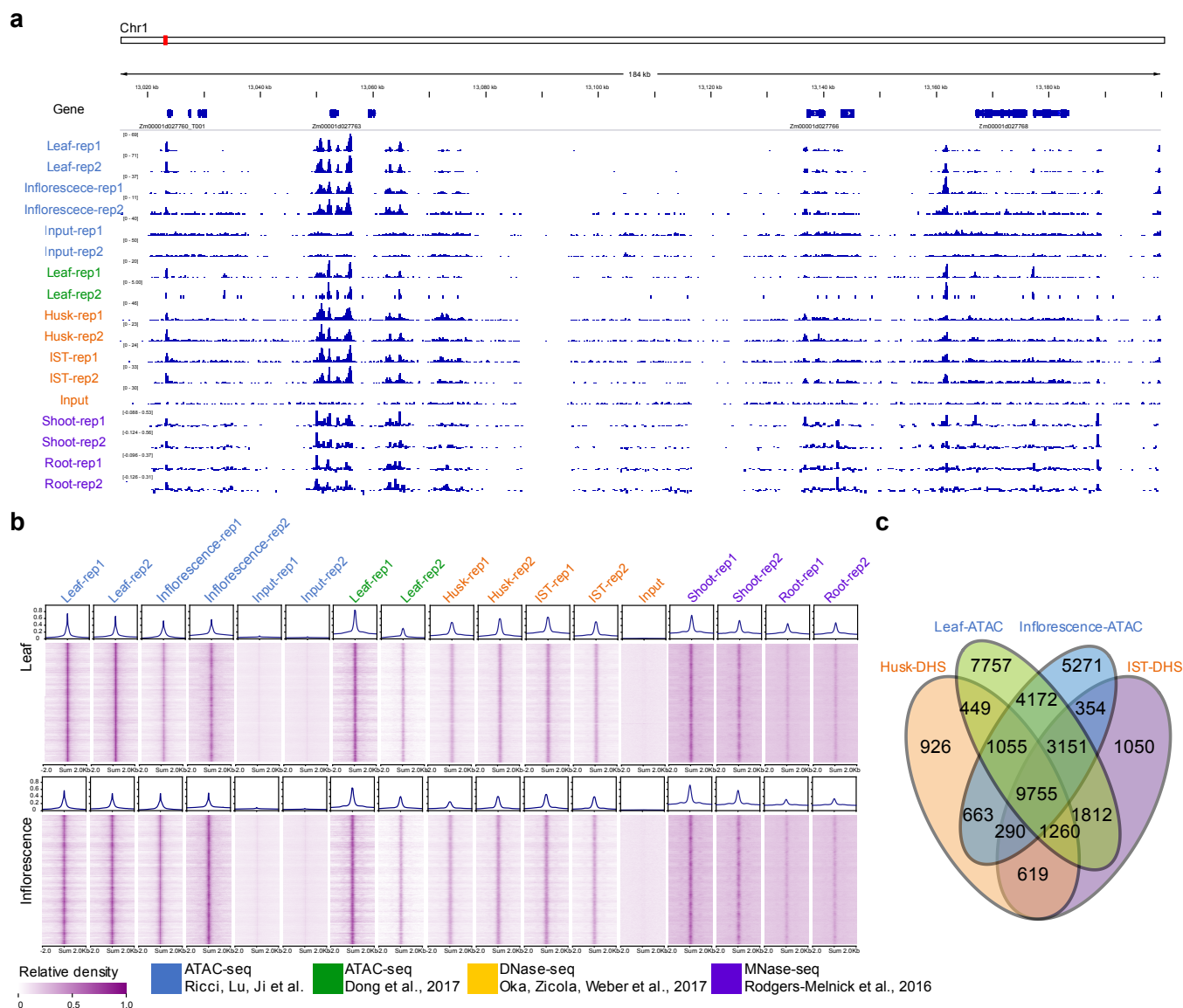


In the format provided by the authors and unedited.

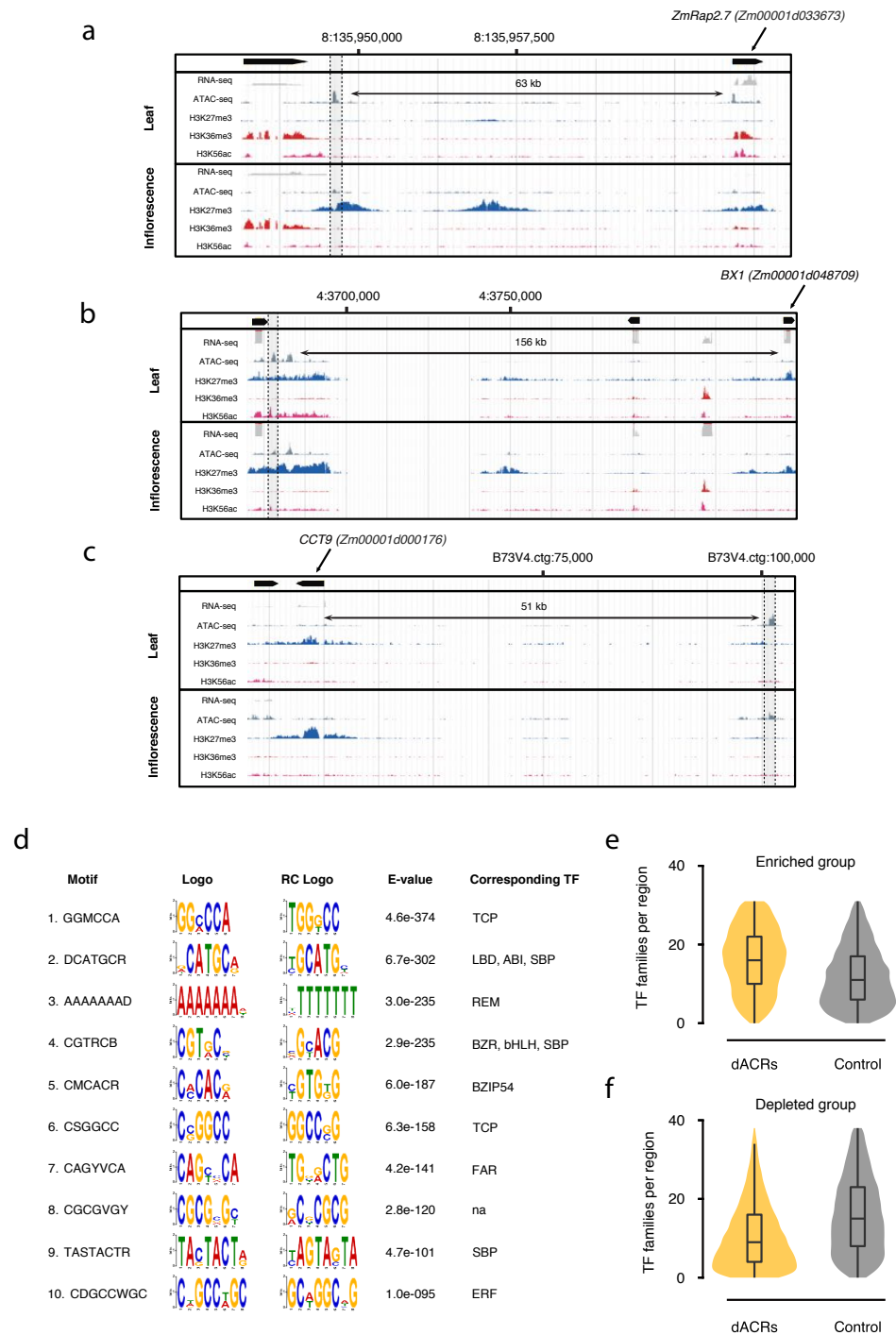
Widespread long-range *cis*-regulatory elements in the maize genome

