

Open chromatin reveals the functional maize genome

Eli Rodgers-Melnick^{a,1}, Daniel L. Vera^{b,c}, Hank W. Bass^b, and Edward S. Buckler^{a,d,1}

^aInstitute for Genomic Diversity, Cornell University, Ithaca, NY 14853; ^bDepartment of Biological Science, Florida State University, Tallahassee, FL 32306-4295; ^cCenter for Genomics and Personalized Medicine, Florida State University, Tallahassee, FL 32306; and ^dUS Department of Agriculture–Agricultural Research Service, Ithaca, NY 14853

Contributed by Edward S. Buckler, April 18, 2016 (sent for review December 21, 2015; reviewed by Paul Chomet, Sylvain Glémin, and Peter M. Visscher)

Cellular processes mediated through nuclear DNA must contend with chromatin. Chromatin structural assays can efficiently integrate information across diverse regulatory elements, revealing the functional noncoding genome. In this study, we use a differential nuclease sensitivity assay based on micrococcal nuclease (MNase) digestion to discover open chromatin regions in the maize genome. We find that maize MNase-hypersensitive (MNase HS) regions localize around active genes and within recombination hotspots, focusing biased gene conversion at their flanks. Although MNase HS regions map to less than 1% of the genome, they consistently explain a remarkably large amount (~40%) of heritable phenotypic variance in diverse complex traits. MNase HS regions are therefore on par with coding sequences as annotations that demarcate the functional parts of the maize genome. These results imply that less than 3% of the maize genome (coding and MNase HS regions) may give rise to the overwhelming majority of phenotypic variation, greatly narrowing the scope of the functional genome.

chromatin | biased gene conversion | maize | recombination | variance partitioning

All cellular processes involving the nuclear DNA, including transcription, recombination, and replication, must contend with local chromatin states. Each of these processes can affect phenotypic variation, either directly or through constraints on natural selection. To date, humans and several model systems with small genomes have had their chromatin landscapes well-characterized (1–4). However, with a limited number of well-studied large, complex genomes, many general principles relating chromatin structure to genome regulation remain unknown. Here, we examine the large genome of maize (*Zea mays* L.), a model crop species. The importance of maize within international agriculture motivates our central question: Which portions of the genome contribute to quantitative trait variation? Several features of maize biology, such as the capacity for large controlled crosses, rapid decay of linkage disequilibrium (LD), high genetic diversity, and substantial spacing between genes, make it an excellent experimental system for pursuing this question.

Fine-scale characterization of the chromatin structural landscape requires an assay that can distinguish accessible (open) from condensed chromatin at the nucleosomal to subnucleosomal scales. In general, chromatin accessibility may be assayed through in situ digestion of the nuclear DNA with a non-sequence-specific nuclease, followed by quantification of the resulting DNA fragments (5). Micrococcal nuclease (MNase) cleaves DNA between nucleosomes, revealing genome-wide nucleosome occupancy (6, 7). However, differential sensitivity to MNase digestion also reveals chromatin accessibility, with genomic regions of open chromatin preferentially recovered under light-digestion relative to heavy-digestion conditions (8). In maize, a differential MNase sensitivity assay with microarray quantification [differential nuclease sensitivity (DNS)-chip] demonstrated that MNase hypersensitive (MNase HS) regions are positively associated with gene expression levels, conserved noncoding sequences, and known transcription factor binding sites (8). However, that study was limited to 11 Mb (0.5%) of the maize genome. In this study, we use differential MNase sensitivity and high-throughput sequencing (DNS-seq) to determine whether the results in the microarray-targeted regions extend to the

entire genome and to test the genome-wide relationship between chromatin structure and complex trait variation.

Here, we report that open chromatin regions are strongly associated with gene expression, epigenetic modifications, and patterns of meiotic recombination. The genetic variants within MNase HS regions consistently explain an unexpectedly large proportion (40%) of the heritable variance in complex traits, despite making up less than 1% of the maize genome. Given that the remaining variance is primarily explained by the 2% of the maize genome from protein coding regions, this study greatly narrows the scope of the functional genome.

Results

MNase Hypersensitive Regions Surround Genes. To enhance the genotype-phenotype map with chromatin structural data, we carried out genome-wide DNS-seq mapping in nuclei from root and shoot tissues of maize seedlings. The MNase HS regions, summarized in Fig. 1, make up 0.6% of the genome. Tissue-specific comparisons reveal considerable overlap between root and shoot MNase HS regions (Fig. 1C). On a megabase scale, local MNase HS frequency is strongly associated with recombination frequency ($R^2 = 0.55$) (Fig. 1A and *SI Appendix*, Fig. S1). We obtain a conservative estimate for the total number of MNase HS regions by merging signals within 150 bp, which is a nucleosome core length, of one another. A total of 126,992 (17.0 Mb) distinct MNase HS regions occur in shoot, and 89,455 (6.9 Mb) occur in root. Combined, these regions cover 19.4 Mb. We measured sequence conservation by

Significance

The maize genome, similar to those of most plant genomes, is 98% noncoding. Much of the remainder is a vast desert of repeats that remain repressed throughout the cell cycle. The plant cell orchestrates its complex activities by restricting access to functional regions with an open chromatin configuration. Here, we identify the small portion (<1%) of the maize genome residing in open chromatin. We demonstrate that open chromatin predicts molecular phenotypes such as gene expression and recombination. Furthermore, we show that genetic variation within open chromatin regions accounts for ~40% of phenotypic variation in agronomic traits. By greatly narrowing the scope of the functional maize genome, this study can help to accelerate the pace of crop improvement through highly focused genomic selection and genome editing.

Author contributions: E.R.-M., D.L.V., H.W.B., and E.S.B. designed research; E.R.-M. and D.L.V. performed research; E.R.-M., D.L.V., H.W.B., and E.S.B. contributed new reagents/analytic tools; E.R.-M., D.L.V., H.W.B., and E.S.B. analyzed data; and E.R.-M., D.L.V., H.W.B., and E.S.B. wrote the paper.

Reviewers: P.C., NRGene; S.G., CNRS, Montpellier; and P.M.V., University of Queensland. The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The sequence reported in this paper has been deposited in the NCBI Sequence Read Archive (accession no. SRA302258).

¹To whom correspondence may be addressed. Email: esb33@cornell.edu or eli.rodgers-melnick@pioneer.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1525244113/-DCSupplemental.

