The American Journal of Human Genetics, Volume *102*

Supplemental Data

A Common Type 2 Diabetes Risk Variant Potentiates Activity of an Evolutionarily Conserved Islet Stretch Enhancer and Increases *C2CD4A* and *C2CD4B* Expression

Ina Kycia, Brooke N. Wolford, Jeroen R. Huyghe, Christian Fuchsberger, Swarooparani Vadlamudi, Romy Kursawe, Ryan P. Welch, Ricardo d'Oliveira Albanus, Asli Uyar, Shubham Khetan, Nathan Lawlor, Mohan Bolisetty, Anubhuti Mathur, Johanna Kuusisto, Markku Laakso, Duygu Ucar, Karen L. Mohlke, Michael Boehnke, Francis S. Collins, Stephen C.J. Parker, and Michael L. Stitzel

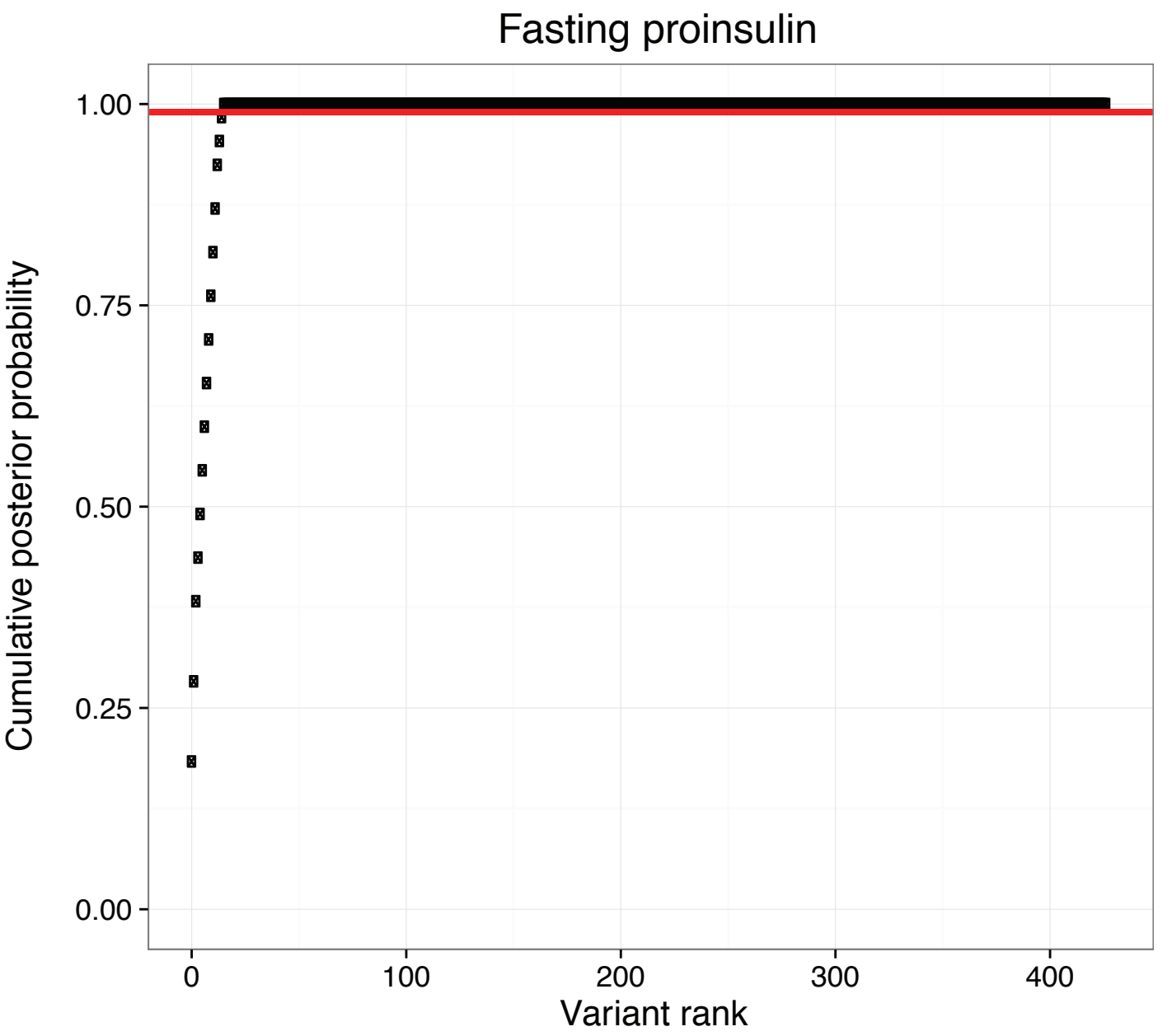


Figure S1. "Credible set" SNP analyses in the C2CD4A/B/VPS13C locus. Cumulative posterior probability plot from Bayesian analyses (Methods) of fasting proinsulin associations in METSIM. Among ~400 variants considered (Table S6), the linked variants identified by fine-mapping analyses in Figure 1 and Table 1 comprise the 99% credible set (indicated by the red line) in these analyses.

