

Figure S1: Quantile-quantile plots of the associations between variants and proinsulin adjusted for BMI. Panel A shows $-\log_{10}(\text{p-values})$ of all analyzed variants adjusted for BMI (X axis). Panel B shows $-\log_{10}(\text{p-values})$ after excluding all variants within 500 kb of a previously-identified signal.

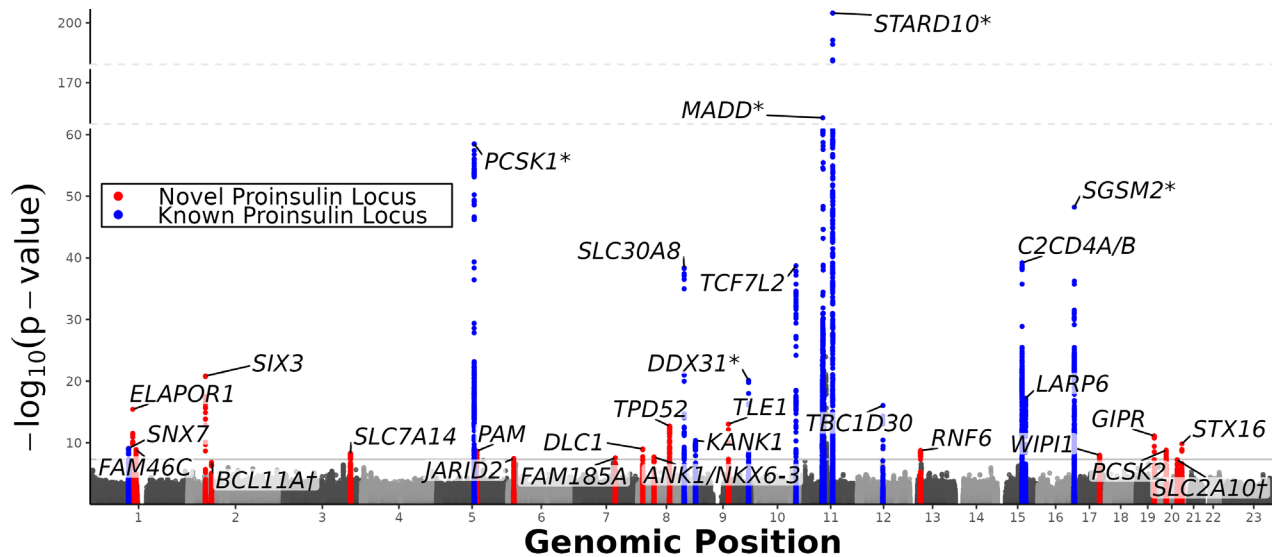
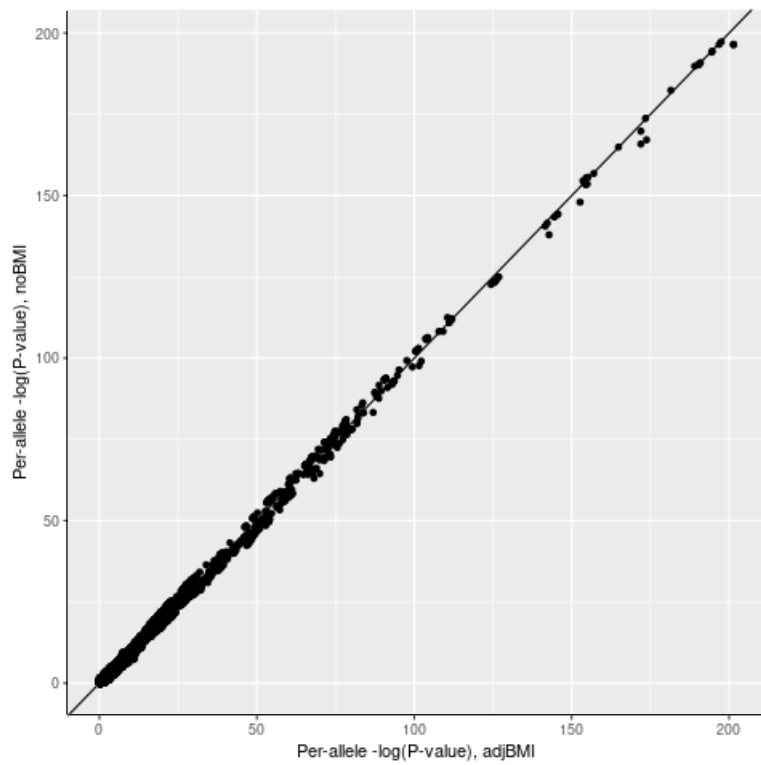


Figure S2: Manhattan plots of the associations between variants and proinsulin adjusted for BMI. Variants within 500 kb of a previously-known proinsulin locus are colored in blue; variants within 500 kb of a new proinsulin locus ($p < 5 \times 10^{-8}$) are colored in red. Loci with asterisks represent multi-signal loci. Loci with dagger represent loci only identified in models not adjusted for BMI.

A



B

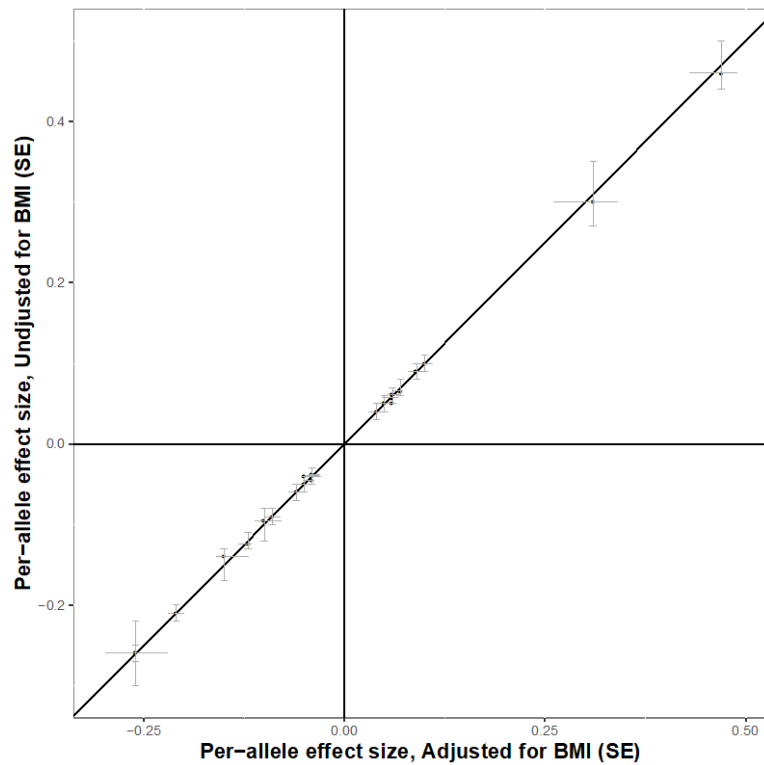


Figure S3: Influence of BMI on genetic associations with fasting proinsulin adjusting for BMI. Adjusting for BMI does little to influence the association between genetic variant and fasting proinsulin (Pearson Correlation, betas = 0.97). Panel A shows $-\log_{10}(p\text{-values})$ of all analyzed variants adjusted for BMI (X axis) against unadjusted for BMI (Y axis). Panel B shows the effect size (SE) of the significant lead variants with and without adjusting for BMI.

