Supplementary Data 1

Genomic Characterization of Posttraumatic Stress Disorder and its Symptom Subdomains in the Million Veteran Program

Fine mapping of risk loci for PCL-Total and EHR-binary

Contents

Figure 1: Summary of risk loci for Total PCL (EUR population) 2

chr2:22430795-22546631 3

chr3:18611283-18824298 4

chr3:49734229-50176259 5

chr5:107346393-107769562 6

chr5:140222641-140346468 7

chr7:1847623-2110850 8

chr8:143309504-143340566 9

chr9:140241209-140278206 10

chr11:133531999-133561399 11

chr17:42785473-44865603 12

chr18:53190188-53393895 13

chr19:18669987-18730695 14

Figure 2: Summary of risk loci for EHR-binary 15

chr7:1847623-1892529 (EUR) 16

chr7:70217631-70234438 (EUR) 17

chr11:28591168-28710118 (EUR) 18

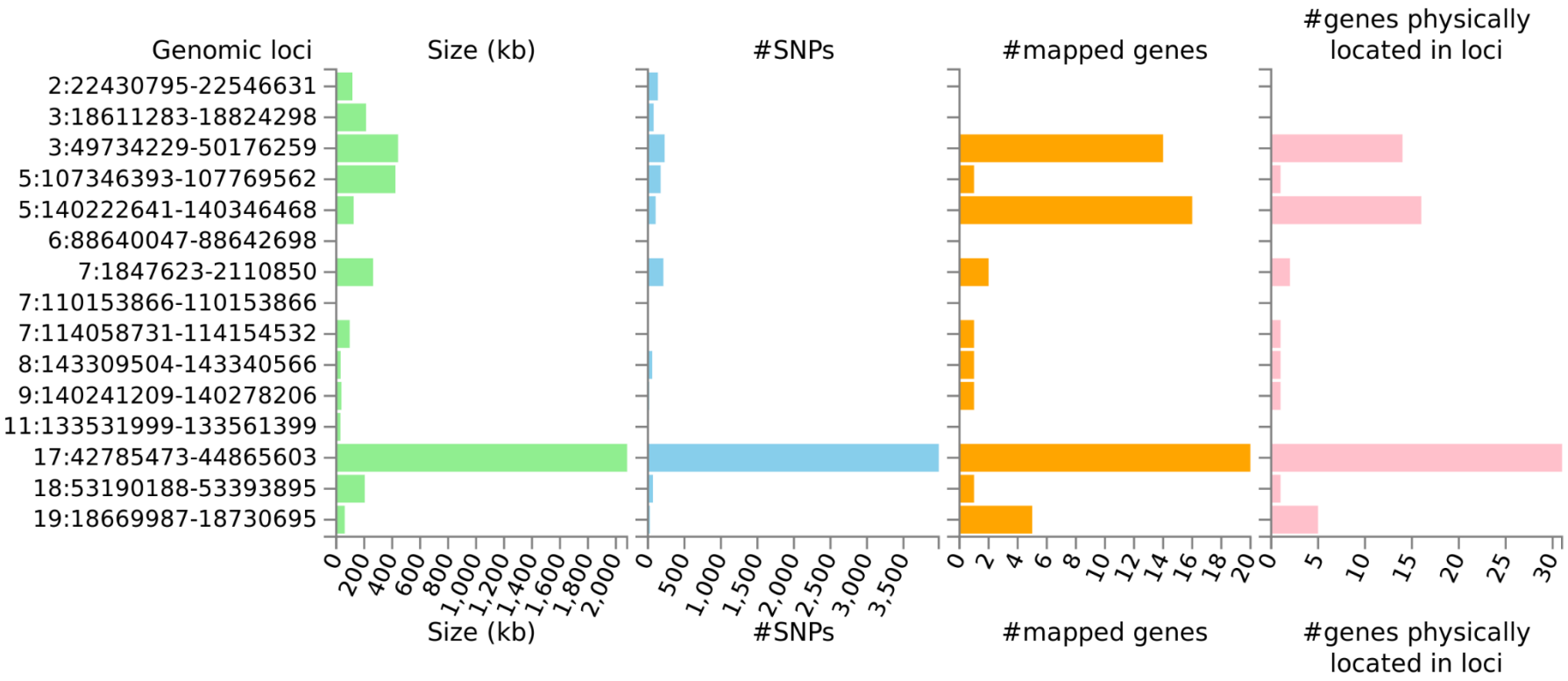
chr3:1256096-1261532 (AFR) 19

chr20:6724577-6724577 (AFR) 20

LEGEND NOTE:

Supplementary Data 1: ALL Figures [top panel]
The top panel shows regional association based on regression analysis of SNPs and PCL phenotype in the locus. The y-axis is -log10 of two-sided p-values for each SNP, wherein the p-value lower than $5x10^{-8}$ represents Bonferroni-threshold for multiple testing (0.05/k SNPs = $p < 5x10^{-8}$)

Figure 1: Summary of risk loci for Total PCL (EUR population)



chr2:22430795-22546631

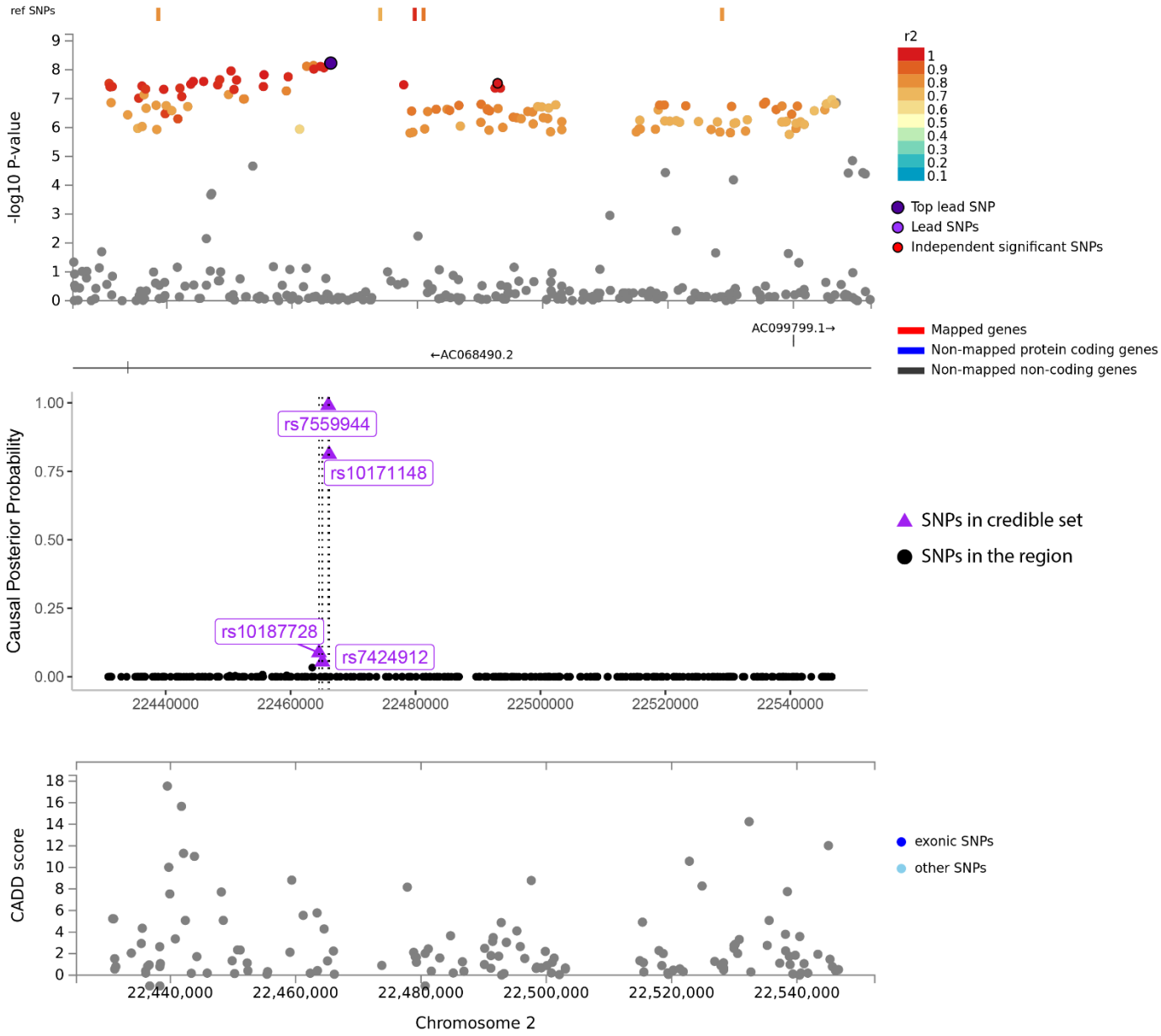


Figure 2: Fine mapping of the risk loci identified from FUMA.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The middle panel finemaps all the SNPs in the region using CAVIAR where SNPs labelled and marked as triangles were identified as credible set. The y-axis is posterior probability of SNPs.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.

