

RESEARCH ARTICLE SUMMARY

PLANT GENETICS

The genetic architecture of cell type-specific cis regulation in maize

Alexandre P. Marand*, Luguang Jiang, Fabio Gomez-Cano, Mark A. A. Minow, Xuan Zhang, John P. Mendieta, Ziliang Luo, Sohyun Bang, Haidong Yan, Cullan Meyer, Luca Schlegel, Frank Johannes, Robert J. Schmitz*

INTRODUCTION: Noncoding genetic variants are a major driver of phenotypic diversity. In *Zea mays* (maize), genetic variation within cis-regulatory regions accounts for ~40% of phenotypic variability in agronomically important traits. Missing from past studies, however, is the role of cell context in shaping regulatory variant effects and a clear definition of the molecular mechanisms underlying phenotypic variation. Thus, resolving the genetic and molecular principles that give rise to phenotypic diversity in a cell state-aware framework is paramount to advancing crop improvement efforts.

RATIONALE: Single-cell genomic methods offer a powerful approach for understanding the genetic sources of gene expression and chromatin

accessibility variation. We generated and analyzed single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq) and single-nuclei RNA sequencing (snRNA-seq) data across 172 genetically and phenotypically diverse inbred maize lines for insight into the regulatory mechanisms underlying phenotypic variability.

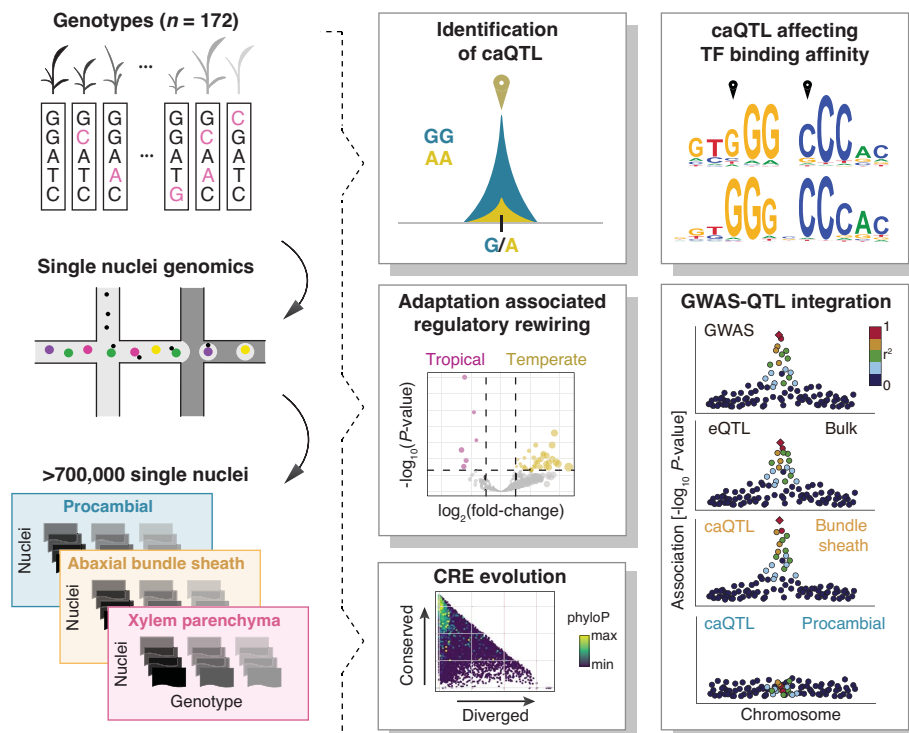
RESULTS: Our single-cell dataset comprises >700,000 nuclei from 33 distinct cell states. We first used this resource to investigate the extent of cis-regulatory variation among diverse genetic backgrounds, identifying binding sites for specific transcription factor (TF) families as being critical determinants of regulatory sequence conservation and functional activity. Through comparisons with 21 teosinte

(a progenitor of modern maize) genomes, we identified 1587 accessible chromatin regions that were unique and fixed in the domesticated maize lineage. These accessible chromatin regions were enriched for *hAT* and *PIF/Harbinger* transposons, implicating co-option of transposon cis-regulatory elements as a major source of new regulatory sequences specific to domesticated maize.

By applying the principles of population genetics, we identified 107,623 cis chromatin accessibility quantitative trait loci (cis-caQTL) within accessible chromatin regions and validated their effects on enhancer activity using self-transcribing active regulatory region sequencing (STARR-seq). We found that cell state-specific cis-caQTL are common and often overlap with phenotype-associated variants identified by genome-wide association studies (GWASs). Deep investigation of caQTL indicated that variants within TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR (TCP)-binding sites are strong determinants of chromatin accessibility. Moreover, all caQTL variants with decreased TCP-binding affinity were concomitant with loss of chromatin accessibility. Construction of cell state-specific gene-regulatory networks indicated that TCP TFs are highly cell state specific and controlled by master cell identity regulators. Analysis of caQTL affecting distal accessible chromatin regions further implicated TCP TFs as being major contributors toward chromatin accessibility variation and, as a result, chromatin interactions.

Through transcriptome-wide association mapping, chromatin accessibility-wide association mapping, and integration of caQTL with expression QTL and GWAS variants, we found that caQTL were commonly associated with flowering-related phenotypes. We hypothesized that the transition of modern maize from tropical to temperate climates may have been a significant contributor of extant chromatin accessibility variation. Indeed, we found that caQTL are associated with signatures of population differentiation, and that these population-divergent variants occur within binding sites for TFs previously implicated in flowering time and floral morphology nonuniformly among cell states.

CONCLUSIONS: These analyses advance our understanding of how cell context and molecular diversity contribute to innovations in phenotype, providing the blueprints for future crop improvement efforts. ■



Single-nuclei caQTL mapping and investigation of cell context-specific effects. scATAC-seq profiles from >700,000 single nuclei across 172 maize inbred lines were used to identify >100,000 cis-caQTL within accessible chromatin regions. This study uncovered TF-binding site perturbations with effects on chromatin accessibility, caQTL colocalized with GWAS and eQTL, regulatory evolution associated with maize domestication, and local adaptation-associated regulatory rewiring at cell type resolution.

The list of author affiliations is available in the full article online.

*Corresponding author. Email: amarand@umich.edu (A.P.M.); schmitz@uga.edu (R.J.S.)

Cite this article as A. P. Marand et al., *Science* 388, eads6601 (2025). DOI: 10.1126/science.ads6601

READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.ads6601>

RESEARCH ARTICLE

PLANT GENETICS

The genetic architecture of cell type-specific cis regulation in maize

Alexandre P. Marand^{1†*}, Luguang Jiang², Fabio Gomez-Cano², Mark A. A. Minow¹, Xuan Zhang¹, John P. Mendieta¹, Ziliang Luo¹, Sohyun Bang³, Haidong Yan^{1†}, Cullan Meyer¹, Luca Schlegel⁴, Frank Johannes⁴, Robert J. Schmitz^{1*}

Gene expression and complex phenotypes are determined by the activity of cis-regulatory elements. However, an understanding of how extant genetic variants affect cis regulation remains limited. Here, we investigated the consequences of cis-regulatory diversity using single-cell genomics of more than 0.7 million nuclei across 172 *Zea mays* (maize) inbreds. Our analyses pinpointed cis-regulatory elements distinct to domesticated maize and revealed how historical transposon activity has shaped the cis-regulatory landscape. Leveraging population genetics principles, we fine-mapped about 22,000 chromatin accessibility-associated genetic variants with widespread cell type-specific effects. Variants in TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR-binding sites were the most prevalent determinants of chromatin accessibility. Finally, integrating chromatin accessibility-associated variants, organismal trait variation, and population differentiation revealed how local adaptation has rewired regulatory networks in unique cellular contexts to alter maize flowering.

Understanding the genetic and molecular principles underlying quantitative trait variation is paramount to improving crop yield and climate resiliency. Toward this goal, genome-wide association studies (GWASs) of genetically distinct inbred *Zea mays* (maize) diversity panels have identified thousands of genetic variants associated with agriculturally important characteristics (1). However, because of linkage disequilibrium (LD), disentangling large numbers of linked genetic variants to derive causal effects remains challenging. In maize, trait-associated variants are mostly nongenic and lie ~16.4 kb from the nearest gene, complicating efforts to infer the molecular mechanisms contributing to phenotypic diversity (2). Thus, complementary approaches are required to pinpoint causal GWAS variants.

Variation in the patterns and magnitude of gene expression is a substantial contributor to quantitative phenotypes. The spatiotemporal patterning and kinetics of gene expression are orchestrated by cis-regulatory elements (CREs), short DNA sequence motifs bound by sequence-specific transcription factors (TFs). Investigation of CREs has been useful in furthering the understanding of the molecular determinants

behind phenotypic manifestation (3–5). Indeed, recent studies have illustrated that CRE sequence evolution leads to morphological innovation and species divergence (6, 7). Although some CRE variants have been associated with phenotypic changes (8–13), the mechanistic links between CRE variation and phenotypic diversity remain unclear. Further complicating interpretations, different cell types are hallmarked by differential chromatin accessibility, which licenses CRE activity. As a result, how regulatory genetic variants affect molecular processes in a cell type-dependent manner remains unexplored. Thus, defining the cellular contexts in which GWAS CRE variants act is a powerful avenue to understanding polygenic trait manifestation (14).

Compared with humans ($r^2 = 0.2$ at 50 kb) (15) and mice ($r^2 = 0.2$ at 200 kb) (16), linkage disequilibrium (LD) in the maize Goodman-Buckler diversity panel (17) is substantially lower ($r^2 = 0.2$ at ~1.4 kb; fig. S1), presenting a unique opportunity to uncover the causal variants associated with molecular phenotypes. Taking advantage of this experimental system, we performed single-nuclei chromatin accessibility profiling in a diverse panel of 172 geographically distributed maize inbreds. We established the TF-binding sites (TFBSs) contributing to accessible chromatin region (ACR) sequence conservation within maize and identified transposon-derived CREs that shaped the regulatory landscape of domesticated maize. Exploiting the quantitative nature of chromatin accessibility, we identified >100,000 single-nucleotide variants (SNVs) within ACRs associated with chromatin accessibility variability (chromatin accessibility quantitative trait loci,

caQTL) and fine-mapped >22,000 SNVs at cell-type resolution. Comprehensive investigation of this dataset uncovered genetic variants within TF footprints for TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR (TCP) as being the most significant determinants of chromatin accessibility and putative chromatin interactions. Finally, by integrating cell state-resolved caQTL with GWAS variants, signatures of population differentiation, and TFBSs, we shed light on the context-dependent regulatory mechanisms that enabled tropical maize to adapt to temperate climates.

Widespread intraspecific cis-regulatory variation

To establish the impact of genetic variation on cis-regulatory function at single-cell resolution in maize, we profiled chromatin accessibility from ~1.37 million seedling nuclei across 172 distinct maize inbred lines grown under identical conditions using single-cell assay for transposase accessible chromatin sequencing (scATAC-seq) with pooling, replication (average of approximately three replicates per genotype), and an in silico genotyping framework (Fig. 1, A to C; figs. S2 and S3; and tables S1 and S2). After strict thresholding for genotype calls, we retained 682,794 high-quality nuclei (average of 3970 nuclei per inbred) and identified a total of 108,843 ACRs (Fig. 1C, fig. S3, and table S3). Variant calls and chromatin accessibility profiles were consistent with prior genotyping (average genetic correlation = 0.91) and among replicated genotypes (average Spearman's correlation coefficient = 0.92), respectively (fig. S3).

