

## **Mendelian randomization analysis identified potential genes pleiotropically associated with polycystic ovary syndrome**

Qian Sun<sup>1\*</sup>, Gao Yuan<sup>1\*</sup>, Jingyun Yang<sup>2,3</sup>, Jiayi Lu<sup>4</sup>, Wen Feng<sup>1</sup>, Wen Yang<sup>1</sup>

<sup>1</sup>Department of Gynecology, The First Affiliated Hospital of Kangda College of Nanjing Medical University, Lianyungang, Jiangsu, China

<sup>2</sup>Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA

<sup>3</sup>Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA

<sup>4</sup>Department of Finance, School of Economics, Shanghai University, Shanghai, China

\*The two authors contributed equally to this paper and share the first authorship

Corresponding to

Wen Yang

Department of Gynecology

The First Affiliated Hospital of Kangda College of Nanjing Medical University

6 Zhenhua Road

Lianyungang, Jiangsu 222061

China

Tel: 86-18961325910

E-mail: 1010393828@qq.com

## ABSTRACT

**Research question:** Polycystic ovary syndrome (PCOS) is a common endocrine disorder with unclear etiology. Are there any genes that are pleiotropically or potentially causally associated with PCOS?

**Design:** We applied the summary data-based Mendelian randomization (SMR) method integrating genome-wide association study (GWAS) for PCOS and expression quantitative trait loci (eQTL) data to identify genes that were pleiotropically associated with PCOS. We performed separate SMR analysis using eQTL data in the ovary and whole blood.

**Results:** Although no genes showed significant pleiotropic association with PCOS after correction for multiple testing, some of the genes exhibited suggestive significance. *RPS26* showed the strongest suggestive pleiotropic association with PCOS in both SMR analyses ( $\beta$ [SE]=0.10[0.03],  $P_{\text{SMR}}=1.72 \times 10^{-4}$  for ovary;  $\beta$ [SE]=0.11[0.03],  $P_{\text{SMR}}=1.40 \times 10^{-4}$  for whole blood). *PM20D1* showed the second strongest suggestive pleiotropic association with PCOS in the SMR analysis using eQTL data for the whole blood, and was also among the top ten hit genes in the SMR analysis using eQTL data for the ovary. Two other genes, including *CTC-457L16.2* and *NEIL2*, were among the top ten hit genes in both SMR analyses.

**Conclusion:** We identified multiple genes that were potentially involved in the pathogenesis of PCOS. Our findings provided helpful leads to a better understanding of the mechanisms underlying PCOS, and revealed potential therapeutic targets for the effective treatment of PCOS.

**Keywords:** polycystic ovary syndrome; pleotropic association; expression quantitative trait loci; summary Mendelian randomization

## **KEY MESSAGE**

Polycystic ovary syndrome (PCOS) is a common endocrine disorder, and its etiology can be multifaceted. We found that multiple genes were potentially involved in the pathogenesis of PCOS. The findings revealed the genetic mechanisms underlying PCOS and potential therapeutic targets for the effective treatment of PCOS.

## INTRODUCTION

Polycystic ovary syndrome (PCOS) is a common endocrine disorder that often occurs in women of childbearing age, characterized by clinical or biochemical manifestations of excessive androgen, persistent anovulation, and polycystic ovarian changes and often accompanied by insulin resistance (IR) and obesity (Bellver et al., 2018). As a highly prevalent endocrine disorder, PCOS affects 5-13% of reproductive-aged women and more than 10% of adolescents (Azziz et al., 2016; Naz et al., 2019; Rotterdam, 2004b; Teede et al., 2010). PCOS is associated with a variety of consequences, such as reproductive issues, metabolic abnormalities, and psychological disorders (Legro et al., 2013). Furthermore, more than half of the PCOS patients suffer from infertility (Joham et al., 2015), and infertile women and adolescent girls with PCOS were reported to have reduced quality of life (QoL) (Barnard et al., 2007; Naumova et al., 2020; Wilson and Pena, 2020).

PCOS is a multifaceted disease, and previous studies have found its association with genetic factors (Zhao et al., 2016), environmental factors such as exposure to Bisphenol A (BPA) (Kandaraki et al., 2011), and life style factors such as smoking (Zhang et al., 2020). However, the exact etiology of PCOS remains to be elucidated. More studies are needed to further explore the pathological mechanisms underlying PCOS to facilitate diagnosis and treatment of this common and jeopardizing disease.

Genetics play an important role in the pathogenesis of PCOS. Previous twin studies demonstrated that genetics explained more than 70% of PCOS pathogenesis (Vink et al., 2006). Multiple genome-wide association studies (GWAS) conducted on Han Chinese, Korean, and European populations have identified multiple genetic variants in association

with PCOS, such as single-nucleotide polymorphisms in/near *FSHR*, *THADA* (Z. J. Chen et al., 2011), *YAPI*, *HMGA2* (Shi et al., 2012), *KHDRBS3* (Lee et al., 2015), *FSHB*, *GATA4/NEIL2* (Hayes et al., 2015), *ERBB4/HER4* and *RAD50* (F. R. Day et al., 2015). However, biological interpretation of the identified genetic variants in the etiology of PCOS remains largely unclear. It is likely that the genetic variants identified in GWAS could exert their effects on diseases/disorders via gene expression because many of them are located in non-coding regions (Visscher et al., 2012). Therefore, exploring the relationship between genetic variation and gene expression can help better understand the regulatory pathways underlying the pathogenesis of PCOS.

Mendelian randomization (MR) is a method for exploring potentially causal association between an exposure and an outcome by using genetic variants as the instrumental variables (IVs) for exposure (Davey Smith and Hemani, 2014). Compared with traditional statistical methods used in the association studies, MR can reduce confounding and reverse causation and is becoming increasingly popular in the exploration of etiological mechanisms (Burgess et al., 2015; Thanassoulis and O'Donnell, 2009). A novel analytical framework through a summary data-based MR (SMR) approach integrating cis- expression quantitative trait loci (cis-eQTL) or cis- DNA methylation QTL (cis-mQTL) and GWAS data was recently proposed (Zhu et al., 2016), and has been employed in identifying gene expressions or DNA methylation loci that are pleiotropically or potentially causally associated with various phenotypes, such as severity of COVID-19, major depression and Alzheimer's disease (D. Liu et al., 2020; Pavlides et al., 2016; Wray et al., 2018), implying that it is a promising tool to explore genes that are pleiotropically associated with complex traits.

In this paper, we applied the SMR method integrating summarized GWAS data for PCOS and cis- eQTL data to prioritize genes that are pleiotropically/potentially causally associated with PCOS.

## **MATERIALS AND METHODS**

### **Data sources**

#### *eQTL data*

In the SMR analysis, cis-eQTL genetic variants were used as the instrumental variables (IVs) for gene expression. We performed SMR analysis using eQTL data in ovary and blood from the V7 release of the GTEx summarized data. The summarized data included 85 participants for ovary and 338 participants for blood (Consortium et al., 2017). The eQTL data can be downloaded at

<https://cnsgenomics.com/data/SMR/#eQTLsummarydata>.

#### *GWAS data for PCOS*

The GWAS summarized data for CCT were provided by a recent genome-wide association meta-analysis of PCOS (F. Day et al., 2018). The results were based on meta-analyses of seven cohorts of European descent, including a total of 10,074 PCOS cases and 103,164 controls. Diagnosis of PCOS was based on the NIH (Zawadzki and Dunaif, 1992) or the Rotterdam Criteria (Rotterdam, 2004a), and the self-report-based data from 23andMe were excluded. The GWAS summarized data can be downloaded at

<https://www.repository.cam.ac.uk/handle/1810/283491>.

#### *SMR analysis*

In the SMR analyses, cis-eQTL was the IV, gene expression was the exposure, and PCOS was the outcome. The analysis was done using the method as implemented in the software SMR. SMR applies the principles of MR to jointly analyze GWAS and eQTL summary statistics to test for pleiotropic association between gene expression and a trait due to a shared and potentially causal variant at a locus. Detailed information regarding the SMR method was reported in a previous publication (Zhu et al., 2016). We also conducted the heterogeneity in dependent instruments (HEIDI) test to evaluate the existence of linkage in the observed association. Rejection of the null hypothesis (i.e.,  $P_{\text{HEIDI}} < 0.05$ ) indicates that the observed association could be due to two distinct genetic variants in high linkage disequilibrium with each other. We adopted the default settings in SMR (e.g.,  $P_{\text{eQTL}} < 5 \times 10^{-8}$ , minor allele frequency [MAF]  $> 0.01$ , removing SNPs in very strong linkage disequilibrium [LD,  $r^2 > 0.9$ ] with the top associated eQTL, and removing SNPs in low LD or not in LD [ $r^2 < 0.05$ ] with the top associated eQTL), and used false discovery rate (FDR) to adjust for multiple testing.