William A. Ricci^{1,15}, Zefu Lu^{2,15}, Lexiang Ji^{3,15}, Alexandre P. Marand², Christina L. Ethridge², Nathalie G. Murphy², Jaclyn M. Noshay⁴, Mary Galli⁵, María Katherine Mejía-Guerra⁶, Maria Colomé-Tatché^{7,8}, Frank Johannes^{8,9}, M. Jordan Rowley¹⁰, Victor G. Corces¹⁰, Jixian Zhai¹¹, Michael J. Scanlon⁶, Edward S. Buckler^{6,12,13,14}, Andrea Gallavotti⁵, Nathan M. Springer⁴, Robert J. Schmitz^{2,9*} and Xiaoyu Zhang^{1*}

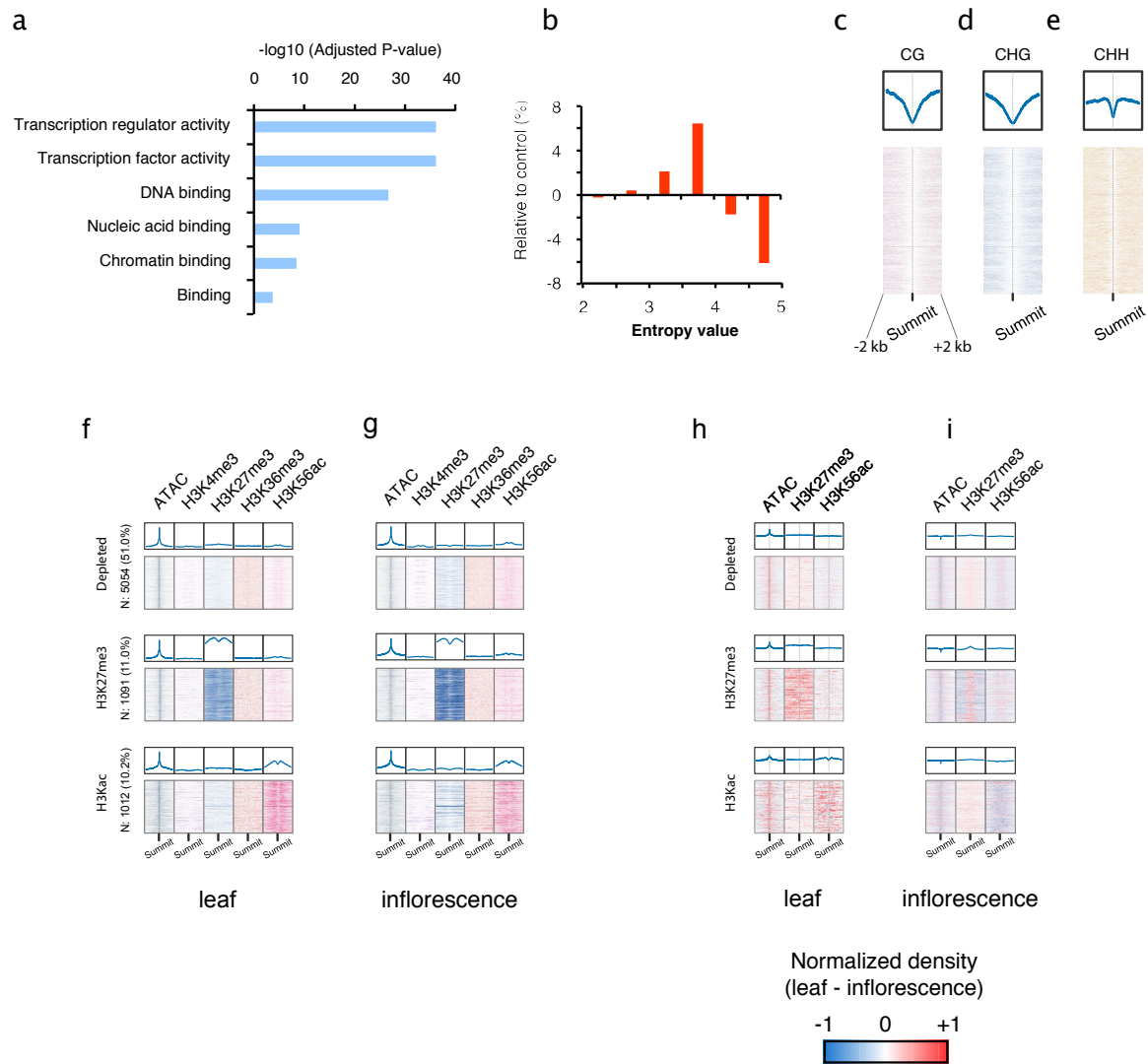
¹Department of Plant Biology, University of Georgia, Athens, GA, USA. ²Department of Genetics, University of Georgia, Athens, GA, USA. ³Institute of Bioinformatics, University of Georgia, Athens, GA, USA. ⁴Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, MN, USA. ⁵Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ, USA. ⁶Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA. ⁷Institute of Computational Biology, Helmholtz Center Munich, German Research Center for Environmental Health, Neuherberg, Germany. ⁸Department of Plant Science, Technical University of Munich, Freising, Germany. ⁹Institute for Advanced Study, Technical University of Munich, Garching, Germany. ¹⁰Department of Biology, Emory University, Atlanta, GA, USA. ¹¹Institute of Plant and Food Science, Department of Biology, Southern University of Science and Technology, Shenzhen, China. ¹²Institute for Genomic Diversity, Cornell University, Ithaca, NY, USA. ¹³US Department of Agriculture–Agricultural Research Service, Robert Holley Center, Ithaca, NY, USA. ¹⁴Department of Plant Breeding and Genetics, Cornell University, Ithaca, NY, USA. ¹⁵These authors contributed equally: William A. Ricci, Zefu Lu, Lexiang Ji. *e-mail: schmitz@uga.edu; xiaoyu@uga.edu



