jaspar(0.955) More Information Similar Motifs Found	motif file (matrix)
14		1e-73	-1.691e+02	2.57%	0.86%	20.8bp (18.6bp)	Rap210/AP2/ERBP/col-Rap210-DAP-Seq(GSE60143)/Homer(0.838) More Information Similar Motifs Found	motif file (matrix)
15		1e-55	-1.266e+02	2.10%	0.75%	20.6bp (20.2bp)	AZF1/CH22/colamp-AZF1-DAP-Seq(GSE60143)/Homer(0.749) More Information Similar Motifs Found	motif file (matrix)
16		1e-31	-7.230e+01	1.40%	0.56%	18.3bp (16.7bp)	LARK/RRM1/Zn/Drosophila_melanogaster-RNCMP00035-PBM/HughesRNA(0.731) More Information Similar Motifs Found	motif file (matrix)
17		1e-22	-5.128e+01	1.03%	0.42%	23.2bp (20.8bp)	PU1a/MA1618.1/jaspar(0.750) More Information Similar Motifs Found	motif file (matrix)
18		1e-19	-4.587e+01	0.79%	0.30%	19.4bp (18.1bp)	GSEF/RRM1/Zn/Homo_sapiens-RNCMP00073-PBM/HughesRNA(0.904) More Information Similar Motifs Found	motif file (matrix)
19		1e-18	-4.293e+01	0.75%	0.28%	18.6bp (18.6bp)	OPI1/Literature/Harrison/Yeast(0.846) More Information Similar Motifs Found	motif file (matrix)
20		1e-13	-3.167e+01	0.79%	0.37%	21.0bp (17.6bp)	RAP211/AP2/ERBP/colamp-RAP211-DAP-Seq(GSE60143) (Homer0.872) More Information Similar Motifs Found	motif file (matrix)

Enriched motifs in Distal-promoter CNSs

Rank	Motif	P-value	log P-value	% of Targets	% of Background	STD(Bg STD)	Best Match/Details	Motif File
1		1e-190	-4.382e+02	43.67%	21.95%	19.3bp (18.6bp)	A2BP1/RRM1/Drosophila_melanogaster-RNCMP00123-PBM/HughesRNA(0.922) More Information Similar Motifs Found	motif file (matrix)
2		1e-122	-2.831e+02	43.67%	25.79%	18.5bp (18.3bp)	JRNKRP/RRM1/Drosophila_melanogaster-RNCMP00138-PBM/HughesRNA(0.777) More Information Similar Motifs Found	motif file (matrix)
3		1e-98	-2.259e+02	38.49%	23.06%	19.3bp (19.0bp)	rex/MA0222.1/jaspar(0.811) More Information Similar Motifs Found	motif file (matrix)
4		1e-91	-2.113e+02	20.97%	9.69%	18.3bp (18.5bp)	Tp121/Homer03/mtS-Tp121-CHP-Seq(GSE55404)/Homer(0.963) More Information Similar Motifs Found	motif file (matrix)
5		1e-84	-1.951e+02	26.61%	14.28%	19.5bp (19.0bp)	Dof2/MA0620.1/jaspar(0.832) More Information Similar Motifs Found	motif file (matrix)
6		1e-29	-6.698e+01	7.50%	3.55%	19.1bp (18.2bp)	PUM1/PUF1/Drosophila_melanogaster-RNCMP00101-PBM/HughesRNA(0.857) More Information Similar Motifs Found	motif file (matrix)
7		1e-21	-4.855e+01	2.92%	0.98%	20.4bp (19.4bp)	wc-1/MA1437.1/jaspar(0.909) More Information Similar Motifs Found	motif file (matrix)
8		1e-8	-1.851e+01	1.64%	0.71%	17.7bp (18.6bp)	GAGA-repeat/Arabidopsis-Promoters/Homer(0.781) More Information Similar Motifs Found	motif file (matrix)

Fig S13. *De novo* motif analysis with the tool Homer indicating the top enriched 6-8 bp motifs found in CNSs at different genomic regions.

Rank	Motif	Best Match/Details
1		CRF4(AP2EREBP)/colamp-CRF4-DAP-Seq(GSE60143) /Homer(0.934) More Information Similar Motifs Found
2		TCP23/MA1066.1/Jaspar(0.904) More Information Similar Motifs Found
3		odd/dmmpmm(Noyes)/fly(0.776) More Information Similar Motifs Found
4		FRS9(ND)/col-FRS9-DAP-Seq(GSE60143)/Homer(0.930) More Information Similar Motifs Found
5		NFIX/MA0671.1/Jaspar(0.742) More Information Similar Motifs Found
6		HOXA1(Homeobox)/mES-Hoxa1-ChIP-Seq(SRP084292) /Homer(0.859) More Information Similar Motifs Found
7		MET31/MA0333.1/Jaspar(0.839) More Information Similar Motifs Found
8		At1g64620(C2C2dof)/colamp-At1g64620-DAP-Seq(GSE60143) /Homer(0.729) More Information Similar Motifs Found
9		REM19(REM)/colamp-REM19-DAP-Seq(GSE60143) /Homer(0.966) More Information Similar Motifs Found

Enriched motifs in Proximal-promoter ATAC peaks

Rank	Motif	Best Match/Details
1		Smp_067420(RRM)/Schistosoma_mansoni-RNCMPT00232-PBM/HughesRNA(0.949) More Information Similar Motifs Found
2		HNRNPR(RRM)/Gallus_gallus-RNCMPT00288-PBM/HughesRNA(0.826) More Information Similar Motifs Found
3		RBMS3(RRM)/Homo_sapiens-RNCMPT00173-PBM/HughesRNA(0.982) More Information Similar Motifs Found
4		MXI1/MA1108.2/Jaspar(0.908) More Information Similar Motifs Found
5		FRS9(ND)/col-FRS9-DAP-Seq(GSE60143)/Homer(0.926) More Information Similar Motifs Found
6		An_0287(RRM)/Aspergillus_nidulans-RNCMPT00287-PBM/HughesRNA(0.760) More Information Similar Motifs Found
7		At5g08330(TCP)/col-At5g08330-DAP-Seq(GSE60143)/Homer(0.993) More Information Similar Motifs Found
8		CG2950(KH)/Drosophila_melanogaster-RNCMPT00007-PBM/HughesRNA(0.685) More Information Similar Motifs Found
9		AT2G33710/MA1245.2/Jaspar(0.938) More Information Similar Motifs Found

Enriched motifs in Proximal-promoter dREG sites

Fig S14. *De novo* motif analysis with Homer indicating the top enriched 6-8 bp motifs found in dREG locations and ATAC peaks at different genomic regions.

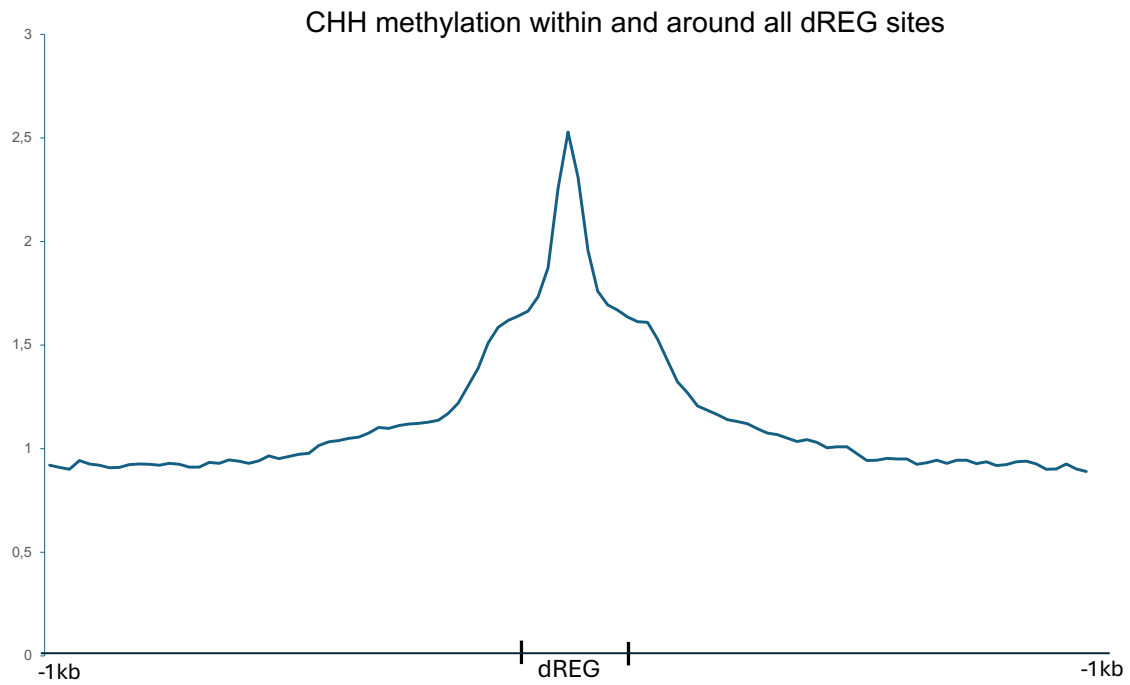


Fig S15. CHH methylation within and around 1 Kb of all detected dREG sites. The Y-axis is an arbitrary unit of signal.

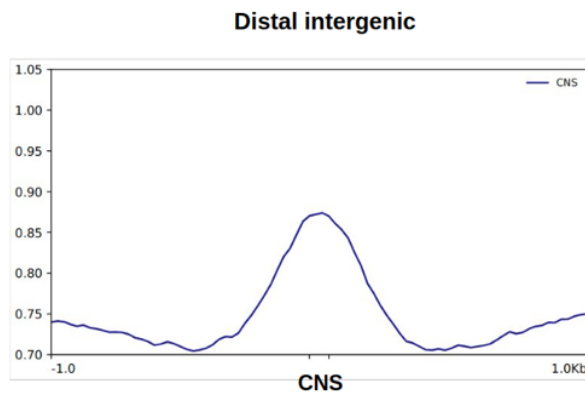
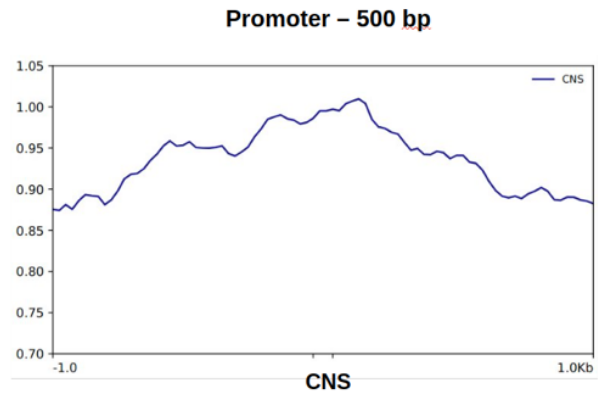
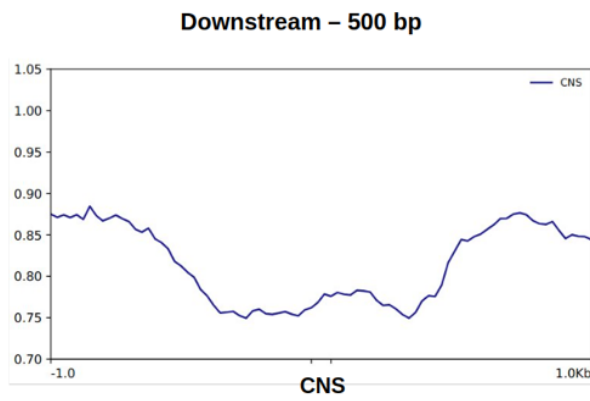
a**b****c**

Fig S16. H3K27me3 signal profile within and around CNS based on their locations.

