Data Supplement for Sanchez-Roige et al., Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorder Identification Test (AUDIT) in Two Population-Based Cohorts. Am J Psychiatry (doi: 10.1176/appi.ajp.2018.18040369)

CONTENTS

Supplementary Methods

UKB genotyping and quality control

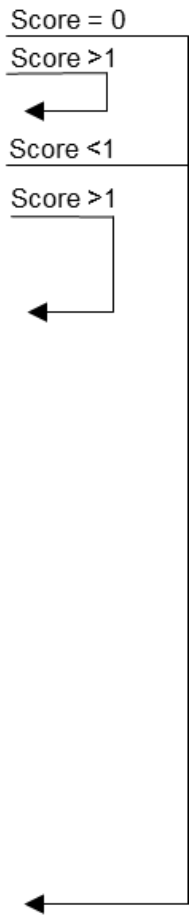
We removed 131,790 related individuals who were third degree relatives or closer (using a kinship coefficient > 0.044). From these related individuals we identified one individual from each group of relatives by creating a genomic relationship matrix and using a genetic-relatedness cut-off of 0.025 and added these back into the sample (N = 55,745). Variants were removed with a call rate < 98%, a minor allele frequency < 0.001 and those that deviated significantly from Hardy-Weinberg equilibrium ($p \leq 5 \times 10^{-6}$).

Gene-based association using transcriptomic data with S-PrediXcan

As input data, we included our meta-analysis GWAS summary statistics, transcriptome tissue data and covariance matrices of the SNPs within each gene model (based on HapMap SNP set; available to download at the PredictDB Data Repository) from 10 brain tissues: anterior cingulate cortex, caudate basal ganglia, cerebellar hemisphere, cerebellum, cortex, frontal cortex, hippocampus, hypothalamus, nucleus accumbens basal ganglia, and putamen basal ganglia. We used a transcriptome-wide significant threshold of $p < 1.07 \times 10^{-6}$, which is the Bonferroni corrected threshold when adjusting for all tissues and genes.

Figure S1. AUDIT gating logic

Item		Item Score
1	How often do you have a drink containing alcohol?	0-4
2	How many drinks containing alcohol do you have on a typical day when you are drinking?	0-4
3	How often do you have six or more drinks on one occasion?	
4-8	How often during the last year have you found that you were not able to stop drinking once you started? How often during the last year have you failed to do what is normally expected from you because of drinking? How often during the last year have you needed a first drink in the morning to get yourself going after a heavy session? How often in the last year have you had a feeling of remorse or guilt after drinking? How often during the last year have you been unable to remember what happened last night because of your drinking?	0-4
9-10	Have you or someone else been injured because of your drinking? Has a relative, friend, doctor, or other health care worker been concerned about your drinking or suggested you cut down?	0, 2, 4



Participants that answered item 1 of the AUDIT ("How often do you have a drink containing alcohol?") as 'Never' (8.1% of participants reported ‘never’ in response to item 1) proceeded to items 9 and 10. In order to complete items 4-8 participants had to score ≥1 for the sum of questions 2 and 3 (45.4% or respondents did not meet this threshold). AUDIT total score was created by taking the sum of items 1-10 for all participants, including those who endorsed currently never drinking alcohol (as they could still endorse past alcohol harm on items 9 and 10)

Figure S2. Distribution of AUDIT scores in the UK Biobank (UKB) cohort before and after log10 transformation: AUDIT total score, AUDIT-C and AUDIT-P dimensions from UKB research participants (N = 121,604)

