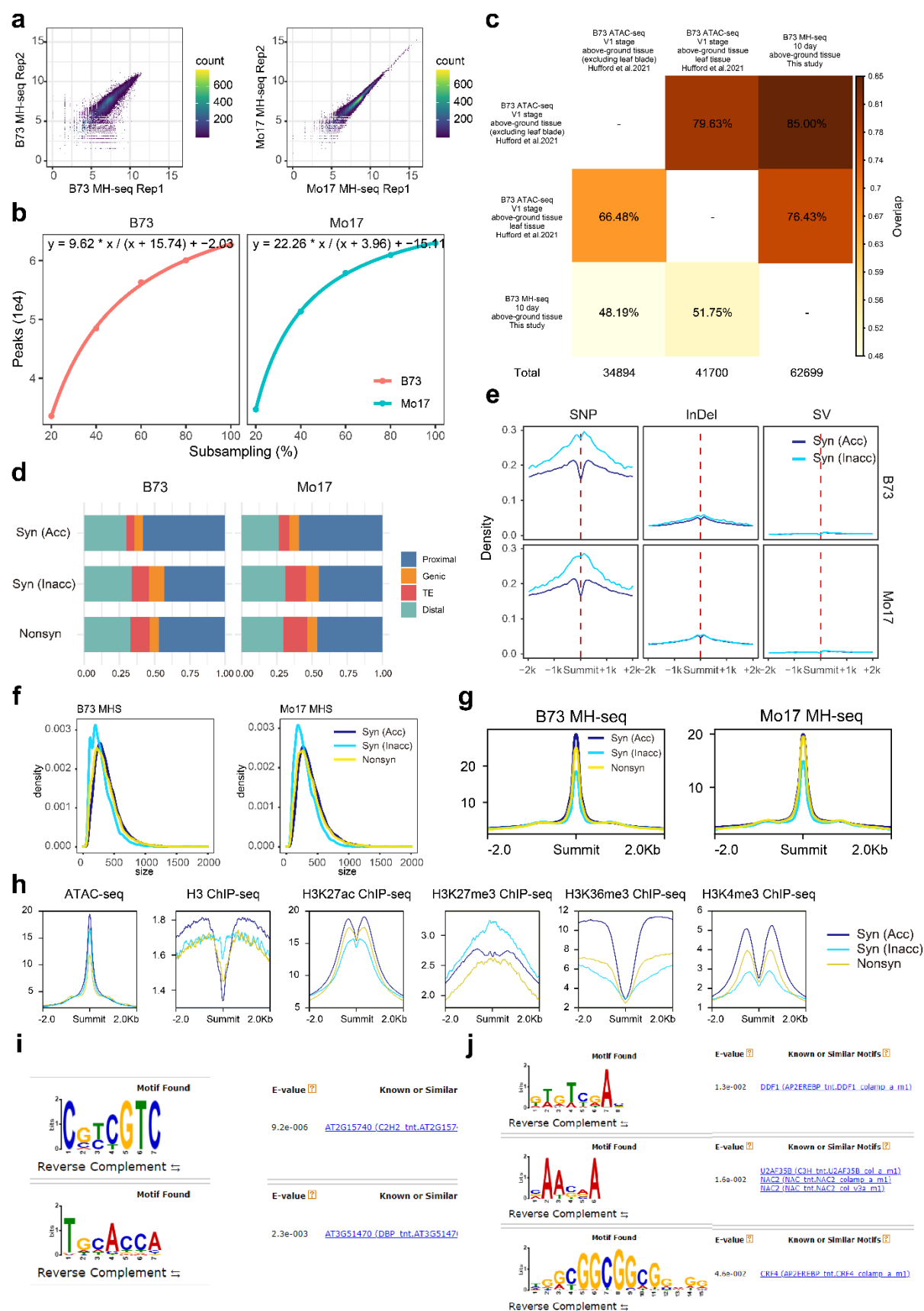


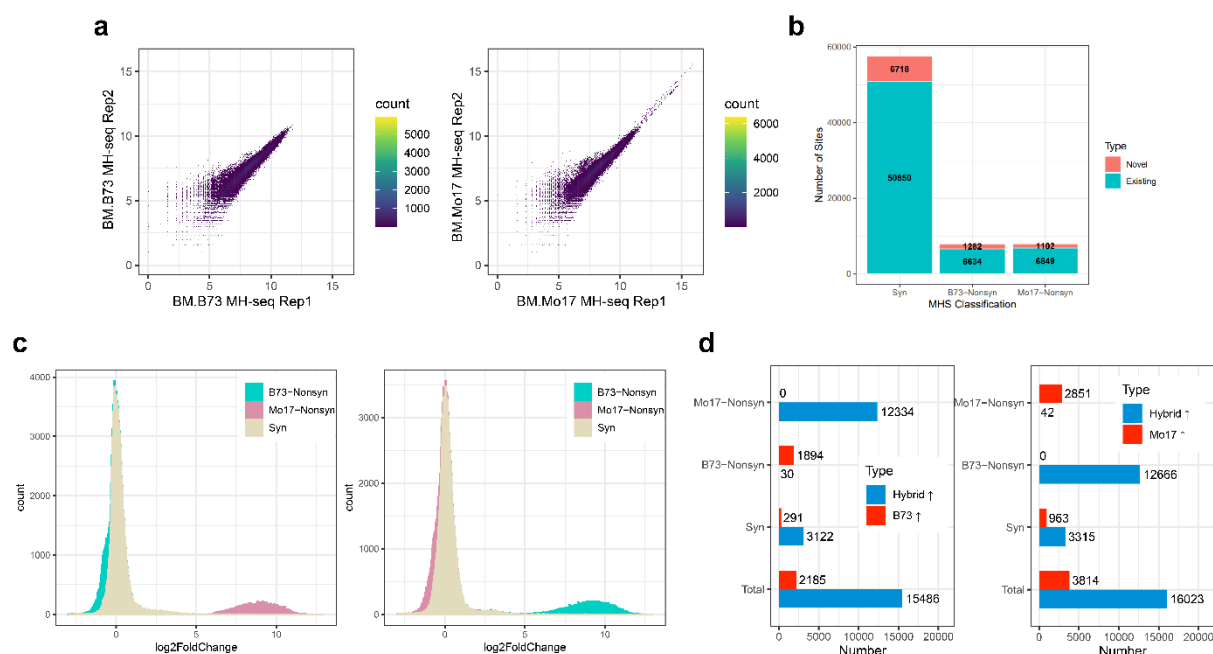
**Chromatin accessibility landscape and its association with heterosis in
maize hybrids**