(a) within-around distal intergenic **(b)** within-around 500 bp of promoters **(c)** within-around 500 bp downstream.

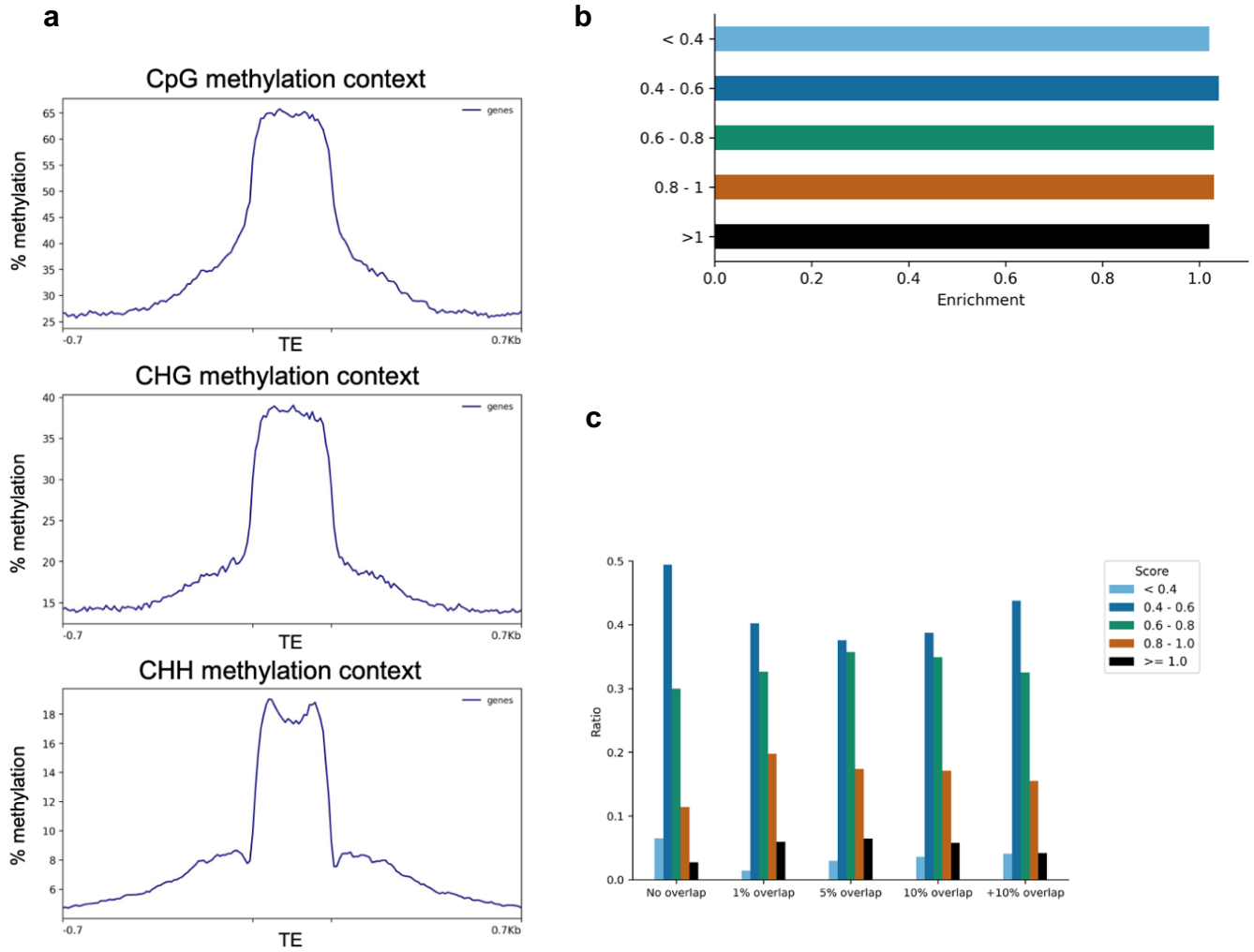


Fig S17. Methylation profile through the selection of dREG candidates. (a) Percentage of methylation by context within and around TEs that overlap dREG peaks by at least 1bp. (b) TE enrichment by dREG score bins. (c) Ratio of dREG score distribution by percentage of TE overlap, showing a similar distribution regardless of overlap with TEs or not.

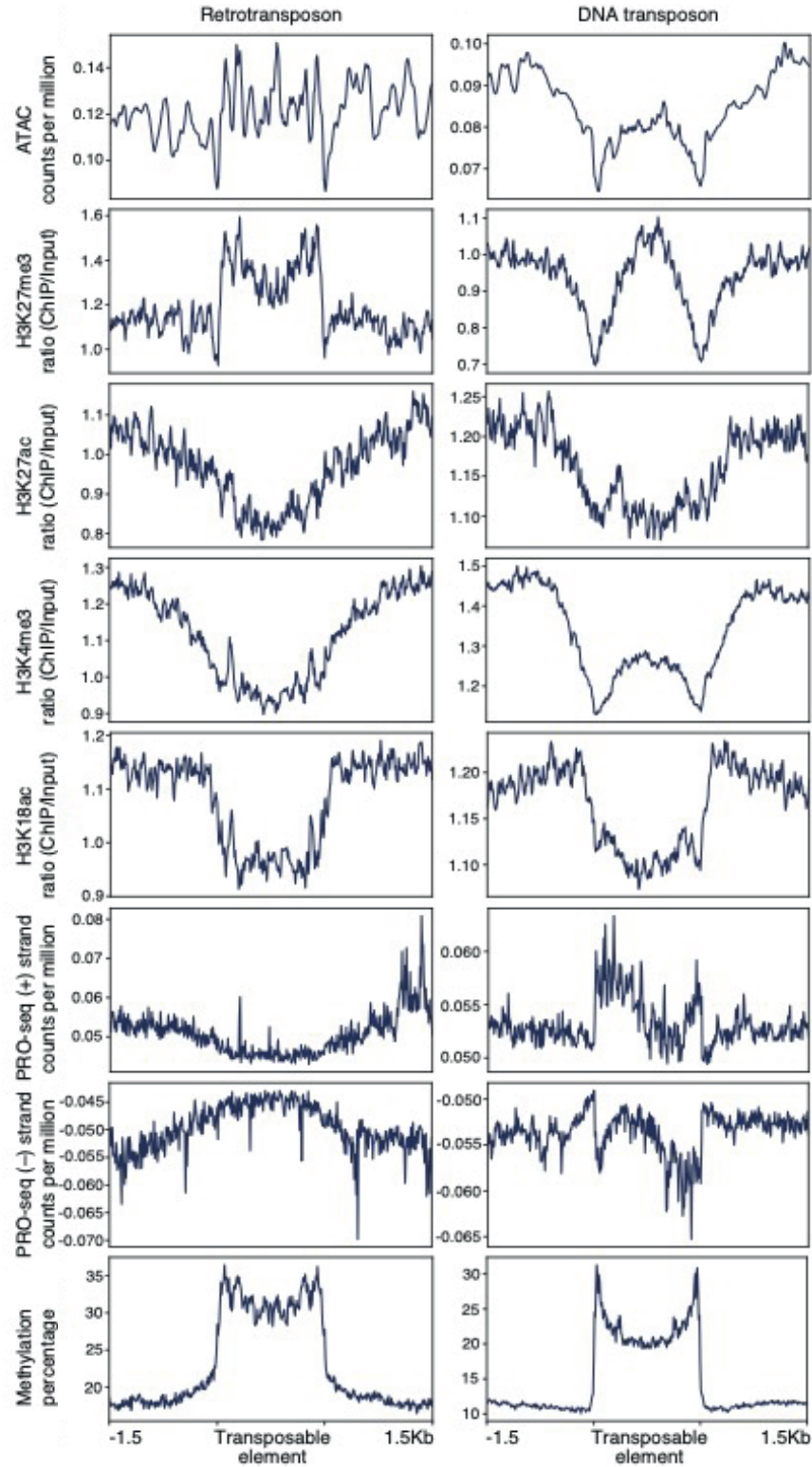


Fig S18. Epigenetic marks for coding and repetitive sequences in the rice genome.
(left) Class I retrotransposons and **(right)** class II DNA transposons.

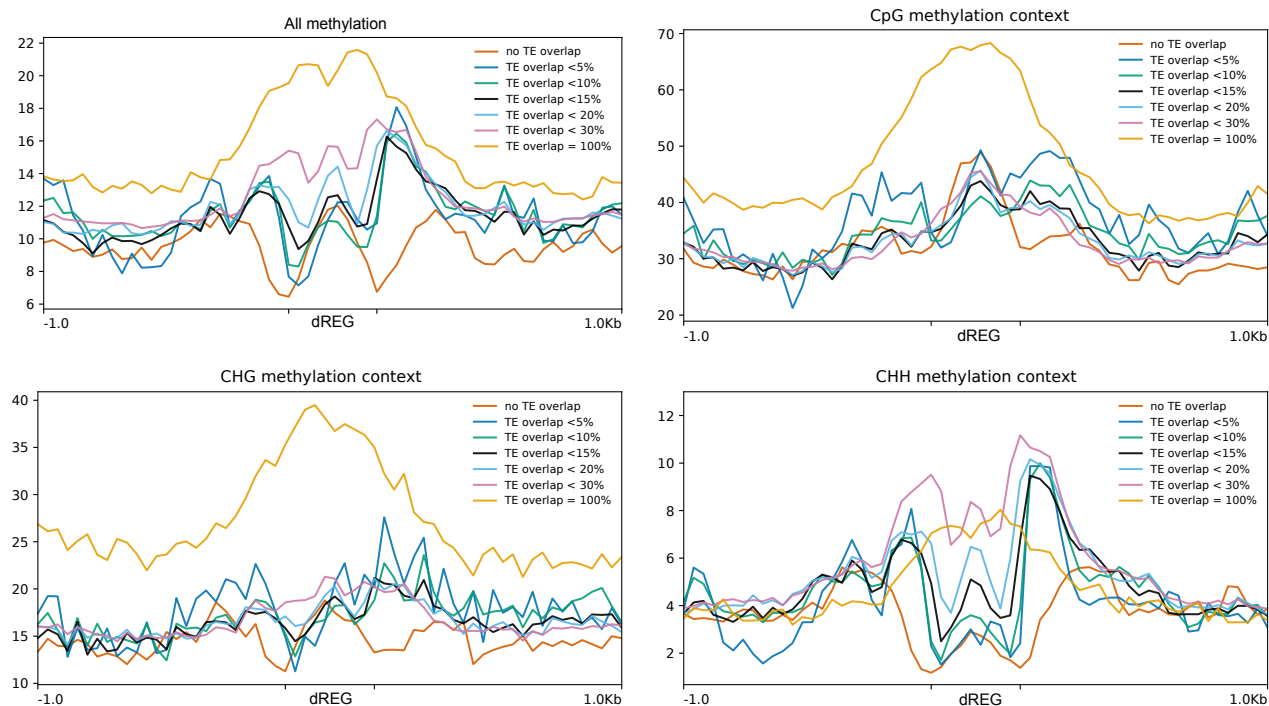
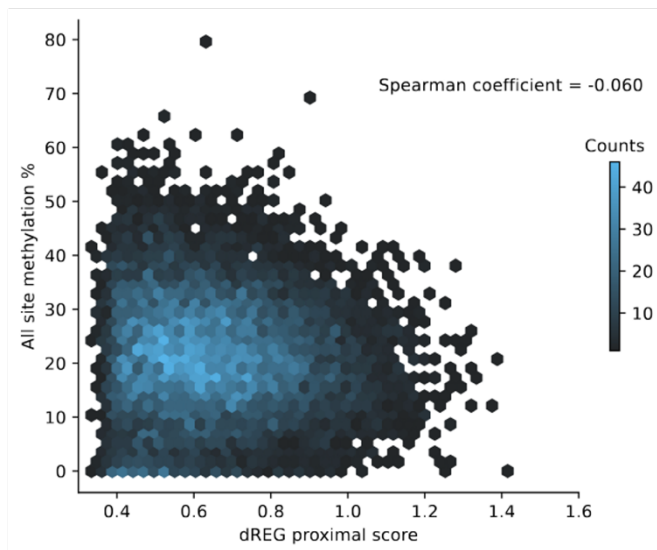