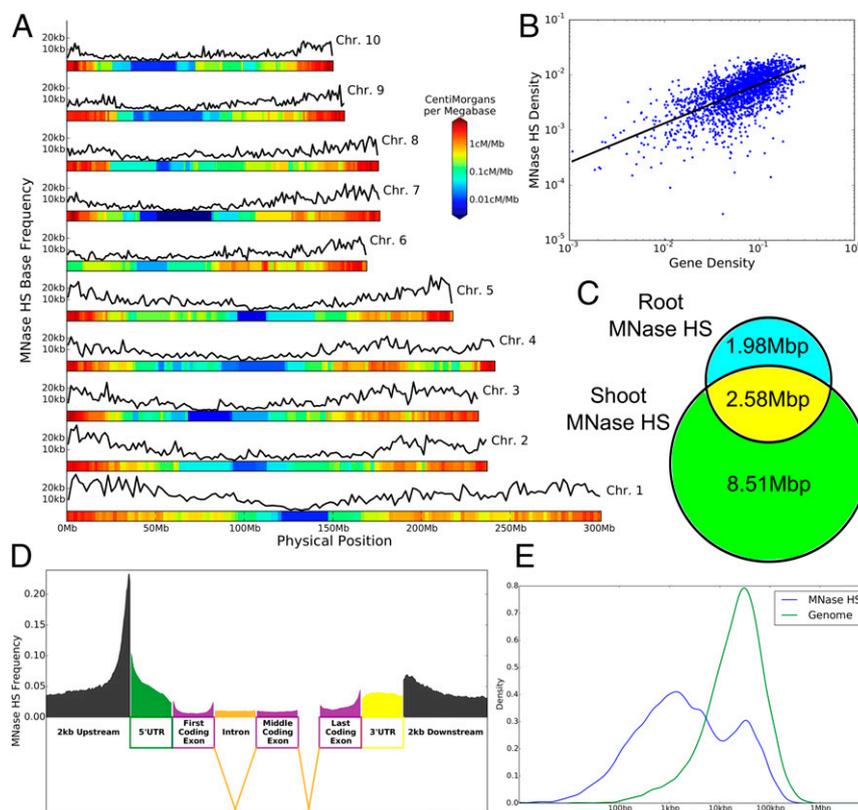


Fig. 1. The distribution of MNase HS regions in maize. (A) The frequency of MNase HS bases across the genome in 1-Mb windows, along with the recombination frequency. (B) The relationship between gene density and MNase HS density in 1-Mb windows. (C) The total sizes of MNase HS regions in the root and shoot and the intersection of the two tissues. (D) MNase HS base frequency within and surrounding protein coding genes. Genic elements were binned according to percentage of total element size, whereas upstream and downstream regions were binned in units of 10 bp. (E) The distribution of distances to the nearest gene boundary for intergenic MNase HS bases (blue) and for all intergenic bases (green).

maximum likelihood scaling of the neutral phylogenetic tree with GERP (9), such that scale values below 1 indicate higher conservation than expected under neutrality. Overall, sequence in MNase HS regions is significantly, albeit slightly, more conserved than 180 bp flanking regions everywhere except coding sequence (*SI Appendix, Fig. S2*), indicating that purifying selection in MNase HS regions is elevated with respect to the noncoding expectation, but remains diffuse. Although slight differences in sequence conservation occur between MNase HS sequences from different tissue types, differences in scaling factors are inconsistent between functional annotations and lower in magnitude than those between MNase HS and non-MNase HS flanking regions (*SI Appendix, Fig. S3*).

Given the known role of chromatin structure in gene function, we anticipated a close relationship between the distribution of MNase HS and genes. On a broad scale (1 Mb), MNase HS density positively correlates with gene density (Spearman $\rho = 0.64$; $P < 1 \times 10^{-16}$) (Fig. 1B). Approximately 80% of MNase HS regions occur outside genes, but they are highly enriched in genic flanks (Fig. 1E). A second peak, ~ 100 kb, in the MNase HS to gene distance frequencies aligns with the overall genomic distribution (Fig. 1E), and these may represent a mixture of gene distal regulatory elements and additional gene proximal elements for unannotated genes. Approximately 70% of all mapped MNase HS sequences occur within the small (11%) portion of the intergenic space that excludes transposable elements (TEs) (*SI Appendix, Table S1*). Within and surrounding genes, averaged MNase HS profiles show stereotypical patterns (Fig. 1D), with prominent hypersensitivity just upstream of gene model transcription start sites (TSSs). Approximately 25% of genes contain statistically significant peaks in shoot or root MNase HS (*Materials and Methods*).

A prominent but broad HS peak, covering and extending beyond the 3' ends of transcript models, is evident, whereas coding and intronic regions show much lower MNase HS.

MNase HS Is Associated with Gene Regulation. Using microarrays spanning 1,688 genes, Vera et al., 2014 (8) showed that gene expression levels and MNase HS strongly covary at TSSs. To examine this relationship in a less-biased and genome-wide manner, we sorted the 36,441 maize genes by their steady-state mRNA levels and used heat map analysis to inspect MNase HS profiles at the TSSs (*SI Appendix, Fig. S4*). We detect similar numbers of expressed genes (reads per kilobase per million ≥ 0.1) in shoot ($n = 25,549$) and root ($n = 26,020$) (*SI Appendix, Table S2*). Gene expression levels and signal strength for MNase HS regions show a clear genome-wide positive correlation around TSSs. Tissue-specific expression patterns are recapitulated by MNase HS signals, confirming that TSS chromatin profiling is predictive of gene expression levels (Fig. 2). TSS-associated MNase HS signals also discriminate between the expression levels of paralogs from the most recent tetraploidization event 5–12 MYA (10). Although large portions of the duplicated genome are differentially expressed (11), the level of MNase HS matches gene expression levels (*SI Appendix, Fig. S5*). These observations establish the relative signals of promoter MNase HS, possibly reflecting the transcription rate, as one of the best epigenomic predictors of gene activity.

MNase HS Regions Are Associated with DNA Hypomethylation and Recombination Hotspots. Heterochromatin in plants is usually distinguished by constitutive hypermethylation of symmetric CpG and

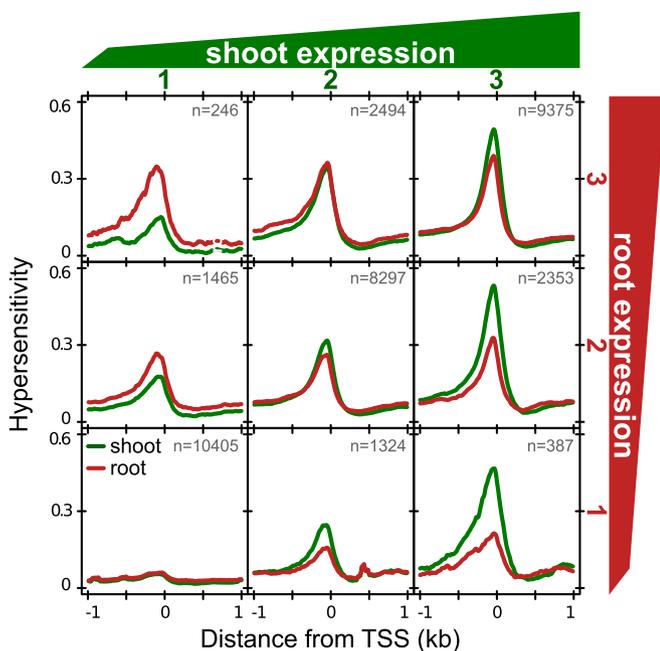


Fig. 2. MNase HS is associated with gene regulation. Mean MNase HS profiles in root (red) and shoot (green) for genes divided into tertiles according to expression levels within root and within shoot.

