nature plants

ARTICLES

https://doi.org/10.1038/s41477-019-0547-0

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. Marand[2], Christina L. Ethridge[2], Nathalie G. Murphy[2], Jaclyn M. Noshay[4], Mary Galli[5], María Katherine Mejía-Guerra[6], Maria Colomé-Tatché[7,8], Frank Johannes[8,9], M. Jordan Rowley[10], Victor G. Corces[10], Jixian Zhai[11], Michael J. Scanlon[6], Edward S. Buckler[6,12,13,14], Andrea Gallavotti[5], Nathan M. Springer[4], Robert J. Schmitz[2,9*] and Xiaoyu Zhang[1*]

Genetic mapping studies on crops suggest that agronomic traits can be controlled by gene–distal intergenic loci. Despite the biological importance and the potential agronomic utility of these loci, they remain virtually uncharacterized in all crop species to date. Here, we provide genetic, epigenomic and functional molecular evidence to support the widespread existence of gene–distal (hereafter, distal) loci that act as long-range transcriptional *cis*-regulatory elements (CREs) in the maize genome. Such loci are enriched for euchromatic features that suggest their regulatory functions. Chromatin loops link together putative CREs with genes and recapitulate genetic interactions. Putative CREs also display elevated transcriptional enhancer activities, as measured by self-transcribing active regulatory region sequencing. These results provide functional support for the widespread existence of CREs that act over large genomic distances to control gene expression.

The long-range transcriptional control of genes by distal *cis*-regulatory elements (CREs) is an important and well-studied feature of metazoan genomes[1]. In contrast, many fundamental questions regarding distal CREs in plants—such as their prevalence, sequence and chromatin attributes, transcriptional regulatory behaviours and mechanisms of action—remain unanswered[2,3]. In maize, agronomic quantitative trait loci (QTL) have been mapped to the intergenic space[4] and a handful of domestication loci that were hypothesized to contain CREs have been fine-mapped to distal regions[5–8]. Genetic evidence demonstrated that these fine-mapped loci controlled their target genes in *cis*. However, currently lacking are molecular characterizations of these loci and demonstrations of direct chromatin interactions between the hypothesized CREs and their target genes.

It has been widely observed that actively engaged CREs reside within accessible chromatin[9]. This is partially due to the interactions between transcription factors (TFs) and DNA, which often disturb nucleosome stability and elevate chromatin accessibility[9,10]. Nucleosomes surrounding accessible chromatin regions (ACRs) often exhibit histone modifications indicative of the transcriptional coregulators that have been recruited to the ACRs. Accordingly, flanking histone modifications provide insight into the regulatory mechanisms of the CREs contained within ACRs. Given that ACRs are enriched at intergenic QTL in the maize genome[11], we decided to take an ACR-centric approach to identify actively engaged CREs within the gene–distal intergenic space. Here, we combined assay for transposase-accessible chromatin sequencing (ATAC-seq) with multiple chromatin assays to demonstrate that distal CRE are abundant in the maize genome.

Results

Gene–dACRs are common in the maize genome. We first profiled chromatin accessibility in young *Zea mays* L., cultivar B73 leaves using ATAC-seq[12,13]. We identified 32,111 ACRs (Fig. 1a,b and Supplementary Table 1), which ranged mostly from 300 to 1,000 base pairs (bp) in length (Fig. 1c) and occupied ~1% of the maize genome. Multiple chromatin accessibility datasets from comparable maize tissues were publicly available[11,14–16], allowing us to compare independent datasets that employed different enzymatic assays (Tn5 (ref. [16]), DNase[14,15] and MNase[11]). Chromatin accessibility signals from the independent experiments were enriched at the ACRs identified in this manuscript (Supplementary Fig. 1a,b). These ACRs recapitulated 88% (18,789/21,384) of the accessible regions identified via DNase treatment[14] (Supplementary Fig. 1c). These results indicated that systematic biases deriving from the Tn5 enzyme were negligible within our experimental context.

We split ACRs based on proximity to their nearest annotated genes (Fig. 1b). We found 12,495 (38.9%) of the ACRs overlapped genes (gACRs, defined as overlapping ≥1 bp with annotated genes) and 9,183 (28.6%) were within 2 kilobases (kb) of genes (pACRs, defined as overlapping ≥1 bp with the 2 kb regions flanking genes, but not overlapping the genes themselves). We also found 10,433

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

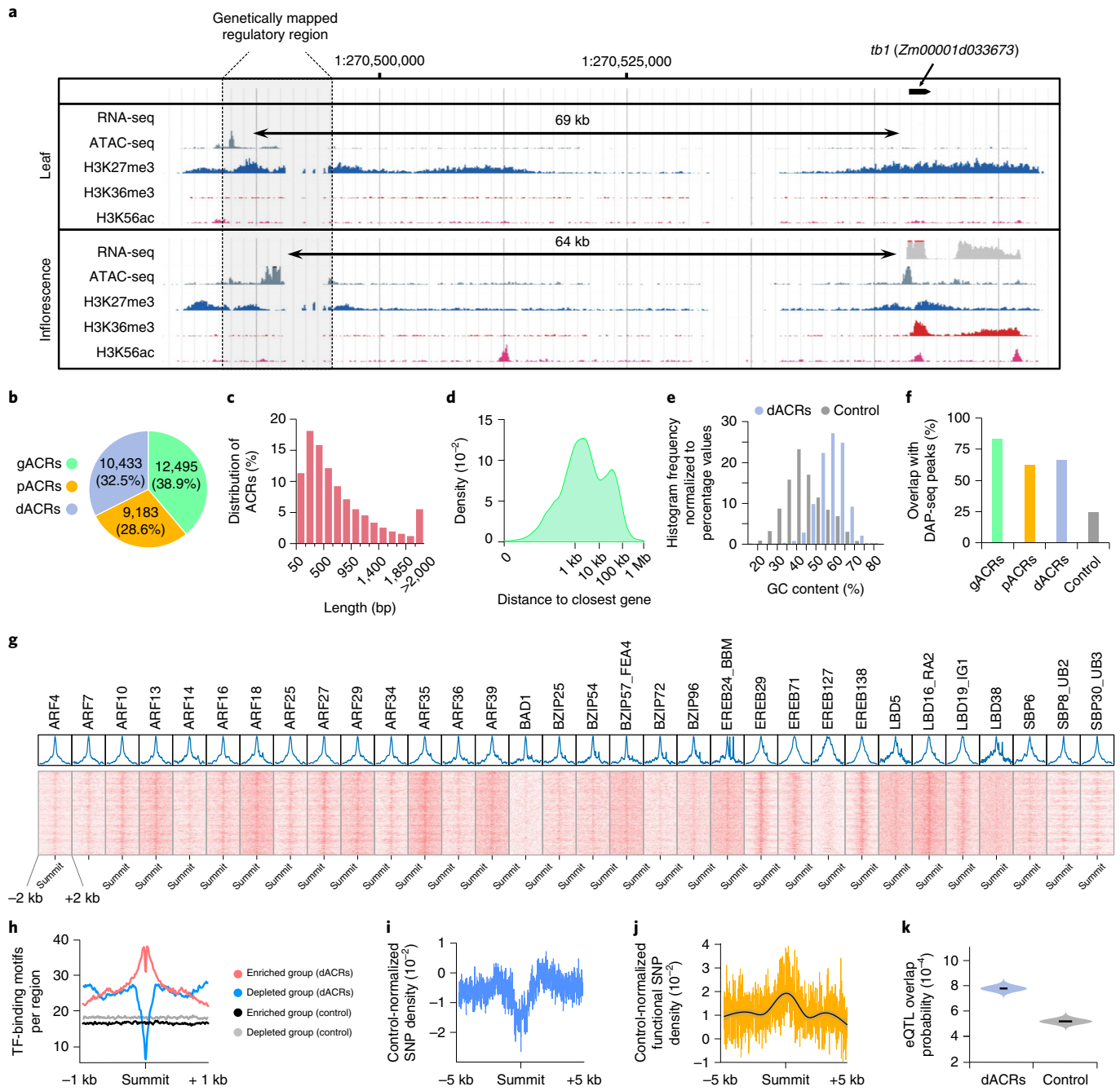


Fig. 1 | ACRs in the maize genome. a, *tb1* is expressed in immature inflorescences and silenced in leaves. The genetically mapped *tb1* CRE (grey shaded area demarcated with dotted lines) displays tissue-dynamic chromatin accessibility and histone modifications. ATAC-seq and chromatin immunoprecipitation sequencing (ChIP-seq) experiments were performed in duplicate and both yielded the same results. **b**, Genome-wide distribution of leaf ATAC-seq peaks in relation to the AGPv4.38 annotated genes. Genic ACRs (gACRs) overlap genes; proximal ACRs (pACRs) fall within 2,000 bp of genes and dACRs are >2,000 bp from genes. **c**, Lengths of total ATAC-seq peaks (base pairs). **d**, Distances of ATAC-seq peaks (excluding gACRs) from the closest annotated gene. **e**, Guanine–cytosine (GC) content at each dACR versus gene-distal uniquely mapping negative control regions. **f**, Percentage of ACR class that overlaps ≥1 DNA affinity purification sequencing (DAP-seq) TF peaks. **g**, Meta-analysis of DAP-seq peak signals for individual TFs at dACR summits. No replicates of this analysis were performed. **h**, Distribution of *Arabidopsis*-derived TF-binding motifs at dACR summits. **i**,**j**, Total single nucleotide polymorphisms (SNPs) among maize inbred lines (**i**) or phenotype-associated SNPs per 10 bp bins flanking dACR summits (**j**). For normalization of **i** and **j**, the negative control distribution was subtracted from the dACR distribution and the difference was plotted. **k**, Probability that the highest signficance SNP of a *cis*-expression QTL (eQTL) overlaps a dACR. The *y* axis shows posterior probability. The centre values correspond to the medians of the distributions. The same set of negative control regions (that is, uniquely mapping, intergenic, nonaccessible regions) are used in panels **e** and **f**.

ACRs (32.5%) occurred >2 kb from their nearest genes (distal ACRs (dACRs)) and 4,091 dACRs exceeded 20 kb from their nearest genes (Fig. 1d). Hypothesized long-range CREs that were previously identified by genetic mapping studies, such as those controlling *tb1* (ref. [7]), *ZmRap2.7* (ref. [6]), *BX1* (ref. [8]) and *ZmCCT9* (ref. [5]), were apparent in the ATAC-seq data (Fig. 1a and Supplementary Fig. 2a–c).

Gene–dACRs probably contain *cis*-regulatory elements. The elevated accessibility at dACRs could be caused by active mechanisms, such as the binding of nucleosome-displacing TFs or chromatin remodellers[9], or by inactive mechanisms, such as the presence of DNA sequences recalcitrant to nucleosome assembly[17]. Our data suggested active mechanisms of dACR formation. The sequence content within dACRs was approximately 15% more GC-rich (better suited for nucleosome formation[17]) than for negative control regions ('control': randomly selected, uniquely mapping, non-ACR intergenic regions; Fig. 1e). Furthermore, dACRs were enriched for TF binding sites, which we identified empirically (using DAP-seq[18,19] for 32 maize TFs) and computationally (using known TF binding motifs from *Arabidopsis thaliana* and de novo motif enrichment). pACRs and dACRs showed similar rates of DAP-seq peak overlap (Fig. 1f) and all 32 DAP-seq TFs were enriched at dACRs (Fig. 1g). Individual dACRs were predicted to contain multiple TF binding sites which corresponded to TFs from multiple families (Fig. 1h and Supplementary Fig. 2d–f).

Several lines of evidence suggested that many dACRs were functionally important and potentially enriched with CREs. First, DNA sequence diversity was markedly reduced at dACRs (Fig. 1i). Second, sequence variation within dACRs was more likely to be associated with phenotypic variation (Fig. 1j) and gene expression variation (Fig. 1k), as determined by genome-wide association data[4,20]. Third, the nearest genes flanking dACRs were enriched for transcriptional regulatory functions and were tissue-specifically expressed (Supplementary Fig. 3a,b).

Gene–dACRs fall into chromatin classes suggestive of their regulatory functions. In mammalian genomes, transcriptional enhancers are associated with specific histone modifications (for example, H3K4me1, H3K27ac and H3K27me3)[21,22]. To determine if a typical chromatin signature existed for maize dACRs, we mapped DNA methylation (mCG, mCHG and mCHH, where 'H' indicates A, C or T, respectively) and histone covalent modifications (H3K4me1, H3K4me3, H3K27me3, H3K36me3, H3K9ac, H3K27ac, H3K56ac and the histone variant H2A.Z) in maize leaves using MethylC-seq and ChIP–seq, respectively. The genic patterns of chromatin accessibility, histone modifications and DNA methylation were similar to those previously described in other plants[11,14,23–29] (Fig. 2a). DNA cytosine methylation in all sequence contexts was markedly reduced at dACRs (Supplementary Fig. 3c–e). In contrast to H3K4me1 found at mammalian enhancers[22], no histone covalent modifications in this study were common to the majority of maize dACRs, although nearly all dACRs were enriched for flanking nucleosomes containing the histone variant H2A.Z.