chr3:18611283-18824298

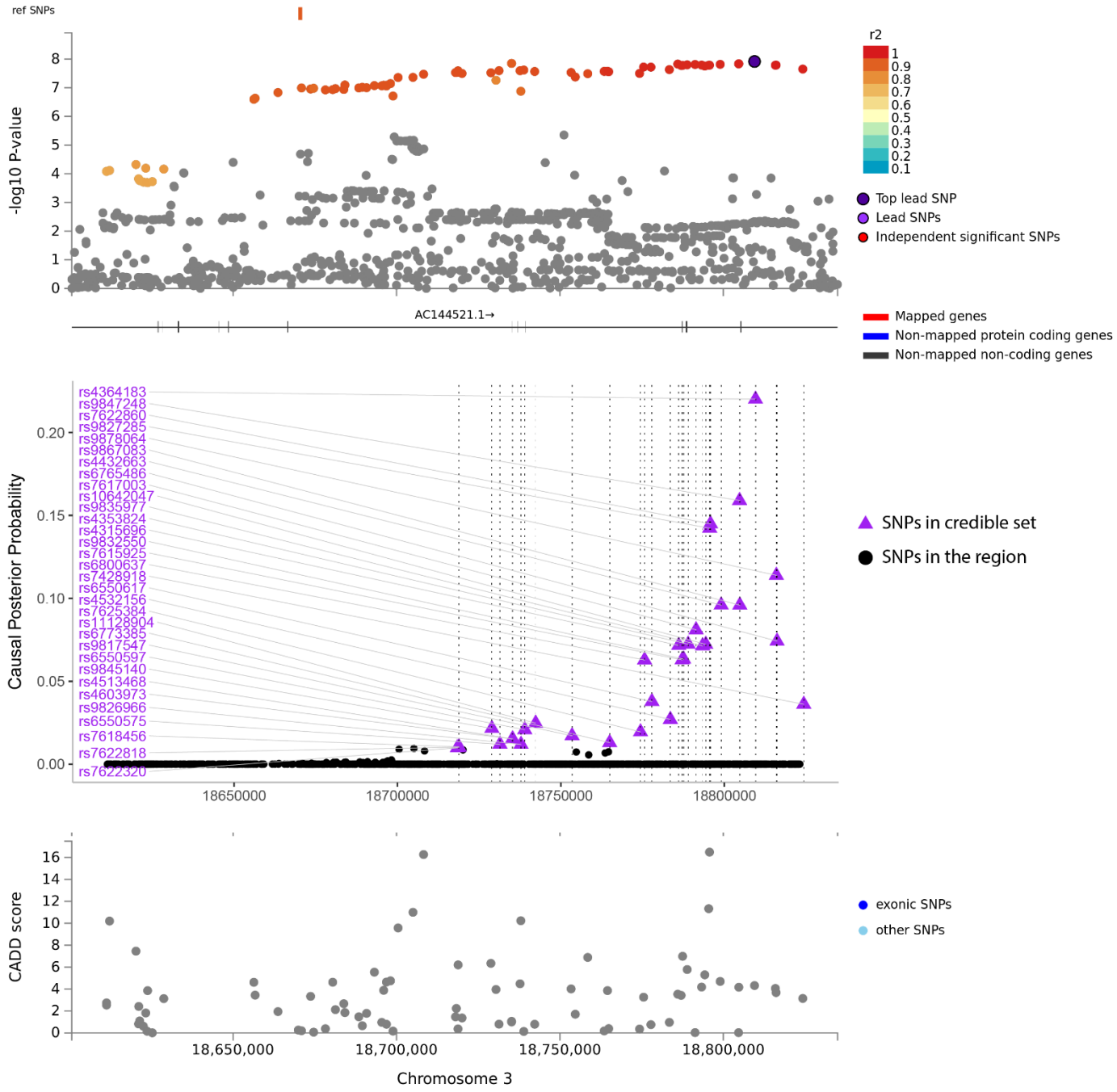


Figure 3: Fine mapping of the risk loci identified from FUMA.

The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The middle panel finemaps all the SNPs in the region using CAVIAR where SNPs labelled and marked as triangles were identified as credible set. The y-axis is posterior probability of SNPs.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.

chr3:49734229-50176259

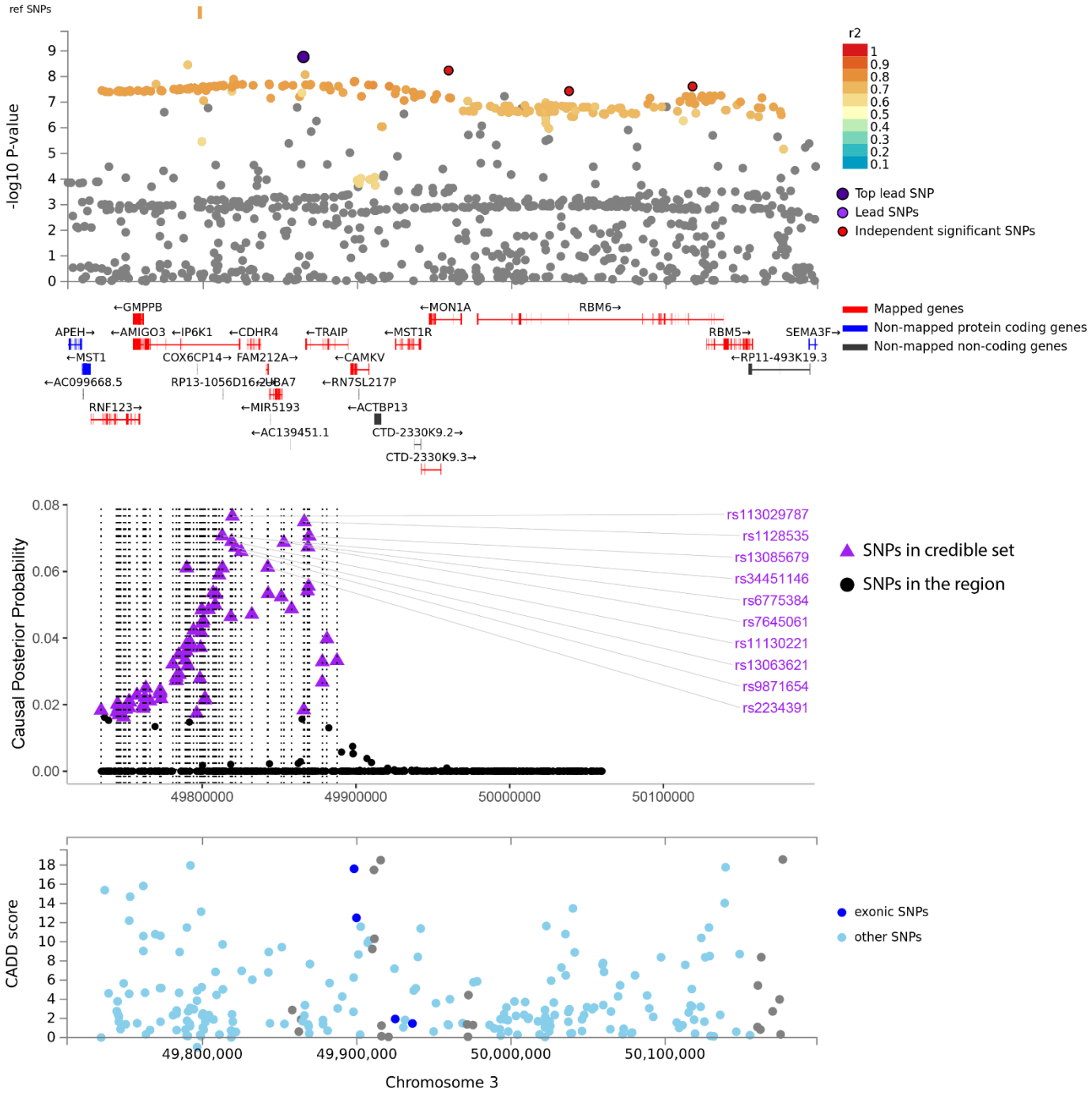


Figure 4: Fine mapping of the risk loci identified from FUMA.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The middle panel finemaps all the SNPs in the region using CAVIAR where SNPs labelled and marked as triangles were identified as credible set. The y-axis is posterior probability of SNPs.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.