Supplementary Fig. 1 | Comparison between this study's ATAC-seq data and published accessibility data. **a**, Representative region showing the genome coverage of Ricci, Lu, Ji et al.'s ATAC-seq, Dong et al.'s ATAC-seq, Oka, Zicola, Weber et al.'s DNase-seq and Rodgers-Melnick et al.'s MNase-seq. In the MNase-seq tracks, the differential nuclease sensitivity (DNS) was calculated according to Rodgers-Melnick et al.'s methods. **b**, heatmap showing the reads or DNS signal enrichments around leaf and inflorescence ACR summits (the ACRs identified in this manuscript). 2 kb upstream to 2 kb downstream of ACR summits were plotted. **c**, Venn diagram showing the overlap among ACRs identified in this study versus the DNase hypersensitive sites identified by Oka, Zicola, Weber et al.



Supplementary Fig. 2 | Chromatin at genetically mapped CREs and TF binding site sequence enrichment in dACRs. a-c, ATAC-seq, ChIP-seq, and RNA-seq signal at agronomic maize genes and their genetically mapped hypothesized CREs (shaded); (a) *ZmRap2.7*, (b) *BX1*, (c) *ZmCCT9*. ATAC-seq and ChIP-seq experiments in a-c were performed in duplicate and yielded the same results both times. **d**, Motifs enriched within dACRs and corresponding TFs or TF families predicted to bind the motifs. **e-f**, count of TF families corresponding to *Arabidopsis*-derived motifs that are (e) enriched in dACRs and (f) depleted in dACRs. Shown in e and f are the results from 10,433 dACRs. Box plots in e and f comprise medians and quartiles.



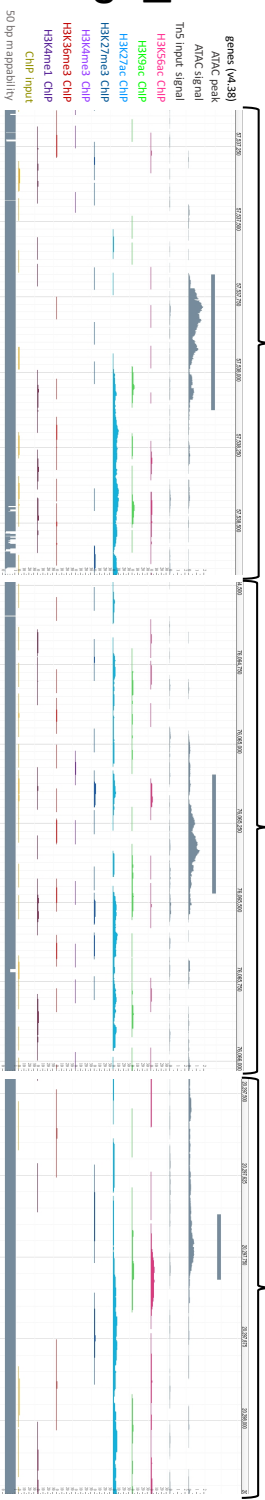
Supplementary Fig. 3 | Characteristics of genes flanking dACRs, DNA methylation depletion within dACRs, and histone modifications at differential dACRs. **a**, Enriched GO terms among genes that flank all dACRs. **b**, Gene expression entropy values of genes flanking all dACRs, relative to genes flanking negative intergenic controls. Control values were subtracted from experimental values. **c-e**, Cytosine methylation of dACRs, aligned by summits. Shown are the three methylation sequence contexts. **f**, ATAC-seq and ChIP-seq signal from leaf at all dACRs identified in leaf, split by the previously described chromatin groups. **g**, ATAC-seq and ChIP-seq signal from inflorescence, plotted at the same loci and sorted in the same order as in **g**. **h**, ATAC-seq and ChIP-seq signal at leaf-specific dACRs with the inflorescence signal subtracted from the leaf signal. Red indicates higher levels in leaf than in inflorescence. **i**, The reciprocal of **h**: ATAC-seq and ChIP-seq signal at inflorescence-specific dACRs with the inflorescence signal subtracted from the leaf signal. Red indicates higher levels in leaf than in inflorescence. ATAC-seq and ChIP-seq experiments shown in this figure were performed in duplicate and yielded the same results both times.