We adopted a reference mapping approach to ensure robust cell type identification (fig. S4, A to C). Specifically, a reference set of high-quality nuclei ($n = 82,283$) from all 172 genotypes were clustered into 52 distinct groups with similar genome-wide chromatin accessibility profiles, providing a reference embedding to project the remaining nuclei (Fig. 1D). To assign cell identity to each cluster, we estimated gene chromatin accessibility scores for all nuclei and assessed chromatin accessibility dynamics for known marker genes using an iterative strategy composed of (i) automated annotation based on cell type and cell cycle-stage enrichment scores, (ii) focused cell type annotation refinement by evaluating differentially accessible chromatin of known cell type and spatial marker genes, and (iii) visual inspection of marker gene enrichment (Fig. 1E and fig. S4D). The final cell type-, cell cycle stage-, and marker-based spatial labels (hereafter referred to as cell state, $n = 33$) were supported through an integration with 30,305 nuclei generated from single-nucleus RNA sequencing (snRNA-seq) (fig. S5), with cell state proportions generally consistent among genotypes (fig. S6).

Genome-wide documentation of intraspecific cis-regulatory variation is limited (18). To understand the relationship between genetic

¹Department of Genetics, University of Georgia, Athens, GA, USA. ²Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI, USA.

³Institute of Bioinformatics, University of Georgia, Athens, GA, USA. ⁴Plant Epigenomics, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Munich, Germany.

*Corresponding author. Email: amarand@umich.edu (A.P.M.); schmitz@uga.edu (R.J.S.)

†Present address: Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI, USA.

‡Present address: College of Grassland Science and Technology, Sichuan Agricultural University, Chengdu, Sichuan, China.

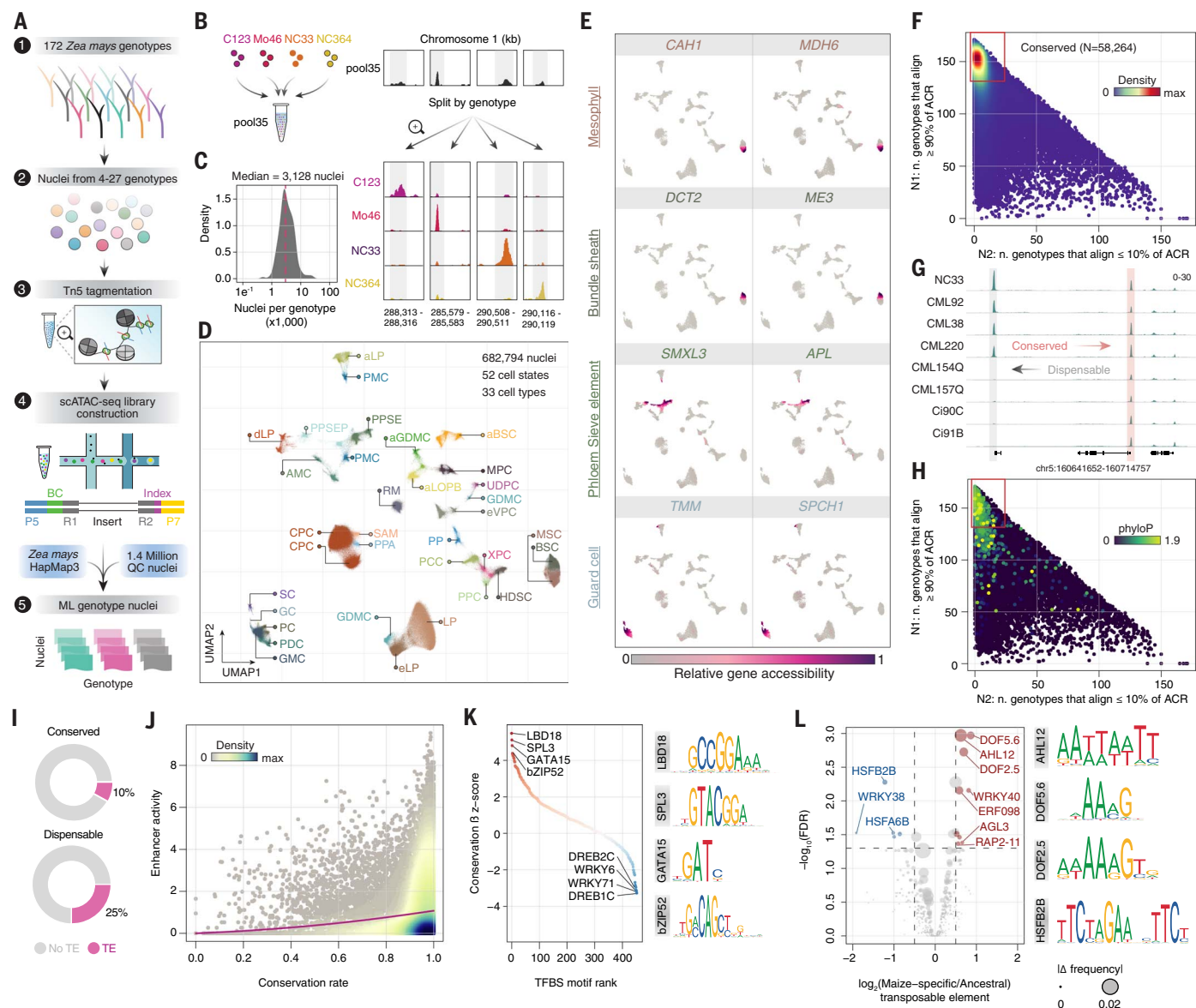


Fig. 1. Widespread cis-regulatory sequence variation. (A) Experimental and computational workflow. (B) Pooling strategy and genotype deconvolution. (C) Distribution of nuclei counts per genotype. (D) Uniform manifold approximation and projection (UMAP) visualization of chromatin accessibility variance among nuclei from diverse maize genotypes. (E) Visualization of chromatin accessibility variance for eight major cell type-specific marker genes. (F) Scatterplot illustrating the number of genotypes with >90% coverage from whole-genome resequencing of an ACR (y axis) versus the number of genotypes with <10% coverage of an ACR (x axis). Points are colored by density. (G) Genome browser examples of conserved and dispensable ACRs identified from (F). (H) Same as (F), except ACRs are colored by weighted average

phyloP scores among grass species. (I) Pie charts comparing the proportion of ACRs overlapping transposable elements from the conserved and dispensable classifications. (J) Scatterplot of ACR conservation rate (x axis) and enhancer activity (y axis) as estimated by STARR-seq. Points are colored by density. (K) Ranked scatterplot of permutation-normalized TFBS effects toward ACR conservation rate adjacent to exemplary TF motifs. (L) Bubble plot of the \log_2 transformation of the ratio between the fraction of maize-specific TE-ACRs and the fraction of ancestrally conserved TE-ACRs containing a given TF motif (x axis) against the \log_{10} -transformed FDR. The size of the bubble reflects the absolute difference in TE-ACR frequency for a given TF motif. Exemplary TF motifs are illustrated on the right.

variation and ACR conservation, we evaluated sequence conservation of 82,098 noncoding ACRs in the context of the B73 reference genome and found 58,264 (71%) ACR sequences conserved among genotypes (Fig. 1, F and G). The remaining dispensable ACRs were associated with significantly lower evolutionary conservation rates (phyloP; 5.74-fold depletion,

Wilcoxon rank sum test, $P < 2.2 \times 10^{-16}$, over-represented in intergenic and gene-proximal regions, and overlapped significantly with certain repetitive sequences (Fisher's exact test, $P < 2.2 \times 10^{-16}$; Fig. 1, H and I; fig. S7, A and B; and table S3). Showcasing the functional importance of conserved ACRs, we found a significant positive relationship between conservation

rate and self-transcribing active regulatory region sequencing (STARR-seq) enhancer activity (likelihood ratio test, $P < 2.2 \times 10^{-16}$; Fig. 1J) (19). We hypothesized that selective constraints on the activities of specific TFBSs could underlie ACR conservation. Indeed, a Monte Carlo permutation ridge regression test integrating resequencing information associated the

occurrence of 64 TFBS motifs with elevated ACR sequence conservation [false discovery rate (FDR) < 0.05; Fig. 1K and table S4]. Most of these highly conserved TFBS motifs recruit critical developmental regulators, including TFs from the *LATERAL ORGAN BOUNDARY DOMAIN (LBD)* (20), *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL)* (21), and *GATA* gene families (22). TFBS motifs from the environmentally responsive *WRKY* and *DEHYDRATION RESPONSE ELEMENT BINDING (DREB)* gene families exhibited strong negative association with conservation, consistent with stimulatory response CREs contributing to divergence and local adaptation (Fig. 1K) (23, 24).

Domestication of maize from its wild progenitor, *Z. mays* ssp. *parviglumis* (teosinte), was partially driven by human selection of extant cis-regulatory variants (10, 13, 25). To identify CREs underlying the transition to cultivated maize, we first isolated ACRs ($n = 9174$) lacking sequence alignments in 21 teosinte genotypes (table S3). Consistent with a recent origin, domesticated maize-specific ACRs were enriched for dispensability (Fisher's exact test, $P < 2.2 \times 10^{-16}$), lower enhancer activity (Wilcoxon rank sum test, $P < 2.2 \times 10^{-16}$), and increased overlap with repetitive elements (Fisher's exact test, $P < 2.2 \times 10^{-16}$; fig. S7, C to E). Many domestication-associated regulatory variants were derived from transposable element (TE) sequences (10, 13, 26, 27). Therefore, we investigated whether proliferation of certain TE families could have contributed to domestication by rewiring TF targets. Of the 9174 ACRs specific to domesticated maize, 1587 were fixed across all inbreds. We found a significant enrichment of *hAT* and *PIF/Harbinger* DNA TE families within 483 fixed ACRs postdating teosinte divergence (chi-square test; FDR < 0.05; fig. S7F). The TEs overlapping domesticated maize-specific ACRs contained TFBSs corresponding to DNA-BINDING WITH ONE FINGER (DOF) (28), ETHYLENE RESPONSE FACTOR (ERF) (29), and SEPALLATA (SEP) (30) TFs, which regulate vasculature and floral development (Fig. 1L). *PIF/Harbinger* insertions are known to affect the floral transition by modulating the vascular expression of *CONSTANS*, *CONSTANS-LIKE*, and *TIMING OF CAB1 (CCT)* clock genes (8, 13), consistent with our genome-wide results. These results highlight that co-option of *hAT* and *PIF/Harbinger* CREs may have played a widespread role in rewiring the regulatory landscape of domesticated maize.

