

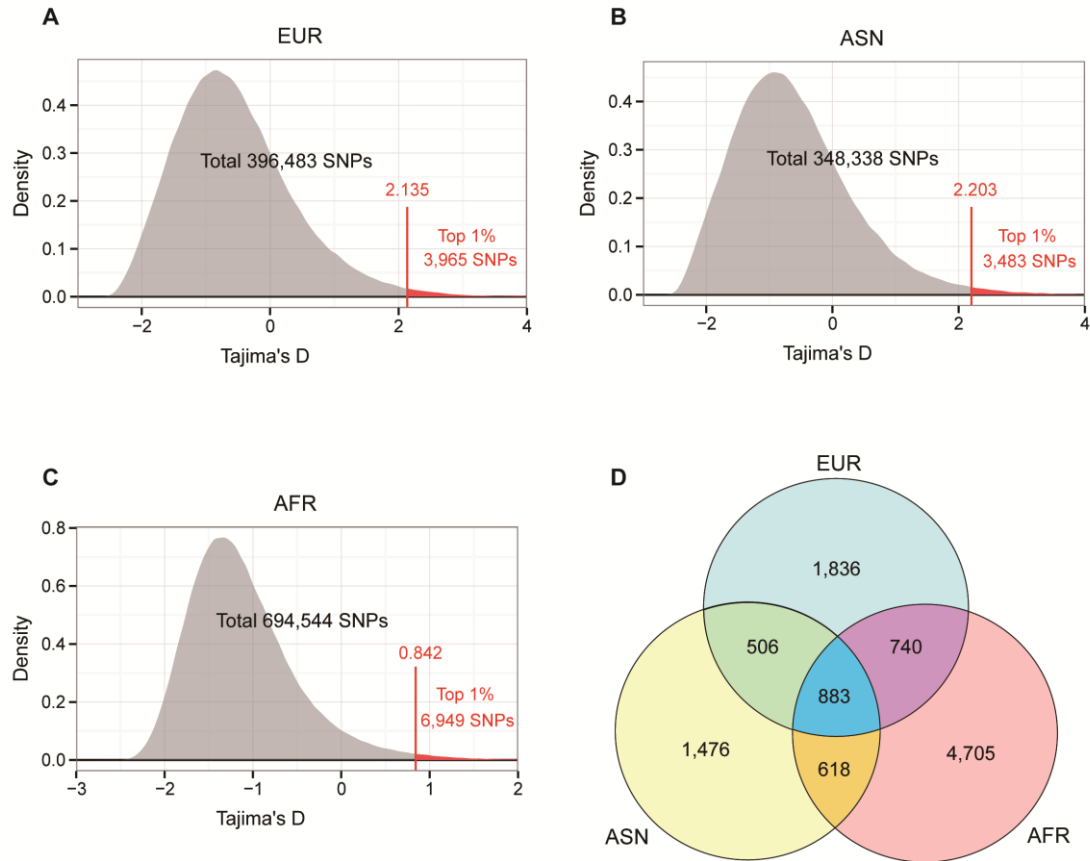


Figure S1. Distribution of Tajima's D for DNase footprints in three subpopulations. (A-C) Density plots of Tajima's D for the (A) EUR, (B) ASN, and (C) AFR panel of the 1000 genomes project, respectively. The top 1% of the distribution of the SNPs was considered to have a footprint of balancing selection and highlighted with red colour. (D) The overlaps of the top 1% SNPs identified for each population.