chr5:107346393-107769562

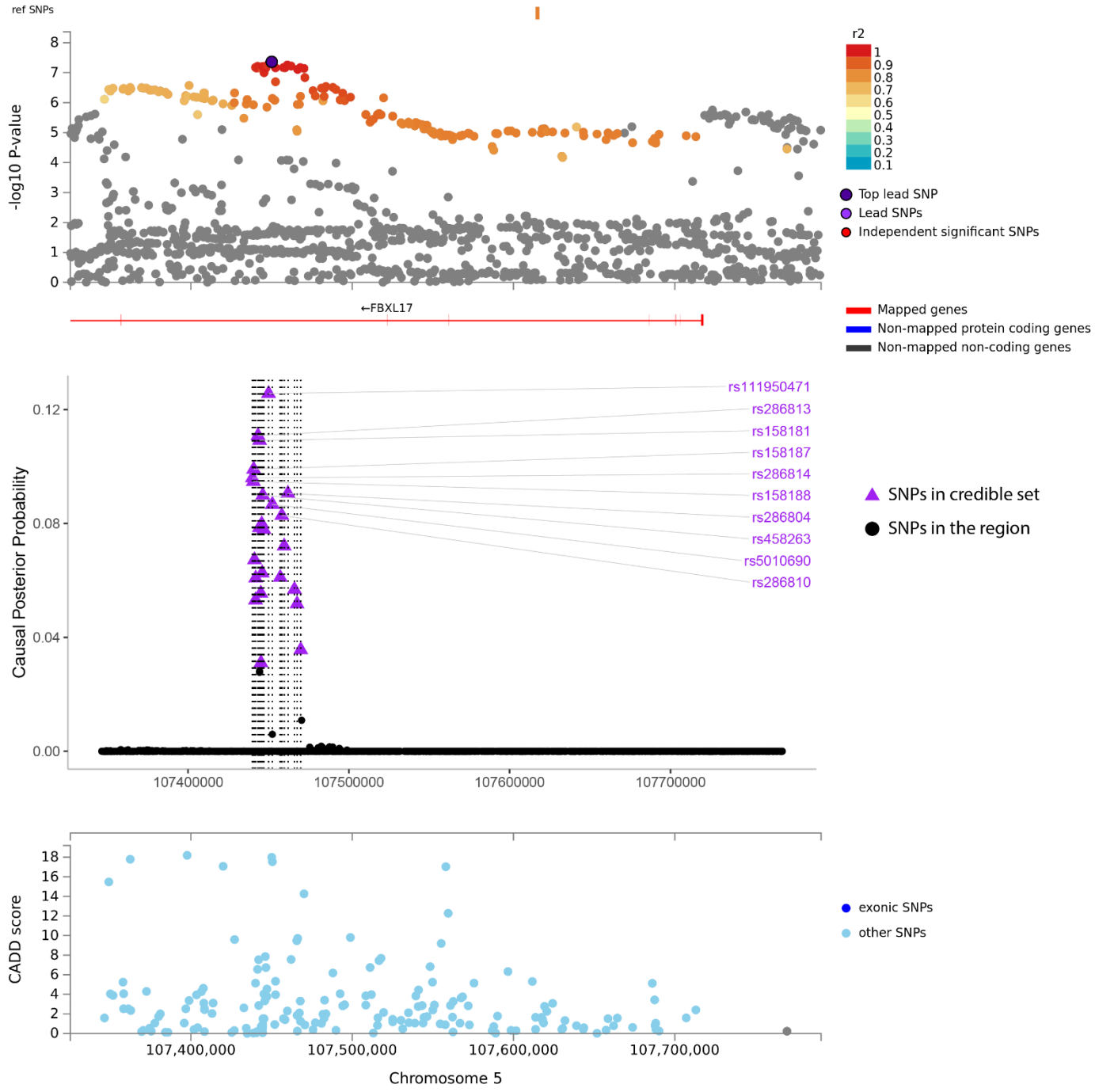


Figure 5: Fine mapping of the risk loci identified from FUMA.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The middle panel finemaps all the SNPs in the region using CAVIAR where SNPs labelled and marked as triangles were identified as credible set. The y-axis is posterior probability of SNPs.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.

chr5:140222641-140346468

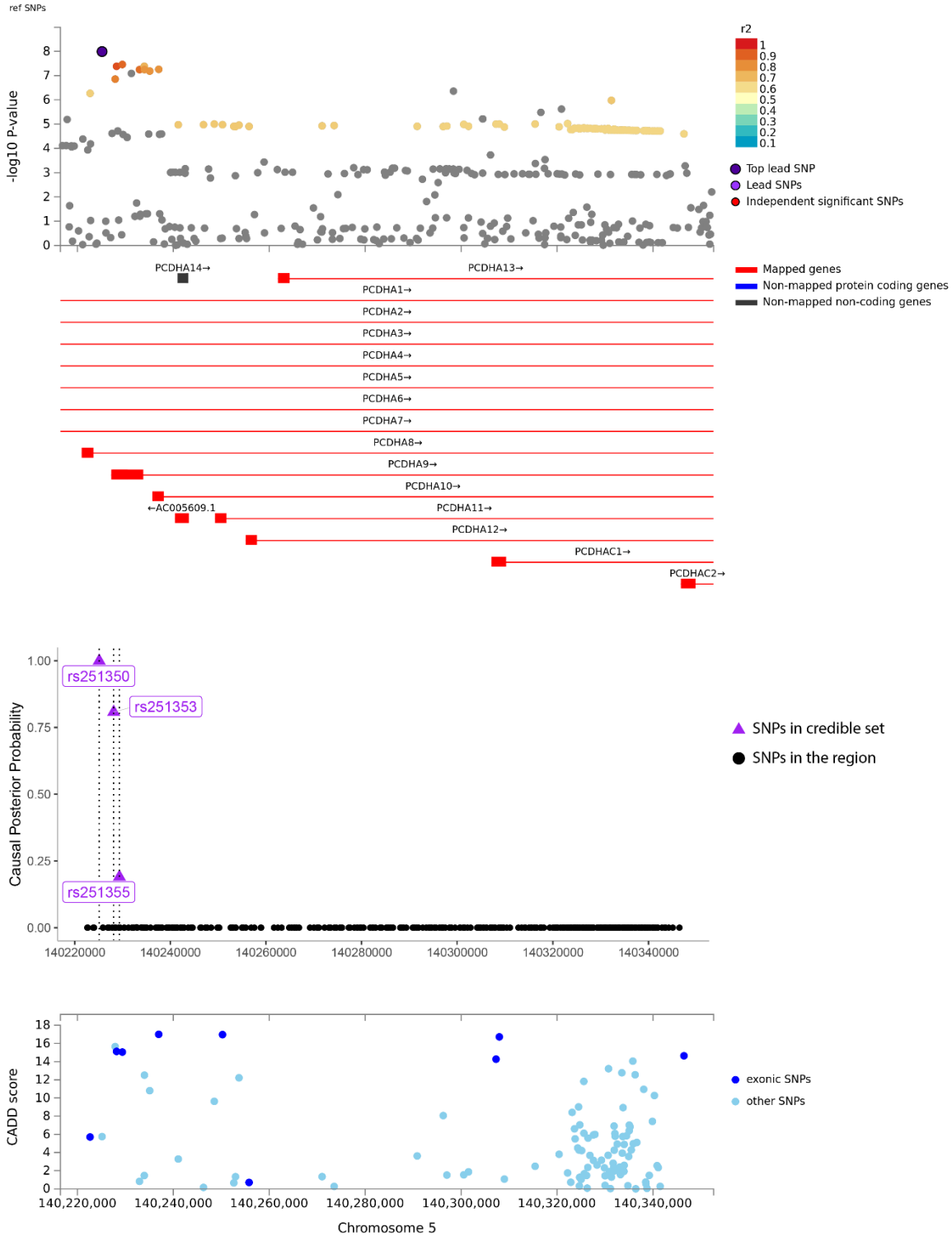


Figure 6: Fine mapping of the risk loci identified from FUMA.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The middle panel finemaps all the SNPs in the region using CAVIAR where SNPs labelled and marked as triangles were identified as credible set. The y-axis is posterior probability of SNPs.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.

chr7:1847623-2110850

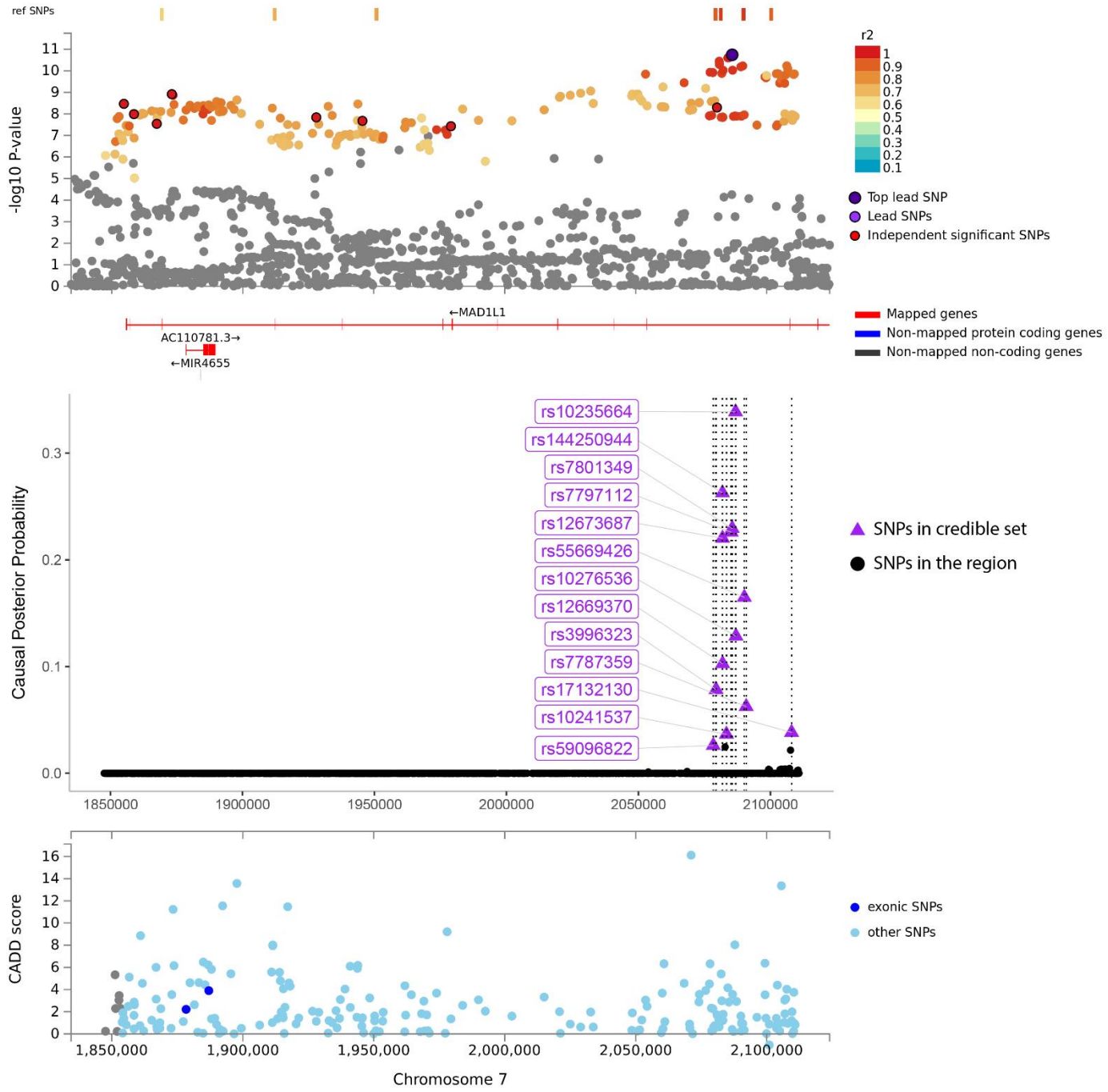


Figure 7: Fine mapping of the risk loci identified from FUMA.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The middle panel finemaps all the SNPs in the region using CAVIAR where SNPs labelled and marked as triangles were identified as credible set. The y-axis is posterior probability of SNPs.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.

