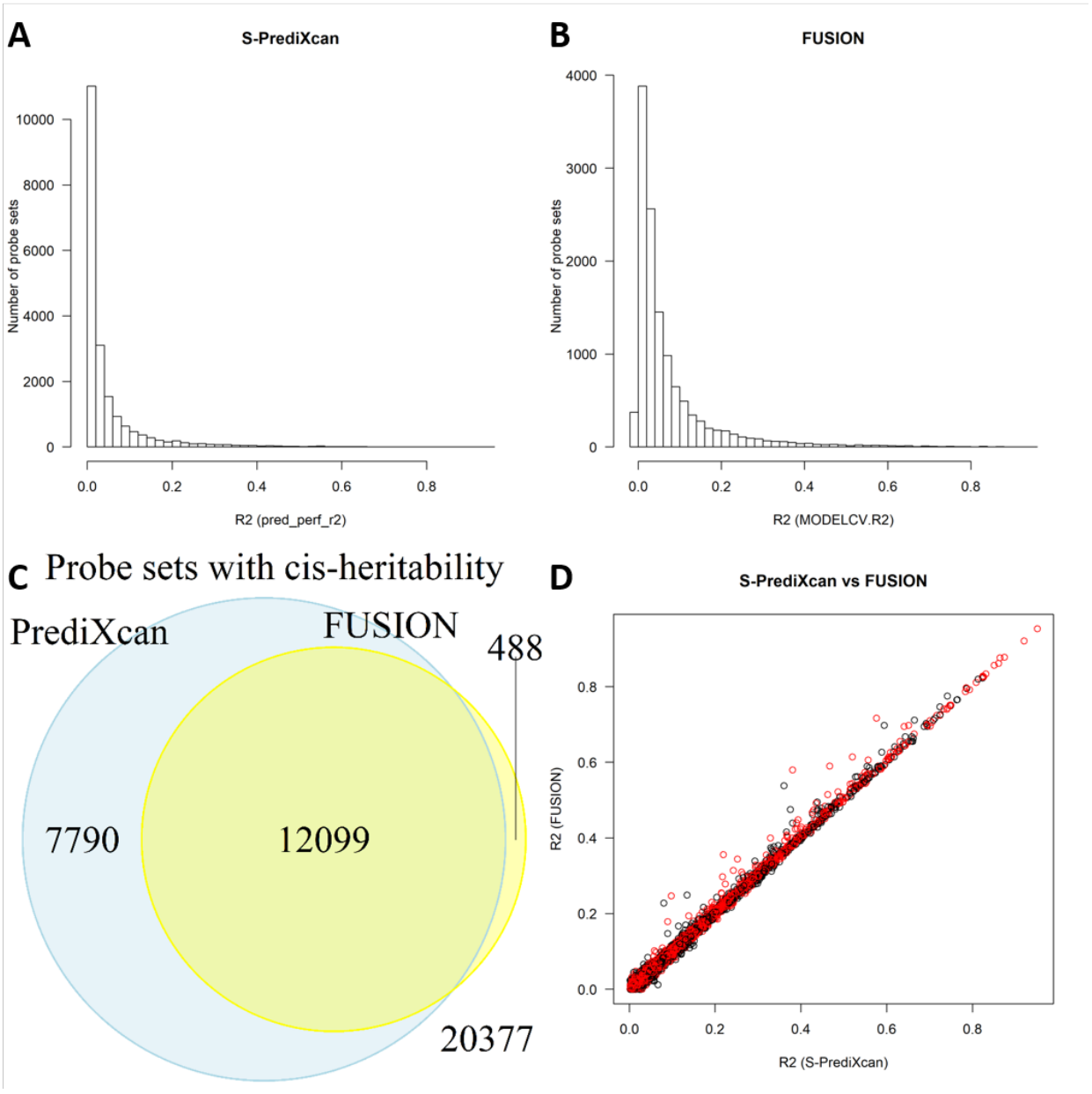
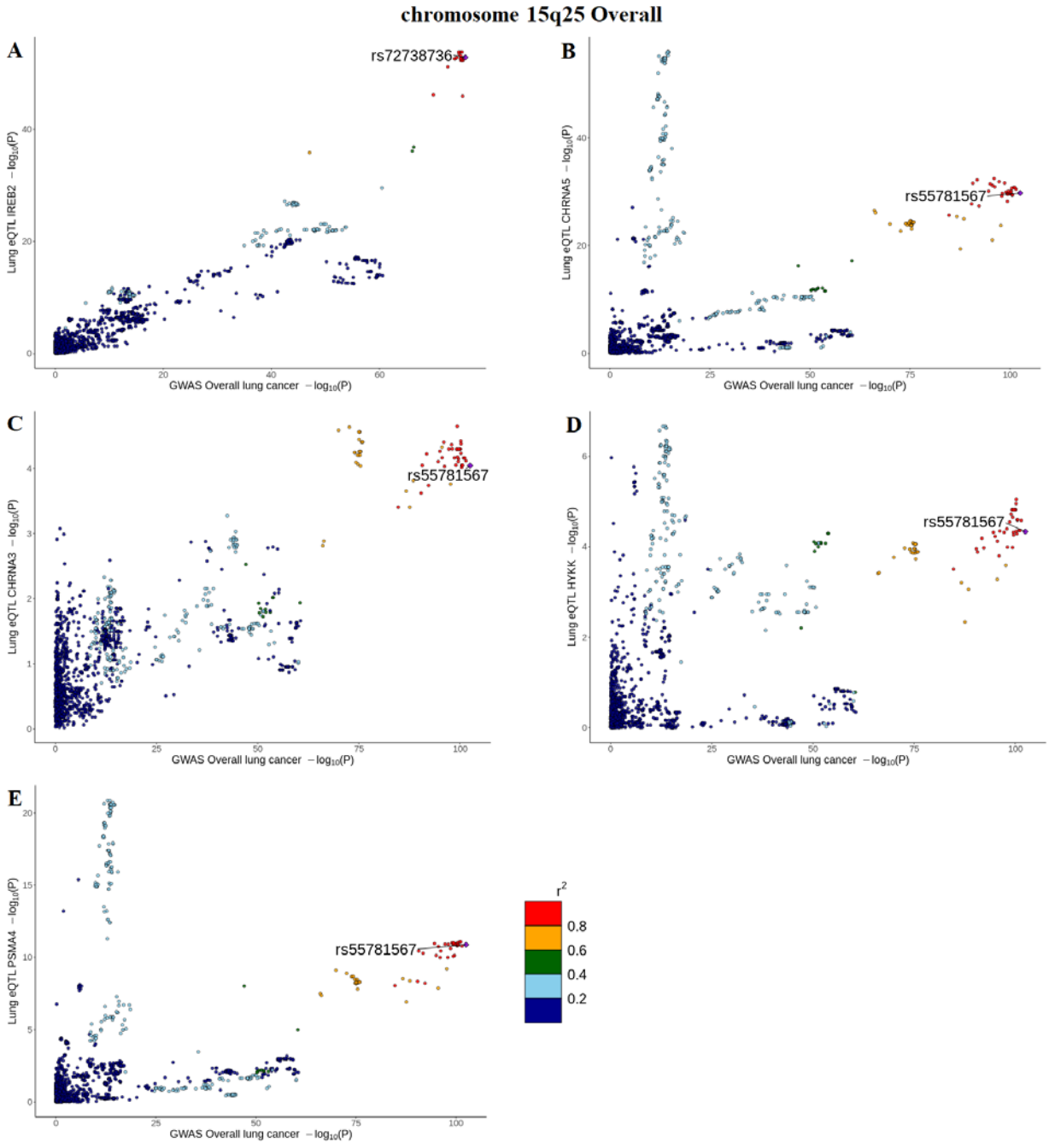
Transcriptome-wide association study reveals candidate causal genes for lung cancer

Yohan Bossé, Zhonglin Li, Jun Xia, Venkata Manem, Robert Carreras-Torres, Aurélie Gabriel, Nathalie Gaudreault, Demetrius Albanes, Melinda C. Aldrich, Angeline Andrew, Susanne Arnold, Heike Bickeböller, Stig E. Bojesen, Paul Brennan, Hans Brunnstrom, Neil Caporaso, Chu Chen, David C. Christiani, John K. Field, Gary Goodman, Kjell Grankvist, Richard Houlston, Mattias Johansson, Mikael Johansson, Lambertus A. Kiemeney, Stephen Lam, Maria