Data cleaning and statistical/bioinformatical analysis was performed using R version 4.0.4 (<https://www.r-project.org/>), PLINK 1.9 (<https://www.cog-genomics.org/plink/1.9/>) and SMR (<https://cnsgenomics.com/software/smr/>).

## **RESULTS**

### **Basic information of the summarized data**

The GWAS summarized data were based on GWAS meta-analysis of 113,238 subjects (10,074 PCOS cases and 103,164 controls). After checking of allele frequencies among the datasets and LD pruning, there were about 6.4 million eligible SNPs included in the



final SMR analysis. The number of participants for GETx eQTL data in ovary is smaller (n=85), compared with that in whole blood (n=385), so is the number of eligible probes (1,530 vs. 4,490). The detailed information was shown in **Table 1**.

### **Pleiotropic association with PCOS**

Information of the top ten probes using eQTL data for the ovary and whole blood was presented in **Table 2**. Although no genes showed significant pleiotropic association with PCOS after correction for multiple testing, some of genes exhibited suggestive significance. Specifically, *RPS26* (ENSG00000197728.5) showed the strongest suggestive pleiotropic association with PCOS in both SMR analyses ( $\beta$ [SE]=0.10[0.03],  $P_{SMR}=1.72\times 10^{-4}$  for ovary;  $\beta$ [SE]=0.11[0.03],  $P_{SMR}=1.40\times 10^{-4}$  for whole blood; **Figure 1**). *PM20D1* (ENSG00000162877.8) showed the second strongest suggestive pleiotropic association with PCOS in the SMR analysis using eQTL data for the whole blood, and it was also among the top ten hit genes in the SMR analysis using eQTL data for the ovary ( $\beta$ [SE]=0.10[0.03],  $P_{SMR}=2.94\times 10^{-3}$  for ovary;  $\beta$ [SE]=0.14[0.04],  $P_{SMR}=8.38\times 10^{-4}$  for blood; **Figure 2**). Two other genes, including *CTC-457L16.2* (ENSG00000262319.1;  $\beta$ [SE]=-0.2[0.06],  $P_{SMR}=1.55\times 10^{-3}$  for ovary;  $\beta$ [SE]=-0.31[0.09],  $P_{SMR}=1.07\times 10^{-3}$  for blood; **Figure 3**) and *NEIL2* (ENSG00000154328.11;  $\beta$ [SE]=-0.2[0.06],  $P_{SMR}=1.02\times 10^{-3}$  for ovary;  $\beta$ [SE]=-0.30[0.10],  $P_{SMR}=2.25\times 10^{-3}$  for blood; **Figure 4**), were among the top ten hit genes in both SMR analyses.

### **DISCUSSION**

In this study, we integrated GWAS summarized data for PCOS and eQTL data in the SMR analysis to explore genes that were pleiotropically or potentially causal associated

with PCOS. We found that multiple genes were potentially involved in the pathogenesis of PCOS. To the best of our knowledge, this is the first study to explore genes in pleiotropic association PCOS through a Mendelian randomization approach. Our findings provided helpful leads to a better understanding of the mechanisms underlying PCOS, and revealed potential therapeutic targets for the effective treatment of PCOS.

In our study, *RPS26* (Ribosomal Protein S26) showed the strongest suggestive pleiotropic association with PCOS in the SMR analyses using both ovary and blood eQTL data (**Table 2**). *RPS26*, located on 12q13.2, is a gene encoding a ribosomal protein which is a component of the 40S subunit and belongs to the S26E family of ribosomal proteins (Filipenko et al., 1998). A recent research found that *RPS26* critically regulated T-cell survival in a p53-dependent manner (C. Chen et al., 2021). Knockout of *RPS26* in mouse oocytes led to retarded follicle development from pre-antral follicles to antral follicles, and arrested chromatin configurations of the oocytes (X. M. Liu et al., 2018). An earlier study using human ovary cDNA library found that *RPS26* was downregulated in PCOS ovary, compared with the normal human ovary (Diao et al., 2004). A recent study examined genes in PCOS-associated regions using a Bayesian colocalization approach (Coloc) and found seven genes, including *RPS26*, that harbored potential causal variants accounting for approximately 30% of known PCOS signals (Censin et al., 2021). These findings, together with ours, suggested that *RPS26* might play a critical role in the etiology of PCOS, and highlighted the potential of this gene as a promising target for the prevention and treatment of PCOS.

*PM20D1* (Peptidase M20 Domain Containing 1) showed the second strongest suggestive pleiotropic association with PCOS in the SMR analysis using blood eQTL data, and was also among the top hit genes in the SMR analysis using ovary eQTL data (**Table 2**).

*PM20D1*, located on 1q32.1, is a gene encoding a bidirectional enzyme capable of catalyzing both the condensation of fatty acids and amino acids to generate N-acyl amino acids and also the reverse hydrolytic reaction, thereby regulating energy homeostasis (Long et al., 2016). Mice with increased circulating *PM20D1* had increased N-acyl amino acids in blood and improved glucose homeostasis, and *PM20D1* knockout in mice resulted in impaired glucose tolerance and insulin resistance (IR) (Long et al., 2018). In human adipocytes, *PM20D1* is one of the most highly up-regulated genes by the antidiabetic thiazolidinedione drug rosiglitazone, suggesting a potential role of this enzyme and/or N-fatty acyl amino acids in obesity and diabetes (Benson et al., 2019). In addition, *PM20D1* has been found to be associated with various diseases such as Parkinson's disease (Pihlstrom et al., 2015) and Alzheimer's disease (Sanchez-Mut et al., 2018). To date, research is scarce on the association of *PM20D1* with PCOS. Given that PCOS is a metabolic disorder that is closely related with obesity, IR, type 2 diabetes (T2D) and metabolic syndrome (Condorelli et al., 2017), it is highly likely that *PM20D1* could be involved in the etiology of PCOS, and more studies are needed to explore the exact roles that *PM20D1* plays in the pathogenesis of PCOS.

*NEIL2* (Nei Like DNA Glycosylase 2) showed the third strongest suggestive pleiotropic association with PCOS in the SMR analysis using ovary eQTL data, and was also among the top hit genes in the SMR analysis using whole blood eQTL data (**Table 2**). *NEIL2*, located on 8p23.1, is a gene encoding a member of the Fpg/Nei family of DNA

glycosylases. The encoded enzyme is primarily involved in DNA repair by cleaving oxidatively damaged bases and introducing a DNA strand break via its abasic site lyase activity (Dou et al., 2003; Hazra et al., 2002). The recent meta-analysis of GWAS on PCOS, on which our SMR analysis was based, found that the genetic variant rs804279 in *GATA4/NEIL2* showed significant association with PCOS (OR=1.14, 95% CI: 1.10-1.18;  $P=3.76\times 10^{-12}$ ); however, significant heterogeneity was observed across the different studies (F. Day et al., 2018). This genetic variant also showed significant association with polycystic ovarian morphology and ovulatory dysfunction (F. Day et al., 2018). Another genetic variant rs8191514 in *NEIL2* was predicted to generate a binding site for twenty transcription factors, and it was in linkage disequilibrium with the PCOS-identified genetic variant rs804279 ( $R^2=0.4$ ,  $D'$ =0.97;  $P<0.0001$ ) (Prabhu et al., 2021). Moreover, 8p23.1, where the *NEIL2* is located, is also the region of *GATA4* (GATA Binding Protein 4), and knock-out of *GATA4* led to abnormal responses to exogenous gonadotropins and impaired fertility in mice (LaVoie, 2014). It also encompasses the promoter region of *FDFT1* (Farnesyl-Diphosphate Farnesyltransferase 1), a gene encoding farnesyl-diphosphate farnesyl transferase that is involved in cholesterol biosynthesis pathway (Ha and Lee, 2020), thereby influencing testosterone biosynthesis. These findings, together with ours, highlight the importance of this region in association with PCOS.

Our study has some limitations. The number of probes used in our SMR analyses was limited, and the sample size in the eQTL analysis was limited, especially for the eQTL data in ovary, which may lead to reduced power in the eQTL analysis. Consequently, we may have missed some important genes implicated in PCOS. Future SMR studies with larger samples for the eQTL analysis are warranted to identify additional genes

underlying the pathogenesis of PCOS. We only performed SMR analyses in participants of European ethnicity, and future studies are needed to explore whether our findings can be generalized to other ethnicities. Moreover, we only explored gene expression probes in association with PCOS, and it is possible that genetic variants exert their effect on PCOS through other epigenetic mechanisms, such as DNA methylation. More studies integrating multi-omics data are needed to more systematically explore the complex mechanisms underpinning PCOS.