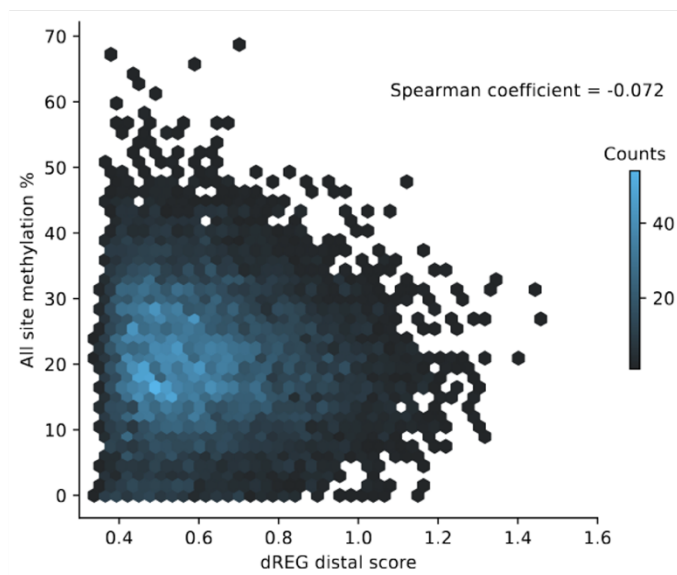
Fig S19. Percentage of methylation by context within and around intergenic dREG locations and their percentage of overlap with transposable elements (TEs). Each line shows the difference in methylation profile depending on the percentage overlap of dREG locations with TEs.

a



$P \text{ value} = 1.59 \times 10^{-09}$

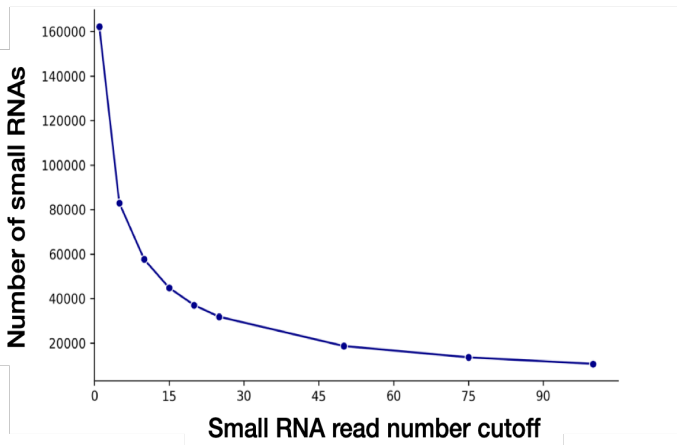
b



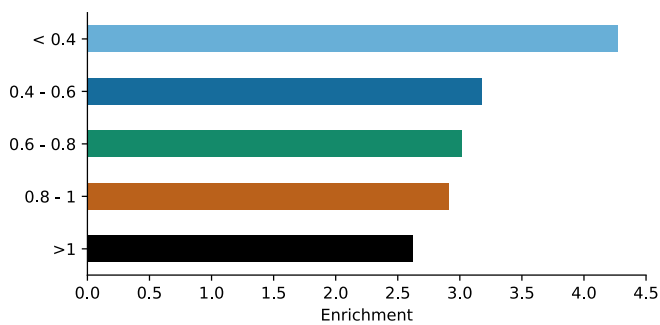
$P \text{ value} = 2,53 \times 10^{-12}$

Fig S20. Correlation of the percentage of cytosine methylation in all three contexts on the different dREG scores for dREG locations proximal (a) and distal (b) to genes. Sperman was used to calculate the correlation.

a



b



c

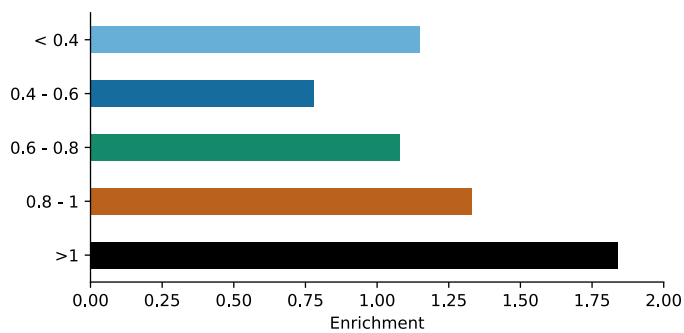


Fig S21. sRNA and ATAC enrichment analysis in dREG sites by bin score. (a) Plot showing the number of annotated small RNAs present depending on a minimum read number cutoff. We have selected a minimum cutoff of 20 reads. **(b)** Annotated small RNAs enrichment by dREG score bins. **(c)** ATAC peak enrichment by dREG score bins.

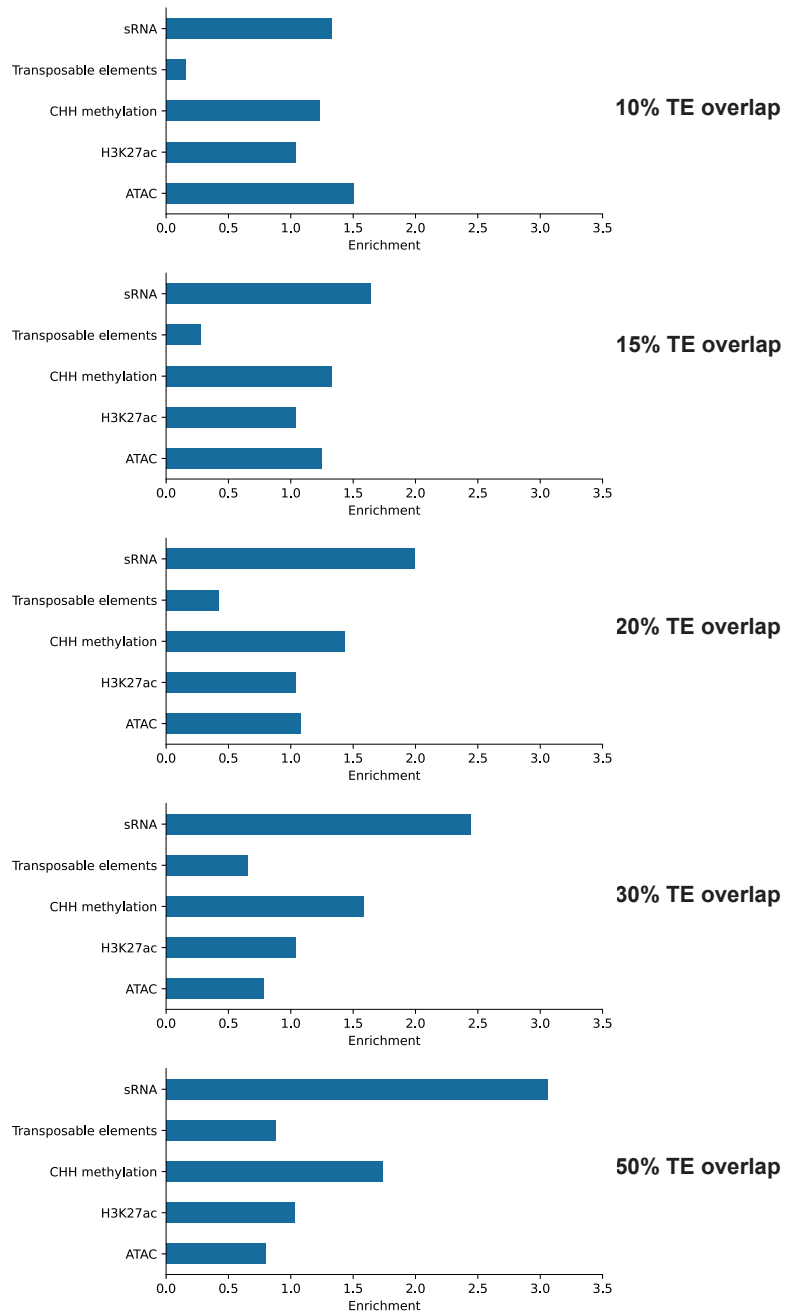


Fig S22. Enrichment analysis for dREG sites with TE overlap. Each bar represents an enrichment in open chromatin (ATAC), small RNA, CHH methylation, and H3K27ac found in dREG regions overlapping with transposable elements at different percentages.