K-means clustering of dACRs by their flanking histone modifications resolved four main groups (Fig. 2b–g and Supplementary Table 1). The majority of dACRs (51.2%) were depleted of flanking histone modifications ('depleted group'; Fig. 2b and Supplementary Figs. 2c and 4). The histones flanking the depleted group dACRs were either lacking modifications or modified at low levels. We found 11.1% of dACRs contained primarily H3K27me3 at flanking histones ('H3K27me3 group'; Figs. 1a, 2c and Supplementary Fig. 4). Similarly to the depleted group dACRs, other histone modifications were sometimes present at low levels, but H3K27me3 was the predominant modification. We also found 10.2% of dACRs were flanked by strong H3K9/K27/K56 acetylation and lacked other histone covalent modifications ('H3Kac group'; Fig. 2d and Supplementary Fig. 4). Additionally, 27.5% of dACRs were flanked by multiple histone modifications typically found together at transcribed genes, including H3K4me1, H3K4me3, H3K36me3 and H3K9/K27/K56ac ('transcribed group', Fig. 2e,f and Supplementary Fig. 4). The assortment and strong directionality of histone modifications at the transcribed group dACRs resembled the chromatin at transcribed genes (Fig. 2a). Furthermore, abundant transcripts colocalized with the histone modifications of the transcribed group dACRs (Fig. 2e,f).

The genes closest to the depleted, H3K27me3 and H3Kac group dACRs were enriched for developmental and transcriptional regulators that were expressed with high tissue specificity (Fig. 2h,i). The genes closest to H3K27me3 group dACRs were transcriptionally repressed, whereas the genes closest to the H3Kac and depleted group dACRs were expressed at low-to-moderate levels (Fig. 2j). In contrast, genes surrounding the transcribed group dACRs lacked significant functional enrichment or expression specificity. Due to the transcribed group's resemblance to genes, we omitted the transcribed group dACRs from subsequent analyses. The omission of this group did not alter the functional enrichment results from Fig. 1 (Supplementary Fig. 5).

We sought to determine if tissue-specific changes in dACR accessibilities correlated with changes in local histone modifications or the expression of nearby genes. We compared ATAC-seq, ChIP–seq and gene expression profiles between leaves and immature inflorescences. Evaluating ChIP–seq signals from both tissues at identical loci revealed that most dACRs (identified in leaf) retained accessibility and the same histone modifications in the second tissue (inflorescences) (Supplementary Fig. 3f,g). However, 15–21% of dACRs that were present in leaves were inaccessible in inflorescences (Fig. 2k and Supplementary Table 2). Tissue-specific dACRs that lost accessibility in one tissue also lost their flanking histone acetylation in the same tissue (Supplementary Fig. 3h,i). This association suggested that the factors responsible for acetylating the flanking histones could be causally linked to chromatin accessibility. In contrast, the relationship between accessibility and H3K27me3 was less clear and potentially decoupled. Tissue-specific dACRs also exhibited relationships with nearby genes. The closest genes to leaf-specific dACRs were more often differentially expressed between leaves and inflorescences (Fig. 2l). This did not hold true for the genes that were buffered from the dACRs by intervening genes. Furthermore, leaf-specific dACRs were more often located upstream, rather than downstream, of differentially expressed genes (Fig. 2m).

Chromatin loops connect gene–dACRs with genes. The locations of dACRs raised the question of how they might regulate target genes over large intergenic distances. To determine if dACRs interacted directly with their target genes through the formation of chromatin loops, we first performed Hi-C[30] on young maize leaves. We focused on the characterization of chromatin loops involving dACRs and genes (Supplementary Tables 3 and 4). Due to technical constraints, we did not search for chromatin loops <20 kb in length; therefore, this was not an exhaustive characterization of all dACR–gene loops. However, 39.2% of dACRs—a sufficiently representative sample of the dACR population—were >20 kb from their nearest genes (Fig. 1d). Although dACRs comprised <0.2% of the intergenic space, more than half (614/1,177) of the identified intergenic–gene loops contained at least one dACR at their intergenic edges (Fig. 3b). Analysis of the Hi-C reads from self-ligated contact pairs demonstrated that the loop enrichment at dACRs was not an artefact arising from chromatin accessibility or mapping biases (Supplementary Fig. 6a,b). Therefore, dACR–gene loops spanning ≥20 kb were a common feature in the maize genome. These loops included interactions between the target genes *tb1*, *ZmRap2.7* and *BX1* and their genetically mapped controlling regions that have been hypothesized to contain long-range CREs (Fig. 3a and Supplementary Fig. 6c–e).

Although the Hi-C results provided evidence for dACR–gene interactions, relatively few chromatin loops were identified due to limited sequencing depth. Furthermore, because the Hi-C experiment was performed on whole leaves (which contained a diversity of cell types) it was not clear whether dACR–gene loops were formed in cells where the genes were expressed, silenced or both. To address these challenges, we performed Hi-C followed by ChIP (HiChIP)[31]

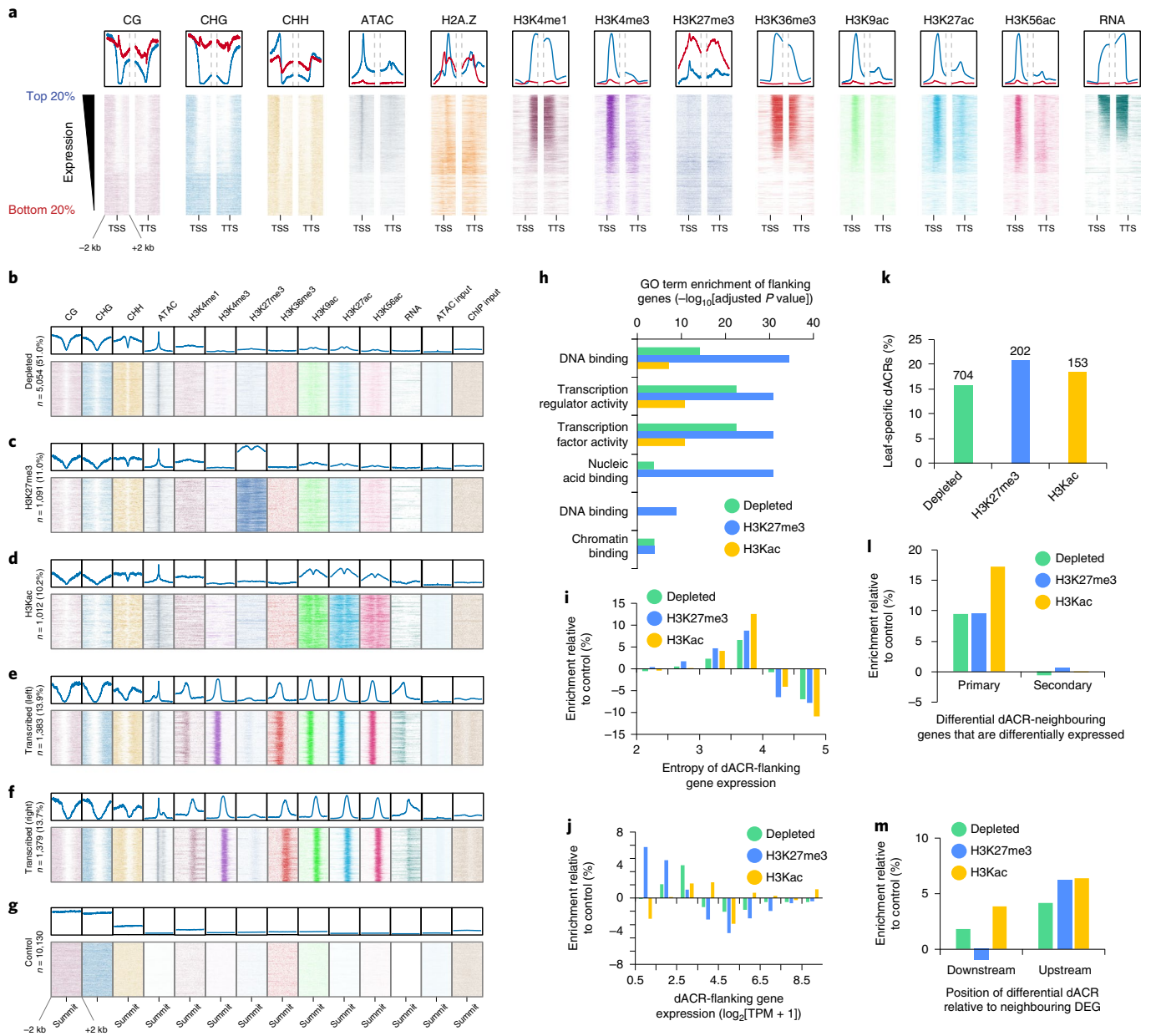


Fig. 2 | Chromatin attributes of dACRs and patterns among dACR-flanking genes. a, Meta-analysis of DNA methylation, ATAC-seq, ChIP–seq and RNA-seq signals at transcription start sites (TSS) and termination sites (TTS) of annotated genes, ranked by expression. Included are 2 kb upstream and downstream of TSS and TTS. Note that the bottom one-third of ranked genes probably correspond to pseudogenes. **b–g**, Chromatin attributes at dACRs, aligned at dACR summits and clustered into four groups: depleted group, H3K27me3 group, H3Kac group and transcribed group. Shown are ±2 kb from ATAC-seq peak summits. ChIP–seq and RNA-seq experiments for **a**–**g** were performed in duplicate and yielded identical results each time. **h**, Gene ontology (GO) term enrichment for the nearest genes flanking the dACRs on both sides. *P* values were determined with a two-sided hypergeometric test using the BiNGO program (see Methods) and adjusted for multiple testing using the Benjamini–Hochberg method. Sample sizes were twice the number of dACRs in each chromatin group because each dACR had two flanking genes. **i**,**j**, Expression Shannon entropy values (**i**) and expression levels (transcripts per million) (**j**) of the nearest genes on both sides of each dACR. **k**, Percentage of total leaf dACRs in each chromatin group present in leaves but absent from inflorescences (that is, the leaf dACR does not overlap an inflorescence dACR). The numbers on top of the bars indicate the total number of leaf-specific dACRs found in each of the categories. **l**, Percentage of first neighbour (primary) and second neighbour (secondary) genes that are differentially expressed among the genes flanking leaf-specific differential dACRs. **m**, Percentage of differentially expressed genes for which the differential dACR occurs downstream or upstream of the 5′ end of the gene. All figures use the same set of negative control regions. For **i**,**j**,**l**,**m**, percentages from genes flanking intergenic negative control regions were subtracted from the percentages of genes flanking dACRs. DEG, differentially expressed gene.

using antibodies targeting histone modifications associated with transcriptional activation (H3K4me3) and silencing (H3K27me3), but largely absent from heterochromatin[25,27,29] (Supplementary Tables 3 and 4). Similar to the Hi-C loops, the intergenic edges of both H3K4me3- and H3K27me3-HiChIP loops were enriched for dACRs (Fig. 3b). Compared to immediately adjacent flanking regions, dACRs were strongly enriched for long-distance interactions (Fig. 3i and Supplementary Fig. 7a), indicating that the dACRs themselves (as opposed to nearby regions) were the focal points of the long-distance interactions. HiChIP detected more loops than