In summary, our SMR study integrating GWAS of PCOS and eQTL data revealed multiple genes were potentially involved in the pathogenesis of PCOS. Some of the genes were reported to be involved in the regulation of T-cell survival, energy homeostasis, and DNA repair. More studies are needed to examine the exact functions of these genes in the etiology of PCOS and to explore additional genes implicated in the pathogenesis of PCOS.

## **ABBREVIATIONS**

BPA, Bisphenol A; eQTL, expression quantitative trait loci; GWAS, genome-wide association study; HEIDI, heterogeneity in dependent instruments; IR, insulin resistance; MR, mendelian randomization; PCOS, polycystic ovary syndrome; QoL, quality of life; SMR, summary data-based Mendelian randomization

## **ACKNOWLEDGEMENTS**

This work was supported by Lianyungang Health and Family Planning Science and Technology Project (No. 201706). Dr. Jingyun Yang's research was supported by

NIH/NIA grants [P30AG10161, R01AG15819, R01AG17917, R01AG36042, U01AG61356 and 1RF1AG064312-01].

## **AVAILABILITY OF DATA AND MATERIALS**

All data generated or analyzed during this study are publicly available as specified in the methods section of this paper. The eQTL data can be downloaded at <https://cnsgenomics.com/data/SMR/#eQTLsummarydata>, and the GWAS summarized data can be downloaded at <https://www.repository.cam.ac.uk/handle/1810/283491>.

## **COMPETING INTERESTS**

No potential conflicts of interest were disclosed by the authors.

## **AUTHORS' CONTRIBUTIONS**

QS, JY, WF and WY designed and registered the study. DL, JL and JY analyzed data and performed data interpretation. QS, YG, DL and JY wrote the initial draft and JY and WY contributed writing to subsequent versions of the manuscript. All authors reviewed the study findings and read and approved the final version before submission.

## REFERENCES

- Azziz R, Carmina E, Chen Z, et al.: Polycystic ovary syndrome, *Nat Rev Dis Primers* 2, 16057.
- Barnard L, Ferriday D, Guenther N, et al.: Quality of life and psychological well being in polycystic ovary syndrome, *Hum Reprod* 22, 2279-2286.
- Bellver J, Rodriguez-Taberner L, Robles A, et al.: Polycystic ovary syndrome throughout a woman's life, *J Assist Reprod Genet* 35, 25-39.
- Benson KK, Hu W, Weller AH, et al.: Natural human genetic variation determines basal and inducible expression of PM20D1, an obesity-associated gene, *Proc Natl Acad Sci U S A* 116, 23232-23242.
- Burgess S, Timpson NJ, Ebrahim S, et al.: Mendelian randomization: where are we now and where are we going?, *International Journal of Epidemiology* 44, 379-388.
- Censin JC, Bovijn J, Holmes MV, et al.: Colocalization analysis of polycystic ovary syndrome to identify potential disease-mediating genes and proteins, *Eur J Hum Genet*.
- Chen C, Peng J, Ma S, et al.: Ribosomal protein S26 serves as a checkpoint of T-cell survival and homeostasis in a p53-dependent manner, *Cell Mol Immunol*.
- Chen ZJ, Zhao H, He L, et al.: Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3, *Nat Genet* 43, 55-59.
- Condorelli RA, Calogero AE, Di Mauro M, et al.: PCOS and diabetes mellitus: from insulin resistance to altered beta pancreatic function, a link in evolution, *Gynecol Endocrinol* 33, 665-667.
- Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, et al.: Genetic effects on gene expression across human tissues, *Nature* 550, 204-213.
- Davey Smith G, and Hemani G: Mendelian randomization: genetic anchors for causal inference in epidemiological studies, *Hum Mol Genet* 23, R89-98.
- Day F, Karaderi T, Jones MR, et al.: Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria, *PLoS Genet* 14, e1007813.
- Day FR, Hinds DA, Tung JY, et al.: Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome, *Nat Commun* 6, 8464.
- Diao FY, Xu M, Hu Y, et al.: The molecular characteristics of polycystic ovary syndrome (PCOS) ovary defined by human ovary cDNA microarray, *J Mol Endocrinol* 33, 59-72.
- Dou H, Mitra S, and Hazra TK: Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2, *J Biol Chem* 278, 49679-49684.
- Filipenko ML, Vinichenko NA, Karpova GG, et al.: Isolation, structural analysis and mapping of the functional gene of human ribosomal protein S26, *Gene* 211, 287-292.
- Ha NT, and Lee CH: Roles of Farnesyl-Diphosphate Farnesyltransferase 1 in Tumour and Tumour Microenvironments, *Cells* 9.
- Hayes MG, Urbanek M, Ehrmann DA, et al.: Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations, *Nat Commun* 6, 7502.
- Hazra TK, Kow YW, Hatahet Z, et al.: Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions, *J Biol Chem* 277, 30417-30420.
- Joham AE, Teede HJ, Ranasinha S, et al.: Prevalence of infertility and use of fertility treatment in women with polycystic ovary syndrome: data from a large community-based cohort study, *J Womens Health (Larchmt)* 24, 299-307.

- Kandaraki E, Chatzigeorgiou A, Livadas S, et al.: Endocrine disruptors and polycystic ovary syndrome (PCOS): elevated serum levels of bisphenol A in women with PCOS, *J Clin Endocrinol Metab* 96, E480-484.
- LaVoie HA: The GATA-keepers of ovarian development and folliculogenesis, *Biol Reprod* 91, 38.
- Lee H, Oh JY, Sung YA, et al.: Genome-wide association study identified new susceptibility loci for polycystic ovary syndrome, *Hum Reprod* 30, 723-731.
- Legro RS, Arslanian SA, Ehrmann DA, et al.: Diagnosis and treatment of polycystic ovary syndrome: an Endocrine Society clinical practice guideline, *J Clin Endocrinol Metab* 98, 4565-4592.
- Liu D, Yang J, Feng B, et al.: Mendelian randomization analysis identified genes pleiotropically associated with the risk and prognosis of COVID-19, *J Infect*.
- Liu XM, Yan MQ, Ji SY, et al.: Loss of oocyte Rps26 in mice arrests oocyte growth and causes premature ovarian failure, *Cell Death Dis* 9, 1144.
- Long JZ, Roche AM, Berdan CA, et al.: Ablation of PM20D1 reveals N-acyl amino acid control of metabolism and nociception, *Proc Natl Acad Sci U S A* 115, E6937-E6945.
- Long JZ, Svensson KJ, Bateman LA, et al.: The Secreted Enzyme PM20D1 Regulates Lipidated Amino Acid Uncouplers of Mitochondria, *Cell* 166, 424-435.
- Naumova I, Castelo-Branco C, Kasterina I, et al.: Quality of Life in Infertile Women with Polycystic Ovary Syndrome: a Comparative Study, *Reprod Sci*.
- Naz MSG, Tehrani FR, Majd HA, et al.: The prevalence of polycystic ovary syndrome in adolescents: A systematic review and meta-analysis, *Int J Reprod Biomed* 17, 533-542.
- Pavlidis JM, Zhu Z, Gratten J, et al.: Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits, *Genome Med* 8, 84.
- Pihlstrom L, Rengmark A, Bjornara KA, et al.: Fine mapping and resequencing of the PARK16 locus in Parkinson's disease, *J Hum Genet* 60, 357-362.
- Prabhu BN, Kanchamreddy SH, Sharma AR, et al.: Conceptualization of functional single nucleotide polymorphisms of polycystic ovarian syndrome genes: an in silico approach, *J Endocrinol Invest*.
- Rotterdam EA-SPCWG: Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome, *Fertil Steril* 81, 19-25.
- Rotterdam EA-SPCwg: Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS), *Hum Reprod* 19, 41-47.
- Sanchez-Mut JV, Heyn H, Silva BA, et al.: PM20D1 is a quantitative trait locus associated with Alzheimer's disease, *Nat Med* 24, 598-603.
- Shi Y, Zhao H, Shi Y, et al.: Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome, *Nat Genet* 44, 1020-1025.
- Teede H, Deeks A, and Moran L: Polycystic ovary syndrome: a complex condition with psychological, reproductive and metabolic manifestations that impacts on health across the lifespan, *BMC Med* 8, 41.
- Thanassoulis G, and O'Donnell CJ: Mendelian randomization: nature's randomized trial in the post-genome era, *JAMA* 301, 2386-2388.
- Vink JM, Sadrzadeh S, Lambalk CB, et al.: Heritability of polycystic ovary syndrome in a Dutch twin-family study, *J Clin Endocrinol Metab* 91, 2100-2104.
- Visscher PM, Brown MA, McCarthy MI, et al.: Five years of GWAS discovery, *Am J Hum Genet* 90, 7-24.
- Wilson NA, and Pena AS: Quality of life in adolescent girls with polycystic ovary syndrome, *J Paediatr Child Health* 56, 1351-1357.



