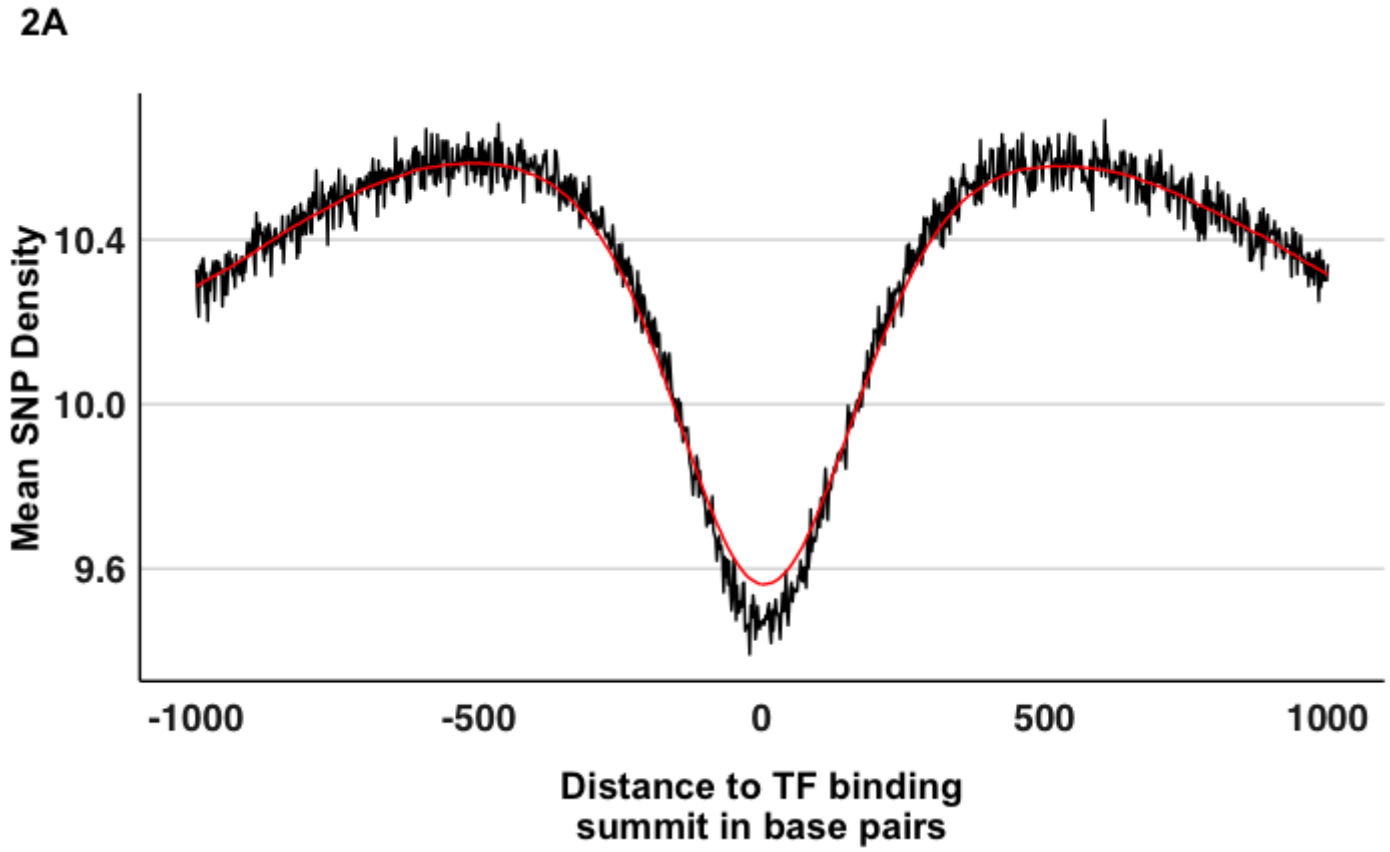
Overlap with functional variation

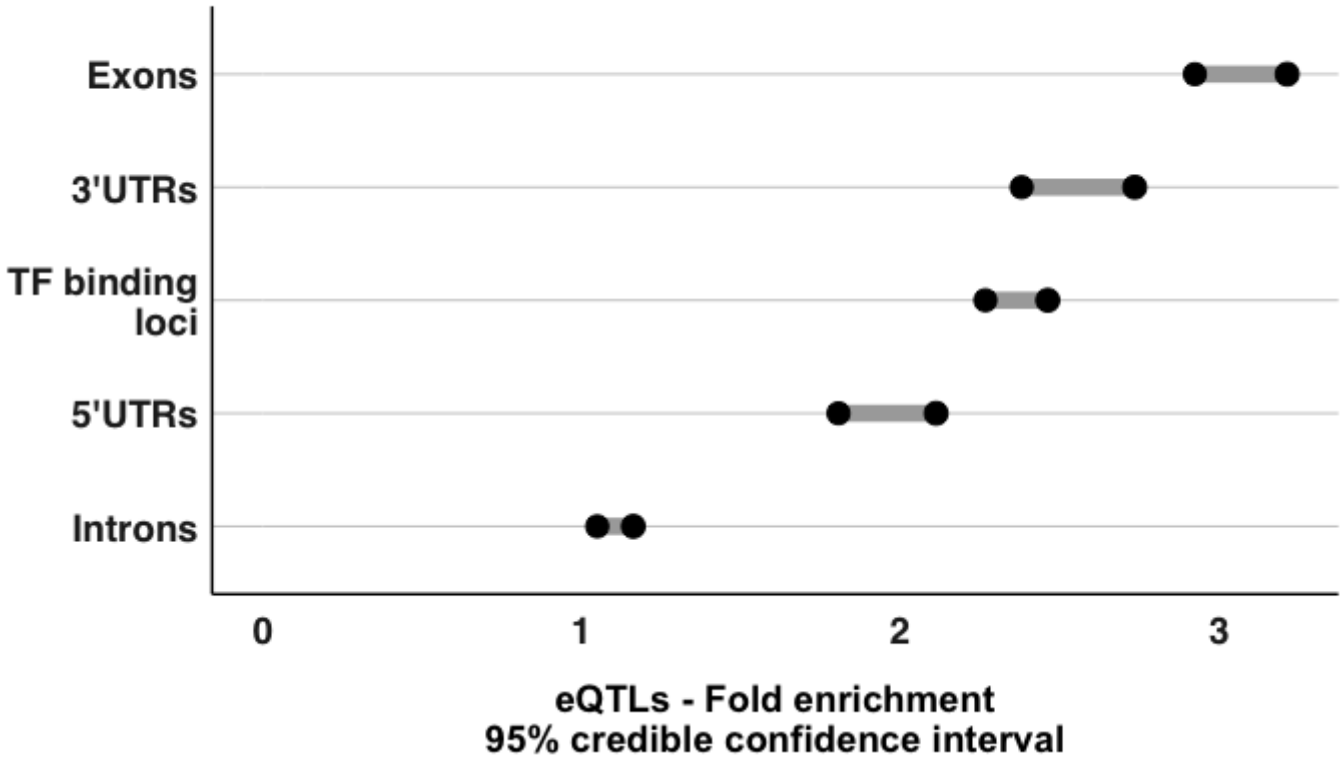
Code

TF binding sites are key determinants of transcriptional regulation, and should harbor low sequence diversity, as well as to preferentially overlap with functional variation. We examined the conservation of the TF binding regions by assessing the overall nucleotide diversity represented in the maize HapMap [1] while controlling for the overall SNP density in function to the distance of TF's peak summit (Fig 2A). The result confirmed that sequence variation is indeed reduced, with variation decreasing in proximity of TF's peak summit.