CHG DNA methylation motifs, where H refers to any nucleotide except G (12), so we expected localized symmetric methylation reductions in HS regions. Plants also contain methylation at asymmetric CHH motifs, which in maize are known to generally mark sites of RNA-dependent DNA methylation (13), so we anticipated elevated CHH methylation within TEs. However, we did not anticipate a dramatic localized relationship of MNase HS regions with CHH methylation. In support of reduced symmetric methylation surrounding MNase HS regions, CpG methylation outside TEs is reduced from a rate of 70% 2 kb away from the nearest MNase HS site to 5% within the MNase HS region (Fig. 3). A similar reduction occurs for CHG motifs. Likewise, a strong MNase HS-associated hypomethylation tendency is seen within TEs, although the difference in CpG and CHG methylation is accompanied by a four times increase in CHH methylation. Strikingly, CpG methylation in coding regions differs from the general pattern in noncoding regions. Coding CpG methylation is elevated within and directly surrounding MNase HS sites, and stronger hypomethylation occurs downstream, relative to upstream, of the MNase HS region.

In maize, meiotic recombination has a strong negative relationship with CpG and CHG methylation (14), linking chromatin status with crossover formation. Investigating this relationship, we find significant and genomic context-specific MNase HS enrichment at 1–30-kb recombination hotspots relative to comparable, adjacent, but nonhotspot regions. Although MNase HS is enriched threefold to fourfold within hotspot TEs, twofold within hotspot non-TE intergenic regions, and twofold within hotspot noncoding genic regions, it is not enriched within the coding regions of recombination hotspots (*SI Appendix, Fig. S6A*). The positive relationship between MNase HS and recombination frequencies extends to gene-distal regions, those >5 kb from the nearest gene, where we find a high association (Spearman $\rho = 0.35$; $P < 1 \times 10^{-16}$) for 1–10-kb bins (*SI Appendix, Fig. S7*). Mean crossover enrichment in these 1–10-kb bins is 16% higher (Wilcoxon rank sum test, $P = 3.6 \times 10^{-5}$) for shoot-specific compared with root-specific MNase HS regions (*SI Appendix, Fig. S8*).

Strong, Sustained GC-Biased Gene Conversion Surrounds MNase HS Regions. In light of the association between seedling MNase HS regions and recombination hotspots, we tested whether MNase HS regions are also enriched for tracts of GC-biased gene conversion. Conversion tracts arise when base mismatches at recombination junctions resolve in favor of G+C nucleotides relative to A+T nucleotides (15, 16). We assessed the historical influence of GC-biased gene conversion on the maize lineage, using PHASTbias with an alignment of 12 monocots and eudicots (17), although limiting the alignment to only grasses had no significant effect (*SI Appendix, Fig. S9*). The mean probability of biased gene conversion increases fivefold in the 2 kb surrounding MNase HS regions (*SI Appendix, Fig. S6B*). Tract sizes are also positively associated with MNase HS frequency within the tract and the immediate flanking regions (Spearman $\rho = 0.107$; $P < 1 \times 10^{-16}$). We find that the highest historical conversion frequencies are located in 300 bp surrounding regions of HS tracts within hotspot coding sequence (mean probability = 0.34) (Fig. 4B and *SI Appendix, Fig. S10B*). Hotspot MNase HS regions in both coding and noncoding genic regions also show a 1.5- to twofold increase in historical conversion frequency compared with adjacent control regions. No similar increases occur within intergenic regions (Fig. 4B). Thus, although hypersensitivity is most enriched within hotspot noncoding regions, genic coding regions have the highest conversion rates (Fig. 4B and *SI Appendix, Fig. S10B*). Using a linear model of mean historical conversion rate on crossover enrichment, MNase HS frequency, genic frequency, and TE frequency in 1–10-kb windows (*SI Appendix, Table S3*), we find that crossover enrichment consistently contributes the most to explained variance (46%), followed by MNase HS frequency (32%). Thus, MNase HS regions may contribute to GC-biased gene conversion beyond their contribution to recombination frequency, perhaps through an increase in mismatch repair-associated conversion resulting from an elevated mutation rate. However, the errors associated with estimates of narrow crossover enrichments in 5,000 recombinant inbred lines hinder clear delineation of GC-biased gene conversion causes.

In general, the conversion pattern surrounding MNase HS regions mirrors the trend toward increasing GC composition in the flanking 1 kb (Fig. 4A and *SI Appendix, Fig. S10A*). Strikingly, the GC content of putatively neutral coding sites increases to nearly 80% in the few hundred base pairs flanking coding MNase HS regions, whereas GC content within coding MNase

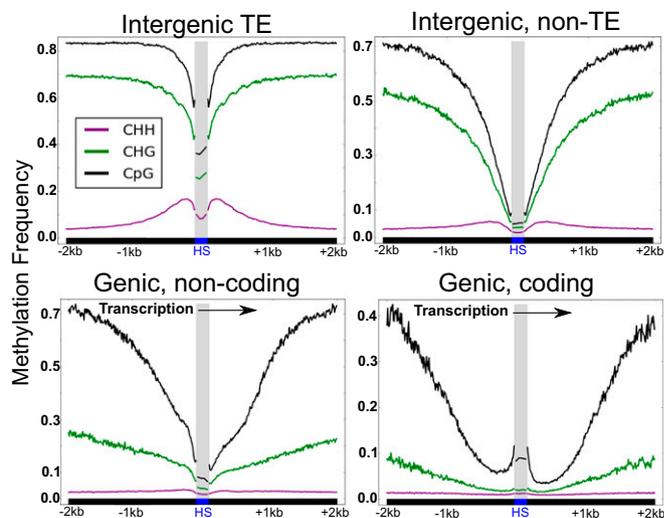


Fig. 3. MNase HS regions are associated with DNA hypomethylation. Mean DNA methylation profiles within and surrounding MNase HS regions.