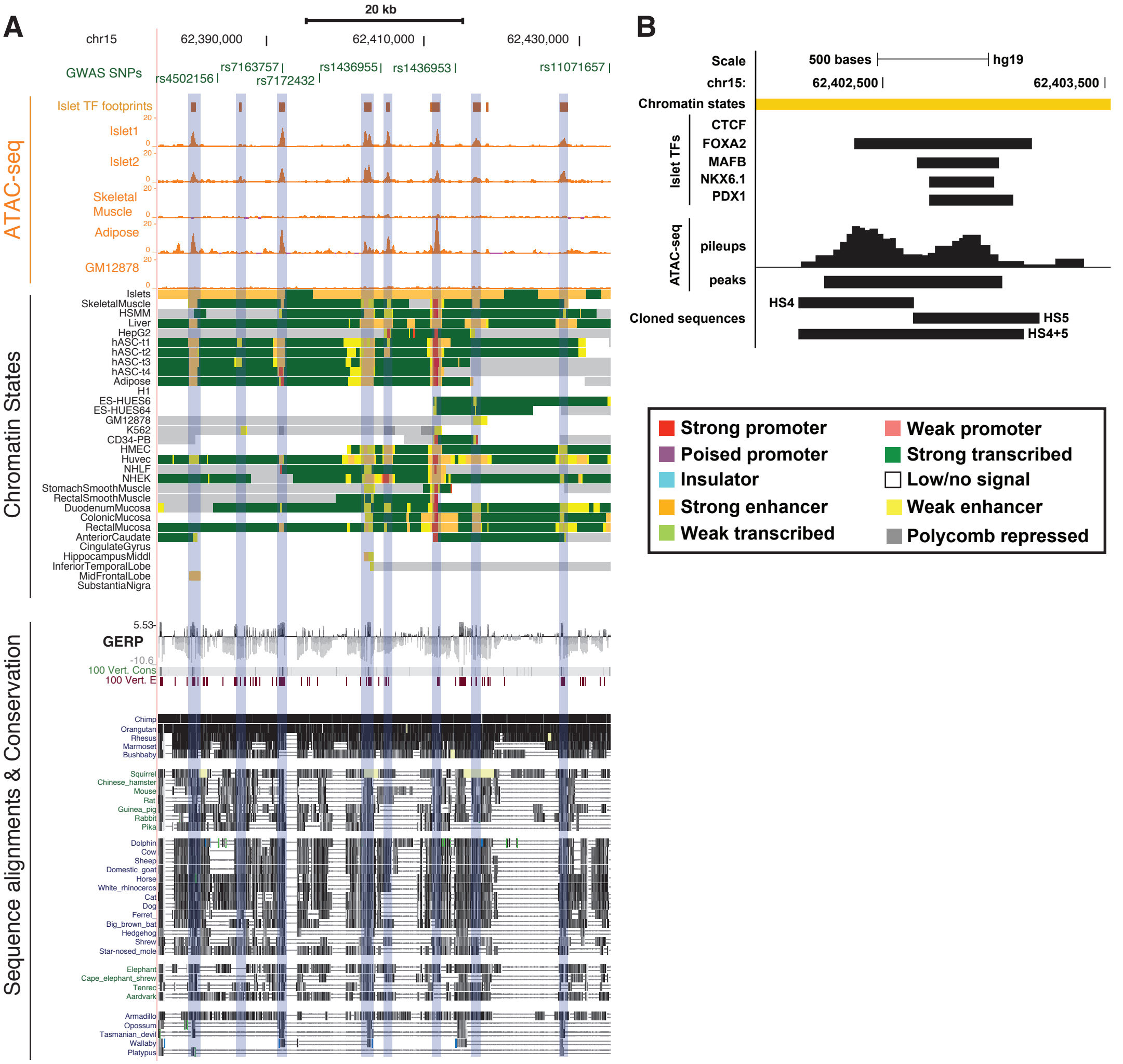
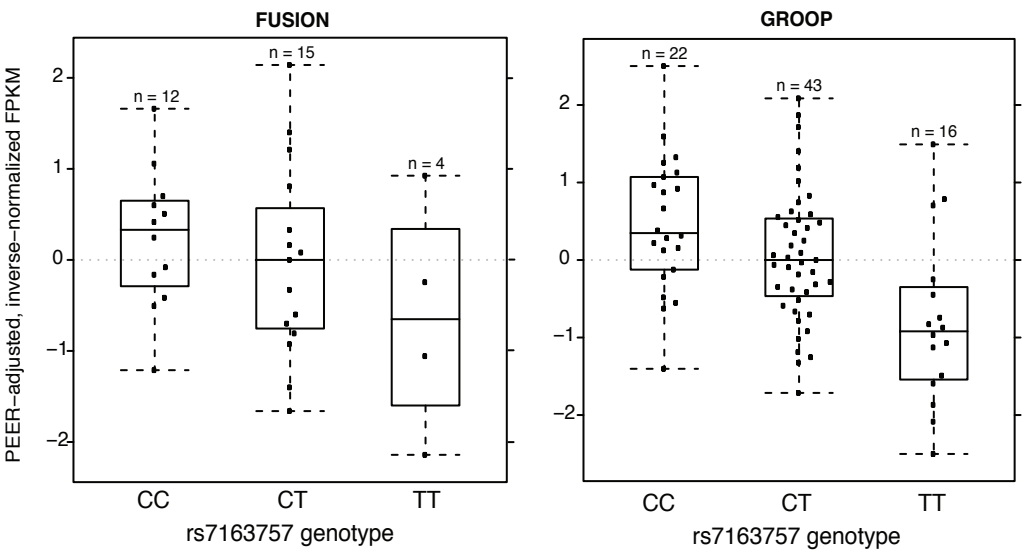
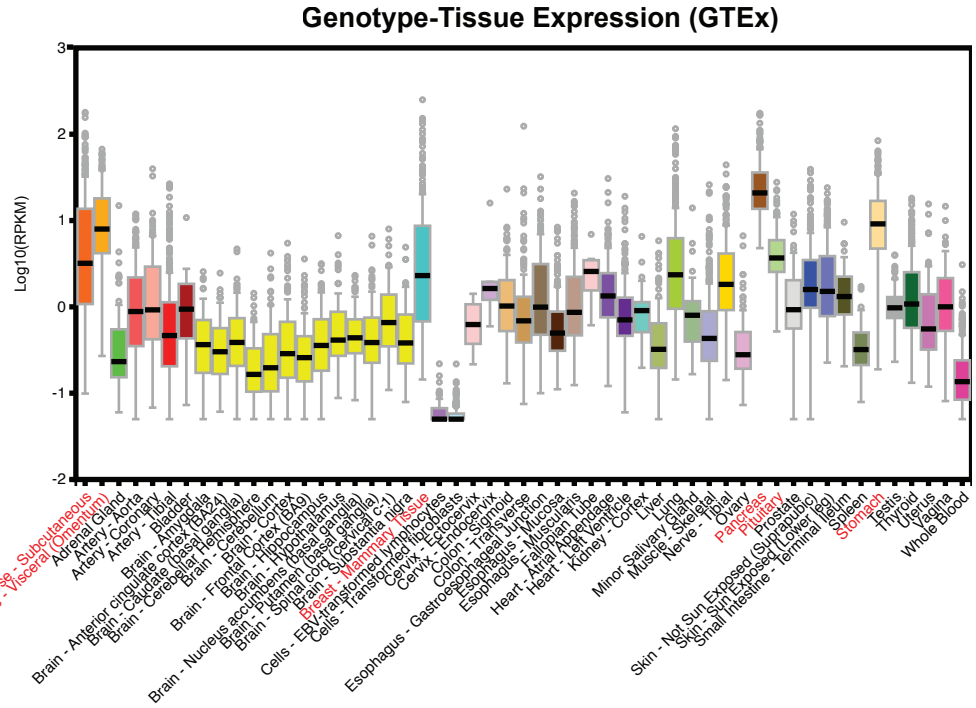