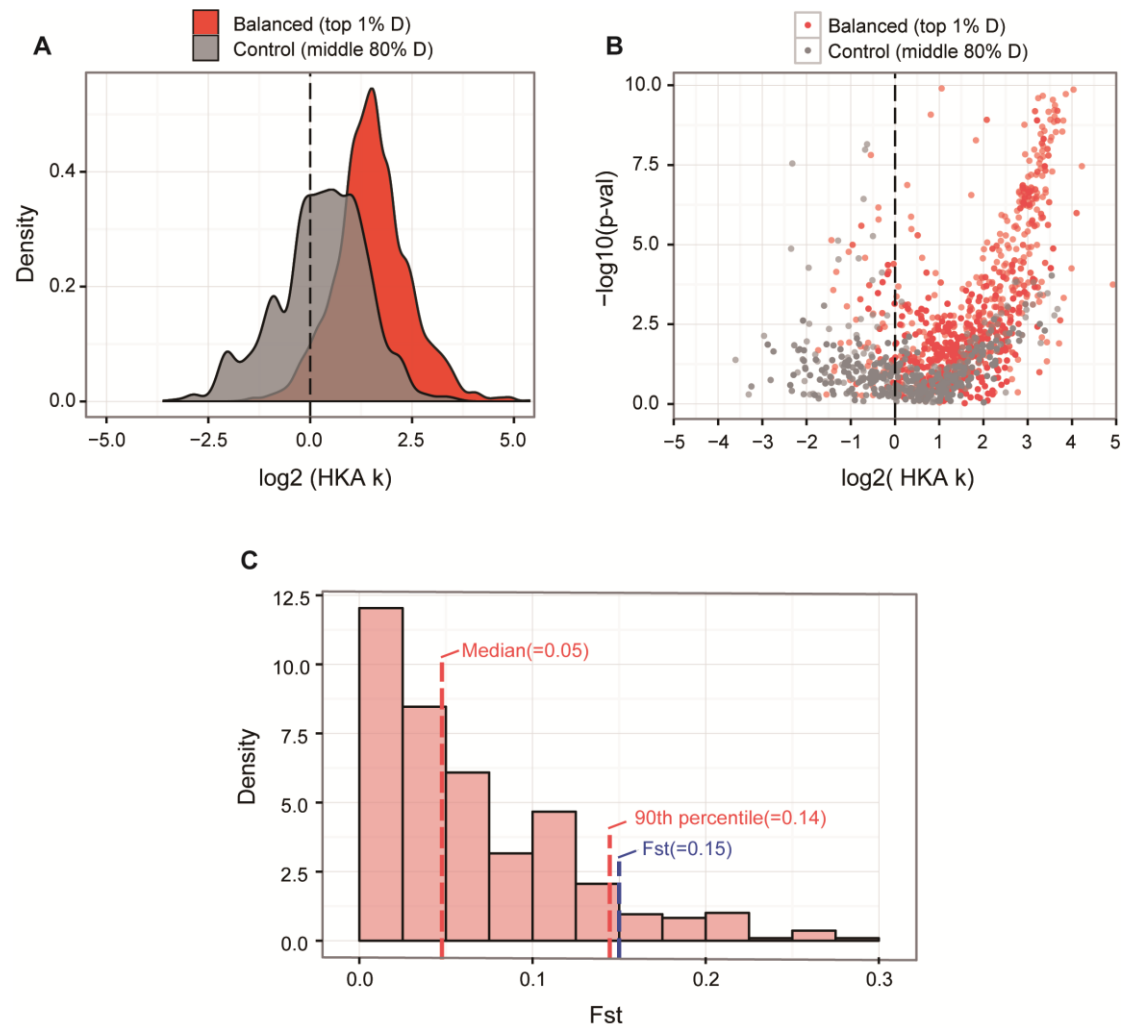


Figure S2. Comparison of Tajima's D with HKA and Fst. (A) The distribution of HKA k in a $\log_2$ scale for the top 1% D SNPs (red) and middle 80% D SNPs as control (grey) in the EUR panel. The selected SNPs showed significantly higher HKA k (95.42 % having $k > 1$) than the control set. (B) Volcano plot showing the P value of the HKA test plotted against the HKA k score for the balanced versus control SNPs. Data points highlighted in red (balanced SNPs with top 1% D) showed a positively skewed distribution with lower P values. (C) Frequency histogram of the F_{st} score. Most balanced SNPs showed low F_{st} values (91.2% having $F_{st} < 0.15$). The value of 0.15 is a typical estimate for neutral loci among continental populations. Outlier scores with $F_{st} > 0.3$ were not represented in the histogram.

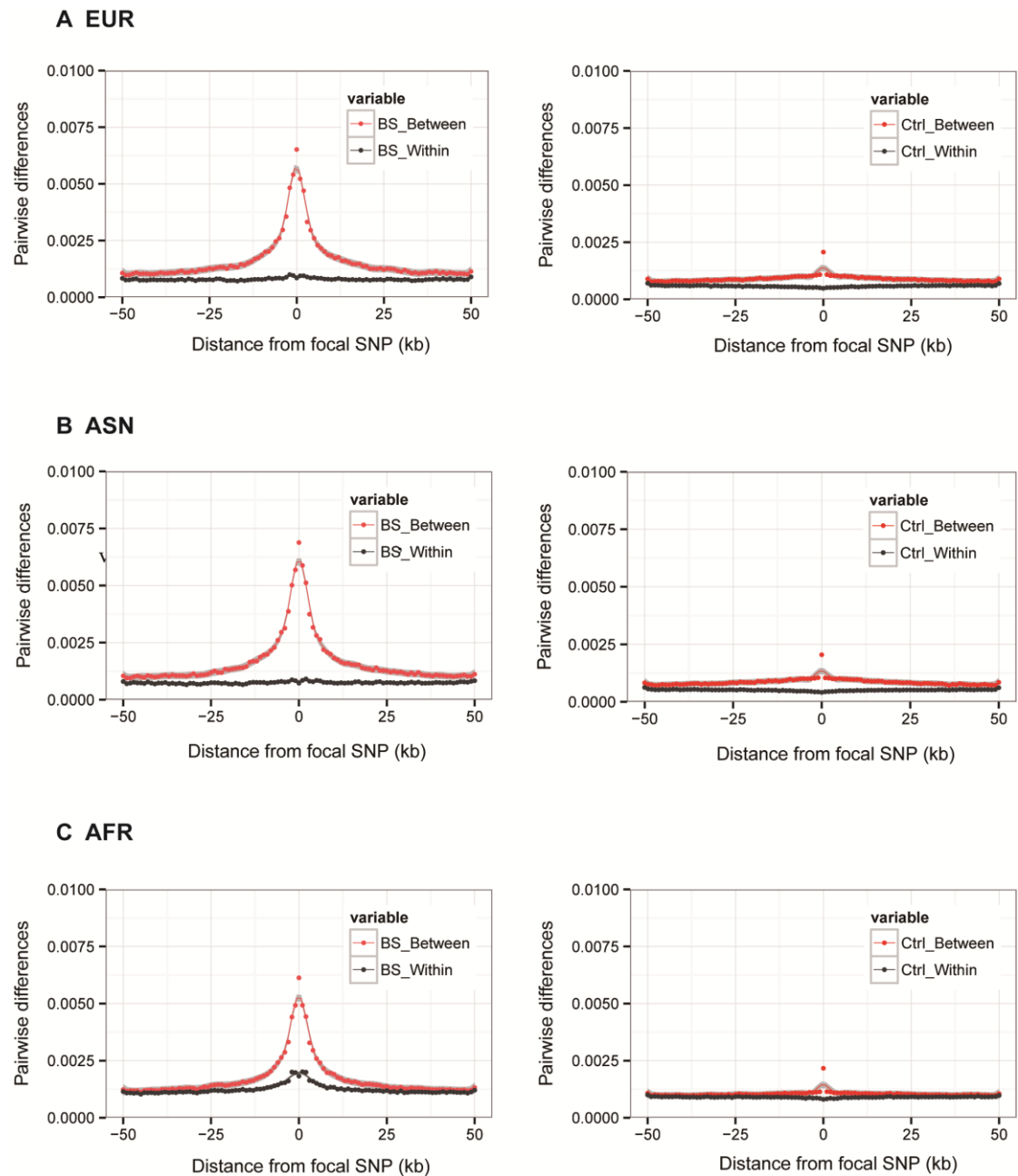


Figure S3. High divergence between haplogroups compared to within haplogroup. The mean pairwise differences of the sequences between haplogroups (red) and within haplogroup (grey) were plotted for the balanced SNPs (left) and control SNPs (right) in the (A) EUR, (B) ASN, and (C) AFR panel, respectively. The pairwise differences were calculated in 1.5-kb sliding windows and averaged for all focal SNPs.