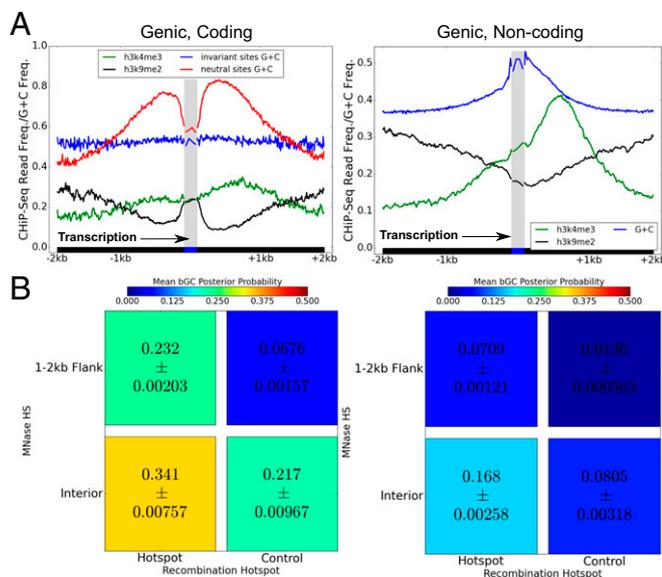


Fig. 4. GC-biased gene conversion strongly affects coding sequence content surrounding MNase HS regions. (A) The frequency of G/C content and the histone modifications H3K9me2 and H3K4me3 surrounding and within MNase HS regions for coding and genic, noncoding (e.g., UTR, intron) sites. Base content for coding sites is divided into invariant and neutral sites. Base positions are plotted relative to the direction of transcription. (B) Mean ranges of GC-biased gene conversion probabilities within MNase HS regions and regions 1–2 kb from the nearest MNase HS region for both recombination hotspots and control regions flanking recombination hotspots. Ranges correspond to 95% credible intervals.

HS regions is similar to that of MNase HS regions elsewhere (50–60%). Moreover, GC content in all genic regions is noticeably increased downstream of MNase HS regions, with respect to the direction of transcription (Fig. 4A).

The ongoing action of enhanced biased gene conversion in coding regions immediately flanking MNase HS is supported by derived allele frequencies being more skewed toward G/C vs. A/T alleles (SI Appendix, Fig. S11). Using the approach of Glémin et al. (18), we used the site frequency spectrum to estimate the current GC-biased gene conversion intensity ($B = 4N_e b$), where N_e is the

effective population size and b is the conversion bias intensity (SI Appendix, Fig. S12). Intensity is greatest ($B = 2.4$) in the 100–200 bp of coding sequence surrounding MNase HS within or proximal to coding sequence, although conversion intensity within the interior of coding MNase HS is also high ($B = 1.9$). Within genic, noncoding sequence, conversion intensity is highest within the MNase HS interior ($B = 0.88$), and it is also centered on MNase HS tracts within gene distal regions of both TEs ($B = 1.65$) and non-TE ($B = 0.45$) regions. Extending beyond current population conversion rates, higher historical substitution rates of A/T to G/C vs. G/C to A/T in these same regions indicate localized GC-biased gene conversion has remained a consistent evolutionary force (SI Appendix, Fig. S13). Although the mechanistic basis for enhanced conversion frequencies is unclear, we find strong associations with epigenetic marks. In particular, the frequency of the histone modification H3K9me2 mirrors base content patterns in both coding and noncoding portions of genes (Fig. 4A). In contrast, H3K4me3 shows the same pattern in coding and noncoding regions. Strikingly, the absence of H3K9me2 clearly separates the high GC (~80%) MNase HS-flanking regions from flanking regions with comparable GC content to the MNase HS interior (~60%) (SI Appendix, Fig. S14). H3K4me3 has a much smaller association with GC content in the opposite direction, and both effects are significant in coding and noncoding segments.

MNase Hypersensitivity Marks Known Quantitative Trait Loci. Many cloned maize quantitative trait loci (QTL) fall within nongenic regulatory regions, prompting us to examine several cases of intergenic QTL for possible underlying explanatory MNase HS regions. The first case is the QTL associated with an expression increase in teosinte branched1 (*tb1*), a domestication gene with alleles for the single stalk form of modern field corn (19). Variation in *tb1* expression levels maps to genetic variation in a region containing two TE insertions, *Hopscotch* and *Tourist*, located ~60 kb upstream of *tb1*. The presence of the proximal *Hopscotch* element causes increased *tb1* expression, whereas the distal region containing *Tourist* represses expression (20). Intriguingly, we observe MNase HS in the 730 bp between these two TEs (Fig. 5A), possibly localizing a distal TE-modulated enhancer as the underlying *tb1* domestication QTL. We note that the *Hopscotch* element itself, too repetitive to measure, may also contain MNase HS sequences.

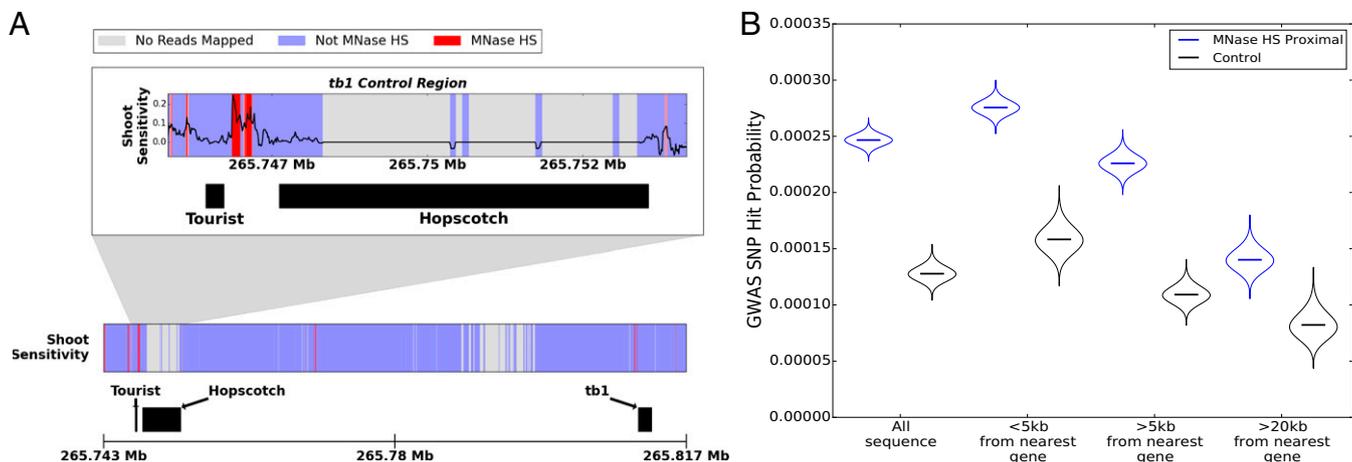


Fig. 5. MNase HS marks known QTL. (A) Profile of shoot hypersensitivity in the *tb1* region. Regions with a MNase HS Bayes factor above 1 are shown in red, and the opacity is proportional to the significance of the difference. The putative regulatory region containing the *tourist* and *hopscotch* TEs is enlarged. Regions without uniquely mapping reads are shown in gray. Regions covered by reads but without sufficient evidence for hypersensitivity are shown in blue. (B) Posterior distributions for the relative frequency of GWAS SNPs among total SNPs within 2 kb of a MNase HS region vs. control regions ≥ 2 kb from the nearest MNase HS region, directly flanking the MNase HS proximal regions.