- Wray NR, Ripke S, Mattheisen M, et al.: Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression, *Nat Genet* 50, 668-681.
- Zawadzki JK, and Dunaif A: Polycystic ovary syndrome Cambridge, 1992, Blackwell Scientific Publications.
- Zhang B, Zhou W, Shi Y, et al.: Lifestyle and environmental contributions to ovulatory dysfunction in women of polycystic ovary syndrome, *BMC Endocr Disord* 20, 19.
- Zhao H, Lv Y, Li L, et al.: Genetic Studies on Polycystic Ovary Syndrome, *Best Pract Res Clin Obstet Gynaecol* 37, 56-65.
- Zhu Z, Zhang F, Hu H, et al.: Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets, *Nat Genet* 48, 481-487.

## FIGURE CAPTIONS

### **Figure 1. Pleiotropic association of *RPS26* with PCOS.**

A) SMR analysis results using eQTL data for ovary; B) SMR analysis results using eQTL data for whole blood

Top plot, grey dots represent the  $-\log_{10}(P)$  values for SNPs from the GWAS of PCOS, with solid rhombuses indicating that the probes pass HEIDI test. Middle plot, eQTL results. Bottom plot, location of genes tagged by the probes.

GWAS, genome-wide association studies; SMR, summary data-based Mendelian randomization; HEIDI, heterogeneity in dependent instruments; eQTL, expression quantitative trait loci; PCOS, polycystic ovary syndrome

### **Figure 2. Pleiotropic association of *PM20D1* with PCOS.**

A) SMR analysis results using eQTL data for ovary; B) SMR analysis results using eQTL data for whole blood

Top plot, grey dots represent the  $-\log_{10}(P)$  values for SNPs from the GWAS of PCOS, with solid rhombuses indicating that the probes pass HEIDI test. Middle plot, eQTL results. Bottom plot, location of genes tagged by the probes.

GWAS, genome-wide association studies; SMR, summary data-based Mendelian randomization; HEIDI, heterogeneity in dependent instruments; eQTL, expression quantitative trait loci; PCOS, polycystic ovary syndrome

### **Figure 3. Pleiotropic association of *CTC-457L16.2* with PCOS.**

A) SMR analysis results using eQTL data for ovary; B) SMR analysis results using eQTL data for whole blood

Top plot, grey dots represent the  $-\log_{10}(P)$  values for SNPs from the GWAS of PCOS, with solid rhombuses indicating that the probes pass HEIDI test. Middle plot, eQTL results. Bottom plot, location of genes tagged by the probes.

GWAS, genome-wide association studies; SMR, summary data-based Mendelian randomization; HEIDI, heterogeneity in dependent instruments; eQTL, expression quantitative trait loci; PCOS, polycystic ovary syndrome

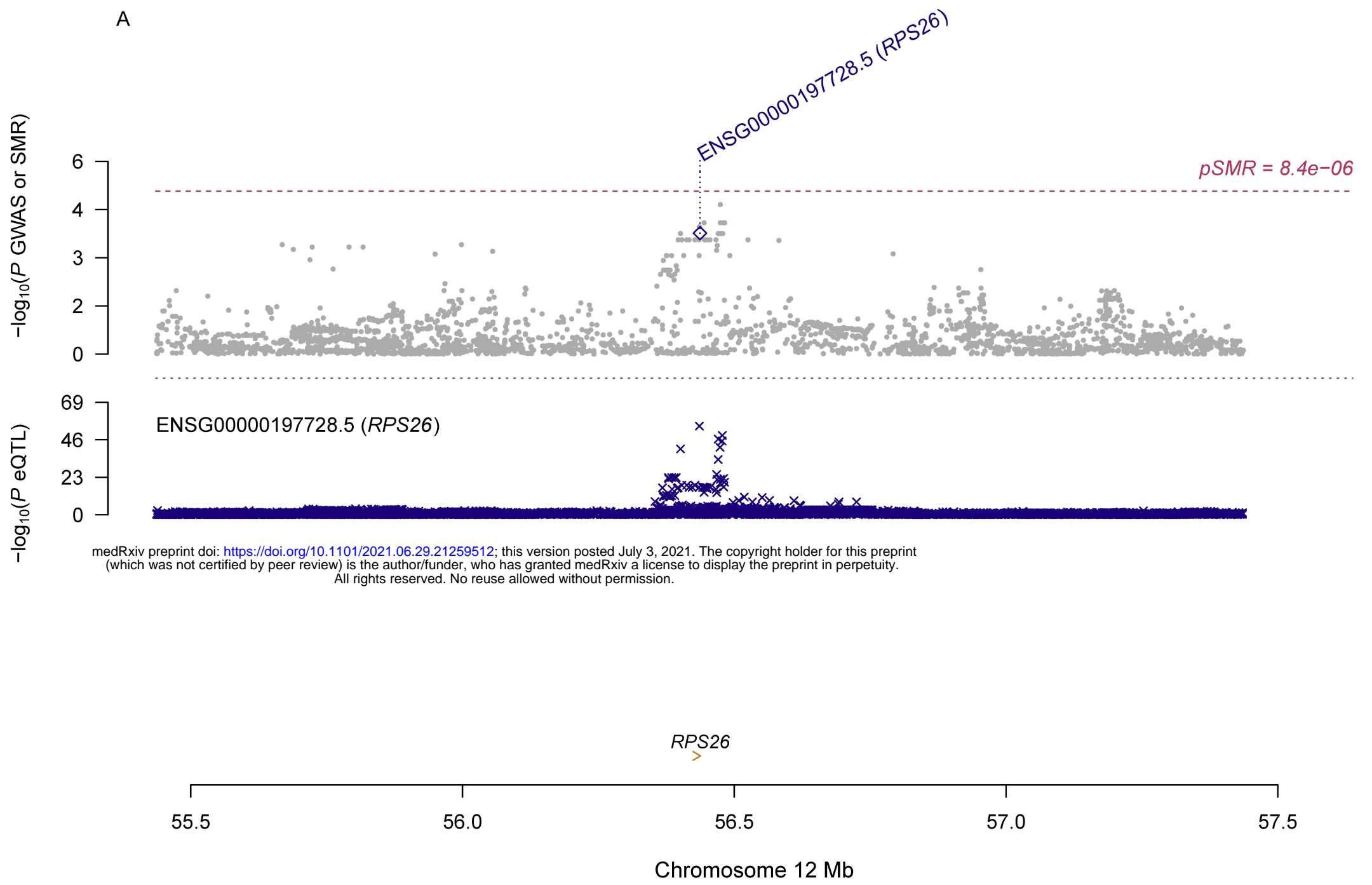
### **Figure 4. Pleiotropic association of *NEIL2* with PCOS.**

A) SMR analysis results using eQTL data for ovary; B) SMR analysis results using eQTL data for whole blood

Top plot, grey dots represent the  $-\log_{10}(P)$  values for SNPs from the GWAS of PCOS, with solid rhombuses indicating that the probes pass HEIDI test. Middle plot, eQTL results. Bottom plot, location of genes tagged by the probes.