Huang *et al.*



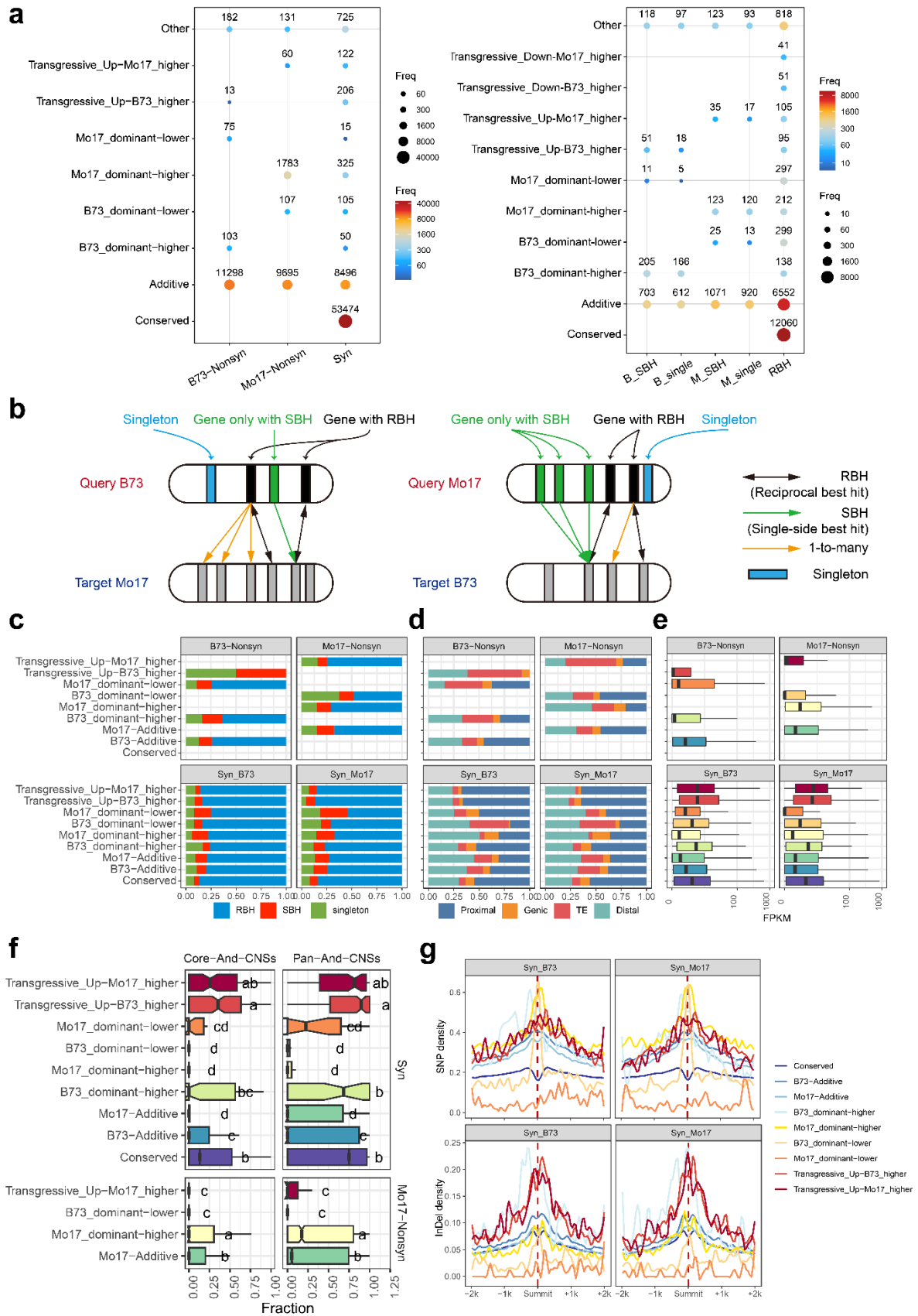
Supplementary Fig. 1. Characterization of different types of MNase hypersensitive sites (MHSs) between B73 and Mo17. a: Correlation between two biological replicates of MH-seq in B73 and Mo17

(Pearson correlation coefficient, all reads are binned with genome-wide 20-kb non-overlapping windows). **b:** The saturation curve of B73 and Mo17 MH-seq data. The real MH-seq data were subsampled, ranging from 20% to 100%. The growth curve was fitting with the model $y = a_1 \times x / (x + a_2) + a_3$. **c:** Overlaps among MH-seq and ATAC-seq peaks¹. The percentage in each grid represents the overlapping rate between the features in rows and the features in columns. The percentage in each cell at the intersection of a row and column represents the overlapping rate between the source feature (row) and the linked feature (column). For example, the cell at the intersection of the MH-seq row and the ATAC-seq (leaf tissue) column indicates that 51.75% of MH-seq peaks overlapped with ATAC-seq peaks. **d:** The composition of the distribution for different types of MHSs on the corresponding genomes. “Genic” means overlapping with the gene coding region; “Proximal” means within 3 kb of the gene coding region; “TE” means overlapping with transposable elements (TE); “Distal” means >3 kb from the gene coding region and not intersecting with TE. **e:** The density of SNP, InDel and SV for the summit region of different types of MHSs. **f:** Feature size distribution for different types of MHSs. **g:** Metaplots showing the enrichment of MH-seq reads around summits for different types of MHSs. **h:** Metaplots showing the enrichment of ATAC-seq reads and epigenetic signatures around summits for different types of MHSs. “Syn (Acc)” means sequence-syntenic MHS that showed consistent accessible chromatin. “Syn (Inacc)” means sequence-syntenic MHS that showed inconsistent accessible chromatin. “Nonsyn” means non-syntenic MHS. **i and j:** Syn (Inacc) MHSs specifically enriched motifs relative to Syn (Acc) (**i**) and Nonsyn (**j**).