Teresa Landi, Philip Lazarus, Loic Le Marchand, Geoffrey Liu, Olle Melander, Gadi Rennert, Angela Risch, Susan M. Rosenberg, Matthew B. Schabath, Sanjay Shete, Zhuoyi Song, Victoria L. Stevens, Adonina Tardon, H-Erich Wichmann, Penella Woll, Shan Zienolddiny, Ma'en Obeidat, Wim Timens, Rayjean J. Hung, Philippe Joubert, Christopher I. Amos, James D. McKay

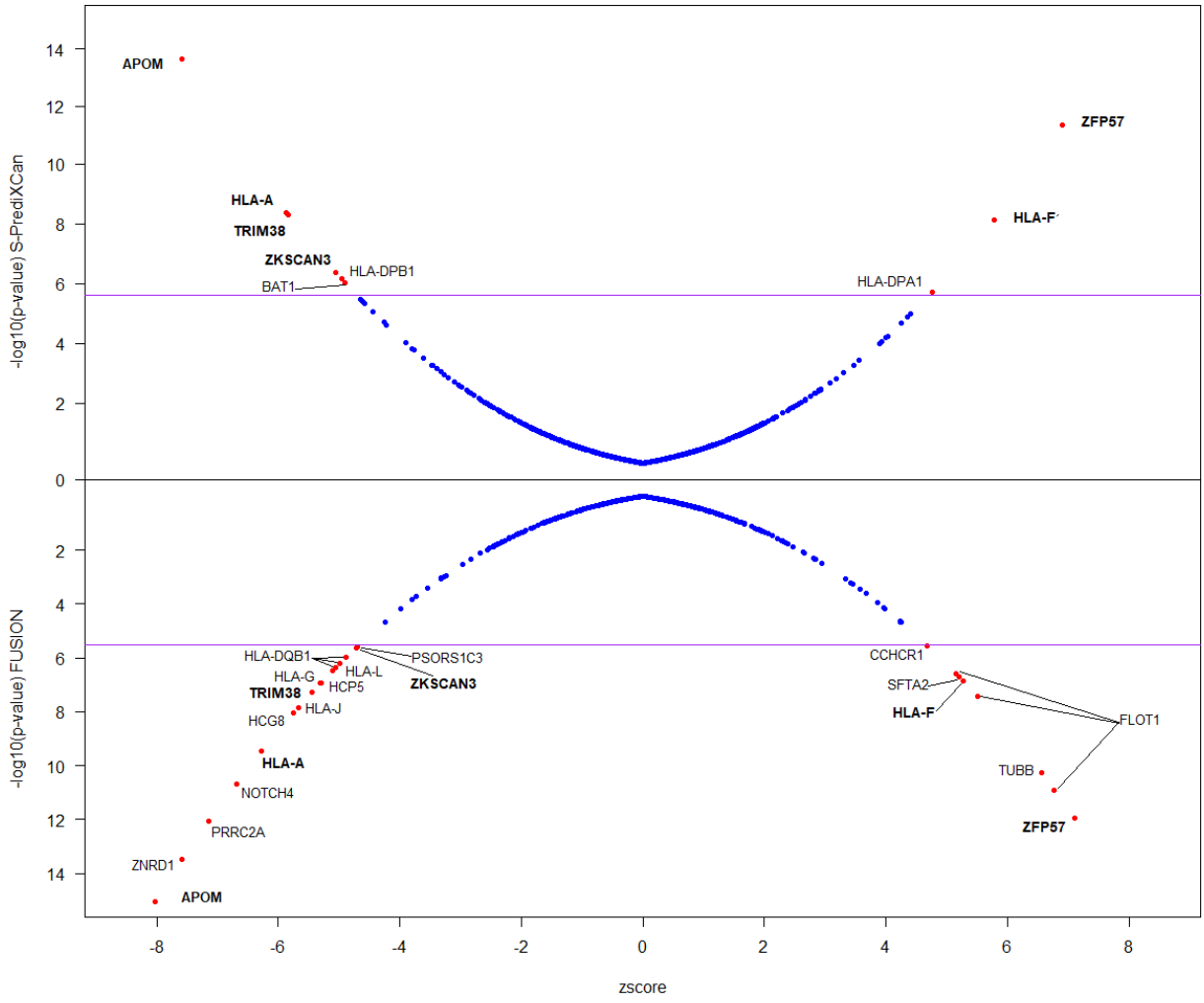
Supplementary Figure 1. Comparison of the *cis*-genetic component of expression in the lung between S-PrediXcan and FUSION. A) Histogram showing the distribution of the gene expression variance explained by