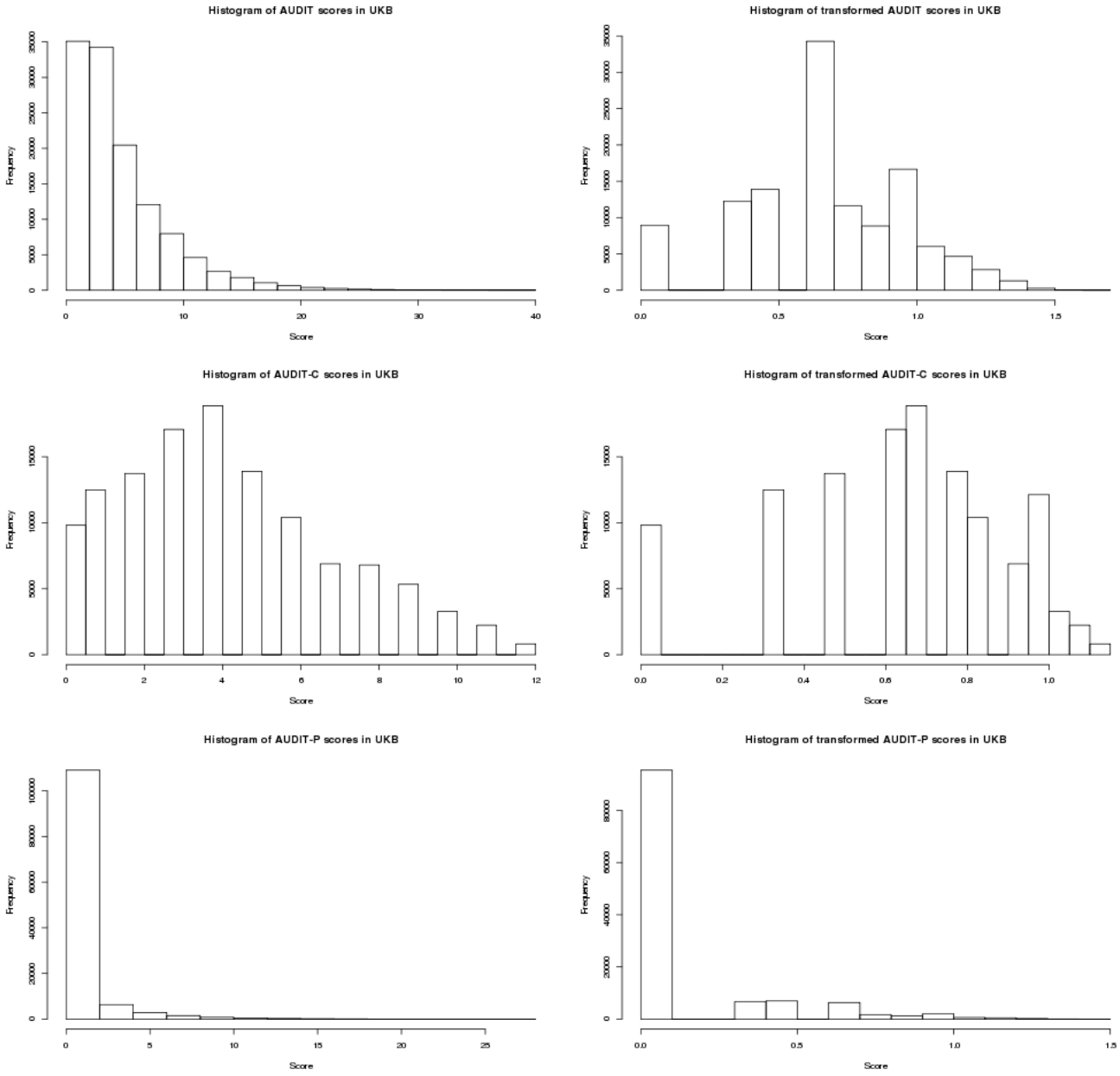


Figure S3. Distribution of AUDIT score in males (blue bars; N=53,215) and females (red bars; 68,393) in the UK Biobank cohort.

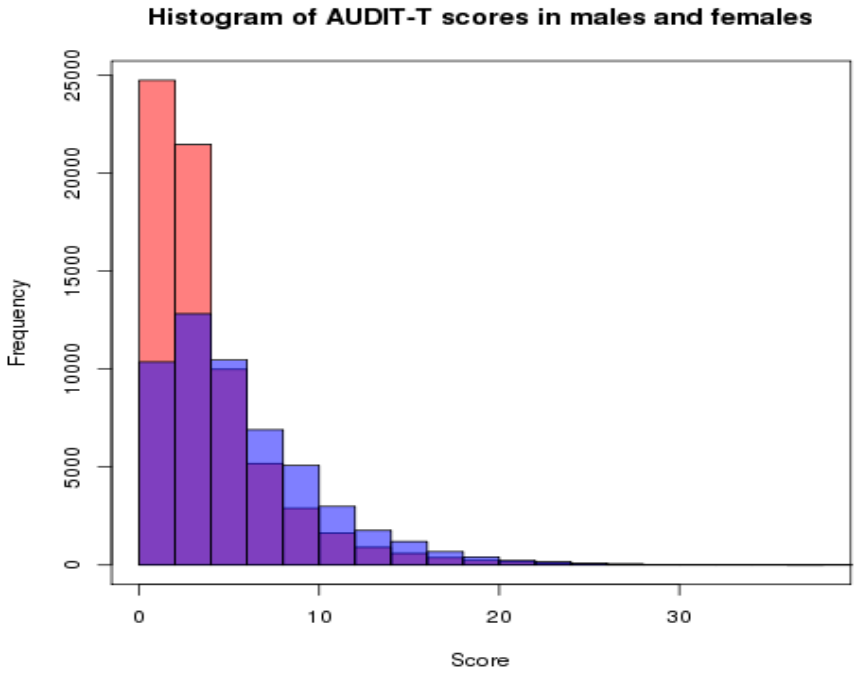


Figure S4. Result of genome-wide association analyses for AUDIT-C and AUDIT-P dimensions. Manhattan and QQ plots of GWAS results showing the strongest associations between the 22 autosomes and AUDIT. AUDIT-C associations are shown above the mid-line and AUDIT-P below.

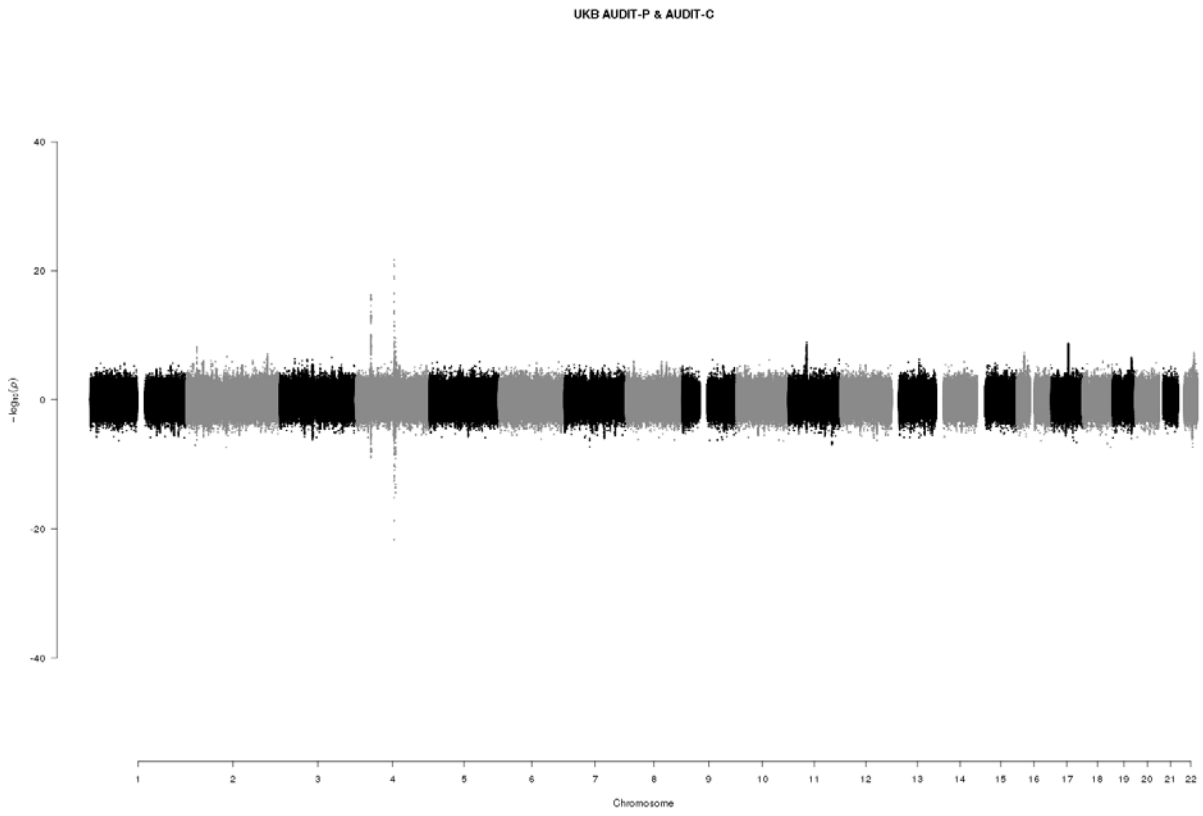


Figure S5. Regional association plot of rs1260326 and LD with SNPs in the surrounding region. LD calculations taken from UKB genotype data release 2 implemented in FUMA. Grey SNPs have $r^2 < 0.1$ with lead SNP.