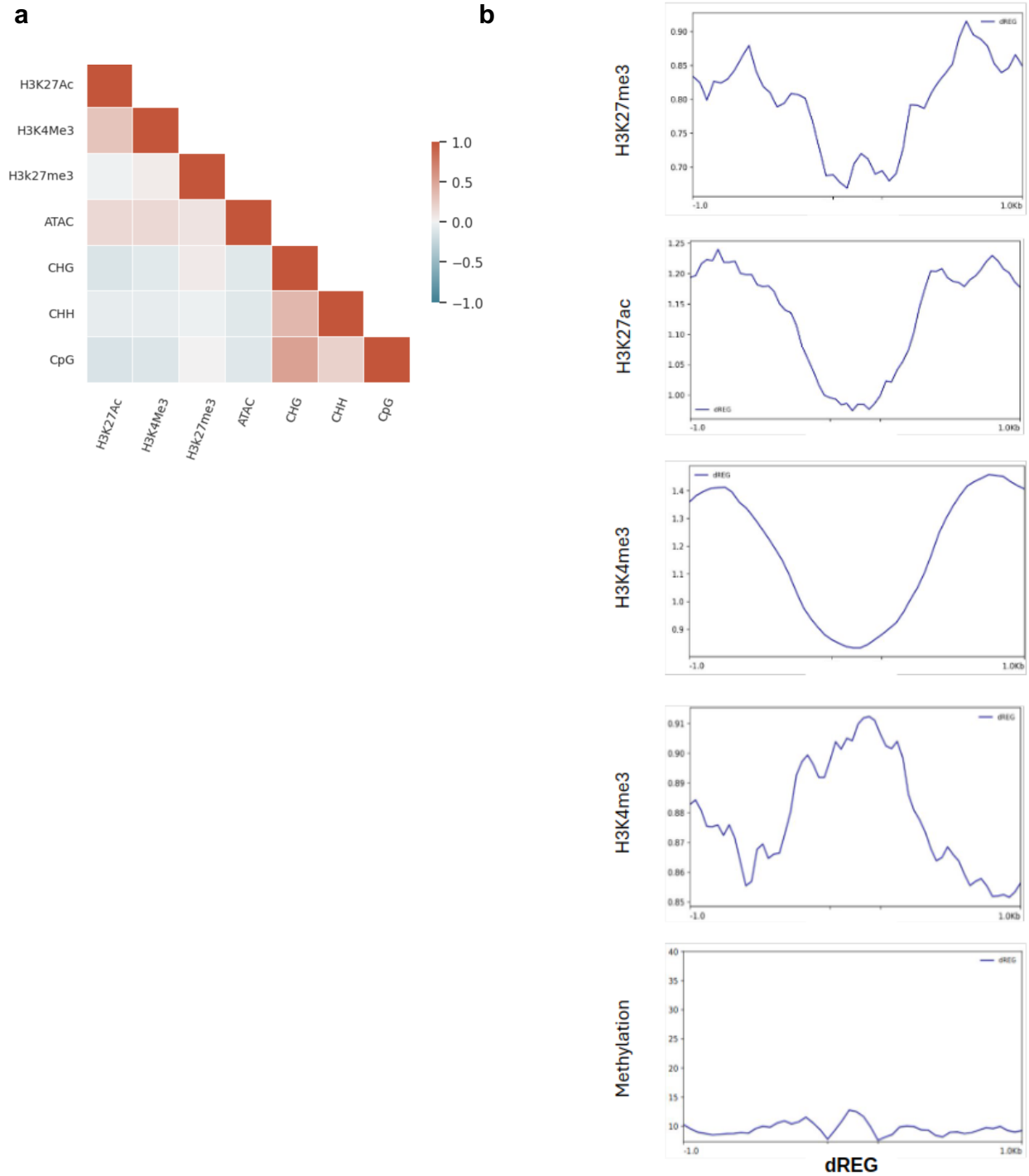


Fig S23. Chromatin marks within and around dREG post-filtering. In **(a)** a Pearson correlation for epigenomic characteristics methylation in all contexts (CpG, CHG, CHH), accessible chromatin (ATAC) and histone-tail modifications. In **(b)** the distribution the genomic features described in (a) but within dREG_{filtered} and +/- 1kilo base-pairs around.

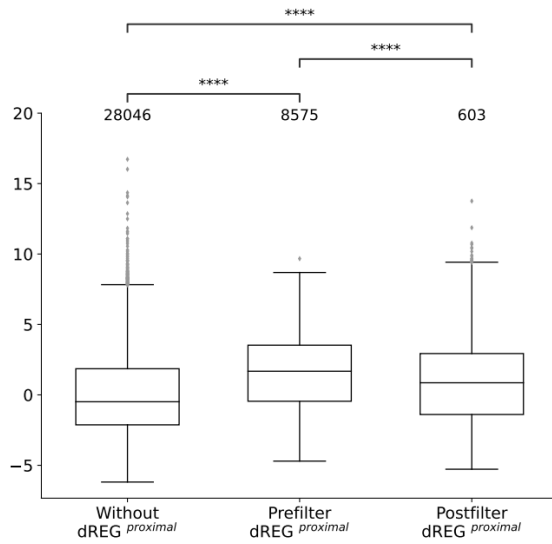
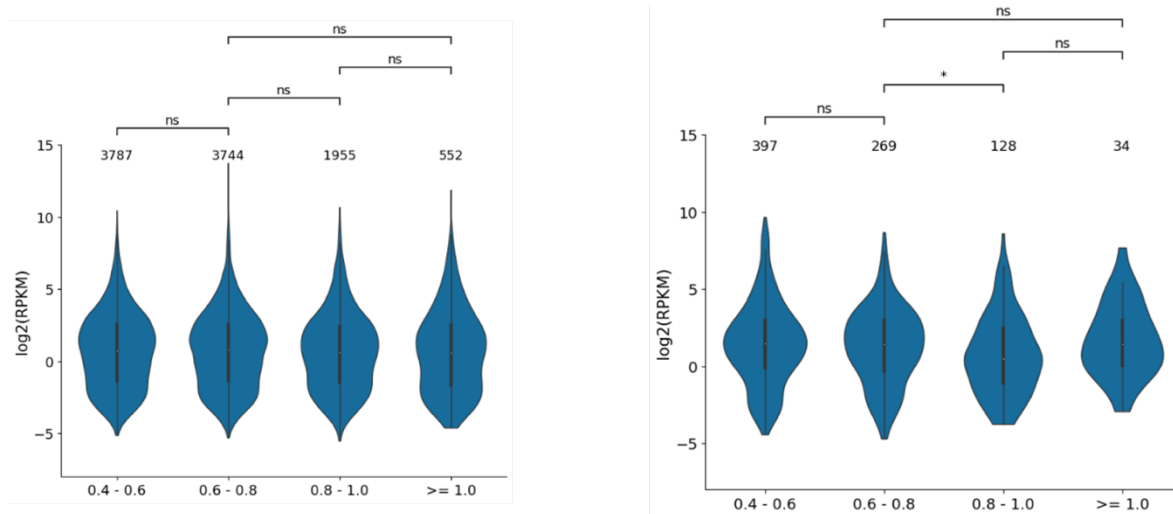
a**b**

Fig S24. Gene expression analysis based on the proximity of dREG sites. (a) Boxplot of gene expression based for genes close or not to dREG_{proximal} regions **(b)** Violin plots showing the expression levels (shown as log2 Reads per kilo base per million mapped reads, RPKM) for genes dREG_{proximal} divided by dREG score bins. dREG_{proximal} locations were divided by prefiltering steps (prefilter dREG_{proximal}) and post filtering steps (postfilter dREG_{proximal}) to retain candidate locations more likely to have a role in promoting gene expression. Asterisk (*) indicate significant differences after all pairwise comparisons using the Mann-Whitney U test with Bonferroni correction. Numbers in boxplot represent sample size of genes. * 0.01 < p < 0.05.

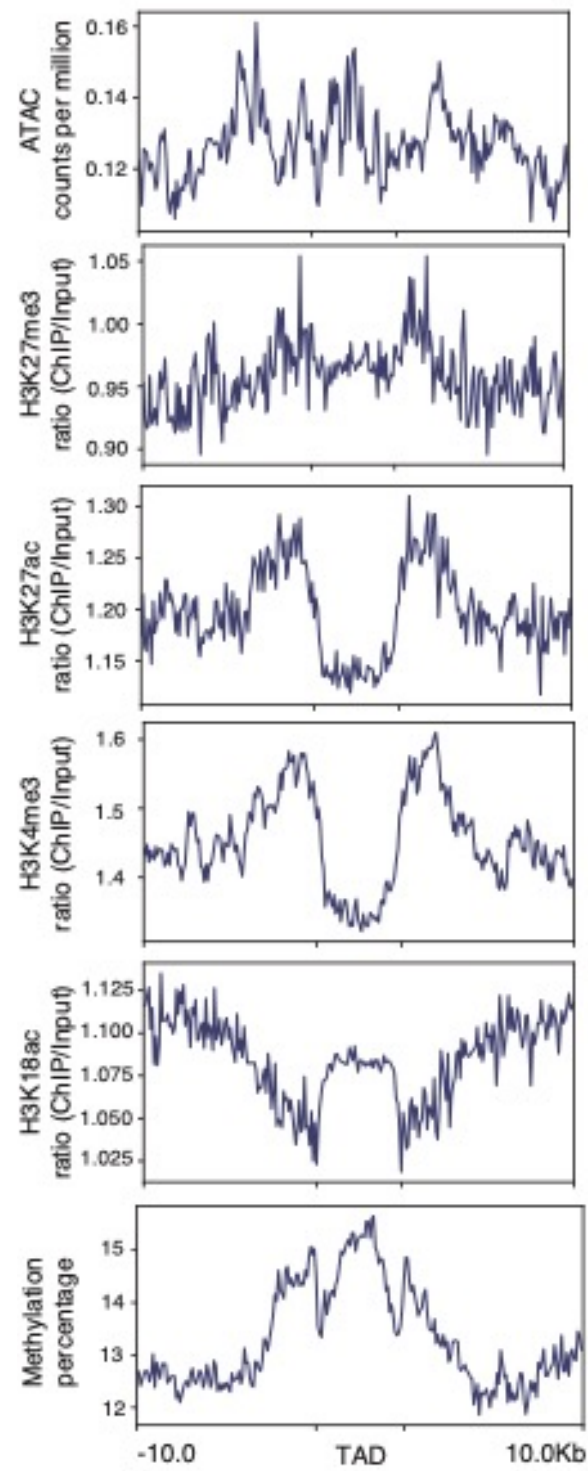
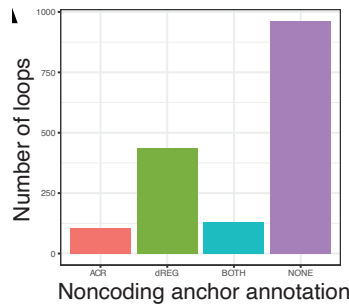
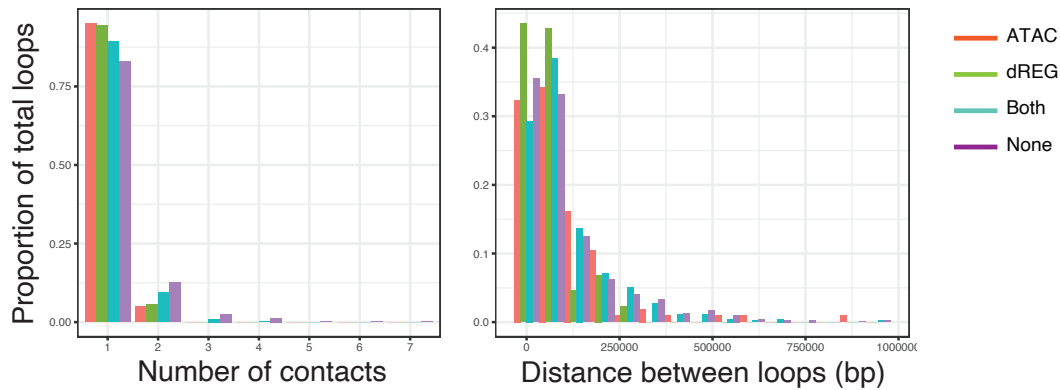


Fig S25. Epigenetic marks surrounding TAD boundaries.

a



b



c

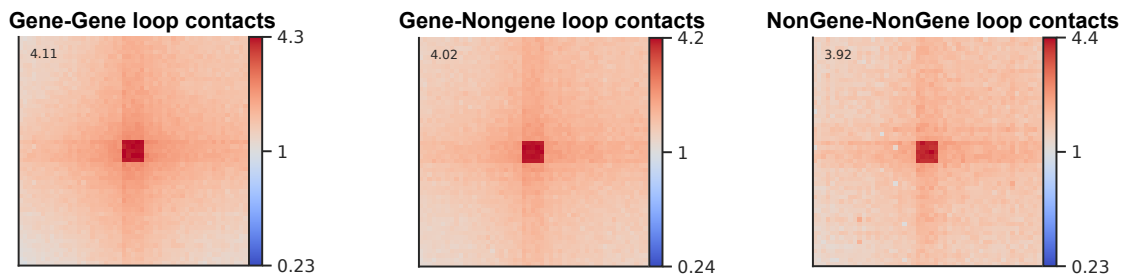


Fig S26. Distribution of loops detected by Pore-C for dREG_{distal} peaks. (a) Total number of noncoding-gene loops where the noncoding anchor contains either an ATAC peak, a dREG peak, both an ATAC and a dREG peak, or none. **(b)** Across the noncoding gene loops the number of noncoding anchor contacts per gene (left) and the distance between the gene and noncoding anchor (right). **(c)** APA scores for chromatin loops stratified by genic content reveal stronger contact enrichment between gene-gene anchors (4.11) compared to gene–non-gene (4.02) and non-gene–non-gene (3.92) interactions.