3:57536767..57538674

10:76064125..76066032

1:20297332..20298113

Depleted
group

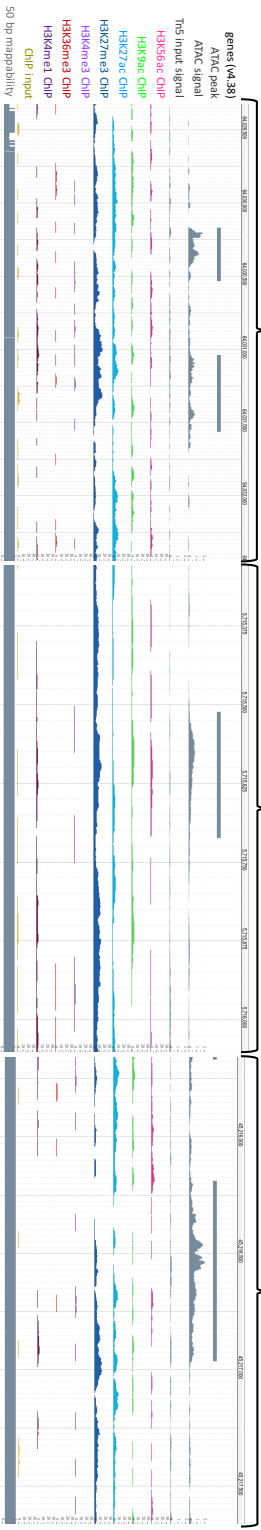


5:64028587..64032446

1:5715099..5716052

2:45215185..45217664

H3K27me3
group

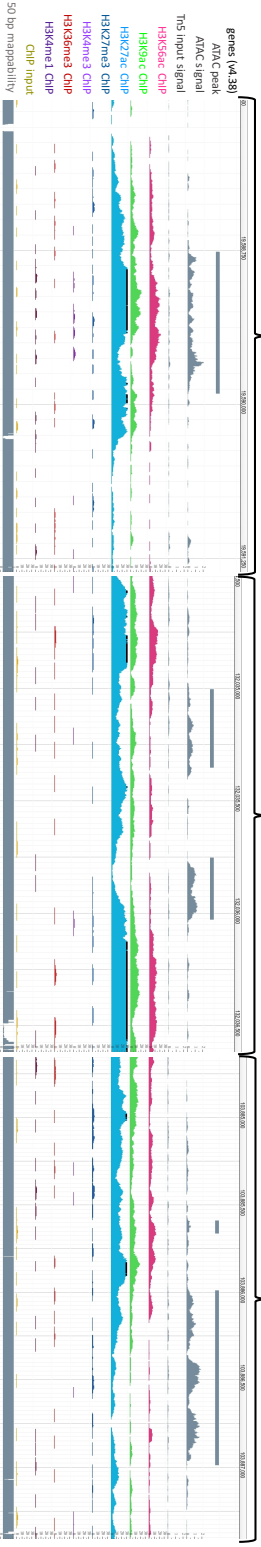


1:19586591..19591360

6:132034000..132036619

8:103884026..103887410

H3Kac
group

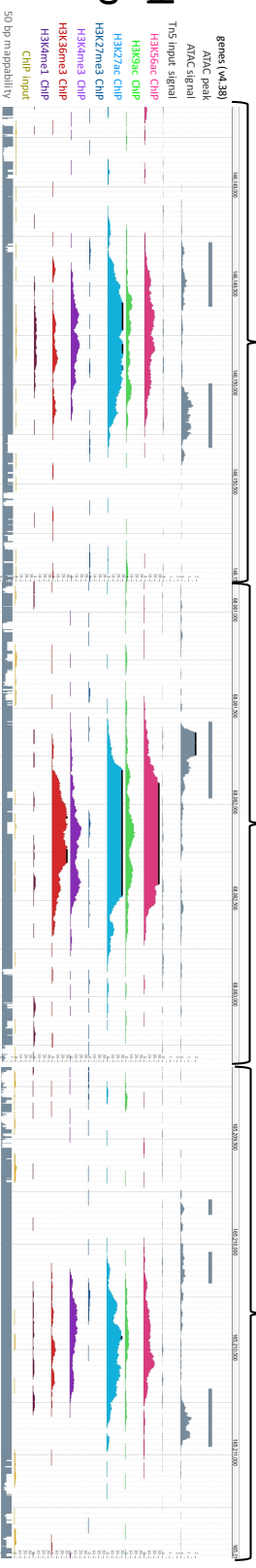


1:146148034..146150995

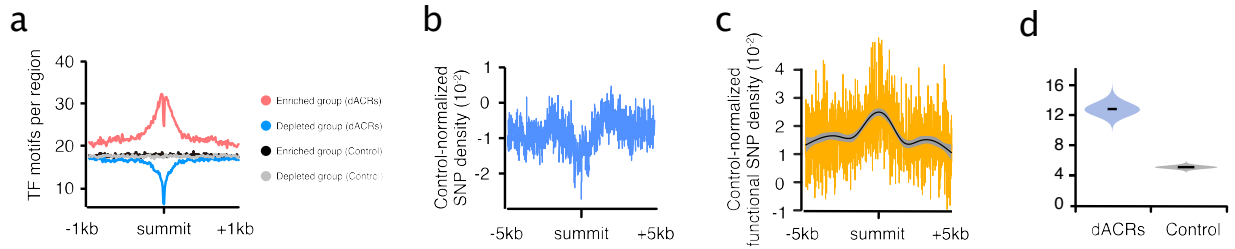
7:68060257..68063341

1:165208574..165211492