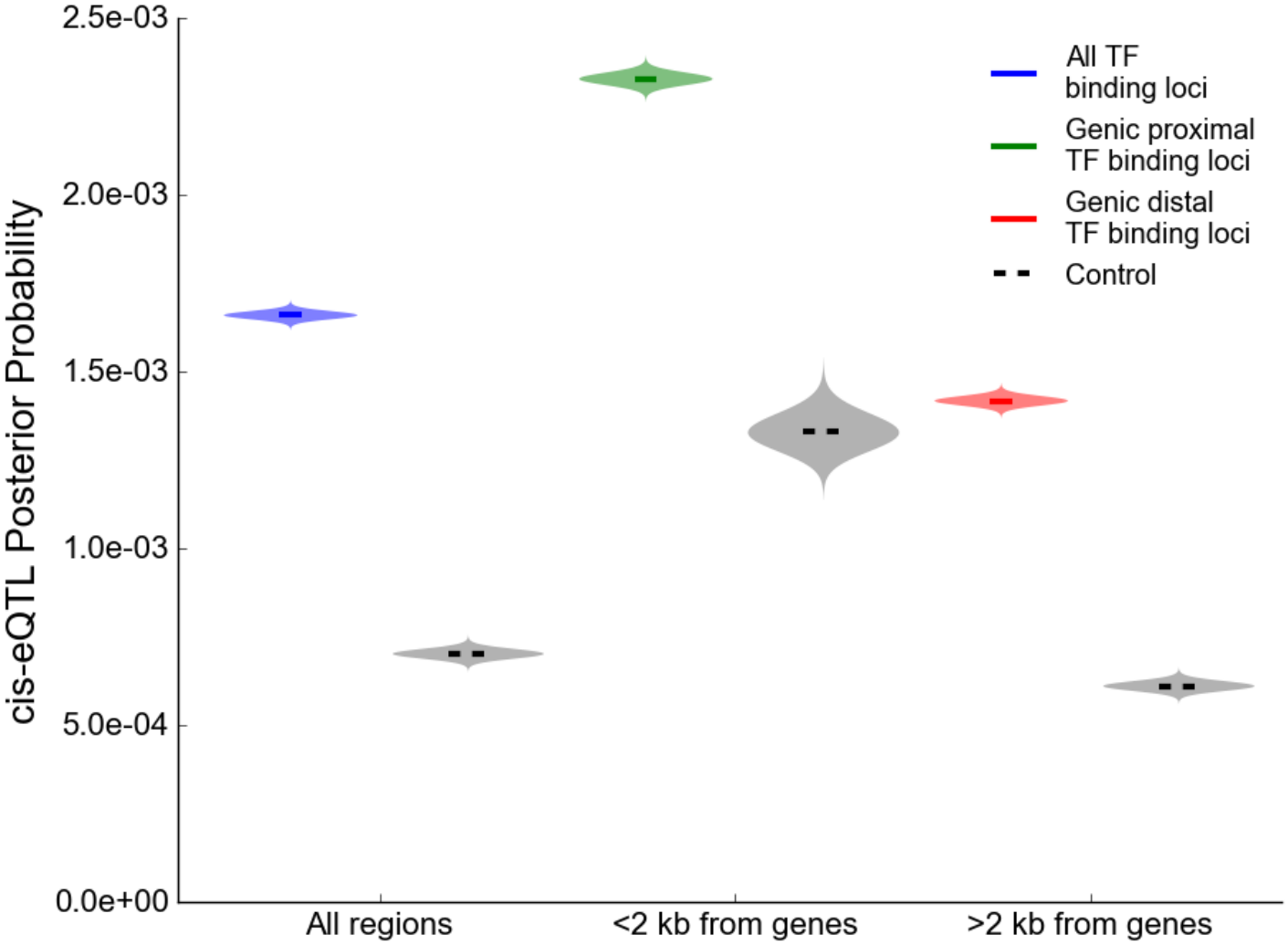
To test the global relationship between the sequence variation associated with variation in gene expression and TF binding, we calculated the statistically significance of the overlap of cis-expression QTLs [2] (hereafter eQTLs) within 2 kb of TF binding loci from the non-eQTLs SNPs (Fisher test, P-value < 0.05). Next, we quantified the enrichment of eQTLs around TF binding loci and compared with adjacent regions at least 5 kb away from the nearest TF binding loci.