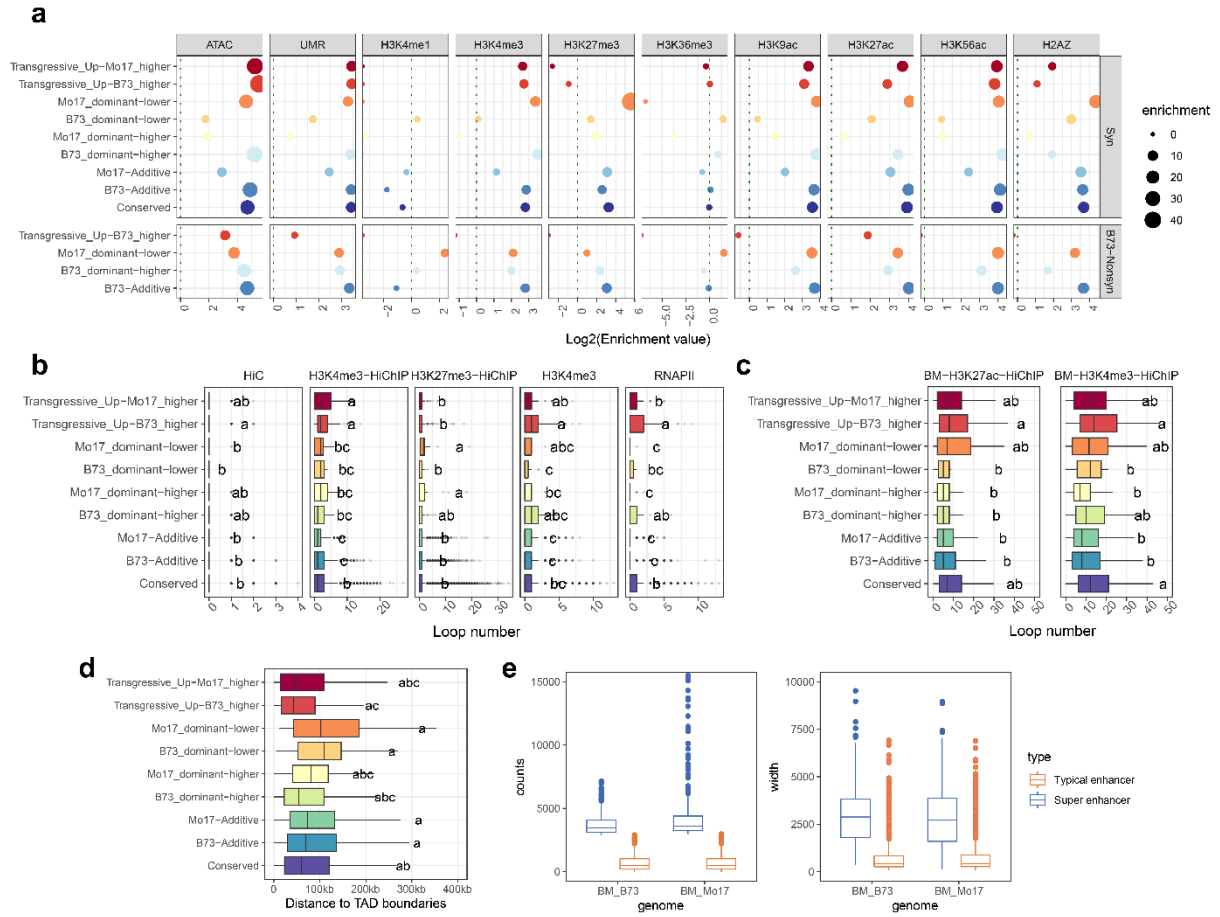
Supplementary Fig. 2. Differential chromatin accessibility of MHSs between hybrids and parents.

a: Correlation between two biological replicates of BM MH-seq in B73 and Mo17 genomes. **b:** Distribution of total and novel MHSs in the BM hybrid. Stacked bar chart showing the composition of total MHSs identified in the BM hybrid across three genomic contexts: Syn (Syntenic regions), B73-Nonsyn (B73-specific non-syntenic regions), and Mo17-Nonsyn (Mo17-specific non-syntenic regions). **c:** Distribution of Log2 of fold change (between hybrid and parents) for syntenic and non-syntenic MHSs. **d:** The number of MHSs that showed significant differential chromatin accessibility between hybrids and parents. For **c** and **d**, Left: BM hybrid versus B73; right: BM hybrid versus Mo17.