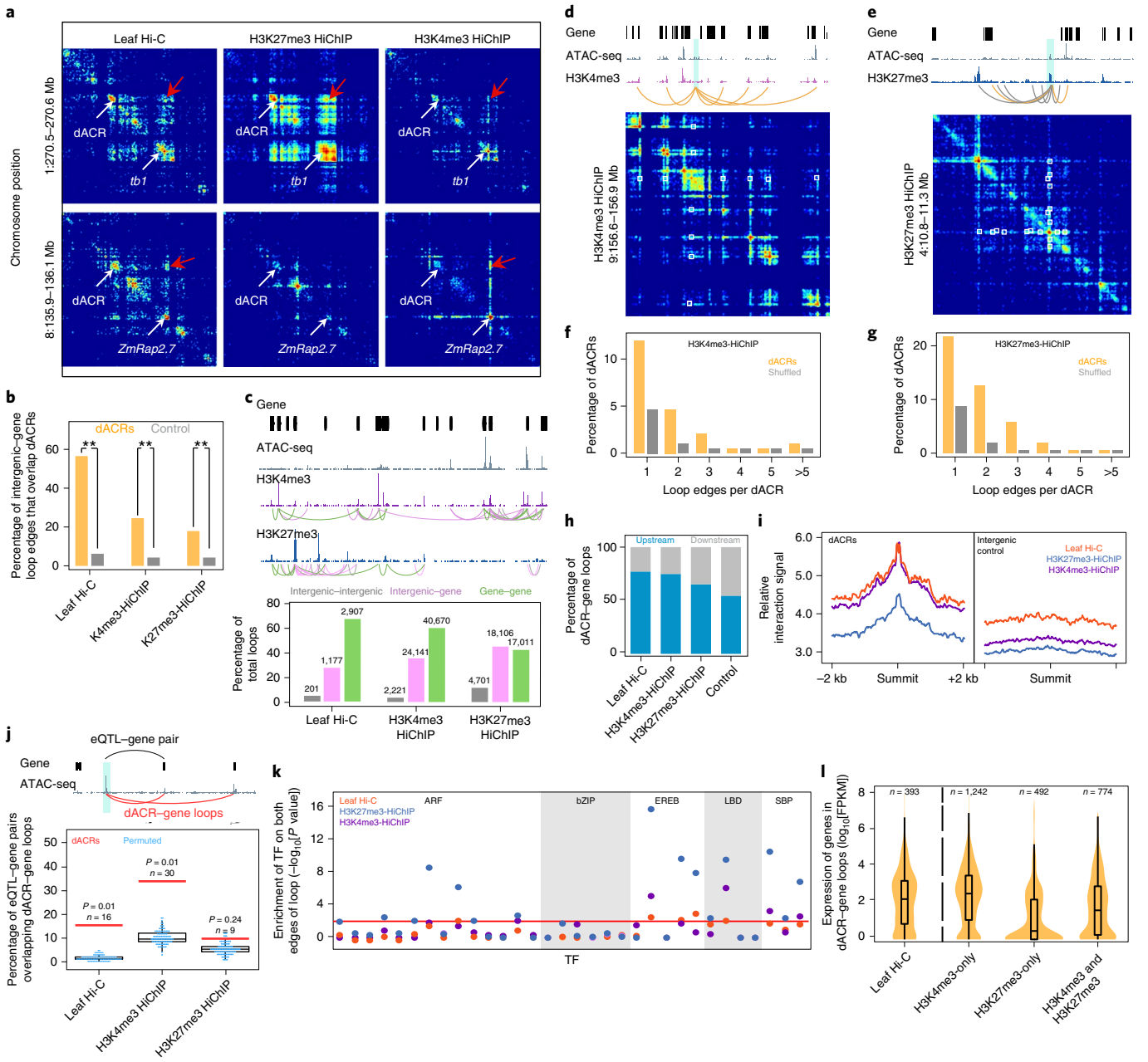


Fig. 3 | Hi-C and HiChIP identify dACR–gene interactions. a, Contact matrix heat maps showing the dACR–gene interactions at *tb1* and *ZmRap2.7*. Red arrows indicate dACR–gene contacts. **b**, Percentage of intergenic–gene loop edges overlapping dACRs. The asterisks denote $P \ll 2.2 \times 10^{-16}$ (Fisher's exact test, two-sided). Leaf Hi-C, $n = 1{,}177$ total loops (within a single biological replicate); H3K4me3 HiChIP, $n = 24{,}141$; and H3K27me3 HiChIP, $n = 18{,}106$. **c**, Representative regions containing various HiChIP loops (top panel) and called loop numbers from Hi-C and HiChIP experiments (bottom panel). Grey curves indicate intergenic–intergenic interactions, pink curves indicate intergenic–gene interactions and green curves indicate gene–gene interactions. **d**,**e**, Regions demonstrating dACR interaction hubs (dACR anchors indicated by shaded pale blue regions in the upper panel). White squares in heat maps indicate loops. Yellow curves indicate dACR–gene loops and grey curves indicate intergenic–intergenic loops. **f**,**g**, Percentages of dACRs involved in multiple dACR–gene loops compared to a control of shuffled dACRs and loops. From a total of 6,939 dACRs (excluding the transcribed group dACRs), 2,809 dACRs looped with one or more genes in H3K4me3-HiChIP (**f**) and 2,001 dACRs looped with one or more genes in H3K27me3-HiChIP (**g**). **h**, The percentages of dACR–gene loops in which the dACR resides either upstream or downstream of the target gene's promoter. dACR–gene pairs that were not crossing gene(s) were used for the analysis. **i**, Virtual circularized chromosome conformation capture intrachromosomal interaction signals at dACR summits and flanking regions. **j**, Top panel: a representative eQTL–gene pair (black curve) connected to Hi-C/HiChIP loops (red curves). Bottom panel: the percentage of eQTL–gene pairs that were connected by loops (red line) compared to genomic-distance-constrained dACR–gene random permutations (blue dots). *P* values were determined by a two-sided permutation test ($n = 100$). **k**, Enrichment of DAP-seq peaks of the same TF in both edges of the same loop (dACR–gene loops only). The red line indicates $P = 0.01$ (Fisher's exact test, two-sided). bZIP, basic leucine zipper; EREB, ethylene responsive element binding; LBD, lateral organ boundaries domain; SBP, squamosa-promoter binding protein. **l**, Expression of genes involved in different dACR–gene loops, separated by HiChIP loop type (n = number of genes shown in the violin distribution). The box plot shows median and quartiles. For the Hi-C and HiChIP experiments in this figure, biological replicates were not performed. FPKM, fragments per kb of transcript per million mapped reads.

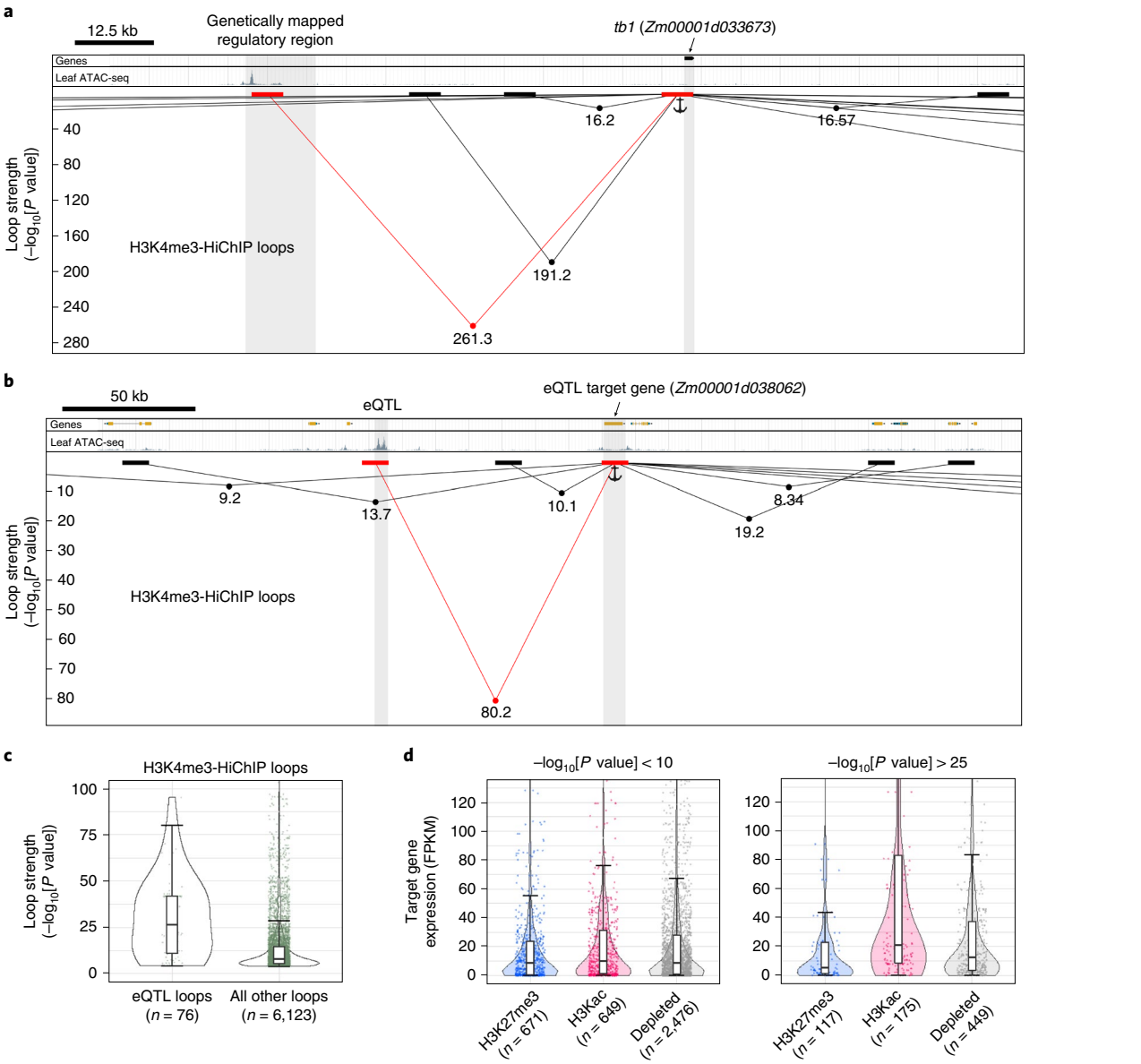


Fig. 4 | Loop strength identifies specific CRE–gene regulatory interactions. a,**b**, Genome browser shot of *tb1* and its fine-mapped distal regulatory region (**a**) and a genetically mapped eQTL and its predicted target gene (**b**). Chromatin loops are represented as line segments with dots indicating $-\log_{10}[P]$. Black and red blocks represent loop edges for all loops interacting with the *tb1* locus (indicated as anchor). The red boxes, connected by red line segments, represent the local loop with the highest contact significance that interacts with the anchor region; the black boxes, connected by black line segments, represent all of the other local loops that interact with the anchor region (that is, not the most significant loop). Panels **a** and **b** were not performed in replicate. **c**, The statistical significance of all H3K4me3-HiChIP loops that link dACR-overlapping eQTL to their target genes versus all other dACR-gene H3K4me3-HiChIP loops. **d**, The expression of target genes at one edge of the loop and dACR at the other end of the loop, split into the three chromatin groups, as classified in Fig. 2. Shown are loops at high and low $-\log_{10}[P]$. Boxplots in **c** and **d** include a median with quartiles and outliers above the top whiskers. All *P* values were determined in the FitHiChIP program using a two-tailed binomial test.

Hi-C and revealed webs of interactions among genes and dACRs (Fig. 3c). Thirty-four percent of all dACRs (excluding the transcribed group dACRs) looped to more than one gene. The dACR–gene loops that did not skip over genes occurred more often upstream than downstream of target genes (Fig. 3h and Supplementary Fig. 7b). In support of the biological relevance of these long-distance interactions, we found that the Hi-C and HiChIP loops could recapitulate links between intergenic eQTL and their target genes[20]. Compared to the background looping rates, dACRs that overlapped eQTL were more likely to loop with the target genes predicted by eQTL (Fig. 3j and Supplementary Fig. 7c). A subset of TFs also showed enrichment for binding (via DAP-seq) at both edges of the same loops (Fig. 3k), suggesting a potential mechanism for sequence-specific loop stabilization.

HiChIP allowed us to distinguish the chromatin-looping status of active (H3K4me3-enriched) and silenced (H3K27me3-enriched) genes within tissues containing mixed cell types. For example, strong chromatin interactions were detected between *tb1* (silenced in leaves) and its distal CRE using H3K27me3-HiChIP, whereas the dACR–gene loop at *ZmRap2.7* (expressed in leaves) was only detected by H3K4me3-HiChIP (Fig. 3a). To systematically explore such relationships, we catalogued dACR–gene loops that were

enriched exclusively for H3K4me3-HiChIP loops, H3K27me3 loops or for an overlap of both (Supplementary Fig. 7d). In the H3K4me3-only loops, H3K4me3 was present at genes but absent from the flanking histones of the dACRs (Supplementary Fig. 7e). In contrast, many of the H3K27me3-only loops (219/632) contained H3K27me3 at both the genes and the interacting dACRs (Supplementary Fig. 7f). Additionally, genes in H3K4me3-only loops were expressed at higher levels than genes in H3K27me3-only loops (Fig. 3l). Although loop identification was not exhaustive, these results demonstrated that dACRs interacted with their target genes via chromatin loops during both transcriptional activation and repression.

Chromatin loop contact strength suggests loops involved in transcriptional regulation. The genes at the aforementioned agronomic loci formed multiple chromatin loops with local regions (Fig. 4a and Supplementary Fig. 8a,b). At each of these genes, the strongest chromatin loop (as measured by the loop statistical significance provided by FitHiChIP[32]) occurred between the genetically mapped control region and the target gene. For example, the chromatin loop connecting *tb1* to its control region 65 kb upstream was stronger than the other loops that also interacted with *tb1*, even those spanning shorter genomic distances (Fig. 4a). Similarly, chromatin loops that connected eQTL to their predicted target genes[20] were stronger than non-eQTL loops (Fig. 4b,c). Furthermore, strong H3K4me3-HiChIP loops preferentially connected highly expressed genes with the H3Kac group dACRs, a relationship that was not apparent for dACRs and genes connected by weaker loops (Fig. 4d). These results suggested that regulatory CRE–gene interactions could be predicted by the strength of the chromatin loops that connected them.

Nearly all the genetically mapped regulatory elements previously discussed resided upstream of their target genes with no intervening genes (Fig. 1a and Supplementary Fig. 2a–c) (the one exception was *BX1*, which had one intervening gene). Among the H3K4me3-HiChIP loops that connected eQTL to their target genes, ~75% of the dACRs were located upstream of the target genes and ~75% of the loops connected dACRs to adjacent genes (that is, no intervening genes). These spatial biases were consistent with the fact that strong loops preferentially contained dACRs located upstream of and adjacent to their interacting genes (Supplementary Fig. 8c,d). Collectively, these results suggested that long-range regulatory interactions were predictable based on loop strength, orientation and location relative to target genes.

Gene–dACRs display elevated transcriptional enhancer capacities. To obtain independent and empirical evidence for the transcriptional regulatory capacities of dACRs, we performed self-transcribing active regulatory region sequencing (STARR-seq)[33] (a massively parallel enhancer reporter assay) in maize mesophyll protoplasts. We used the enrichment of transcriptional output ('STARR-RNA') over DNA input ('STARR-input') as a quantitative readout of transcriptional enhancer activity[33] (Fig. 5a). We first performed STARR-seq on a ~150 kb bacterial artifical chromosome that contained the *tb1* control region (encompassing the region shown in Fig. 1a). The *tb1* control region had previously been demonstrated to function as an enhancer in maize protoplasts[7] and could serve as a positive control for STARR-seq. A *Hopsctoch* long-terminal repeat, previously identified as the enhancer-containing element within the *tb1* control region, showed pronounced elevation of STARR-seq activity compared to the adjacent genomic regions (Fig. 5b). This demonstrated that the STARR-seq assay was sufficiently sensitive to detect a previously validated maize enhancer.