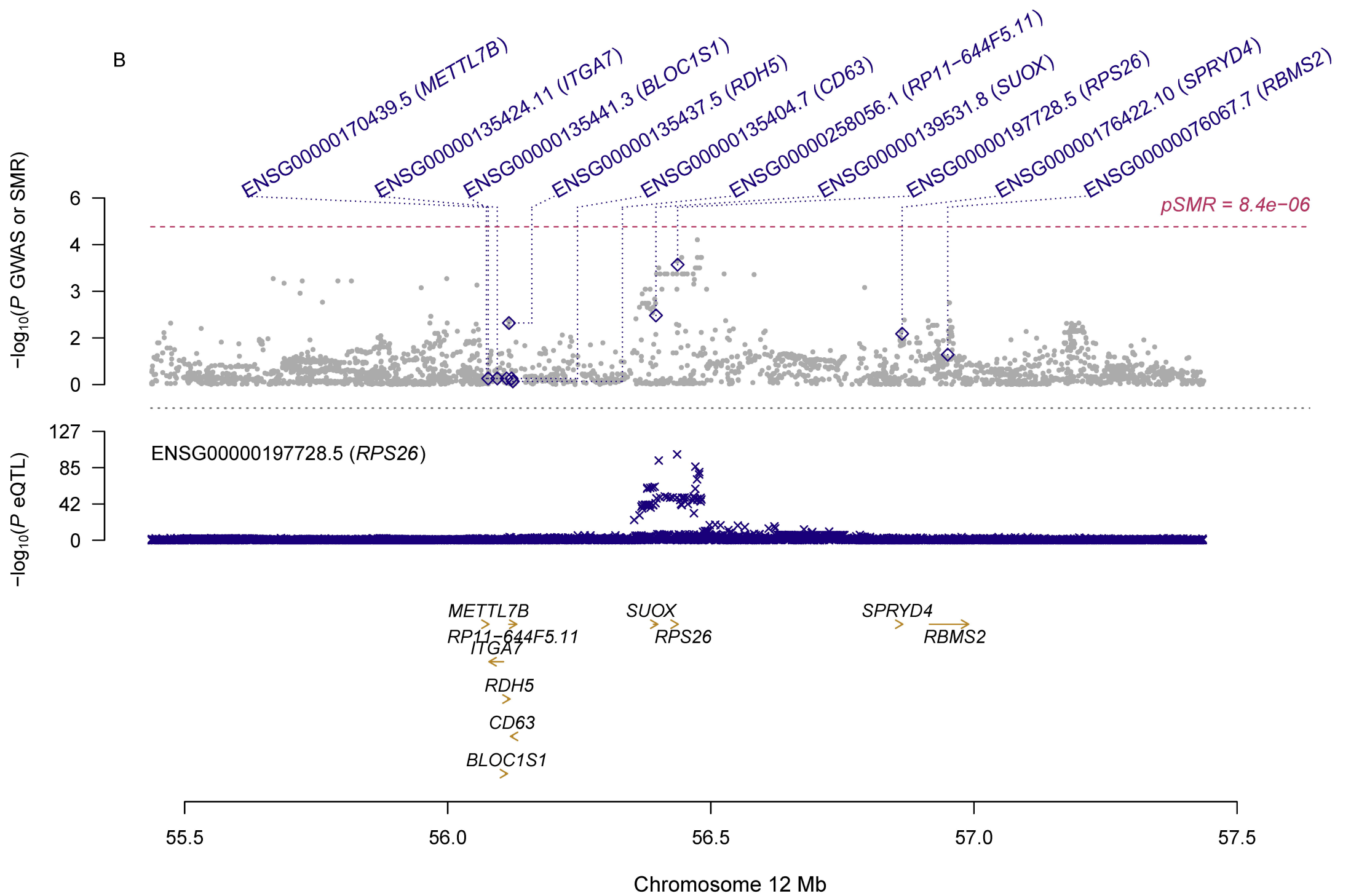
GWAS, genome-wide association studies; SMR, summary data-based Mendelian randomization; HEIDI, heterogeneity in dependent instruments; eQTL, expression quantitative trait loci; PCOS, polycystic ovary syndrome

A

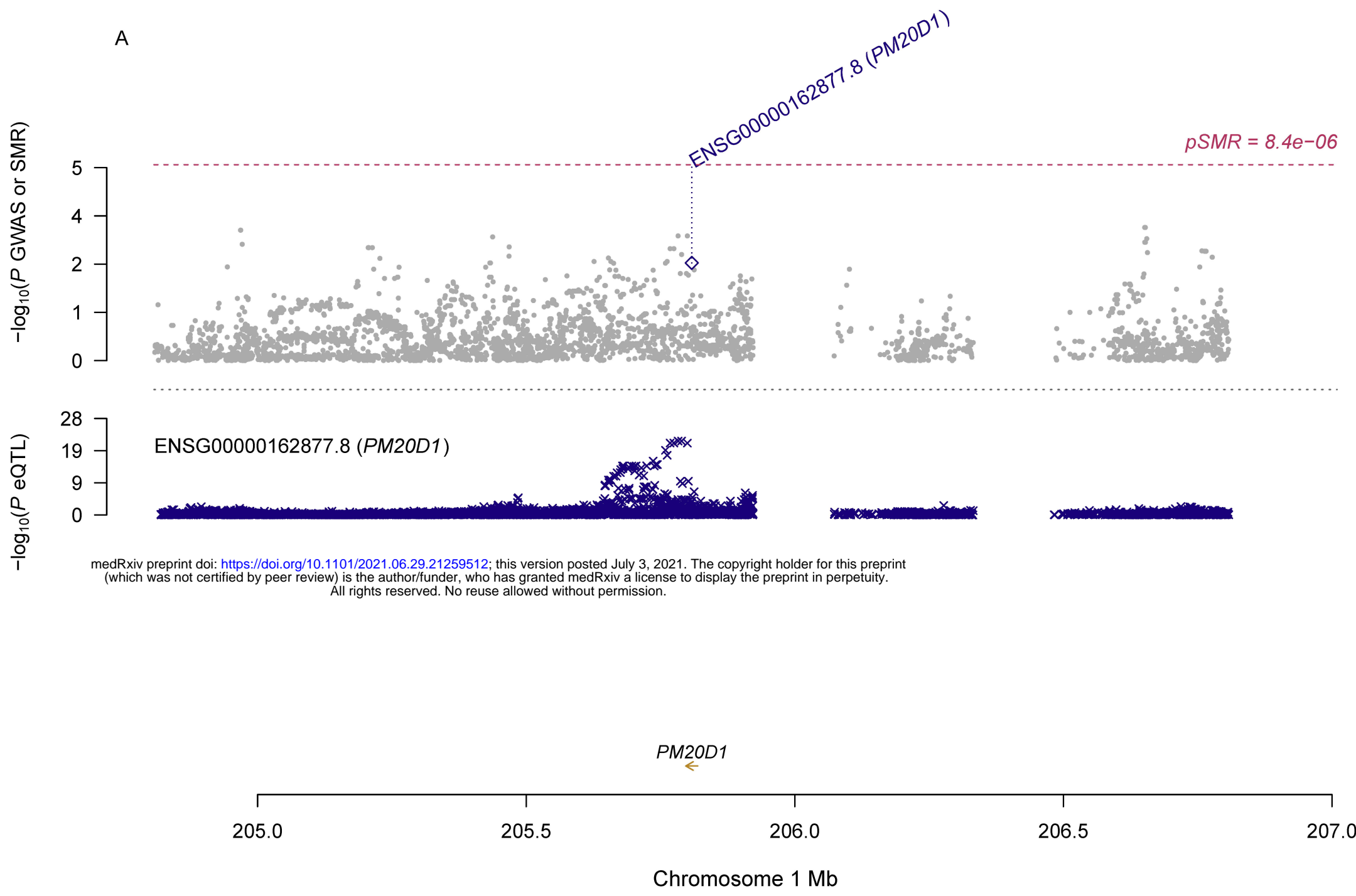


medRxiv preprint doi: <https://doi.org/10.1101/2021.06.29.21259512>; this version posted July 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

