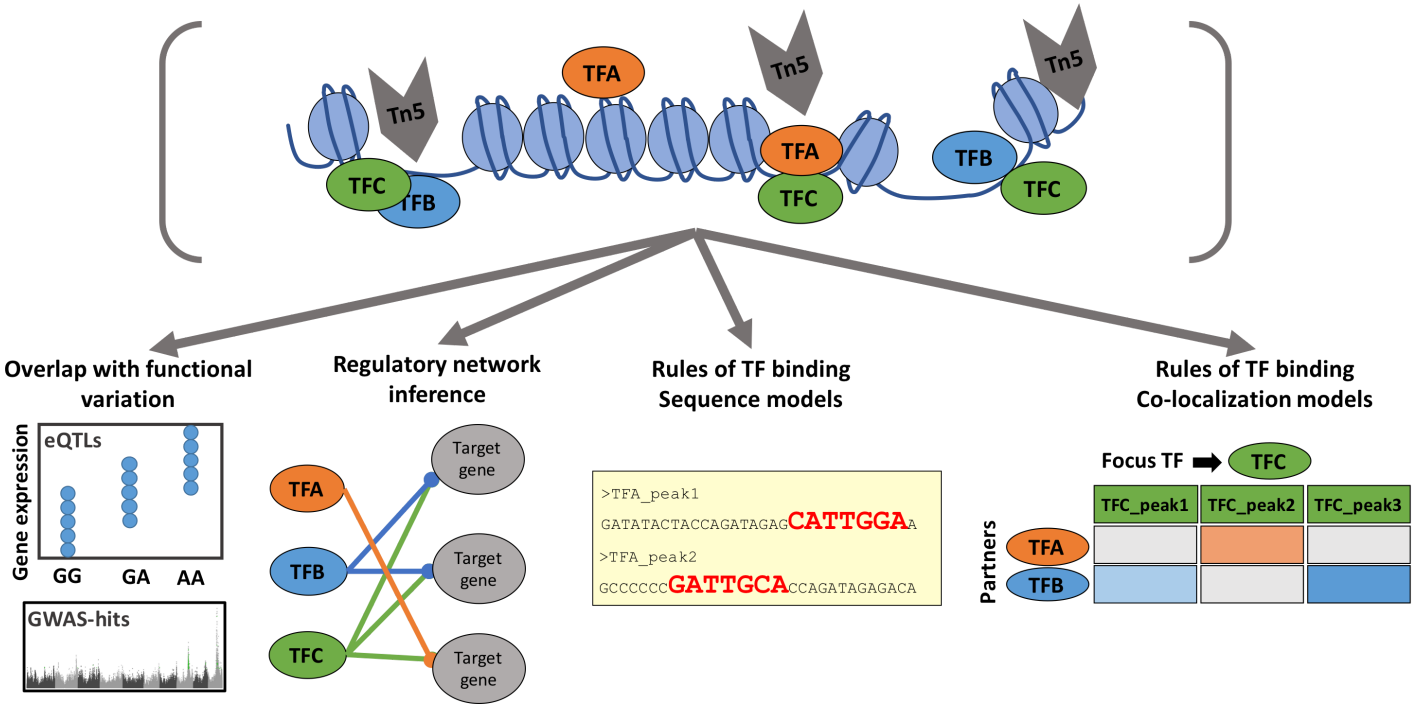
ChIP-seq and ATAC-seq overlap

Code

Here we present a large collection of non-coding functional elements in the maize genome that includes accessible chromatin regions, and to the best of our knowledge, the largest inventory for in-vivo transcription factors (TF) binding profiles in plants (Fig 1A).