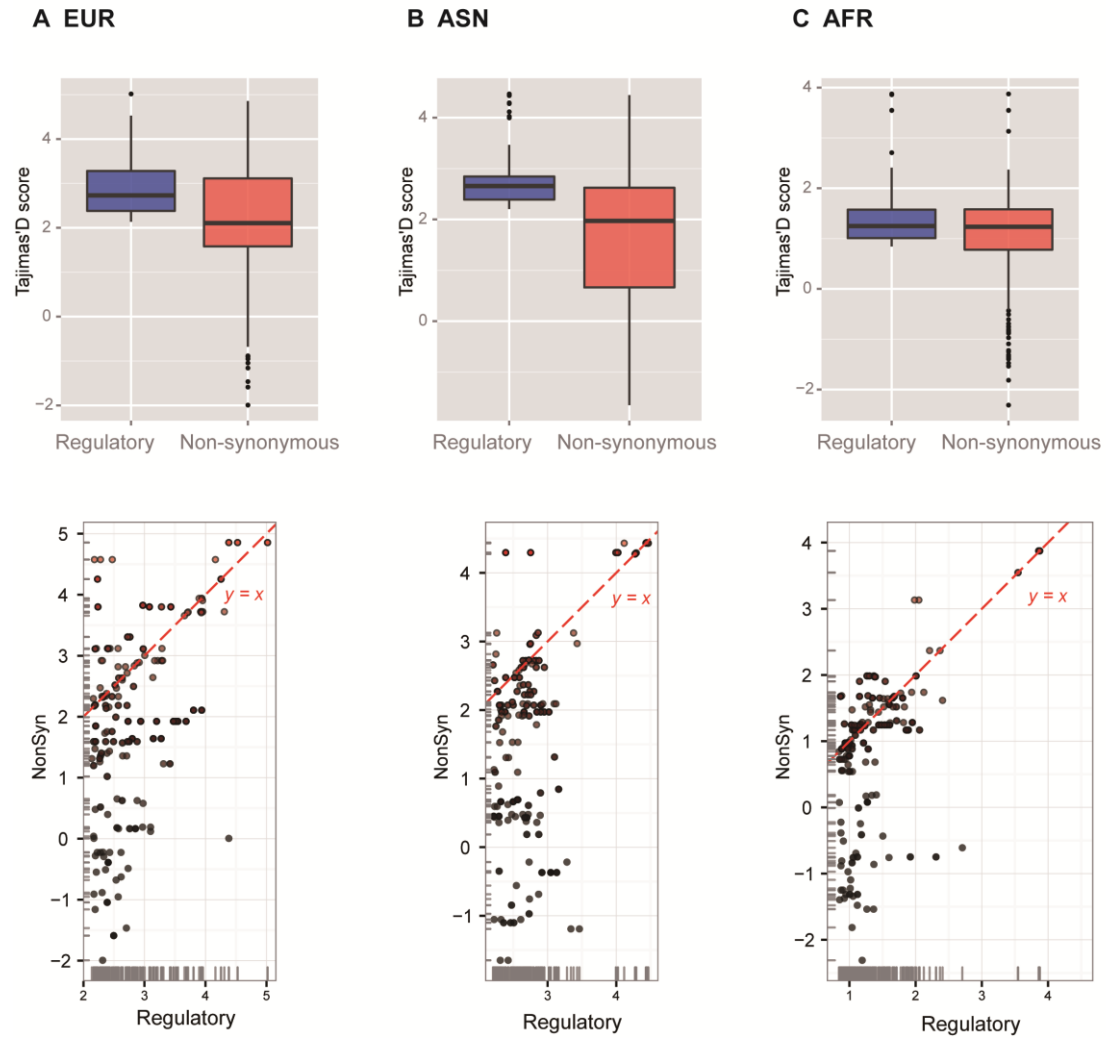


Figure S4. Comparison of Tajima's D between the selected regulatory and linked non-synonymous SNPs. Box plots (upper) and scatter plots (lower) of the Tajima's D scores for the balanced SNPs in *cis*-regulatory regions and their linked non-synonymous SNPs in linkage disequilibrium ($r^2 = 0.9$) in the (A) EUR, (B) ASN, and (C) AFR panel, respectively. Among the linked non-synonymous SNPs of each regulatory SNP, we selected that has the highest Tajima's D in the linkage disequilibrium block. One unit on the x-axis is the same length as one unit on the y-axis.