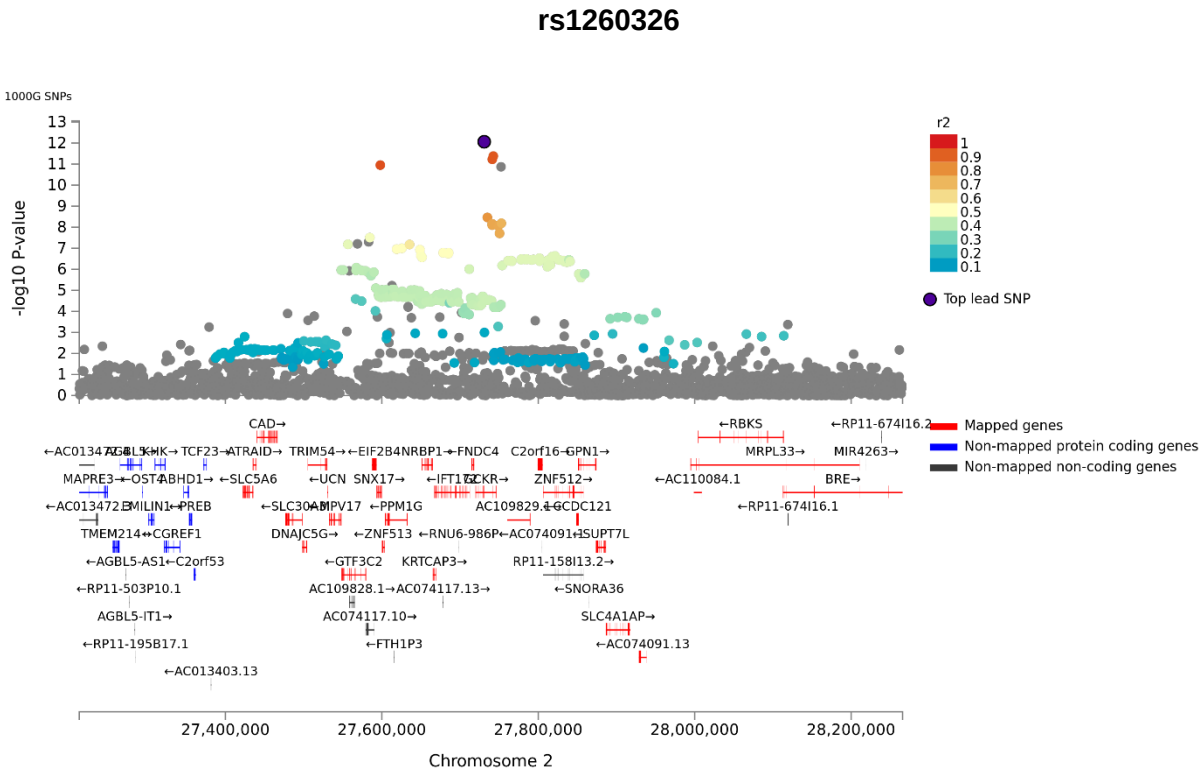


Figure S6. Regional association plot of rs4953148 and LD with SNPs in the surrounding region. LD calculations taken from UKB genotype data release 2 implemented in FUMA. Grey SNPs have $r^2 < 0.1$ with lead SNP. RP11_89/K21.1 are alternative names for the LINC01833 locus.

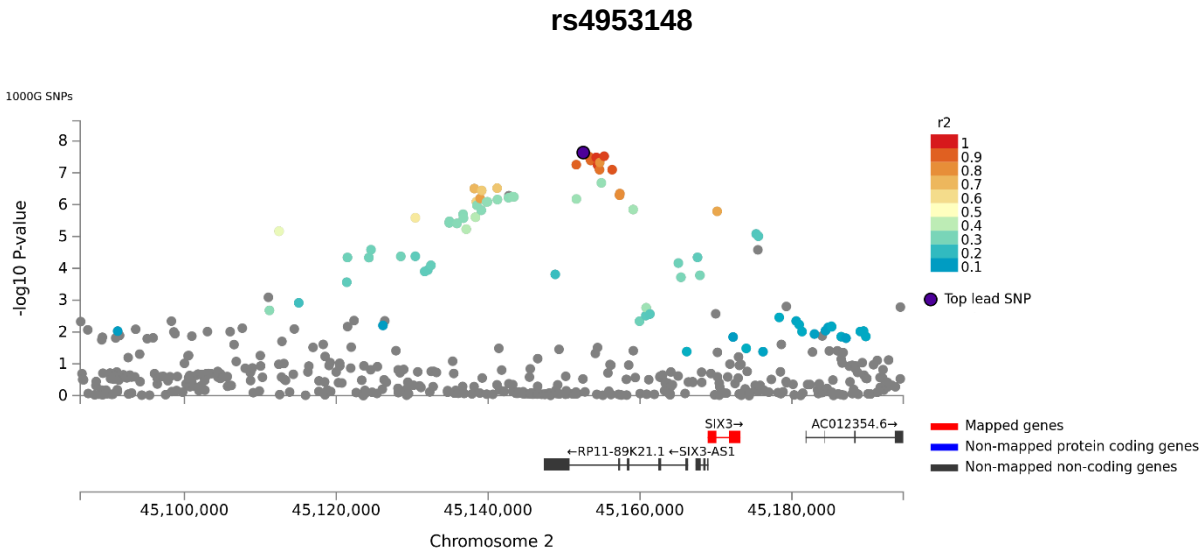


Figure S7. Regional association plot of rs1920650 and LD with SNPs in the surrounding region. LD calculations taken from UKB genotype data. LD calculations taken from UKB genotype data release 2 implemented in FUMA. Grey SNPs have $r^2 < 0.1$ with lead SNP.