Genetic determinants of chromatin accessibility

Single-cell data allow the investigation of cis-regulatory variation at cell type resolution. We performed caQTL mapping by modeling normalized chromatin accessibility as a function of SNV allelic dosage and a suite of latent and measured covariates. We uncovered 4.6 million cis-caQTL associated with 23,959 and

23,123 ACRs (hereafter referred to as caQTL-ACRs; dynamic beta-adjusted FDR < 0.1 to 0.01) at the bulk tissue and cell state-resolved scales, respectively (Fig. 2, A to C, and fig. S8). As expected, we found that SNVs within ACRs (22%; 107,623/481,947) were enriched for caQTL (Fisher's exact test, odds ratio = 1.93, $P < 2.2 \times 10^{-16}$; fig. S8F). Several caQTL-ACRs represented known CRE variants with trait effects, including those regulating *ZmCCT9* (8), *ZmRAP2.7* (9), *gtl* (12), and *ZmRAVL1* (11) (fig. S9A). Additionally, bulk caQTL overlapped significantly (26%; chi-square test, $P < 2.2 \times 10^{-16}$) with previously identified expression QTL (eQTL) (31) with ~85% effect direction concordance (Fig. 2, D and E, and fig. S9, B to F). Most alternate alleles (76%) were associated with both reduced chromatin accessibility and transcription, indicating that most alternate allele caQTL disrupt transcription-activating CREs (Fig. 2E). However, we also found matched caQTL-eQTL variants associated with increased chromatin accessibility and transcript abundance. These variants were enriched within polycomb response elements recognized by the BASIC PENTACYSTEINE (BPC) family of TFs (BPC5, BPC6, and BPC1) relative to background (32, 33), suggesting that variants inhibiting BPC-based polycomb repressive complex 2 targeting occur frequently and increase local chromatin accessibility and gene expression (Fisher's exact test, $P < 7.1 \times 10^{-5}$; fig. S10A).

We posited that cellular heterogeneity in bulk tissue masks caQTL from rare cellular contexts. Indeed, 44% of caQTL-ACRs identified by cell state-resolved caQTL mapping were not found in bulk tissue (fig. S10B). These caQTL-ACRs were associated with greater chromatin accessibility specificity (Wilcoxon rank sum test, $P < 3.5 \times 10^{-12}$; fig. S10C), suggesting that many caQTL are cell state specific. To determine the prevalence of caQTL specificity, we explicitly modeled caQTL across cellular contexts through multivariate adaptive shrinkage (34). We found that 53% of cell state-resolved caQTL were restricted to three or fewer cell states, in contrast to ~1.2% of caQTL shared across all cellular contexts (fig. S10, D and E). caQTL identified in select contexts and less abundant cell states were more frequently missed in bulk tissue (fig. S10, F and G). Cell state-specific caQTL also exhibited more frequent changes (11%) in effect directions between cell states, unlike shared and bulk tissue caQTL (<1%; fig. S10, H to J). Highlighting the utility of single-cell profiling, we found that cell state-specific caQTL-ACRs often overlapped with variants associated with complex traits (Fig. 2, F and G). For example, in the last intron of *light harvesting complex a2*, we identified a mesophyll-specific caQTL-ACR missed in bulk tissue in LD ($r^2 = 0.6$) with a GWAS hit for plant height and embedded within a selective sweep in temperate nonstiff stalk germplasm

(Fig. 2, H and I) (35). These results highlight that cell state-specific caQTL are common in maize and can underpin trait variance.

caQTL perturb TFBSs

A total of 107,623 caQTL were physically located within caQTL-ACRs, leading us to hypothesize that causal caQTL perturb TFBSs. Focusing on cell state-resolved caQTL-ACRs, we first fine-mapped 22,053 putatively causal caQTL (95% credible sets; average of 3054 per cell state; fig. S11A and table S5). Next, within each cellular context, we identified differential TF footprints between the two most frequent haplotypes, reflecting alleles with differential TF occupancy (Fig. 2J). This analysis revealed an average of ~25,000 allele-specific TF footprints per cell state, including examples of allele- and cell state-specific chromatin accessibility over TF footprints (Fig. 2K and fig. S11A). Consistent with functional regulatory elements, TF footprints had higher STARR-seq enhancer activity compared with nearby controls (50-bp shift; Wilcoxon rank sum test, $P < 6.9 \times 10^{-101}$; fig. S11B). Moreover, nearly half (47%) of fine-mapped caQTL were in TF footprints, a 50-fold enrichment compared with randomized control regions (Monte Carlo permutation, $P < 1 \times 10^{-4}$; fig. S11C). To validate the effects of the fine-mapped variants within TF footprints, we used STARR-seq data to quantify enhancer activity from both reference and alternate alleles. Fine-mapped TF footprint variants had significantly greater absolute effect sizes on STARR-seq enhancer activity compared with non-caQTL variants (Wilcoxon rank sum test, $P < 2.6 \times 10^{-73}$; fig. S11D). We also found that TF footprint alleles associated with negative-effect caQTL had a greater frequency of decreased enhancer activity (chi-square test, $P < 9.8 \times 10^{-8}$; Fig. 2L). Cell state-resolved analysis of TF footprint enrichment was consistent with known TF usage among cell states; footprint enrichment was observed for REVEILLE1 (RVE1) in mesophyll (36), DNA-BINDING WITH ONE FINGER (DOF) in vasculature (37), KANADI (KAN) in abaxial bundle sheath (38), and E2F in dividing leaf primordia cells (fig. S11, E to G) (39). We observed coincidence of cell state specificity among fine-mapped caQTL, differential TF footprints, and chromatin accessibility, suggesting that variants with cell state-specific effects precisely alter chromatin accessibility by perturbing cell state-specific TF binding (Monte Carlo permutation, $P < 1 \times 10^{-4}$; fig. S11H). To test this, we assessed the predicted allelic effects of fine-mapped SNVs on TF-binding affinity. Fine-mapped caQTL were associated with significantly greater absolute allele effects compared with non-caQTL SNVs (Monte Carlo permutation, $P < 1 \times 10^{-4}$), indicating that putatively causal caQTL affect chromatin accessibility by influencing TF-binding affinity (fig. S11, I and J).

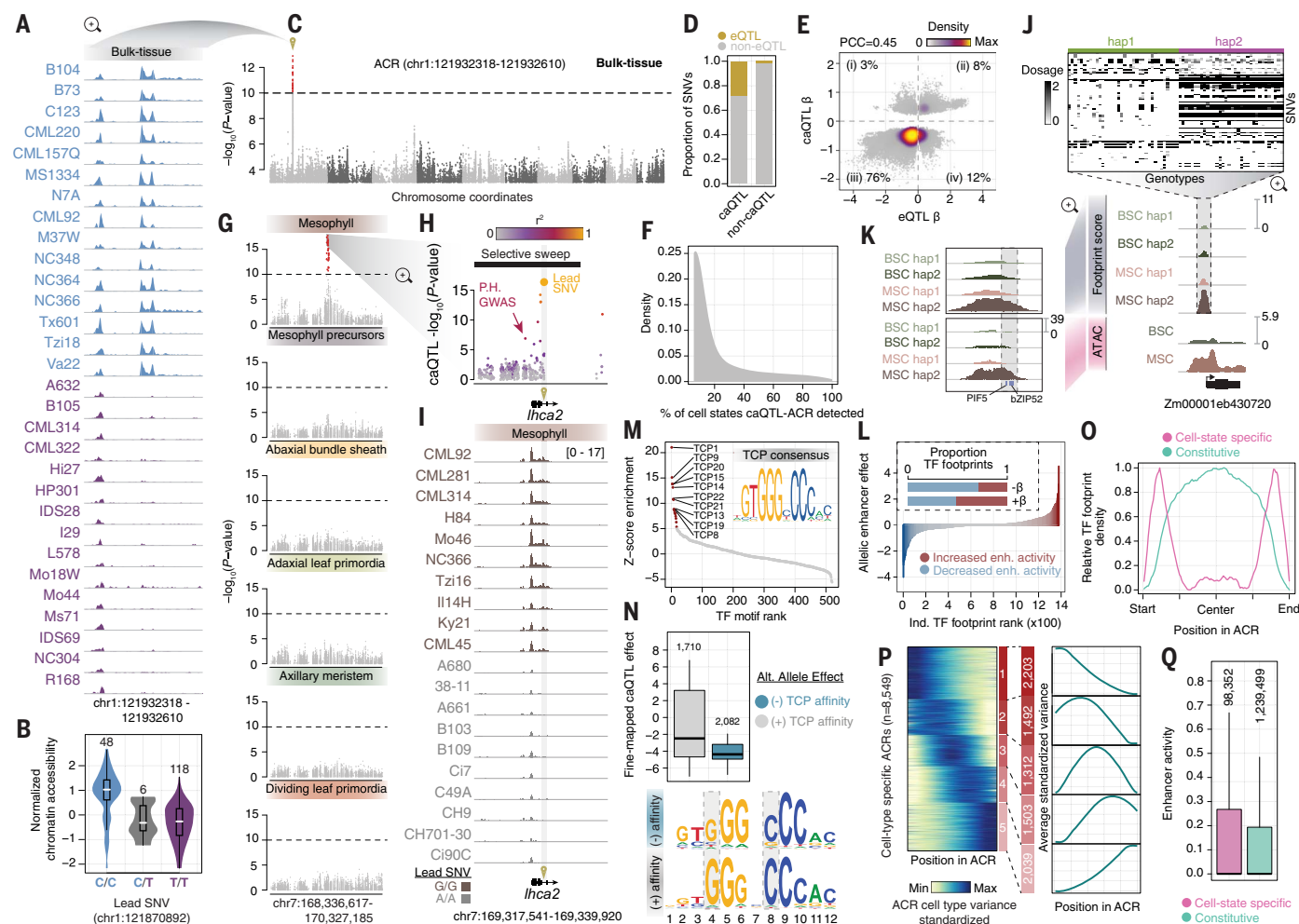


Fig. 2. Genetic determinants of chromatin accessibility variation.

(A) Genome browser view of a caQTL-ACR in 30 exemplary bulk tissue aggregates colored by the lead SNV genotype in (B). (B) Violin plot of the distribution of normalized chromatin accessibility for three caQTL genotypes associated with the ACR in (A). (C) Manhattan plot of SNV significance by genomic position for a single ACR phenotype in pseudobulk tissue aggregates. (D) Coincidence of caQTL and eQTL between the closest ACR gene pairs from bulk tissue association mapping. (E) Comparison of caQTL and eQTL effect sizes and directions for SNVs from matched ACR gene pairs. The percentages of variants are denoted in each quadrant. (F) Distribution of caQTL-ACRs detection across cell states. (G) Manhattan plot for a cell state-specific caQTL. (H) Variant association plot versus genomic position for the ACR shown in (I). Points are colored by linkage disequilibrium (r^2) relative to the lead variant. A previous GWAS hit for plant height is shown by the red arrow. The location of a selective sweep is shown as a black bar. (I) Genome browser view of 20 exemplary