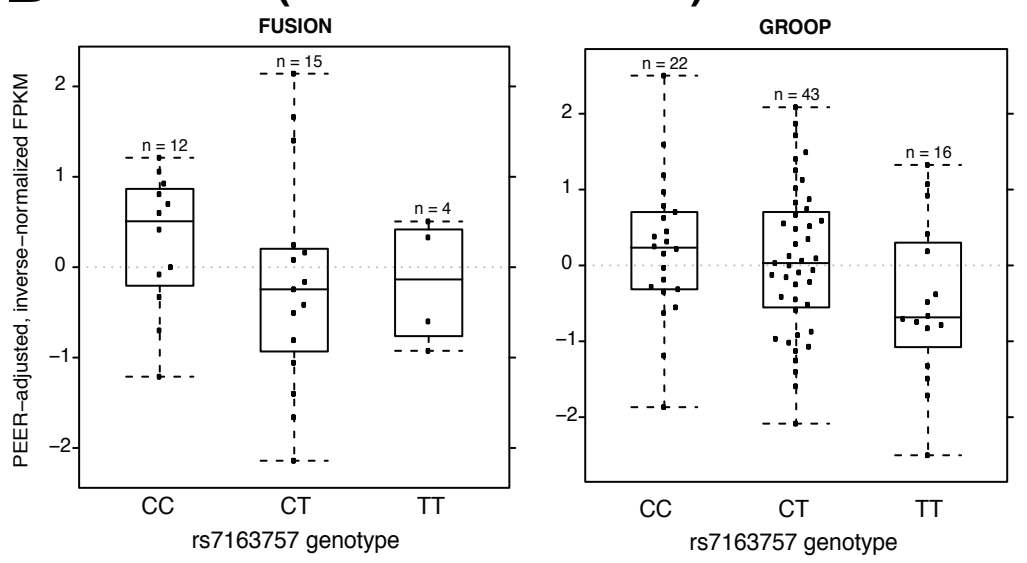
Figure S2. Cross-tissue chromatin states and cross-species conservation of DNA sequences in C2CD4A/B SE. (A) UCSC Genome Browser view of open chromatin sites (ATAC-seq, orange), chromatin states, DNA sequence conservation metrics (GERP, 100 Vert cons, 100 Vert El), and DNA sequence alignments of the C2CD4A/B/VPS13C regulatory region on chromosome 15. Index SNPs associated with T2D or quantitative measures of islet dysfunction are shown in green. Gray rectangles highlight the constituent open chromatin sites (HS1-9) in this islet SE. (B) UCSC Genome Browser view of human islet stretch enhancer constituent open chromatin sites of HS4, HS5 and HS4+5 (collectively HS4/5#). Strong enhancer chromatin state (orange), islet transcription factor (TF) ChIP-seq, ATAC-seq pileup and MACS peak calls are indicated. Chromatin states colors correspond to those in Figure 2. Genome browser coordinates correspond to hg19.