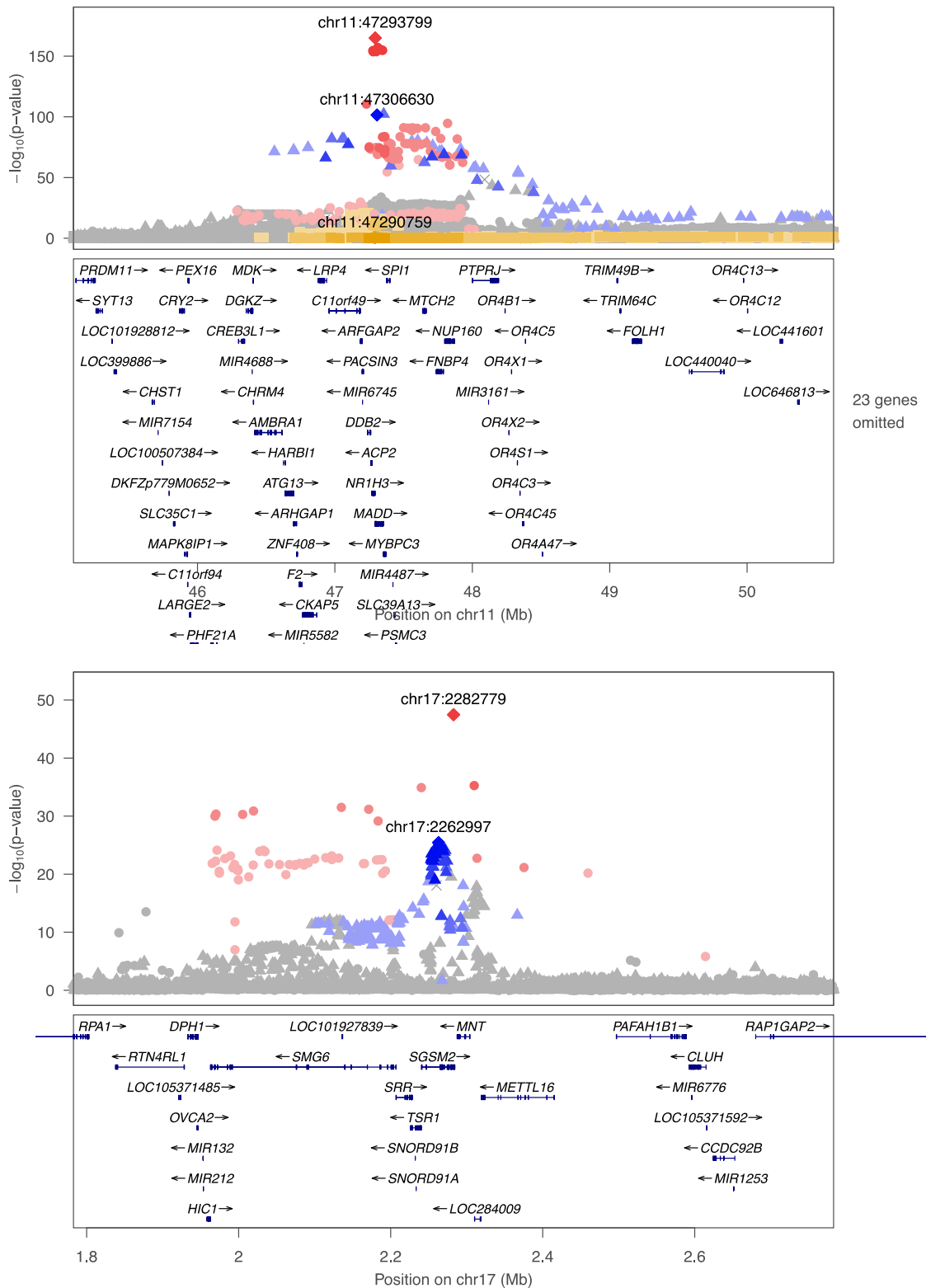


Figure S4: Conditionally distinct signals in four multiple-signal loci. Top, three signals in *MADD*. Bottom, two signals in *SGSM2*. The primary signal for the locus is shown in red, the secondary signal for the locus is shown in blue, and the tertiary locus is shown in yellow. Continued on the next page.

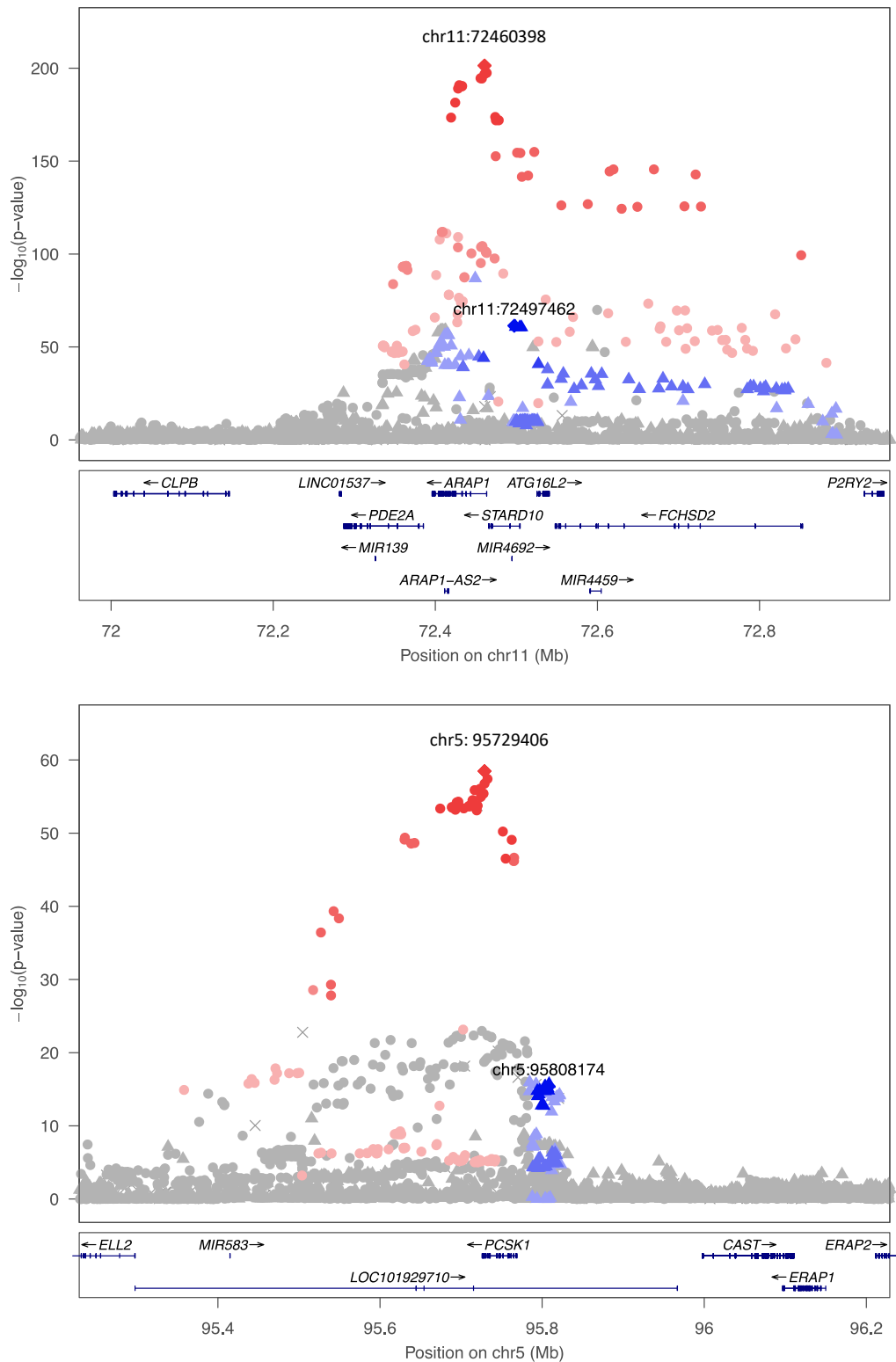


Figure S4 (continued): Conditionally distinct signals in four multiple-signal loci. Top, two signals in *STARD10*. Bottom, two signals in *PCSK1*. The primary signal for the locus is shown in red, the secondary signal for the locus is shown in blue.