We hypothesized that only select TFBS variants affect chromatin accessibility. To investigate this, we compared the enrichment of TF footprint motifs between fine-mapped caQTL and non-caQTL SNVs. Fine-mapped caQTL within TFBSs recognized by TCP family members had the strongest effects (Monte Carlo permutation, $FDR < 1 \times 10^{-5}$; Fig. 2M). Moreover, fine-mapped alleles with reduced TCP-binding affinity, particularly at the 4th and 8th base positions, were exclusively associated

with strong decreases in chromatin accessibility (Wilcoxon rank sum test, $P < 5.6 \times 10^{-8}$; Fig. 2N). Investigating the expression patterns of TCP TFs revealed greater cell state specificity than for other TFs, suggesting that distinct TCP TFs promote accessible chromatin in unique cellular contexts (fig. S12, A and B). We reasoned that characterizing the regulation of TCP TFs could provide additional context for their role in promoting accessible chromatin. To this end, we generated TF-ACR gene-

genotypes at the *lhca2* locus aligned with the genomic coordinates of (H). Tn5 integrations are scaled per million. (J) Example of haplotype clustering and TF footprinting in bundle sheath and mesophyll cells. (K) Magnification of allele-specific footprint scores (top) and chromatin accessibility (bottom). (L) Allelic difference in enhancer activity for 1379 TF footprints. Inset: Proportion of TF footprints with decreased or increased enhancer activity grouped by caQTL effect direction. (M) TF footprint motifs ranked by enrichment with fine-mapped caQTL. Red dots indicate $FDR < 1 \times 10^{-5}$. (N) Top: Distribution minus outliers of fine-mapped caQTL effects for SNVs increasing and decreasing TCP-binding affinity. Bottom: Motif logos from variants predicted to increase and decrease TCP-binding affinity. (O) Distribution of cell state-specific and constitutive TF footprints. (P) Heatmap illustrating standardized ACR variance across cell types clustered by k-means. The average cluster standardized variances are shown to the right. (Q) Distribution minus outliers of enhancer activity for cell state-specific and constitutive TF footprints.

regulatory networks for each cell state using metacells from the integrated scATAC-seq/snRNA-seq embedding (fig. S12C). As expected, unique cell states were hallmarked by distinct core TFs based on centrality scores, such as GLOSSY3 in protoderm, GOLDEN2-LIKE TFs in mesophyll, and DOF TFs in procambial cells (fig. S12D). Although the most connected TFs in each cell state were generally non-TCPs, 90% (18/20) of cell states had at least one TCP TF in the top 10% of TFs ranked by centrality

(fig. S12E). Additionally, the centrality scores of these TCPs were highly specific to the cognate cell state (fig. S12F). Finally, we found that TCPs are regulated by TFs with significantly greater numbers of targets relative to other TF genes, suggesting that cell state-specific TCP transcription is coordinated by master regulators of cell identity (Wilcoxon rank sum test, $P < 2.24 \times 10^{-35}$; fig. S12G).

Next, we noted that TF footprints with cell specificity tended to occur on ACR edges (Fig. 2K). Indeed, cell state-specific TF footprints were biased toward ACR boundaries compared with TF footprints present across multiple cell states globally (Fig. 2O). To measure chromatin accessibility dynamism at ACR boundaries, we estimated and k-means clustered the cell state variance in chromatin accessibility along the length of the ACR sequence for each cell state-specific ACR ($n = 8549$). Approximately 86% of cell state-specific ACRs were most variable in boundary regions (clusters 1, 2, 4, and 5; Fig. 2P). Quantifying TF footprint frequencies spatially within ACRs (fig. S13A) revealed that the collection of TF footprint motifs in ACR peripheries are highly unique to each cell state (fig. S13B). These boundary TFBSs had greater STARR-seq enhancer activity compared with constitutively accessible TFBSs (Wilcoxon rank sum test, $P < 1.2 \times 10^{-12}$; Fig. 2Q), indicating that the dynamic ACR boundaries unveil potent cell state-specific CREs. Consistent across cell states, the top 20 most enriched TF footprints within ACR centers all corresponded to TCP TFs (fig. S13C). Taken together, these results suggest that ACRs are licensed or maintained by TCPs, whereas dynamic accessible chromatin boundaries are influenced by other cell state-specific TFs.

Genetic determinants of long-range chromatin interactions

More than 56% of bulk-scale caQTL were associated with more than one ACR (fig. S14A), suggesting that a single caQTL might affect spatial chromatin interactions. Similar to past studies in humans (40), we identified 29,281 long-range caQTL (29,293 ACR-ACR linkages), defined as a caQTL variant associated with chromatin accessibility changes at more than one ACR (Fig. 3A). On average, long-range caQTL were associated with 3.1 ACRs and commonly linked gene-distal regions (fig. S14, B and C). Relative to randomly permuted caQTL-ACR links, long-range caQTL were enriched for co-accessible ACRs (Monte Carlo permutation, $P < 1 \times 10^{-4}$) and predicted chromatin interactions based on a deep learning convolution neural network (Monte Carlo permutation, $P < 1 \times 10^{-4}$) (41) and chromatin conformation capture sequencing (HiC) and chromatin conformation capture coupled to chromatin immunoprecipitation (HiChIP) chromatin loops (Monte Carlo permutation, $P < 1 \times 10^{-4}$; Fig. 3, B and C, and

fig. S14D). For example, we identified a caQTL within the promoter of *abscisic acid insensitive 8* (*abi8*), an ABI-VP1 TF with a hypothesized negative regulatory role in leaf blade morphogenesis (42). The *abi8* promoter variant was associated with ACRs 10 and 50 kb downstream coincident with H3K4me3- and H3K27me3-HiChIP chromatin loops, respectively (Fig. 3D). Additionally, long-range caQTL with fine-mapped variants showed stronger overlap with chromatin loops than those lacking fine-mapped caQTL (Fisher's exact test, $P < 1.49 \times 10^{-6}$). These long-range fine-mapped caQTL exhibited larger effect sizes and smaller SEs than non-fine-mapped caQTL, indicating that large-effect caQTL are better powered to detect distal chromatin interactions (fig. S14E). As a result, the number of long-range QTL in this study is likely underestimated.

We then aimed to understand the regulatory basis of long-range caQTL. To do so, we compared TFBSs found in ACRs of long-range interactions supported by chromatin loops that colocalized with either fine-mapped caQTL or SNVs with nonsignificant effects on chromatin accessibility (Fig. 3E). Like variants that affect local chromatin accessibility, caQTL affecting long-range interactions were enriched for TCP family TFBSs (Fig. 3E). For example, a local caQTL affecting a predicted TCP-binding site in the promoter of the maize homolog of *DELAY OF GERMINATION 1* (*DOG1*) was associated with accessible chromatin variation of two distal co-accessible ACRs 88 and 90 kb upstream (Fig. 3F). Investigating long-range caQTL further, we found that long-range caQTL within intergenic ACRs were more prevalent than other genomic contexts (chi-square test, $P < 1.7 \times 10^{-221}$), associated with a greater number of interactions (interactive capacity), and had stronger STARR-seq enhancer activity than ACRs near genes (Wilcoxon rank sum test, $P < 2.0 \times 10^{-28}$; Fig. 3, G and H, and fig. S14F). To evaluate whether specific TFBSs affect the number of interactions, we fit a generalized linear model of interactive capacity as a function of motif occurrences within long-range caQTL-ACRs and compared the model coefficients with 100 random permutations. TFBSs from the TCP family were strongly associated with increased interactive capacity (Fig. 3I). These results implicate gene-distal enhancers containing TCP-binding sites as potential determinants of long-range chromatin loops.

Identifying the trans regulators driving chromatin interactions remains challenging in plants. As an orthogonal approach, we reasoned that co-accessibility scores between distal ACRs could be used as a quantitative proxy for chromatin interactions. Using co-accessibility scores as a response variable, we performed QTL mapping for co-accessible ACRs supported by HiC and/or HiChIP loops ($n = 802,504$). This analysis revealed 12 loci associated with genome-wide co-accessibility (Fig. 3J), including six

transcription factors, three chromatin-associated proteins, two phosphosignaling proteins, and a cell cycle-associated gene (table S6). We identified a haplotype containing *tcp-transcription factor 36* (*tcp36*) on chromosome 10 as being one of the top candidates associated with variation in genome-wide ACR co-accessibility (fig. S14G). The coding sequence for *tcp36* contained several missense mutations and a premature stop codon (fig. S14H). In addition to identifying interesting candidate genes for future exploration, these results provide support for the association between TCP TFs and global chromatin accessibility variation.

Selection on regulatory variants shaped complex traits

To understand the relative contribution of regulatory variation toward organismal phenotypes and to define the regulatory modules underlying complex traits, we performed chromatin accessibility genome-wide association (CAWA) and transcriptome-wide association (TWA) mapping. As expected, phenotypes clustered by physiological relationships and trait associations captured known gene-trait interactions (Fig. 4A). For example, transcript abundances and regulatory region chromatin accessibilities for *LONELY GUY 7* (*ZmLOG7*), *CUP-SHAPED COTYLEDON 2* (*ZmCUC2*), and *knotted 1* (*kn1*) were associated with shoot apical meristem (SAM) morphology (43); leaf width (44); and ear, tassel, and leaf architecture (45), respectively. CAWA deconvolution identified the TFBSs with the largest impact on organismal traits: SPL and homeodomain motifs associated with SAM and ear morphology, WRKY motifs with flowering time traits, and AUXIN RESPONSE FACTOR (ARF) motifs with leaf width (fig. S15A).

Considering that complex traits are the synergistic outcome of cellular phenotypes, we compared the percentage of trait variation explained by cell state-resolved fine-mapped caQTL relative to permutations of non-caQTL SNVs. Fine-mapped caQTL explained a significant proportion of trait variance for 96% (47/49) of complex traits (Fig. 4B). Moreover, organismal traits were often influenced by multiple cell states at varying degrees, indicating that cell-distinct processes affect phenotypic variation nonuniformly (Fig. 4B). Highlighting the utility of partitioning signal by cell state, caQTL identified only by cell state-resolved profiling explained a greater proportion of trait variation for 29% (14/49) of complex traits relative to bulk caQTL (fig. S15A). Furthermore, fine-mapped caQTL from meristem and primordia cells explained the most variance for 69% (34/49) of traits (fig. S15, B and C), including flowering time and SAM, ear, and tassel morphologies, supporting prior observations that adult architectural traits associate with regulators of SAM identity (46).