chr8:143309504-143340566

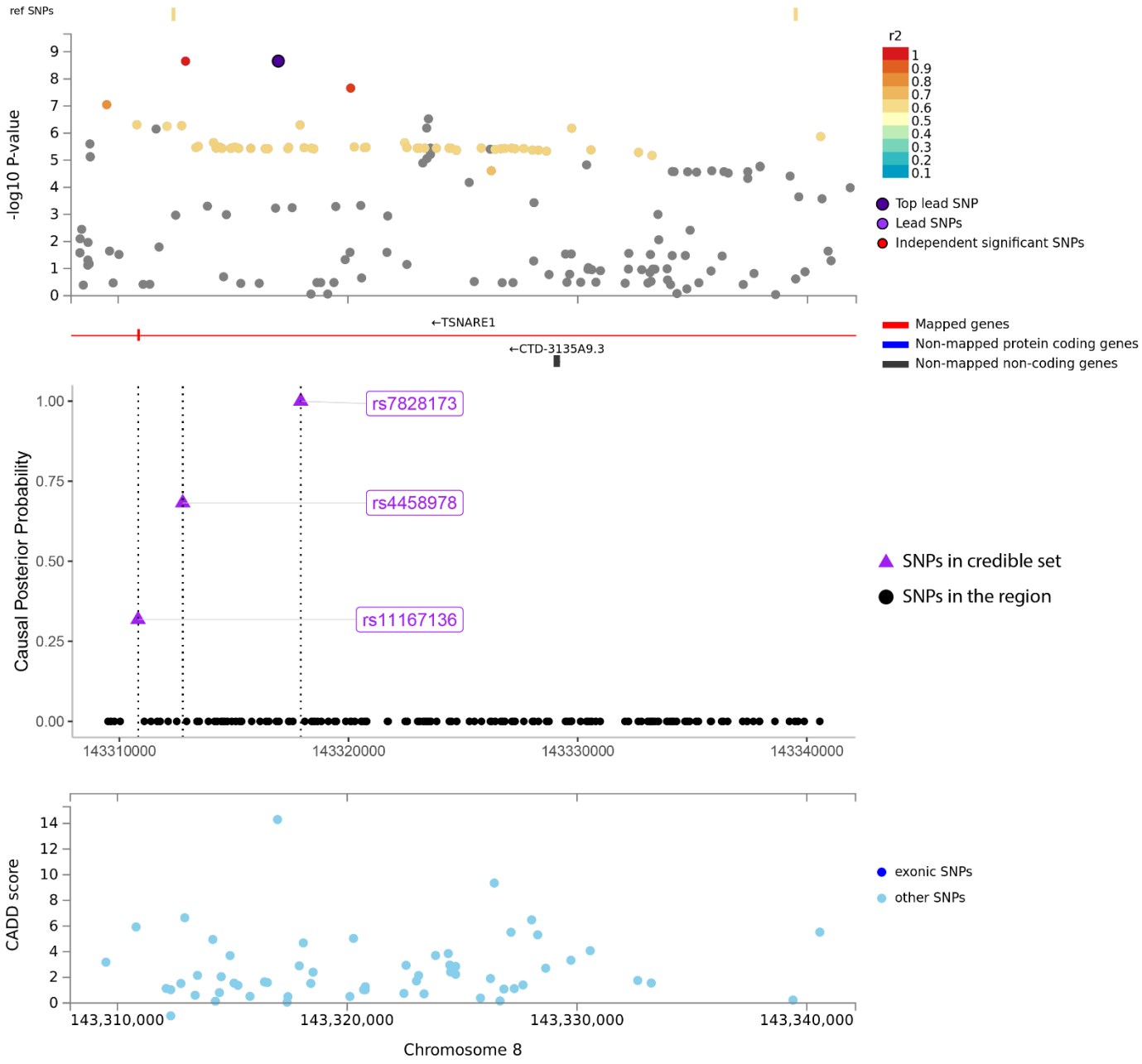


Figure 8: Fine mapping of the risk loci identified from FUMA.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The middle panel finemaps all the SNPs in the region using CAVIAR where SNPs labelled and marked as triangles were identified as credible set. The y-axis is posterior probability of SNPs.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.

chr9:140241209-140278206

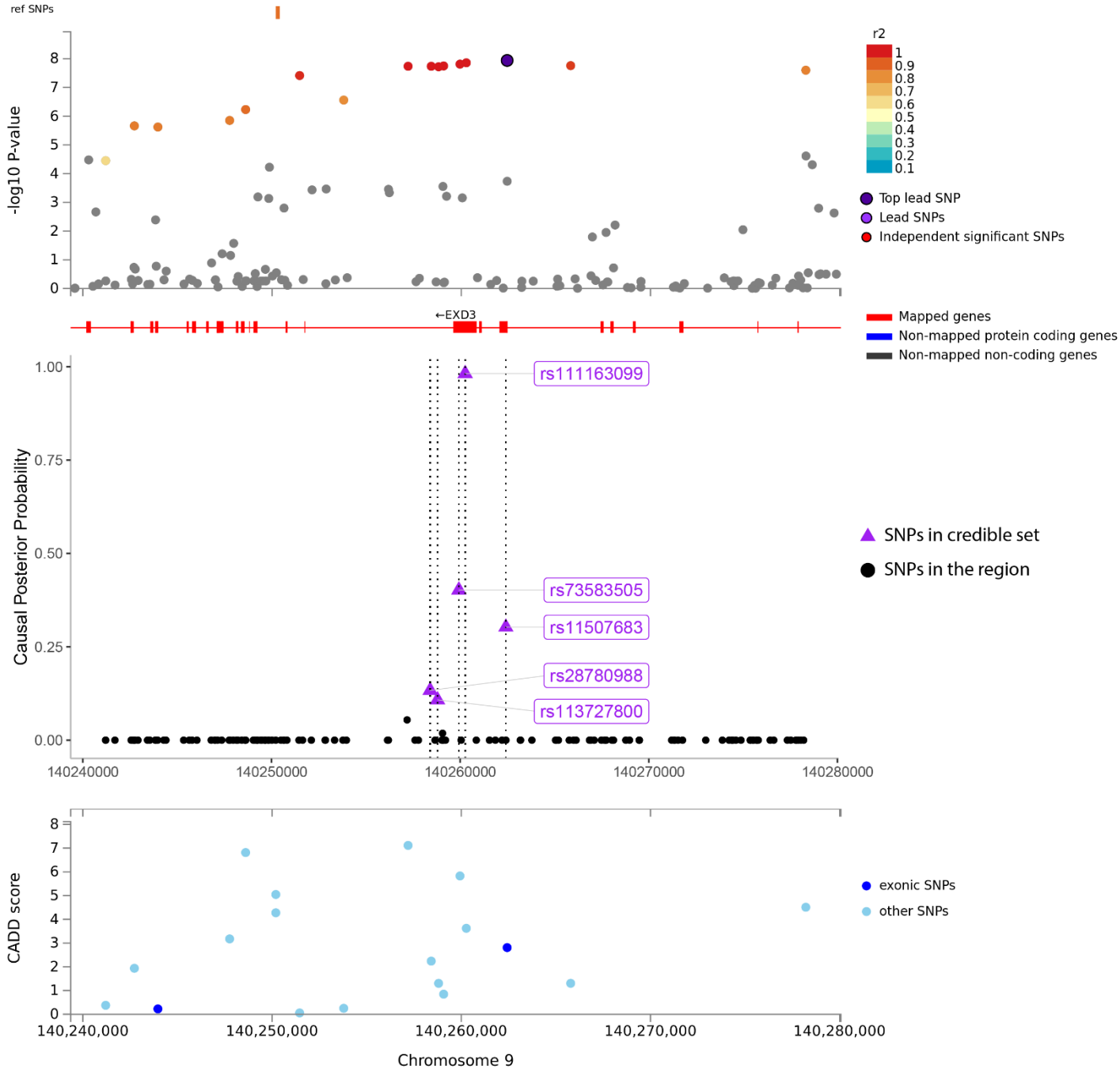


Figure 9: Fine mapping of the risk loci identified from FUMA.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The middle panel finemaps all the SNPs in the region using CAVIAR where SNPs labelled and marked as triangles were identified as credible set. The y-axis is posterior probability of SNPs.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.