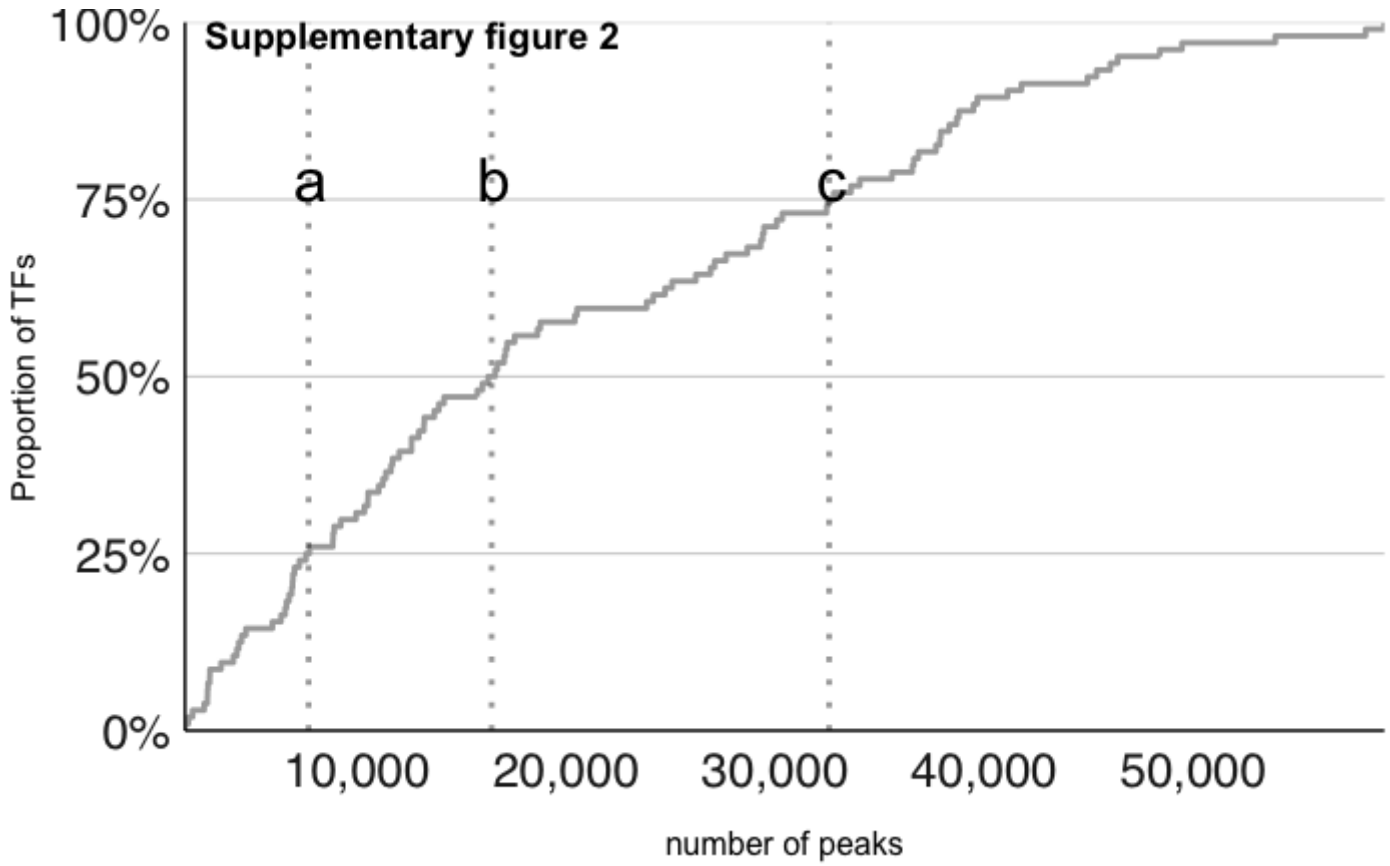
Figure 1A



To map TF binding in a native chromatin context, we developed a modified ChIP-seq assay that eliminates the need for specific antibodies and reduces the amount of material required for library preparation, two major limitations for quality and scalable assays 1 (http://doi.org/10.1101/gr.136184.111). This allowed us to generate libraries for a large number of TFs (104 sequence specific TFs, from 25 TF families) expressed in leaves (B73, leaves of 9-day-old seedlings) (Supplementary table 1). To determine the quality of each library we evaluated the success of the immunoprecipitation step, following quality-control measures from the ENCODE and modENCODE consortia guidelines 1 (http://doi.org/10.1101/gr.136184.111) (Supplementary table 2). Next, we selected experiments that showed reproducibility between biological replicates (Pearson correlation coefficient for aligned reads, $r > 0.8$) to proceed with peak calling (Supplementary figure 1).