We then performed STARR-seq with a leaf ATAC-seq library as the input. This allowed us to quantify the enhancer activities of all ACRs in parallel (Fig. 5a). The enhancer activities of dACRs (excluding the transcribed group) were significantly greater than the activities of control regions (control regions were intergenic, non-ACRs containing sufficient STARR-input coverage, matched for length and GC content) (Mann–Whitney, $P < 10^{-314}$) (Fig. 5c,d). dACRs and pACRs showed similar enhancer activities, with activities (regression coefficients) of dACRs and pACRs twice that of control regions (Fig. 5c). Further analyses suggested that many dACRs functioned as bona fide transcriptional enhancers. In 95% of cases, the enhancer activities of candidate DNA fragments were independent of their orientations relative to the minimal promoter within the STARR-seq vector (Fig. 5e,f). dACRs participating in long-distance chromatin loops were significantly more active than dACRs that were not within loop edges (Fig. 5g). The H3Kac group dACRs showed significantly greater enhancer activity than those of the depleted and H3K27me3 dACR groups (Fig. 5h). Lastly, the binding of specific classes of TFs (via DAP-seq) was associated with increased enhancer activities and was enriched in highly active dACRs (Fig. 5i,j). Taken together, these results demonstrated that dACRs generally contained the capacity to act as transcriptional enhancers and that H3Kac group dACRs looping to genes showed the greatest enhancer capacities.

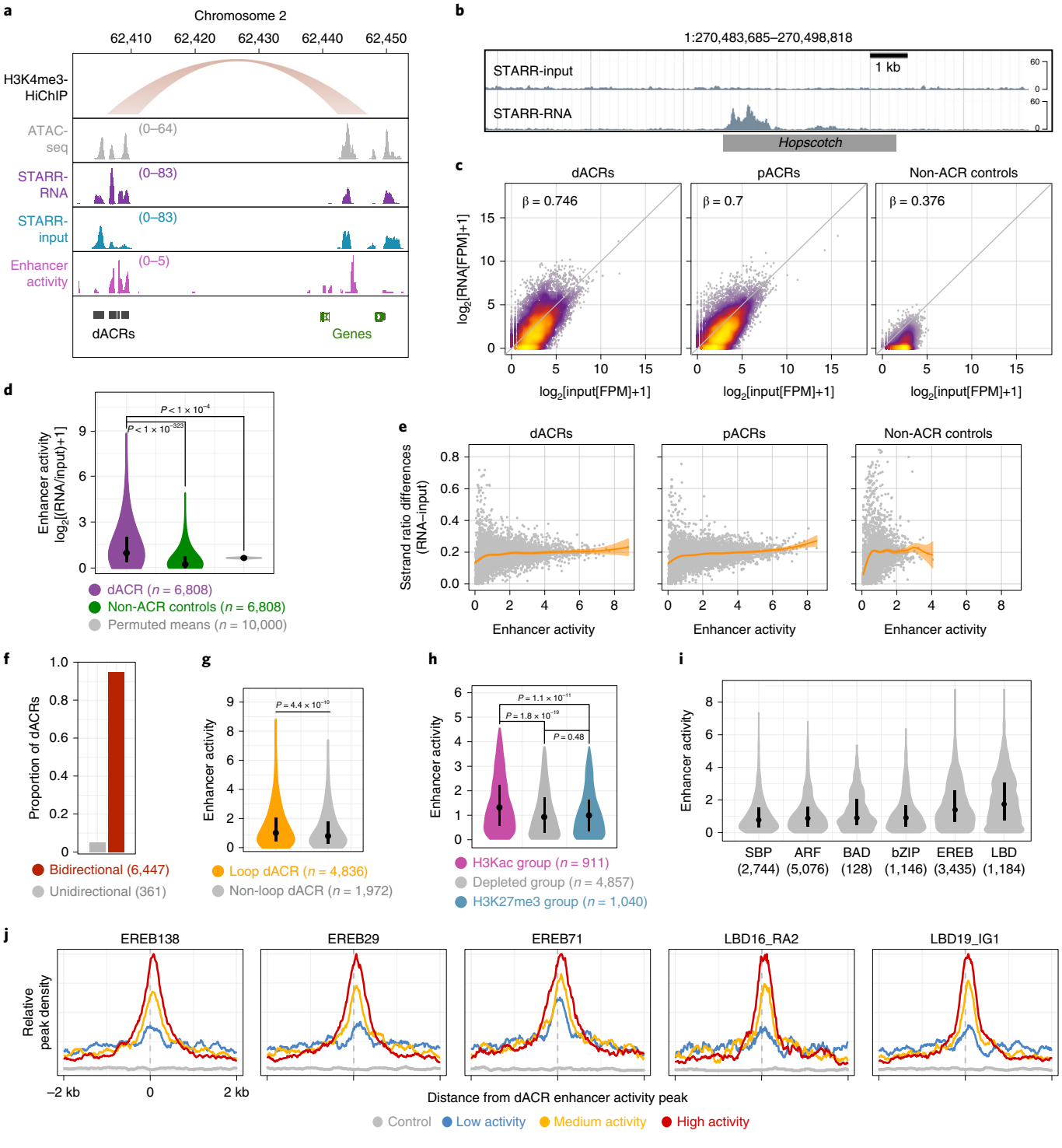
Discussion

Decades of studies on individual loci in the compact genome of *A. thaliana* suggested that CREs were predominantly located within or near genes[34]. However, emerging evidence in maize suggests that CREs can control genes located dozens of kilobases away. The most notable examples are several fine-mapped agronomic loci (including *tb1* (ref. [7]), *ZmRap2.7* (ref. [6]), *BX1* (ref. [8]) and *ZmCCT9* (ref. [5])), which are hypothesized to contain CREs that act over large genomic

Fig. 5 | dACRs display elevated transcriptional enhancer capacities. a, Representative region showing a H3K4me3-HiChIP loop, ATAC-seq, RNA from STARR-seq, input from STARR-seq and the estimated enhancer activity using the $\log_2$-transformed ratio of STARR-seq signal to input (RNA/input). **b**, STARR DNA input from a bacterial artificial chromosome (top track) and its corresponding RNA output (bottom track) at the *Hopscotch* positive control locus characterized by Studer et al.[7]. **c**, STARR-RNA versus STARR-input fragments per million (FPM) across dACRs (including H3Kac, depleted and H3K27me3 groups of dACRs and excluding the transcribed group of dACRs; left panel), pACRs (middle panel) and intergenic control regions (right panel). Regression coefficients are from a generalized linear model. **d**, Distributions of enhancer activities (max $\log_2$[RNA/input] FPM) for dACRs (excluding the transcribed group) compared to matched control regions (two-sided Mann–Whitney, $P < 1 \times 10^{-323}$) and mean enhancer activities of permuted random mappable regions matched in length to dACRs ($n = 6{,}808$ regions per iteration, $n = 10{,}000$ Monte Carlo iterations). **e**, Absolute difference in strand ratios between STARR-RNA and STARR-input fragments for dACRs (left), pACRs (middle) and control regions (right) relative to enhancer activity. **f**, Proportion of dACRs with bidirectional and unidirectional activity determined by a beta-binomial model. The number of dACRs are shown in parenthesis. **g**, Distribution of enhancer activities for dACRs coincident or non-coincident with HiChIP loop edges (Mann–Whitney, $P < 4.4 \times 10^{-10}$). **h**, distribution of enhancer activities among the different dACR chromatin group classifications. Hypotheses tests were performed using Mann–Whitney. **i**, Distribution of enhancer activities overlapping binding site peaks of DAP-seq-profiled TF families (n = number of dACRs containing DAP-seq peaks). BAD, BRANCH ANGLE DEFECTIVE 1. **j**, Average density of DAP-seq peaks centred on enhancer activity summits within dACRs. dACRs are split by enhancer activity. The sample sizes used for metaplots in **j** were the same as in **i**. The STARR-seq experiment described in this figure was performed as a single biological replicate. Boxplots shown in **d**,**g**,**h**,**i** comprise medians (black dots) and quartiles. Violin plots depict 0–99% of the entire distribution.

distances. Wallace et al.[4] compiled thousands of agronomic QTL and found that approximately one-third of QTL were >5 kb from their nearest annotated genes. These QTL suggest a potentially substantial role for distal regulatory elements in controlling agronomic phenotypes. However, the QTL could also derive from unannotated genes or gene presence–absence variation. Rodgers-Melnick et al.[11] and ourselves (Fig. 1j) demonstrated that dACRs are enriched for intergenic QTL, indicating that many of these QTL contain euchromatin, probably in the form of unannotated genes, non-coding transcription units or regulatory elements. We used histone modification data to identify 7,157 dACRs that did not resemble transcription units (Fig. 2b–d). These dACRs, which are unlikely to be annotation artefacts, are the most likely candidates for long-range CREs in the maize genome.

Multiple lines of evidence indicate that the non-transcription unit dACRs contain CREs: (1) the dACRs overlap fine-mapped hypothesized CREs; (2) dACRs display DNA sequence constraint, manifested as elevated GC content and depleted SNP frequency; (3) dACRs are enriched for TF binding sites and (4) eQTL; (5) dACRs loop to genes in *cis*, and these loops recapitulate genetically predicted interactions. The dACR–gene loops occur in a spatially non-random manner (dACRs containing putative CRE are



primarily adjacent to and upstream of target genes) and this resembles CREs in the proximal promoters of genes; (6) dACRs with acetylated flanking histones preferentially loop to highly expressed genes, thereby establishing a connection between the chromatin status and transcriptional status at distant loci and (7) dACRs contain sequence elements capable of acting as transcriptional enhancers. Collectively, these results indicate abundant CRE-containing dACRs in the maize genome.

dACRs display chromatin attributes that are useful for their discovery and classification. We found that all dACRs were depleted of DNA methylation (Fig. 2b–f). This finding was previously reported by Rodgers-Melnick et al.[11] and Oka et al.[14] using MNase- and DNase-based assays, respectively. Because regions of depleted DNA methylation in plant genomes are developmentally stable, DNA methylation status can potentially be used to locate CREs within tissue-specific dACRs that are not detectable in bulk accessibility assays[35]. Flanking histone modifications allowed us to separate dACRs into transcribed, H3Kac, H3K27me3 and modification-depleted groups (Fig. 2). The non-transcribed group dACRs appear analogous to metazoan transcriptional enhancers, which are acetylated when active, enriched for H3K27me3 when inactive, and neither enriched for acetylation nor H3K27me3 when in a primed state[1]. Our maize results, which demonstrate an association of H3Kac dACRs with highly expressed genes and an association of H3K27me3 dACRs with polycomb-silenced genes (Figs. 3 and 4), suggest that the chromatin marks at dACRs in maize are analogous to those in metazoans. However, the absence of H3K4me1 at maize dACRs (also found previously by Oka et al.[14]) contrasts with metazoan enhancers and suggests mechanistic differences in how TFs interact with chromatin pathways.

The prevalence of distal CREs raises the question of how long-range chromatin loops are established and maintained between CREs and their target genes. The loops may form as a consequence of compartmental segregation, in which euchromatic regions (primarily genes and ACRs) self-associate and exclude the intervening heterochromatin, thereby forming loops that span heterochromatin[36–38]. Alternatively, sequence-specific architectural proteins may play a role in loop formation or stabilization[36–38]. Both of these mechanisms appear to be common throughout eukaryotes[36,37] and the dACR–gene loops described here can be explained by a combination of both. We speculate that the pervasive gene–gene and dACR–gene loops are a consequence of compartmental segregation. Loops with contact strengths elevated above local backgrounds may be brought together via compartmental segregation and then further stabilized by sequence-specific architectural proteins. The dACR–gene loops that are likely to contain specific CRE–gene interactions (such as the fine-mapped agronomic loci and the eQTL–gene interactions) display the greatest contact strengths (Fig. 4 and Supplementary Fig. 8). This leads us to speculate that specific CRE–gene interactions are stabilized by sequence-specific factors, such as TFs that form multimers. The contact strengths of loops can therefore be used to distinguish specific regulatory loops from non-specific compartmental loops. Furthermore, because the predicted regulatory loops preferentially reside upstream of and adjacent to their target genes, putative distal CREs can be assigned to target genes with reasonable confidence, even in lieu of Hi-C data.

A companion study in this issue (Lu et al., 2019 [DOI to come]) demonstrates that distal CREs exist across a wide range of evolutionarily diverse angiosperms and are particularly abundant in plants with large genomes. Even within the compact *A. thaliana* genome, distal CREs are common in pericentromeric regions with low gene densities. A multispecies comparison of homologous ACRs (Lu et al., 2019 [DOI to come]) suggests that most distal CREs originate in gene–proximal regions (for example, the promoter) and become gene–distal as a result of transposable element proliferation. This is consistent with our observation that distal CREs in maize preferentially reside upstream of and adjacent to their target genes. Collectively, the results of both manuscripts indicate that long-range transcriptional regulation by CREs is a common phenomenon among angiosperms.

Methods

Experimental design. All experiments, except for Hi-C, HiChIP and STARR-seq, were replicated. We did not perform blind experiments or analyses. All assays, except for STARR-seq, were performed on the same tissues at the same developmental stages and grown in the same conditions. However, separate batches of plants were grown for separate experiments. Biological replicates were performed on separately grown batches of plants.

Plant material and growth conditions. *Z. mays* L. cultivar B73 was grown from seed collected from field-grown ears during summer 2017 in Athens, USA. ATAC-seq, RNA-seq, ChIP–seq, MethylC-seq, Hi-C and HiChIP experiments were all performed on seedling tissue grown under the following conditions: kernels were sown in Sungro Horticulture professional growing mix (Sungro Horticulture Canada). Soil was saturated with tap water and placed under a 50/50 mixture of 4,100 K (Sylvania Supersaver Cool White Delux F34CWX/SS, 34 W) and 3,000 K (GE Ecolux with starcoat, F40CX30ECO, 40 W) light. The photoperiod was 16 h of light and 8 h of dark. The temperature was approximately 25 °C during light hours. The relative humidity was approximately 54%. Seedlings were grown for approximately 6 d and harvested 4–6 h after Zeitgeber Time 0 (ZT0, lights on). Seedlings were harvested when the first leaf had emerged 2–3 cm above the apical tip of the coleoptile. The seedlings were cut 3 mm above the coleoptile–mesocotyl boundary, excluding the shoot apical meristem, and the second leaf was removed from within the sheath of the first leaf. Only the inner second leaves, which contained the third and fourth leaves sheathed inside, were used for experiments.