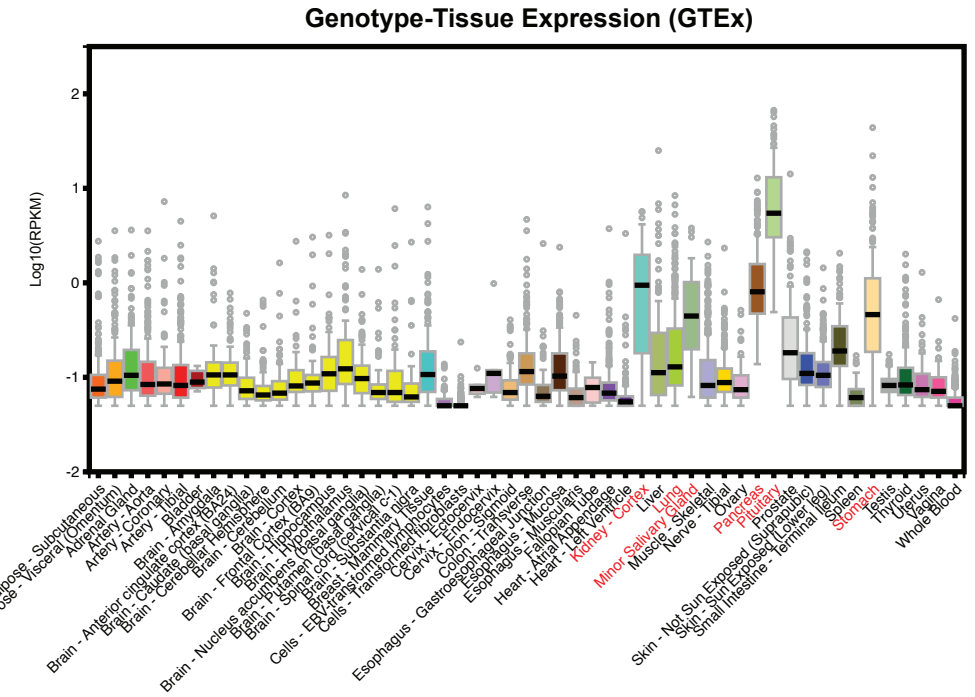
A *C2CD4B* (ENSG00000205502)



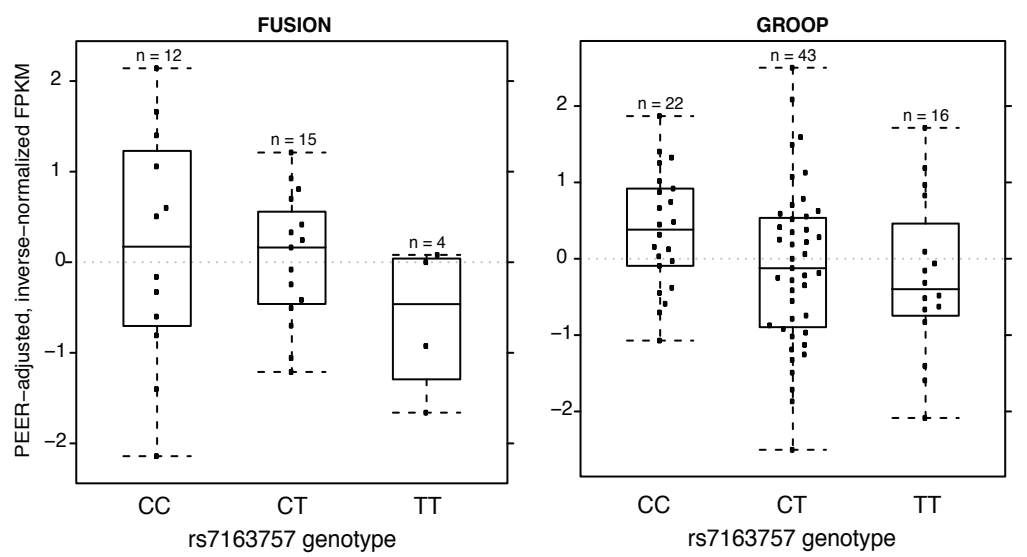


B *C2CD4A* (ENSG00000198535)





C *VPS13C* (ENSG00000129003)