2B



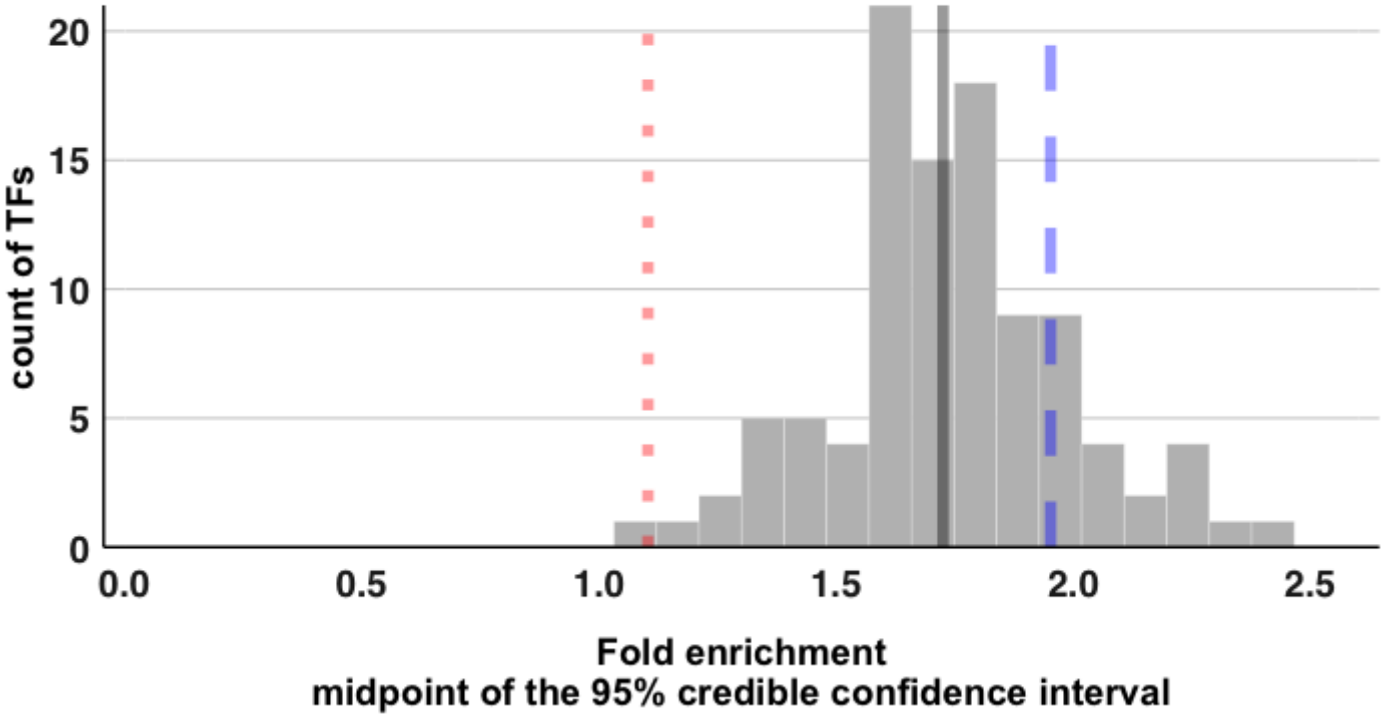
We found twofold enrichment of eQTLs around TF binding loci (95% credible interval 2.26-2.46), similar to the enrichment around 5'UTRs (95% credible interval 1.80-2.11) and 3'UTRs (95% credible interval 2.37-2.73) (Fig 2B). Furthermore, the enrichment over control regions was robust to the distance between TF binding loci and genes (Supplementary figure 1).