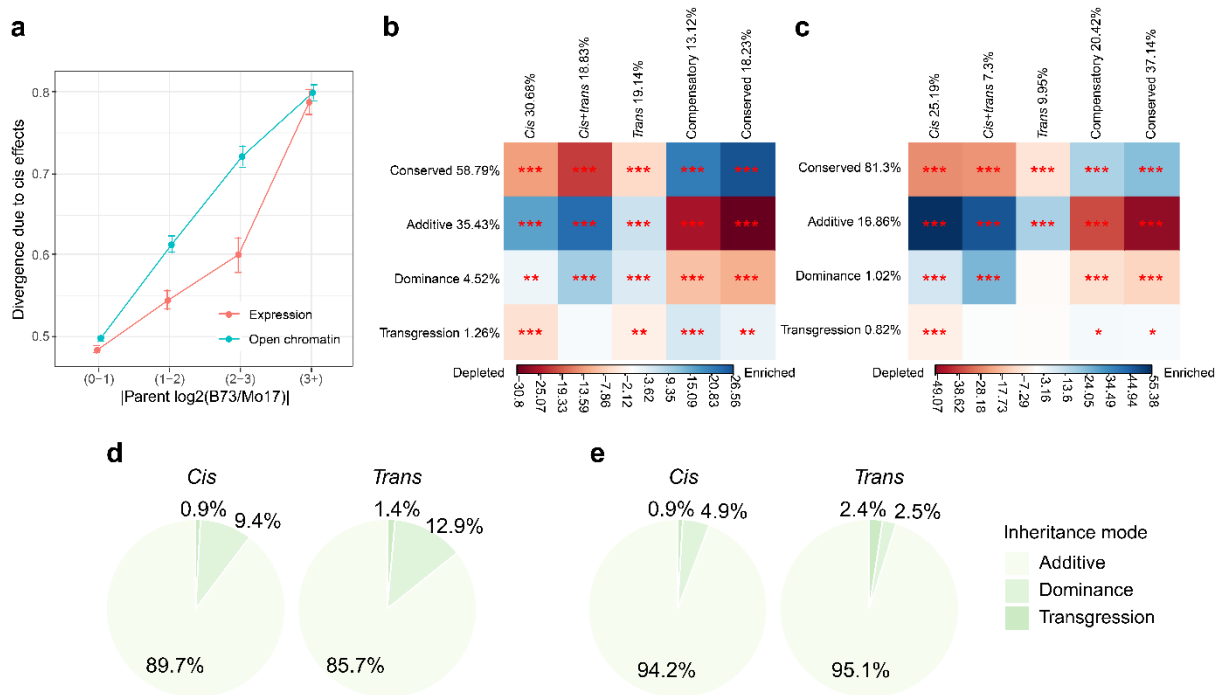


Supplementary Fig. 3. Analysis of inheritance modes for MHSs and genes. a: Assigning inheritance modes for MHSs and genes. Left panel: The number of MHSs assigned to each inheritance mode in the

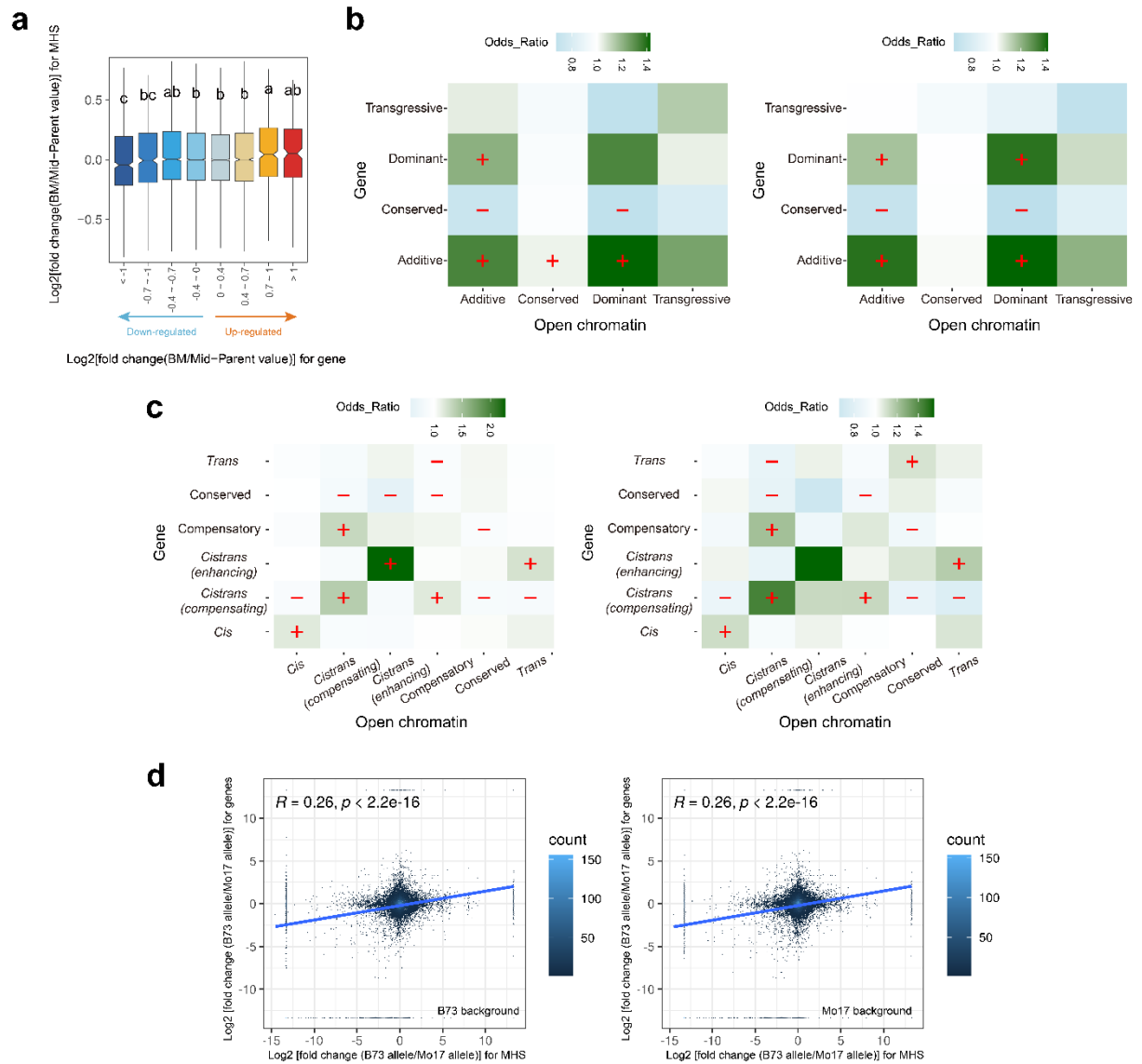
syntenic or non-syntenic region. Right panel: The number of genes assigned to each inheritance mode in each category. **b:** Schematic illustration of three different types of genes categorized by GeneTribe, namely “reciprocal best hits” (RBHs), “single-side best hits” (SBHs) and singleton genes. The relationships between the B73 and Mo17 genomes were directional. **c:** The proportion of genome distribution for MHSs in each inheritance mode. **d:** The proportion of flanking gene types for MHSs in each inheritance mode. **e:** The expression levels in parental genomes of flanking genes for MHSs in each inheritance mode. For “Syn_B73” and “B73-Nonsyn”, the expression level in the B73 parental genome was shown; for “Syn_Mo17” and “Mo17-Nonsyn”, the expression level in the Mo17 parental genome was shown. The middle bars show the median, and the bottle and top of each box indicate the 1/4 and 3/4 percentiles, respectively. The whiskers extend to 1.5 times the interquartile range. **f:** Fraction of MHSs of each inheritance mode overlapping with CNS (based on the Mo17 reference). Sample sizes (*n*) represent the MHS numbers in different inheritance modes. The middle bars show the median, and the bottle and top of each box indicate the 1/4 and 3/4 percentiles, respectively. The whiskers extend to 1.5 times the interquartile range. Multiple comparisons were conducted by the LSD method and performed for syntenic and non-syntenic MHSs respectively. Different letters above the plots indicate significant differences ($P < 0.05$, Bonferroni correction) in a pairwise comparison. **g:** The SNPs/InDels density around the summit of MHSs in each inheritance. For **c-g**: “Syn_B73” represents syntenic MHSs on the B73 coordinates; “Syn_Mo17” represents syntenic MHSs on the Mo17 coordinates; “B73-Nonsyn” represents B73 non-syntenic MHSs; “Mo17-Nonsyn” represents Mo17 non-syntenic MHSs. Source data are provided as a Source Data file.