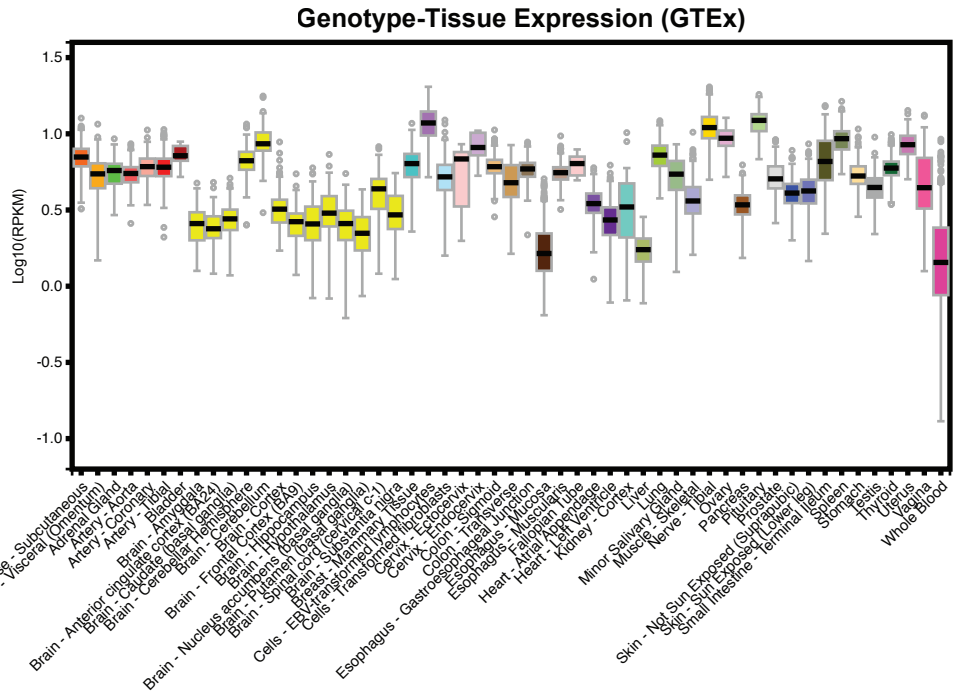


Figure S3. Expression of putative rs7163757 target genes C2CD4A, C2CD4B, VPS13C in islets and Genotype Tissue Expression (GTEx) Consortium tissues. Left, Box plots of PEER-adjusted, inverse-normalized RNA-seq expression (FPKM) in FUSION (n=32) and Groop (n=81) cohort islets. Right, Box plots downloaded from the Genotype-Tissue Expression (GTEx) Portal (https://www.gtexportal.org/home/) showing RNA-seq expression levels (log(10)RPKM) among profiled tissues.