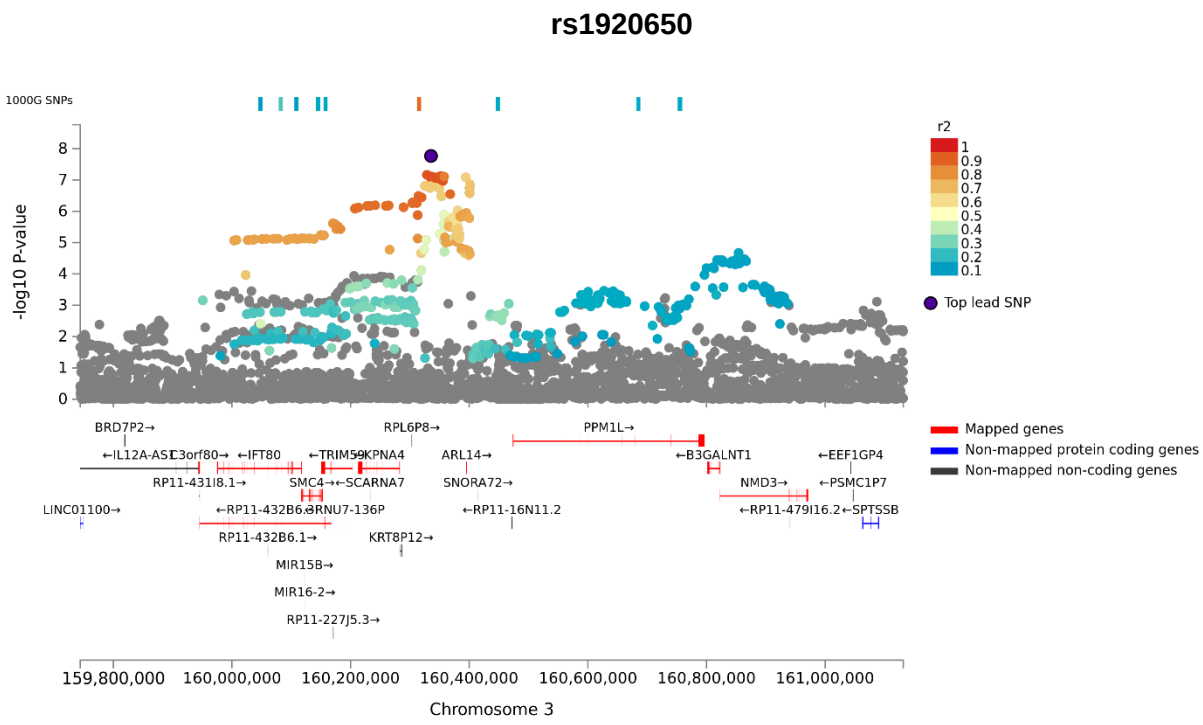


Figure S8. Regional association plot of rs11940694 (top lead SNP) and rs4975012 (lead SNP) and LD with SNPs in the 1000kb surrounding region. LD calculations taken from UKB genotype data release 2 implemented in FUMA. Grey SNPs have $r^2 < 0.1$ with lead SNP. Each SNP colour coded according to r^2 with closest lead SNP.

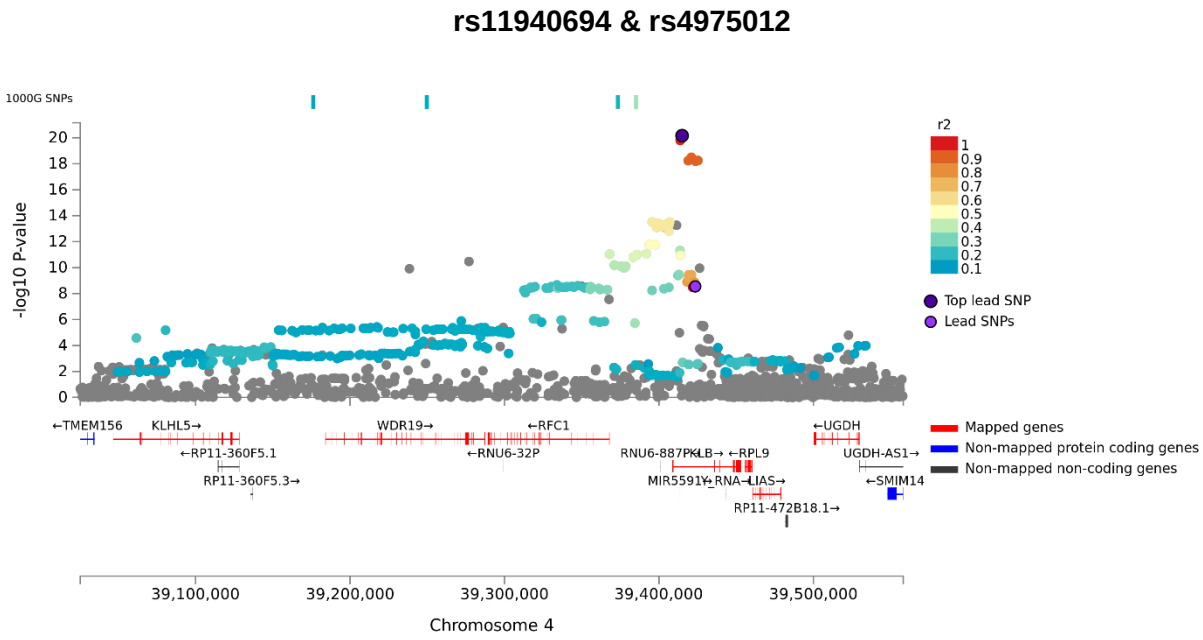


Figure S9. Regional association plot of rs146788033 (lead SNP), rs11733695 (top lead SNP), rs3114045 (lead SNP), rs188514326 (lead SNP) and rs13135092 (lead SNP) and LD with SNPs in the surrounding region. LD calculations taken from UKB genotype data release 2 implemented in FUMA. Grey SNPs have $r^2 < 0.1$ with lead SNP. Each SNP colour coded according to r^2 with closest lead SNP.