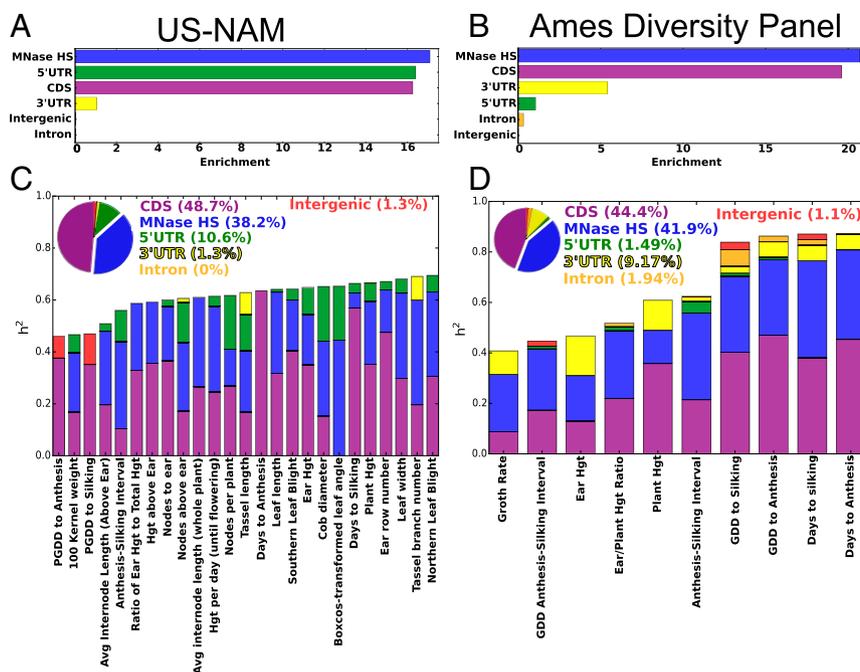


Fig. 6. MNase HS regions explain nearly half of quantitative trait variation. (A and B) The enrichments of variance in functional categories within (A) US-NAM and (B) Ames Diversity Panel (C and D) The average contributions of SNPs with the given annotations to heritable variance and the partitioning of variance to individual traits within (C) US-NAM and (D) Ames Diversity Panel. Exploded slices in the pie charts denote the MNase HS.

A second case involves the vegetative to generative transition 1 (*vgt1*) QTL, associated with altered expression of a maize flowering-time repressor gene, AP2 domain transcription factor, *ZmRap2.7*. Insertion of a miniature inverted-repeat transposable element (MITE) into conserved noncoding sequence in the QTL is associated with reduced expression of *ZmRap2.7* and early flowering time (21, 22). Correspondingly, we observe a strong MNase HS signal overlapping the MITE insertion point (*SI Appendix, Fig. S15A*). Finally, a third case involves the prolificacy1.1 (*prol1.1*) QTL, found to contain several prominent MNase HS regions (*SI Appendix, Fig. S15A*). This QTL is presumed to contain a *cis*-regulatory element that increases expression of grassy tillers1 (*gt1*), a gene that acts to suppress secondary ear outgrowth (23). These examples reveal the potential for chromatin structural data to narrow the focus from ~100-kb windows to small intervals as a means to help identify actual enhancers near genes.

MNase HS Regions Explain Quantitative Traits. To more globally test the relationship between MNase HS and complex trait variation, we quantified the enrichment of genome-wide association study (GWAS) hits across 41 traits within 2 kb of MNase HS regions. We find twofold enrichment of GWAS hits in MNase HS proximal regions (95% credible interval, 1.73–2.16) compared with adjacent regions at least 2 kb away from the nearest MNase HS site (Fig. 5B). Furthermore, this enrichment is nearly unchanged as a function of distance to the nearest gene (Fig. 5B), revealing comparable MNase HS efficacy for mapping functionally significant genomic loci in gene-proximal versus gene-distal regions. To further investigate the role of open chromatin in complex organismal trait variation, we partitioned the heritable phenotypic variance into annotation-specific components using methods previously applied to human variation (24). We classified SNPs into coding (CDS), MNase HS, 5' UTR, 3' UTR, intronic, and intergenic regions, in that order of priority. A genetic relationship matrix was constructed for each SNP set, allowing for jointly estimated heritability explained by genetic variation

within each annotation. We examined the United States nested association mapping (US-NAM) population, a set of 5,000 recombinant inbred lines from B73 × 25 diverse inbreds in controlled crosses. We found that the CDS explains the highest proportion of heritable variance (mean = 47.6%) for traits with moderate to high heritability ($h^2 > 0.4$), whereas MNase HS regions explain a remarkably large majority of remaining variance (mean = 39.3%) (Fig. 6C). This 18× enrichment of variation attributed to SNPs within MNase HS (Fig. 6A) is comparable to that (16×) for CDS SNPs. Similar results, 21× enrichment for MNase HS SNPs, are obtained from variance partitioning of flowering time and plant height phenotypes of a different mapping population, the Ames Diversity Panel (Fig. 6B and D). Splitting variance among MNase HS categories, we find that the largest contribution of MNase HS heritability is intergenic for both mapping panels (*SI Appendix, Fig. S16*).

Considering the MNase HS variance in shoots versus roots, we expected that the shoot chromatin profiles might better explain trait variation because of the aboveground nature of the measured phenotypes. Indeed, shoot-only MNase HS regions explain an average of 60% (US-NAM panel) or 75% (Ames panel) of the MNase HS heritable variance, whereas the root-only regions explain less than 20% (*SI Appendix, Fig. S17 C and D*). However, given the different sizes of the SNP sets within each tissue category, the level of enrichment varies. In US-NAM, the MNase HS regions common to both root and shoot show the most enrichment (24×) compared with those unique to shoot (19×) or root (16×). In the Ames panel, the MNase HS regions unique to shoot show the most enrichment (26×) compared with those either unique to root (23×) or common to both (8.3×) (*SI Appendix, Fig. S17 A and B*).

Several types of reliability testing for variance partitioning were performed. We note that SNP error rates do not appear to bias the observed results, as revealed by a lack of enhancement of SNP quality scores among MNase HS SNPs relative to SNPs from other categories (*SI Appendix, Fig. S18D*). In addition, little to no variance components are allocated to all SNP sets when phenotypes are permuted both within and between families,

another sign of unbiased categories (*SI Appendix, Fig. S1C*). As a test of robustness to patterns of LD, we tested partitioning with the separate category of intergenic SNPs 2–5 kb away from the nearest MNase HS region. This category explains little heritability in either population (*SI Appendix, Fig. S19 A and B*). We also reversed the order of annotation priority in SNP classes, but obtained essentially the same results for the Ames panel, but with more variance allocated to the UTRs within the US-NAM panel for several phenotypes (*SI Appendix, Fig. S19 C and D*). Finally, we performed variance partitioning on 250 simulated phenotypes for each annotation, in which only SNPs within that annotation contributed to phenotypic variance. Using this method, for which we know causal locations, we find that most variance is unambiguously allocated to the correct SNP set in both populations (*SI Appendix, Fig. S18 A and B*).