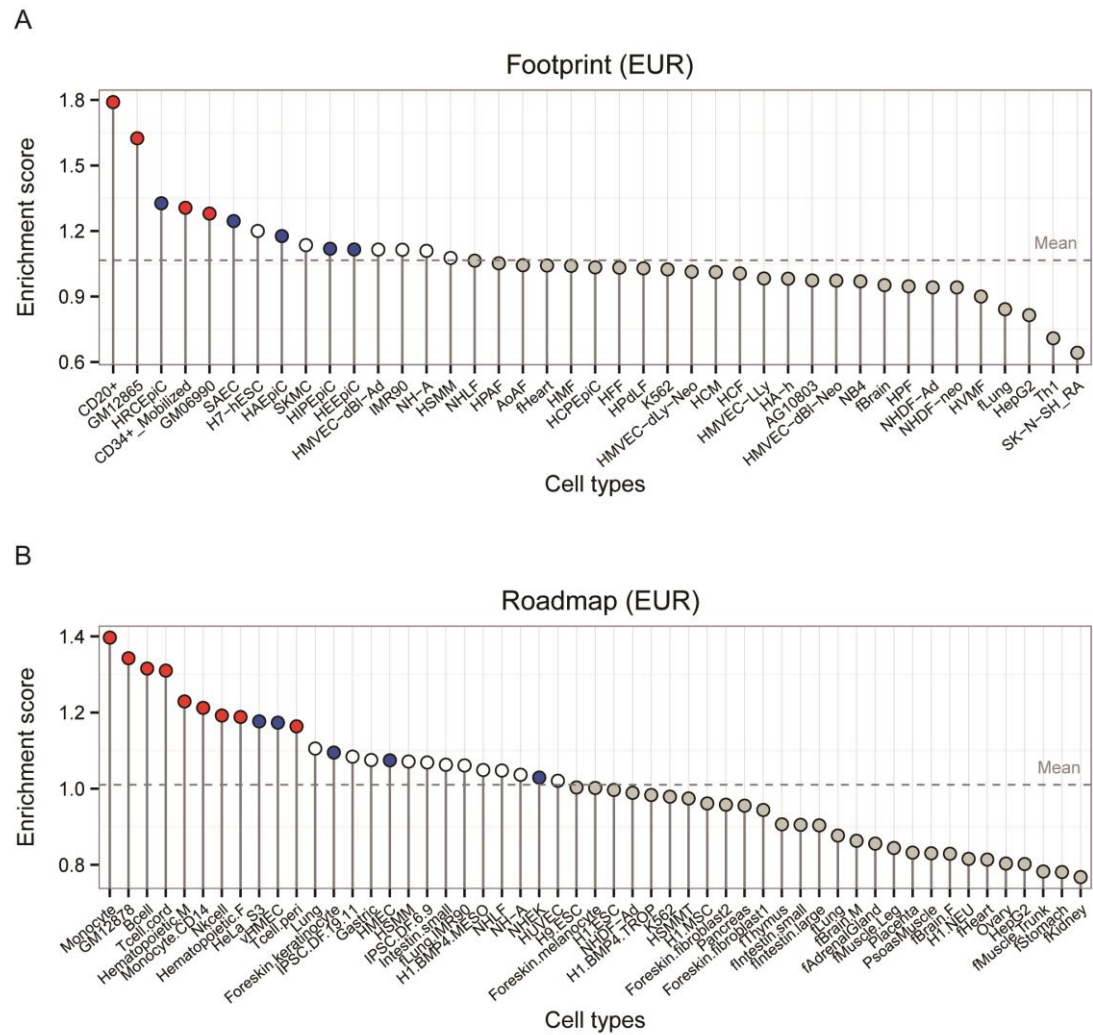
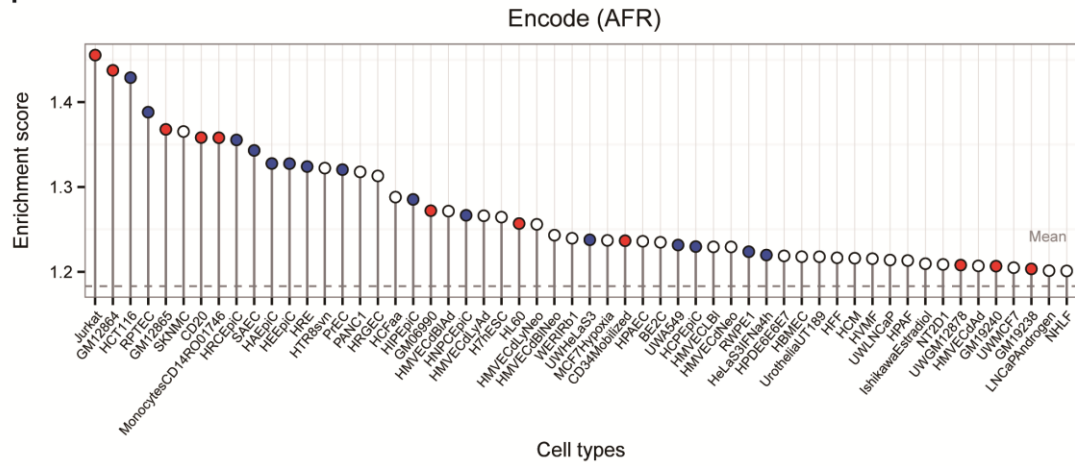
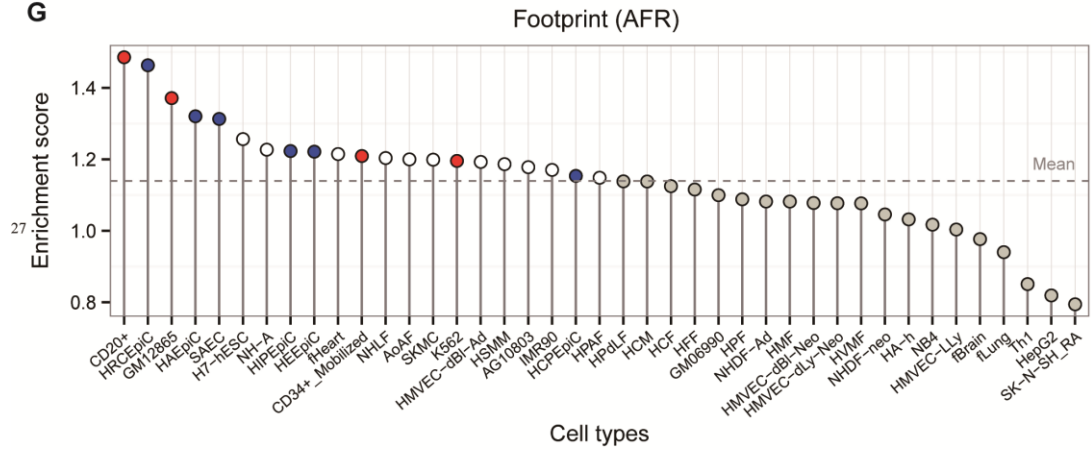


Figure S5. Enrichment of the balanced SNPs in DNase footprints. Cleveland's dot plots showing enrichment of the balanced SNPs in the *cis*-regulatory regions of diverse cell types including those for which DNase footprints are available and those used in the ENCODE project and Epigenome Roadmap project in the (A-B) EUR, (C-E) ASN, and (F-H) AFR panel. The EUR ENCODE data is provided in Fig. 1B.