In brief, peaks were called with SPP 2 (http://doi.org/10.1038/nbt.1508), followed by the irreproducibility discovery rate (IDR) statistical framework (IDR < 1%) to determine sets of reproducible peaks between biological replicates 3 (http://doi.org/10.1214/11-AOAS466). The number of ChIP-seq peaks varies between TFs, with a median value of ~16K (Interquartile range, IQR_{25-75} 7,664-32,566 peaks), for a total of 2,147,346 peaks (Supplementary figure 2, Supplementary table 3).

Peaks from different TFs were clustered based on their summit positions resulting in 144,890 non-overlapping binding loci, corresponding to ~2% of the maize genome (Supplementary figure 2).

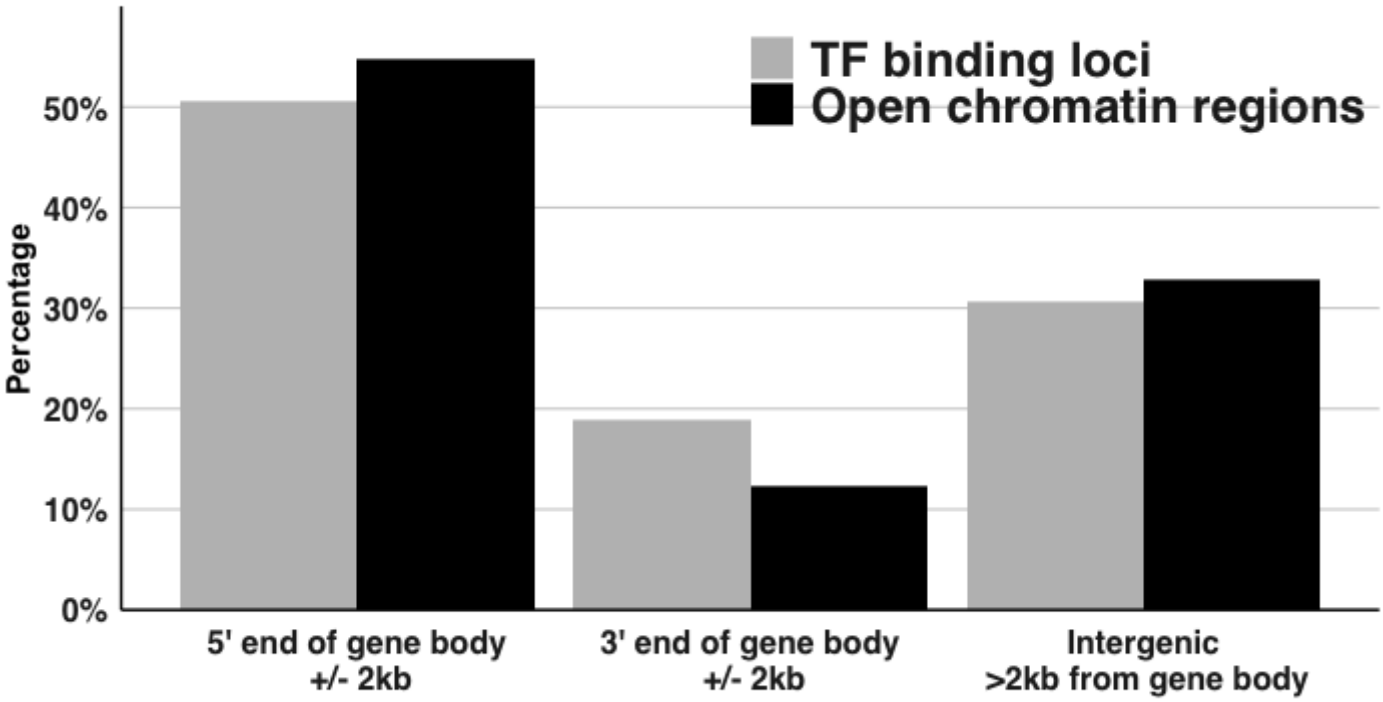