Discussion

In this study, we demonstrate that relatively open chromatin structure, mapped with MNase HS profiling, marks functionally important regions that link genotype to phenotype. MNase HS regions delineate molecular phenotypes such as recombination breakpoints, enhancers, and other possible remote controllers of gene expression. We demonstrate how to use MNase HS regions as epigenomic annotations to resolve the variation underlying organism-level quantitative traits at high resolution.

We observe more MNase HS sites, in both number and total coverage, for the shoot compared with the root sample. The elevated signal in the shoot may reflect its greater overall tissue complexity with respect to cell type differentiation and development potential (reviewed in ref. 25). Moreover, the seedling harvest time coincided with major shoot developmental events, including the onset of autotrophy, the juvenile to adult vegetative phase change, and the genome-wide epigenetic change coupled to transposon expression and gene silencing pathways (26). Taken together, the detection of more open chromatin in shoots versus roots is likely reflective of more numerous genomic activities distributed across the genome, collectively measured by DNS-seq.

Regarding gene regulation, we find a prominent and consistent positive relationship between gene expression and chromatin accessibility surrounding the TSS. The strength of this relationship is so strong that one might be able to predict transcription rates directly from the DNS profiles without measuring transcript levels. However, three possible scenarios are considered in which the promoter DNS signals and transcript abundance could appear to be uncoupled, yet maintain the DNS–transcription rate relationship on a per cell basis. In one case, genes exhibiting low promoter MNase HS signals may produce high mRNA levels, a pattern that could result from a gene that is very highly expressed in a small proportion of cells. In a contrasting case, a gene with high promoter MNase HS signals may produce low mRNA levels, a pattern that could result from high, posttranscriptional mRNA turnover. However, another case could involve the mapping of MNase HS regions to the wrong gene, as might occur at promoters from nearby genes. Overall, however, most genes clearly show a robust positive relationship between RNA abundance and promoter MNase HS signals, likely related to transcriptional initiation and its dependence on open chromatin.

DNA methylation rates also closely mirror the chromatin accessibility landscape. Cytosine hypomethylation of open chromatin is common across eukaryotes, as reduced methylation within CpG contexts occurs in plants and animals (1, 4, 27). In contrast to the ubiquitous hypomethylation of cytosines in symmetric motifs, we observe a twofold to fourfold increase in CHH methylation in the region immediately surrounding and MNase HS regions in TEs. CHH hypermethylation within DNaseI HS regions was previously observed within *Arabidopsis thaliana* (27), and it co-occurs with accessible chromatin in the RNA-dependent DNA methylation-targeted regions of maize (13). Given that RNA-dependent DNA

methylation requires active transcription of siRNAs, MNase HS regions may universally mark transcriptionally active DNA, including regions coupled to genomic silencing pathways.

Beyond gene expression, we find compelling evidence that open chromatin marks recombination hotspots. This relationship differs from that in humans, where PRDM9 marks the locations of recombination hotspots but without a strong relationship to open chromatin (28, 29). An unexpected finding with possible implications for mechanisms of recombination in plants is the relationship between open chromatin and GC-biased gene conversion. In coding sequence, where we observe the highest conversion rates, the allele frequencies of derived neutral alleles with conversion advantages are increased nearly 2.5-fold above those of disfavored alleles. Moreover, the patterns of substitution within the phylogeny indicate that GC-biased gene conversion has remained a consistent, localized force around MNase HS regions. This situation differs from that in humans, where the high conversion sites are inconsistent between populations and species (18), likely because of the rapid evolution of PRDM9 motifs (30). Because GC-biased gene conversion imposes a fitness-independent selective force, slightly deleterious alleles may become fixed if their selective disadvantages are sufficiently less extreme than their conversion advantages (31). Indeed, GC-biased gene conversion may increase disease burden in humans by up to 60% (32). However, high levels of GC-biased gene conversion may also favor increased heterosis when deleterious mutations are highly recessive (33). Our estimates of the population conversion intensity indicate that GC-biased gene conversion is strong enough to overcome genetic drift, especially within coding regions. Therefore, this localized, nonadaptive force may substantially increase maize genetic load and contribute to the heterotic patterns observed in breeding germplasm through complementation of these deleterious variants.

The mechanistic basis of conversion rate differences in coding versus noncoding genic sequence is unknown. One possibility is that optimal codon selection increases GC content in coding regions. However, codon selection does not explain the conversion bias localizing around MNase HS regions, specifically those lacking H3K9me2. If the same recombination-promoting mechanisms occur in coding and noncoding segments, the differences in base content may result from histone mark stability over deep evolutionary time within coding sequence. The support for any direct relationship between H3K9me2 and recombination is nonetheless tenuous. In *A. thaliana*, knockout of DNA polymerase α causes localized reduction of H3K9me2 and concomitant increases in recombination frequency (34). H3K4me3 also correlates with meiotic recombination hotspots in *A. thaliana* (35), and from our results, shows a slight positive relationship to recombination in maize. However, one or more chromatin marks correlated with those we assayed; for instance, H2A.Z, which is strongly associated with *A. thaliana* hotspots (35), could be the causative factor. The interrelationships among chromatin structure, epigenetic marks, and recombination control remain largely undefined, but their investigation bears on evolutionary paradigms and agricultural breeding strategies.

Relationship to Quantitative Traits. We observe consistent, robust relationships between genetic variation in MNase HS regions and complex trait variation, establishing an epigenetic framework for the discovery and analysis of enhancers and other genome-wide regulatory sites. As a group or genomic annotation, MNase HS regions represent a promising collection of gene regulatory candidates underlying quantitative loci, including long-distance expression QTL. Several lines of evidence support this idea. First, MNase HS regions colocalize with several well-studied intergenic QTL previously fine-mapped as regulatory regions: *tb1*, *vgt1*, and *pro1.1*. Notably, for these intergenic QTL, B73 has the high expression alleles. Alternate lines with structural variation

corresponding to the less-expressed alleles may lack the same open chromatin signals as a result of structural disruption or epigenetic modification such as heterochromatin spreading. Comparative genomics using MNase HS regions will become increasingly feasible as multiple other maize inbred lines (haplotypes) are sequenced. Second, the relationship of MNase HS regions to functional polymorphisms is supported by the enrichment of GWAS SNPs within the proximity of MNase HS regions. A previous study of maize quantitative traits found evidence for significant GWAS hit enrichment in gene-proximal intergenic regions (36). However, GWAS studies can only provide evidence for causal variants in LD with tagged SNPs, so nearby genetic causal variants cannot be dismissed. In contrast, our results show twofold GWAS hit enrichment near open chromatin, regardless of genic proximity. This twofold enrichment is smaller than the ~20-fold enrichment for explained variance from partitioning analysis. Two nonexclusive explanations may explain this discrepancy. First, the stochastic history of mutation and recombination will produce inconsistent GWAS resolution, blurring the true location of causal mutations. Second, the relationship of GWAS hit quantity to explained variation depends on the distribution of allelic effects, possibly skewed toward larger values in open chromatin.