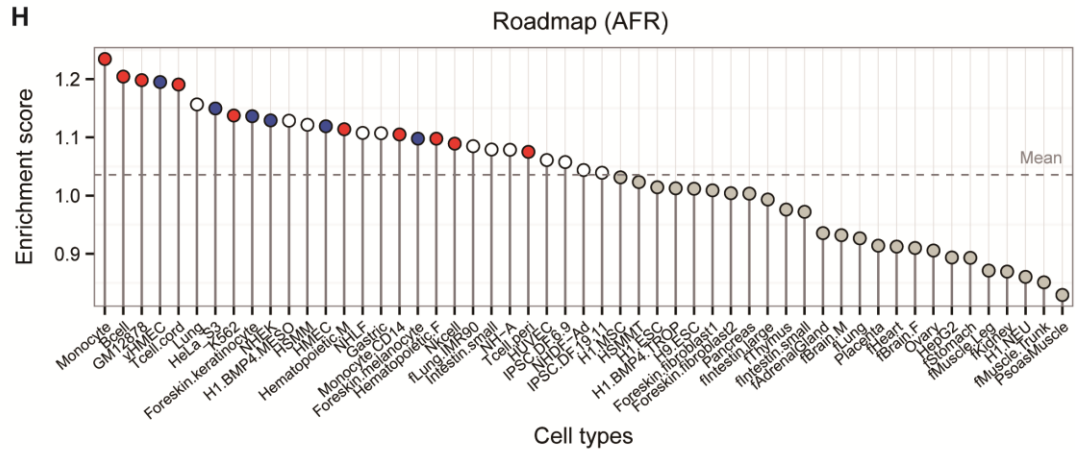
F



G



H



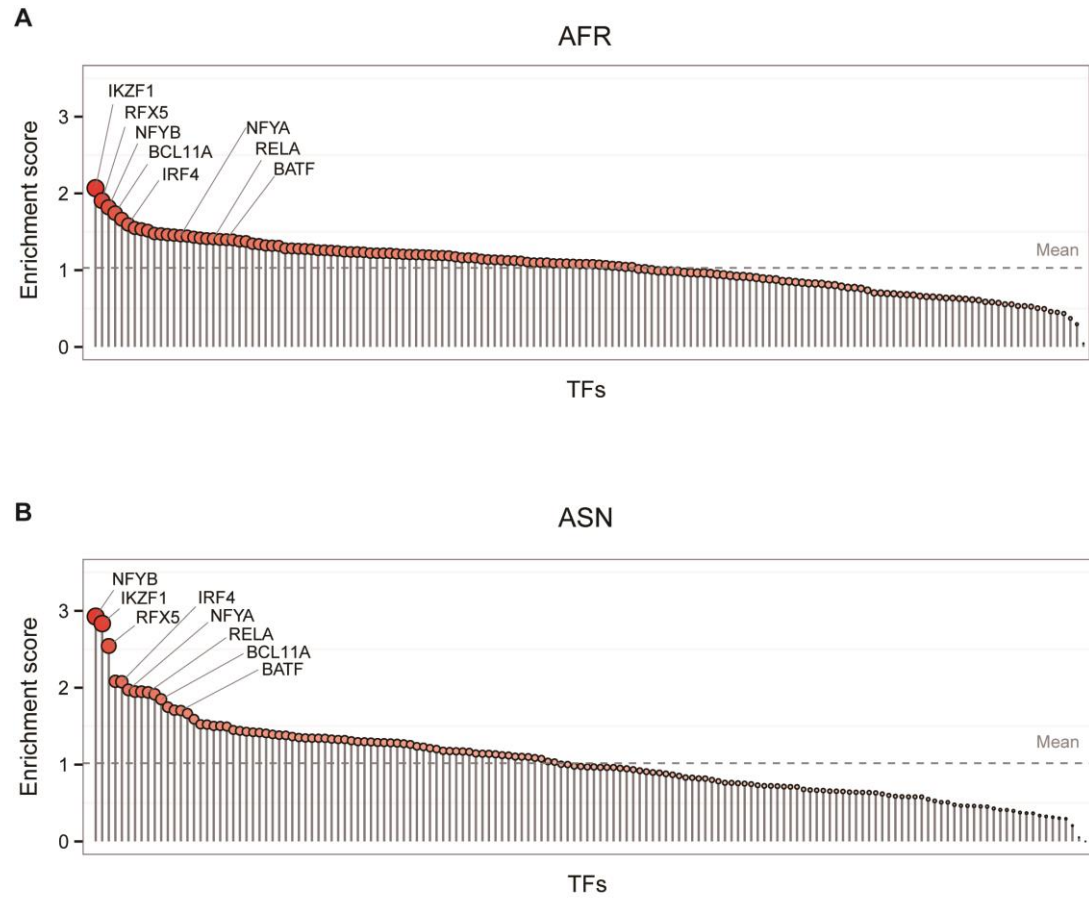
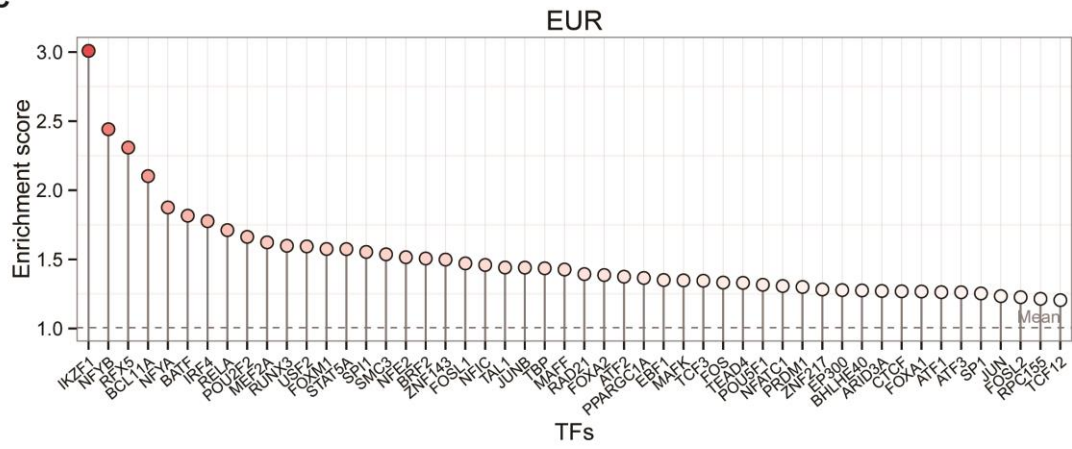
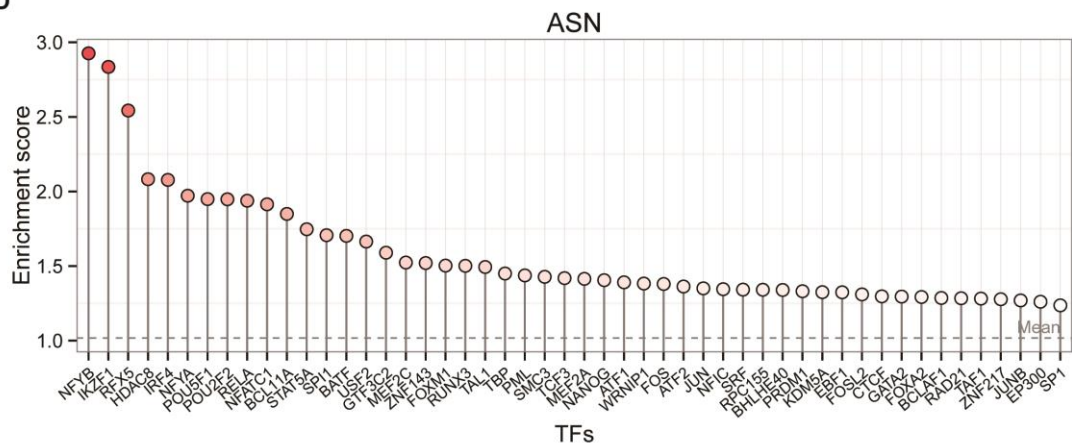


Figure S6. Enrichment of the balanced SNPs in TF binding regions. Cleveland's dot plots showing enrichment of the balanced SNPs in the binding regions of 161 TFs in the (A) ASN and (B) AFR panel. The same plot for the EUR panel is provided in Fig. 1C. The point size is proportional to the enrichment score. Only the top 50 TFs were displayed for the (C) EUR, (D) ASN, and (E) AFR population.