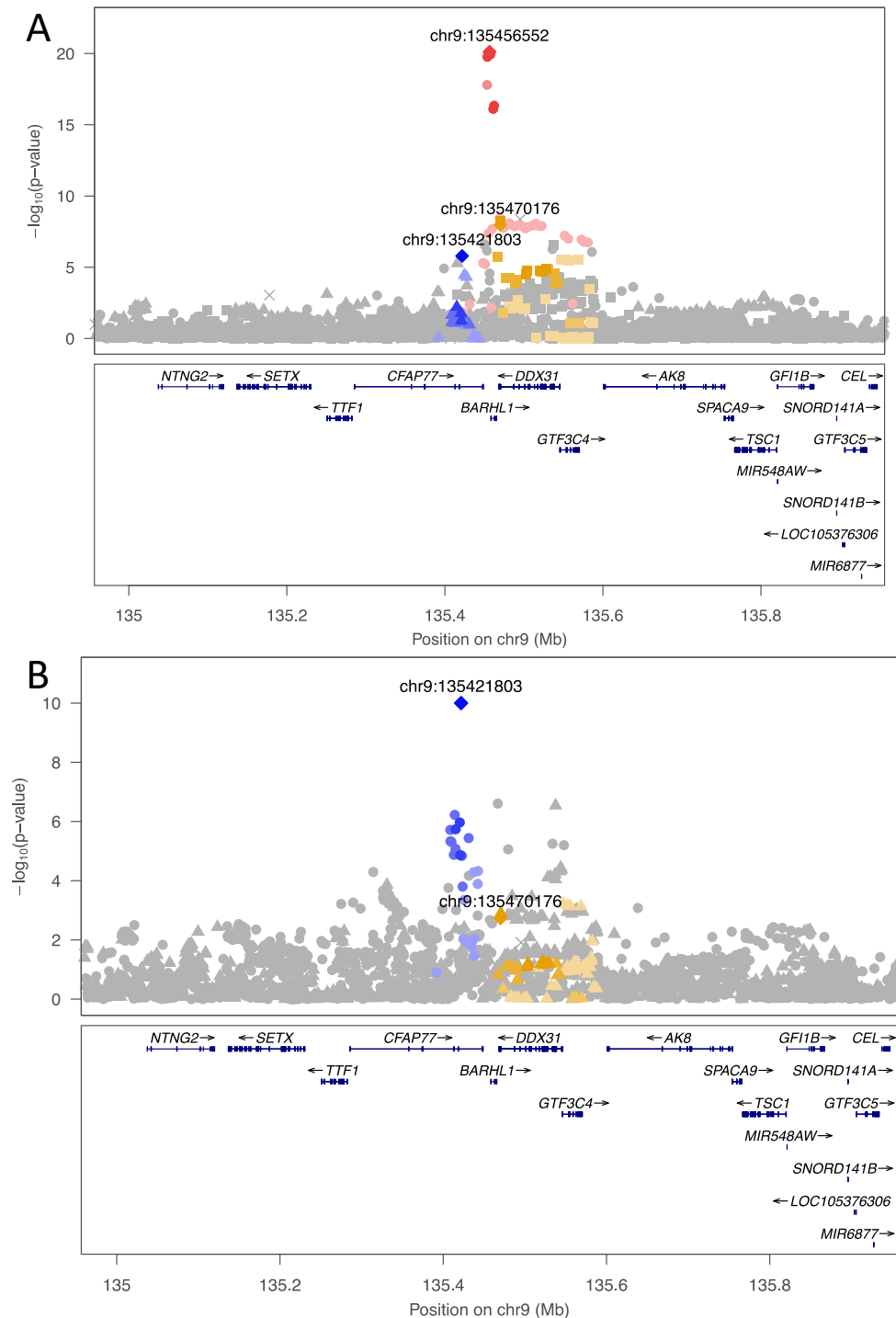


Figure S5: *DDX31* proinsulin signals. A: The two conditionally distinct proinsulin signals at the *DDX31* locus, shown in red and blue. In yellow, the female-specific *DDX31* signal identified in Strawbridge (2011), lead variant chr9:135470176, that was not identified in this meta-analysis. B: Proinsulin results after conditioning on the first signal. The secondary *DDX31* proinsulin signal is in blue and the female-specific *DDX31* signal identified by Strawbridge is in yellow. The Strawbridge lead is in low LD ($r^2 = 0.09$) with the primary *DDX31* lead variant (chr9:135456552), and does not reach the significance threshold after conditioning on the primary signal or conditioning on both the primary and secondary signals.

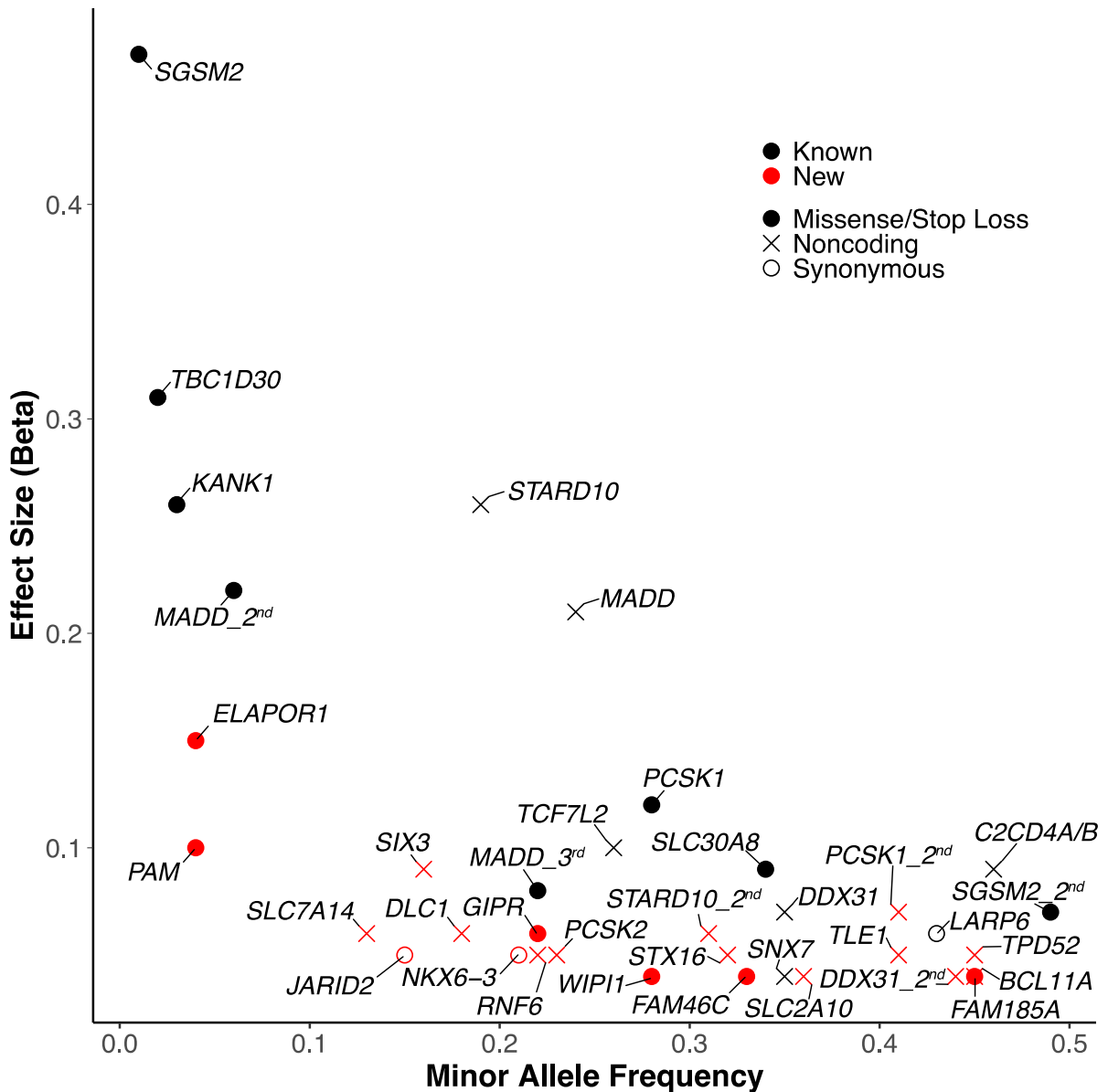


Figure S6a: Effect size versus minor allele frequency of lead proinsulin-associated variants. Dots representing signals that have been identified previously are shown in black and novel proinsulin signals are shown in red. Filled circles indicate loci with a missense or stop loss variant in the extended credible set; open circles indicate loci with a synonymous variant in the extended credible set; x indicates loci with only noncoding variants in the extended credible set.