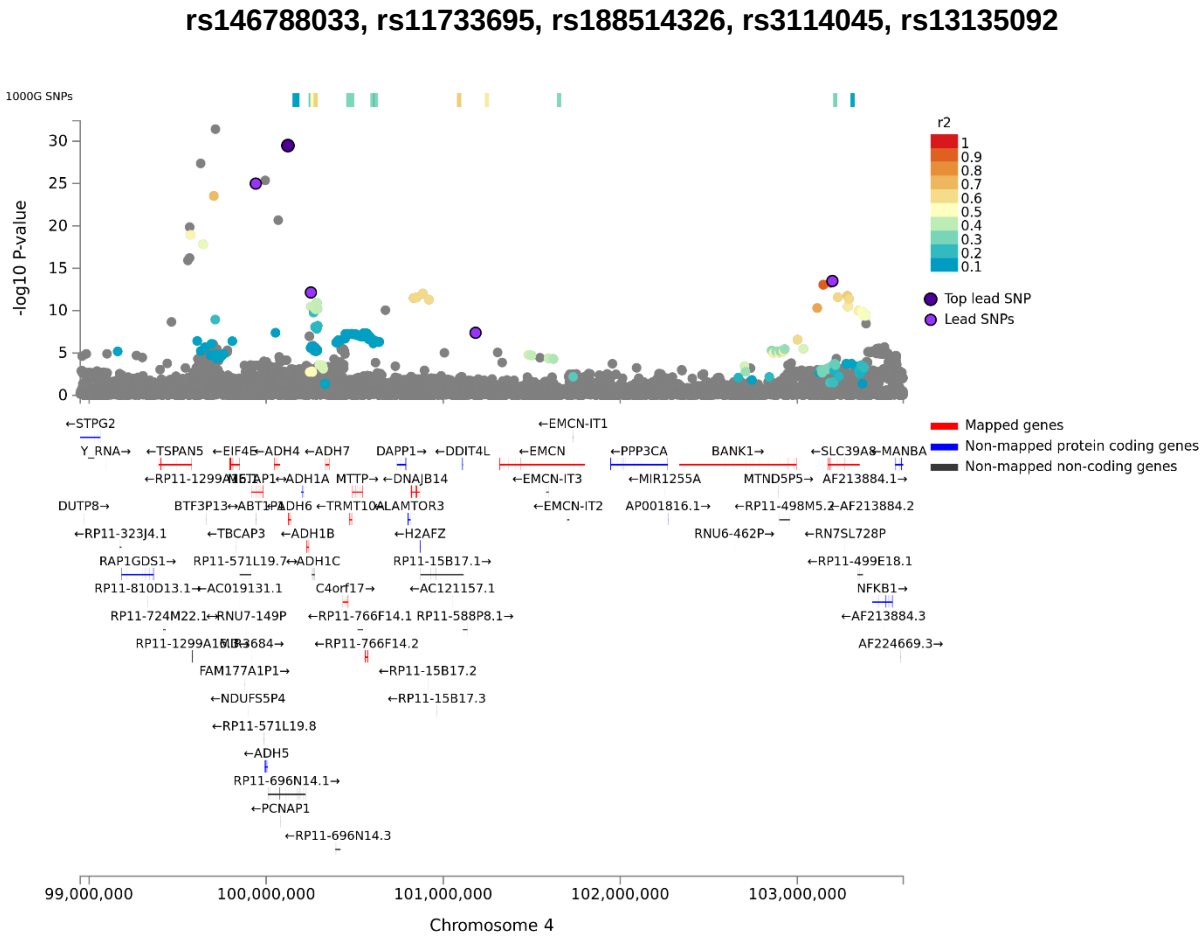


Figure S10. Regional association plot of rs35040843 and LD with SNPs in the surrounding region. LD calculations taken from UKB genotype data release 2 implemented in FUMA. Grey SNPs have $r^2 < 0.1$ with lead SNP. Each SNP colour coded according to r^2 with closest lead SNP.

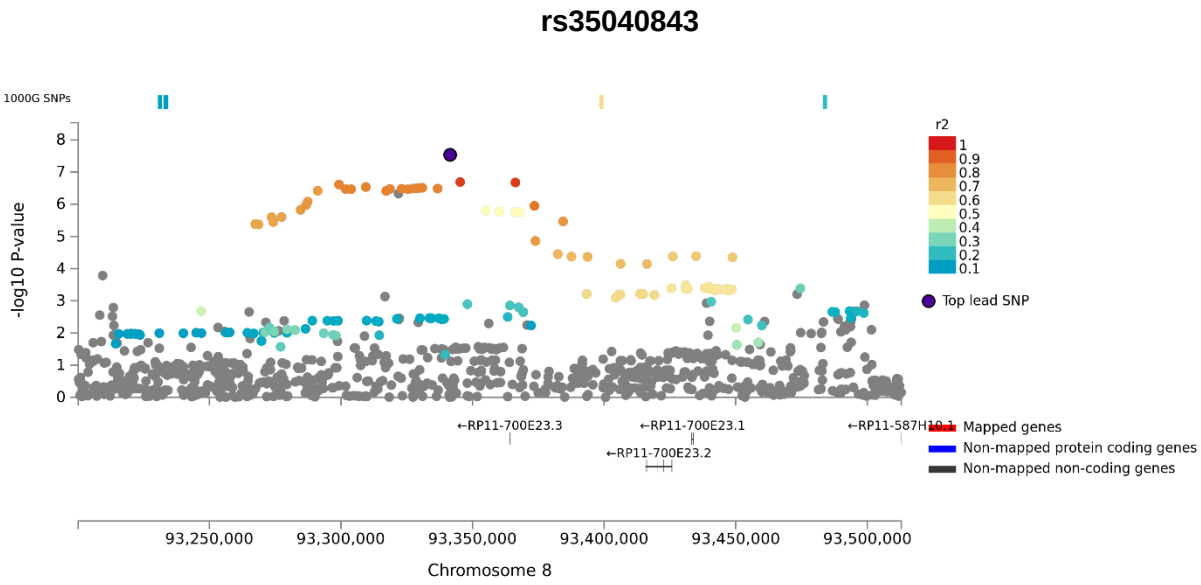


Figure S11. Regional association plot of rs7078436 and LD with SNPs in the surrounding region. LD calculations taken from UKB genotype data release 2 implemented in FUMA. Grey SNPs have $r^2 < 0.1$ with lead SNP. Each SNP colour coded according to r^2 with closest lead SNP.