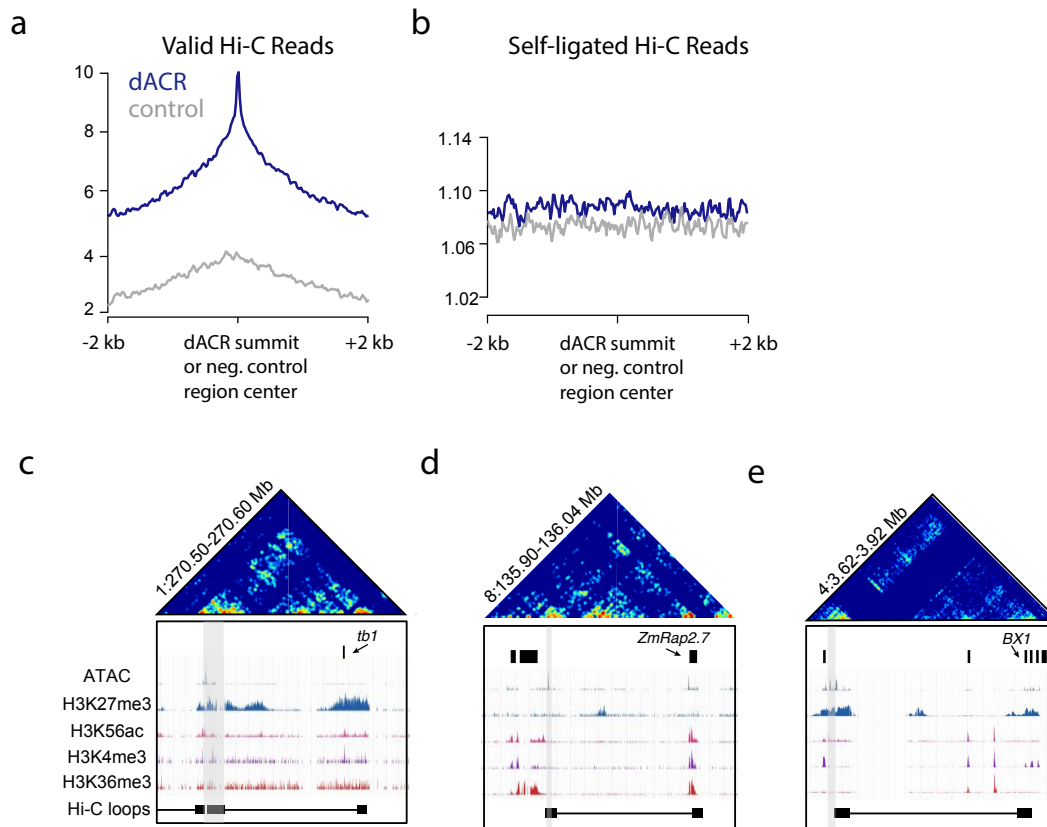
Transcribed
group



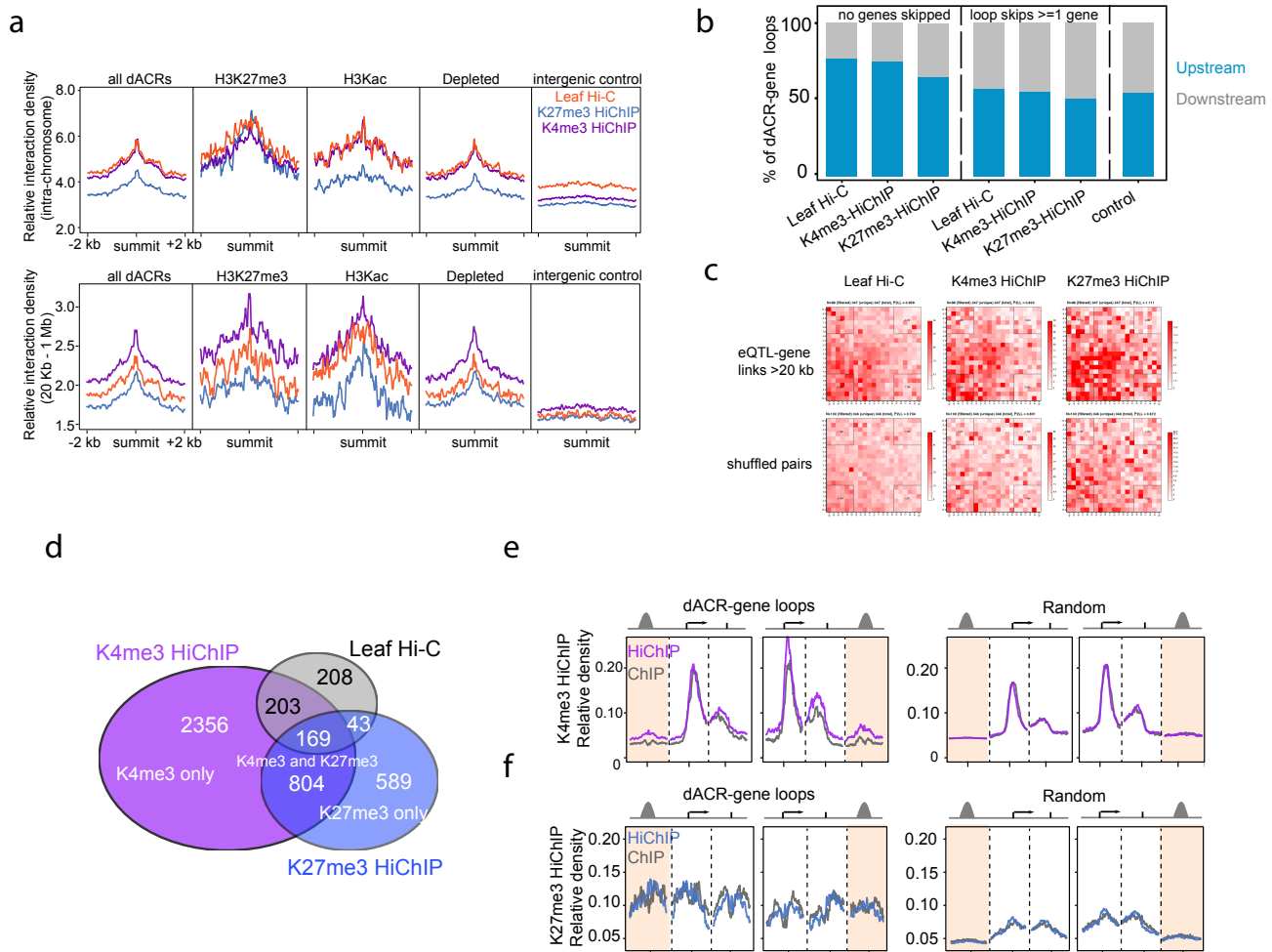
Supplementary Fig. 4 | Randomly selected loci of the dACR chromatin groups. For each of the four dACR groups described in fig. 2b-g, three loci were randomly selected. All y-axes are set to the same scales for ChIP-seq data. The strings above brackets indicate the coordinates in the B73 v4 reference genome. ATAC-seq and ChIP-seq experiments were performed in duplicate and yielded the same results both times.