chr11:133531999-133561399

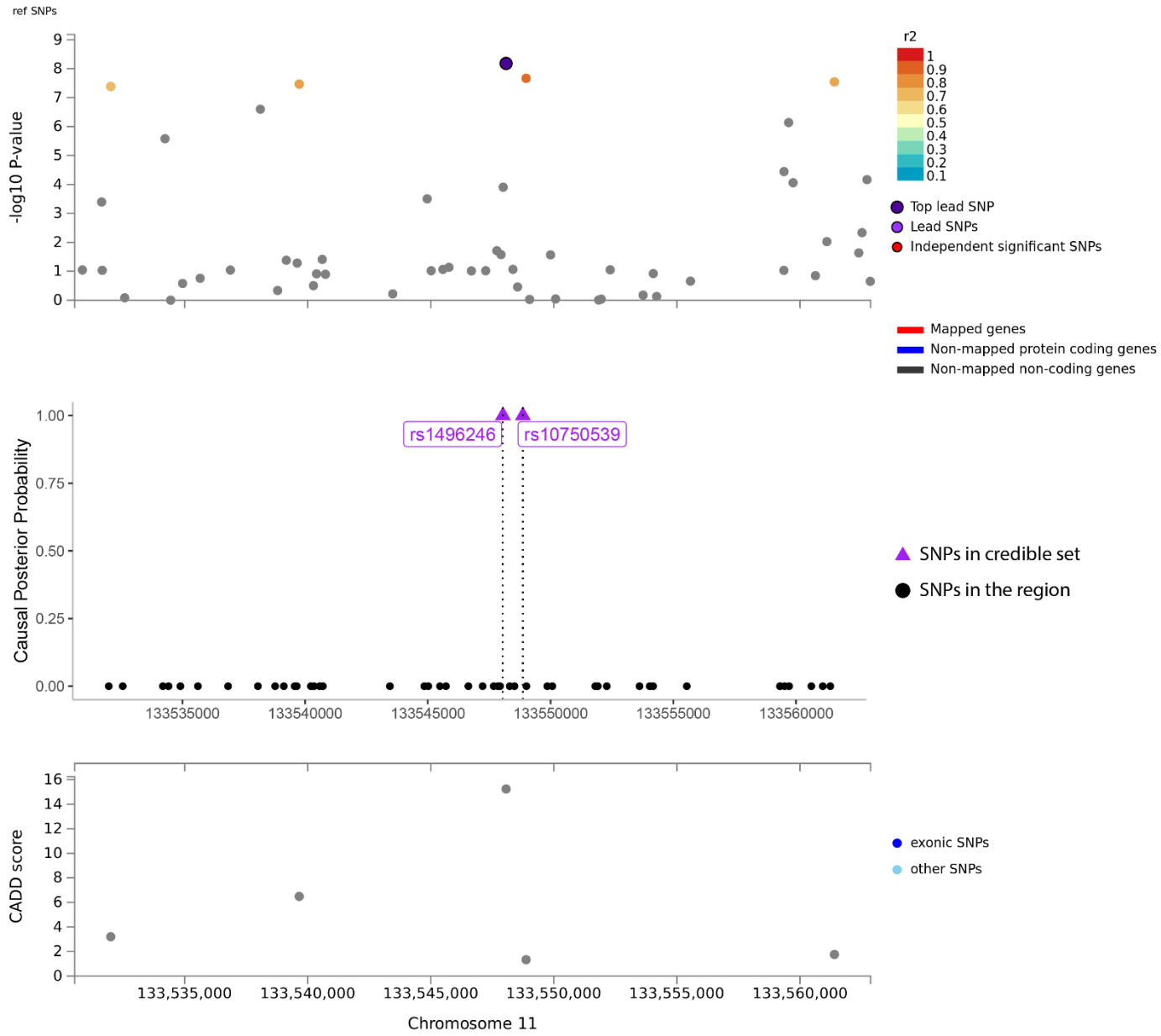


Figure 10: Fine mapping of the risk loci identified from FUMA.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The middle panel finemaps all the SNPs in the region using CAVIAR where SNPs labelled and marked as triangles were identified as credible set. The y-axis is posterior probability of SNPs.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.

chr17:42785473-44865603

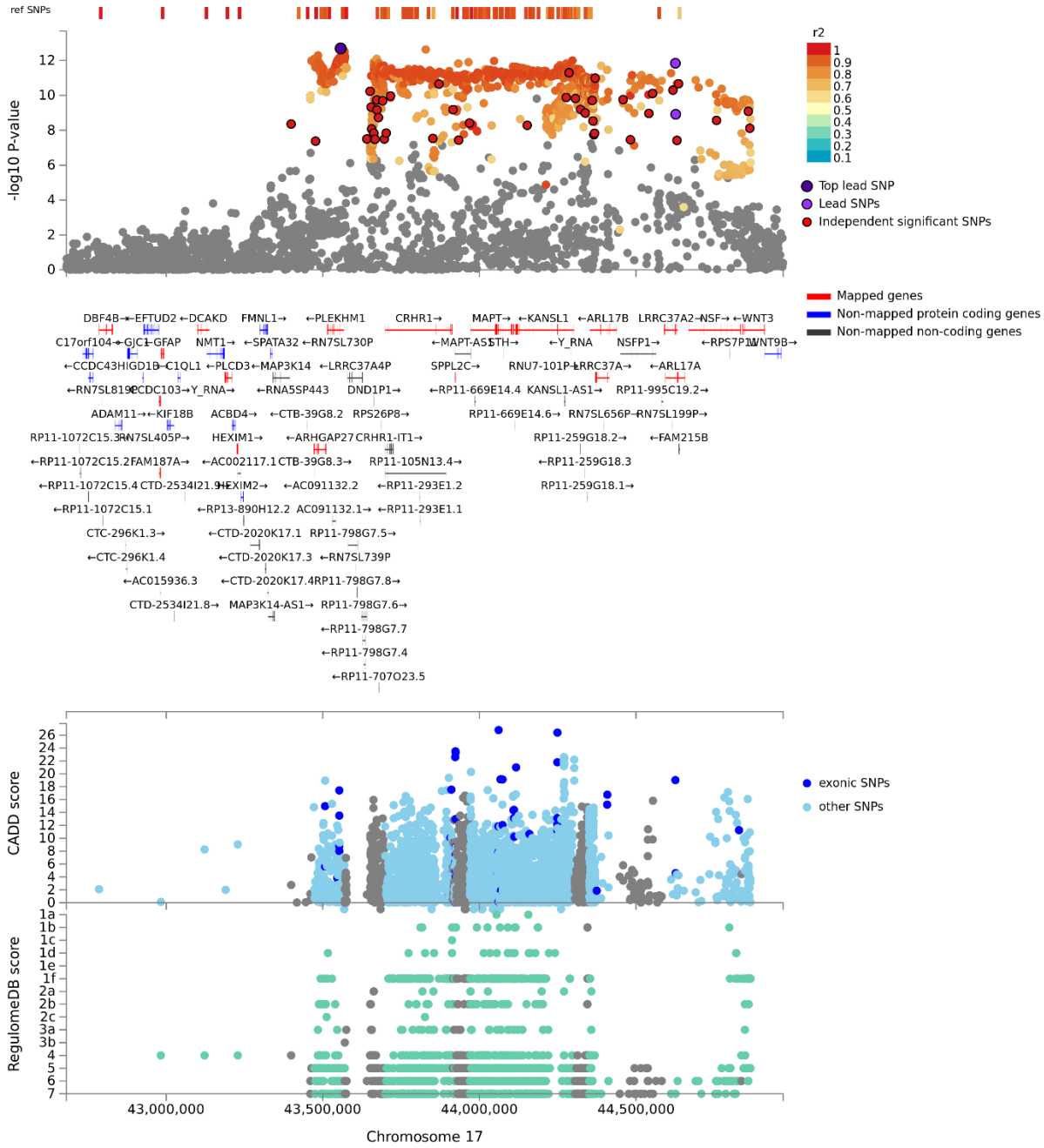


Figure 11: Risk loci identified from FUMA.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The middle panel shows the CADD score (y-axis) for the SNPs in the region.
The bottom panel shows scores of transcriptional activity as catalogued in RegulomeDB. The legend of the scores is shown on the right. The x-axis in all panels represents genomic location (base pair) of the locus.
Note: Finemapping is unable to prioritize variants in this region due to high association and complex LD pattern.

Regulome DB

RegulomeDB Categorical Scores

Category	Description
	Likely to affect binding and linked to expression of a gene target
1a	eQTL + TF binding + matched TF motif + matched DNase footprint + DNase peak
1b	eQTL + TF binding + any motif + DNase footprint + DNase peak
1c	eQTL + TF binding + matched TF motif + DNase peak
1d	eQTL + TF binding + any motif + DNase peak
1e	eQTL + TF binding + matched TF motif
1f	eQTL + TF binding/DNase peak
	Likely to affect binding
2a	TF binding + matched TF motif + matched DNase footprint + DNase peak
2b	TF binding + any motif + DNase footprint + DNase peak
2c	TF binding + matched TF motif + DNase peak
	Less likely to affect binding
3a	TF binding + any motif + DNase peak
3b	TF binding + matched TF motif
	Minimal binding evidence
4	TF binding + DNase peak
5	TF binding or DNase peak
6	Motif hit
	No binding evidence
7	No evidence
NA	the variant does not exist in RegulomeDB

*External link to RegulomeDB from SNP table (when one of the SNPs is clicked) will open a new tab. rsID does not always match since RegulomeDB used dbSNP build 141 (the rsID in FUMA is dbSNP build 146). Genomic position (bp on hg19) shown in the link of RegulomeDB is the position shown in the SNP table - 1, since RegulomeDB used 0 based coordinate.