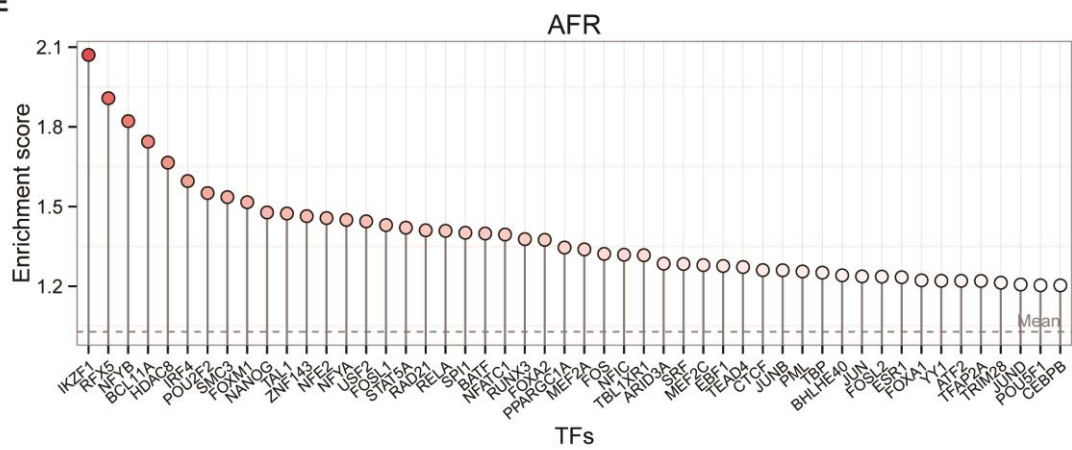
C



D



E



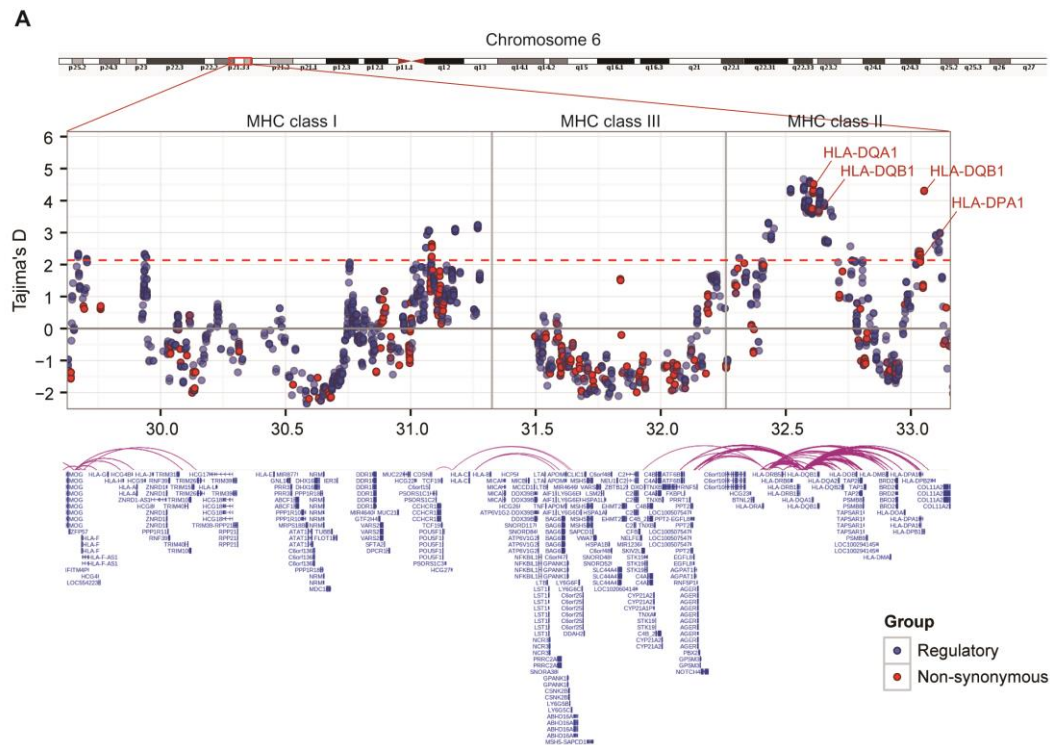


Figure S7. Chromosomal distribution of balancing selection signatures for MHC class II and SPRR/LCE genes. (A) Tajima's D signals at the MHC locus on chromosome 6 between the genetic markers MOG and COL11A2 (from 6p22.1 to 6p21.3 on hg19) for regulatory (blue) and non-synonymous SNPs (red). The horizontal dotted red line indicates the cutoff used for whole-genome detection of balanced SNPs (top 1% D). The promoter-enhancer chromatin interactions (Methods) were visualized by using the WashU Epigenome Browser (<http://epigenomegateway.wustl.edu/browser/>) in the bottom panel. (B) Tajima's D signals in the chromosome 1q21 region containing genes involved in epidermal differentiation for regulatory (blue) and non-synonymous SNPs (red). The chromatin interactions were displayed by using the WashU Epigenome Browser. (C) A zoom-in view of (B) into chr1:152350000-153250000.