B

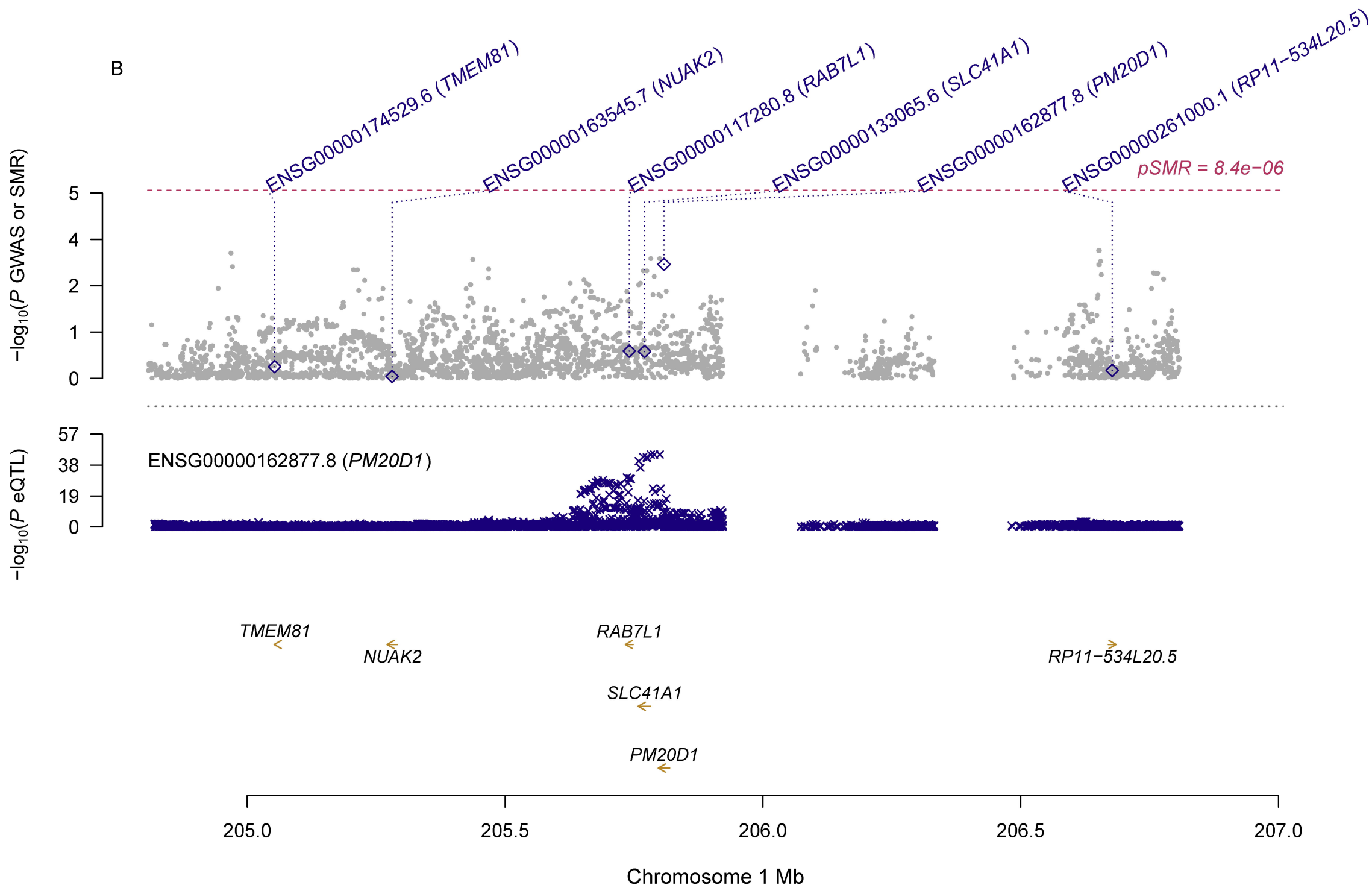


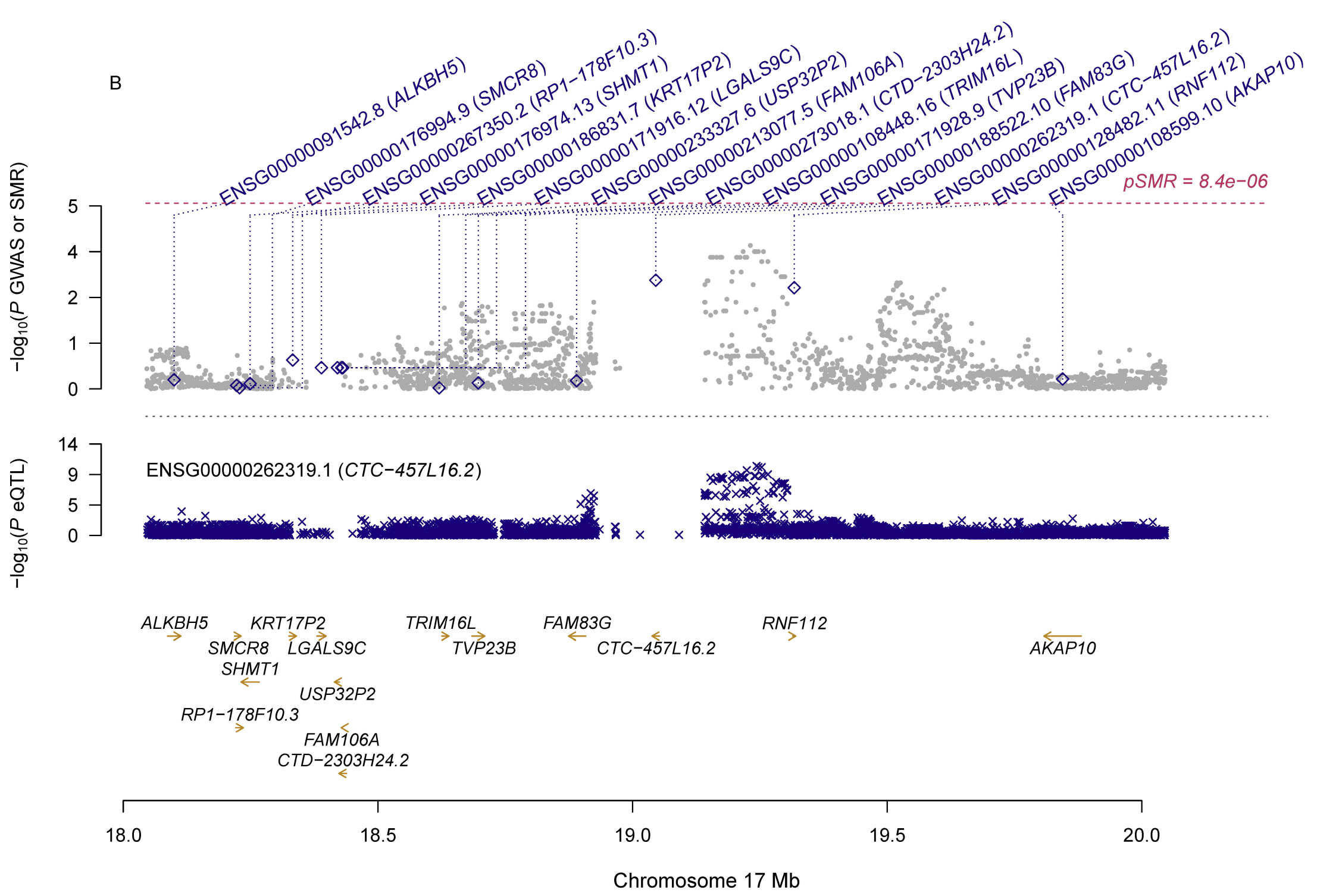
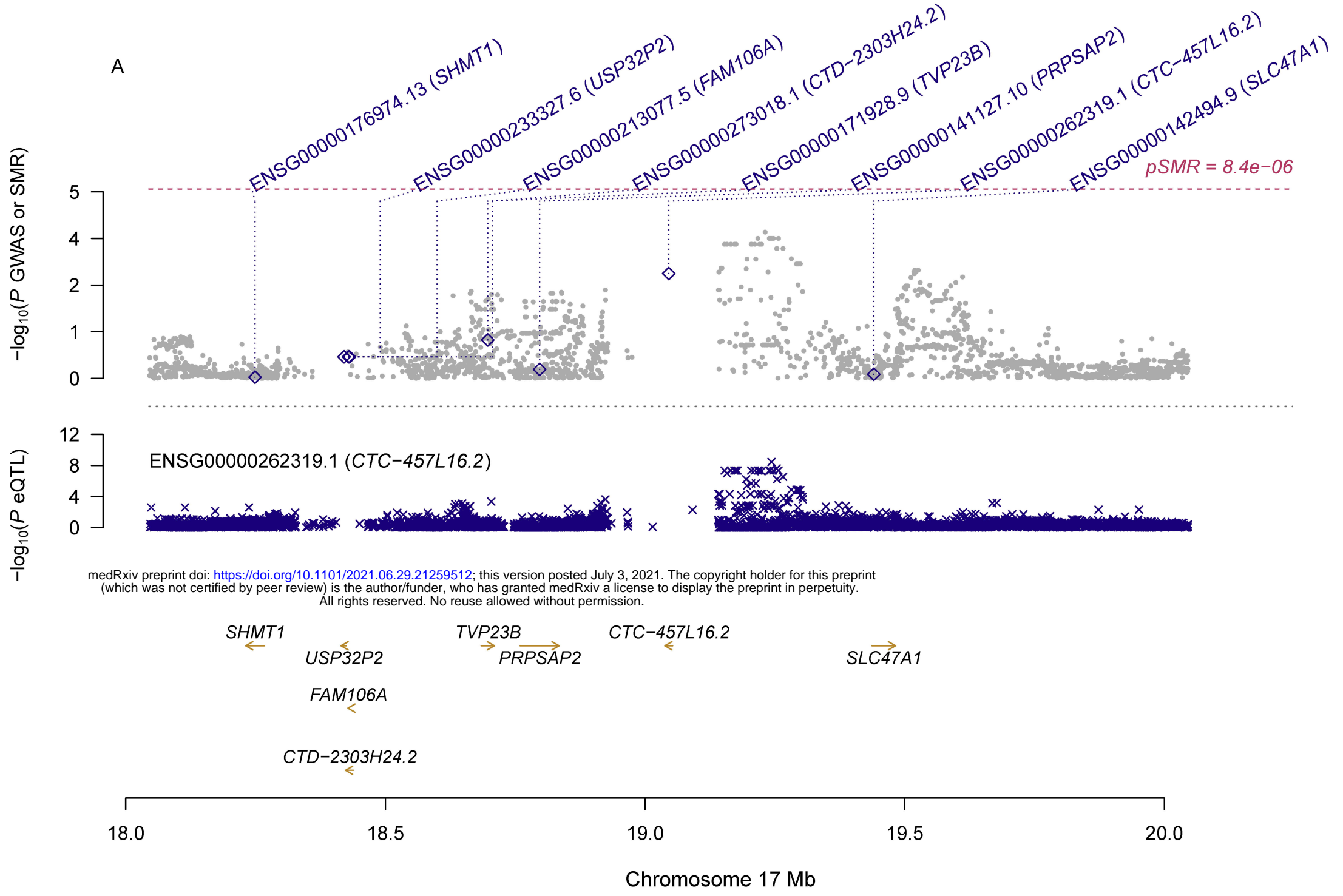
A