Supplementary Fig. 4. Analysis of genomic features for MHSs. a: The enrichment of epigenomic signatures for MHSs in different inheritance modes. **b and c:** The number of chromatin loops in B73 (**b**) and BM F1 hybrid (**c**) associated with different inheritance modes of syntenic MHSs. **d:** Distances to topologically associating domains boundaries for different inheritance modes of syntenic MHSs. **e:** Genomic characteristics associated with putative super enhancers (or CRE-clusters) and typical enhancers identified in the BM hybrid in the B73 and Mo17 genome background (BM_B73 and BM_Mo17). Left: Normalized MH-seq read counts. Right: Constituent width. Sample sizes (n) = 552, 21548, 552 and 21532 for BM_B73 super-enhancer, BM_B73 typical-enhancer, BM_Mo17 super-enhancer and BM_Mo17 typical-enhancer, respectively. For **b**, **c** and **d**, Sample sizes (n) represent the MHS numbers in different inheritance modes. The middle bars show the median, and the bottle and top of each box indicate the 1/4 and 3/4 percentiles, respectively. The whiskers extend to 1.5 times the interquartile range. Multiple comparisons were conducted by the LSD method. Different letters above the plots indicate significant differences ($P < 0.05$, Bonferroni correction) in a pairwise comparison. Source data are provided as a Source Data file.

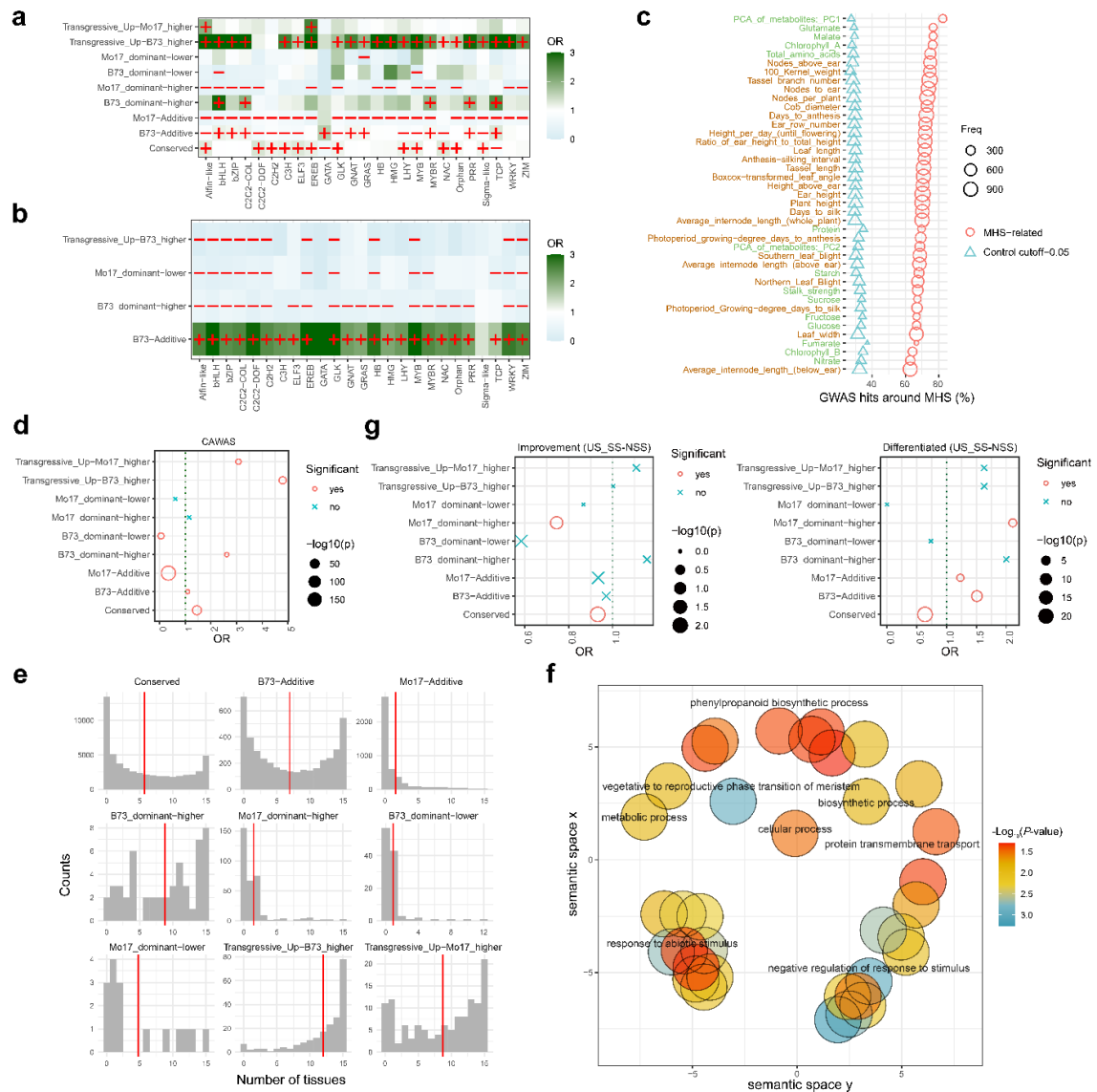


Supplementary Fig. 5. Analysis of regulatory mechanisms for MHSs and genes. **a:** The proportion of overall regulatory divergence that can be explained by *cis*-regulatory divergence, with error bars representing 95 percent confidence intervals. Total divergence ($\log_2$ of parental fold change) was divided into four bins: 0–1, 1–2, 2–3, 3+. **b and c:** The enrichment or depletion of each inheritance mode in each regulatory category for gene expression (**b**) and open chromatin (**c**). The color shows the magnitude of enrichment (blue) and depletion (red) based on residuals of the Chi-square test of independence. Statistical significance was derived from Fisher's exact test, as indicated by * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. **d and e:** The proportions of additive and non-additive genes (**d**) and open chromatin (**e**) for *cis*-only and *trans*-only regulatory categories.