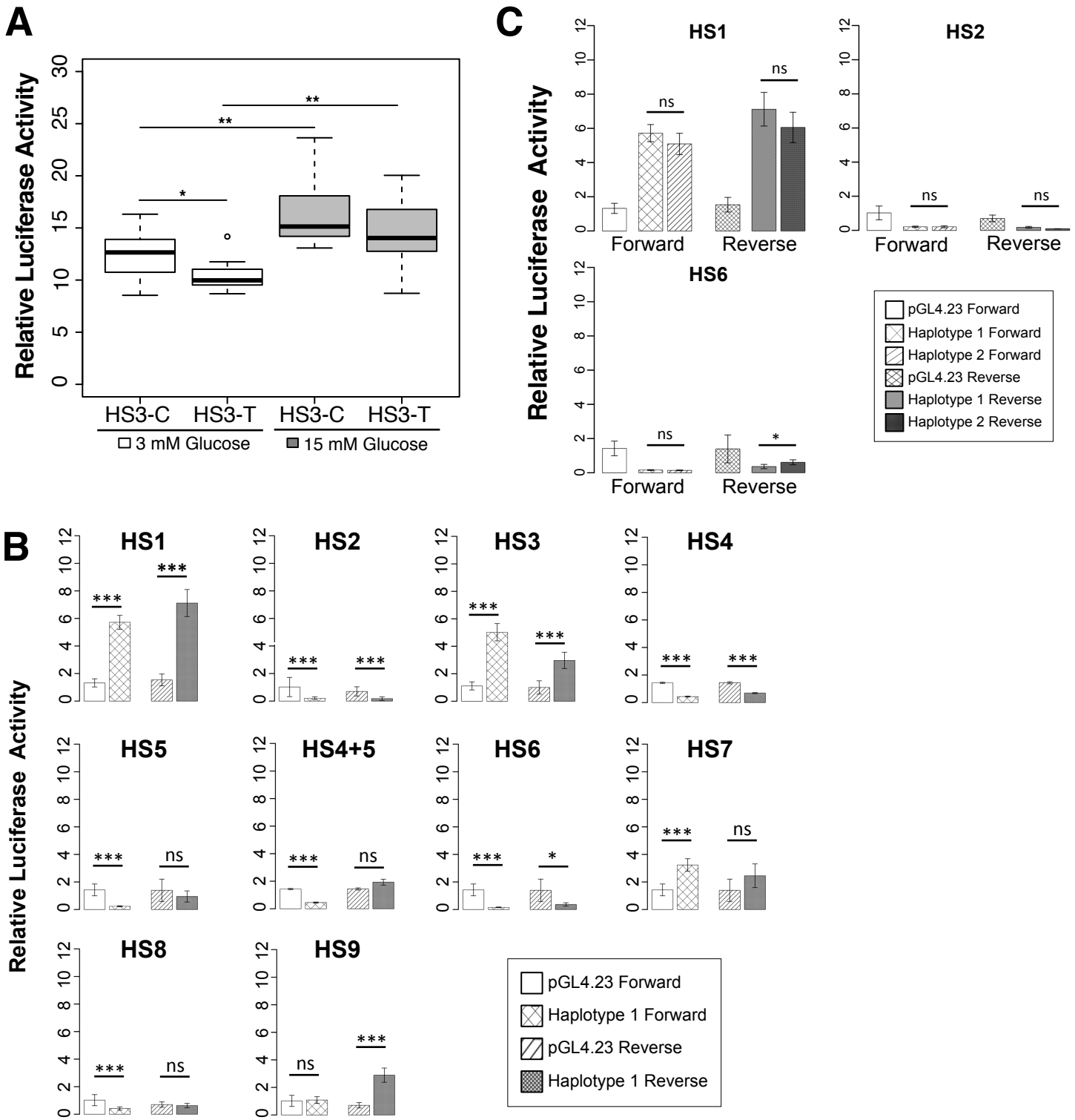


Figure S4. HS3 (containing the rs7163757 variant) is the only enhancer constituent displaying differential enhancer activity of alleles/haplotypes. (A) Luciferase assays comparing enhancer activity of rs7163757 risk (C) and non-risk (T) alleles in INS-1(832/13) cells under low (3 mM) and high (15 mM) glucose conditions. Both sequences exhibited glucose-stimulated increases in enhancer activity. The rs7163757 risk (C) allele showed significantly higher enhancer activity in 3mM glucose and a similar trend in 15 mM glucose. **(B)** Luciferase reporter assay of islet stretch enhancer constituents (cloned in forward and reverse orientation) of HS1-9 (haplotype 1 only). **(C)** Luciferase assays comparing haplotype 1 and haplotype 2 activity for HS1, HS2, and HS6. All cloned sequences were significantly different than the control vectors, except for HS6 Haplotype 2 Reverse (p-values not indicated). Luciferase activity for each element tested in A and B was normalized to the activity of the empty forward and reverse vectors and is plotted as fold activity over these empty vector controls on the y-axis. Data in A and B represent the mean values of four to sixteen clones per sequence/haplotype ± SEM. *, **, *** denote $p<0.05$, <0.01, <0.001 according to two-sided, unpaired t-test. ns = non-significant differences.

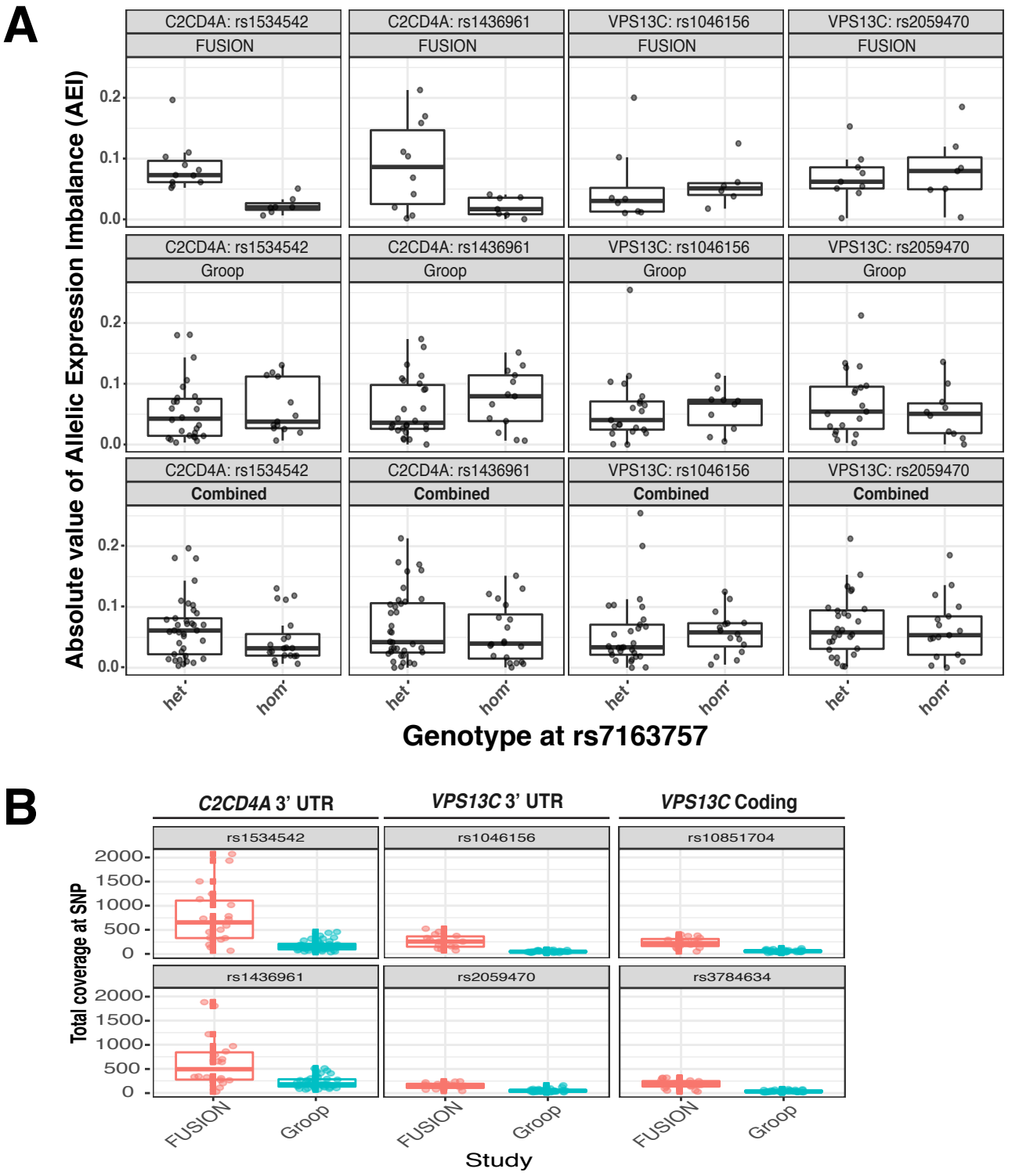


Figure S5. Allelic Expression imbalance (AEI) of C2CD4A and VPS13C tSNPs in islet cohorts. (A) Allelic imbalance of RNA-seq reads overlapping tSNPs in C2CD4A or VPS13C mRNA for FUSION, Groop, and combined islet cohorts. Dot and box plots compare allelic imbalance between rs7163757 heterozygous (het) and homozygous (hom) islet samples. Number of heterozygous and homozygous samples for each tSNP are indicated in Table S11; * denotes q<0.05, two-sided Wilcoxon Rank Sum Test adjusted for multiple testing (see Methods and Table S11). **(B)** Dot and box plots indicating RNA-seq sequence coverage at tSNPs for FUSION and Groop islet samples. 3'UTR = 3' untranslated region.