For experiments on young inflorescences (which were ear primordia, hereafter inflorescence primordia), B73 maize was grown in the field or greenhouse. Plants were harvested approximately 1 month after sowing and inflorescence primordia were dissected from shoots. Inflorescence primordia were harvested from any node of the shoot if the length was 3–8 mm from the base to the apical tip of the inflorescence primordia.

ATAC-seq. ATAC-seq was performed as described previously[13]. For each replicate, approximately 200 mg of maize second leaves and several inflorescence primordia were harvested and immediately chopped with a razor blade and placed in 2 ml of pre-chilled lysis buffer (15 mM Tris–HCl pH 7.5, 20 mM sodium chloride, 80 mM potassium chloride, 0.5 mM spermine, 5 mM 2-mercaptoethanol and 0.2% TritonX-100). The chopped slurry was filtered twice through miracloth and once through a 40-µm cell strainer. The crude nuclei were stained with 4,6-diamidino-2-phenylindole and loaded into a flow cytometer (Beckman Coulter MoFlo XDP). Nuclei were purified by flow sorting and washed in accordance with Lu et al.[13].

The sorted nuclei (50,000 nuclei per reaction) were incubated with 2 µl of transposome in 40 µl of tagmentation buffer (10 mM TAPS–sodium hydroxide pH 8.0, 5 mM magnesium chloride) at 37 °C for 30 min without rotation. The integration products were purified using a Qiagen MinElute PCR Purification Kit and then amplified using Phusion DNA polymerase for 11 cycles. PCR cycle number was determined as described previously[12]. Amplified libraries were purified with AMPure beads to remove primers.

To make the ATAC-seq control, nuclei were sorted and genomic DNA was extracted from maize leaves using the Qiagen DNeasy Plant Mini Kit (cat. no. 69106). Then ~1 ng of gDNA was incubated with 2 µl of transposomes in 40 µl of tagmentation buffer at 37 °C for 30 min. All procedures after this were identical to the standard ATAC-seq library protocol described here.

RNA-seq. Second leaves and inflorescence primordia were flash-frozen with liquid nitrogen immediately after collection. Samples were ground to a powder with a mortar and pestle in liquid nitrogen. Total RNA was extracted and purified with TRIzol Reagent (Thermo Fisher Scientific) following the manufacturer's instructions. For each tissue and replicate, 1.3 µg of total RNA was prepared for sequencing with the Illumina Truseq mRNA Stranded Library Kit (Illumina) following the manufacturer's instructions.

ChIP–seq. ChIP was performed following the general protocol of Zhang et al.[29]. For a single chromatin extraction, which yielded sufficient chromatin for several ChIPs, approximately 500 mg of leaves and five inflorescence primordia were used. Immediately after harvesting, the tissue was chopped into 0.5 mm cross-sections and crosslinked in accordance with the referenced protocol. Samples were immediately flash-frozen in liquid nitrogen after crosslinking. Nuclei were extracted and lysed in 300 µl of lysis buffer. The lysed nuclei suspension was sonicated on a Diagenode Bioruptor on the high setting: 30 cycles of 30 s on, 30 s off. Tubes were centrifuged at 12,000*g* for 5 min and supernatants transferred to new tubes. At this point, ChIP input aliquots were collected.

Dynabeads Protein A (Thermo Fisher Scientific, cat. no. 10002D) were washed with ChIP dilution buffer and then rotated with antibodies at a concentration of

1.5 µg of antibody (see Table 11 for antibodies used) per 100 µl of ChIP dilution buffer for 4 h at 4 °C. The antibody-coated beads were washed three times with ChIP dilution buffer.

Sonicated chromatin was diluted tenfold in ChIP dilution buffer to bring the SDS buffer concentration down to 0.1%. For all samples and replicates, 460 µl of diluted chromatin was incubated with 750 µg of Dynabeads Protein A coated with 1.5 µg of antibody. Samples were rotated at 4 °C overnight, then washed, reverse-crosslinked and treated with proteinase K in accordance with the referenced protocol. DNA was purified by a standard phenol–chloroform extraction followed by ethanol precipitation.

The DNA samples were end-repaired using the End-It DNA End-Repair Kit (epicentre) following the manufacturer's protocol. DNA was cleaned up on AMPure beads (Beckman Coulter) with a size selection of 100 bp and larger. Samples were eluted into 43 µl of Tris–HCl and underwent a 50µl A-tailing reaction in NEBNext dA-tailing buffer with Klenow fragment (3′–>5′ exo-) at 37 °C for 30 min. A-tailed fragments were ligated to Illumina Truseq adaptors and purified with AMPure beads. Fragments were amplified with Phusion polymerase in a 50 µl reaction following the manufacturer's instructions. The following PCR programme was used: 95 °C for 2 min, 98 °C for 30 s, then 15 cycles of 98 °C for 15 s, 60 °C for 30 s, 72 °C for 4 min and once at 72 °C for 10 min. PCR products were purified with AMPure beads to remove primers.

MethylC-seq. Several B73 second leaves were immediately flash-frozen after harvesting and ground to a powder in liquid nitrogen. DNA was extracted and purified with the DNeasy Plant Mini Kit (Qiagen) and 130 ng were used for MethylC-seq library preparation. MethylC-seq libraries were prepared as detailed in Urich et al.[39]; however, we used a final PCR amplification of eight cycles.

DAP-seq. DAP-seq experiments involving maize auxin response factor (ARF) samples were performed as detailed in Galli et al.[19]. All other TF were processed according to Bartlett et al.[40], with the exception that 1 µg of pIX-HALO-TF plasmid DNA was used for protein expression, 1 µg of adaptor-ligated library prepared from B73 inflorescence genomic DNA was used for DNA binding and 1 µg of maize leaf genomic DNA was used for EREB71 and EREB127 binding.

Hi-C and HiChIP. We performed HiChIP as detailed in Mumbach et al.[31], but with modifications in the nuclear isolation, enzymatic reactions and ChIP steps. Hi-C was performed identically to HiChIP, except after sonication, the chromatin was immediately reverse-crosslinked and the DNA purified. Fourteen B73 second leaves were harvested 4 h after ZT0 and were immediately crosslinked in 1% formaldehyde. Crosslinking was performed similarly to the ChIP protocol, except that the crosslinking times were extended: −84659.725 Pa for 20 min, followed by atmospheric pressure for 10 min, then −84659.725 Pa for 10 min, then −84659.725 Pa with glycine for 5 min and then washed six times in ultrapure water and flash-frozen in liquid nitrogen.

Approximately two-thirds of the flash-frozen tissue was used for nuclei extraction. The leaves were chopped with a razor blade for 5 min in ice-cold Hi-C lysis buffer (1 mM EDTA, 1 × cOmplete Mini EDTA-free Protease Inhibitor Cocktail, 10 mM Tris–HCl pH 7.5, 10 mM sodium chloride, 0.2% NP-40, 5 mM 2-mercaptoethanol, 0.1 mM phenylmethyl sulfonyl fluoride) and the slurry was passed through a 40-µm cell strainer. The filtrate was centrifuged at 2,000*g* at 4 °C for 2 min and the pellet was resuspended in 1 ml of Hi-C lysis buffer and strained a second time through a 40 µm cell strainer into a new tube. The suspension was centrifuged at 2,000*g* for 1 min and the pellet was resuspended in another 1 ml Hi-C lysis buffer. Nuclei concentration was determined via staining with 4,6-diamidino-2-phenylindole and viewing on a hemocytometer.

The nuclei suspension was split into two tubes, each containing approximately 4 million nuclei. The two tubes underwent identical Hi-C enzymatic reactions in parallel: the restriction digests, Klenow fill-in reactions and ligation reactions were performed as in Mumbach et al.[31]. Two-hundred units of DpnII restriction enzyme (NEB, R0543T) were used to digest each tube of 4 million nuclei. Tubes were rotated for 2 h at 37 °C for restriction digestion. The Klenow fill-in was performed with 50 units of DNA Polymerase I, Large Klenow Fragment (NEB, M0210). Ligation was performed with 4,000 units of T4 DNA Ligase (NEB, M0202). Tubes were rotated at 22 °C for 4 h for ligation. Nuclei were pelleted and lysed in 150 µl of nuclei lysis buffer (10 mM EDTA, 1% (v/v) SDS, 50 mM Tris–HCl pH 8.0, 0.1 mM phenylmethyl sulfonyl fluoride, 1 × cOmplete Mini EDTA-free Protease Inhibitor Cocktail). The samples were diluted twofold with the addition of 150 µl ChIP dilution buffer (1.2 mM EDTA, 167 mM sodium chloride, 16.7 mM Tris–HCl pH 8.0, 1.1% (v/v) Triton X-100, 0.1 mM phenylmethyl sulfonyl fluoride, 1 × cOmplete Mini EDTA-free Protease Inhibitor Cocktail) and sonicated on a Diagenode Bioruptor on the high setting: five cycles of 30 s on, 30 s off. Tubes were centrifuged at 16,000*g* for 5 min and the supernatants were transferred to new tubes. The supernatants were pooled together and diluted fivefold with ChIP dilution buffer to bring the SDS concentration to 0.1%.

The diluted, ligated chromatin was added to Dynabeads Protein A (Thermo Fisher Scientific, cat. no. 10002D), which had been previously incubating with antibodies as follows: Dynabeads were washed three times with ChIP dilution buffer and then rotated with antibodies at a concentration of 1.5 µg of antibody per 100 µl of ChIP dilution buffer for 4 h at 4 °C. We then incubated 4.5 µg of H3K27me3 antibody (Millipore 07-449) with 2,250 µg of beads, and also 3 µg of H3K4me3 antibody (Millipore 07-473) with 1,500 ug of beads. After incubation, the antibody-coated beads were washed three times with ChIP dilution buffer and the diluted chromatin was added to the beads. Then 1,380 µl (15 µg as measured by Qubit with DNA HS reagent) of chromatin was added to the H3K27me3 beads and 920 µl (9.6 µg) of chromatin was added to the H3K4me3 beads. Samples were rotated for 14 h at 4 °C. Chromatin washes, reverse crosslinking, proteinase K digestion and elution were performed in an identical fashion as in the ChIP protocol[29]. DNA was purified with the Monarch PCR and DNA Cleanup Kit (New England Biolabs) following the manufacturer's protocol. Each ChIP sample was eluted into 20 µl of ultrapure water. For each Hi-C and HiChIP sample, biotinylated DNA was captured, tagmented and amplified using PCR as in Mumbach et al.[31].

STARR-seq. The STARR-seq plasmid backbone features the core region of the cauliflower mosaic virus 35S promoter[41,42], followed by an open reading frame encoding green fluorescent protein derived from pMDC107, the cloning site containing a CcdB suicide gene, followed by a transcriptional polyA site derived from the *A. thaliana* ribulose bisphosphate carboxylase small chain 1A gene. The plasmid backbone is derived from pMD19 (simple) (http://www.snapgene.com/resources/plasmid_files/ta_and_gc_cloning_vectors/T-Vector_pMD19_(Simple)/). Our STARR-seq plasmid sequence and additional information can be found at Addgene, deposit number 117379 (https://www.addgene.org/117379/).

The genomic DNA input for the STARR-seq assay was a ~150 kb bacterial artifical chromosome (BAC) CH201–136H12 (https://www.maizegdb.org/data_center/bac?id=613738) and an ATAC-seq library derived from maize second leaves. Libraries for the BAC or ATAC inputs were prepared in an identical manner, although scaled down tenfold for the BAC. To generate the starting ATAC library, we followed the same method detailed in the ATAC-seq methods section; however, the protocol was scaled up to 1 million nuclei instead of 50,000. The tagmented product was split into eight 50 µl PCRs. A single primer was used for amplification (5′-AGATGTGTATAAGAGACAG-3′) instead of the standard Nextera primers. The following PCR programme was used: 72 °C for 5 min, 98 °C for 30 s, seven cycles of 98 °C for 10 s, 55 °C for 30 s, 72 °C for 30 s and then 72 °C for 2 min. The PCR product was size-selected on an 0.8% agarose gel to a range of 400–700 bp. The gel product was purified with the Monarch PCR and DNA Cleanup Kit (New England Biolabs). The eluate was split into a second PCR round of eight 50 µl reactions with the same number of cycles as indicated above. The purpose of multiple parallel and serial PCRs was to reduce amplification biases. The PCR products were combined and concentrated with the Monarch PCR and DNA Cleanup Kit.

The STARR-seq plasmid was double-digested with the restriction enzymes SacI and KpnI and the upper band was gel-purified. The sticky ends of the gel product were blunted by incubating with Large Klenow Fragment (New England Biolabs) following the manufacturer's instructions. The ATAC fragments and vector backbone were assembled with the NEBuilder Hifi DNA Assembly Mastermix (New England Biolabs) according to the manufacturer's instructions. The reaction product was precipitated using ethanol, washed with 70% ethanol and dissolved in 15 µl of ultrapure water. We electropulsed 80 µl of MegaX DH10B T1R Electrocomp Cells (Thermo Fisher Scientific) with 2 µl of HiFi Assembly product at 2,000 V and 25 µF. The cells were grown for 16 h in 1 l of lysogeny broth with 100 µg ml^{-1} carbenicillin. Plasmids were isolated with the NucleoBond Xtra Midi EF kit (Macherey-Nagel) following the manufacturer's instructions and the purified product was dissolved in ultrapure water to a concentration exceeding 1 µg µl^{-1}.