SNPs for probe sets with significant prediction models (FDR<0.05) in S-PrediXcan. B) Histogram showing the distribution of the gene expression variance explained by SNPs for probe sets with significant *cis*-heritability (P<0.01) in FUSION. C) Venn diagram showing the number of probe sets with significant *cis*-genetic component of expression that overlap between S-PrediXcan and FUSION. D) Scatter plot showing the prediction performance for the 12,099 probe sets in common between S-PrediXcan and FUSION. Red dots indicate probe sets evaluated with different prediction models in S-PrediXcan and FUSION (enet vs LASSO).



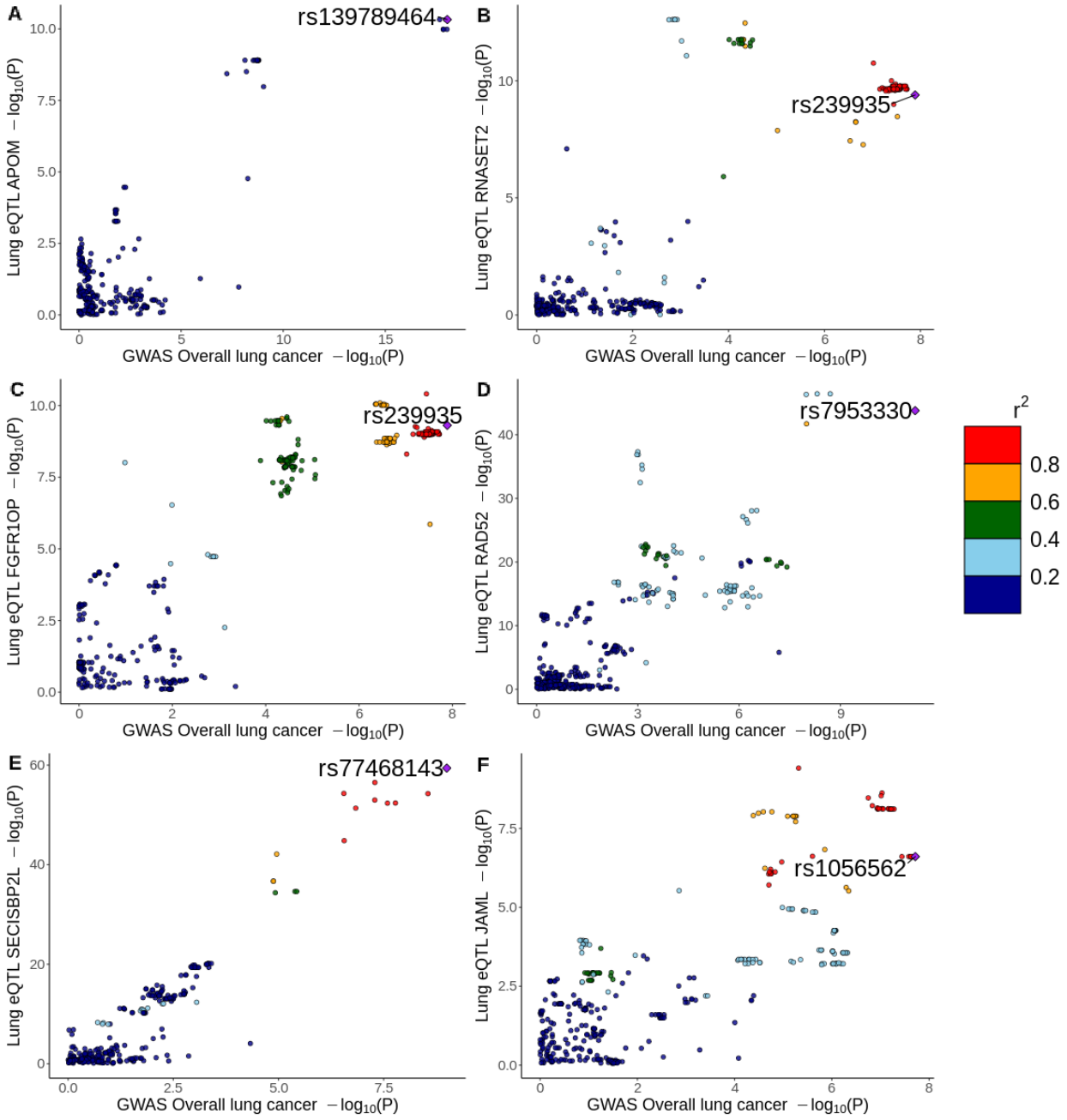
Supplementary Figure 2. LocusCompare plots for significant TWAS genes on chromosome 15q25. Association signals for SNPs within 50 Kb up and downstream of target genes are illustrated. A) *IREB2*, B) *CHRNA5*, C) *CHRNA3*, D) *HYKK*, and E) *PSMA4*.



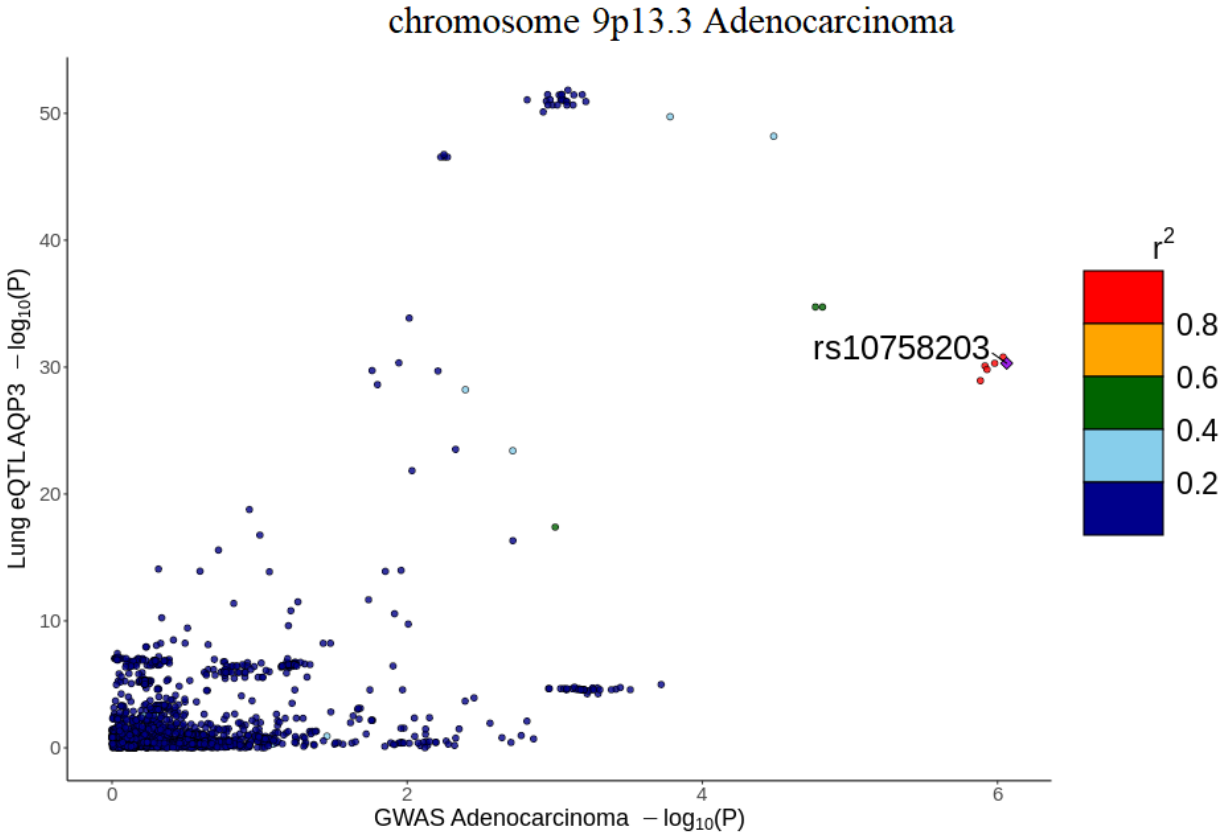
Supplementary Figure 3. TWAS results for probe sets located in the MHC locus (6p21). Results for S-PrediXcan (top) and FUSION (bottom) are illustrated in a mirror view to show similarity and differences between the two TWAS approaches. Each point