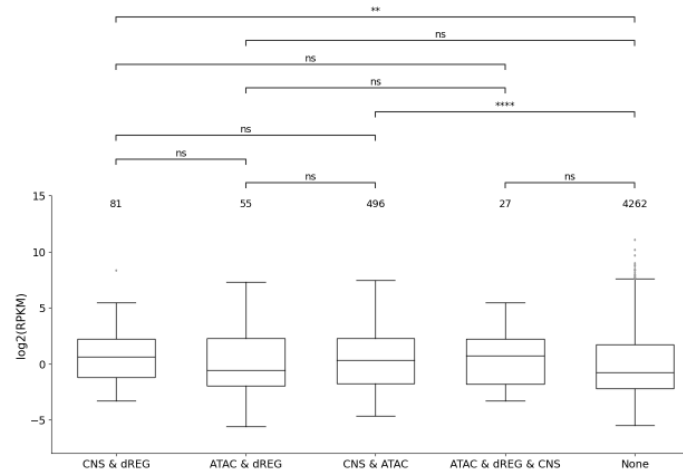
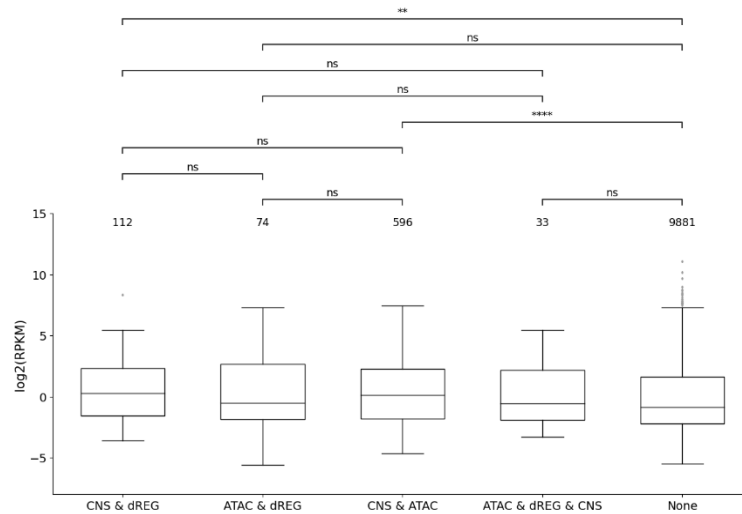
a**Non-genic contact loop with genic loop****b****Non-genic contact loop with more than one genic loop**

Fig S27. Expression levels analysis of genes presents in genic loops with a contact of a noncoding region. Boxplot showing the expression levels (shown as log₂ Reads per kilo base per million mapped reads, RPKM) for genes present in genic loops that contact a non-genic loop containing a combination of sequence features compared to loops without ATAC, CNS or dREG (none). **(a)** for one contact and **(b)** for more than one contact. Asterisk (*) indicate significant differences after all pairwise comparisons using the Mann-Whitney U test with Bonferroni correction. ** $1.00\text{e-}03 < p \leq 1.00\text{e-}02$; **** $p \leq 1.00\text{e-}04$ Numbers in boxplot represent sample size of genes.

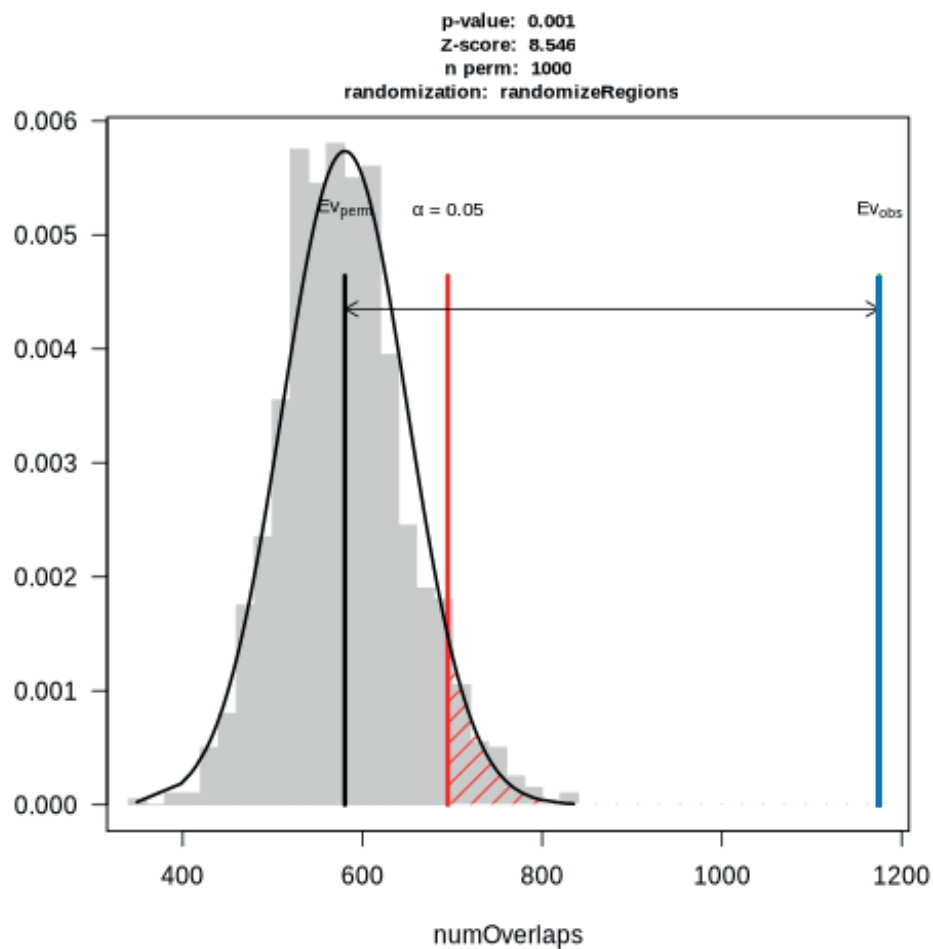


Fig S28. Association analysis between expression quantitative trait loci (eQTLs) regions and dREG_{filtered} sites identified in Figure 5c. Analysis was performed on 1000 permutation tests. The grey curves represent the number of overlaps of eQTLs found in randomized dREG_{filtered} regions with same length and chromosome location. The green line indicates number of eQTLs regions found in dREG_{filtered} sites, and black line represents the mean of the randomized evaluations, and the red line corresponds to the significance limit of p-value = 0.05. The p-value corresponding to the dREG_{filtered} sites is 0.001.

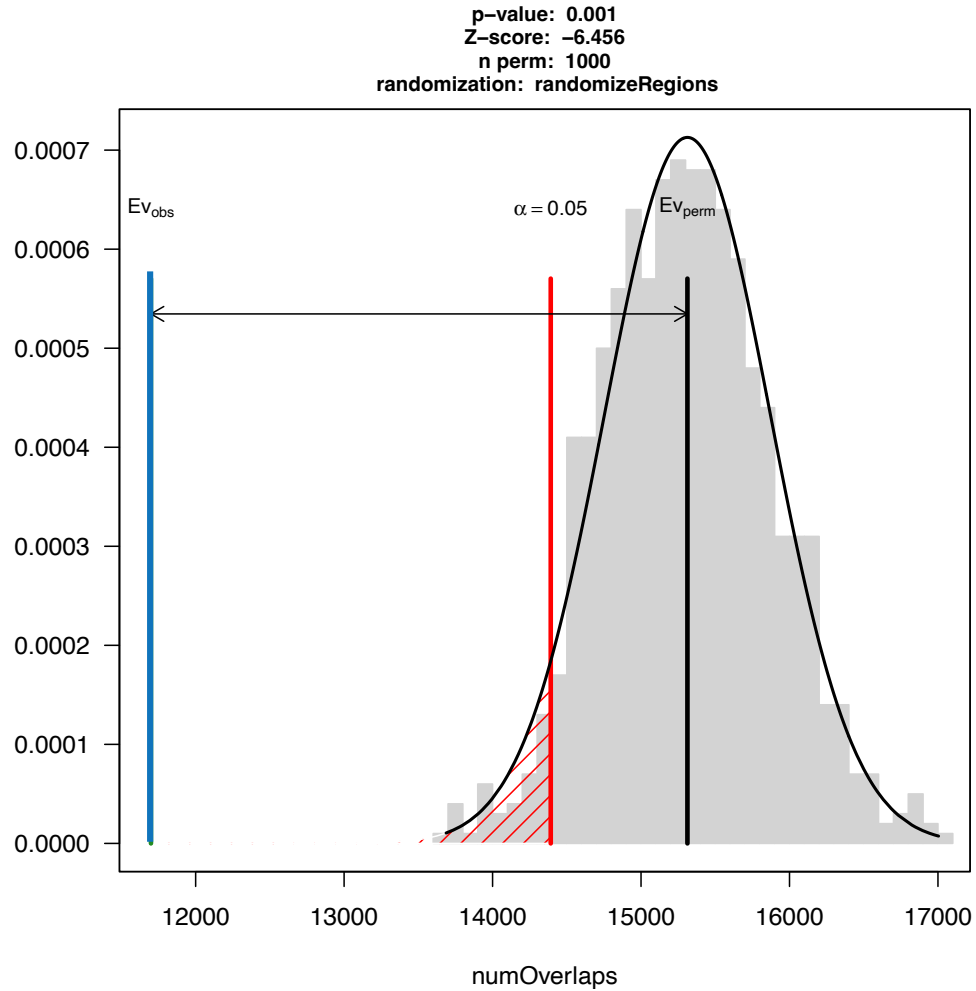


Fig S29. Association analysis between eQTLs regions and ATAC intergenic regions. Analysis was performed on 1000 permutation tests. The grey curves represent the number of overlaps of eQTLs found in randomized ATAC intergenic regions with same length and chromosome location. The green line indicates number of eQTLs regions found in ATAC sites, the black line represents the mean of the randomized evaluations, and the red line corresponds to the significance limit of p-value = 0.05. The p-value corresponding to ATAC intergenic regions is 0.001