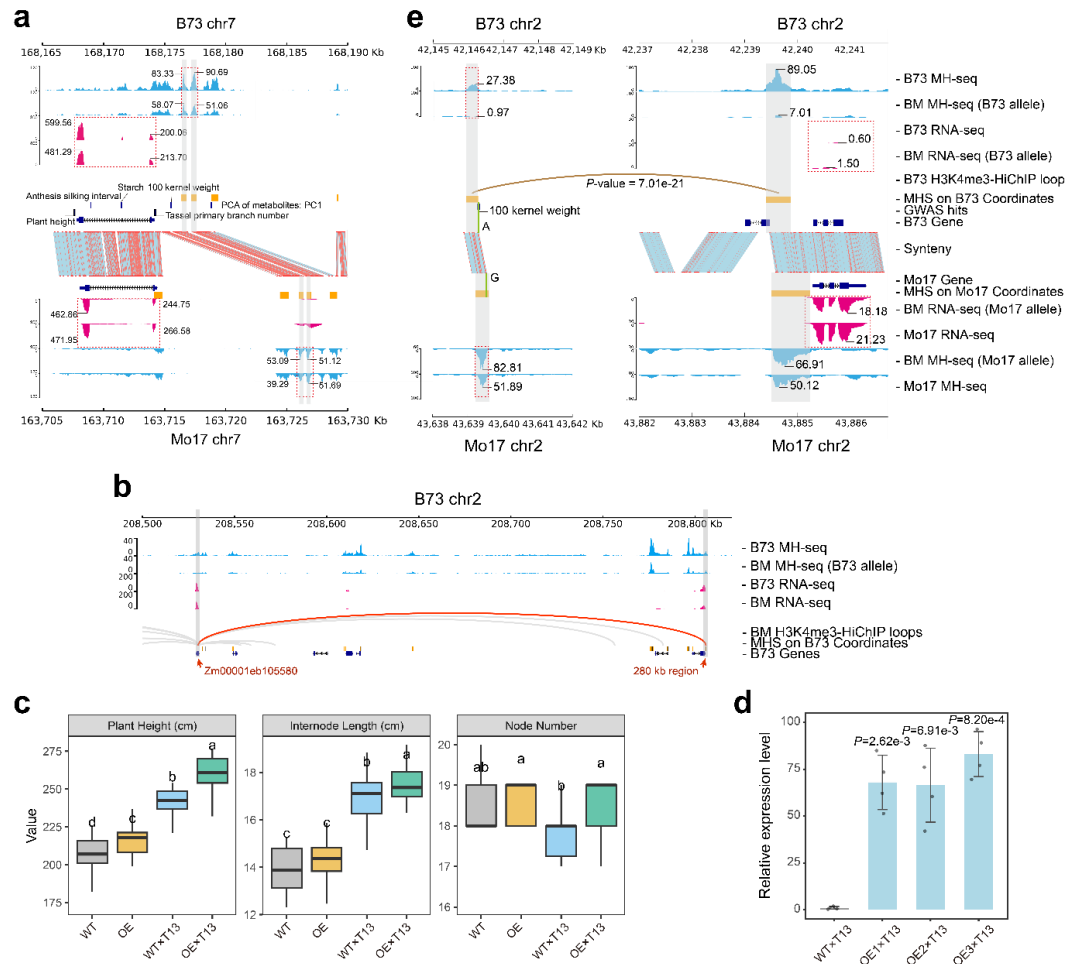


Supplementary Fig. 6 Analysis of the association between open chromatin and gene expression.

a: Comparisons of log2-fold change for MHS near different groups of genes during the hybridization. The middle bars show the median, and the bottle and top of each box indicate the 1/4 and 3/4 percentiles, respectively. The whiskers extend to 1.5 times the interquartile range. Sample sizes (n) represent the gene numbers in different groups. Multiple comparisons were conducted by the LSD method. Different letters above the plots indicate significant differences ($P < 0.05$, Bonferroni correction) in a pairwise comparison. **b** and **c:** The association between genes and their nearby open chromatin in inheritance modes (**b**) and regulatory mechanisms (**c**). The results in B73 (Left panel) and Mo17 (Right panel) genome backgrounds are shown. For each gene expression classification, separate logistic regressions were conducted, odds ratios were mean-centered. Absolute Z-scores greater than 2 are marked with plus (enriched) or minus (depleted) signs. **d:** The Pearson correlation coefficient between allele-specific gene expression divergence and allele-specific chromatin accessibility divergence.