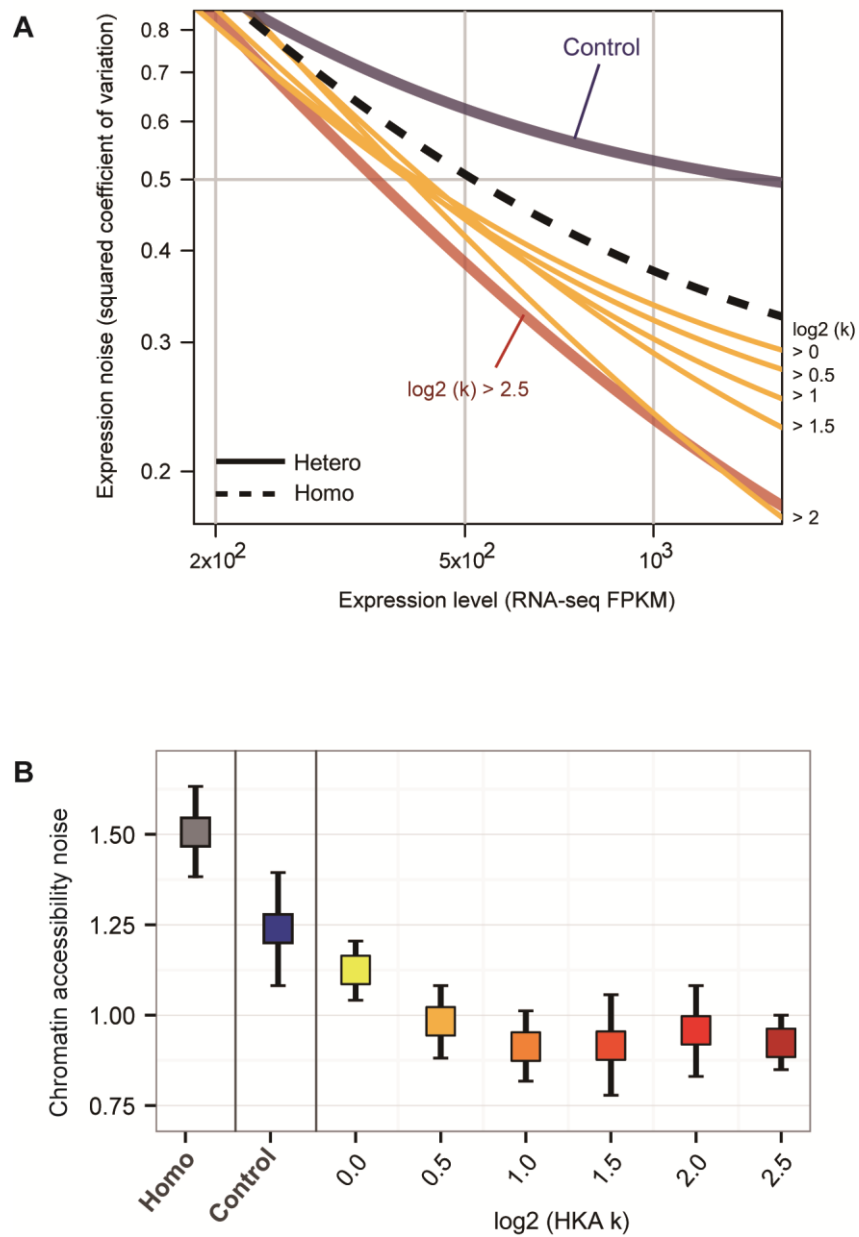


Figure S8. Noise in gene expression and chromatin accessibility as a function of HKA k.

(A) Progressive reduction of the expression noise (CV^2) in proportion to the HKA k of the associated regulatory SNP. The curves show a fit to RNA sequencing data for 62 single cells with the solid and dashed lines representing heterozygous and homozygous loci in GM12878, respectively. The heterozygous curves are divided by the HKA k levels. (B) Cell-to-cell variability in chromatin accessibility for accessible regions categorized according to the HKA k levels in the same manner as in (A). Error bars represent one standard deviation of the variability obtained by bootstrapping cells (see Methods).

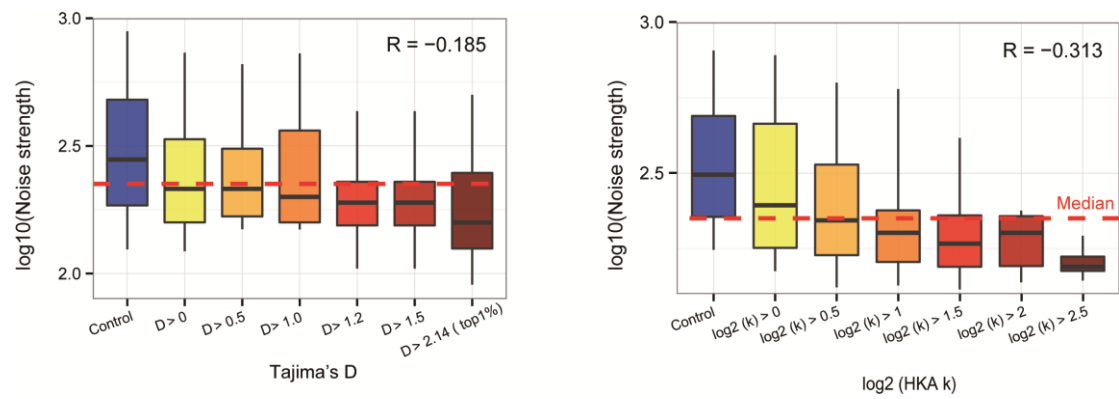


Figure S9. Negative correlation between the selection strength and gene expression noise.

The noise strength of target gene expression as measured based on GM12878 single-cell RNA sequencing according to the selection strength at the corresponding regulatory SNPs as estimated by Tajima's D (left) or HKA k (right). Only heterozygous loci in GM12878 were used for the analyses.