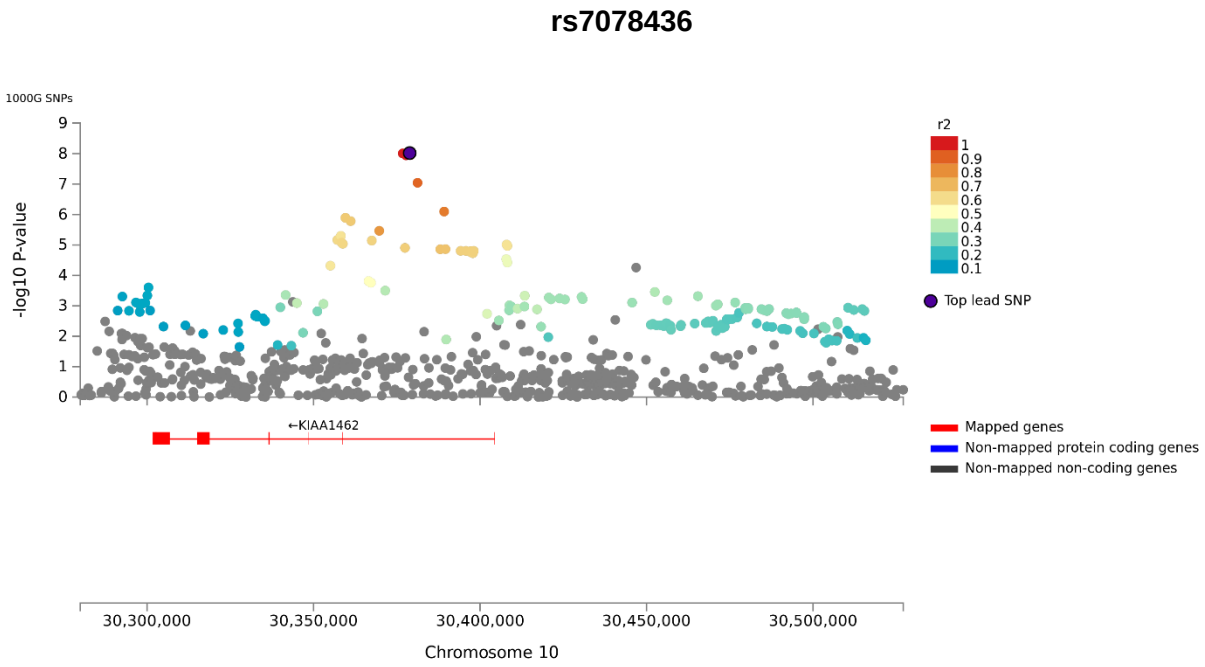


Figure S12. Regional association plot of rs2293576 and LD with SNPs in the surrounding region. LD calculations taken from UKB genotype data release 2 implemented in FUMA. Grey SNPs have $r^2 < 0.1$ with lead SNP. Each SNP colour coded according to r^2 with closest lead SNP.

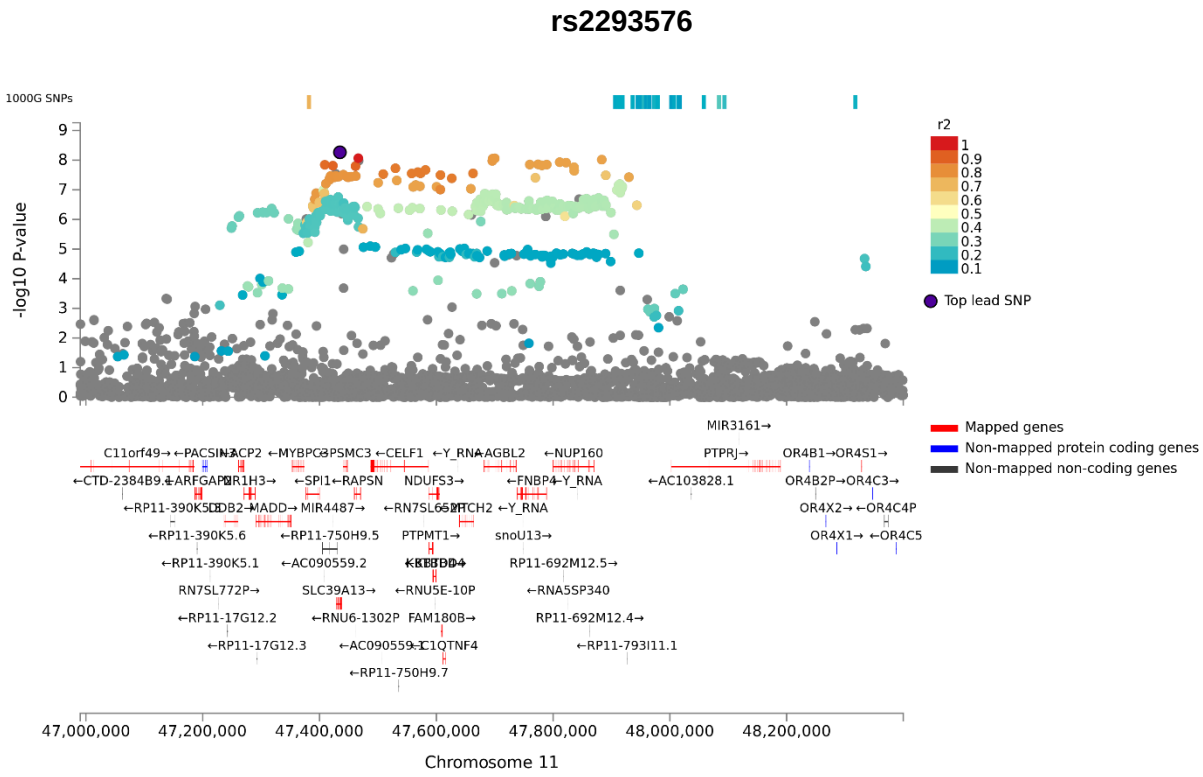


Figure S13. Regional association plot of rs62062288 and LD with SNPs in the surrounding region. LD calculations taken from UKB genotype data release 2 implemented in FUMA. Grey SNPs have $r^2 < 0.1$ with lead SNP. Each SNP colour coded according to r^2 with closest lead SNP.