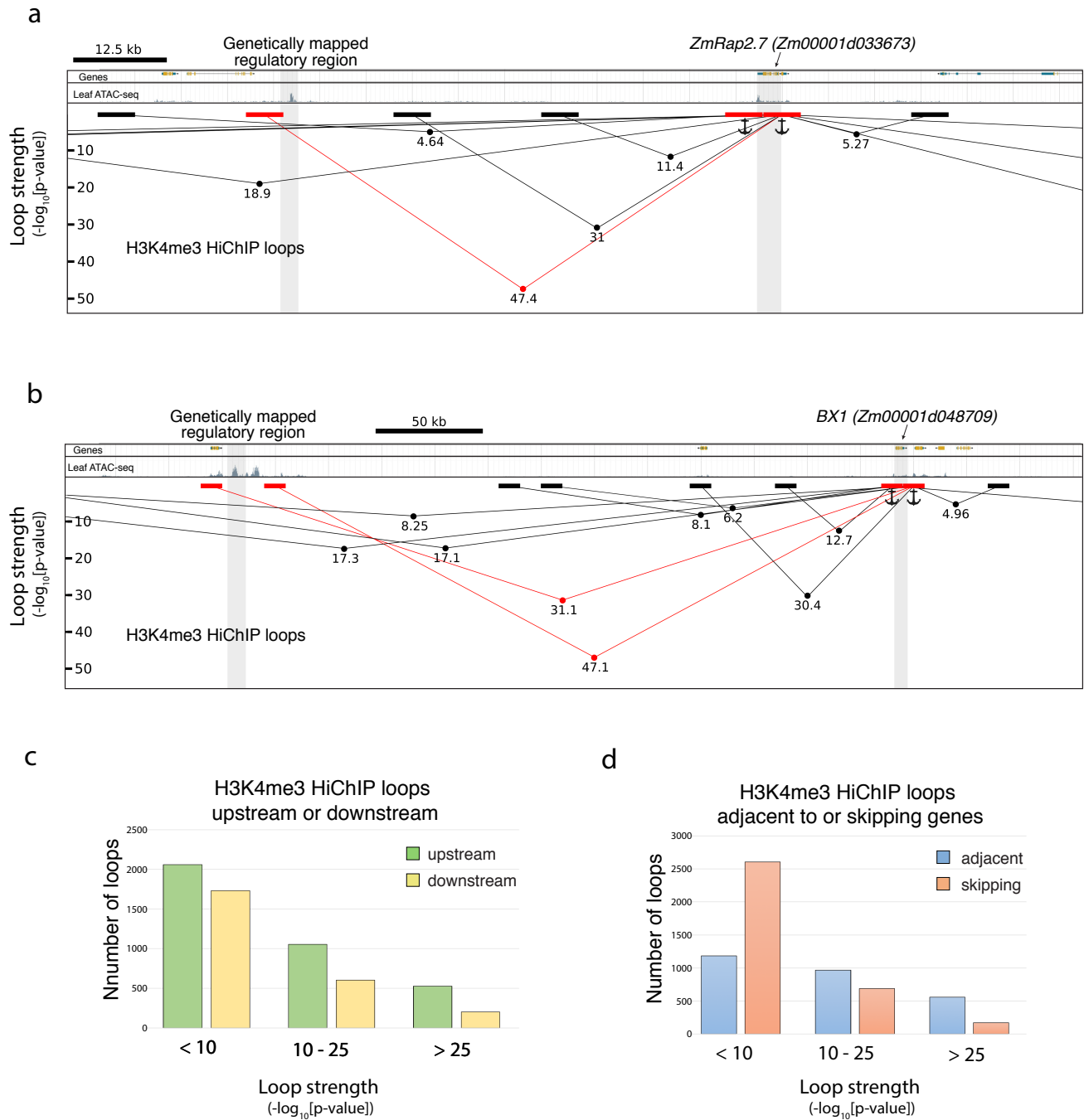
Supplementary Fig. 5 | The same analyses as in fig. 1h-k, but with the transcribed group dACRs omitted. **a**, Distribution of known TF binding motifs from Arabidopsis at dACR summits. **b**, Number of total SNPs among maize inbred lines or **c**, phenotype-associated SNPs per 100 bp bins flanking dACR summits. The negative control distribution was subtracted from the dACR distribution. **d**, Probability that a *cis*-eQTL's highest-significance SNP overlaps a dACR. Y-axis shows posterior probability. The center values correspond to the medians of the distributions. All figures use the same set of negative control regions (i.e. uniquely mapping, intergenic, non-accessible regions). Used for all figures were 10,130 negative control regions and 7,157 non-transcribed-group dACRs.



Supplementary Fig. 6 | Self-ligated Hi-C reads at dACRs and Hi-C loops identified between genes and their genetically mapped hypothesized CREs. a, Reads from valid Hi-C contact pairs (as determined by the Hi-C Pro pipeline) plotted at all dACRs (excluding the transcribed group) and plotted at all intergenic negative control regions. **b,** Reads from the self-ligated contact pairs plotted against the same regions. **c-e,** Hi-C detects interactions between the agronomic genes (c) *tb1*, (d) *ZmRap2.7*, (e) *BX1* and their genetically mapped controlling regions (shaded gray). Called Hi-C loops are shown in the bottom track. Heat maps display contacts at 500 bp resolution. The Hi-C loop passes over the *tb1* gene by a small margin, but H3K27me3 HiChIP loops directly overlap the gene (Fig. 4). The Hi-C experiment was analyzed as a single biological replicate.



Supplementary Fig. 7 | Hi-C and HiChIP interactions at dACRs. **a**, Virtual 4C intrachromosomal interaction signals at dACR summits. Top panels show all intrachromosomal interactions and bottom panels show only interactions between 20 kb and 1 Mb. **b**, The percentages of dACR-gene loops in which the dACR resides either upstream or downstream of the target gene's promoter. dACR-gene pairs were split into those crossing over other genes and those not crossing any other genes. Sample sizes (left to right): 837, 5357, 4051, 449, 5876, 4218, 1000. **c**, Aggregate Peak Analysis (APA) showing the Hi-C/HiChIP interaction signals at the eQTL-gene pairs, compared to randomly permuted, genomic-distance-constrained dACRs and genes. The eQTL-gene pairs show enhanced Hi-C/HiChIP interactions. **d**, The overlap of dACR-gene loops from Hi-C, K4me3 HiChIP and K27me3 HiChIP. **e**, H3K4me3 ChIP signal was enriched at the TSS of expressed genes, but absent from the dACRs that were looped with the genes. **f**, H3K27me3 ChIP signal was enriched at the both the dACR and TSS of dACR-gene K27me3 HiChIP loops. Hi-C and HiChIP experiments were performed as single biological replicates.



Supplementary Fig. 8 | Loop strength at agronomic QTL and loop numbers for various orientations relative to target genes. a, Browser shot of the *ZmRAP2.7* locus and its genetically fine-mapped predicted CRE. **b**, Browser shot of the *BX1* locus and its genetically fine-mapped predicted CRE. **c**, Number of H3K4me3 HiChIP loops with dACRs positioned upstream of target genes' promoters versus downstream, separated by loop statistical significance. **d**, Number of dACR-gene H3K4me3 HiChIP loops which link adjacent dACRs and genes, versus linking dACRs and genes that are separated by one or more genes in between (skipping). All p-values shown in figures were determined in the FitHiChIP program utilizing a two-tailed binomial test.