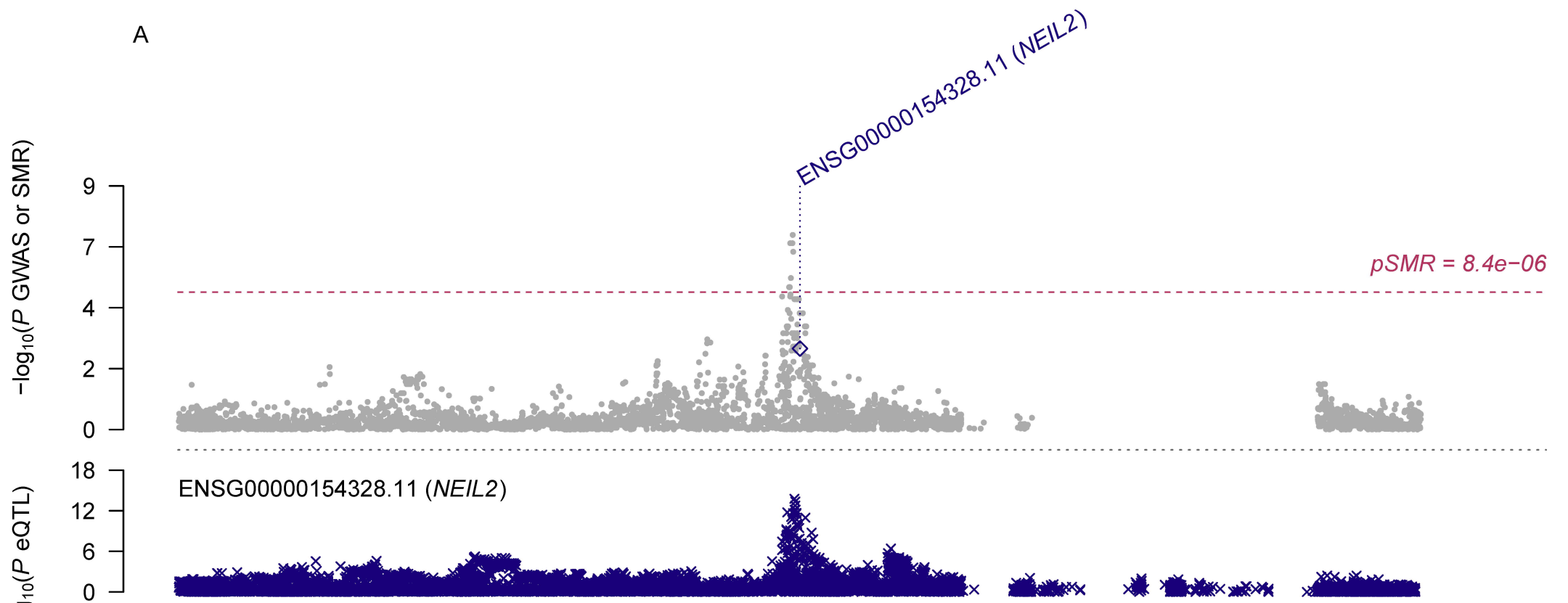
medRxiv preprint doi: <https://doi.org/10.1101/2021.06.29.21259512>; this version posted July 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

B

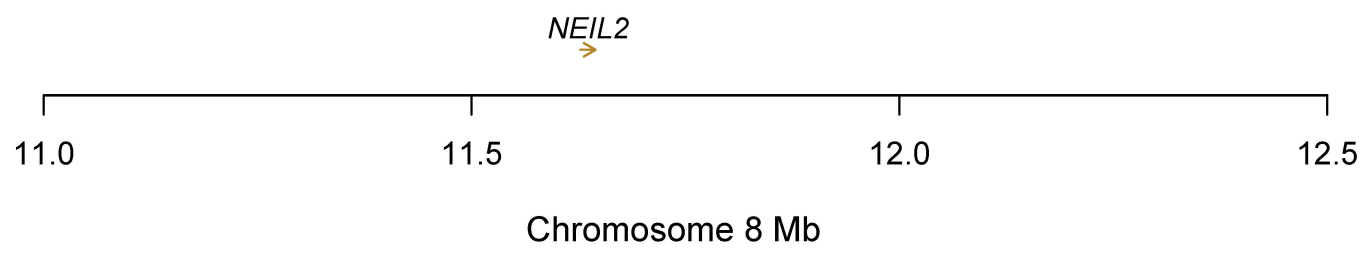




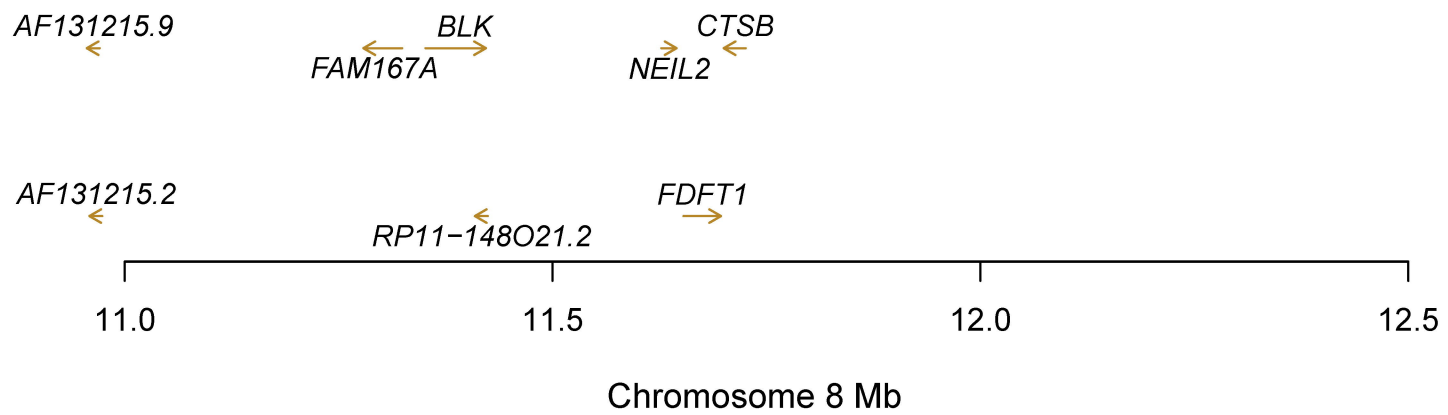
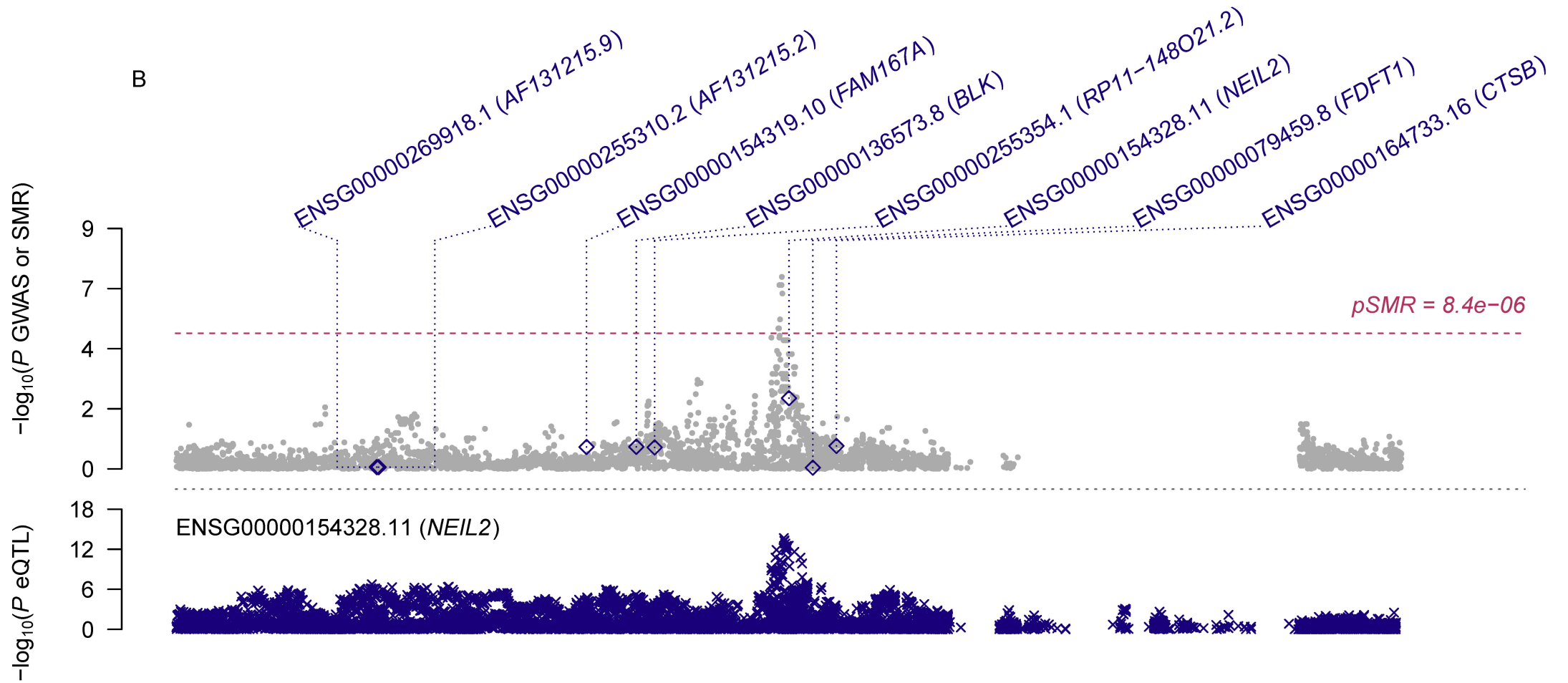
A