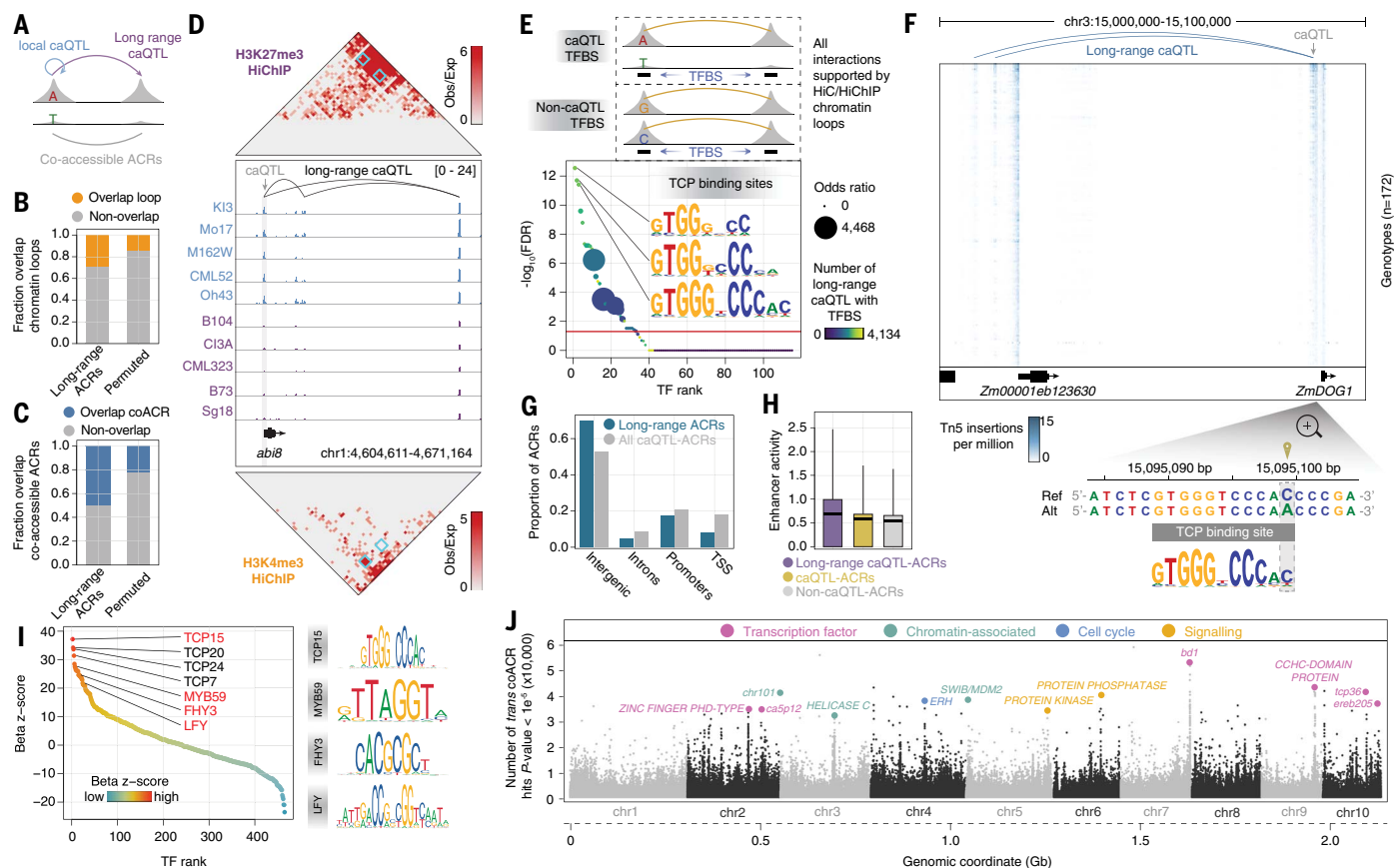


Fig. 3. Long-range caQTL perturb chromatin loops. (A) Schematic of local and long-range caQTL (top) and co-accessible ACRs (bottom). (B) Fraction of predicted (based on long-range caQTL genetic evidence) and permuted ACR-ACR interactions overlapping experimentally resolved chromatin loops. (C) Fraction of predicted and permuted ACR-ACR interactions overlapping with co-accessible ACRs. (D) Top: Normalized H3K27me3 HiChIP contact matrix. Middle: Exemplary long-range caQTL interaction between the promoter of *abi8* and two ACRs downstream. Bottom: Normalized H3K4me3 HiChIP contact matrix. Blue rectangles denote the interaction frequency between the promoter of *abi8* and the downstream ACRs. (E) Top, illustration of test sets (TFBS counts containing long-range caQTL or non-caQTL SNVs conditioned on HiC/HiChIP looping and reciprocal presence of TFBSs at both anchors). Bottom, bubble plot of TF enrichment ranks by $-\log_{10}(\text{FDR})$. Points are colored by the number of

interactions and sized by the odds ratio from Fisher's exact tests. Red line indicates $-\log_{10}(\text{FDR} = 0.05)$. Inset, TFBS logo of top three enriched motifs corresponding to 17 TCP TF family members. (F) Top, density heatmap of Tn5 insertion counts across genotypes (rows). Bottom, magnification of caQTL alleles overlapping a predicted TCP-binding site. (G) Proportion of ACRs in each genomic context split by long-range caQTL-ACRs with local caQTL and all caQTL-ACRs. (H) Distribution of enhancer activity across long-range caQTL-ACRs, all caQTL-ACRs, and non-caQTL-ACRs. Black bars indicate distribution means. (I) TF motifs ranked by beta coefficient Z scores (relative to 100 permutations) from a generalized linear model of long-range caQTL-ACR interactive capacity. (J) Manhattan plot of the number of co-accessible ACRs associated ($P < 1 \times 10^5$) with a single SNV. Candidate genes are annotated by different colors.

To determine the confluence of chromatin and trait genetic architectures, we used statistical colocalization to identify shared causal variants, revealing 811 caQTL-GWAS trait pairs (fig. S16 and table S7). The most abundant colocalized caQTL-GWAS variants were related to flowering traits (Fig. 4C). For example, the alternate allele of an SNV (chr1:261397272) located inside an ACR 5 kb upstream of *lbd10* was associated with decreased chromatin accessibility in epidermal cells and an increase of ~30 growing degree days to tassel (fig. S15D). However, because many causal variants are likely not represented in our study (such as unseen insertions or deletions), we reasoned that finding shared molecular QTL in strong LD ($r^2 > 0.8$) with trait-associated variants

could be useful for investigating trait-regulatory mechanisms. To this end, we uncovered an additional 326 unique caQTL-eQTL variants in LD with 1342 GWAS hits (table S7). Highlighting the utility of this method to resolve regulatory mechanisms, we identified a GWAS hit for plant height in LD ($r^2 = 0.93$) with a shared adaxial leaf primordia caQTL and eQTL ~35 kb upstream of the heat shock gene *hsp19* (Zm00001eb400020; Fig. 4D). The reference allele was associated with agronomically favorable decreased plant height, greater *hsp19* transcript abundance, and an expanded accessible chromatin domain. Investigating integrated molecular QTL-GWAS variants more generally, we found that shared caQTL-eQTL were most frequently in LD with GWAS hits

for flowering-related traits (fig. S15E). Thus, we hypothesized that flowering variability in the diversity panel may be caused by regulatory adaptation.

Many physiological processes, including flowering, changed during the migration of maize from tropical to temperate latitudes. Comparing tropical and temperate maize lines, we found 10,740 ACRs with differential accessibility and refer to these regions as subpopulation-specific ACRs (FDR < 0.1; Fig. 4E). These ACRs were strongly enriched in signatures of selection that accompanied the adaptation to temperate climates and GWAS variants associated with flowering time (Fig. 4F, fig. S15G, and table S8) (47). Moreover, different TFBSs were enriched in subpopulation-specific ACRs depending on

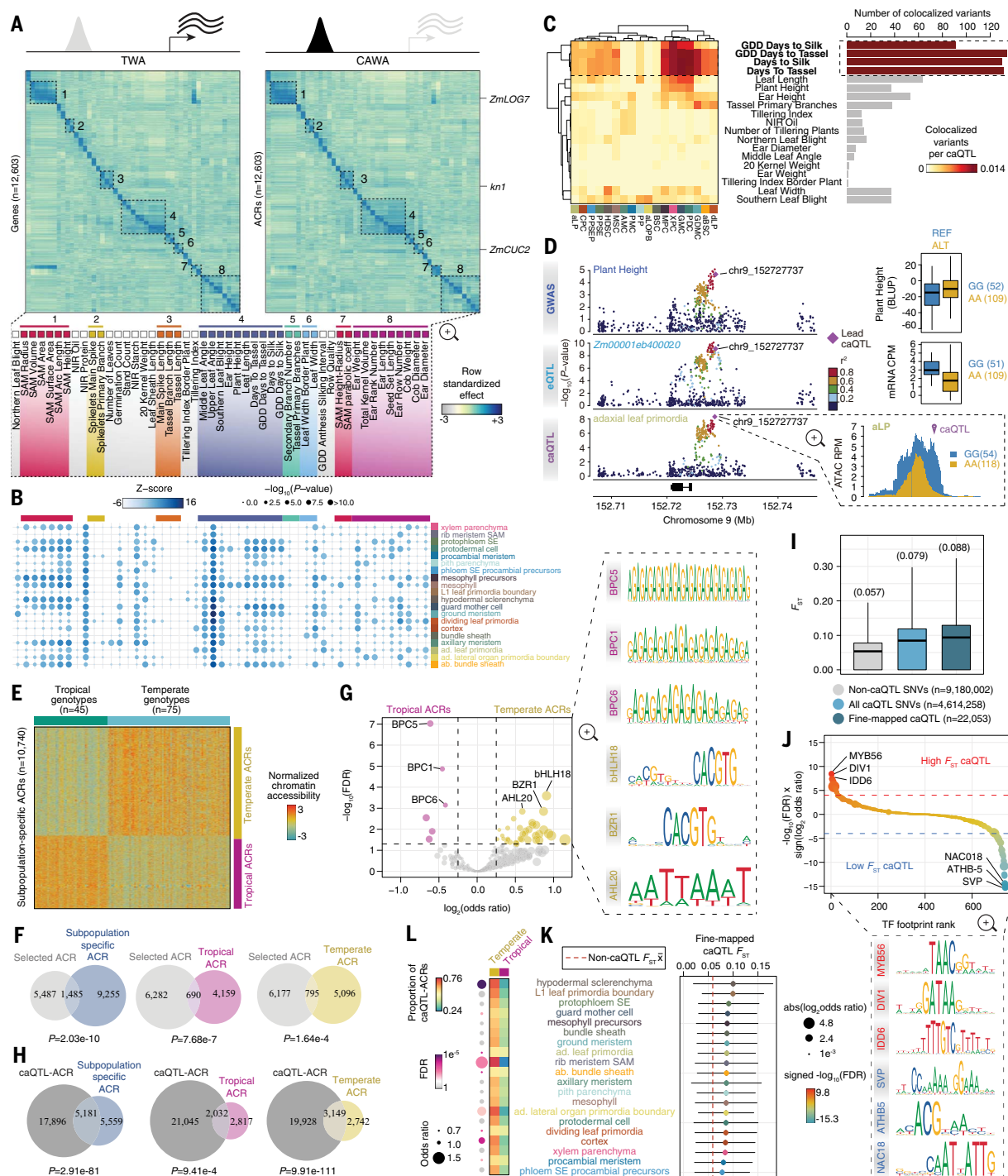


Fig. 4. Integration of caQTL with eQTL and GWAS of complex traits.