We also profiled chromatin accessibility in mesophyll cells (B73, leaves of 9-day-old seedlings) using ATAC-seq 4 (http://doi.org/10.1038/nmeth.2688) (Supplementary table 4). Aligned reads from two biological replicates were well correlated (Pearson correlation coefficient, r = 0.89), and the data was consolidated into 38,713 open chromatin regions that correspond to ~1% of the maize genome, in agreement with previous observations from MNAse-seq data 5 (http://doi.org/10.1073/pnas.1525244113).

Distribution of TF binding loci with respect to genes and open chromatin distribution are similar, with ~70% of the TF binding loci, and 67% of the open chromatin regions located in the gene proximal regions (+/- 2.5kb of gene), and preferentially towards the 5’ end of genes (Fig 1B).

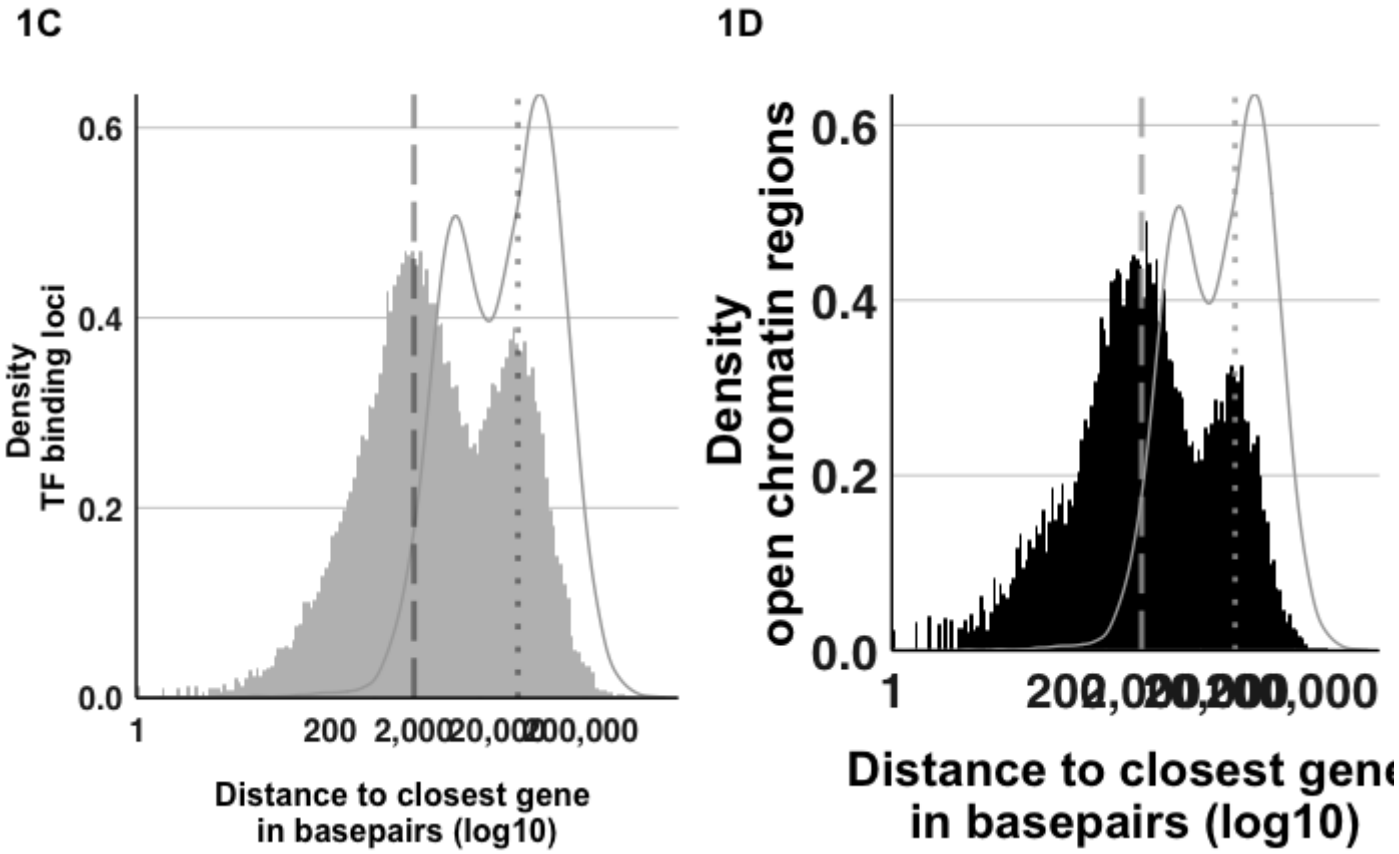
Distribution of TF binding loci and open chromatin regions relative to gene bodies