For the generation and transfection of maize mesophyll protoplasts, we followed the maize-specific guidelines of the Jen Sheen lab (https://molbio.mgh.harvard.edu/sheenweb/protocols_reg.html); however, we used the polyethylene glycol transfection method detailed in Yoo et al.[43]. Maize seedlings were sowed and grown under conditions detailed in the plant growth and materials section; however, once the coleoptiles emerged approximately 1 cm above the soil surface, trays of plants were transferred to total dark conditions and etiolated for approximately 1 week. Protoplasts were extracted, transfected and then incubated on petri dishes for 14 h at 22 °C at a concentration of 1 million cells ml^{-1}. An estimated 15 million protoplasts were transformed by STARR-seq plasmids. Protoplasts were pelleted by centrifugation at 100*g* for 2 min and the cell pellets were immediately flash-frozen in liquid nitrogen.

Total RNA was extracted from protoplasts via the Monarch total RNA miniprep kit (New England Biolabs) using the cultured mammalian cell protocol in the manufacturer's instructions. On-column DNase treatment was performed. Total RNA was eluted into RNase-free water. To enrich for polyA RNA, 276 µg of total protoplast RNA was incubated with 4 mg of Dynabeads Oligo (dT)25 (ambion cat. no. 61002) following the manufacturer's protocol. We then eluted 5.5 µg of polyA RNA into 160 µl of RNase-free water.

The polyA RNA was incubated in a 200 µl Turbo DNase reaction (Turbo DNA-free kit, Thermo Fisher Scientific) at 37 °C for 25 min. DNase was inactivated by the addition of 20 µl DNase inactivation reagent. The reaction was cleaned up and concentrated with the Monarch RNA cleanup kit (New England Biolab) following the manufacturer's instructions.

The Superscript IV reverse-transcriptase kit (Thermo Fisher Scientific, cat. no. 18091050) was used for cDNA first-strand synthesis. We split 3.7 µg of polyA RNA into ten reactions, and 0.25 µg of polyA RNA was used for a no reverse-transcriptase negative control. The cDNA was primed with a plasmid-specific primer (5′- TTGAGGTCTACACAAAAGCAAAGGG-3′). The samples were treated with RNaseH following cDNA synthesis. The cDNA was Monarch-purified and eluted into 40 µl of 10 mM Tris–HCl.

PCR was performed on the first-strand cDNA with Phusion polymerase. The cDNA library was split into 16 50-µl PCRs with the following parameters: 98 °C for 1 min, ten cycles of 98 °C for 15 s, 63 °C for 30 s, 72 °C for 1 min, then once at 72 °C for 2 min. The reactions were pooled, Monarch-purified and selected for size (300–800 bp, which encompassed the entire range of the library) on a 0.8% agarose gel to remove primers. The purpose of the size selection was to eliminate primers and small fragments that resulted from RNA splicing. The DNA was purified from the gel with the Monarch DNA gel extraction kit. This product was split into eight more 50-µl PCR with the same parameters, except for a total of four cycles. This product was similarly size-selected on a 0.8% gel and DNA purified. To determine how much plasmid input to amplify, quantitative PCR was used to determine a computerized tomography of similar value to the cDNA. The plasmid input was subjected to the same PCR protocol and the samples were sequenced on an Illumina Nextseq500 platform with paired-end 35 bp reads.

Sequencing Information. Sequencing of ATAC-seq, RNA-seq, ChIP–seq, DAP-seq, Hi-C, HiChIP and STARR-seq was performed at the University of Georgia Genomics Facility using an Illumina NextSeq 500 instrument. MethylC-seq was performed at the University of Minnesota, Twin Cities using an Illumina HiSeq 2500 instrument. ATAC-seq, MethylC-seq, Hi-C, HiChIP and STARR-seq were sequenced in paired-end 35 bp, 125 bp, 75 bp, 75 bp and 35 bp, respectively. RNA-seq leaf and inflorescence replicates were sequenced in single-end 75 bp and 150 bp, respectively. ChIP–seq and DAP-seq were sequenced in single-end 75 bp. Information on read counts and alignment statistics can be found in Supplementary Tables 5–10.

Data processing, quantification and statistical analyses. *Definition of intergenic negative control regions.* To create the intergenic negative control regions, we first generated all possible simulated 75 bp fragments in the *Z. mays* v4 AGPv4 reference genome[44] by extending 75 bp downstream from every position in the genome. Then the uniquely mappable regions were identified by remapping all simulated fragments with the same parameters for ChIP–seq analysis. Genomic regions with mapped reads were considered as uniquely mappable. Annotated genes and their 2 kb flanking regions, as well as gene–dACR, were removed. Negative control regions with the same length distribution to dACR were then generated by the 'shuffle' command in BEDTools[45], constrained to only the genomic space that was determined to be uniquely mappable.

ATAC-seq raw data processing and alignment. Raw reads were trimmed with Trimmomatic v.0.33 (ref. [46]). Reads were trimmed for NexteraPE with a maximum of two seed mismatches, a palindrome clip threshold of 30 and a simple clip threshold of ten. Reads shorter than 30 bp were discarded. Trimmed reads were aligned to the *Z. mays* AGPv4 reference genome[44] using Bowtie v.1.1.1 (ref. [47]) with the following parameters: 'bowtie -X 1000 -m 1 -v 2 --best –strata'. Aligned reads were sorted using SAMtools v.1.3.1 (ref. [48]) and clonal duplicates were removed using Picard version v.2.16.0 (http://broadinstitute.github.io/picard/).

RNA-seq raw data processing, alignment and expression quantification. Raw reads were trimmed with Trimmomatic v.0.33 (ref. [46]) with default parameters. The remaining reads were aligned to the *Z. mays* AGPv4 reference genome[44] using HISAT2 v.2.0.5 (ref. [49]). Gene expression values were computed using StringTie v.1.3.3b[50] with the maize annotation version AGPv4.38. Genes determined to have at least a twofold expression change and statistically significant differences in expression (adjusted *P* value cutoff of 0.05) by DESeq2[51] were identified as differentially expressed genes.

ChIP–seq raw data processing and alignment. Raw reads were trimmed with Trimmomatic v.0.33 (ref. [46]) with default parameters. The remaining reads were aligned to the *Z. mays* AGPv4 reference genome[44] using Bowtie v.1.1.1 (ref. [47]) with the following parameters: 'bowtie -m 1 -v 2 --best --strata --chunkmbs 1024 -S'. Aligned reads were sorted using SAMtools v.1.2 and duplicated reads were removed using SAMtools v.0.1.19 (ref. [48]).

MethylC-seq raw data processing, alignment and calculation of methylation status. Quality-filtering and adaptor-trimming were performed using cutadapt v.1.9.dev1. Reads were aligned to the *Z. mays* AGPv4 reference genome[44] using Methylpy 1.3 as described in Schultz et al.[52]. Chloroplast DNA was used as a control to calculate the sodium bisulfite reaction non-conversion rate of unmodified cytosines. The conversion rates were >99.7%. A binomial test was used to determine the methylation status of cytosines with a minimum coverage of three reads.

DAP-seq raw data processing and alignment. DAP-seq analyses were performed as described Galli et al.[19]. Raw reads were trimmed using Trimmomatic[46] with the following parameters: ILLUMINACLIP:TruSeq3-SE:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50. Trimmed reads were mapped to the *Z. mays* AGPv4 reference genome[44] using Bowtie2 v.2.2.8 (ref. [47]). Mapped reads were filtered for reads containing >MAPQ30 using SAMTools (samtools view –b –q 30)[48].

Hi-C and HiChIP raw data processing and interaction-calling. The Hi-C library quality was determined following the principles of Rao et al.[53]. Raw data were processed with the HiC-pro v.2.8.0 pipeline[54]. We independently aligned the paired-end 75-bp reads using Bowtie2 with the iterative mapping strategy. Alignments with MAPQ > 5 were kept for further analysis. Read pairs within the same restriction enzyme fragments and PCR duplicates were removed. Raw interaction matrices for selected windows were generated with analyzeHiC from Homer v.4.10.0 (ref. [55]) with the parameters '-res 200 -superRes 2000 -raw'. The validated contact pairs were then transformed to Juicer hic files with hicpro2juicebox. Loop-calling for the leaf Hi-C experiment was performed using Juicer v.0.7.0 HICCUPS[56] with 5 kb and 10 kb bin sizes and a maximum genomic distance of 2 Mb.

HiChIP raw data were also processed with the HiC-pro pipeline v.2.8.0 (ref. [54]). Alignments with MAPQ > 5 were kept for further analyses. The ChIP pulldown efficiency was determined by analysing dangling end and self-ligation reads. The valid read pairs were used for loop-calling. H3K4me3 and H3K27me3-HiChIP loops were identified using FitHiChIP[32] with 5 kb bin sizes, bias correction by coverage, false discovery rate (FDR) <0.01, a minimum genomic distance of 20 kb and a maximum genomic distance of 2 Mb.

STARR-seq data processing. Raw reads from the STARR-RNA and STARR-input libraries were trimmed for adaptor, quality and minimum length with Trimmomatic v.0.36 (ref. [46]) (SLIDINGWINDOW:3:20 LEADING:0 TRAILING:0 MINLEN:30) and mapped to the *Z. mays* AGPv4 reference genome[44] using Bowtie v.1.2.2 (ref. [47]) with non-default parameters (-t -v 1 -X 2000 --best --strata -m 1 -S). All reads overlapping BAC-contaminated regions were removed (see Supplementary Table 12). Fragments were inferred using the start and end positions from the paired-end alignments. STARR peaks were identified from fragments using MACS2 with non-default settings (--keep-dup –bw 1000) by defining the STARR-RNA and -input libraries as treatment and control, respectively[57]. The FDR was controlled to $\alpha < 0.05$ via the Benjamini–Hochberg method. Enhancer activity was determined at base-pair resolution as the ratio of RNA to input FPM. dACR enhancer activity was estimated as the maximum ratio of RNA to input within the dACR interval. This was done instead of calculating the activity of the entire dACR to account for the fact that only a small portion of a dACR may contain the *cis*-regulatory element of interest. Control regions ($n = 6{,}808$) were identified from random mappable regions, matched to dACR peak lengths and a similar composition of input FPM (median difference between dACR and control input FPM = 0.008). For calling dACR transcriptional directionality, forward to reverse ratios of fragments overlapping dACR were modelled as beta-binomial distributions independently for RNA and input fragments. Significant departure of RNA forward to reverse fragment ratios from input ratios was estimated through empirical construction of *P* values by Markov Chain Monte Carlo sampling ($n = 10{,}000$) of the two beta-binomial distributions. H3K4me3 and H3K27me3-HiChIP loops were used to define dACR as upstream or downstream of their target genes. Enhancer activities for DAP-seq-enriched dACR were estimated similarly as previous dACR activity analyses. We split dACR into three equal-sized groups based on activity (low, medium, high) for DAP-seq peak density analysis.

Identification of ACRs. MACS2 (ref. [57]) was used to define ACRs with the '--keep-dup all' function and with ATAC-seq input samples (Tn5 transposition into naked gDNA) as a control. The ACRs identified by MACS2 were further filtered using the following steps: (1) peaks were split into 50 bp windows with 25 bp steps; (2) the accessibility of each window was quantified by calculating and normalizing the Tn5 integration frequency in each window with the average integration frequency across the whole genome to generate an enrichment fold value; (3) windows with enrichment fold values passing a cutoff (25-fold) were merged together by allowing 150 bp gaps and (4) possible false positive regions were removed by filtering small regions with only one window for lengths >50 bp. The sites within ACRs with the highest Tn5 integration frequencies were defined as ACR 'summits'.

Identification of differential ACRs. To call differential ACRs, MACS2 (ref. [57]) was first used with '--keep-dup all'. The identified ACRs were filtered as such: (1) they were kept if they overlapped with the filtered ACRs (for example, Leaf versus Inflorescence differential ACRs should overlap Leaf ACRs) and (2) the Tn5 integration frequency of each peak was calculated and normalized with the integration frequency of 100 kb regions centred around the peak. Differential ACRs that passed a fold-change cutoff (Leaf versus Inflorescence, fourfold; Inflorescence versus Leaf, twofold) were selected.

Identification of DAP-seq peaks. Peaks were called using GEM v.2.5 (ref. [58]) using the GST-HALO negative control sample and a blacklist of peak regions appearing

in all samples for background subtraction. Peak-calling was performed with the following parameters: --d Read_Distribution_default.txt --k_min 6 --k_max 20 --outNP –sl. The default FDR (0.01) was used for all samples except ARF, which used an FDR of 0.00001 (--q 5). The final peaks were merged together using BEDTools 2.25 (ref. [45]).

Heatmap and metaplot analysis. Two-hundred 10-bp bins were created for both upstream and downstream regions, starting from transcription start sites and transcription polyA sites based on the *Z. mays* AGPv4.38 genome annotation[44]. For analyses flanking ATAC-seq-identified peak summits, 200/500 10-bp bins were created. For MethylC-seq, weighted methylation levels were computed for each predetermined bin[52]. For ChIP–seq and RNA-seq analyses, the number of reads per bin were normalized by total aligned reads in each library. Average values were calculated for samples with two replicates. Histone modifications were further normalized by subtracting the H3 ChIP signal from the values. Normalized values <0 were set to 0. Finally, the 95th quantile value of each sample was set as an upper limit. The average values of each bin were used to construct metaplots.

Identification of dACR groups by K-means clustering. For K-means clustering, we only used dACRs that had ≥70% mapping coverage (from the 75 bp simulated reads; see Definition of intergenic negative control regions) in the ± 2 kb region flanking the dACR summits. We used this filtering step to ensure that none of the dACRs analysed were directly adjacent to unmappable regions.