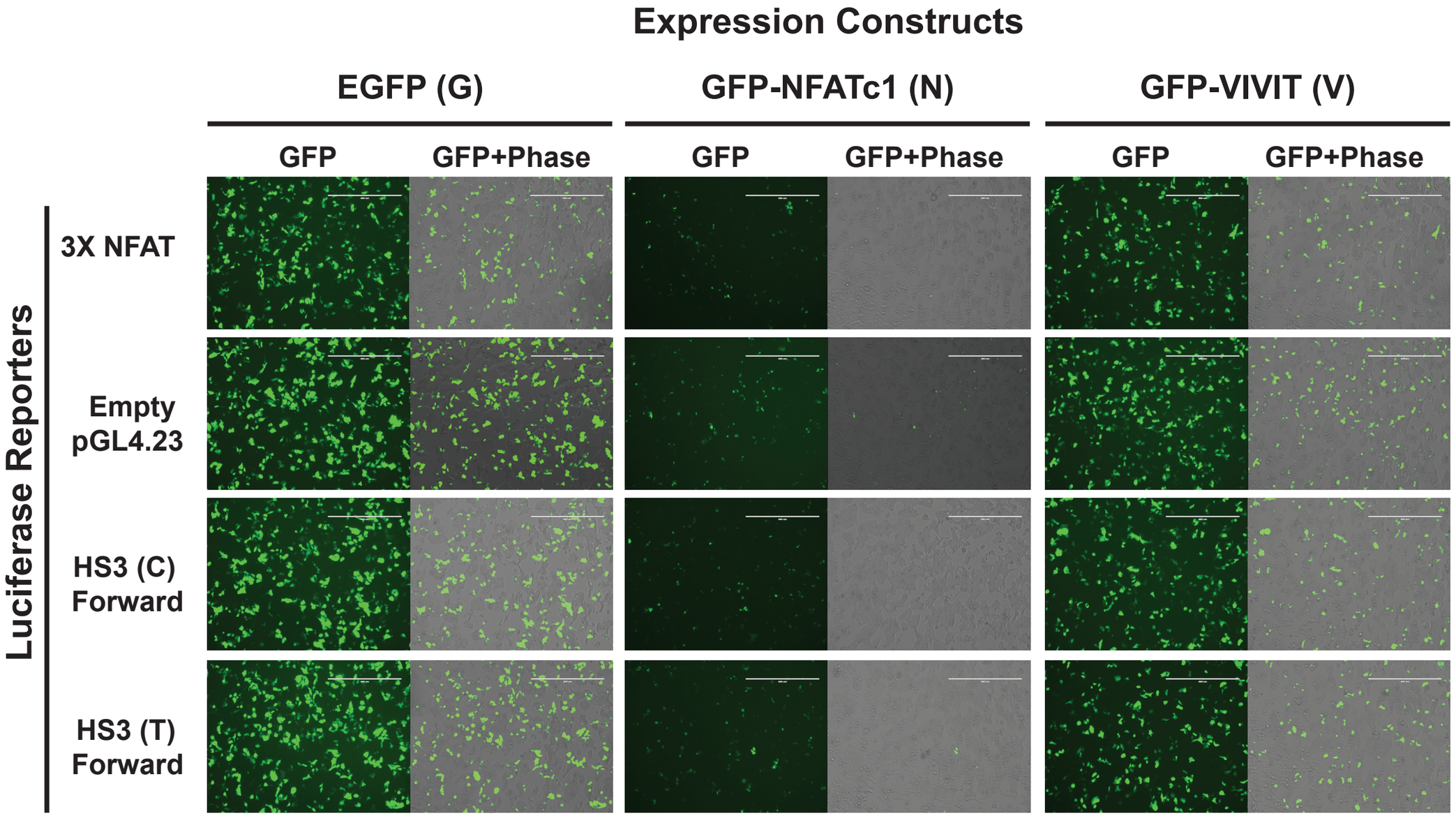


Figure S6. GFP fusion protein levels for co-transfected EGFP, NFAT, and VIVIT expression constructs to determine NFAT effects on HS3 enhancer activity. Representative images for each co-transfection are shown. For all HS3 enhancer sequences co-transfected with EGFP/EGFPC1-huNFATc1EE-WT/EGFPN1-VIVIT eight biological replicates were measured over four independent experiments. EGFP signal intensity for EGFP transfected cells (G, left) EGFPC1-huNFATc1EE-WT (N, middle), EGFPN1-VIVIT (V, right) is shown for cells co-transfected with the following luciferase reporter plasmids: pGL3-NFAT-Luciferase (3X NFAT, top) empty pGL4.23 empty vector (row 2), and HS3 sequence cloned in the forward orientation containing the rs7163757 risk (C) allele (row 3) or non-risk (T) allele (row 4). GFP levels for each expression construct co-transfected with luciferase reporters containing HS3 sequences cloned in reverse orientation (not shown) were comparable to those shown for the HS3 Forward luciferase reporters.