Data Table S5. Summary statistics of MethylC-seq

No	Sample	Length (nt)	Total reads	Aligned reads*	Non-conversion (%)	Genome coverage
1	B73 leaf	125*2	636,294,670	310,612,031	0.36%	36.38

* Uniquely aligned non-clonal reads

Data Table S6. Summary statistics of RNA-seq

No	Sample	Length (nt)	Total reads	Uniquely aligned	%	Multiple aligned	%
1	B73 leaf rep1	75	46,421,549	37,629,479	81.1%	5,849,616	12.6%
2	B73 leaf rep2	75	45,672,397	35,856,199	78.5%	5,663,861	12.4%
3	B73 inflorescence rep1	150	55,742,012	46,845,242	84.0%	5,640,111	10.1%
4	B73 inflorescence rep2	150	74,438,862	62,322,494	83.7%	7,667,537	10.3%
5	B73 inflorescence rep3	150	72,358,047	57,554,374	79.5%	7,001,809	9.7%
6	B73 inflorescence rep4	150	58,987,597	47,535,124	80.6%	5,639,592	9.6%

Data Table S7. Summary statistics of ATAC-seq, STARR-seq and ChIP-seq

Data type	Sample	Length (nt)	Total reads	Aligned reads	%
ATAC-seq	B73 leaf rep1	35*2	95,964,379	77,472,312	80.73%
ATAC-seq	B73 leaf rep2	35*2	571,321,741	451,685,647	79.06%
ATAC-seq	B73 inflorescence rep1	35*2	76,811,222	56,582,127	73.66%
ATAC-seq	B73 inflorescence rep2	35*2	174,903,747	87,212,919	49.86%
STARR-seq	B73 DNA input	35*2	41,763,217	35,679,215	85.43%
STARR-seq	B73 RNA	35*2	69,363,243	55,613,533	80.18%
ChIP-seq	B73 leaf H2AZ rep1	75	27,732,267	26,486,801	95.51%
ChIP-seq	B73 leaf H2AZ rep2	75	19,252,409	18,399,538	95.57%
ChIP-seq	B73 leaf H3K4me1 rep1	75	48,198,327	46,208,231	95.87%
ChIP-seq	B73 leaf H3K4me1 rep2	75	48,459,950	46,294,226	95.53%
ChIP-seq	B73 leaf H3K4me3 rep1	75	14,706,000	13,658,703	92.88%
ChIP-seq	B73 leaf H3K4me3 rep2	75	14,206,461	11,497,814	80.93%
ChIP-seq	B73 leaf H3K9ac rep1	75	23,017,373	22,278,839	96.79%
ChIP-seq	B73 leaf H3K9ac rep2	75	42,990,103	41,542,119	96.63%
ChIP-seq	B73 leaf H3K27ac rep1	75	50,627,599	49,003,001	96.79%
ChIP-seq	B73 leaf H3K27ac rep2	75	51,044,191	49,430,997	96.84%
ChIP-seq	B73 leaf H3K27me3 rep1	75	24,095,651	23,474,756	97.42%
ChIP-seq	B73 leaf H3K27me3 rep2	75	24,046,522	23,478,832	97.64%
ChIP-seq	B73 leaf H3K36me3 rep1	75	49,208,879	47,247,742	96.01%
ChIP-seq	B73 leaf H3K36me3 rep2	75	75,396,238	72,412,977	96.04%
ChIP-seq	B73 leaf H3K56ac rep1	75	14,631,000	13,604,127	92.98%
ChIP-seq	B73 leaf H3K56ac rep2	75	30,604,762	29,571,854	96.63%
ChIP-seq	B73 leaf H3 rep1	75	46,542,831	43,533,860	93.54%
ChIP-seq	B73 leaf H3 rep2	75	15,137,000	14,418,690	95.25%
ChIP-seq	B73 leaf input rep1	75	22,716,315	21,335,884	93.92%
ChIP-seq	B73 leaf input rep2	75	20,824,166	19,559,465	93.93%
ChIP-seq	B73 inflorescence H3K4me3 rep1	75	42,571,199	41,517,523	97.52%
ChIP-seq	B73 inflorescence H3K4me3 rep2	75	44,288,396	42,644,295	96.29%
ChIP-seq	B73 inflorescence H3K27me3 rep1	75	78,963,021	75,414,392	95.51%
ChIP-seq	B73 inflorescence H3K27me3 rep2	75	72,768,893	70,398,382	96.74%
ChIP-seq	B73 inflorescence H3K36me3 rep1	75	41,183,792	40,282,556	97.81%
ChIP-seq	B73 inflorescence H3K36me3 rep2	75	56,234,785	54,357,330	96.66%
ChIP-seq	B73 inflorescence H3K56ac rep1	75	20,524,329	19,544,866	95.23%
ChIP-seq	B73 inflorescence H3K56ac rep2	75	38,591,965	37,530,383	97.25%
ChIP-seq	B73 inflorescence H3 rep1	75	13,843,701	13,474,764	97.33%
ChIP-seq	B73 inflorescence input rep1	75	33,165,873	32,379,611	97.63%

Data Table S8. Summary statistics of Hi-C

	B73 leaf rep1	B73 leaf rep2	B73 leaf rep3	B73 leaf rep4	B73 leaf rep5	B73 leaf rep6	B73 leaf merged	Paired reads (%)
Total_pairs_processed	91,218,236	112,000,000	118,000,000	173,000,000	106,000,000	94,261,611	694,479,847	/
Unmapped_pairs	491,771	1,900,777	4,645,751	6,207,808	6,691,319	5,078,380	25,015,806	3.60%
Low_qual_pairs (multi-mapping)	62,042,480	76,337,699	79,263,318	118,000,000	67,622,383	61,371,165	464,637,045	66.90%
Unique_paired_alignments	25,618,714	29,800,633	28,106,664	40,094,016	25,855,705	22,968,300	172,444,032	24.83%
Multiple_pairs_alignments	0	0	0	0	0	0	0	/
Pairs_with_singleton	3,065,271	3,838,670	6,021,255	8,932,041	5,574,669	4,843,766	32,275,672	/
Low_qual_singleton	0	0	0	0	0	0	0	/
Unique_singleton_alignments	0	0	0	0	0	0	0	/
Multiple_singleton_alignments	0	0	0	0	0	0	0	/
Reported_pairs	25,618,714	29,800,633	28,106,664	40,094,016	25,855,705	22,968,300	172,444,032	24.83%
Valid_interaction_pairs	22,037,183	23,372,211	21,046,197	30,062,199	16,966,690	15,706,013	129,190,493	18.60%
Valid_interaction_pairs_FF	5,485,723	5,815,606	5,249,394	7,497,467	4,221,396	3,911,314	32,180,900	/
Valid_interaction_pairs_RR	5,481,926	5,817,225	5,241,079	7,493,350	4,220,781	3,908,401	32,162,762	/
Valid_interaction_pairs_RF	5,456,262	5,782,669	5,209,937	7,441,862	4,197,800	3,893,111	31,981,641	/
Valid_interaction_pairs_FR	5,613,272	5,956,711	5,345,787	7,629,520	4,326,713	3,993,187	32,865,190	/
Dangling_end_pairs	3,437,915	6,209,275	6,809,185	9,673,006	8,606,742	7,024,555	41,760,678	6.01%
Religation_pairs	114,893	128,287	101,862	132,755	104,326	82,793	664,916	0.10%
Self_Cycle_pairs	28,302	90,282	148,646	225,167	177,199	154,441	824,037	0.12%
Single-end_pairs	0	0	0	0	0	0	0	0.00%
Dumped_pairs	421	578	774	889	748	498	3,908	0.00%
valid_interaction	22,037,183	23,372,211	21,046,197	30,062,199	16,966,690	15,706,013	129,190,493	18.60%
valid_interaction_rmdup	21,469,859	22,706,721	19,911,029	28,006,068	16,503,930	15,274,208	123,871,815	17.84%
trans_interaction	9,035,156	9,646,159	6,236,837	9,411,588	5,297,148	5,191,594	44,818,482	6.45%
cis_interaction	12,434,703	13,060,562	13,674,192	18,594,480	11,206,782	10,082,614	79,053,333	11.38%
cis_shortRange (<20k)	1,976,020	2,129,440	2,138,604	2,881,181	1,649,003	1,424,887	12,199,135	1.76%
cis_longRange (>20k)	10,458,683	10,931,122	11,535,588	15,713,299	9,557,779	8,657,727	66,854,198	9.63%