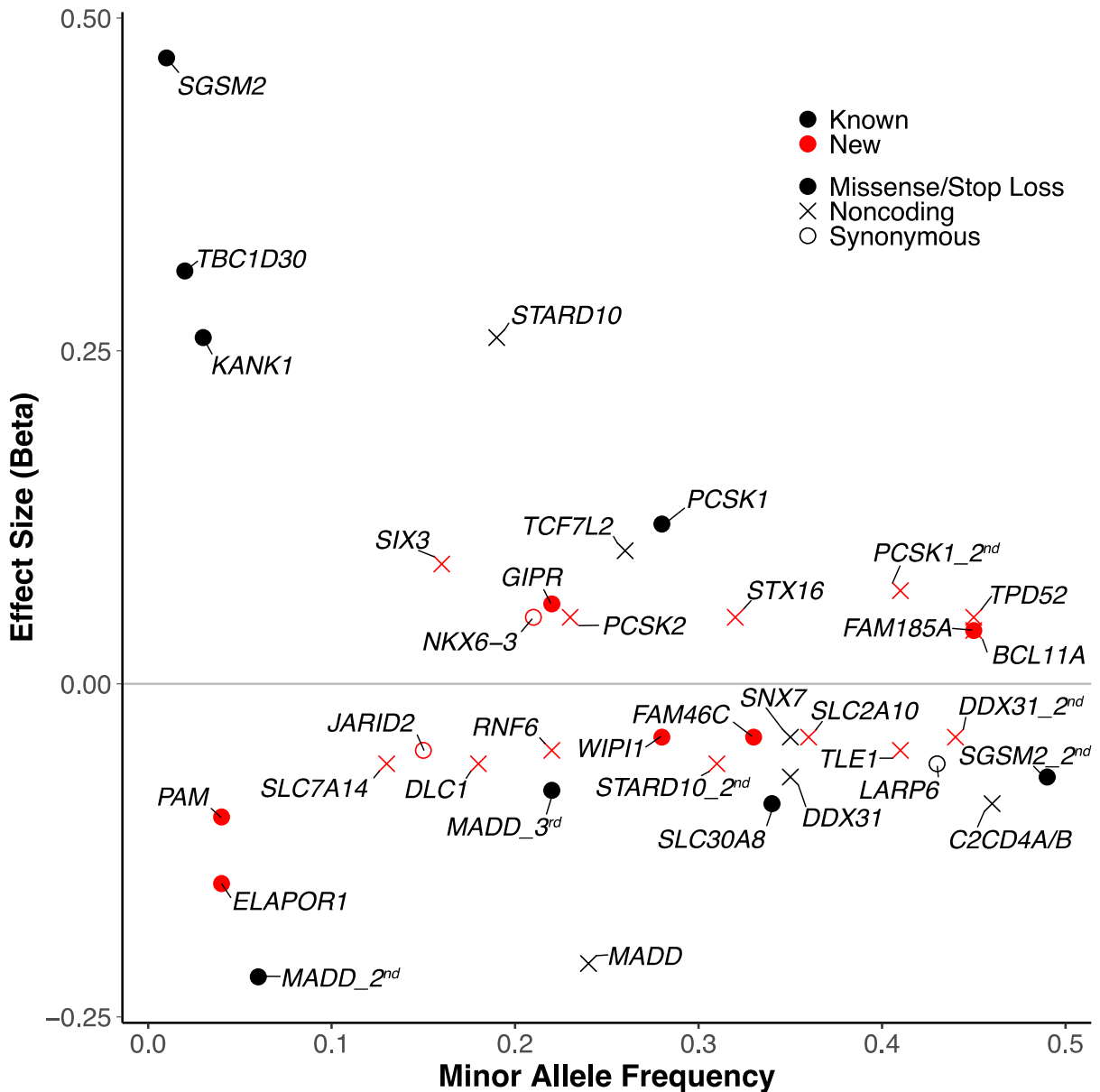
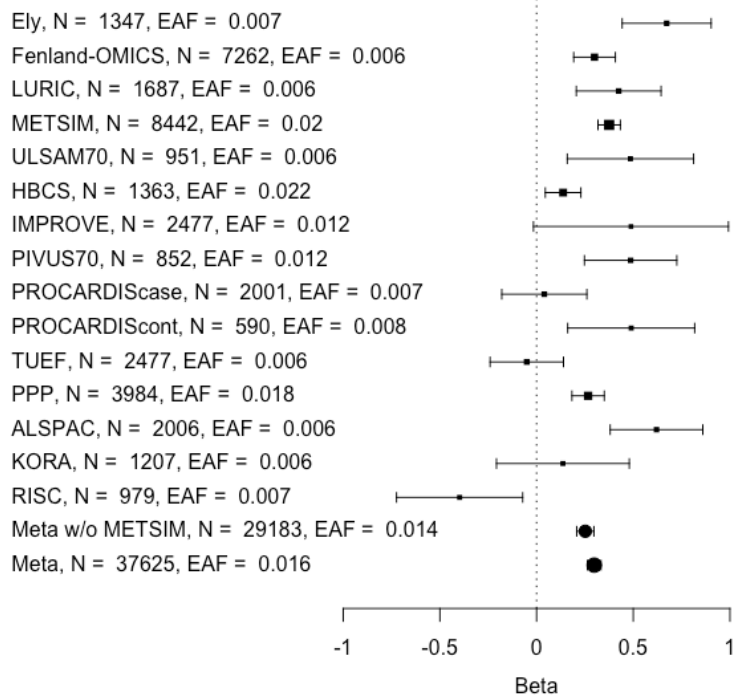


Figure S6b: Effect size versus minor allele frequency of lead proinsulin-associated variants. Effect size is flipped to the minor allele. Dots representing signals that have been identified previously are shown in black and novel proinsulin signals are shown in red. Filled circles indicate loci with a missense or stop loss variant in the extended credible set; open circles indicate loci with a synonymous variant in the extended credible set; x indicates loci with only noncoding variants in the extended credible set.

TBC1D30



SGSM2

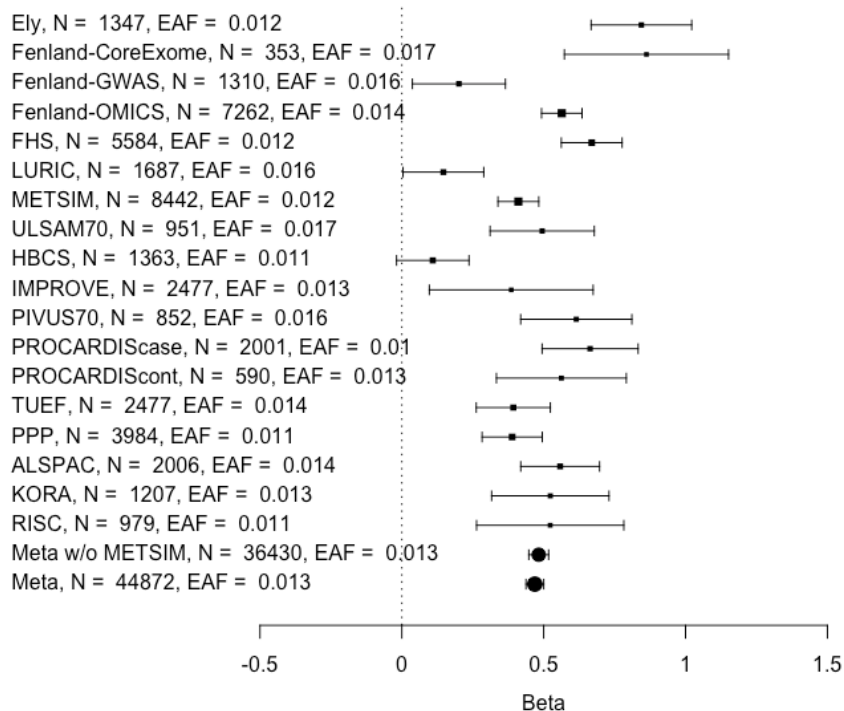
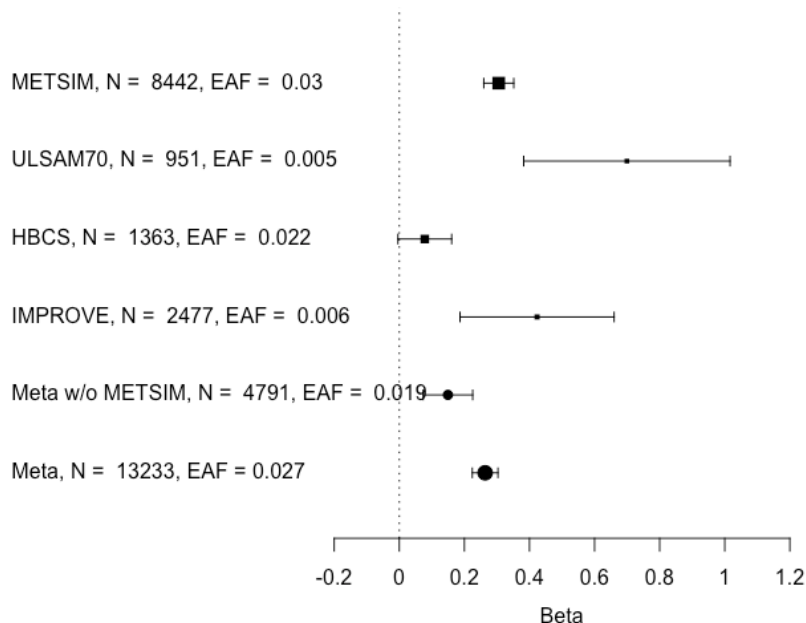


Figure S7: Replication of low frequency proinsulin-associated variants first identified in Huyghe (2013). We replicated the four low-frequency proinsulin-associated variants described in the METSIM exome array study. Continued on next page.