chr18:53190188-53393895

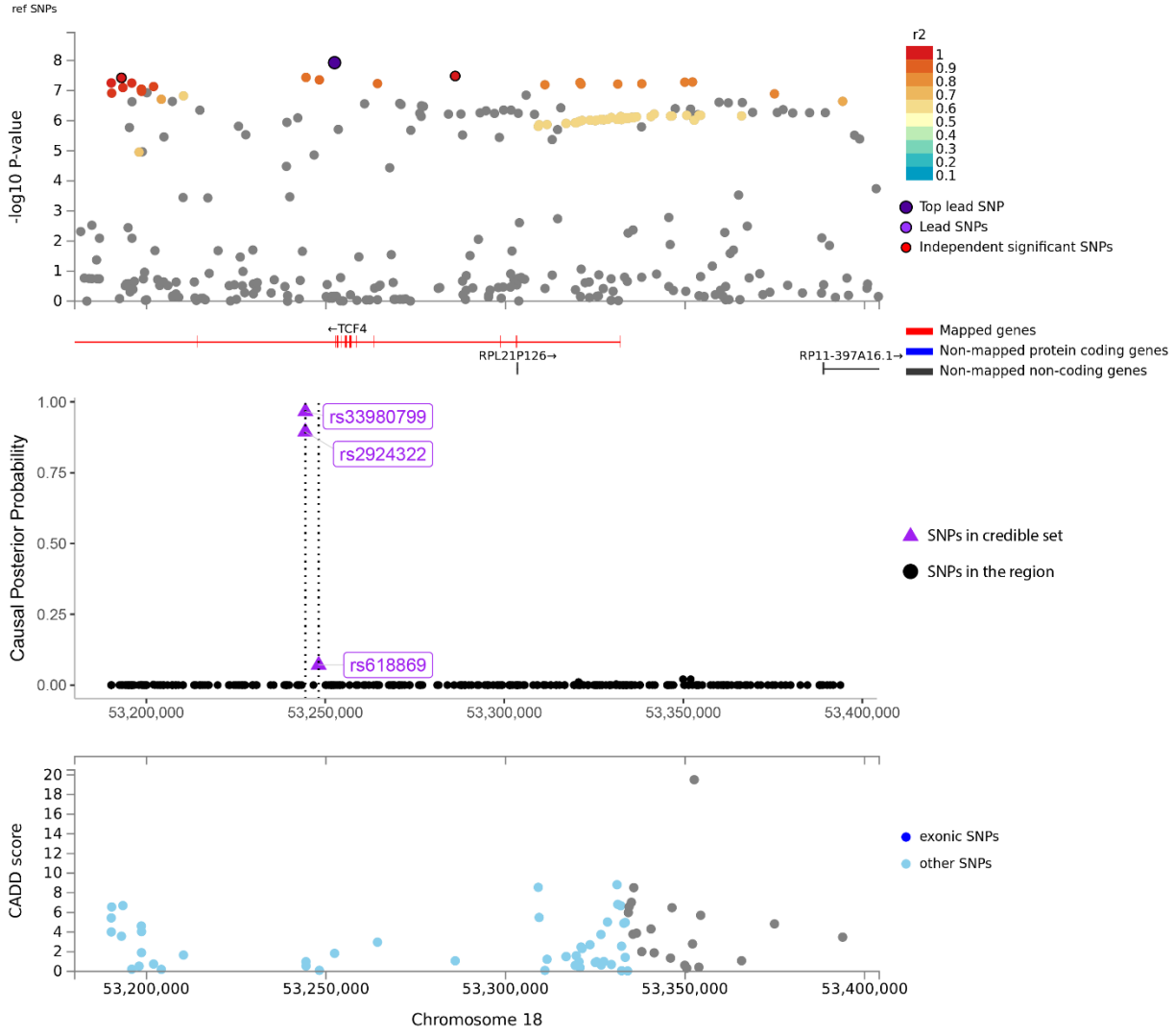


Figure 12: Fine mapping of the risk loci identified from FUMA.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The middle panel finemaps all the SNPs in the region using CAVIAR where SNPs labelled and marked as triangles were identified as credible set. The y-axis is posterior probability of SNPs.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.

chr19:18669987-18730695

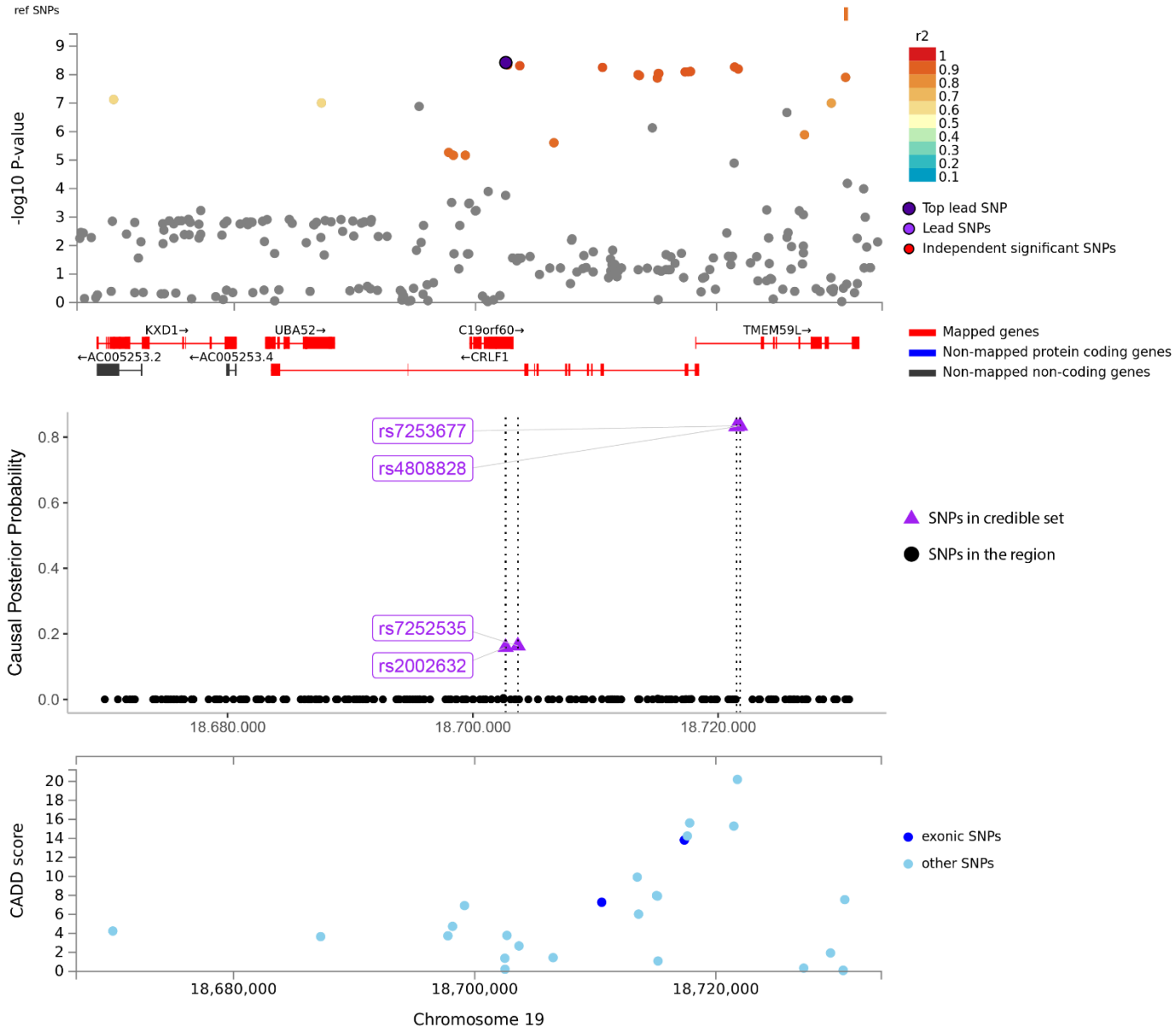
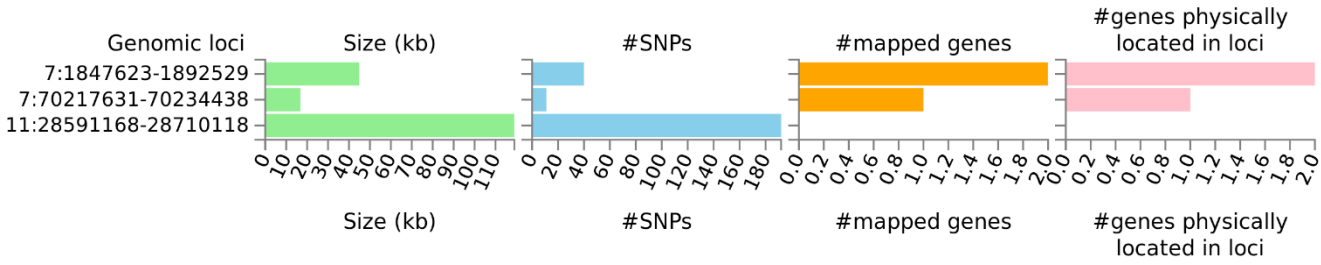


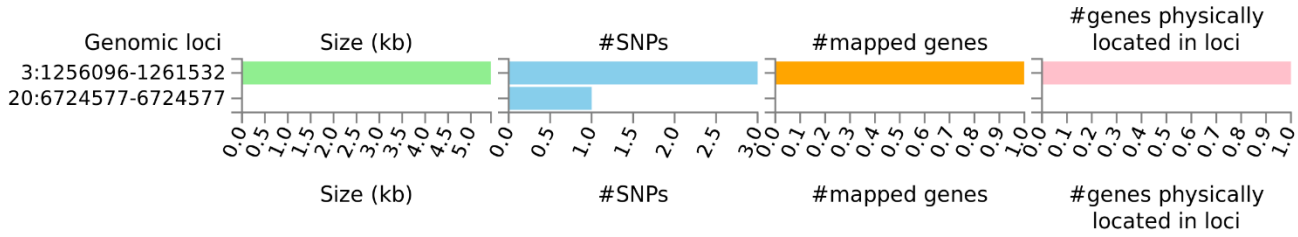
Figure 13: Fine mapping of the risk loci identified from FUMA.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The middle panel finemaps all the SNPs in the region using CAVIAR where SNPs labelled and marked as triangles were identified as credible set. The y-axis is posterior probability of SNPs.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.