Data Table S9. Summary statistics of HiChIP

	H3K4me3	Paired reads (%)	H3K27me3	Paired reads (%)
Total_pairs_processed	326,790,696	/	373,824,484	/
Unmapped_pairs	6,891,622	2.11%	7,327,488	1.96%
Low_qual_pairs (multi-mapping)	203,735,261	62.34%	246,489,441	65.94%
Unique_paired_alignments	104,559,206	32.00%	107,187,533	28.67%
Multiple_pairs_alignments	0	/	0	/
Pairs_with_singleton	11,604,607	/	12,820,022	/
Low_qual_singleton	0	/	0	/
Unique_singleton_alignments	0	/	0	/
Multiple_singleton_alignments	0	/	0	/
Reported_pairs	104,559,206	32.00%	107,187,533	28.67%
Valid_interaction_pairs	85,952,971	26.30%	89,774,346	24.02%
Valid_interaction_pairs_FF	21,376,712	/	22,309,423	/
Valid_interaction_pairs_RR	21,374,430	/	22,295,825	/
Valid_interaction_pairs_RF	21,081,403	/	22,037,893	/
Valid_interaction_pairs_FR	22,120,426	/	23,131,205	/
Dangling_end_pairs	16,067,350	4.92%	15,042,473	4.02%
Religation_pairs	686,585	0.21%	747,142	0.20%
Self_Cycle_pairs	1,849,735	0.57%	1,620,087	0.43%
Single-end_pairs	0	0.00%	0	0.00%
Dumped_pairs	2,565	0.00%	3,485	0.00%
valid_interaction	85,952,971	26.30%	89,774,346	24.02%
valid_interaction_rmdup	71,724,261	21.95%	52,485,887	14.04%
trans_interaction	19,351,764	5.92%	13,984,287	3.74%
cis_interaction	52,372,497	16.03%	38,501,600	10.30%
cis_shortRange (<20k)	13,515,142	4.14%	9,972,259	2.67%
cis_longRange (>20k)	38,857,355	11.89%	28,529,341	7.63%

Data Table S10. Summary statistics of DAP-seq

No	Sample	Length (nt)	Total read*	Aligned reads	%	MAPQ > 30	%
1	BAD1	75	5,417,421	5,385,760	99.42%	2,925,671	54.32%
2	BZIP25	75	13,642,865	13,582,792	99.56%	7,363,265	54.21%
3	BZIP54	75	13,196,410	13,139,452	99.57%	7,206,373	54.85%
4	BZIP57_FEA4	75	25,985,097	25,818,283	99.36%	13,864,652	53.70%
5	BZIP72	75	10,895,725	10,852,766	99.61%	5,894,672	54.31%
6	BZIP96	75	10,330,472	10,273,254	99.45%	5,551,721	54.04%
7	EREB127	75	4,357,850	4,316,436	99.05%	2,378,062	55.09%
8	EREB138	75	10,362,804	10,288,525	99.28%	6,421,780	62.42%
9	EREB24_BBM	75	28,297,699	28,129,669	99.41%	14,489,873	51.51%
10	EREB29	75	8,757,185	8,698,027	99.32%	5,334,257	61.33%
11	EREB71-rep1	75	11,787,059	11,676,857	99.07%	6,280,947	53.79%
12	LBD16_RA2	75	27,975,475	27,779,002	99.30%	15,681,385	56.45%
13	LBD19_IG1	75	16,700,403	16,589,953	99.34%	10,886,229	65.62%
14	LBD38	75	14,255,281	14,177,747	99.46%	7,684,190	54.20%
15	LBD5	75	16,642,239	16,541,913	99.40%	8,937,884	54.03%
16	SBP30_UB3	75	10,169,207	10,091,075	99.23%	5,924,112	58.71%
17	SBP6	75	10,682,140	10,603,571	99.26%	6,052,893	57.08%
18	SBP8_UB2	75	10,690,459	10,608,533	99.23%	6,298,605	59.37%

* After quality trimming with trimmomatic

Data Table S11. Antibodies used for ChIP and HiChIP

epitope	source	cat #	used in
H2A.Z	in-house		ChIP
H3	Abcam	Cat# ab1791	ChIP
H3K4me1	Abcam	Cat# ab8895	ChIP
H3K4me3	Millipore	Cat# 07-473	ChIP, HiChIP
H3K9me2	Cell Signaling	Cat# 9753L	ChIP
H3K9ac	Active Motif	Cat# 61251	ChIP
H3K23ac	Millipore	Cat# 07-355	ChIP
H3K27me3	Millipore	Cat# 07-449	ChIP, HiChIP
H3K27ac	Abcam	Cat# ab4729	ChIP
H3K36me3	Abcam	Cat# ab9050	ChIP
H3K56ac	Millipore	Cat# 07-677-1	ChIP
RNA polII	Abcam	Cat# ab26721	ChIP

Data Table S12. Coordinates of regions omitted from STARR-seq analyses

2:19452127..19589055
8:135887933..136051059
4:3582270..3781476
1:270433990..270585643
8:67224311..67419905
6:120633193..120880806

Table s12. The STARR-seq sequencing data showed some cross-contamination with six bacterial artificial chromosomes derived from the B73 v4.0 reference genome. The six regions listed contained the contamination and were omitted from STARR-seq analyses.