KANK1



MADD

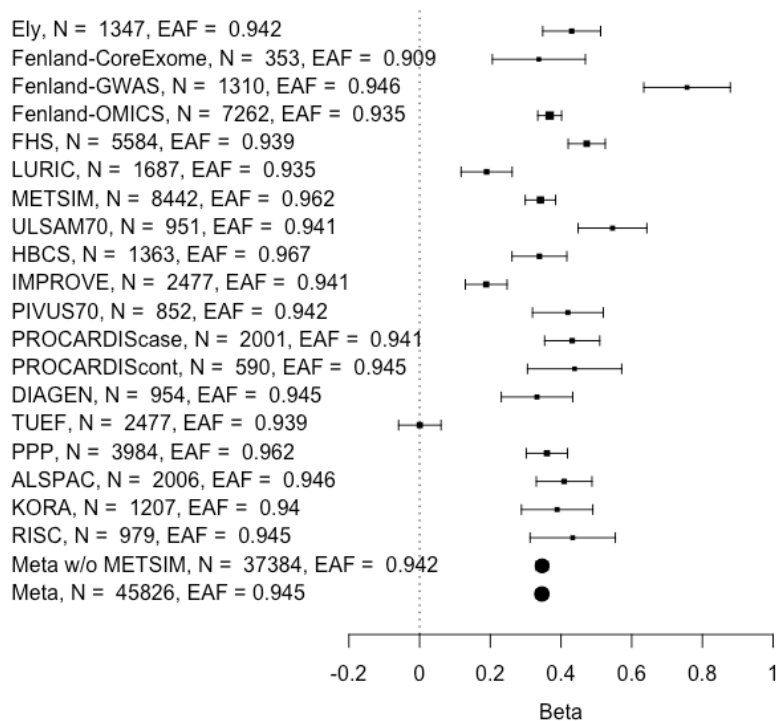


Figure S7 (Continued): Replication of low frequency proinsulin-associated variants first identified in Huhghe (2013). We replicate the four low-frequency proinsulin-associated variants described in the METSIM exome array study.

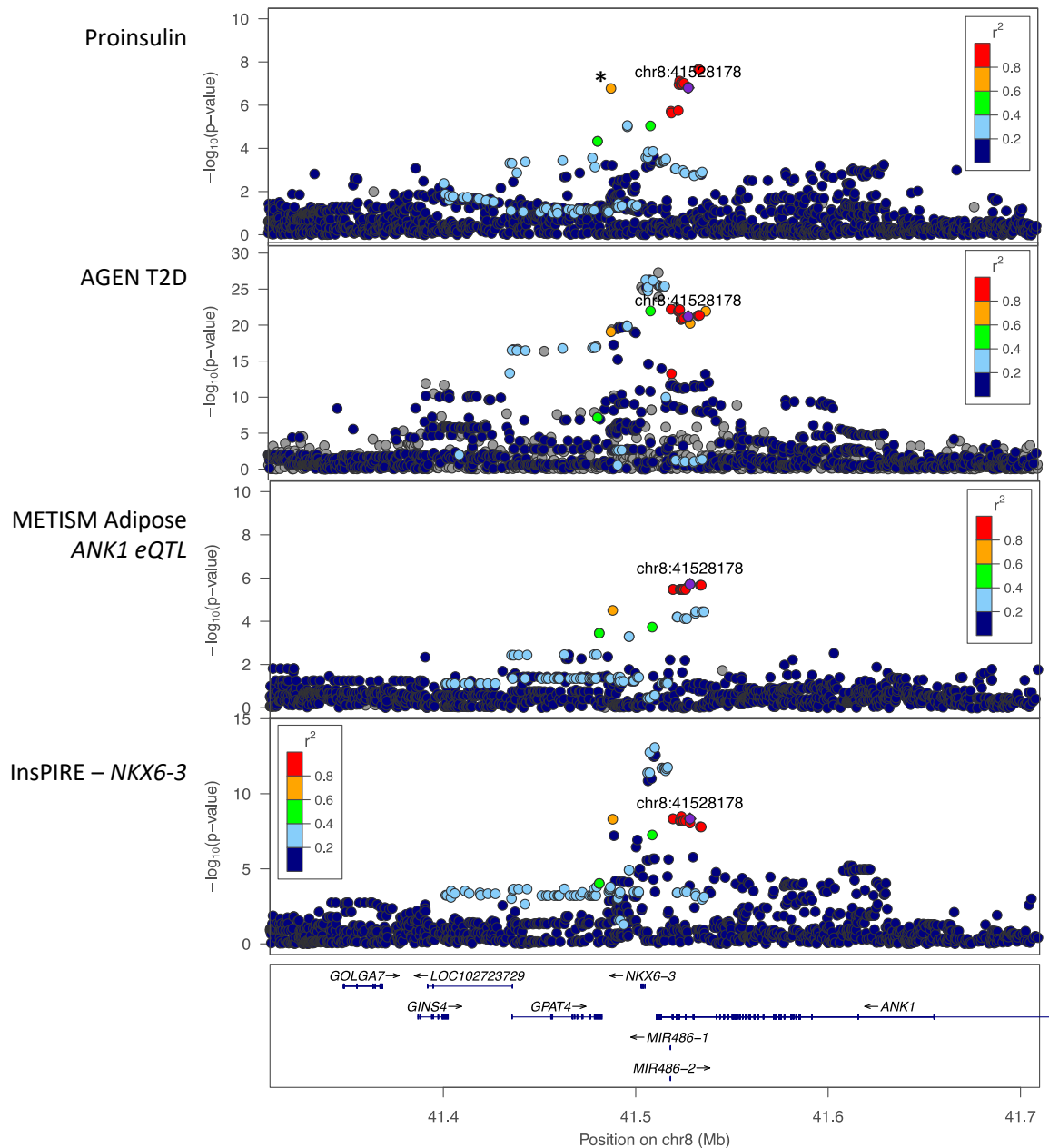


Figure S8: *ANK1/NKX6-3* locus associated with proinsulin and T2D in AGEN EAS, the InsPIRE *NKX6-3* islet eQTL and METSIM *ANK1* adipose eQTL data. Plots shown colored by lead variant for METSIM adipose eQTL. The proinsulin signal colocalizes with the *ANK1* adipose eQTL signal and the secondary AGEN T2D signal. Continued on next page. * chr8:41523745 shown in yellow (beta = 0.042, MAF = 0.23) has a larger sample size than the lead variant chr8:41533514 or LD proxy chr8:41528178 (beta = 0.046, MAF = 0.21) (n = 45,826 vs 44,872). Continued on next page.