represents a probe set. The x-axis shows the TWAS z-scores. The P values for gene expression-lung cancer associations are on the y-axis in -log10 scale. Annotations for the significant probe sets are indicated. The six genes in common between S-PrediXcan and FUSION with consistent direction of effect are in bold. Purple lines represent the Bonferroni significant thresholds.



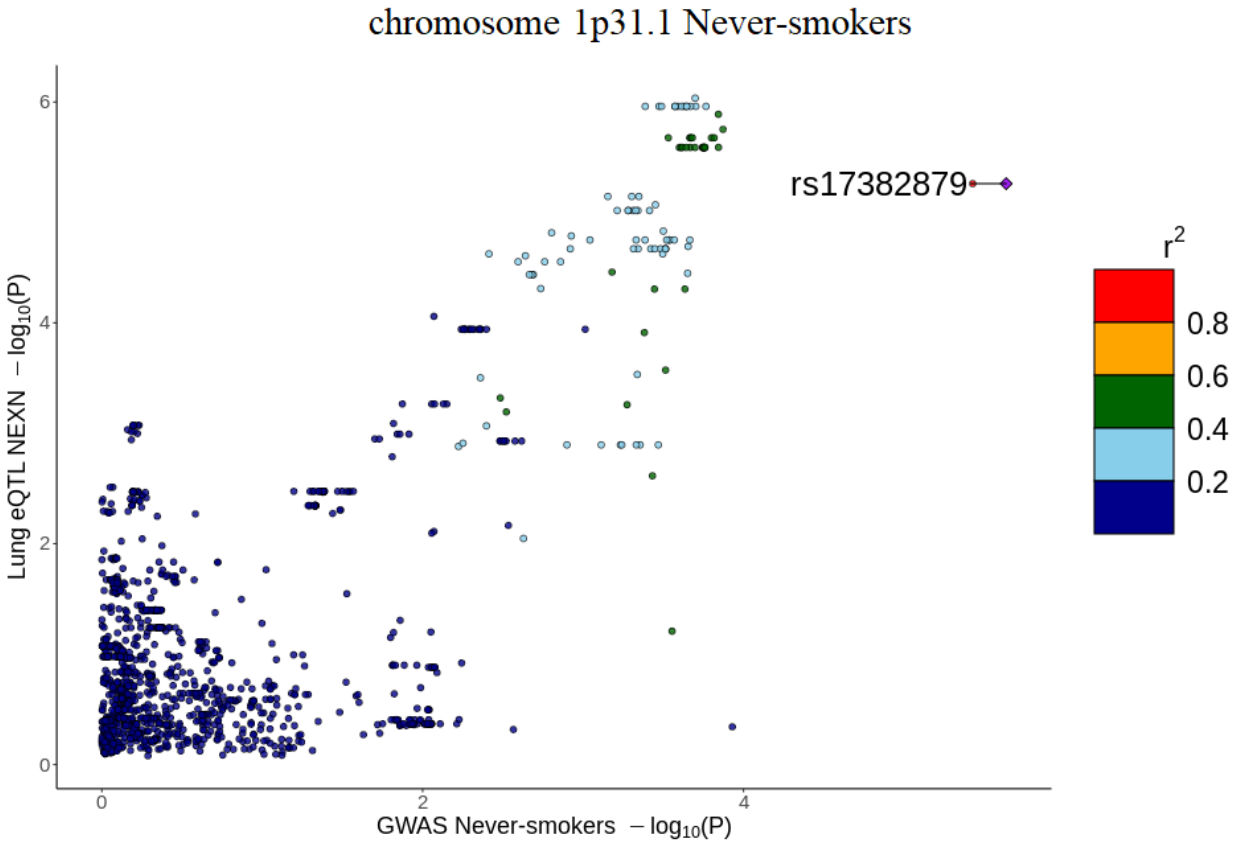
Supplementary Figure 4. LocusCompare plots for TWAS hits for overall lung cancer. Association signals for SNPs within 50 Kb up and downstream of target genes are illustrated. A) *APOM* on 