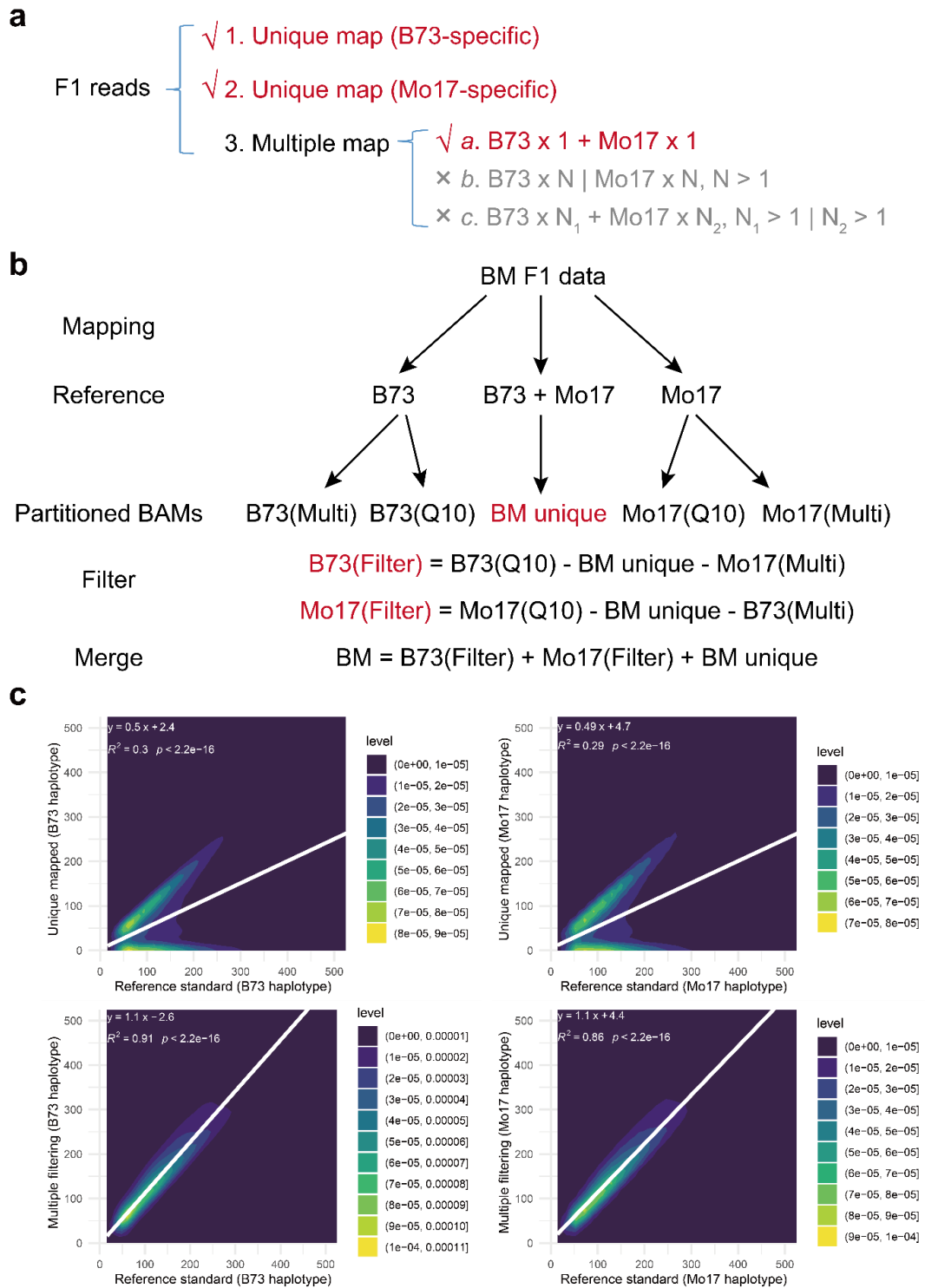


Supplementary Fig. 7. Enrichment of MHSs in TFBSs, GWAS hits, CAWAS hits, and improvement regions. **a** and **b**: The enrichment analysis of syntenic MHSs (A) and non-syntenic MHSs (B) in various types of transcription factor binding sites (TFBSs), performed by Fisher's exact test. **c**: MHS is significantly enriched in genome-wide association studies (GWAS) hits compared to random genome control. 1000 simulations were performed, and the genomic locations of the actual B73 MHS distribution were randomly permuted. The 95th quantile for GWAS hits was defined as a genome-wide random control cutoff. **d**: The enrichment analysis of different types of syntenic MHSs in chromatin accessibility genome-wide association (CAWAS) hits. **e**: Histogram showing the MHSs with different inheritance modes that conserved or specifically existed in ATAC-seq peaks across 15 major maize tissues. The red line indicates the mean number of tissues. **f**: Enriched gene ontology terms for putative targeted genes regulated by the transgressive MHSs. Semantically similar GO terms should remain close together in the plot. Circle color indicates the $-\log_{10}$ -transformed P -value. **g**: The enrichment analysis of different types of syntenic MHSs in improvement regions and differentiated regions of the SS $\times$ NSS heterotic group. SS, stiff stalk; NSS, non-stiff stalk. Source data are provided as a Source Data file.