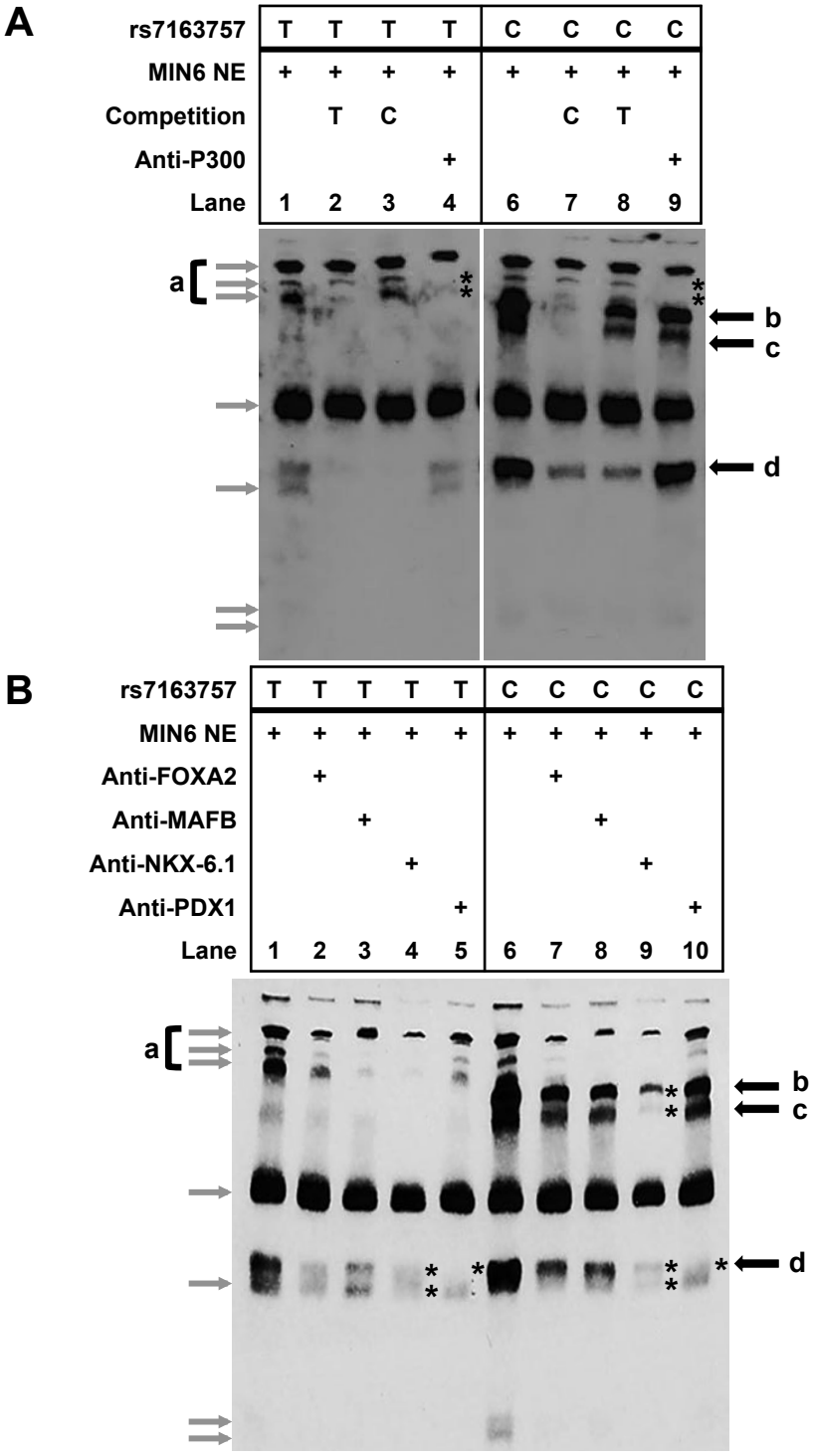


Figure S7. Beta cell nuclear extract binding to sequences containing different rs7163757 alleles. EMSA of biotin-labeled probes containing rs7163757 non-risk (T, left) or risk (C) allele and incubated with MIN6 nuclear extract (NE). These experiments were performed independently from each other and the EMSA shown in Figure 6E. Arrows are labeled to match bands corresponding to Figure 6E. **(A)** Bands labeled 'b', 'c' and 'd' indicate proteins that bind more strongly to the (C) allele. Bands 'b' and 'c' are competed away more strongly by excess of unlabeled (C) probe than (T) probe. The asterisks in lanes 4 and 9 indicate that including P300 antibody disrupts two of the non-allele-specific bands labeled 'a'. **(B)** Including antibodies to FOXA2, MAFB, NKX6.1, and PDX1 disrupted several of the bands. All antibodies appear to disrupt bands labeled 'a'. Antibodies to NKX6.1 most strongly disrupted band 'c' and also appeared to disrupt bands 'b' and 'd' (asterisks in lanes 4 and 9). Antibodies to PDX1 disrupted band 'd' (asterisks in lanes 5 and 10). Antibodies to FOXA2 and MAFB inconsistently disrupted band 'd'.