6p21, B) *RNASET2* on 6q27, C) *FGFR1OP* on 6q27, D) *RAD52* on 12p13.33, E) *SECISBP2L* on 15q21.1, and F) *JAML* on 11q23.3.



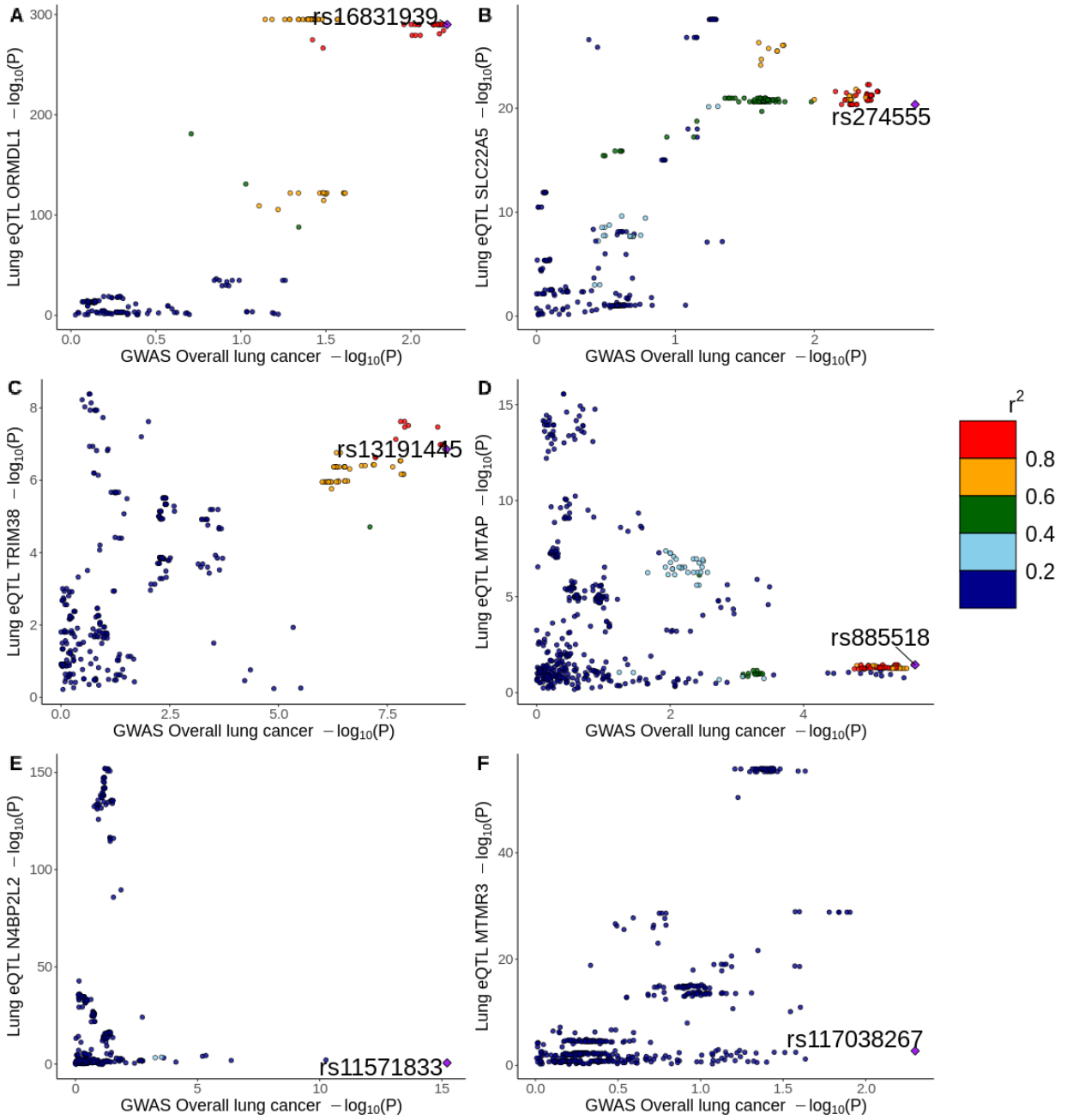
Supplementary Figure 5. LocusCompare plot for *AQP3* on chromosome 9p13.3. Association signals for SNPs within 50 Kb up and downstream of *AQP3* are illustrated. The GWAS lead variant for adenocarcinoma (rs10758203) is also strongly associated with *AQP3* expression in lung tissues.



Supplementary Figure 6. LocusCompare plot for *NEXN* on chromosome 1p31.1. Association signals for SNPs within 50 Kb up and downstream of *NEXN* are