Supplementary figure 1



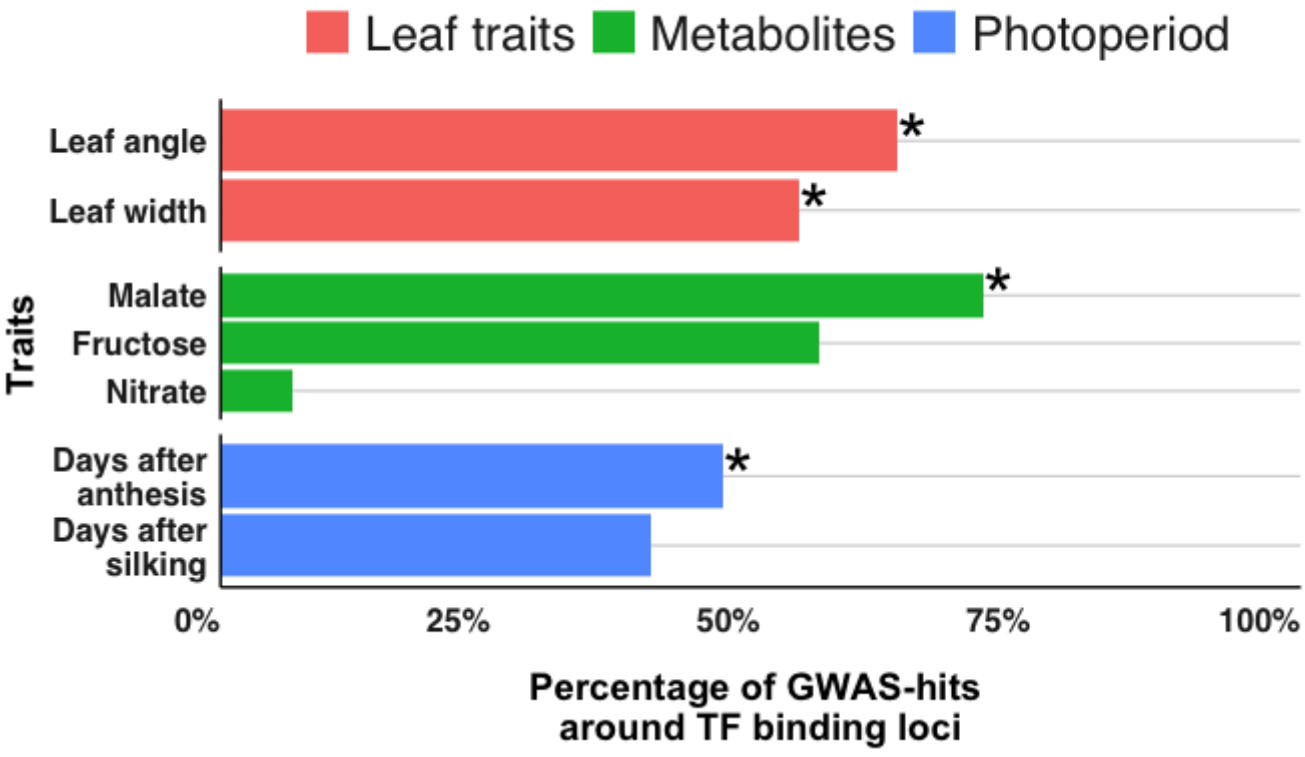
At the level of individual TFs, we confirmed the trend observed for the whole set, with TFs enriched more than intronic regions and similar to the observed enrichment in 5'UTRs (Supplementary figure 2, supplementary table X).