Normalized values of 200 10-bp bins from upstream and downstream of distal ACR summits from heatmap analysis were extracted for the histone modifications H3K27me3, H3K36me3, H3K4me3 and H3K56ac. The values were concatenated into a single matrix with 1,600 columns. Finally, using the matrix as the input, dACRs were separated into different groups by the K-means method in R (https://www.r-project.org/) with ten random sets and 30 maximum iteration cycles[59]. The number of clusters were determined by the total within-cluster sum of square and subsequently manual inspection of identified histone patterns.

Identification of gene expression tissue specificity. Gene expression tissue specificity was determined by a modified entropy formula as described previously[60]. RNA-seq raw data from 23 *Z. mays* tissues (first replicate from each tissue) were downloaded from accession number GSE50191 (ref. [61]). Raw data were processed as described in the RNA-seq raw data processing section of this publication. Transcripts per million values were used as the input to calculate an entropy value for each annotated gene.

GO enrichment analysis. GO enrichment analysis was performed using BiNGO v.3.0.3 (ref. [62]) with the *Z. mays* AGPv4 GO annotation from maizeGDB[63]. GO terms under the 'molecular function' category were used for the analyses.

eQTL analysis. To test for a significant relationship between dACRs and nucleotides identified as genetic regulators of gene expression (that is, eQTL), we quantified the enrichment of best eQTL hits (relative to all SNPs) within ACRs. First, we obtained maize eQTL from a recent study[20]. We used the union of eQTL with higher effect and lowest *P* value for each gene in the maize genome across leaf tissues[20]. The set of all SNP were obtained from the maize hapmap 3.2.1 (ref. [64]) for all taxa in the RNA-set using a minimum read count of five (the same filtering criteria applied to run the eQTL analysis). We plotted the posterior distribution of eQTL SNP frequency, relative to all SNP, using a beta-binomial distribution with a Beta(1,1) prior. To test if enrichment was present within the dACRs, we estimated the same distributions for a group of control regions that were both gene–distal and uniquely mappable (see Definition of intergenic negative control regions).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data generated from this study has been uploaded to the Gene Expression Omnibus database and can be retrieved through accession number GSE120304. Additionally, the data from this study can be viewed interactively on the publicly accessible epigenome browser http://epigenome.genetics.uga.edu/PlantEpigenome/. The STARR-seq plasmid sequence and additional information can be found at Addgene, deposit number 117379 (https://www.addgene.org/117379/).

Code availability

The code used for analyses can be accessed at https://github.com/schmitzlab/Widespread-Long-range-Cis-Regulatory-Elements-in-the-Maize-Genome/.

Received: 8 May 2019; Accepted: 9 October 2019;
Published online: 18 November 2019

References

1. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
2. Weber, B., Zicola, J., Oka, R. & Stam, M. Plant enhancers: a call for discovery. *Trends Plant Sci.* **21**, 974–987 (2016).
3. Marand, A. P., Zhang, T., Zhu, B. & Jiang, J. Towards genome-wide prediction and characterization of enhancers in plants. *Biochim. Biophys. Acta Gene Regul. Mech.* **1860**, 131–139 (2017).
4. Wallace, J. G. et al. Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genet.* **10**, e1004845 (2014).
5. Huang, C. et al. Zm*CCT9* enhances maize adaptation to higher latitudes. *Proc. Natl Acad. Sci. USA* **115**, E334–E341 (2018).
6. Salvi, S. et al. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc. Natl Acad. Sci. USA* **104**, 11376–11381 (2007).
7. Studer, A., Zhao, Q., Ross-Ibarra, J. & Doebley, J. Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat. Genet.* **43**, 1160–1163 (2011).
8. Zheng, L. et al. Prolonged expression of the BX1 signature enzyme is associated with a recombination hotspot in the benzoxazinoid gene cluster in *Zea mays*. *J. Exp. Bot.* **66**, 3917–3930 (2015).
9. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
10. Iwafuchi-Doi, M. et al. The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol. Cell* **62**, 79–91 (2016).
11. Rodgers-Melnick, E., Vera, D. L., Bass, H. W. & Buckler, E. S. Open chromatin reveals the functional maize genome. *Proc. Natl Acad. Sci. USA* **113**, E3177–E3184 (2016).
12. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
13. Lu, Z., Hofmeister, B. T., Vollmers, C., DuBois, R. M. & Schmitz, R. J. Combining ATAC-seq with nuclei sorting for discovery of *cis*-regulatory regions in plant genomes. *Nucleic Acids Res.* **45**, e41 (2017).
14. Oka, R. et al. Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biol.* **18**, 137 (2017).
15. Zhao, H. et al. Proliferation of regulatory DNA elements derived from transposable elements in the maize genome. *Plant Physiol.* **176**, 2789–2803 (2018).
16. Dong, P. et al. 3D chromatin architecture of large plant genomes determined by local A/B compartments. *Mol. Plant* **10**, 1497–1509 (2017).
17. Segal, E. et al. A genomic code for nucleosome positioning. *Nature* **442**, 772–778 (2006).
18. O'Malley, R. C. et al. Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* **166**, 1598 (2016).
19. Galli, M. et al. The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family. *Nat. Commun.* **9**, 4526 (2018).
20. Kremling, K. A. G. et al. Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* **555**, 520–523 (2018).
21. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107**, 21931–21936 (2010).
22. Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
23. Zhang, W. et al. High-resolution mapping of open chromatin in the rice genome. *Genome Res.* **22**, 151–162 (2012).
24. Zhang, W., Zhang, T., Wu, Y. & Jiang, J. Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in *Arabidopsis*. *Plant Cell* **24**, 2719–2731 (2012).
25. Zhang, X., Bernatavichute, Y. V., Cokus, S., Pellegrini, M. & Jacobsen, S. E. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol.* **10**, R62 (2009).
26. Bewick, A. J. et al. On the origin and evolutionary consequences of gene body DNA methylation. *Proc. Natl Acad. Sci. USA* **113**, 9111–9116 (2016).
27. Roudier, F. et al. Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *EMBO J.* **30**, 1928–1938 (2011).
28. Sullivan, A. M. et al. Mapping and dynamics of regulatory DNA and transcription factor networks in *A. thaliana*. *Cell Rep.* **8**, 2015–2030 (2014).
29. Zhang, X. et al. Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biol.* **5**, e129 (2007).
30. Belton, J. M. et al. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
31. Mumbach, M. R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).

32. Bhattacharyya, S., Chandra, V., Vijayanand, P. & Ferhat, A. Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nat. Commun.* **10**, 4221 (2019).
33. Arnold, C. D. et al. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074–1077 (2013).
34. Bennetzen, J. L. & Wang, X. Relationships between gene structure and genome instability in flowering plants. *Mol. Plant* **11**, 407–413 (2018).
35. Crisp, P. A., Noshay, J. M., Anderson, S. N. & Springer, N. M. Opportunities to use DNA methylation to distil functional elements in large crop genomes. *Mol. Plant* **12**, 282–284 (2019).
36. Rowley, M. J. et al. Evolutionarily conserved principles predict 3D chromatin organization. *Mol. Cell* **67**, 837–852 (2017).
37. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet*. **19**, 789–800 (2018).
38. Rowley, M. J. et al. Condensin II counteracts cohesin and RNA polymerase II in the establishment of 3D chromatin organization. *Cell Rep.* **26**, 2890–2903 (2019).
39. Urich, M. A., Nery, J. R., Lister, R., Schmitz, R. J. & Ecker, J. R. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat. Protoc.* **10**, 475–483 (2015).
40. Bartlett, A. et al. Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat. Protoc.* **12**, 1659–1672 (2017).
41. Benfey, P. N. & Chua, N. H. The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. *Science* **250**, 959–966 (1990).
42. Ow, D. W., Jacobs, J. D. & Howell, S. H. Functional regions of the cauliflower mosaic virus 35S RNA promoter determined by use of the firefly luciferase gene as a reporter of promoter activity. *Proc. Natl Acad. Sci. USA* **84**, 4870–4874 (1987).
43. Yoo, S. D., Cho, Y. H. & Sheen, J. *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572 (2007).
44. Jiao, Y. et al. Improved maize reference genome with single-molecule technologies. *Nature* **546**, 524–527 (2017).
45. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
46. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
47. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
48. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
49. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
50. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).
51. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
52. Schultz, M. D. et al. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212–216 (2015).
53. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
54. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
55. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
56. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
57. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
58. Guo, Y., Mahony, S. & Gifford, D. K. High resolution genome wide binding event finding and motif discovery reveals transcription factor spatial binding constraints. *PLoS Comput. Biol.* **8**, e1002638–e1002638 (2012).
59. Hartigan, J. A. & Wong, M. A. Algorithm AS 136: a K-means clustering algorithm. *J. R. Stat. Soc. Ser. C.* **28**, 100–108 (1979).
60. Zhang, X. et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126**, 1189–1201 (2006).
61. Walley, J. W. et al. Integration of omic networks in a developmental atlas of maize. *Science* **353**, 814–818 (2016).
62. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449 (2005).
63. Harper, L., Gardiner, J., Andorf, C. & Lawrence, C. J. in *Plant Bioinformatics: Methods and Protocols* (ed Edwards, D.) 187–202 (Springer, 2016).
64. Bukowski, R. et al. Construction of the third-generation *Zea mays* haplotype map. *Gigascience* **7**, 1–12 (2018).

Acknowledgements

This work was funded by the National Science Foundation (NSF) grant no. IOS-1546867 to R.J.S. and X.Z.; grant no. NSF IOS-1238142 to X.Z and M.J.S.; and grant no. NSF IOS-1456950 and NSF IOS-1546873 to A.G. F.J. and R.J.S. acknowledge support from the Technical University of Munich–Institute for Advanced Study funded by the German Excellent Initiative and the European Seventh Framework Programme under grant agreement no. 291763. F.J. is also supported by the SFB/Sonderforschungsbereich924 of the Deutsche Forschungsgemeinschaft. R.J.S. is a Pew Scholar in the Biomedical Sciences, supported by The Pew Charitable Trusts. M.C.-T. acknowledges support from the Impuls-und Vernetzungsfonds of the Helmholtz-Gemeinschaft (grant no. VH-NG-1219). J.Z. and his team is supported by the Programme for Guangdong Introducing Innovative and Entrepreneurial Teams (grant no. 2016ZT06S172). This work was supported in part by the National Institutes of Health Pathway to Independence Award no. K99/R00 GM127671 (M.J.R.) and the US Public Health Service Award (R01) no. GM035463 (V.G.C.) from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions

W.A.R., Z.L., M.J.R., V.G.C., J.Z., M.J.S., E.S.B., N.M.S., R.J.S. and X.Z. conceived and designed the experiments. W.A.R., Z.L., C.L.E., N.G.M., J.M.N. and M.G. performed the experiments. C.L.E., N.G.M., M.G. and A.G. performed the DAP-seq experiments. W.A.R., Z.L., L.J., A.P.M., M.K.M.-G., M.C.-T., F.J. and X.Z. performed the computational analyses. W.A.R., Z.L., L.J., A.P.M., M.K.M.-G. created the figures. W.A.R., A.P.M., R.J.S. and X.Z. wrote the manuscript. W.A.R., R.J.S. and X.Z. revised the manuscript. All authors read and approved the final manuscript.

Competing interests

R.J.S. and X.Z. are cofounders of REquest Genomics, LLC, a company that provides epigenomics services.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41477-019-0547-0.

Correspondence and requests for materials should be addressed to R.J.S. or X.Z.

Peer review information *Nature Plants* thanks Dao-Xiu Zhou and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Robert J Schmitz & Xiaoyu Zhang

Last updated by author(s): Sep 19, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Flow cytometry of nuclei was performed with a MoFlo XDP in conjunction with Summit (v6.3.1) software.
Data analysis	Trimmomatic (v0.36); HISAT2 (v2.0.5); StringTie (v1.3.3b); DESeq2; Bowtie (v1.2.2); Bowtie2(v2.1.1); Picard (v2.16.0); MACS2(v2.1.2); BEDTools (v2.25.0); BEDTools (v2.26.0); SAMtools (v1.3.1); R (v3.4.3); RStudio (v1.1.383); Methylpy (v1.3); Cutadapt (v1.9.dev1); HiC-pro pipeline (v2.8.0); Homer (v4.10.0); Juicer (v0.7.0); FitHiChIP (v5.1); GEM (v2.5)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The raw data generated for this study has been uploaded to the Gene Expression Omnibus (GEO) database and can be retrieved through accession number GSE120304. Additionally, the data from this study can be viewed interactively on the publicly accessible genome browser http://epigenome.genetics.uga.edu/PlantEpigenome/. Processed data—including ACR coordinates, differential dACRs, called chromatin loops, and dACR-gene loop intervals—are available in Supplementary Data Tables S1-S4. The STARR-seq plasmid sequence and additional information can be found at Addgene, deposit number 117379 (https://www.addgene.org/117379/), where it is available for purchase. We have also used publicly available data for some analyses. Independently produced ATAC-seq datasets (Dong et al. 2017) (used in supplementary fig. 1) are available at the NCBI SRA project accession PRJNA391551. DNase-seq datasets (Oka et al. 2017) (used in supplementary fig. 1) are available at the NCBI SRA project accession PRJNA369253. MNase-seq datasets (Rodgers-Melnick et al. 2016) (used in supplementary fig. 1) are available at the NCBI SRA project accession PRJNA297204. Data for the maize SNP analysis (fig. 1i and supplementary fig. 5b) were from the maize