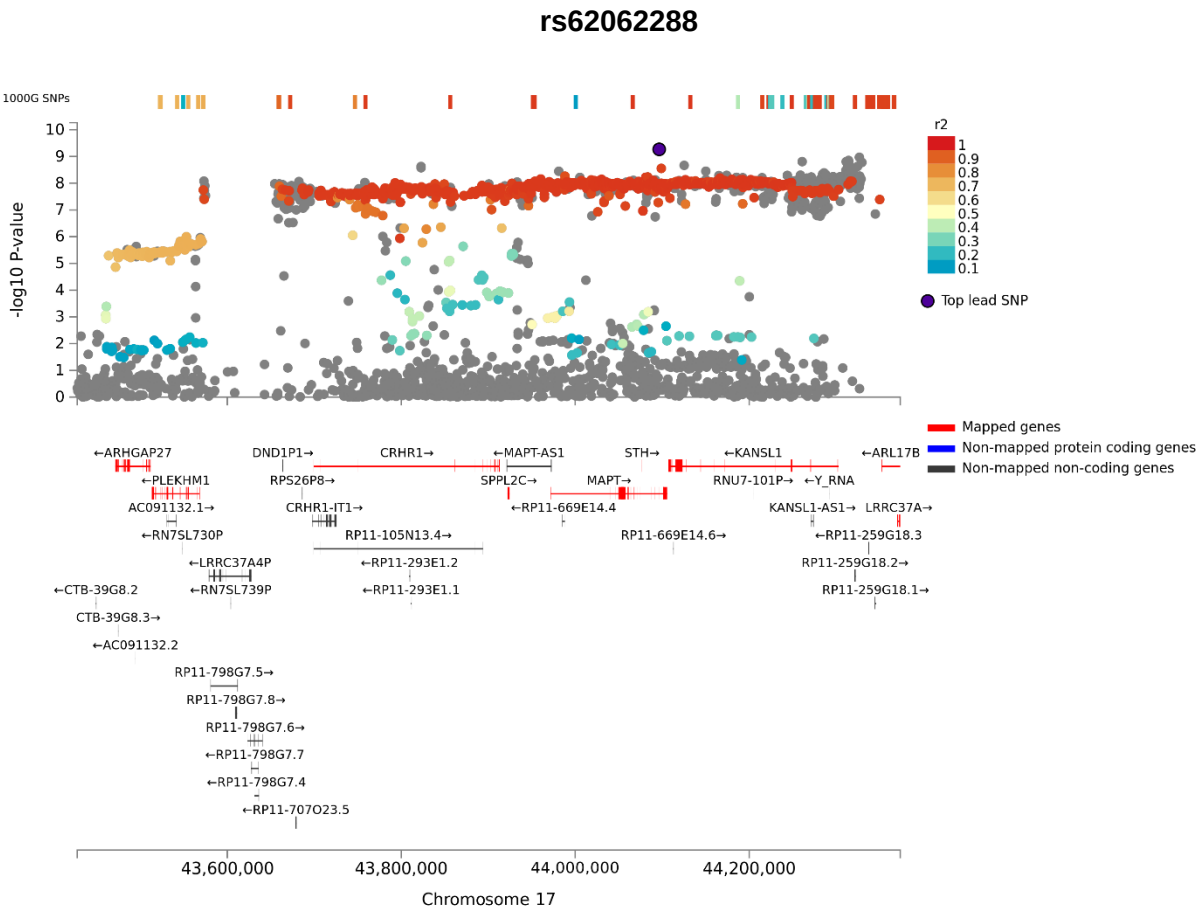


Figure S14. Regional association plot of rs492602 and LD with SNPs in the surrounding region. LD calculations taken from UKB genotype data release 2 implemented in FUMA. Grey SNPs have $r^2 < 0.1$ with lead SNP. Each SNP colour coded according to r^2 with closest lead SNP.

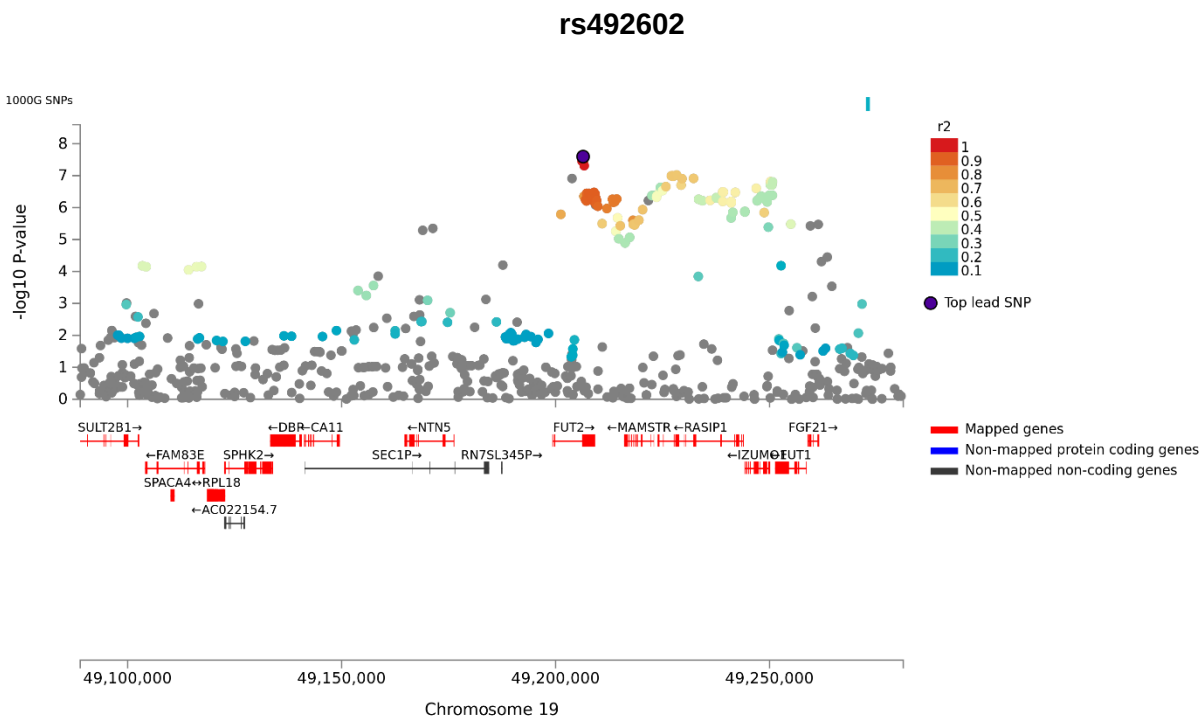


Figure S15. Results of gene-based association analyses for the GWAS meta-analysis of AUDIT total scores. Line denotes genome-wide significance ($p < 2.7 \times 10^{-6}$). LD calculations taken from UKB genotype data.