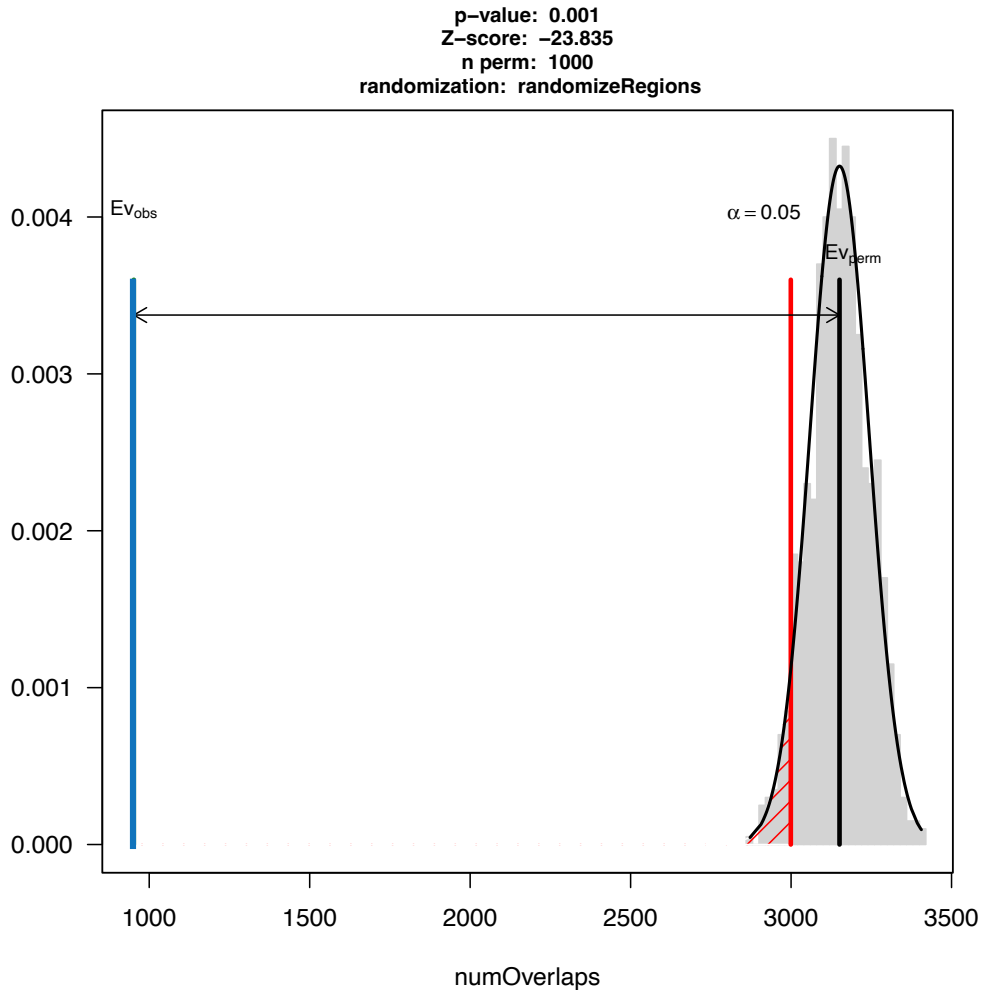


Fig S30. Association analysis between eQTLs regions and CNS intergenic regions.

Analysis was performed on 1000 permutation tests. The grey curves represent the number of overlaps of eQTLs found in randomized CNS intergenic regions with the same length and chromosome location. The green line indicates number of eQTLs regions found in CNS sites, the black line represents the mean of the randomized evaluations, and the red line corresponds to the significance limit of p-value = 0.05. The p-value corresponding to CNS intergenic regions is 0.001.

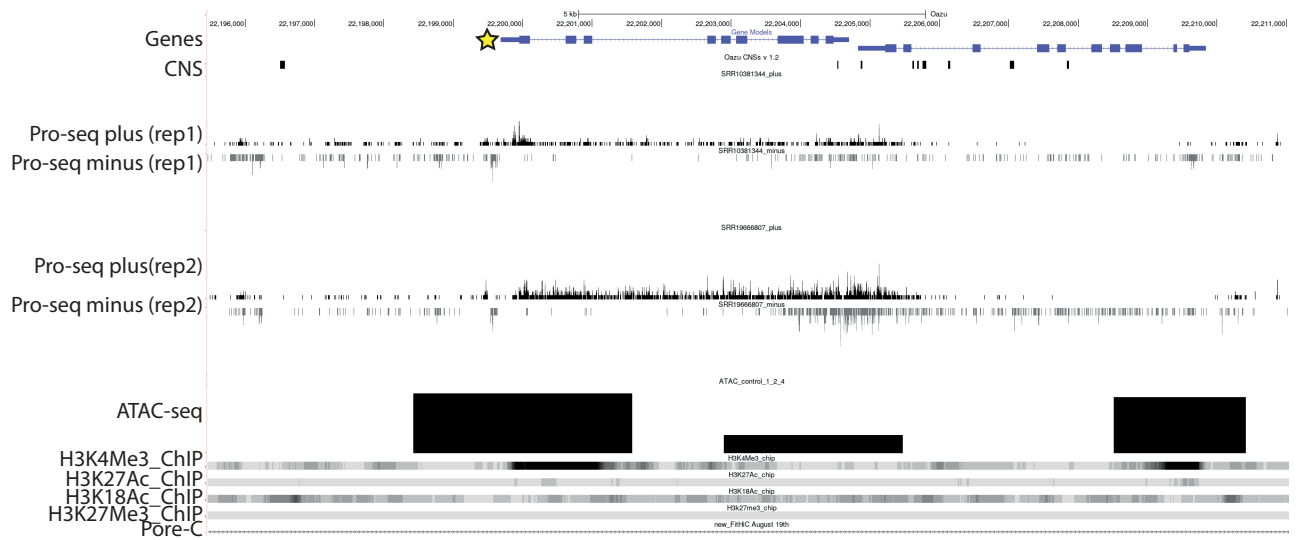


Fig S31. Genome Browser view of an Azucena locus (GAPDH). This is showing the different genomic tracks mentioned in this study at and around the locus GAPDH (OsAzu_08g0017550). Yellow star indicates the locus.

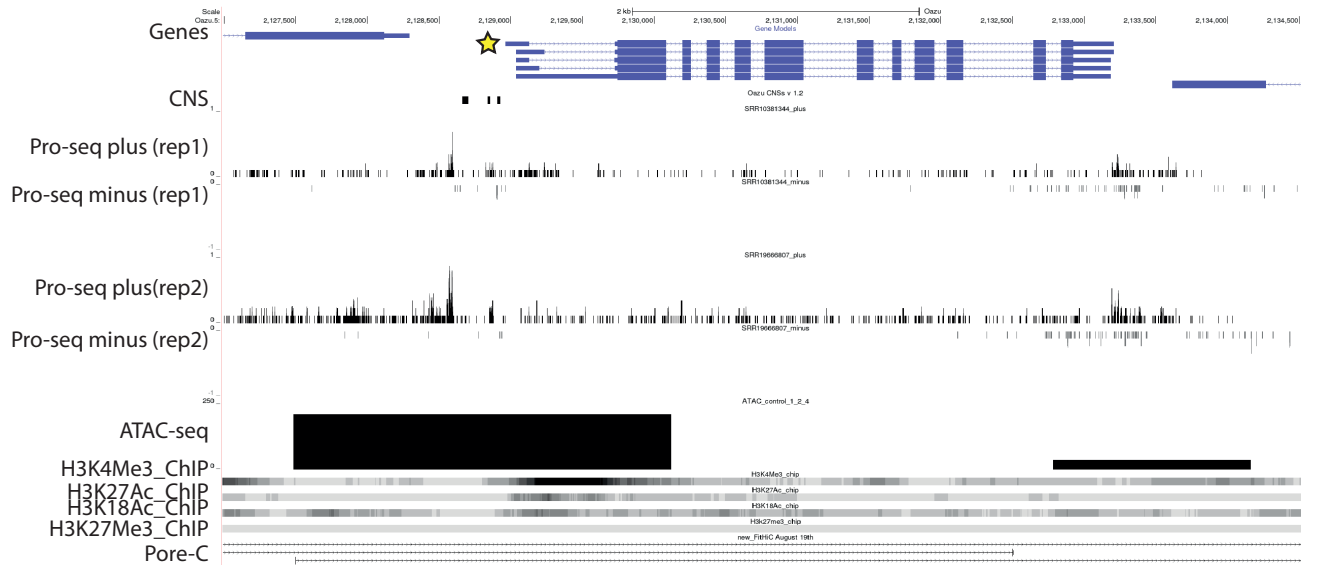


Fig S32. Genome Browser view of the Azucena locus (WRKY5). This is showing the different genomic tracks mentioned in this study at and around the locus WRKY5. Yellow star indicates the locus.

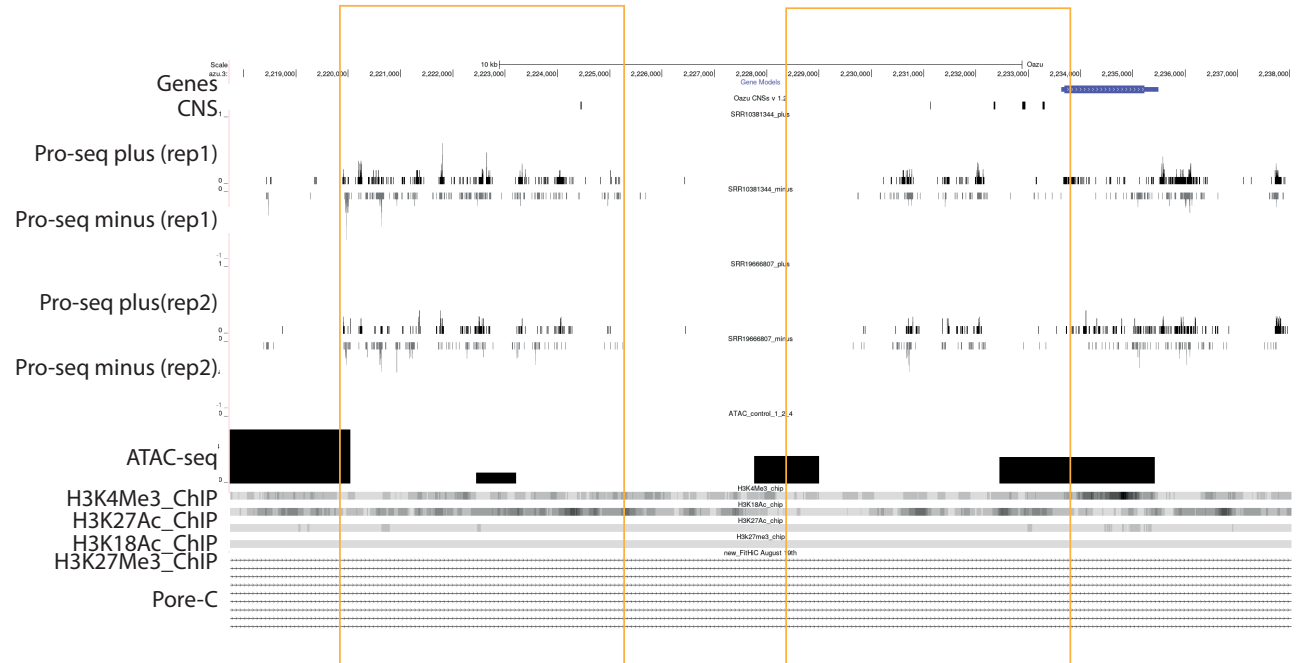


Fig S33. Genome Browser view of an Azucena intergenic region. This is showing the reproducibility of the two PRO-seq replicates in intergenic regions on chromosome 3. The different genomic tracks mentioned in this study are shown here. The yellow rectangles highlight regions with intergenic bi-directional PRO-seq signal in both replicates, where some regions overlap with ATAC-seq (left) and with CNSs (right).