(A) Heatmaps of standardized effects from CAVA and TWA for matched ACRs (left) and genes (right). Phenotype traits of the columns (identical order for both heatmaps) are shown below. (B) Permutation tests ($\times 10,000$) of percent variance explained by caQTL partitioned by cell state compared with non-caQTL SNVs. (C) Left: Heatmap illustrating the fraction of caQTL colocalizing with a given GWAS trait. Right: Bar plot of the number of colocated variants for each GWAS trait. (D) Top: Manhattan plots of plant height GWAS. Middle: Expression QTL of bulk seedlings. Bottom: caQTL for adaxial leaf primordia. Variants are colored by LD to the lead caQTL. Trait distributions split by the lead variant genotypes are shown to the right. (E) Heatmap of subpopulation-specific normalized chromatin accessibility (rows) across 120 tropical and temperate maize genotypes. (F) Venn diagram

of ACRs within genomic regions under selection that overlap subpopulation-specific ACRs, tropical-specific ACRs, and temperate-specific ACRs. (G) Volcano plot of differentially accessible motifs between tropical and temperate ACRs. (H) Venn diagram of caQTL-ACRs that overlap subpopulation-specific ACRs, tropical ACRs, and temperate ACRs. (I) Distribution of F_{ST} values of non-caQTL SNVs, all caQTL SNVs, and fine-mapped caQTL SNVs. (J) TF footprint ranked by the signed significance ($-\log_{10}$ FDR) for enrichment of high- or low- F_{ST} SNVs. (K) Distribution (SD) and mean F_{ST} values of fine-mapped caQTL split by cell state. The dashed red line indicates the genome-wide average for non-caQTL SNVs. (L) Heatmap illustrating the proportion of caQTL-ACRs from each cell state that are temperate or tropical specific. Cell states with significant deviations from the expected proportions (FDR < 0.05) are denoted by colored bubbles, where the size of the bubble indicates the odds ratio.

the subpopulation. For example, we found that transcription-silencing motifs recognized by BPC TFs (32) were uniquely associated with tropical ACRs (Fig. 4G and table S9). By contrast, BRASSINAZOLE-RESISTANT 1 (BZR1), a master developmental regulator; AT-HOOK MOTIF CONTAINING NUCLEAR LOCALIZED 20 (AHL20), a DNA binding protein associated with delayed flowering (48); and a BASIC HELIX-LOOP-HELIX 18 (bHLH18) TF specifically expressed in floral tissue (49) were distinctly associated with temperate ACRs (Fig. 4G and table S9). These TFBS enrichments are consistent with a modified floral transition being integral for temperate adaptation. Despite including genetic relatedness covariates during caQTL mapping, subpopulation-specific ACRs were strongly coincident with caQTL-ACRs (Fig. 4H), suggesting that local adaptation underlies a significant proportion of chromatin accessibility variation. Indeed, comparison of allele frequencies between tropical and temperate inbreds revealed higher F_{ST} values for caQTL and fine-mapped caQTL compared with non-caQTL SNVs (Wilcoxon rank sum test, $P < 2.2 \times 10^{-16}$; Fig. 4I). These results highlight prevalent stratifying selection as a source of CRE functional diversity within maize subpopulations.

To define the regulatory grammar promoting temperate and tropical differentiation, we compared TF footprints localized with fine-mapped caQTL categorized as high or low F_{ST} . High- F_{ST} fine-mapped caQTL were enriched for TF footprints previously associated with flowering time (MYB56, IDD6) (50, 51) and floral cell identity (DIV1) (52) and depleted of TF footprints associated with short vegetative phase (SVP) (53), ABA responsiveness (ATHB-5), (54), and embryogenesis (NAC018) (55) (Fig. 4J and table S10). We hypothesized that local adaptation would affect chromatin accessibility in specific cell states. Indeed, F_{ST} distributions for fine-mapped caQTL differed among cell states and were heavily biased toward ACRs with increased temperate chromatin accessibility, particularly in the rib meristem and hypodermal sclerenchyma (Kruskal-Wallis rank sum test, Wilcoxon rank sum test, $P < 9.4 \times 10^{-12}$; Fig. 4, K and L). These results suggest that temperate adaptation involved widespread changes to chromatin accessibility that rewired flowering and developmental TF networks nonuniformly across cell states.

Discussion

Variable chromatin accessibility underlies cell type-specific transcriptional regulation and, as a result, functionally diverse cell states. Although we previously established widespread chromatin accessibility variation among cell states within maize (56), the extent to which cell type-level chromatin accessibility differed between genetically diverse individuals remained unclear. Here, we present a large-scale investigation of

chromatin accessibility variation in 20 cell states across 172 geographically dispersed maize inbreds. Exploiting low LD in maize, we fine mapped >22,000 putatively causal variants. Fine-mapped caQTL are highly cell state specific and have the strongest effects when occurring in TCP TFBSs. The patterns of TCP gene expression, association with global co-accessibility, and significant effects on local chromatin accessibility collectively implicate TCP TFs as being regulators of chromatin accessibility. The identification of TCP TFs as prevalent and potent promoters of chromatin accessibility and contributors toward chromatin interactions adds to a growing body of evidence implicating them as major regulators of chromatin architecture (57). Our gene-regulatory network analysis also found that cell state-specific TCPs are targeted by master cell identity TF regulators, implicating the upstream factors that are important for promoting chromatin accessibility in diverse cell contexts. The identification of TCPs as promoters of accessible chromatin is a potentially exciting advance; only a few plant pioneer factors have been identified, whereas all studied plants lack a homolog to the metazoan-specific chromatin looping factor CTCF. We anticipate follow-up genetic and biochemical experiments to investigate the mechanistic roles of TCP factors in chromatin accessibility and long-distance chromosome looping.

Integrating caQTL with GWAS variants uncovered the cellular contexts associated with phenotypic traits. Further, the combination of CAWA, TWA, and GWAS provided preliminary insights into the underlying regulatory mechanisms behind complex trait variation. Examining the ACR differences between temperate and tropical maize, we demonstrated that caQTL are associated with signatures of population differentiation and selection, confirming regulatory variation as a key target of local adaptation and human breeding efforts. As such, this resource provides exciting avenues for follow-up exploration and direct real-world application. First, we identified colocalized cis variants predicted to causally affect chromatin accessibility and complex traits, which narrows candidate molecular breeding and genetic engineering targets for crop improvement. Second, we revealed how individual TFBS variants affect both the cell state specificity and enhancer activity of TFBSs, which can inform synthetic CRE design. Third, we defined the ACRs and cognate CREs that are fixed in domesticated maize. The absence of extant genetic variation over these regions both highlights their agronomic importance and occludes their study through traditional quantitative genetic techniques. Therefore, these fixed regions represent potent exploratory targets for engineering allelic variation through gene editing, both to understand their function and to shepherd new crop phenotypes. Future reverse genetic and biochemical exploita-

tion of these data will undoubtedly shape crop improvement and deepen our mechanistic understanding of phenotypic variation.

Our study addressed several outstanding questions, serving as a foundation for similar caQTL studies. However, we defined caQTL in only 20 of 33 cell states due to limited power in rare cell types. With technological increases in throughput, future studies will be better powered to resolve caQTL from low-frequency cell types. Additionally, we sampled 7-day-old seedlings, a tissue selected because it includes the SAM, which is known to forecast adult traits. However, our sampling strategy cannot find caQTL related to other phenotypically relevant organs, including roots, inflorescences, and seeds, which is also complicated by the small effect sizes of complex traits mapped in our study. These challenges ultimately affect the recovery of colocalized caQTL with known GWAS variants. Expanding the number of tissues and developmental stages profiled by scATAC-seq and snRNA-seq, as well as the number of individuals, will be useful to more comprehensively characterize the molecular underpinnings of trait variance. Another consideration is the use of a single reference genome; our analyses are limited to ACRs and caQTL that can be identified in the context of B73. We anticipate that advances in pangenomics and additional reference genomes will facilitate more comprehensive future discovery (58). Finally, although the size of the diversity panel was well powered to identify cis-caQTL, it was limited for trans-QTL detection; uncovering trans-regulatory interactions remains a challenge in most eukaryotic genomes. Future studies with larger cohorts and datasets that preidentify trans-regulated ACRs will provide a promising avenue to reveal hidden trans-QTL.

Methods summary

Replicated material was collected from ~7-day-old seedlings from 172 diverse maize inbreds from the Goodman-Buckler maize diversity panel (17) across several randomized batches (table S2). Nuclei were isolated using a previously described approach using flow cytometry sorting (59). Flow-sorted nuclei were spun in a swinging bucket centrifuge [5 min, 500 relative centrifugal force (rcf)], resuspended in 10 μ l of LBO1, visualized, counted on a hemocytometer with a fluorescence microscope, and adjusted to a final concentration of between 3200 and 12,000 nuclei per microliter using diluted nuclei buffer (10X Genomics, catalog no. 1000176). For each library preparation, 5 μ l of suspended nuclei was loaded per well on the Next GEM Chip H (10X Genomics, catalog no. 1000162). Single-cell ATAC-seq libraries were prepared according to the manufacturer's instructions (10X Genomics, catalog no. 1000176, Chromium Next GEM v1.1) using the Chromium Controller (10X Genomics, catalog no. 120223).

The resulting scATAC-seq libraries were sequenced using an Illumina S4 flow cell (NovaSeq 6000) in dual-index mode. Raw sequencing output was processed with Cell Ranger Single Cell Software (v2.2.0; 10x Genomics) and aligned to the maize AGPv5 reference genome (60) with *chromap* (v0.2.3) (61).

Singlet nuclei were identified by removing barcodes associated with multiple genotype identities from *souporcell*, a minimum of 500 unique Tn5 insertions within ACRs, and fraction of reads in peaks and fraction of reads near transcription start sites above background levels determined for each library independently. Cell state identities were identified by analysis of marker genes, differential gene chromatin accessibility, and integration with tissue-matched snRNA-seq data. Chromatin accessibility profiles were generated in each cell state by aggregating nuclei by genotype and cell state identity. Sample covariates were estimated based on genetic relatedness, library pooling information, cell cycle stage, and latent characteristics present in the chromatin accessibility counts matrix. Chromatin accessibility QTL were detected with *tensorQTL* (v1.0.9) (62) for each distinct cell state and in the pseudobulk datasets using cell context-specific FDR thresholds based on average read counts.