illustrated. The GWAS lead variant for lung cancer in never-smokers (rs17382879) is also associated with *NEXN* expression in lung tissues and in LD with the top eQTL-SNPs.



Supplementary Figure 7. LocusCompare plots for the top TWAS genes in known GWAS lung cancer risk loci that show P_{TWAS}<0.05 in both S-PrediXcan and FUSION. Association signals for SNPs within 50 Kb up and downstream of target genes are illustrated. A) *ORMDL1* on 2q32.2, B) *SLC22A5* on 5q31, C) *TRIM38* on 6p22.2, D) *MTAP* on 9p21.3, E) *N4BP2L2* on 13q13.1, and F) *MTMR3* on 22q12.2.



Supplementary Figure 8. The lung cancer gene map. The map is an ideogram of the 22 autosomal human chromosomes. The 45 lung cancer risk loci derived from published GWAS on lung cancer[1] are depicted in blue. The new adenocarcinoma susceptibility locus identified in this study (9p13.3-*AQP3*) is illustrated in red. Annotation is provided for loci for which insightful results were obtained about candidate causal genes. The colors of gene names correspond to their association with overall lung cancer (black), adenocarcinoma (blue), squamous cell carcinoma (green), and never-smokers (magenta). Bonferroni-corrected TWAS genes are in bold.