1B

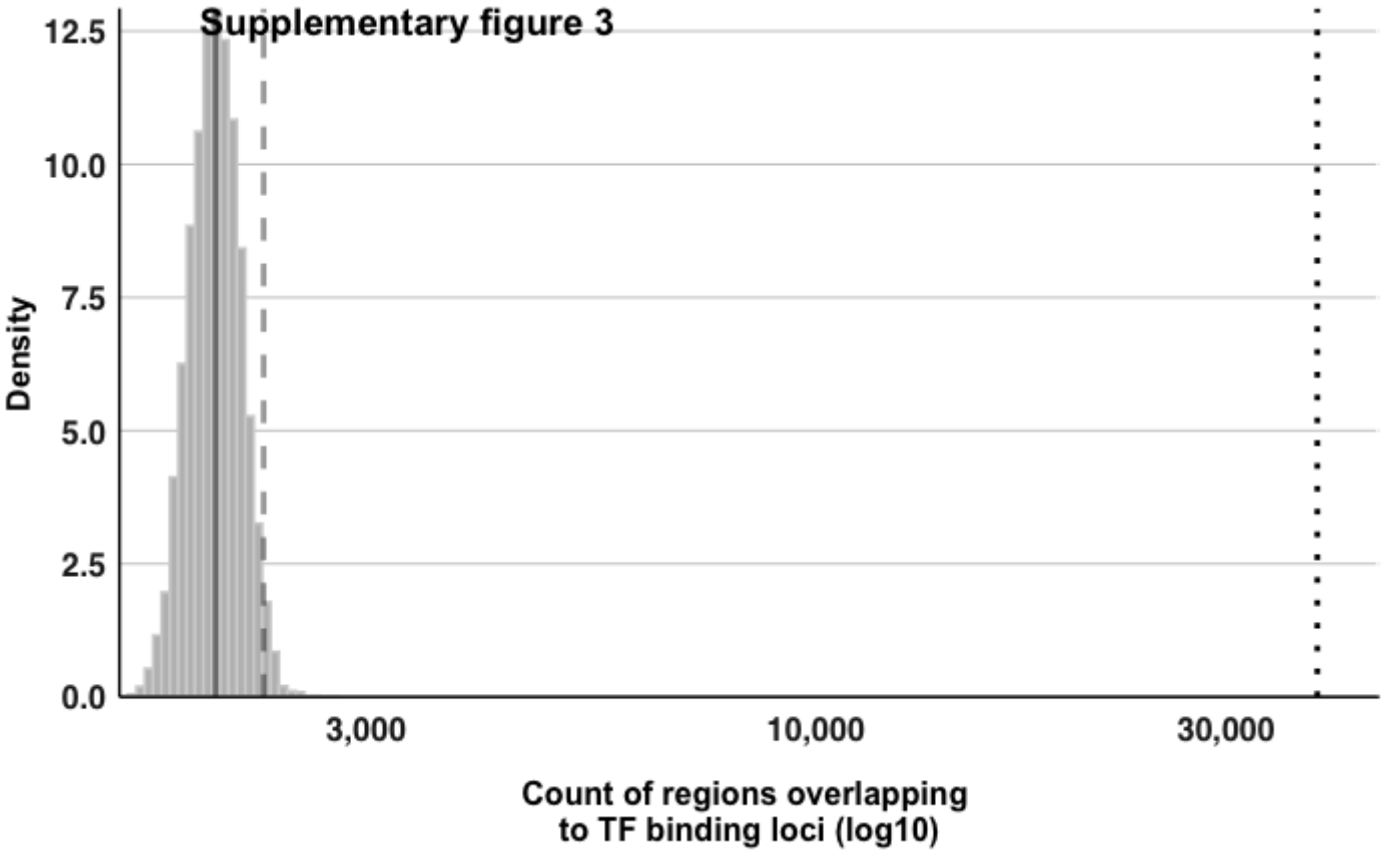


Considering only regions that did not overlap with gene bodies, the distribution of the frequency of the distances between open regulatory regions to its closest gene appears to be bimodal (Fig 1C-D). TF bindings in proximal regions were abundant ~2kb from gene bodies, consistent with a role in regulation of gene expression, and distal regions located at a wide range of distances, with ~15% of the TF binding loci, and ~17% of the open chromatin regions in the range of 10-100kb away from gene bodies, consistent with known long-distance QTLs 6 (http://doi.org/10.1073/pnas.0704145104),7 (http://doi.org/10.1371/journal.pgen.1004745). The proportion of proximal vs. distal regions did not appear to correspond to the available “space” between genes, as genic regions are spread out at larger distances providing more opportunity for distal regulation that the observed (Fig 1C-D).