REFERENCES AND NOTES

- Y. Xiao, H. Liu, L. Wu, M. Warburton, J. Yan, Genome-wide association studies in maize: Praise and stargaze. *Mol. Plant* **10**, 359–374 (2017). doi: [10.1016/j.molp.2016.12.008](https://doi.org/10.1016/j.molp.2016.12.008); pmid: [28039028](https://pubmed.ncbi.nlm.nih.gov/28039028/)
- J. G. Wallace et al., Association mapping across numerous traits reveals patterns of functional variation in maize. *PLOS Genet.* **10**, e1004845 (2014). doi: [10.1371/journal.pgen.1004845](https://doi.org/10.1371/journal.pgen.1004845); pmid: [25474422](https://pubmed.ncbi.nlm.nih.gov/25474422/)
- J. Engelhorn et al., Genetic variation at transcription factor binding sites largely explains phenotypic heritability in maize. *bioRxiv*, 2023.2008.2008.551183 (2024).
- A. P. Marand, A. L. Eveland, K. Kaufmann, N. M. Springer, Cis-regulatory elements in plant development, adaptation, and evolution. *Annu. Rev. Plant Biol.* **74**, 111–137 (2023). doi: [10.1146/annurev-arplant-070122-030236](https://doi.org/10.1146/annurev-arplant-070122-030236); pmid: [36608347](https://pubmed.ncbi.nlm.nih.gov/36608347/)
- R. J. Schmitz, E. Grotewold, M. Stam, Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. *Plant Cell* **34**, 718–741 (2022). doi: [10.1093/plcell/koab281](https://doi.org/10.1093/plcell/koab281); pmid: [34918159](https://pubmed.ncbi.nlm.nih.gov/34918159/)
- F. He et al., Cis-regulatory evolution spotlights species differences in the adaptive potential of gene expression plasticity. *Nat. Commun.* **12**, 3376 (2021). doi: [10.1038/s41467-021-23558-2](https://doi.org/10.1038/s41467-021-23558-2); pmid: [34099660](https://pubmed.ncbi.nlm.nih.gov/34099660/)
- E. Z. Kwon et al., Progressive loss of function in a limb enhancer during snake evolution. *Cell* **167**, 633–642.e11 (2016). doi: [10.1016/j.cell.2016.09.028](https://doi.org/10.1016/j.cell.2016.09.028); pmid: [27768887](https://pubmed.ncbi.nlm.nih.gov/27768887/)
- C. Huang et al., *ZmCCT9* enhances maize adaptation to higher latitudes. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E334–E341 (2018). doi: [10.1073/pnas.1718058115](https://doi.org/10.1073/pnas.1718058115); pmid: [29279404](https://pubmed.ncbi.nlm.nih.gov/29279404/)
- S. Salvi et al., Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11376–11381 (2007). doi: [10.1073/pnas.070415104](https://doi.org/10.1073/pnas.070415104); pmid: [17595297](https://pubmed.ncbi.nlm.nih.gov/17595297/)
- A. Studer, Q. Zhao, J. Ross-Ibarra, J. Doebley, Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat. Genet.* **43**, 1160–1163 (2011). doi: [10.1038/ng.942](https://doi.org/10.1038/ng.942); pmid: [21946354](https://pubmed.ncbi.nlm.nih.gov/21946354/)
- J. Tian et al., Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. *Science* **365**, 658–664 (2019). doi: [10.1126/science.aax5482](https://doi.org/10.1126/science.aax5482); pmid: [31416957](https://pubmed.ncbi.nlm.nih.gov/31416957/)
- C. J. Whipple et al., grassy tillers1 promotes apical dominance in maize and responds to shade signals in the grasses. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E506–E512 (2011). doi: [10.1073/pnas.1102819108](https://doi.org/10.1073/pnas.1102819108); pmid: [21808030](https://pubmed.ncbi.nlm.nih.gov/21808030/)
- Q. Yang et al., CACTA-like transposable element in *ZmCCT* attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16969–16974 (2013). doi: [10.1073/pnas.1310949110](https://doi.org/10.1073/pnas.1310949110); pmid: [24089449](https://pubmed.ncbi.nlm.nih.gov/24089449/)
- M. A. A. Minow, A. P. Marand, R. J. Schmitz, Leveraging single-cell populations to uncover the genetic basis of complex traits. *Annu. Rev. Genet.* **57**, 297–319 (2023). doi: [10.1146/annurev-genet-022123-110824](https://doi.org/10.1146/annurev-genet-022123-110824); pmid: [37562412](https://pubmed.ncbi.nlm.nih.gov/37562412/)
- International HapMap Consortium, A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005). doi: [10.1038/nature04226](https://doi.org/10.1038/nature04226); pmid: [16255080](https://pubmed.ncbi.nlm.nih.gov/16255080/)
- C. C. Laurie et al., Linkage disequilibrium in wild mice. *PLOS Genet.* **3**, e144 (2007). doi: [10.1371/journal.pgen.0030144](https://doi.org/10.1371/journal.pgen.0030144); pmid: [17722986](https://pubmed.ncbi.nlm.nih.gov/17722986/)
- S. A. Flint-Garcia et al., Maize association population: A high-resolution platform for quantitative trait locus dissection. *Plant J.* **44**, 1054–1064 (2005). doi: [10.1111/j.1365-3113.2005.02591.x](https://doi.org/10.1111/j.1365-3113.2005.02591.x); pmid: [16359397](https://pubmed.ncbi.nlm.nih.gov/16359397/)
- J. M. Noshay et al., Stability of DNA methylation and chromatin accessibility in structurally diverse maize genomes. *G3* **11**, jkab190 (2021). doi: [10.1093/g3journal/jkab190](https://doi.org/10.1093/g3journal/jkab190); pmid: [34849810](https://pubmed.ncbi.nlm.nih.gov/34849810/)
- W. A. Ricci et al., Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants* **5**, 1237–1249 (2019). doi: [10.1038/s41477-019-0547-0](https://doi.org/10.1038/s41477-019-0547-0); pmid: [31740773](https://pubmed.ncbi.nlm.nih.gov/31740773/)
- A. Husbands, E. M. Bell, B. Shuai, H. M. Smith, P. S. Springer, LATERAL ORGAN BOUNDARIES defines a new family of DNA-binding transcription factors and can interact with specific bHLH proteins. *Nucleic Acids Res.* **35**, 6663–6671 (2007). doi: [10.1093/nar/gkm775](https://doi.org/10.1093/nar/gkm775); pmid: [17913740](https://pubmed.ncbi.nlm.nih.gov/17913740/)
- X. Chen et al., SQUAMOSA promoter-binding protein-like transcription factors: Star players for plant growth and development. *J. Integr. Plant Biol.* **52**, 946–951 (2010). doi: [10.1111/j.1744-7909.2010.00987.x](https://doi.org/10.1111/j.1744-7909.2010.00987.x); pmid: [20977652](https://pubmed.ncbi.nlm.nih.gov/20977652/)
- C. Schwechheimer, P. M. Schröder, C. E. Blaby-Haas, Plant GATA Factors: Their Biology, Phylogeny, and Phylogenomics. *Annu. Rev. Plant Biol.* **73**, 123–148 (2022). doi: [10.1146/annurev-arplant-072221-092913](https://doi.org/10.1146/annurev-arplant-072221-092913); pmid: [35130446](https://pubmed.ncbi.nlm.nih.gov/35130446/)
- P. K. Agarwal, P. Agarwal, M. K. Reddy, S. K. Sopory, Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. *Plant Cell Rep.* **25**, 1263–1274 (2006). doi: [10.1007/s00299-006-0204-8](https://doi.org/10.1007/s00299-006-0204-8); pmid: [16858552](https://pubmed.ncbi.nlm.nih.gov/16858552/)
- M. A. Khoso et al., WRKY transcription factors (TFs): Molecular switches to regulate drought, temperature, and salinity stresses in plants. *Front. Plant Sci.* **13**, 1039329 (2022). doi: [10.3389/fpls.2022.1039329](https://doi.org/10.3389/fpls.2022.1039329); pmid: [36426143](https://pubmed.ncbi.nlm.nih.gov/36426143/)
- D. M. Wills et al., From many, one: Genetic control of prolificacy during maize domestication. *PLOS Genet.* **9**, e1003604 (2013). doi: [10.1371/journal.pgen.1003604](https://doi.org/10.1371/journal.pgen.1003604); pmid: [23825971](https://pubmed.ncbi.nlm.nih.gov/23825971/)
- J. M. Noshay et al., Assessing the regulatory potential of transposable elements using chromatin accessibility profiles of maize transposons. *Genetics* **217**, 1–13 (2021). doi: [10.1093/genetics/yaa003](https://doi.org/10.1093/genetics/yaa003); pmid: [33683350](https://pubmed.ncbi.nlm.nih.gov/33683350/)
- H. Zhao et al., Proliferation of regulatory DNA elements derived from transposable elements in the maize genome. *Plant Physiol.* **176**, 2789–2803 (2018). doi: [10.1104/pp.17.01467](https://doi.org/10.1104/pp.17.01467); pmid: [29463772](https://pubmed.ncbi.nlm.nih.gov/29463772/)
- S. Yanagisawa, The Dof family of plant transcription factors. *Trends Plant Sci.* **7**, 555–560 (2002). doi: [10.1016/S1360-1385\(02\)02362-2](https://doi.org/10.1016/S1360-1385(02)02362-2); pmid: [12475498](https://pubmed.ncbi.nlm.nih.gov/12475498/)
- F. Licausi, M. Ohme-Takagi, P. Perata, APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: Mediators of stress responses and developmental programs. *New Phytol.* **199**, 639–649 (2013). doi: [10.1111/nph.12291](https://doi.org/10.1111/nph.12291); pmid: [24010138](https://pubmed.ncbi.nlm.nih.gov/24010138/)
- C. Smaczniak, R. G. Immink, G. C. Angenent, K. Kaufmann, Developmental and evolutionary diversity of plant MADS-domain factors: Insights from recent studies. *Development* **139**, 3081–3098 (2012). doi: [10.1242/dev.074674](https://doi.org/10.1242/dev.074674); pmid: [22872082](https://pubmed.ncbi.nlm.nih.gov/22872082/)
- K. A. F. Kremling et al., Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* **555**, 520–523 (2018). doi: [10.1038/nature25966](https://doi.org/10.1038/nature25966); pmid: [29539638](https://pubmed.ncbi.nlm.nih.gov/29539638/)
- J. Xiao et al., Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in *Arabidopsis*. *Nat. Genet.* **49**, 1546–1552 (2017). doi: [10.1038/ng.3937](https://doi.org/10.1038/ng.3937); pmid: [28825728](https://pubmed.ncbi.nlm.nih.gov/28825728/)
- Y. Zhou et al., Telobox motifs recruit CLF/SWN-PRC2 for H3K27me3 deposition via TRB factors in *Arabidopsis*. *Nat. Genet.* **50**, 638–644 (2018). doi: [10.1038/s41588-018-0109-9](https://doi.org/10.1038/s41588-018-0109-9); pmid: [29700471](https://pubmed.ncbi.nlm.nih.gov/29700471/)
- S. M. Urbat, G. Wang, P. Carbonetto, M. Stephens, Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019). doi: [10.1038/s41588-018-0268-8](https://doi.org/10.1038/s41588-018-0268-8); pmid: [30478440](https://pubmed.ncbi.nlm.nih.gov/30478440/)
- B. Wang et al., Genome-wide selection and genetic improvement during modern maize breeding. *Nat. Genet.* **52**, 565–571 (2020). doi: [10.1038/s41588-020-0616-3](https://doi.org/10.1038/s41588-020-0616-3); pmid: [32341525](https://pubmed.ncbi.nlm.nih.gov/32341525/)
- P. Sijacic, M. Bajic, E. C. McKinney, R. B. Meagher, R. B. Deal, Changes in chromatin accessibility between *Arabidopsis* stem cells and mesophyll cells illuminate cell type-specific transcription factor networks. *Plant J.* **94**, 215–231 (2018). doi: [10.1111/tpj.13882](https://doi.org/10.1111/tpj.13882); pmid: [29513366](https://pubmed.ncbi.nlm.nih.gov/29513366/)
- S. Miyashima et al., Mobile PEAR transcription factors integrate positional cues to prime cambial growth. *Nature* **565**, 490–494 (2019). doi: [10.1038/s41586-018-0839-y](https://doi.org/10.1038/s41586-018-0839-y); pmid: [30626969](https://pubmed.ncbi.nlm.nih.gov/30626969/)
- G. Wu et al., KANAD1 regulates adaxial-abaxial polarity in *Arabidopsis* by directly repressing the transcription of ASYMMETRIC LEAVES2. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16392–16397 (2008). doi: [10.1073/pnas.0803997105](https://doi.org/10.1073/pnas.0803997105); pmid: [18849474](https://pubmed.ncbi.nlm.nih.gov/18849474/)
- S. M. de Jager et al., Dissecting regulatory pathways of G1/S control in *Arabidopsis*: Common and distinct targets of CYCD3;1, E2Fa and E2Fc. *Plant Mol. Biol.* **71**, 345–365 (2009). doi: [10.1007/s11003-009-9527-5](https://doi.org/10.1007/s11003-009-9527-5); pmid: [19662336](https://pubmed.ncbi.nlm.nih.gov/19662336/)
- N. Kumasaka, A. J. Knights, D. J. Gaffney, High-resolution genetic mapping of putative causal interactions between regions of open chromatin. *Nat. Genet.* **51**, 128–137 (2019). doi: [10.1038/s41588-018-0278-6](https://doi.org/10.1038/s41588-018-0278-6); pmid: [30478436](https://pubmed.ncbi.nlm.nih.gov/30478436/)
- L. Schlegel et al., GenomicLinks: Deep learning predictions of 3D chromatin interactions in the maize genome. *NAR Genom. Bioinform.* **6**, lqae123 (2024). doi: [10.1093/nargab/lqae123](https://doi.org/10.1093/nargab/lqae123); pmid: [39318505](https://pubmed.ncbi.nlm.nih.gov/39318505/)
- L. Dong et al., Transcriptomic analysis of leaf sheath maturation in maize. *Int. J. Mol. Sci.* **20**, 2472 (2019). doi: [10.3390/ijms20102472](https://doi.org/10.3390/ijms20102472); pmid: [31109136](https://pubmed.ncbi.nlm.nih.gov/31109136/)
- J. W. Satterlee, J. Strable, M. J. Scanlon, Plant stem-cell organization and differentiation at single-cell resolution. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 33689–33699 (2020). doi: [10.1073/pnas.2018781117](https://doi.org/10.1073/pnas.2018781117); pmid: [33318187](https://pubmed.ncbi.nlm.nih.gov/33318187/)
- R. Johnston et al., Transcriptomic analyses indicate that maize ligule development recapitulates gene expression patterns that occur during lateral organ initiation. *Plant Cell* **26**, 4718–4732 (2014). doi: [10.1105/tpc.114.132688](https://doi.org/10.1105/tpc.114.132688); pmid: [25516601](https://pubmed.ncbi.nlm.nih.gov/25516601/)
- R. A. Kerstetter, D. Laudencia-Chinguanco, L. G. Smith, S. Hake, Loss-of-function mutations in the maize homeobox gene, *knot1*, are defective in shoot meristem maintenance. *Development* **124**, 3045–3054 (1997). doi: [10.1242/dev.124.16.3045](https://doi.org/10.1242/dev.124.16.3045); pmid: [9272946](https://pubmed.ncbi.nlm.nih.gov/9272946/)
- S. Knauer et al., A high-resolution gene expression atlas links dedicated meristem genes to key architectural traits. *Genome Res.* **29**, 1962–1973 (2019). doi: [10.1101/gr.250878.119](https://doi.org/10.1101/gr.250878.119); pmid: [31744902](https://pubmed.ncbi.nlm.nih.gov/31744902/)
- H. Liu et al., Genomic, transcriptomic, and phenomic variation reveals the complex adaptation of modern maize breeding. *Mol. Plant* **8**, 871–884 (2015). doi: [10.1016/j.molp.2015.01.016](https://doi.org/10.1016/j.molp.2015.01.016); pmid: [25620769](https://pubmed.ncbi.nlm.nih.gov/25620769/)
- R. Tayengwa, P. Sharma Koirala, C. F. Pierce, B. E. Werner, M. M. Neff, Overexpression of *AtAHL20* causes delayed flowering in *Arabidopsis* via repression of FT expression. *BMC Plant Biol.* **20**, 559 (2020). doi: [10.1186/s12870-020-02733-5](https://doi.org/10.1186/s12870-020-02733-5); pmid: [33308168](https://pubmed.ncbi.nlm.nih.gov/33308168/)
- A. V. Klepikova, A. S. Kasianov, E. S. Gerasimov, M. D. Logacheva, A. A. Penin, A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J.* **88**, 1058–1070 (2016). doi: [10.1111/tpj.13312](https://doi.org/10.1111/tpj.13312); pmid: [27549386](https://pubmed.ncbi.nlm.nih.gov/27549386/)
- L. Chen, A. Bernhardt, J. Lee, H. Hellmann, Identification of *Arabidopsis* MYB56 as a novel substrate for CRL3^{BFM} E3 ligases. *Mol. Plant* **8**, 242–250 (2015). doi: [10.1016/j.molp.2014.10.004](https://doi.org/10.1016/j.molp.2014.10.004); pmid: [25618823](https://pubmed.ncbi.nlm.nih.gov/25618823/)
- J. Colasanti, Z. Yuan, V. Sundaresan, The indeterminate gene encodes a zinc finger protein and regulates a leaf-generated signal required for the transition to flowering in maize. *Cell* **93**, 593–603 (1998). doi: [10.1016/S0092-8674\(00\)81188-5](https://doi.org/10.1016/S0092-8674(00)81188-5); pmid: [9604934](https://pubmed.ncbi.nlm.nih.gov/9604934/)
- D. Zhang et al., Transcription factor *DIVARICATA1* positively modulates seed germination in response to salinity stress. *Plant Physiol.* **195**, 2997–3009 (2024). doi: [10.1093/plphys/kiae231](https://doi.org/10.1093/plphys/kiae231); pmid: [38687890](https://pubmed.ncbi.nlm.nih.gov/38687890/)
- U. Hartmann et al., Molecular cloning of SVP: A negative regulator of the floral transition in *Arabidopsis*. *Plant J.* **21**,