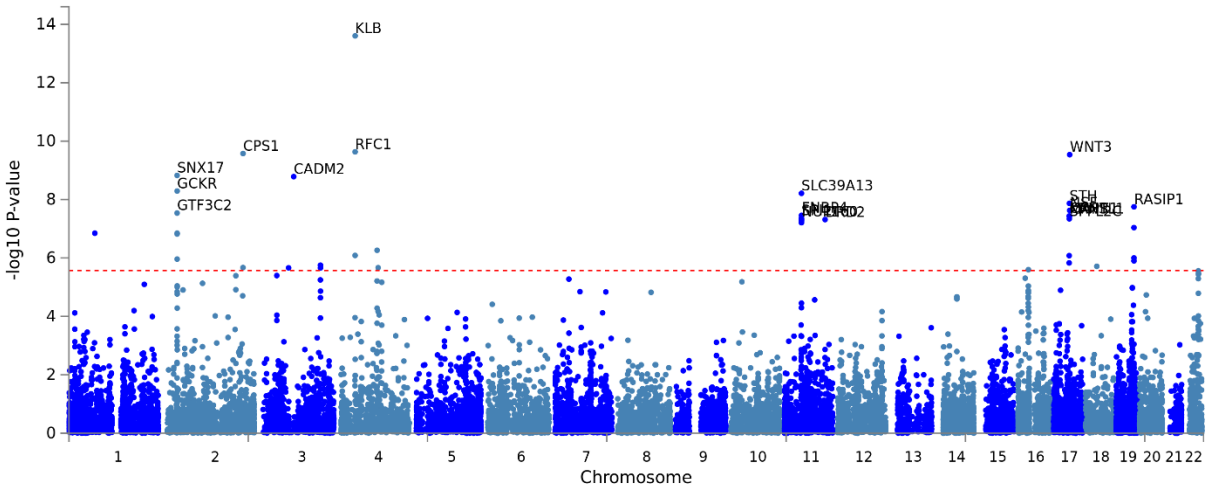


Figure S16. Results of gene-based association analyses for the GWAS of AUDIT-C. Line denotes genome-wide significance (p [0.05/18775] = 2.7 x 10^{-6})

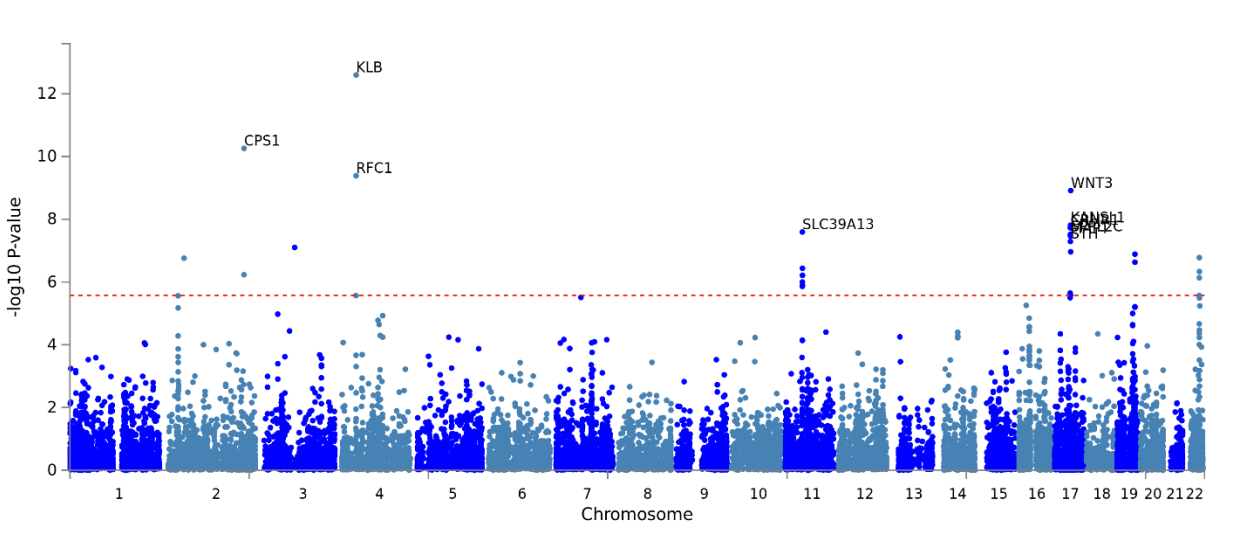


Figure S17. Results of gene-based association analyses for the GWAS of AUDIT-P. Line denotes genome-wide significance (p [0.05/18775] = 2.7 x 10^{-6})

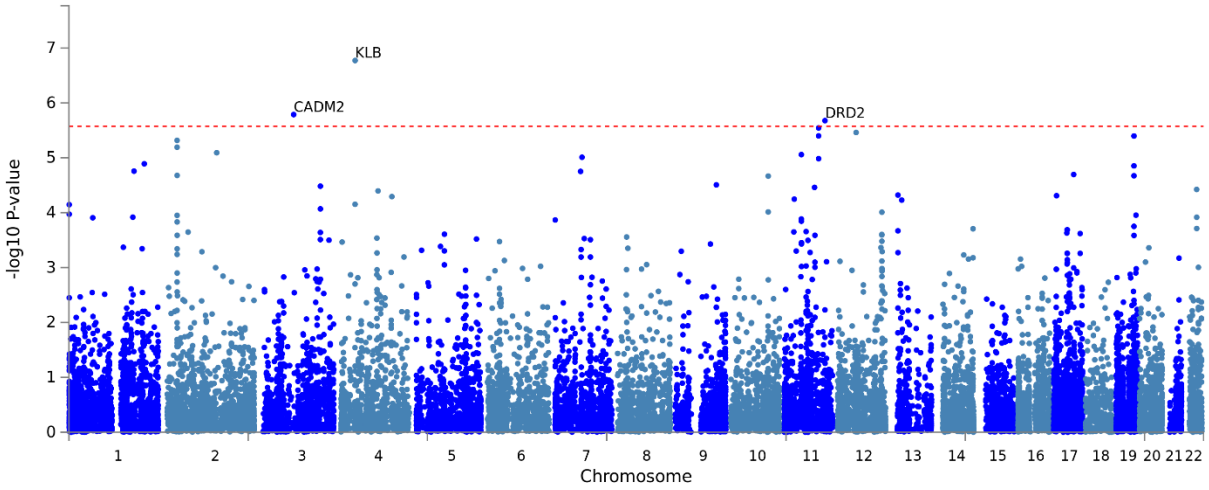


Figure S18. Overlap of MAGMA gene-based association analyses for the three AUDIT dimensions