Supplementary figure 2

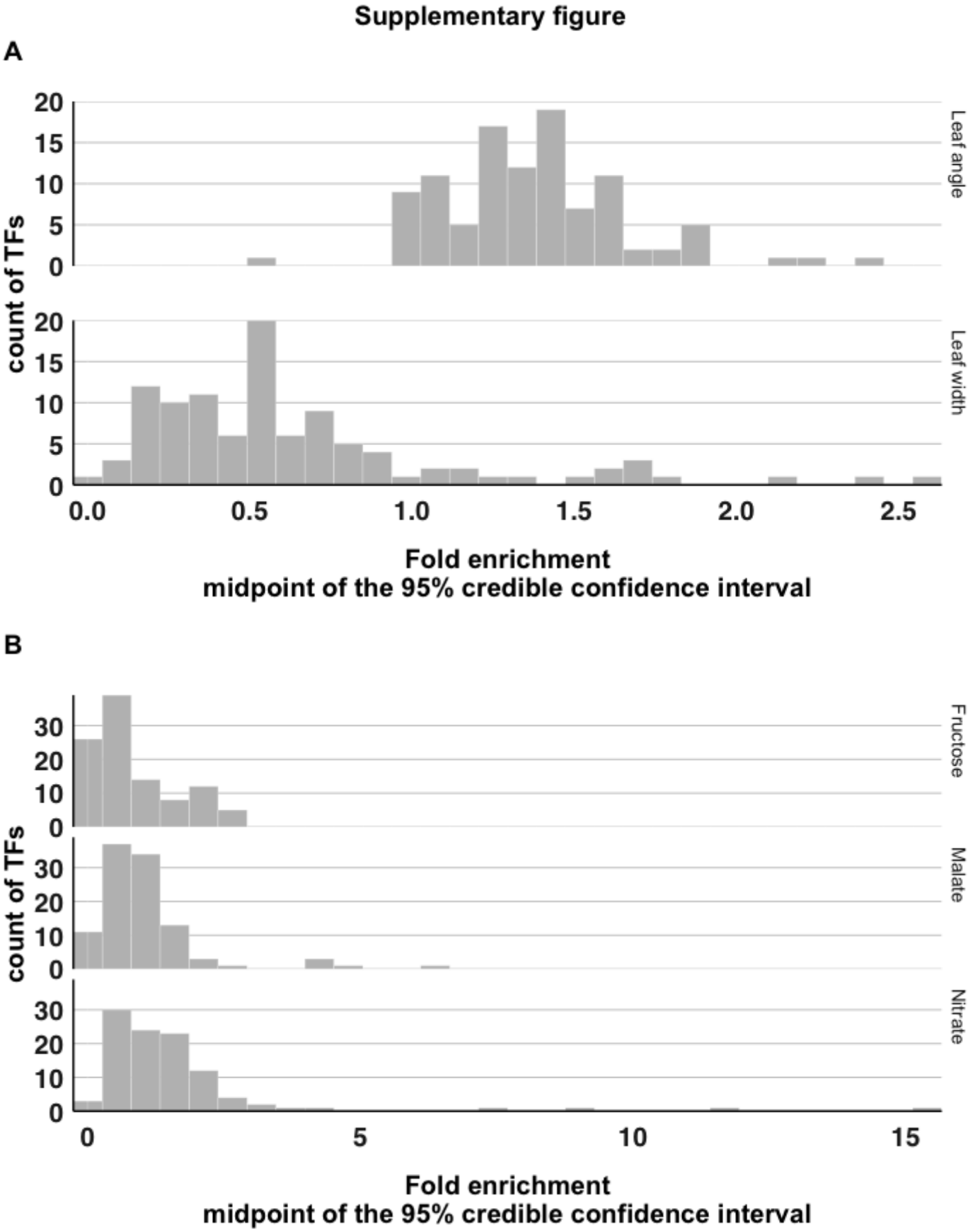


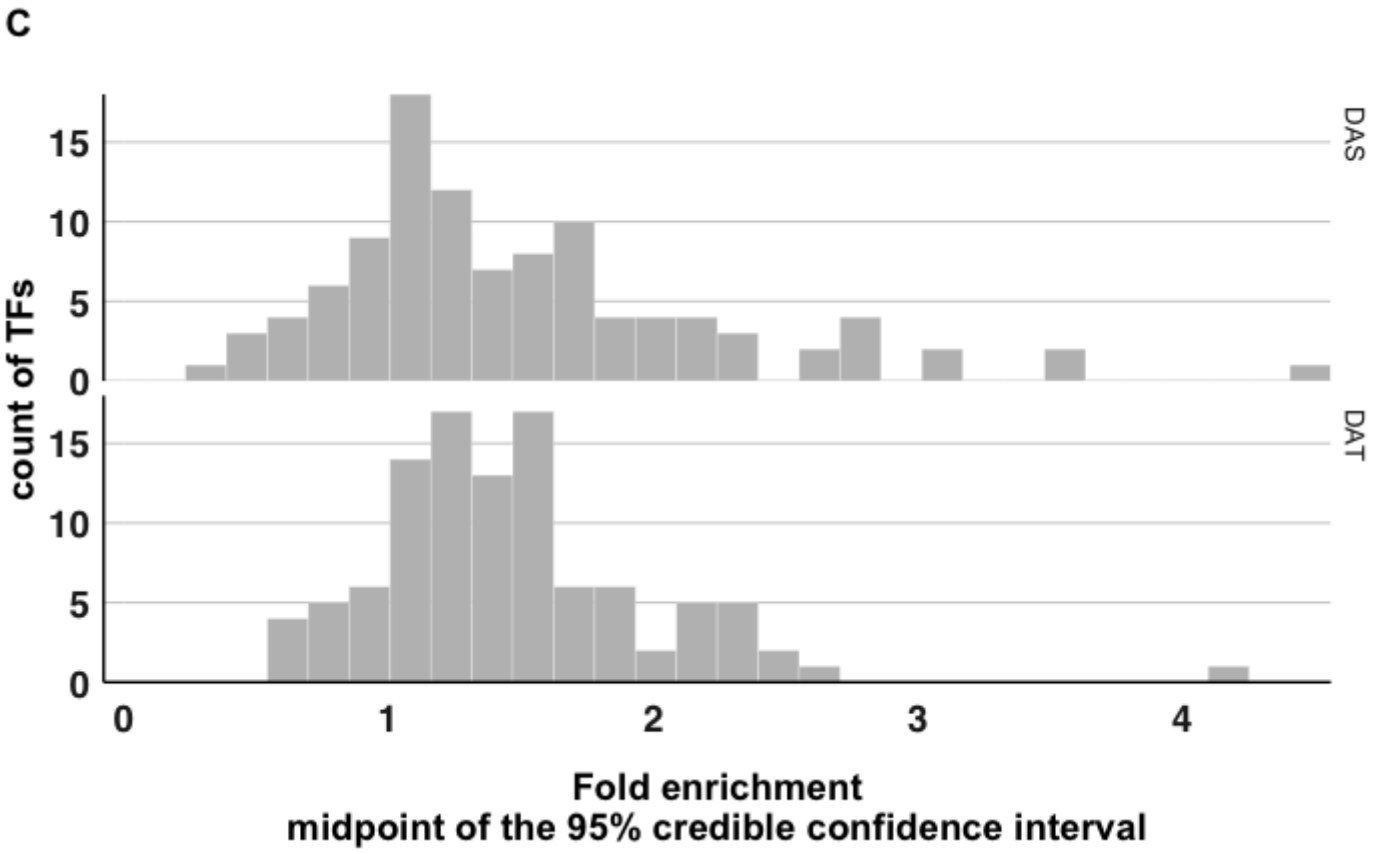
To asses the relationship between TF binding and complex trait variation, we calculated enrichment in GWAS hits for seven traits (metabolites [3], leaf architecture [4], and photoperiodicity [5]) measured in the US NAM population. We found statistical significance for the enrichment in GWAS-hits for four of the traits, including leaf angle and leaf width (Fig 2C).