- 351–360 (2000). doi: [10.1046/j.1365-313x.2000.00682.x](https://doi.org/10.1046/j.1365-313x.2000.00682.x); pmid: [10758486](https://pubmed.ncbi.nlm.nih.gov/10758486/)
54. H. Johannesson, Y. Wang, J. Hanson, P. Engström, The *Arabidopsis thaliana* homeobox gene *ATHB5* is a potential regulator of abscisic acid responsiveness in developing seedlings. *Plant Mol. Biol.* **51**, 719–729 (2003). doi: [10.1023/A:1022567625228](https://doi.org/10.1023/A:1022567625228); pmid: [12678559](https://pubmed.ncbi.nlm.nih.gov/12678559/)
55. T. Kunieda *et al.*, NAC family proteins NARS1/NAC2 and NARS2/NAM in the outer integument regulate embryogenesis in *Arabidopsis*. *Plant Cell* **20**, 2631–2642 (2008). doi: [10.1105/tpc.108.060160](https://doi.org/10.1105/tpc.108.060160); pmid: [18849494](https://pubmed.ncbi.nlm.nih.gov/18849494/)
56. A. P. Marand, Z. Chen, A. Gallavotti, R. J. Schmitz, A cis-regulatory atlas in maize at single-cell resolution. *Cell* **184**, 3041–3055.e21 (2021). doi: [10.1016/j.cell.2021.04.014](https://doi.org/10.1016/j.cell.2021.04.014); pmid: [33964211](https://pubmed.ncbi.nlm.nih.gov/33964211/)
57. C. Liu, Y. J. Cheng, J. W. Wang, D. Weigel, Prominent topologically associated domains differentiate global chromatin packing in rice from *Arabidopsis*. *Nat. Plants* **3**, 742–748 (2017). doi: [10.1038/s41477-017-0005-9](https://doi.org/10.1038/s41477-017-0005-9); pmid: [28848243](https://pubmed.ncbi.nlm.nih.gov/28848243/)
58. P. J. Bradbury *et al.*, The Practical Haplotype Graph, a platform for storing and using pangenomes for imputation. *Bioinformatics* **38**, 3698–3702 (2022). doi: [10.1093/bioinformatics/btac410](https://doi.org/10.1093/bioinformatics/btac410); pmid: [35748708](https://pubmed.ncbi.nlm.nih.gov/35748708/)
59. A. P. Marand, X. Zhang, J. Nelson, P. A. Braga Dos Reis, R. J. Schmitz, Profiling single-cell chromatin accessibility in plants. *STAR Protoc.* **2**, 100737 (2021). doi: [10.1016/j.xpro.2021.100737](https://doi.org/10.1016/j.xpro.2021.100737); pmid: [34430912](https://pubmed.ncbi.nlm.nih.gov/34430912/)
60. M. B. Hufford *et al.*, De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655–662 (2021). doi: [10.1126/science.abg5289](https://doi.org/10.1126/science.abg5289); pmid: [34353948](https://pubmed.ncbi.nlm.nih.gov/34353948/)
61. H. Zhang *et al.*, Fast alignment and preprocessing of chromatin profiles with Chromap. *Nat. Commun.* **12**, 6566 (2021). doi: [10.1038/s41467-021-26865-w](https://doi.org/10.1038/s41467-021-26865-w); pmid: [34772935](https://pubmed.ncbi.nlm.nih.gov/34772935/)
62. A. Taylor-Weiner *et al.*, Scaling computational genomics to millions of individuals with GPUs. *Genome Biol.* **20**, 228 (2019). doi: [10.1186/s13059-019-1836-7](https://doi.org/10.1186/s13059-019-1836-7); pmid: [31675989](https://pubmed.ncbi.nlm.nih.gov/31675989/)
63. A. P. Marand *et al.*, Scripts and software for: The genetic architecture of cell type-specific cis-regulation in maize, Zenodo (2024); <https://doi.org/10.5281/zenodo.14230112>.
64. A. P. Marand *et al.*, Data for: The genetic architecture of cell type-specific cis-regulation in maize, Dryad (2024); <https://doi.org/10.5061/dryad.nk98sf82v>.

ACKNOWLEDGMENTS

We thank the GACRC for providing valuable computational support and resources, USDA GRIN for maize germplasm, the Duke University School of Medicine Sequencing and Genomic Technologies Shared Resource for Illumina sequencing services, E. Buckler for advice on genotyping strategies, J. Gage for modeling approaches, and N. Springer for feedback on analyses. **Funding:** This work was supported by the National Institutes of Health (grant 1R00/K99GM144742 to A.P.M.), the National Science Foundation (grant IOS-1905869 to A.P.M. and grants IOS-2026554, MCB-2120132, and IOS-1856627 to R.J.S.), and the University of Georgia Office of Research (R.J.S.). **Author contributions:** Conceptualization: A.P.M., R.J.S.; Funding acquisition: A.P.M., R.J.S.; Investigation: A.P.M., L.J., F.G.-C., M.A.A.M., J.P.M., Z.L., S.B., C.M., L.S., F.J., R.J.S.; Methodology: A.P.M., X.Z.; Project administration: A.P.M., R.J.S.; Supervision: A.P.M., R.J.S., F.J.;

Visualization: A.P.M., L.J., F.G.-C.; Writing – original draft: A.P.M., M.A.A.M., R.J.S. **Competing interests:** R.J.S. is a cofounder of RQuest Genomics, LLC, a company that provides epigenomic services. The remaining authors declare no competing interests.

Data and materials availability: All data supporting the conclusions of this study are provided in the main text or supplementary materials. Raw and processed data generated by this study are available in the National Center for Biotechnology Information Gene Omnibus (NCBI GEO) under accession number GSE275410. Coverage tracks from the different inbred lines can be viewed on JBrowse (https://epigenome.genetics.uga.edu/PlantEpigenome/?data=maize_v5). Scripts and software used in the analysis of the data generated by this study are publicly available at Zenodo (63). Additional processed datasets are available on Dryad (64). **License information:** Copyright © 2025 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. <https://www.science.org/about/science-licenses-journal-article-reuse>

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ads6601
Materials and Methods
Supplementary Text
Figures S1 to S16
References (65–100)
Data S1 to S10

Submitted 23 August 2024; accepted 4 February 2025
10.1126/science.ads6601