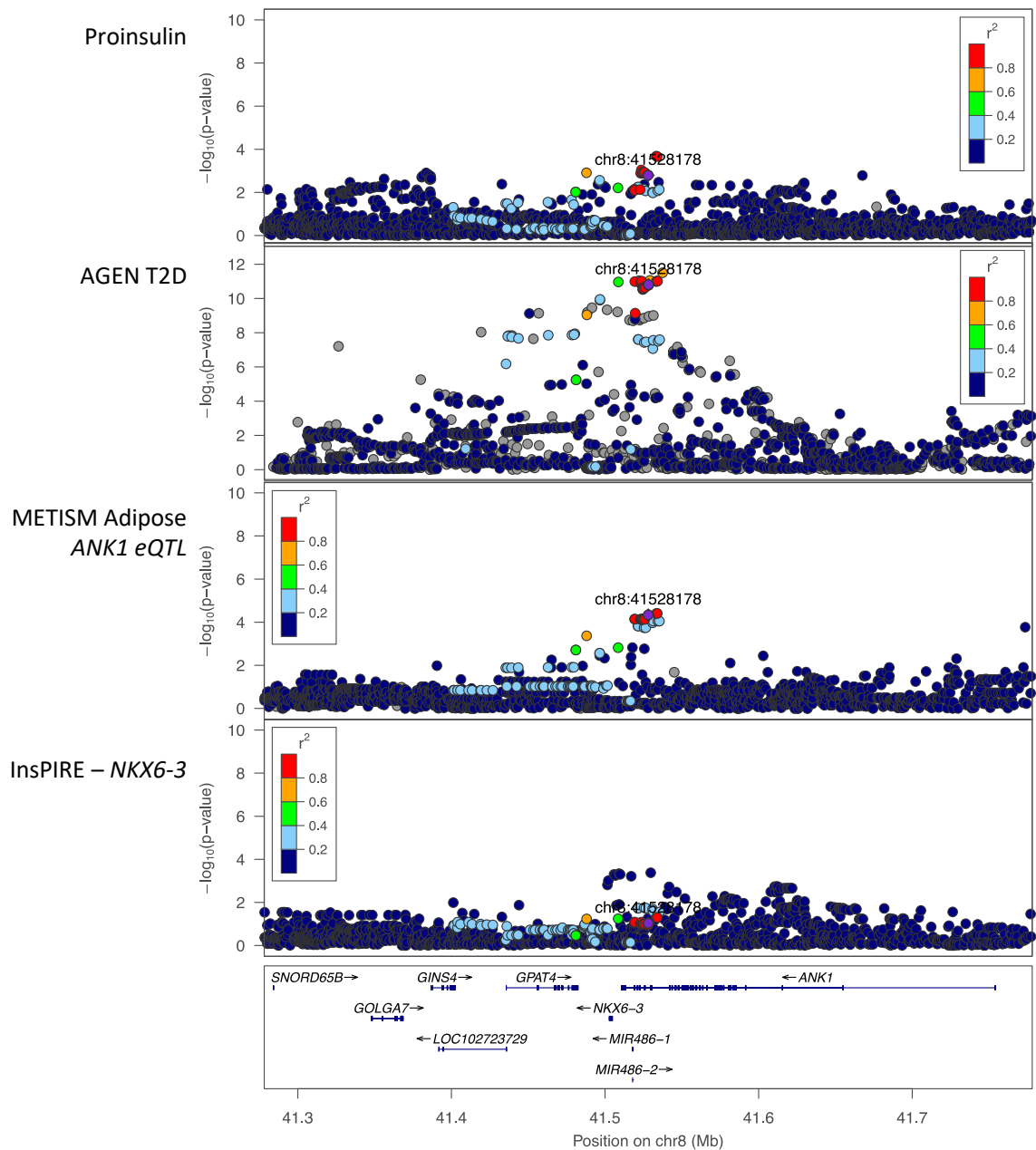


Figure S8 Continued: *ANK1/NKX6-3* locus in proinsulin and T2D AGEN EAS, the InsPIRE *NKX6-3* eQTL and METSIM *ANK1* eQTL data. All results shown have been conditioned on InsPIRE lead variant chr8:41509915, which colocalizes with the first AGEN T2D signal. Plots shown colored by lead variant for METSIM adipose eQTL. The proinsulin signal colocalizes with the *ANK1* adipose eQTL signal and the secondary AGEN T2D signal. Continued on next page.

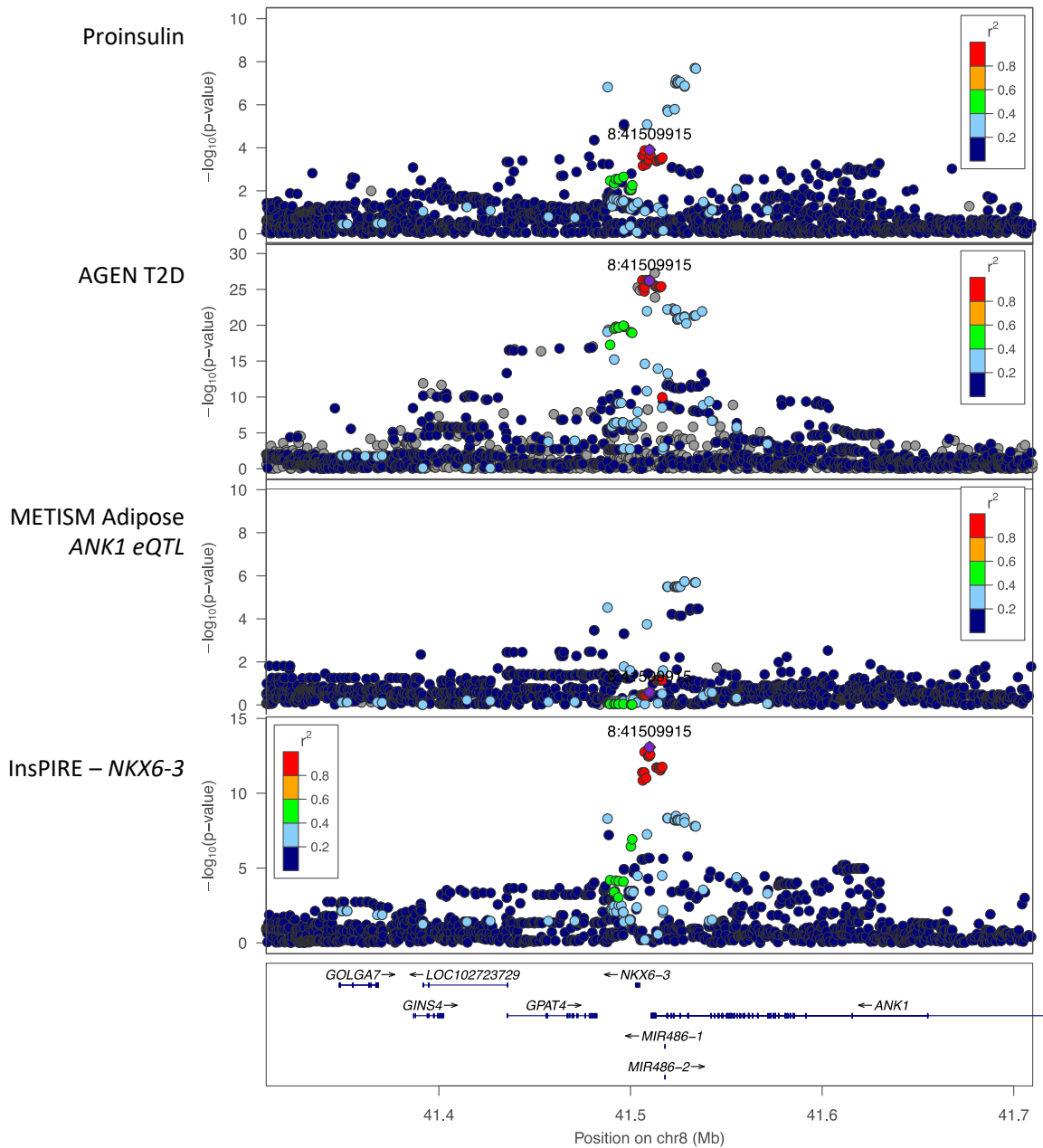


Figure S8 continued: *ANK1/NKX6-3* locus associated with proinsulin and AGEN, the InsPIRE *NKX6-3* eQTL and METSIM *ANK1* eQTL data. Plots shown colored by lead variant for InsPIRE islet eQTL. The proinsulin signal does not colocalize with the InsPIRE *NKX6-3* signal. Continued on next page.