Our findings show that SNPs within MNase HS regions explain ~40% of the heritable variance of quantitative traits in multiple maize mapping populations. The scale and scope of this finding are remarkable, and thus worthy of considerable scrutiny, prompting us to examine the robustness and validity of our results. First, we note that the pattern of relative contributions to heritability is consistent across traits and populations. The US-NAM panel of 5,000 lines from controlled crosses between 26 diverse founders results in large linkage blocks that permit complete imputation of SNPs. The other population, a non-designed diversity panel, contains high amounts of missing data, but both populations benefit from the comparably rapid rate of LD decay in maize (37, 38). Using permutation testing and simulated phenotypes based on empirical genotypes, we show that the high variance apportioned to MNase HS and CDS regions is not explained by intrinsic bias toward these sites. SNPs outside of a given annotation class can inflate its contribution if they explain heritability and closely flank the annotation of interest (24); however, we find no evidence for such distortion when comparing intergenic MNase HS regions with those from a subset 2–5 kb from the nearest MNase HS region. Nonetheless, the considerable structural variation within maize invites the possibility that MNase HS-proximal indels, rather than genetic variants within the HS region, are often the causal polymorphisms behind the explained variance. Future studies of interline MNase HS variation will therefore focus on the relationship between chromatin state, structural variation, and QTL.

Although we observe a large phenotypic contribution from the MNase HS portion of the genome (40% of heritable variance), this value is only half the heritable variance explained by DNaseI HS regions for 11 human diseases (24). However, as a percentage of the genome, the MNase HS regions in the current study cover more than 50-fold less sequence than the human DNaseI regions. Although the plant genomes currently lack open chromatin profiling data as extensive as that for the human ENCODE project (39), the results of this study, along with previous genomic DNaseI HS profiles of rice and *A. thaliana* chromatin (3, 4, 27), suggest that the genome-wide extent of open chromatin in plants may not substantially scale with increasing genome size. Furthermore, DNase I HS profiles in animals are increasingly shifting from hypersensitivity to general sensitivity, further complicating the definition, let alone comparison, of the distributions of chromatin states across kingdoms. Given that our results support intergenic functional variation explained almost exclusively by open chromatin, these chromatin accessibility assays in

plants are, a priori, at least as informative as protein-coding regions when defining the functional genome.

In summary, we show how DNS mapping can be used to delineate the functional portion of a large, complex genome, using maize as a model genetic system. Biochemical DNS footprints produced in situ are highly localized, and maize offers both a low LD and extensive structured mapping populations. Combined, these organismal and experimental attributes allowed us to measure the effects of local chromatin structure on heritable phenotypic variation at an unprecedented depth and breadth, using only seedling shoot and root tissues of one genotype. Furthermore, we illuminate the relationship of open chromatin to recombination, opening the door for future studies into the targeting of crossovers and the evolutionary consequences of strong, consistent GC-biased gene conversion. In agriculture, epigenomic profiling with DNS-seq can strategically guide the predictive accuracy of genomic selection, narrow candidate regions for experimentation with reverse genetics, and define the functions of intergenic chromatin toward organismal fitness. Overall, DNS profiling has multiple applications ranging from predicting transcription rates and recombination sites to defining enhancers and QTL candidates. As a genomic annotation, they bring an invaluable resource to bear on biological, agricultural, and societal problems, including contemporary and future challenges related to population growth and climate change.

Materials and Methods

Plant Material. Seeds from the maize (*Zea mays* L., cultivar B73) were obtained from field-grown ears from the Buckler laboratory at Cornell University. Seeds were germinated in Fafard Seedling Mix in the greenhouse (Department of Biological Science, Florida State University). Tissue was harvested at 11:00 AM–12:00 PM, 9 d after planting. Shoot tissue was harvested by cutting and collecting the tissue just above the soil line, and root tissue was harvested by rapidly rinsing the root system in water and cutting off the kernel-attached roots. Harvested tissues were immediately flash frozen in liquid nitrogen and stored at -80°C .

Nucleus Isolation and Digestion. As modified from Vera et al. (8), ten grams of tissue were ground under liquid nitrogen with a mortar and pestle and cross-linked by stirring for 10 min in 100 mL ice-cold fixation buffer (15 mM Pipes-NaOH at pH 6.8, 0.32 M sorbitol, 80 mM KCl, 20 mM NaCl, 0.5 mM EGTA, 2mM EDTA, 1 mM DTT, 0.15 mM spermine, and 0.5 mM spermidine) containing 1% formaldehyde. Fixation was stopped by adding glycine to 125 mM. Nuclei were isolated by adding Triton X-100 to 1% final by addition of 0.1 vol of a 10% (vol/vol) Triton X-100 stock, followed by stirring for 10 min. The suspension was filtered through one layer of Miracloth (Calbiochem) and placed in 50-mL centrifuge tubes. In these centrifuge tubes, 35 mL nuclear suspensions were underlaid with 15 mL Percoll cushion composed of 50% (vol/vol) Percoll (GE) in BFA. Nuclei suspensions were centrifuged at $3,000 \times g$ for 15 min at 4°C . The nuclei at the Percoll interface were transferred to a 50-mL tube and diluted twofold with MNase digestion buffer (50 mM Tris-HCl at pH 7.5, 320 mM sucrose, 4 mM MgCl_2 , and 1 mM CaCl_2). Nuclei suspensions were centrifuged at $2,000 \times g$ for 10 min at 4°C , and nuclei pellets were resuspended in 2.5 mL MNase digestion buffer. Nuclei were aliquoted into 500-mL aliquots, flash frozen in liquid nitrogen, and stored at -80°C . Nuclei were thawed at room temperature and digested by adding MNase to 10 U/mL (light) or 100 U/mL (heavy), and incubated at room temperature for 5 min. Digestions were stopped with 10 mM EGTA. Nuclei were de-cross-linked by incubation overnight at 65°C in the presence of 1% SDS and 100 $\mu\text{g}/\text{mL}$ proteinase K. DNA was extracted by phenol-chloroform extraction followed by EtOH precipitation. Digested DNA was resuspended in 40 $\mu\text{g}/\text{mL}$ RNase A and electrophoresed in a 1% agarose gel. DNA fragments smaller than 200 bp were excised and gel extracted after ethidium bromide staining with the Qiaex II gel extraction kit (Qiagen), following the manufacturer's instructions.