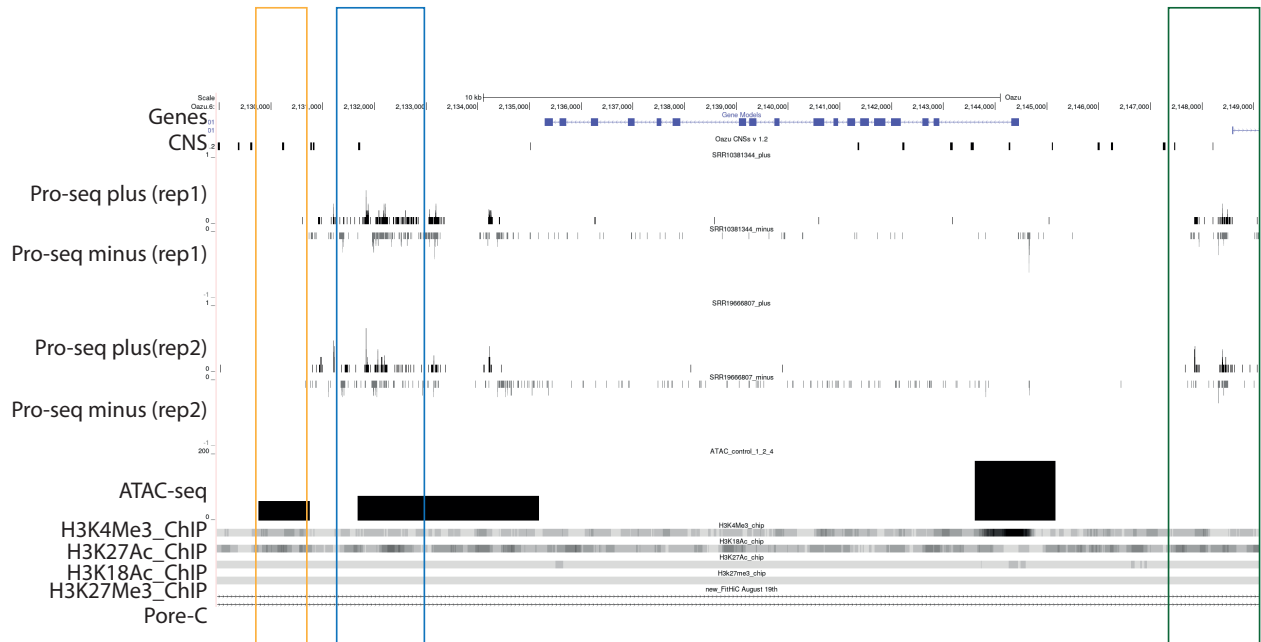


Fig S34. Genome Browser view of an Azucena intergenic region with three types of regulatory elements. This is showing an intergenic region on chromosome 8 with CNSs, ATAC peaks, and bi-directional PRO-seq signal. The different genomic tracks mentioned in this study are shown here. The yellow rectangles highlight an intergenic region with co-occurrence of CNS and ATAC; the blue rectangle a co-occurrence of CNS, bi-directional PRO-seq and ATAC; the green rectangle a co-occurrence of bi-directional PRO-seq and CNS. In all three highlighted cases we see an enrichment for the enhancer histone modification H3K18Ac.

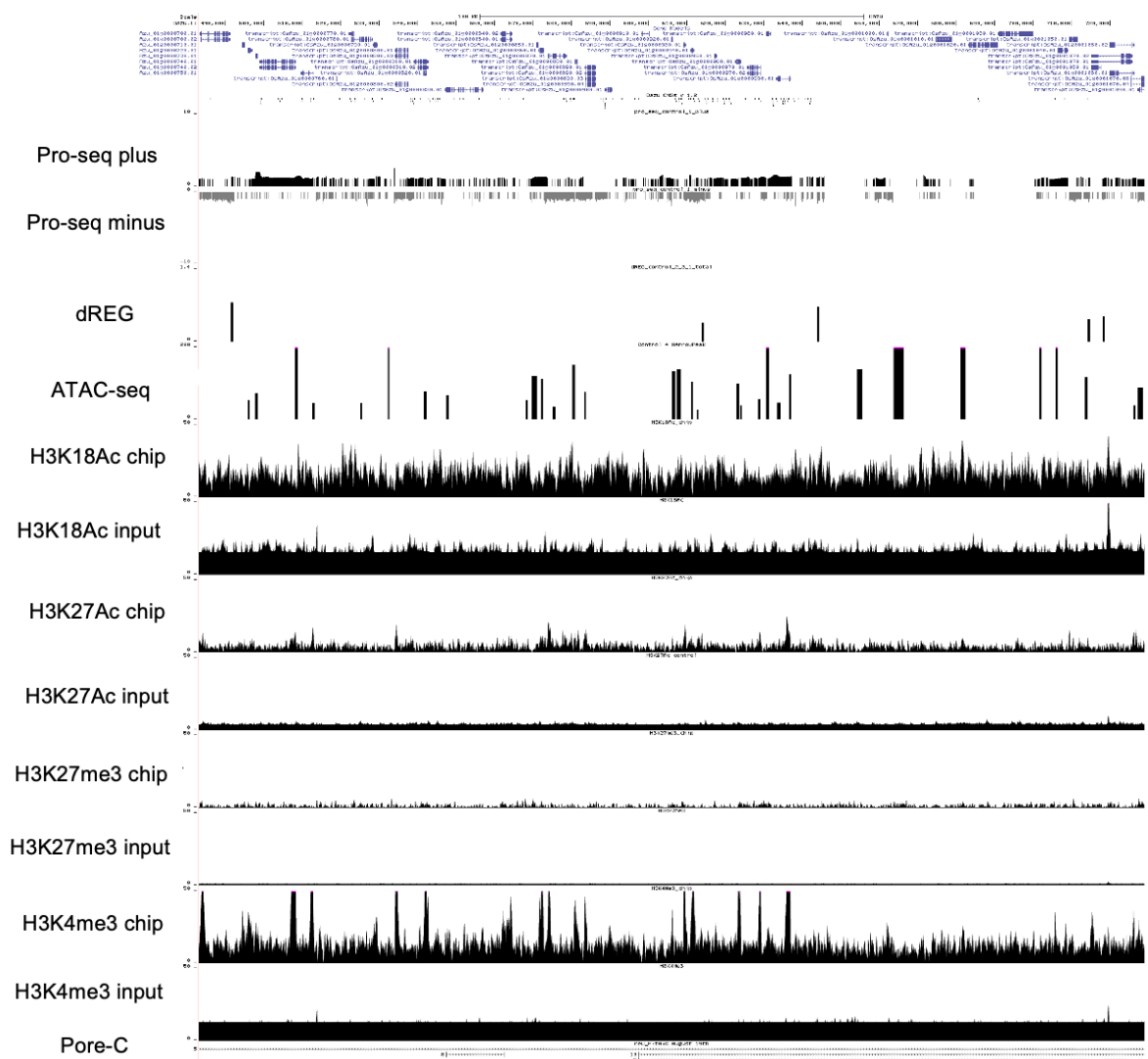


Fig S35. Genome Browser view of the Azucena genome #1. This is showing the different genomic tracks mentioned in this study.

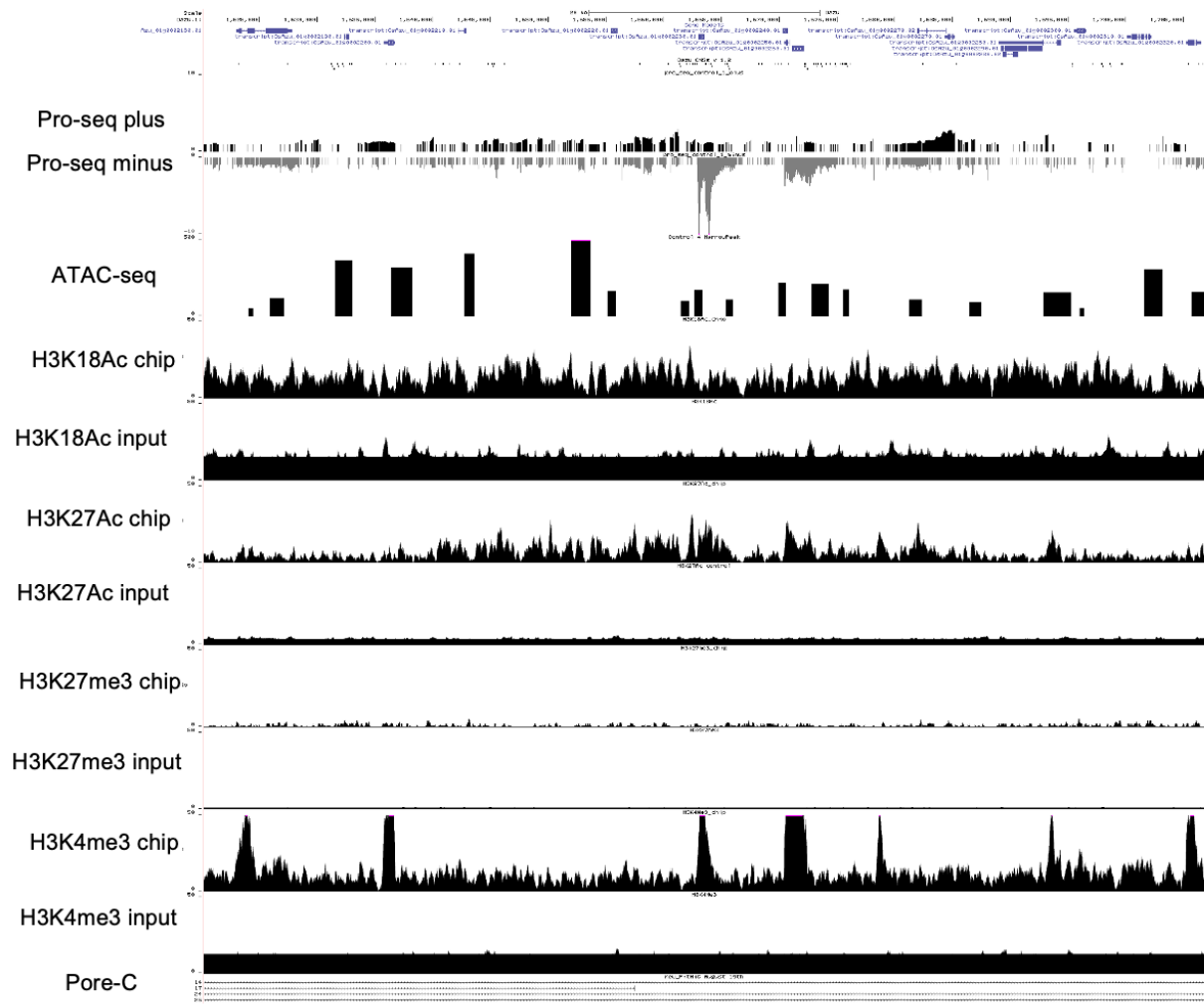


Fig S36. Genome Browser view of the Azucena genome #2. This is showing the different genomic tracks mentioned in this study.