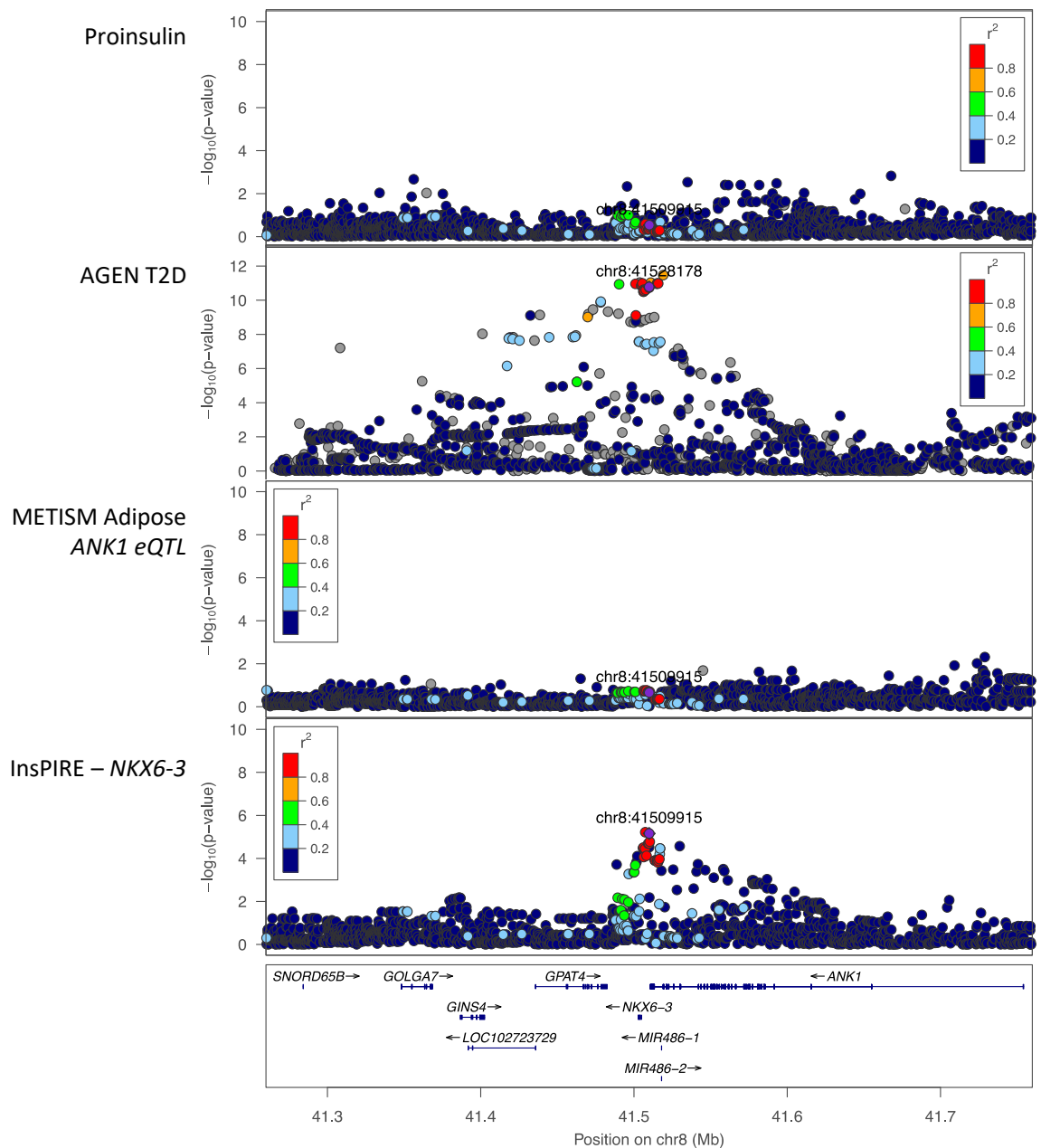


Figure S8 continued: *ANK1/NKX6-3* locus in proinsulin and AGEN, the InsPIRE *NKX6-3* eQTL and METSIM *ANK1* eQTL data. All results shown have been conditioned on METSIM lead variant chr8:41533514, which colocalizes with proinsulin and the second AGEN signal. Plots shown colored by lead variant for InsPIRE islet eQTL. The proinsulin signal does not colocalize with the InsPIRE *NKX6-3* signal.

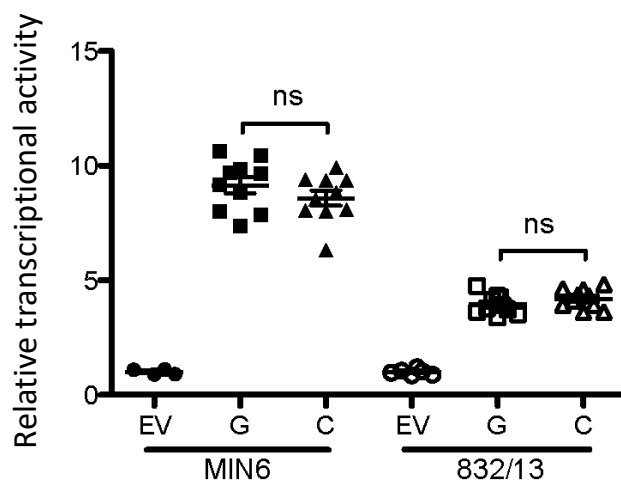


Figure S9: rs10501320 does not exhibit allelic differences in transcriptional activity in the forward orientation. 411-bp fragments including the rs10501320 G or C alleles were cloned upstream of a minimal promoter driving luciferase expression in the forward orientation with respect to the promoter. Values represent fold-change of firefly luciferase/*Renilla* activity normalized to empty pGL4.23 vector in MIN6 and 832/13 cells. EV: empty vector; G/C: alleles at the lead variant rs10501320. In Error bars represent the SEM of four or five independent clones tested across two trials. P-values are calculated from two-sided t-tests.

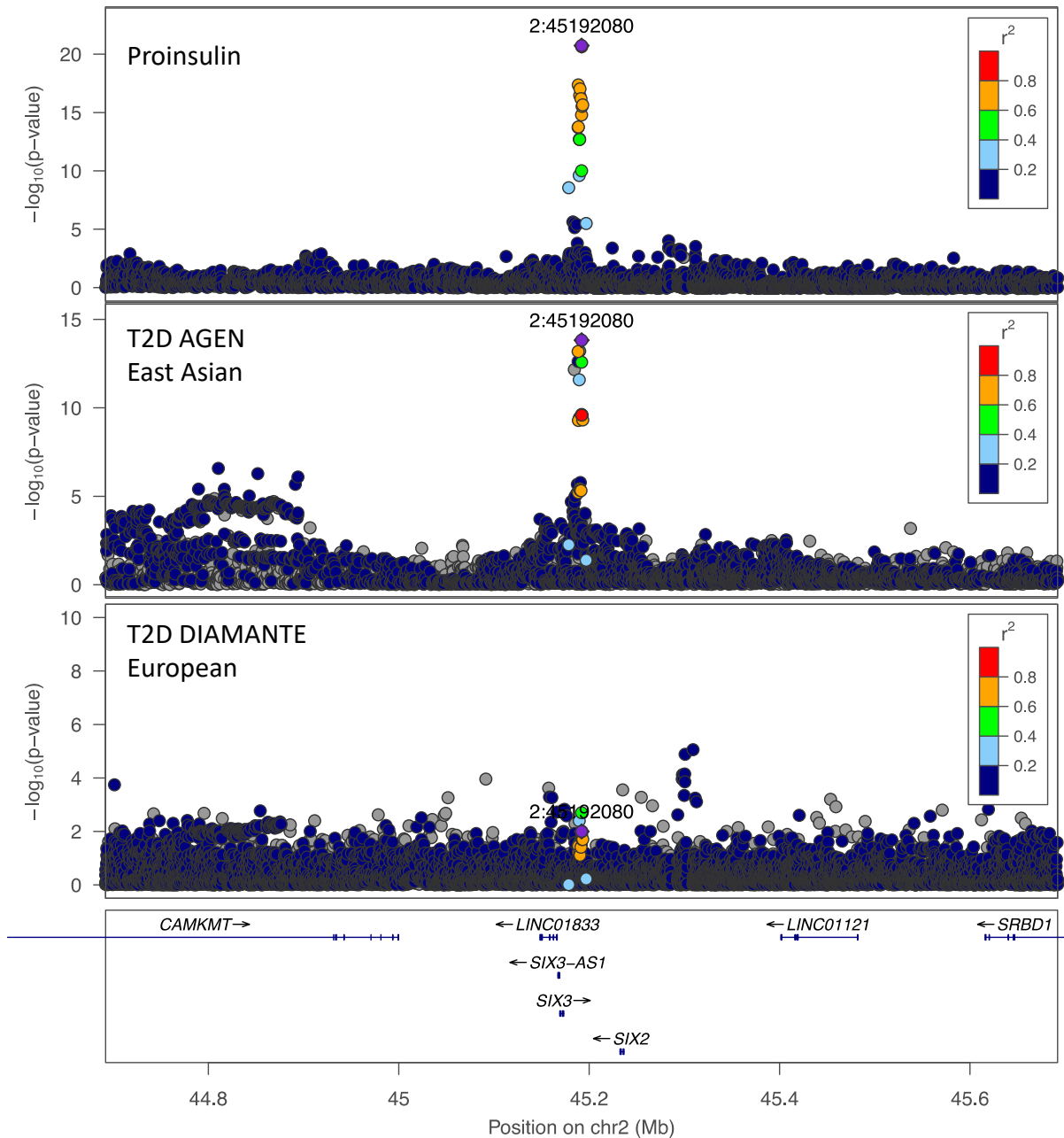


Figure S10: *SIX3* locus associations with proinsulin, T2D DIAMANTE European, and East Asian AGEN T2D. Among East Asians, the *SIX3* locus was associated with T2D; however, despite a common allele frequency in other ancestries, analyses of other ancestries have failed to identify an association between T2D and the *SIX3* locus. Findings in this proinsulin analysis indicates that the glycemic associations at *SIX3* are not specific to East Asians and predicts that the direction of effect on glucose levels and T2D in Europeans will be consistent between that shown in East Asians. Continued on next page.