Library Preparation and Sequencing. After nuclei isolation and digestion, gel-extracted DNA was used to prepare sequencing libraries using the NEBNext Ultra DNA Library Prep Kit for Illumina (NEB), using manufacturer instructions. Indexed libraries were pooled and sequenced on eight Illumina HiSeq 2500 lanes with paired-end 50-cycle sequencing. Short-read data are deposited in the NCBI short read archive (SRP064243).

Read Assembly and Calling of Hypersensitive Sites. After the computational trimming of adaptor sequences using CutAdapt (40), paired-end reads were mapped to the maize B73 AGPv3 reference genome, using Bowtie2 with options “no-mixed,” “no-discordant,” “no-unal,” and “dovetail” (41) for each replicate digest and for the genomic DNA. BED files were made from the resulting BAM files, using bedtools bamtobed, filtered for minimal alignment quality (≥ 10), and read coverage in 10-bp intervals was calculated using coverageBed (42). The DNS values were obtained by subtracting the mean normalized depth (in reads per million) of the heavy digest replicates from those of the light digest replicates. In this way, positive DNS values correspond to MNase hypersensitive

footprints (as defined by ref. 8; and referred to here as MNase HS regions), whereas negative DNS values correspond to nuclease hyper-resistant footprints (MRF, as per ref. 8). A Bayes factor criterion was used to classify as significantly hypersensitive.

ACKNOWLEDGMENTS. We thank Toby Kellogg from the Donald Danforth Center for providing samples of *Vossia cuspidata* and *Coelorachis tuberculosa*. This work was supported by National Science Foundation Grants IOS-0820619 and IOS-1238014 (to E.S.B.) and IOS-1025954 and IOS-1444532 (to H.W.B.) and the US Department of Agriculture–Agricultural Research Service.

- Thurman RE, et al. (2012) The accessible chromatin landscape of the human genome. *Nature* 489(7414):75–82.
- Kharchenko PV, et al. (2011) Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* 471(7339):480–485.
- Zhang W, Zhang T, Wu Y, Jiang J (2012) Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in *Arabidopsis*. *Plant Cell* 24(7):2719–2731.
- Zhang W, et al. (2012) High-resolution mapping of open chromatin in the rice genome. *Genome Res* 22(1):151–162.
- Tsompana M, Buck MJ (2014) Chromatin accessibility: A window into the genome. *Epigenetics Chromatin* 7(1):33.
- Axel R (1975) Cleavage of DNA in nuclei and chromatin with staphylococcal nuclease. *Biochemistry* 14(13):2921–2925.
- Yuan GC, et al. (2005) Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* 309(5734):626–630.
- Vera DL, et al. (2014) Differential nuclease sensitivity profiling of chromatin reveals biochemical footprints coupled to gene expression and functional DNA elements in maize. *Plant Cell* 26(10):3883–3893.
- Davydov EV, et al. (2010) Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* 6(12):e1001025.
- Swigonová Z, et al. (2004) Close split of sorghum and maize genome progenitors. *Genome Res* 14(10A):1916–1923.
- Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc Natl Acad Sci USA* 108(10):4069–4074.
- Chan SW, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat Rev Genet* 6(5):351–360.
- Gent JJ, et al. (2014) Accessible DNA and relative depletion of H3K9me2 at maize loci undergoing RNA-directed DNA methylation. *Plant Cell* 26(12):4903–4917.
- Rodgers-Melnick E, et al. (2015) Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proc Natl Acad Sci USA* 112(12):3823–3828.
- Chen JM, Cooper DN, Chuzhanova N, Férec C, Patrinos GP (2007) Gene conversion: Mechanisms, evolution and human disease. *Nat Rev Genet* 8(10):762–775.
- Serres-Giardi L, Belkhir K, David J, Glémin S (2012) Patterns and evolution of nucleotide landscapes in seed plants. *Plant Cell* 24(4):1379–1397.
- Capra JA, Hubisz MJ, Kostka D, Pollard KS, Siepel A (2013) A model-based analysis of GC-biased gene conversion in the human and chimpanzee genomes. *PLoS Genet* 9(8):e1003684.
- Glémin S, et al. (2015) Quantification of GC-biased gene conversion in the human genome. *Genome Res* 25(8):1215–1228.
- Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. *Nature* 386(6624):485–488.
- Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat Genet* 43(11):1160–1163.
- Castelletti S, Tuberosa R, Pindo M, Salvi S (2014) A MITE transposon insertion is associated with differential methylation at the maize flowering time QTL *Vgt1*. *G3 (Bethesda)* 4(5):805–812.
- Salvi S, et al. (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc Natl Acad Sci USA* 104(27):11376–11381.
- Wills DM, et al. (2013) From many, one: Genetic control of prolificacy during maize domestication. *PLoS Genet* 9(6):e1003604.
- Gusev A, et al.; Schizophrenia Working Group of the Psychiatric Genomics Consortium; SWE-SCZ Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; SWE-SCZ Consortium (2014) Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am J Hum Genet* 95(5):535–552.
- Poethig RS (2003) Phase change and the regulation of developmental timing in plants. *Science* 301(5631):334–336.
- Li H, Freeling M, Lisch D (2010) Epigenetic reprogramming during vegetative phase change in maize. *Proc Natl Acad Sci USA* 107(51):22184–22189.
- Sullivan AM, et al. (2014) Mapping and dynamics of regulatory DNA and transcription factor networks in *A. thaliana*. *Cell Reports* 8(6):2015–2030.
- Myers S, et al. (2006) The distribution and causes of meiotic recombination in the human genome. *Biochem Soc Trans* 34(Pt 4):526–530.
- Baudat F, et al. (2010) PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* 327(5967):836–840.
- Myers S, et al. (2010) Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* 327(5967):876–879.
- Glémin S (2010) Surprising fitness consequences of GC-biased gene conversion: I. Mutation load and inbreeding depression. *Genetics* 185(3):939–959.
- Lachance J, Tishkoff SA (2014) Biased gene conversion skews allele frequencies in human populations, increasing the disease burden of recessive alleles. *Am J Hum Genet* 95(4):408–420.
- Glémin S (2011) Surprising fitness consequences of GC-biased gene conversion. II. Heterosis. *Genetics* 187(1):217–227.
- Liu J, et al. (2010) Mutation in the catalytic subunit of DNA polymerase alpha influences transcriptional gene silencing and homologous recombination in *Arabidopsis*. *Plant J* 61(1):36–45.
- Choi K, et al. (2013) *Arabidopsis* meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. *Nat Genet* 45(11):1327–1336.
- Wallace JG, et al. (2014) Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genet* 10(12):e1004845.
- Romay MC, et al. (2013) Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol* 14(6):R55.
- Chia JM, et al. (2012) Maize HapMap2 identifies extant variation from a genome in flux. *Nat Genet* 44(7):803–807.
- ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74.
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J* 17(1):10–12.
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.
- Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842.