2C

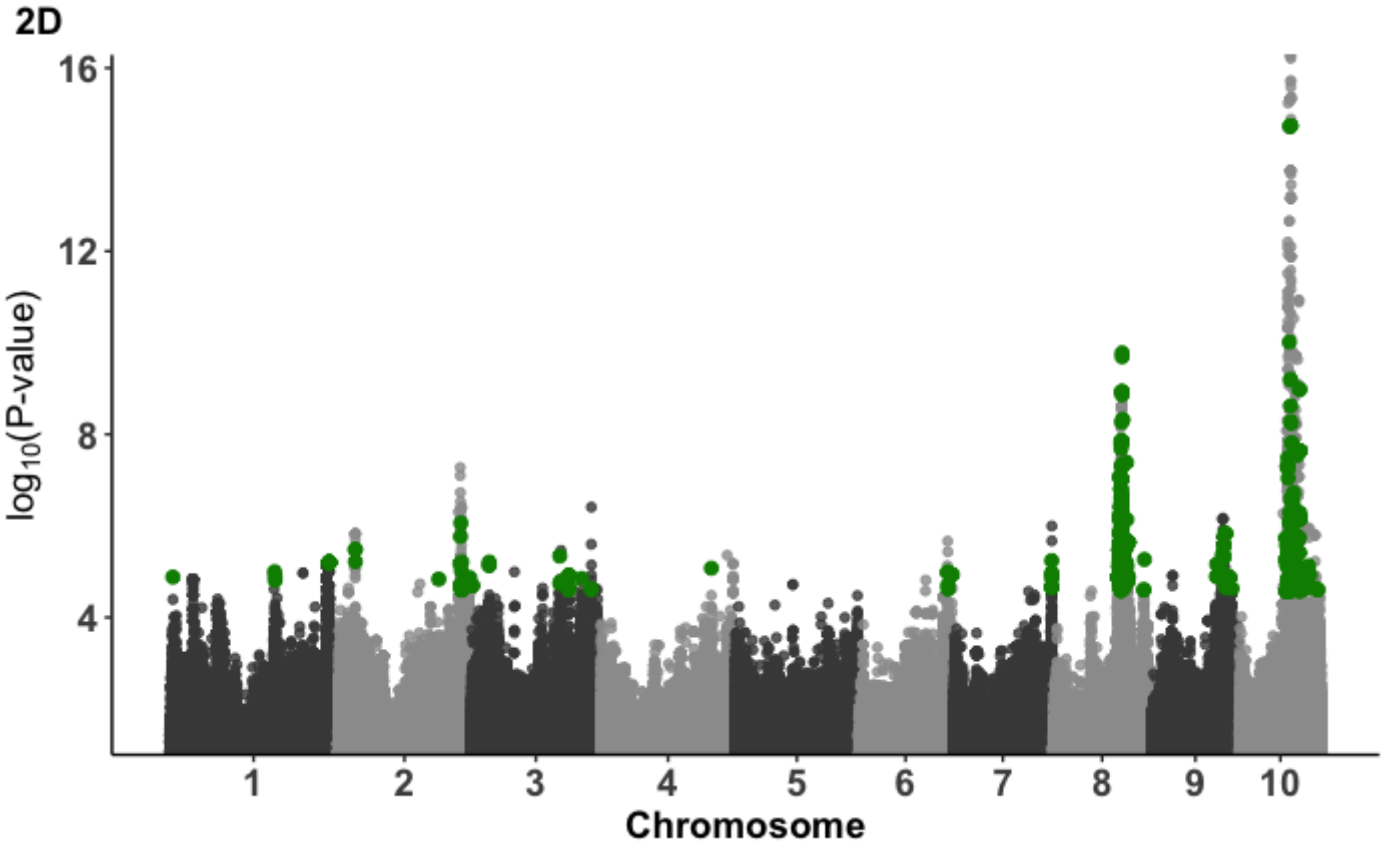


The analyses of individual TFs in comparison to nearby regions, revealed that a large number of TFs overlap with GWAS hits for those traits that are expected to be influenced by a large number of regulators (complex traits) (e.g., DAS and DAT). More simple traits, such as metabolites, show few TFs enriched for GWAS hits, as those are expected to be under the control of few regulators (e.g., Malate, Nitrate) (Supplementary figure 3).





Examination of the TFs enriched in GWAS hits for days after anthesis (DAT), and days after silking (DAS) revealed that 51% of the TFs enriched in DAT, and 35% in DAS had binding sites in near proximity to VGT1, a major quantitative trait loci for flowering time, located at a distance of ~72 kb from the TSS of the RELATED TO AP2.7 (RAP2.7) gene [6] (Fig 1A). From the group of TFs binding around VGT1 and/or enriched for variants associated to photoperiod, a group of six(PRR5, ELF3, COL3, COL8, COL18 and DOF3/PBF1) have been previously associated with flowering time variations [7-8] (Fig 2D).



Our observation indicates that TF binding regions are conserved, and their frequent overlap with cis-eQTLs and GWAS hits suggests that TF binding regions are functional portions of the maize genome. Taken together, our finding proves the potencial of our data to connect sequence variation in cis regulatory regions to trans-regulators, and to detect functional regulatory variants implicated in complex phenotypes.

1. Bukowski, R., Guo, X., Lu, Y., Zou, C., He, B., Rong, Z., et al. (2018). Construction of the third-generation Zea mays haplotype map. GigaScience, 7(4), 1–12. (http://doi.org/10.1093/gigascience/gix134)
2. Kremling, K. A. G., Chen, S.-Y., Su, M.-H., Lepak, N. K., Romay, M. C., Swarts, K. L., et al. (2018). Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. Nature, 555(7697), 520–523. (http://doi.org/10.1038/nature25966)
3. Zhang, N., Gibon, Y., Wallace, J. G., Lepak, N., Li, P., Dedow, L., et al. (2015). Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. Plant Physiology, 168(2), 575–583. (http://doi.org/10.1104/pp.15.00025)
4. Tian, F., Bradbury, P. J., Brown, P. J., Hung, H., Sun, Q., Flint-Garcia, S., et al. (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. Nature Genetics, 43(2), 159–162. (http://doi.org/10.1038/ng.746)
5. Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., et al. (2009). The Genetic Architecture of Maize Flowering Time. Science, 325(5941), 714–718. (http://doi.org/10.1126/science.1174276)
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. (http://doi.org/10.1073/pnas.0704145104)
7. Li, Y.-X., Li, C., Bradbury, P. J., Liu, X., Lu, F., Romay, C. M., et al. (2016). Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. The Plant Journal : for Cell and Molecular Biology, 86(5), 391–402. (http://doi.org/10.1111/tpj.13174)
8. Alter, P., Bircheneder, S., Zhou, L.-Z., Schlüter, U., Gahrtz, M., Sonnewald, U., & Dresselhaus, T. (2016). Flowering Time-Regulated Genes in Maize Include the Transcription Factor ZmMADS1. Plant Physiology, 172(1), 389–404. (http://doi.org/10.1104/pp.16.00285)