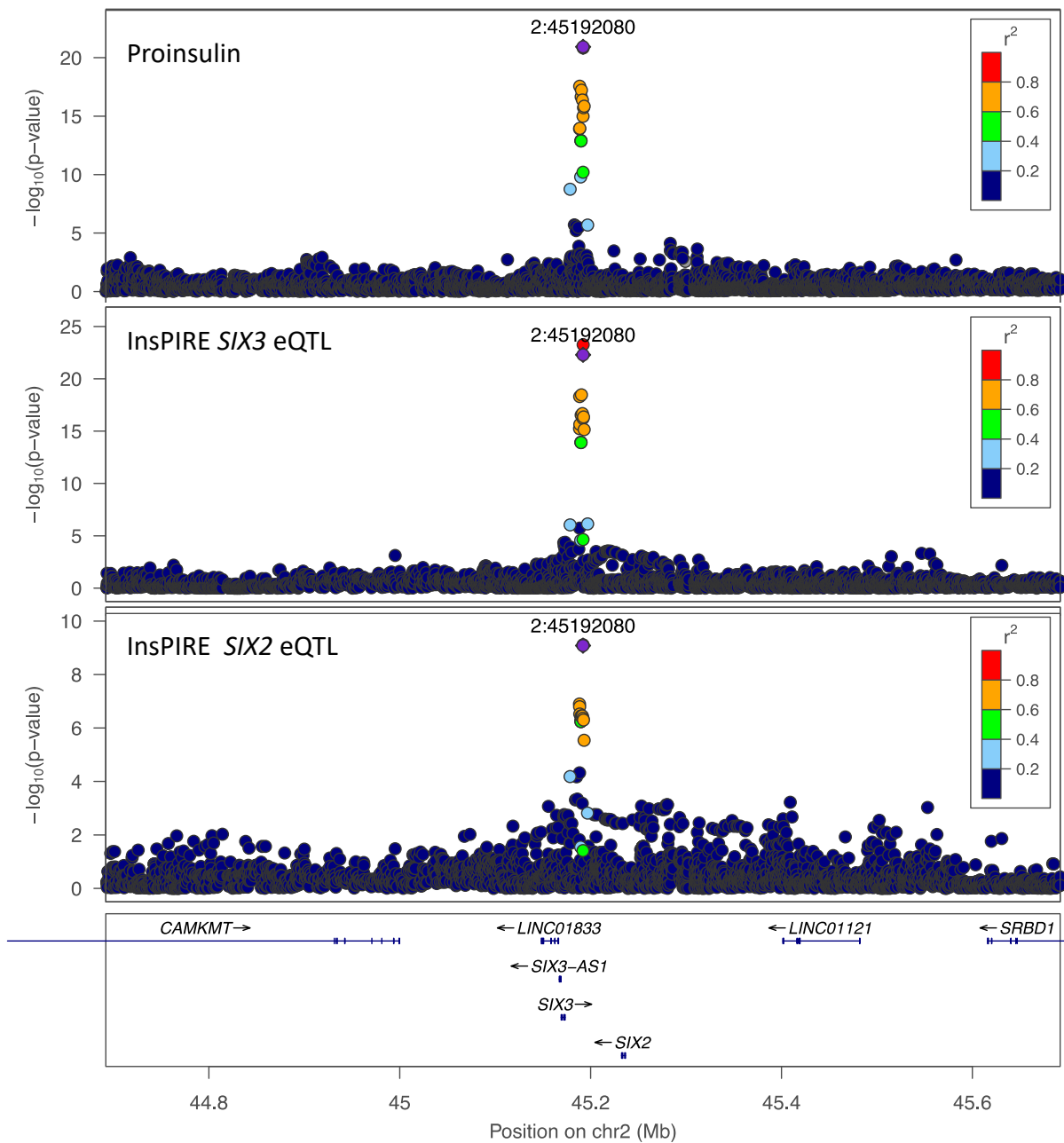


Figure S10 (continued): *SIX3* locus associated with proinsulin and InsPIRE eQTL for *SIX3* and *SIX2* expression levels.

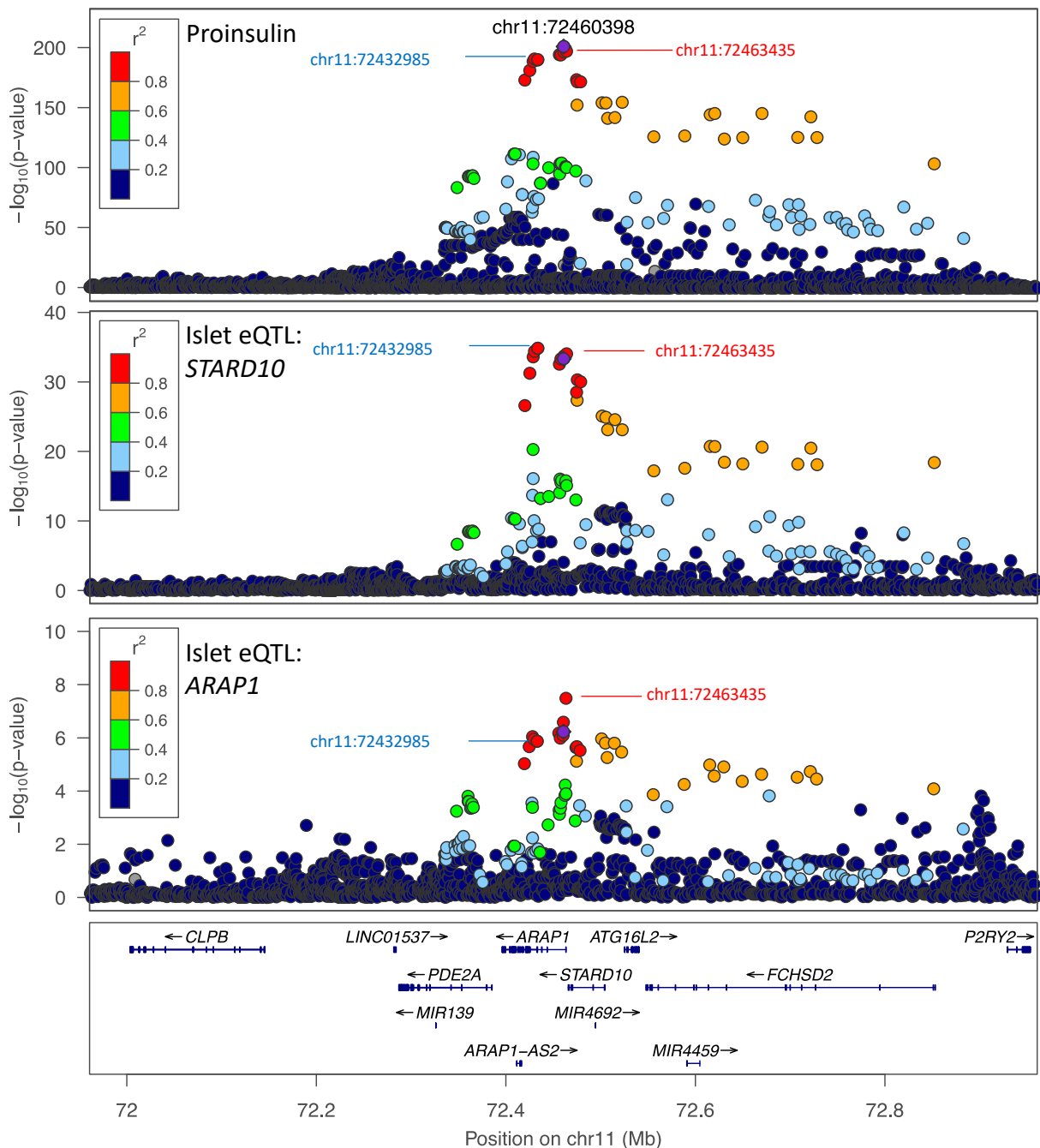


Figure S11: *STAR10* locus associations with proinsulin and InsPIRE islet *STAR10* and *ARAP1* expression. Although evidence of colocalization is stronger with *ARAP1* eQTL, the strength of association between the lead variant and *STAR10* expression is stronger. Plots are colored by the lead proinsulin variant (chr11:72460398, rs77464186), denoted by purple diamond. Variant labeled in blue (chr11:72432985) is the lead *STAR10* eQTL variant; variant labeled in red (chr11:72463435) is the lead *ARAP1* eQTL variant.