Figure 2: Summary of risk loci for EHR-binary

A: EUR population



B: AFR-AM population



chr7:1847623-1892529 (EUR)

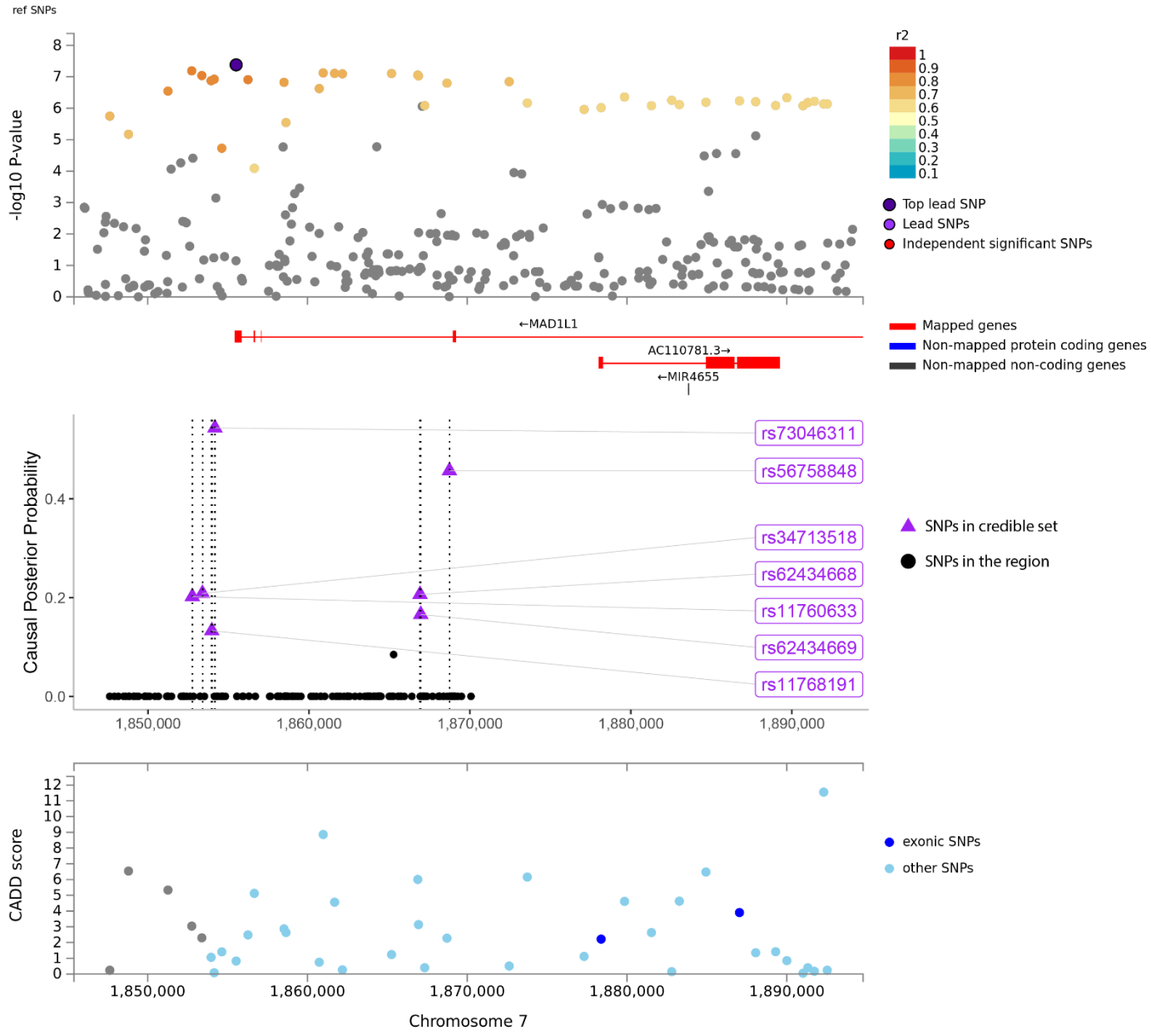


Figure 15: Fine mapping of the risk loci identified from FUMA.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The middle panel finemaps all the SNPs in the region using CAVIAR where SNPs labelled and marked as triangles were identified as credible set. The y-axis is posterior probability of SNPs.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.

chr7:70217631-70234438 (EUR)

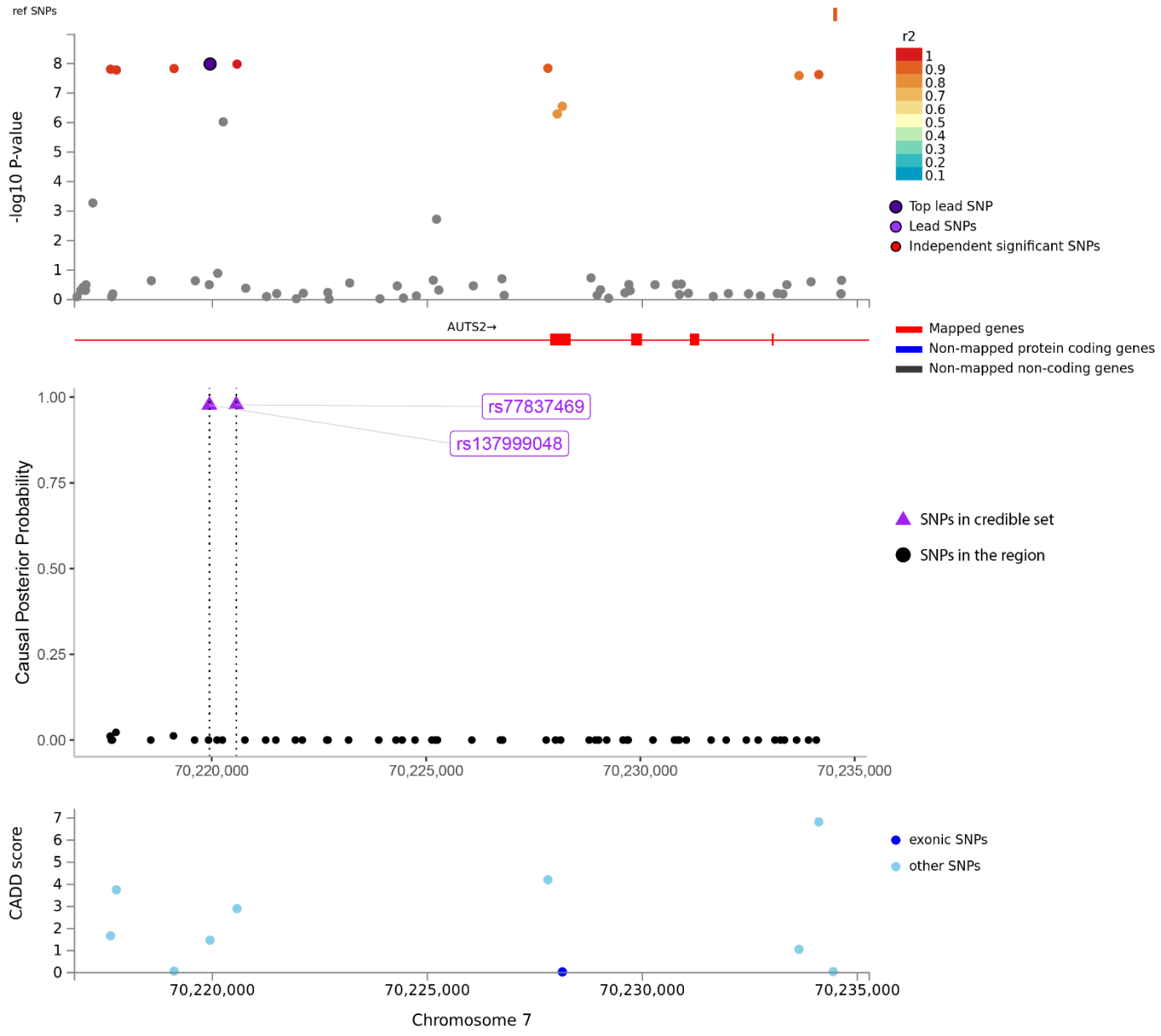


Figure 16: Fine mapping of the risk loci identified from FUMA.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The middle panel finemaps all the SNPs in the region using CAVIAR where SNPs labelled and marked as triangles were identified as credible set. The y-axis is posterior probability of SNPs.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.

chr11:28591168-28710118 (EUR)