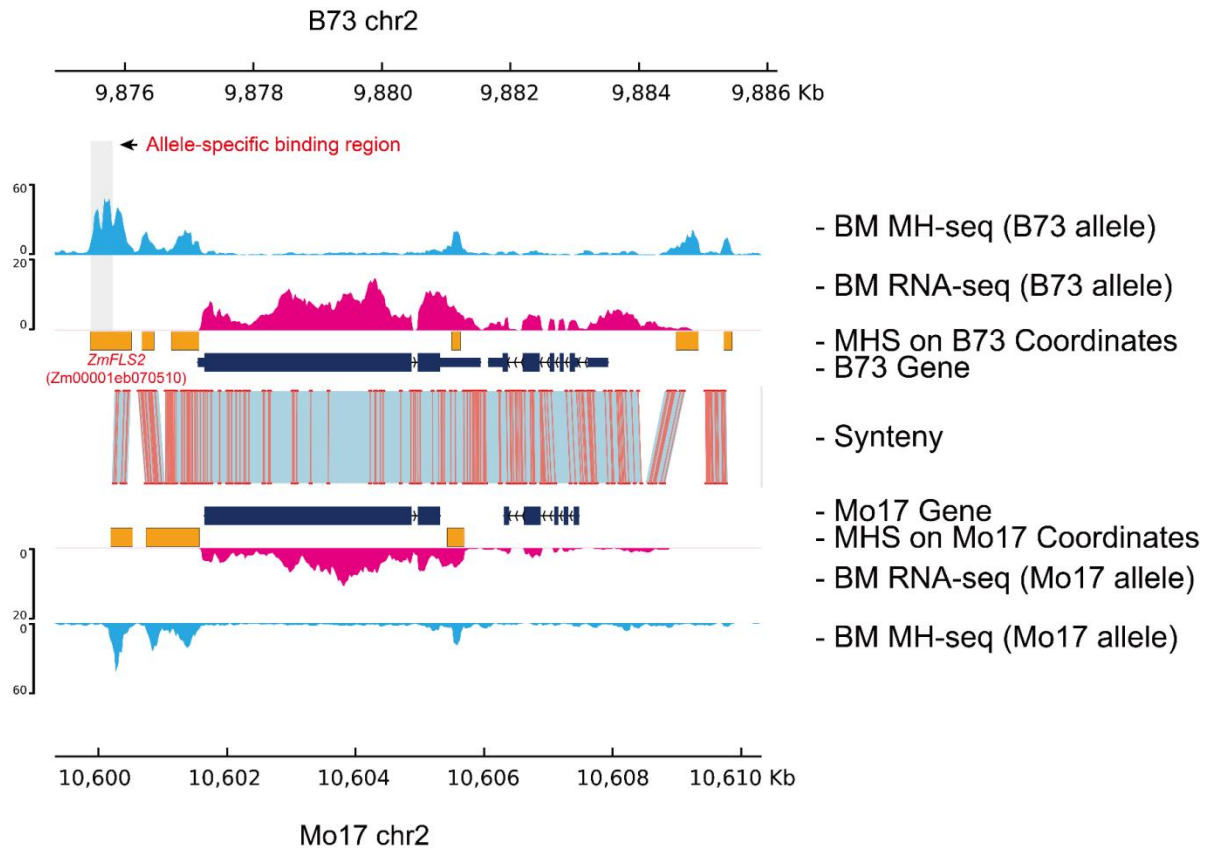


Supplementary Fig. 8. Examples of non-additivity open chromatin contribute to heterosis.

a: Case of proximal non-additivity MHS-regulated gene expression. **b:** The chromatin interaction between the 280 kb region and the *sm2* promoter in the F1. **c:** Phenotypic analysis for plant height, internode length and node number across the wild-type LH244 (WT), *sm2*-overexpression (OE) line, and their respective F1 hybrids (WT×T13 and OE×T13) at Hainan in 2026 ($n = 25, 20, 22$ and 29 for WT, OE, WT×T13 and OE×T13, respectively). Sample sizes (n) represent the MHS numbers in different inheritance modes. The middle bars show the median, and the bottle and top of each box indicate the 1/4 and 3/4 percentiles, respectively. The whiskers extend to 1.5 times the interquartile range. Multiple comparisons were conducted by the LSD method. Different letters above the plots indicate significant differences ($P < 0.05$, Bonferroni correction) in a pairwise comparison. **d:** Relative expression level of *sm2* in OE×T13 and WT×T13 lines. Data are presented as mean values $\pm$ SD from independent biological replicates ($n = 4$ for all lines). Pairwise two-sides t -tests were performed between WT and each *sm2*-OE line. **e:** Case of distal non-additivity MHS-regulated gene expression according to the active chromatin interaction. For **a**, **b** and **e**: MH-seq, RNA-seq, chromatin interaction loops, GWAS hits and B73-Mo17 syntenic blocks information are shown in the tracks. For BM MH-seq and RNA-seq data, B73 and Mo17 allele abundance are shown separately; the sum of the B73 allele and Mo17 allele is the total abundance in the hybrid. For syntenic blocks, red lines indicate the SNPs/InDels between B73 and Mo17. Red dashed boxes indicate genes or MHSs in a non-additive mode of inheritance. Source data are provided as a Source Data file.



Supplementary Fig. 9. Descriptions and rationale for the read filtering and quantification steps of the B73×Mo17 hybrid data. **a:** The cases when a read from the F1 is aligned to the F1 genome. **b:** The workflow to filter and quantify F1 data. **c:** The correlation coefficients (R^2) between actual quantification results and the reference standard for the simulated B73×Mo17 hybrid MH-seq dataset. The P -values ($<2.2e-16$) were calculated by two-sided t test. The traditional method, which often retained only uniquely mapped reads for quantification, showed a relatively lower correlation level ($R^2 \sim 0.29-0.30$) than our multiple filtering strategy in this study ($R^2 \sim 0.86-0.91$).



Supplementary Figure 10. Validation of the allele-specific binding region of *ZmFLS2* with B73xMo17 hybrid MH-seq data. BM MH-seq signal around previously reported allele-specific binding region near *ZmFLS2* (10.1186/s13059-023-02909-w; Fig. 5c). The allele-specific binding pattern was aligned with our MH-seq quantification results, the allele-specific MH-seq signal was found in the B73 allele (indicated by the arrow). MH-seq, RNA-seq, and B73-Mo17 syntenic blocks information are shown in the tracks. For BM MH-seq and RNA-seq data, B73 and Mo17 allele abundance are shown separately; the sum of the B73 allele and Mo17 allele is the total abundance in the hybrid. For syntenic blocks, red lines indicate the SNPs/InDels between B73 and Mo17. Red dashed boxes indicate genes or MHSs in a non-additive mode of inheritance.

Supplementary Table 1. The reproducibility of MH-seq data.

	Rep1	Rep2	IDR filtered (proportion %)
B73	129464	79963	62699 (48.43-78.42%)
Mo17	77123	87473	62984 (72.00-81.67%)
BM (B73 ref)	90285	83404	60316 (66.81-72.32%)
BM (Mo17 ref)	87969	84008	60588 (68.87-72.12%)

Supplementary Table 2. The effective utilization rate of F1 and parent MH-seq data.

	Rep	Q10 Reads		Total	Effective data
B73	1	69073021		200417392	34.46%
	2	76939091		194991302	39.46%
Mo17	1	141258598		363755470	38.83%
	2	84847558		219203560	38.71%
	Rep	Q10 Reads		Total	Effective data
BM	1	89727915	56832232	313289286	46.78%
	2	59546046	35022900	235361070	40.18%

Supplementary Table 3. Primers used for RT-qPCR.

Primer name	Sequence (5' to 3')
SM2-qPCR-F	TGATGCCAGTTCTCGTCGACC
SM2-qPCR-R	CGTATCGCTGTGGCGATGTTG

Supplementary Table 4. Primers used in the LUC experiment.

Primer name	Sequence (5' to 3')
SM2-LUC-F	TATAGGGCGAATTGGGCAATGAAATGAATGGAATATTCAG
SM2-LUC-R	TAGAACTAGTGGATCGCCTTGCGCACAAGCTTC

Supplementary reference

1. Hufford, M. B. *et al.* De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655–662 (2021).