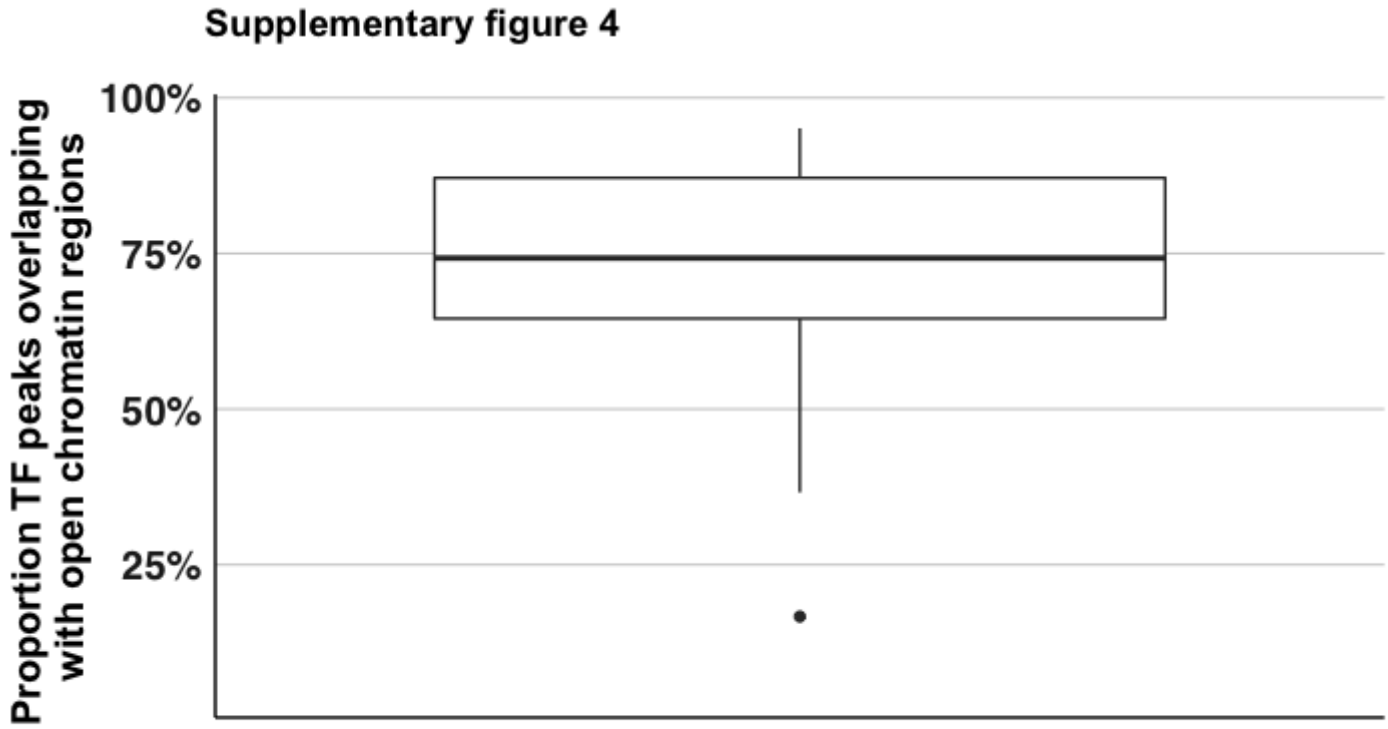
Figure 1C-D - Distribution regulatory regions that are non-overlapping with gene bodies, as a function of gene distance



Layering ATAC-seq and ChIP-seq data further confirmed that TF peaks and open chromatin overlap, with a non-random relationship (H0: 10000 sets of randomly positioned loci, P-value < 10-5) (Supplementary figure 3).