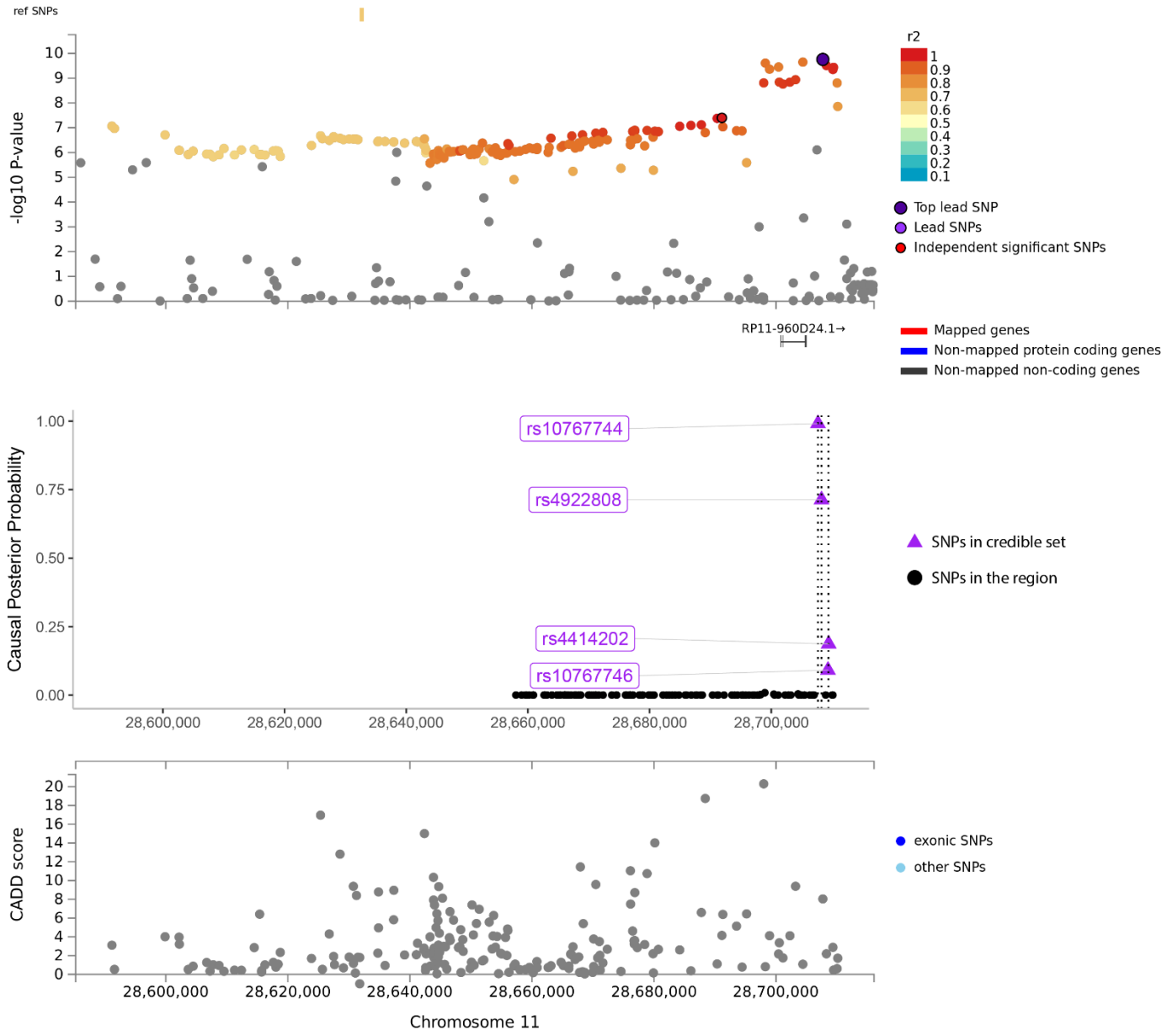


Figure 17: Fine mapping of the risk loci identified from FUMA.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The middle panel finemaps all the SNPs in the region using CAVIAR where SNPs labelled and marked as triangles were identified as credible set. The y-axis is posterior probability of SNPs.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.

chr3:1256096-1261532 (AFR)

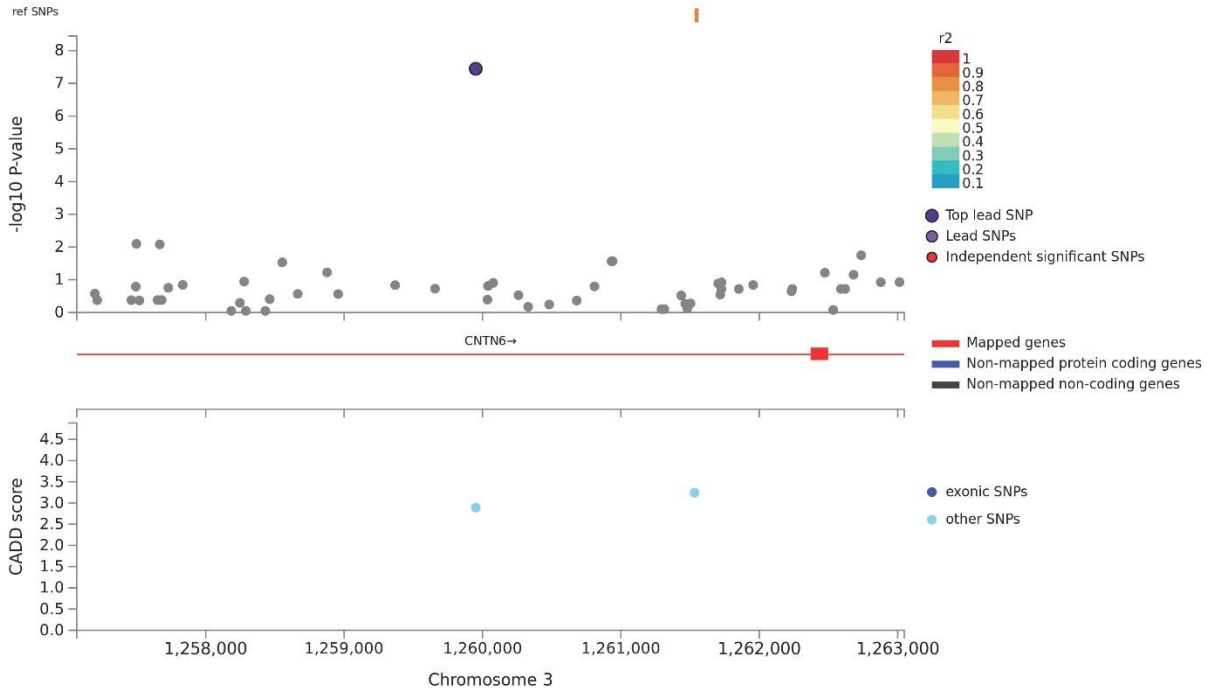


Figure 18: Risk loci identified from FUMA in African American population.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.
Note: Fine-mapping is not applicable at this loci since there are no SNPs in LD with the lead SNP.

chr20:6724577-6724577 (AFR)

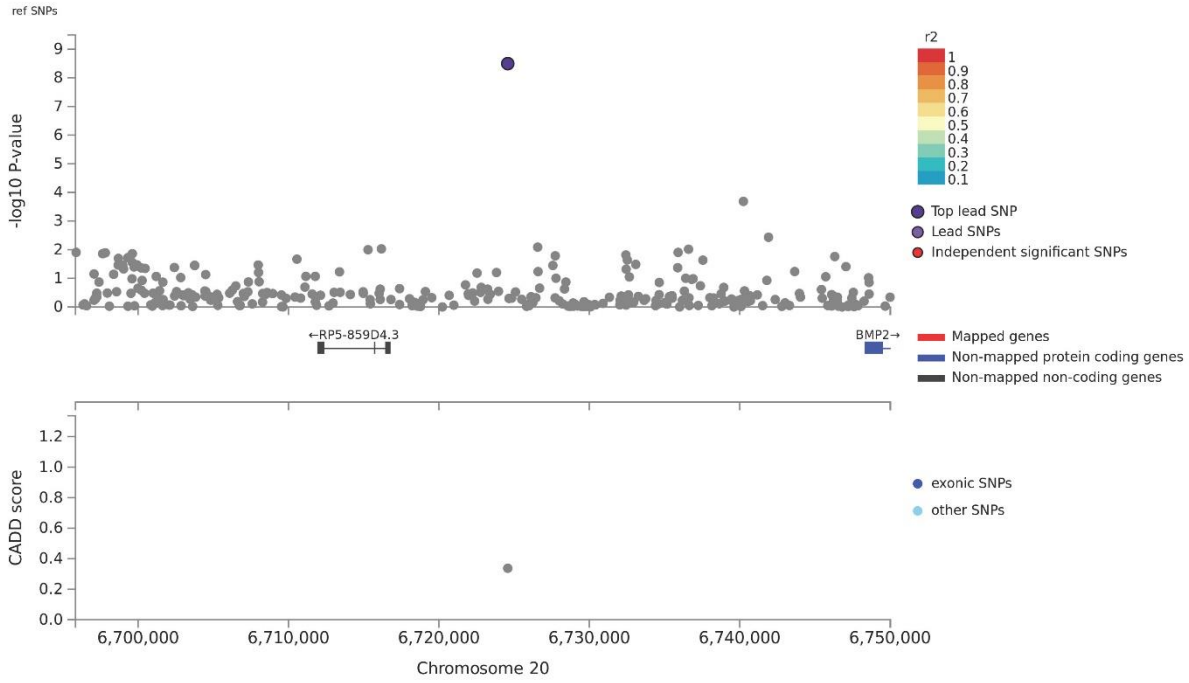


Figure 19: Risk loci identified from FUMA in African American population.
The top panel shows regional association of SNPs in the locus. The y-axis is -log10 of p-values for each SNP.
The bottom panel shows the CADD score (y-axis) for the SNPs in the region. The x-axis in all panels represents genomic location (base pair) of the locus.
Note: Fine-mapping is not applicable at this loci since there are no SNPs in LD with the lead SNP.