HapMap 3.1.1 (Bukowski et al. 2018), accessible at NCBI BioProject PRJNA399729. Coordinates of the GWAS QTLs used in fig. 1j and supplementary fig. 5c are available as supplementary data from Wallace et al. (2014). dACR-eQTL overlap analysis (fig 1k, 4c, and supplementary fig. 5d) used eQTLs generated from raw data (Kremling et al. 2018) available at NCBI BioProject PRJNA383416. Gene expression entropy analyses (fig. 2i and supplementary fig. 3b) used data from the gene expression atlas (Walley et al. 2016), available at NCBI BioProject PRJNA217053.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The effective sample size is the number of loci studied, which are described throughout the manuscript. Generally, analyses are performed on thousands of loci.
Data exclusions	From the STARR-seq analysis, we noticed some regions of the genome that were apparently contaminated from a separate experiment that was performed in parallel. We provide the coordinates of these genomic regions, totaling ~ 1 Mb of the genome, in supplemental table S12. These genomic regions were excluded from all analyses, with the exception of fig. 5b.
Replication	All experiments, except for Hi-C, HiChIP and STARR-seq, were performed in duplicate. All assays, except for STARR-seq, were performed on the same tissues at the same developmental stages and grown in the same conditions. However, separate batches of plants were grown for separate experiments. Biological replicates were performed on separately grown batches of plants, grown on different days. All attempts at replication were successful.
Randomization	Batches of plants were grown in approximately two-fold excess and plants were randomly sampled from different locations on the growth flat. Randomly selected plants were pooled together for each replicate.
Blinding	No blinding was used since measurements were not vulnerable to observer bias. Whenever separate groups were compared (i.e. dACR chromatin groups), the same analysis pipelines were performed in parallel.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	For all antibodies, 1.5 µg antibody was incubaed with 750 µg Dynabeads® Protein A. H3: Abcam, cat # ab1791, lot # GR71822-1 H3K4me1: Abcam, cat # ab8895, lot # 889421 H3K4me3: Millipore, cat # 07-473, lot # 2839113 H3K9ac: Active Motif, cat # 61251, lot # 4812001 H3K27ac: Abcam, cat # ab4729, lot UNKNOWN H3K27me3: Millipore, cat # 07-449, lot # DAM1703508 H3K36me3: Abcam, cat # ab9050, lot # 826243 H3K56ac: Millipore, cat # 07-677-1, lot # 2514206
Validation	All antibodies used here have been validated by manufacturers. Furthermore, these antibodies have a long historical use in the ENCODE project and in plant genomics. H3K56ac, H3K27me3, and H3K27ac antibody have been independently validated with peptide array. We have also validated antibodies with computational analysis. The histone marks H3K56ac, H3K27me3,

H3K36me3, and H3K4me3 were used for cluster analysis. Gene metaplot analysis demonstrates that these marks have distinct enrichment profiles, indicating that cross-reactivity was not problematic in our experiments.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

http://epigenome.genetics.uga.edu/PlantEpigenome/ for the interactive genome browser.
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120304 (token:kbstyyswzxupdgl)

Files in database submission

chip_B73_ear_H3K27me3_rep1.fastq.bz2
chip_B73_ear_H3K27me3_rep2.fastq.bz2
chip_B73_ear_H3K36me3_rep1.fastq.bz2
chip_B73_ear_H3K36me3_rep2.fastq.bz2
chip_B73_ear_H3K4me3_rep1.fastq.bz2
chip_B73_ear_H3K4me3_rep2.fastq.bz2
chip_B73_ear_H3K56ac_rep1.fastq.bz2
chip_B73_ear_H3K56ac_rep2.fastq.bz2
chip_B73_ear_H3_rep1.fastq.bz2
chip_B73_ear_input_rep1.fastq.bz2
chip_B73_leaf_H2AZ_rep1.fastq.bz2
chip_B73_leaf_H2AZ_rep2.fastq.bz2
chip_B73_leaf_H3K27ac_rep1.fastq.bz2
chip_B73_leaf_H3K27ac_rep2.fastq.bz2
chip_B73_leaf_H3K27me3_rep1.fastq.bz2
chip_B73_leaf_H3K27me3_rep2.fastq.bz2
chip_B73_leaf_H3K36me3_rep1.fastq.bz2
chip_B73_leaf_H3K36me3_rep2.fastq.bz2
chip_B73_leaf_H3K4me1_rep1.fastq.bz2
chip_B73_leaf_H3K4me1_rep2.fastq.bz2
chip_B73_leaf_H3K4me3_rep1.fastq.bz2
chip_B73_leaf_H3K4me3_rep2.fastq.bz2
chip_B73_leaf_H3K56ac_rep1.fastq.bz2
chip_B73_leaf_H3K56ac_rep2.fastq.bz2
chip_B73_leaf_H3K9ac_rep1.fastq.bz2
chip_B73_leaf_H3K9ac_rep2.fastq.bz2
chip_B73_leaf_H3_rep1.fastq.bz2
chip_B73_leaf_H3_rep2.fastq.bz2
chip_B73_leaf_input_rep1.fastq.bz2
chip_B73_leaf_input_rep2.fastq.bz2
mapped_chip_B73_ear_H3K27me3_rep1.bed.bz2
mapped_chip_B73_ear_H3K27me3_rep2.bed.bz2
mapped_chip_B73_ear_H3K36me3_rep1.bed.bz2
mapped_chip_B73_ear_H3K36me3_rep2.bed.bz2
mapped_chip_B73_ear_H3K4me3_rep1.bed.bz2
mapped_chip_B73_ear_H3K4me3_rep2.bed.bz2
mapped_chip_B73_ear_H3K56ac_rep1.bed.bz2
mapped_chip_B73_ear_H3K56ac_rep2.bed.bz2
mapped_chip_B73_ear_H3_rep1.bed.bz2
mapped_chip_B73_ear_input_rep1.bed.bz2
mapped_chip_B73_leaf_H2AZ_rep1.bed.bz2
mapped_chip_B73_leaf_H2AZ_rep2.bed.bz2
mapped_chip_B73_leaf_H3K27ac_rep1.bed.bz2
mapped_chip_B73_leaf_H3K27ac_rep2.bed.bz2
mapped_chip_B73_leaf_H3K27me3_rep1.bed.bz2
mapped_chip_B73_leaf_H3K27me3_rep2.bed.bz2
mapped_chip_B73_leaf_H3K36me3_rep1.bed.bz2
mapped_chip_B73_leaf_H3K36me3_rep2.bed.bz2
mapped_chip_B73_leaf_H3K4me1_rep1.bed.bz2
mapped_chip_B73_leaf_H3K4me1_rep2.bed.bz2
mapped_chip_B73_leaf_H3K4me3_rep1.bed.bz2
mapped_chip_B73_leaf_H3K4me3_rep2.bed.bz2
mapped_chip_B73_leaf_H3K56ac_rep1.bed.bz2
mapped_chip_B73_leaf_H3K56ac_rep2.bed.bz2
mapped_chip_B73_leaf_H3K9ac_rep1.bed.bz2
mapped_chip_B73_leaf_H3K9ac_rep2.bed.bz2
mapped_chip_B73_leaf_H3_rep1.bed.bz2
mapped_chip_B73_leaf_H3_rep2.bed.bz2
mapped_chip_B73_leaf_input_rep1.bed.bz2
mapped_chip_B73_leaf_input_rep2.bed.bz2

Genome browser session
(e.g. UCSC)

http://epigenome.genetics.uga.edu/PlantEpigenome/

Methodology

Replicates

Replicates were biological replicates. Two replicates were performed for each histone modification and tissue (listed below).

Sequencing depth

All ChIP-seq reads were single-end 75 bp. Single all profile histone marks are euchromatic, and only a small fraction of the maize genome (a few percent) are euchromatic, the effective genome coverage is much greater than the numbers indicated here (which correspond to the global genome coverage).
sample read_length_(nt) total_reads aligned_reads %_unique_aligned coverage_(x)
B73 leaf H2AZ rep1 75 "27,732,267" "26,486,801" 95.51% 0.902959125
B73 leaf H2AZ rep2 75 "19,252,409" "18,399,538" 95.57% 0.627256977
B73 leaf H3K4me1 rep1 75 "48,198,327" "46,208,231" 95.87% 1.575280602
B73 leaf H3K4me1 rep2 75 "48,459,950" "46,294,226" 95.53% 1.57821225
B73 leaf H3K4me3 rep1 75 "14,706,000" "13,658,703" 92.88% 0.465637602
B73 leaf H3K4me3 rep2 75 "14,206,461" "11,497,814" 80.93% 0.391970932
B73 leaf H3K9ac rep1 75 "23,017,373" "22,278,839" 96.79% 0.759505875
B73 leaf H3K9ac rep2 75 "42,990,103" "41,542,119" 96.63% 1.416208602
B73 leaf H3K27ac rep1 75 "50,627,599" "49,003,001" 96.79% 1.670556852
B73 leaf H3K27ac rep2 75 "51,044,191" "49,430,997" 96.84% 1.685147625
B73 leaf H3K27me3 rep1 75 "24,095,651" "23,474,756" 97.42% 0.800275773
B73 leaf H3K27me3 rep2 75 "24,046,522" "23,478,832" 97.64% 0.800414727
B73 leaf H3K36me3 rep1 75 "49,208,879" "47,247,742" 96.01% 1.610718477
B73 leaf H3K36me3 rep2 75 "75,396,238" "72,412,977" 96.04% 2.468624216
B73 leaf H3K56ac rep1 75 "14,631,000" "13,604,127" 92.98% 0.463777057
B73 leaf H3K56ac rep2 75 "30,604,762" "29,571,854" 96.63% 1.008131386
B73 leaf H3 rep1 75 "46,542,831" "43,533,860" 93.54% 1.484108864
B73 leaf H3 rep2 75 "15,137,000" "14,418,690" 95.25% 0.49154625
B73 leaf input rep1 75 "22,716,315" "21,335,884" 93.92% 0.727359682
B73 leaf input rep2 75 "20,824,166" "19,559,465" 93.93% 0.666799943
B73 inflorescence H3K4me3 rep1 75 "42,571,199" "41,517,523" 97.52% 1.415370102
B73 inflorescence H3K4me3 rep2 75 "44,288,396" "42,644,295" 96.29% 1.453782784
B73 inflorescence H3K27me3 rep1 75 "78,963,021" "75,414,392" 95.51% 2.570945182
B73 inflorescence H3K27me3 rep2 75 "72,768,893" "70,398,382" 96.74% 2.399944841
B73 inflorescence H3K36me3 rep1 75 "41,183,792" "40,282,556" 97.81% 1.373268955
B73 inflorescence H3K36me3 rep2 75 "56,234,785" "54,357,330" 96.66% 1.853090795
B73 inflorescence H3K56ac rep1 75 "20,524,329" "19,544,866" 95.23% 0.66630225
B73 inflorescence H3K56ac rep2 75 "38,591,965" "37,530,383" 97.25% 1.279444875
B73 inflorescence H3 rep1 75 "13,843,701" "13,474,764" 97.33% 0.459366955
B73 inflorescence input rep1 75 "33,165,873" "32,379,611" 97.63% 1.103850375

Antibodies

H3: Abcam, cat # ab1791, lot # GR71822-1
H3K4me1: Abcam, cat # ab8895, lot # 889421
H3K4me3: Millipore, cat # 07-473, lot # 2839113
H3K9ac: Active Motif, cat # 61251, lot # 4812001
H3K27ac: Abcam, cat # ab4729, lot UNKNOWN
H3K27me3: Millipore, cat # 07-449, lot # DAM1703508
H3K36me3: Abcam, cat # ab9050, lot # 826243
H3K56ac: Millipore, cat # 07-677-1, lot # 2514206

Peak calling parameters

Parameter: MACS2, H2AZ/H3K27me3/H3K36me3/H3K4me1, "--nomodel --extsize 147 --broad --broad-cutoff 0.1" and FDR<0.05;
H3K9ac, H3K27ac, H3K56ac, H3K4me3, "--nomodel --extsize 147" and FDR < 0.05.

Data quality

ChIP peaks were not used for the study. The ChIP signal was only quantitatively assessed. However, we provide ChIP peaks here as a quality-control measurement. All ChIP experiments show high signal to background noise. Refer to the genome browser link to assess the ChIP quality. Called peak numbers are listed below:
98929 H2AZ_peaks.broadPeak Leaf
116168 H3K27ac_peaks.narrowPeak Leaf
48163 H3K27me3_peaks.broadPeak Leaf
42861 H3K36me3_peaks.broadPeak Leaf
71551 H3K4me1_peaks.broadPeak Leaf
51228 H3K4me3_peaks.narrowPeak Leaf
67525 H3K56ac_peaks.narrowPeak Leaf
68435 H3K9ac_peaks.narrowPeak Leaf
69347 H2AZ_peaks.broadPeak Ear
59354 H3K27me3_peaks.broadPeak Ear
45852 H3K36me3_peaks.broadPeak Ear

Software

Trimmomatic (v0.36);Bowtie (v1.1.1); Bowtie2(v2.1.1); Picard (v2.16.0); BEDTools (v2.26.0); samtools (v1.3.1); MACS2 (v2.1.2).

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Nuclei were sorted in order to purify them from other organelles prior to ATAC-seq. Approximately 200 mg freshly collected tissue was immediately chopped with a razor blade in ~ 1ml of pre-chilled lysis buffer (15 mM Tris-HCl pH 7.5, 20 mM NaCl, 80 mM KCl, 0.5 mM spermine, 5 mM 2-Mercaptoethanol, 0.2% TritonX-100). The chopped slurry was filtered twice through miracloth and once through a 40 µm filter. The crude nuclei were stained with DAPI and loaded into the flow cytometer.
Instrument	Beckman Coulter MoFlo XDP
Software	Summit version 6.3.1
Cell population abundance	For each library preparation, 50,000 nuclei were used.
Gating strategy	There are multiple DAPI signal peaks with high quality nuclei, reflecting the copy number of plant genomes. The nuclei with DAPI signal >= that of 2×n genomes were collected for the ATAC-seq.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.