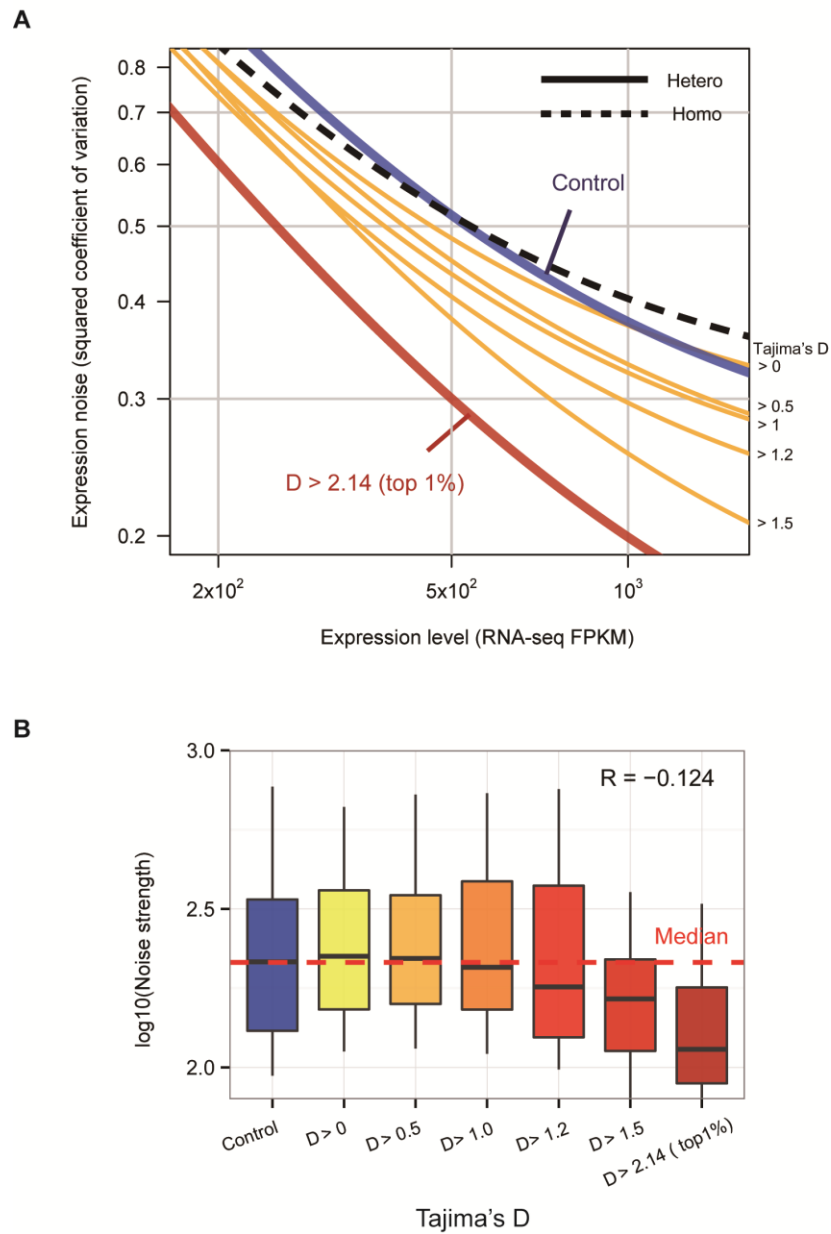
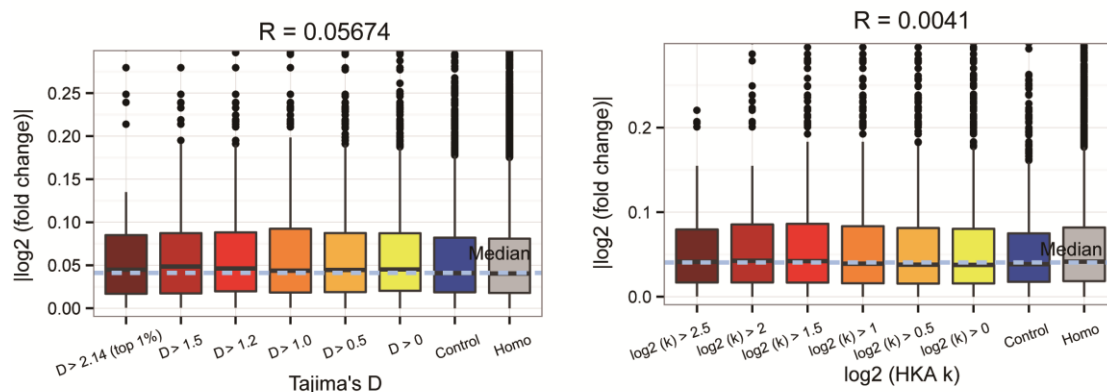
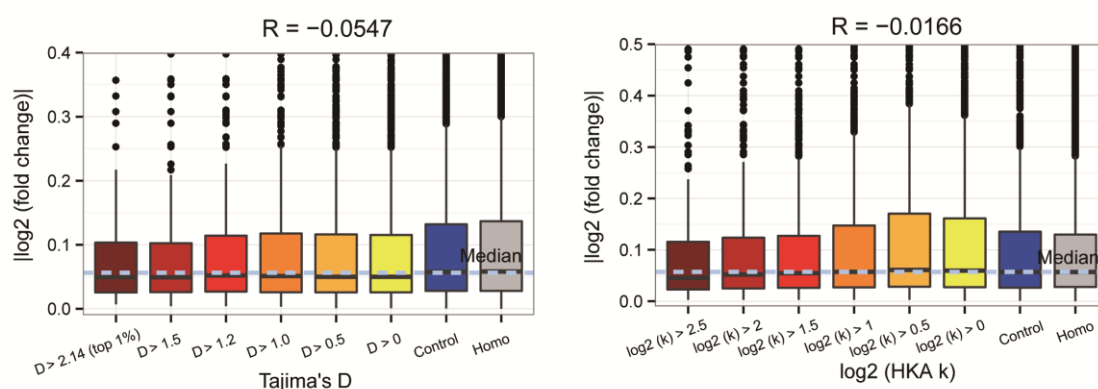


Figure S10. Lower expression noise of genes mapped to footprint SNPs by a different chromatin interaction dataset. (A) Progressive reduction of the expression noise (CV^2) in proportion to the HKA k of the associated regulatory SNP. The curves show a fit to GM12878 single-cell RNA sequencing with the solid and dashed lines representing heterozygous and homozygous loci in the GM12878 cells, respectively. The heterozygous curves are divided by Tajima's D . (B) The noise strength of target gene expression as measured based on GM12878 single-cell RNA sequencing according to the selection strength at the corresponding regulatory SNPs that are heterozygous in GM12878.

A. Doxorubicin



B. Radiation



C. TNF-alpha

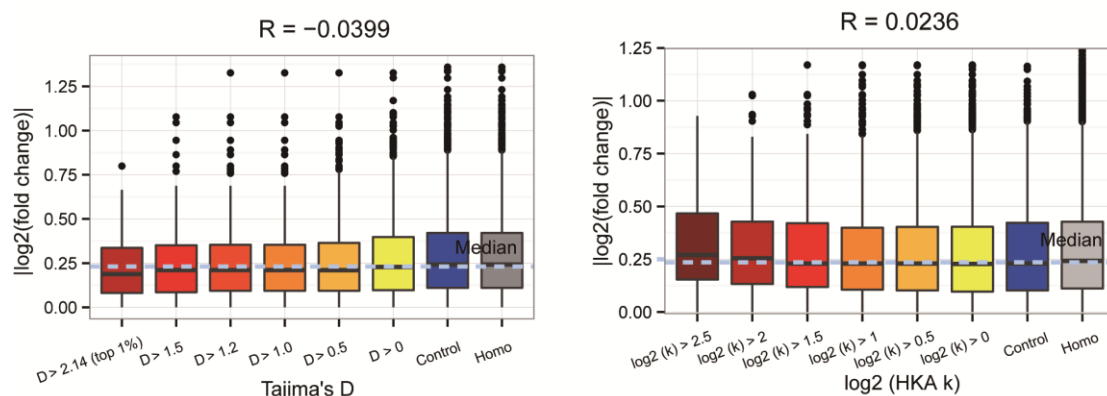


Figure S11. No association between the selection strengths at footprint SNPs and transcriptional responsiveness. (A-B) Transcriptional responsiveness of each gene was estimated as the change in gene expression level in response to the treatment of (A) doxorubicin or (B) ionizing radiation. The maximum change was obtained from multiple treatment experiments or time points. (C) Change in chromatin accessibility was calculated as the fold change of peak density in response to TNF α . (A-C) The grouping was based on Tajima's D (left) or HKA k (right) for the associated regulatory SNPs.