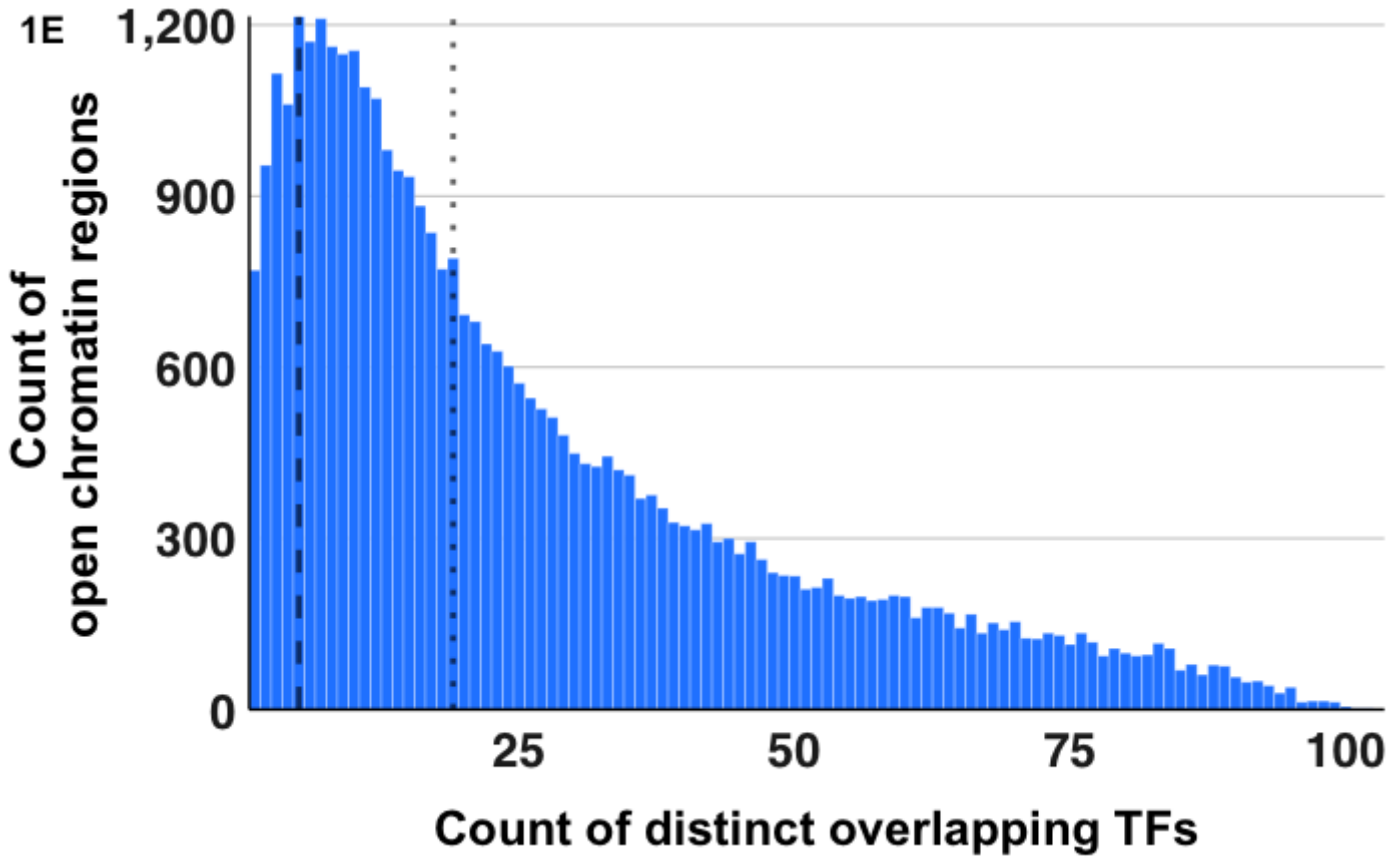


From the total set of TF peaks, 64% were overlapping with open chromatin regions, with a median value of 74% (IQR_{25-75} 64%-87%) of the total peaks for a given TF (Supplementary figure 4).



For open chromatin regions, 98% overlap with at least one TF (mode=5, median=19 distinct TFs) for each region (Fig 1E), and 95% overlap with TFs from at least two distinct families (median=10) for each region. The large number of distinct TFs clustered within TF binding loci - often from diverse TF families – that co-localize with open chromatin regions suggests a large number of possible TF combinations. The similarity between TF binding profiles and open chromatin confirm the reliability of the ChIP-seq data, and the relevance of the identified binding sites within the in-vivo chromatin context.

Distribution of open chromatin regions non-overlapping with gene bodies, as a function of gene distance



Hide

#References:
#1. Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Research, 22(9), 1813–1831.
#2. Kharchenko, P. V., Tolstorukov, M. Y., & Park, P. J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nature Biotechnology, 26(12), 1351–1359.
#3. Li, Q., Brown, J. B., Huang, H., & Bickel, P. J. (2011). Measuring reproducibility of high-throughput experiments. The Annals of Applied Statistics, 5(3), 1752–1779.
#4. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods, 10(12), 1213–1218.
#5. Rodgers-Melnick, E., Vera, D. L., Bass, H. W., & Buckler, E. S. (2016). Open chromatin reveals the functional maize genome. Proceedings of the National Academy of Sciences of the United States of America, 113(22), E3177–84.
#6. Salvi, S., Sponza, G., Morgante, M., Tomes, D., Niu, X., Fengler, K. A., et al. (2007). Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proceedings of the National Academy of Sciences, 104(27), 11376–11381.
#7. Lemmon, Z. H., Bukowski, R., Sun, Q., & Doebley, J. F. (2014). The role of cis regulatory evolution in maize domestication. PLoS Genetics, 10(11), e1004745.