medRxiv preprint doi: <https://doi.org/10.1101/2021.06.29.21259512>; this version posted July 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



B



**Table 1. Basic information of the GWAS and eQTL data.**

<b>Data Source</b>	<b>Total number of participants</b>	<b>Number of eligible genetic variants or probes</b>
<b>eQTL data</b>		
<b>Ovary</b>	85	1,530
<b>Whole blood</b>	338	4,490
<b>GWAS data</b>	113,238	6,400,755

GWAS: genome-wide association studies; QTL, quantitative trait loci



**Table 2. The top ten probes identified in the SMR analysis.**

eQTL data	Probe ID	Gene	CHR	Top SNP	P <sub>eQTL</sub>	P <sub>GWAS</sub>	Beta	SE	P <sub>SMR</sub>	P <sub>HEIDI</sub>	N <sub>SNP</sub>
<b>Ovary</b>	ENSG00000197728.5	<i>RPS26</i>	12	rs1131017	3.89×10 <sup>-55</sup>	1.60×10 <sup>-4</sup>	0.10	0.03	1.72×10 <sup>-4</sup>	1.87×10 <sup>-1</sup>	20
	ENSG00000261390.1	<i>RP11-345M22.2</i>	16	rs7196490	1.02×10 <sup>-11</sup>	1.20×10 <sup>-5</sup>	-0.28	0.08	3.18×10 <sup>-4</sup>	8.51×10 <sup>-1</sup>	20
	ENSG00000154328.11	<i>NEIL2</i>	8	rs804271	1.71×10 <sup>-14</sup>	3.70×10 <sup>-4</sup>	-0.20	0.06	1.02×10 <sup>-3</sup>	1.87×10 <sup>-1</sup>	20
	ENSG00000225706.1	<i>RP11-75C9.1</i>	9	rs77301379	7.45×10 <sup>-20</sup>	5.30×10 <sup>-4</sup>	0.13	0.04	1.41×10 <sup>-3</sup>	8.71×10 <sup>-1</sup>	20
	ENSG00000262319.1	<i>CTC-457L16.2</i>	17	rs55885404	3.59×10 <sup>-9</sup>	1.80×10 <sup>-4</sup>	-0.20	0.06	1.55×10 <sup>-3</sup>	9.58×10 <sup>-1</sup>	12
	ENSG00000172000.3	<i>ZNF556</i>	19	rs4807349	2.34×10 <sup>-10</sup>	5.70×10 <sup>-4</sup>	-0.24	0.08	2.24×10 <sup>-3</sup>	1.84×10 <sup>-1</sup>	20
	ENSG00000162877.8	<i>PM20D1</i>	1	rs823080	3.53×10 <sup>-22</sup>	8.50×10 <sup>-4</sup>	0.10	0.03	2.94×10 <sup>-3</sup>	9.31×10 <sup>-1</sup>	20
	ENSG00000186300.7	<i>ZNF555</i>	19	rs917653	3.31×10 <sup>-12</sup>	8.30×10 <sup>-4</sup>	-0.15	0.05	3.18×10 <sup>-3</sup>	8.66×10 <sup>-1</sup>	17
	ENSG00000165623.5	<i>UCMA</i>	10	rs78415540	1.12×10 <sup>-20</sup>	3.80×10 <sup>-3</sup>	-0.09	0.03	3.69×10 <sup>-3</sup>	1.46×10 <sup>-1</sup>	20
	ENSG00000100211.6	<i>CBY1</i>	22	rs62229988	3.09×10 <sup>-13</sup>	2.50×10 <sup>-3</sup>	0.21	0.07	3.90×10 <sup>-3</sup>	6.00×10 <sup>-1</sup>	20
<b>Whole blood</b>	ENSG00000197728.5	<i>RPS26</i>	12	rs1131017	5.82×10 <sup>-101</sup>	1.60×10 <sup>-4</sup>	0.11	0.03	1.40×10 <sup>-4</sup>	3.60×10 <sup>-1</sup>	20
	ENSG00000162877.8	<i>PM20D1</i>	1	rs1772143	3.17×10 <sup>-45</sup>	5.20×10 <sup>-4</sup>	0.14	0.04	8.38×10 <sup>-4</sup>	8.78×10 <sup>-1</sup>	20
	ENSG00000262319.1	<i>CTC-457L16.2</i>	17	rs55885404	2.28×10 <sup>-11</sup>	1.80×10 <sup>-4</sup>	-0.31	0.09	1.07×10 <sup>-3</sup>	9.96×10 <sup>-1</sup>	16
	ENSG00000100201.14	<i>DDX17</i>	22	rs2267390	5.13×10 <sup>-18</sup>	4.90×10 <sup>-4</sup>	-0.77	0.24	1.08×10 <sup>-3</sup>	3.23×10 <sup>-2</sup>	20
	ENSG00000127483.13	<i>HP1BP3</i>	1	rs9426742	2.54×10 <sup>-16</sup>	6.90×10 <sup>-4</sup>	0.60	0.19	1.53×10 <sup>-3</sup>	9.17×10 <sup>-1</sup>	20
	ENSG00000242550.1	<i>SERPINB10</i>	18	rs9955526	4.76×10 <sup>-37</sup>	1.00×10 <sup>-3</sup>	0.22	0.07	1.64×10 <sup>-3</sup>	1.11×10 <sup>-1</sup>	20
	ENSG00000128482.11	<i>RNF112</i>	17	rs7502682	2.34×10 <sup>-20</sup>	9.90×10 <sup>-4</sup>	0.23	0.07	1.71×10 <sup>-3</sup>	2.58×10 <sup>-1</sup>	20
	ENSG00000254415.3	<i>SIGLEC14</i>	19	rs872629	2.70×10 <sup>-30</sup>	1.20×10 <sup>-3</sup>	0.20	0.07	1.97×10 <sup>-3</sup>	2.71×10 <sup>-1</sup>	20
	ENSG00000154328.11	<i>NEIL2</i>	8	rs804270	2.11×10 <sup>-14</sup>	1.00×10 <sup>-3</sup>	-0.30	0.10	2.25×10 <sup>-3</sup>	1.33×10 <sup>-1</sup>	20
ENSG00000105501.7	<i>SIGLEC5</i>	19	rs872629	7.55×10 <sup>-14</sup>	1.20×10 <sup>-3</sup>	0.52	0.18	3.15×10 <sup>-3</sup>	1.83×10 <sup>-1</sup>	18	

The GWAS summarized data were provided by the study of Shungin et al. and can be downloaded at <https://www.repository.cam.ac.uk/handle/1810/283491>. The GTE<sub>x</sub> eQTL data can be downloaded at <https://cnsgenomics.com/data/SMR/#eQTLsummarydata>.

P<sub>eQTL</sub> is the P-value of the top associated cis-eQTL in the eQTL analysis, and P<sub>GWAS</sub> is the P-value for the top associated cis-eQTL in the GWAS analysis. Beta is the estimated effect size in SMR analysis, SE is the corresponding standard error, P<sub>SMR</sub> is the P-value for SMR analysis, P<sub>HEIDI</sub> is

the P-value for the HEIDI test and  $N_{\text{snp}}$  is the number of SNPs involved in the HEIDI test. FDR was calculated at  $P=10^{-3}$  threshold.

**Bold font** means statistical significance after correction for multiple testing using FDR.

CHR, chromosome; HEIDI, heterogeneity in dependent instruments; SNP, single-nucleotide polymorphism; SMR, summary data-based Mendelian randomization; QTL, quantitative trait loci; FDR, false discovery rate; GWAS, genome-wide association studies