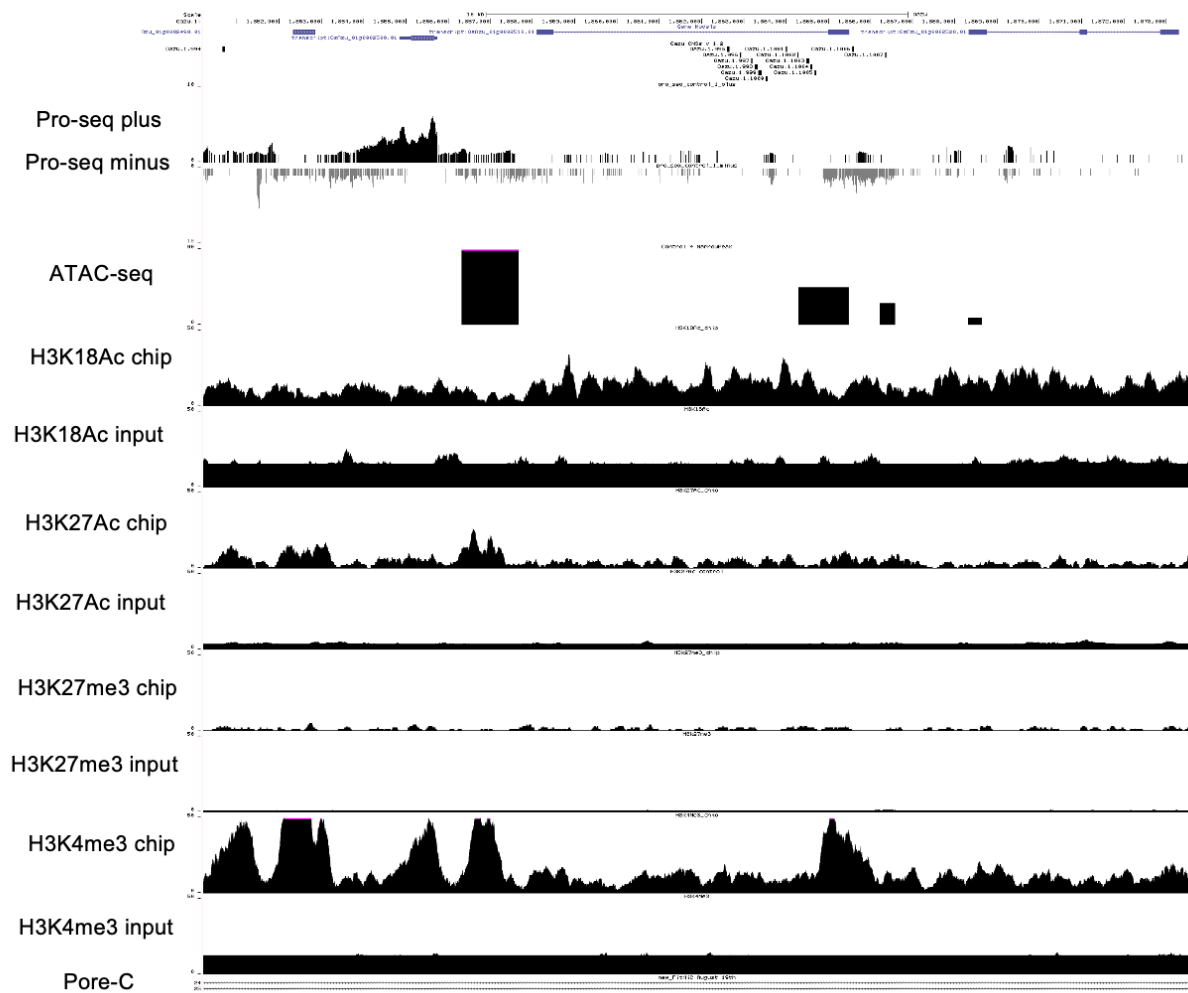


Fig S37. Genome Browser view of the Azucena genome #3. This is showing the different genomic tracks mentioned in this study.

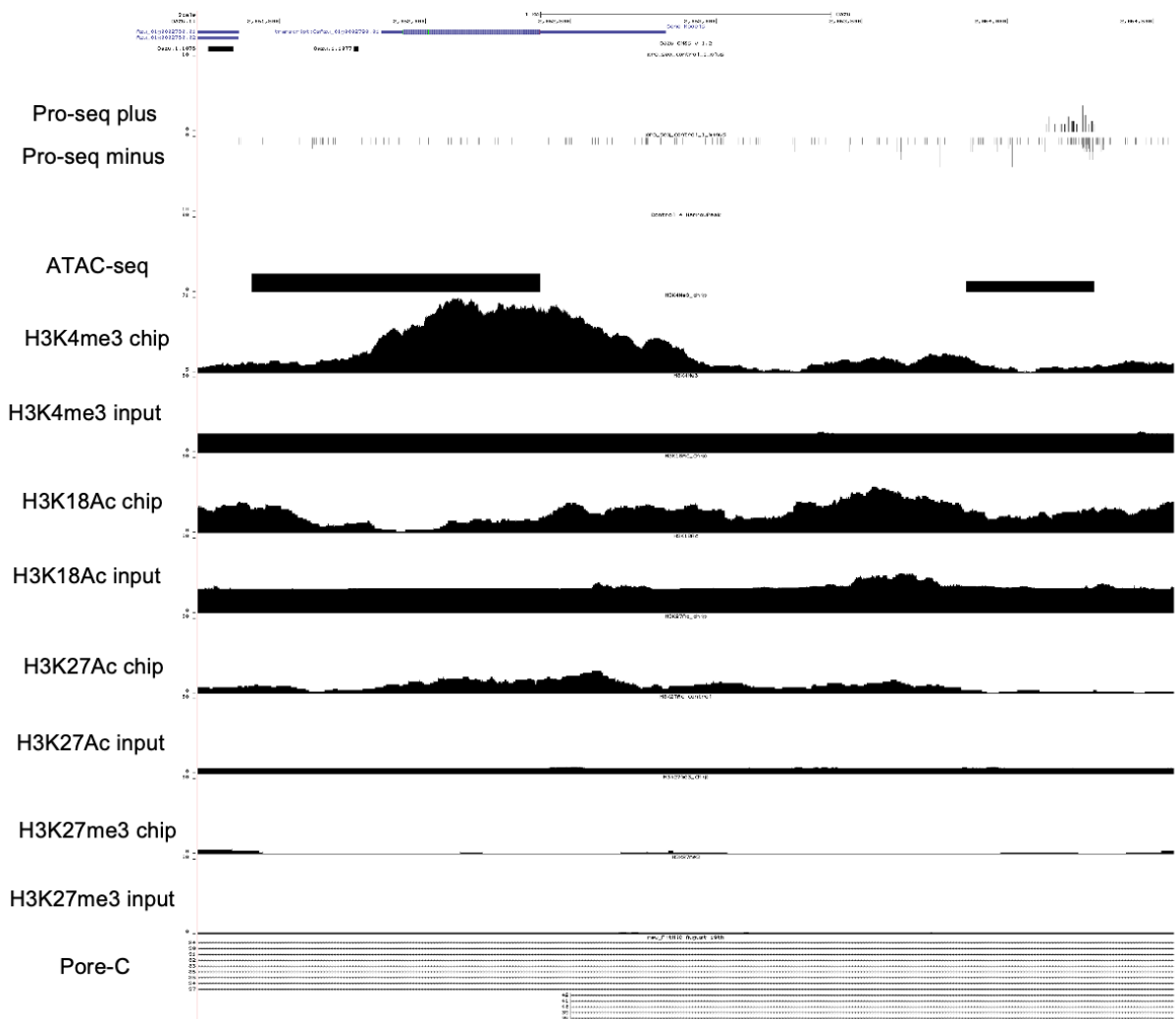


Fig S38. Genome Browser view of the Azucena genome #4. This is showing the different genomic tracks mentioned in this study.

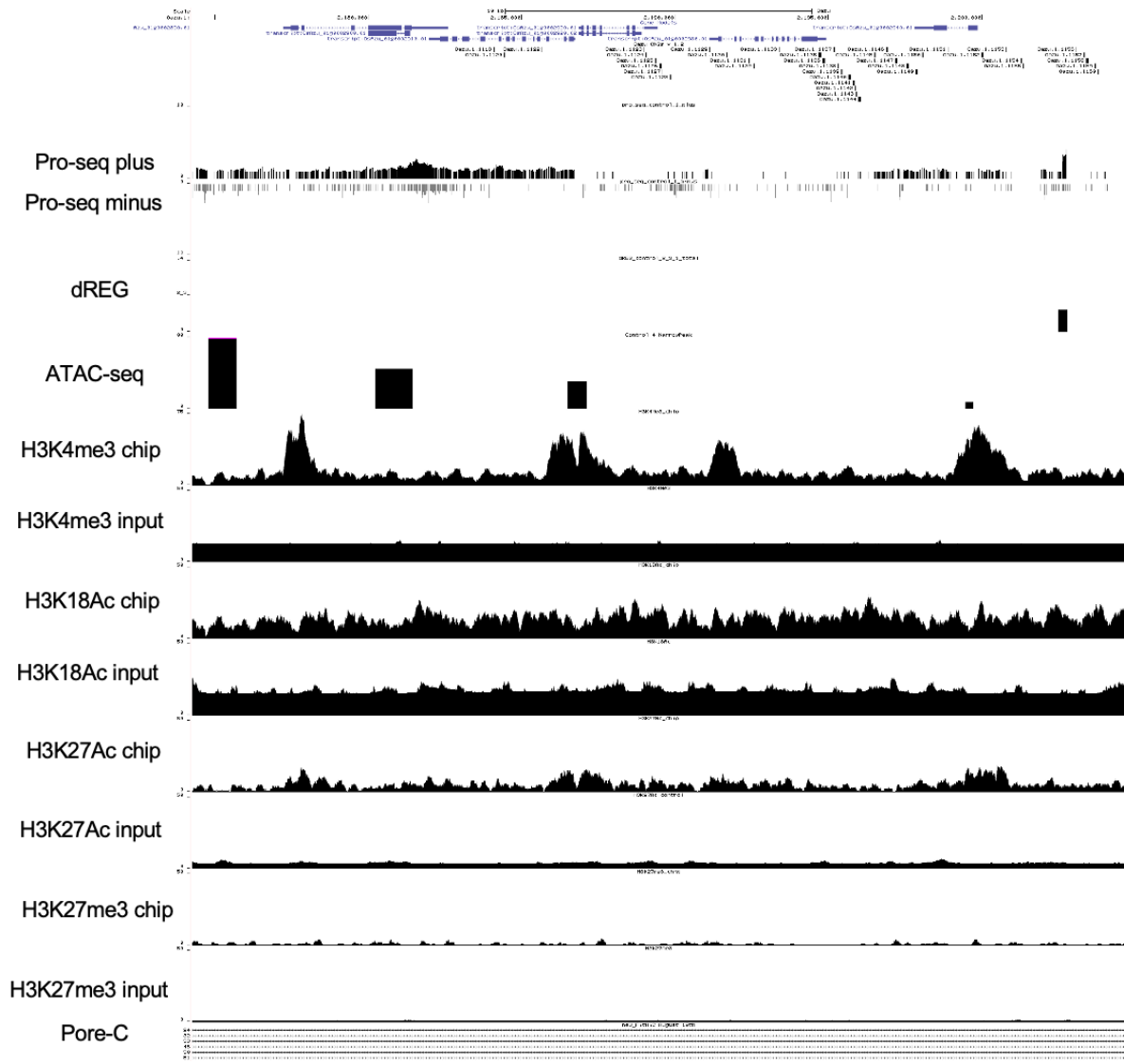


Fig S39. Genome Browser view of the Azucena genome #5. This is showing the different genomic tracks mentioned in this study.