Appendix 1

Circulating white blood cell traits and small cell lung cancer risk: a Mendelian randomisation study—the statistical code

UVMR:

```
library(ggplot2)
library(data.table)
library(dplyr)
library(forestRwQ)
mydata <- fread("MVMR chemokins.csv") %>% as.data.frame()
mydata <- mydata[,c("exposure", "P-value","OR","LCI95","UCI95","se")]
mydata$pva1 <- round(mydata$pva1 ,4)
mydata$se <- round(mydatasse ,4)
head(mydata)
forestRwQ(
  left_dataframe = mydata[, 1:4],
  OR = mydata$OR,
  ci_low = mydata$LCI95,
  ci_high = mydata$UCI95,
  #piot width= 30
  oR_precision = 3,
  xlim=c(0.6),
  #right_dataframe =mydata[, 8, drop=F],
  # head_text ="finngen_R9"
  # head position = c(-0.15,0,0,0),
  # group_column = mydata$method,
  #signif_column = mydata$pval,
  #signif_threshold = 0.05,
  #save_format= "png"
   )
```

MVMR:

```
library(TwoSampleMR)
id_exposure
<-c("ebi-a-GCST90001561","ebi-a-GCST90001984","ebi-a-GCST90002061","ebi-a-GCST90002104")
id_outcome <- "ebi-a-GCST004746"
exposure_dat <- mv_extract_exposures(id_exposure)
dim(exposure_dat)
outcome_dat <- extract_outcome_data(exposure_dat$SNP, id_outcome)
mvdat <- mv_harmonise_data(exposure_dat,  outcome_dat)
res <- mv_multiple(mvdat)
res_OR<-generate_odds_ratios(res$result)
res_OR
write.csv(exposure_dat,file='exposure_dat.csv')
write.csv(mvdat,file='mvdat.csv')
write.csv(outcome_dat,file='outcome_dat.csv')
write.csv(res,file='res.csv')
write.csv(res_OR,file='res_OR.csv')
```


https://dx.doi.org/10.21037/tcr-24-1211

Abstract

Introduction

Materials and methods

IEU database
Outcom
: ebi-a-GCST004746
CD14+ CD16+ monocyte: ebi-a-GCST90001580
CX3CR1 on CD14+ CD16+ monocyte: ebi-a-GCST90001996
CD64 on CD14+ CD16+ monocyte: ebi-a-GCST90002011
CD14- CD16+ monocyte: ebi-a-GCST90001584
HLA DR on CD14- CD16+ monocyte: ebi-a-GCST90001984
CD14- CD16- Absolute Count ebi-a-GCST90001581
CCR2 on CD14- CD16-monocyte: ebi-a-GCST90002003
Eosinophil counts ebi-a-GCST90002299
White blood cell count (lymphocyte) ebi-a-GCST90018742
White blood cell count (basophil) ebi-a-GCST90025997
CD4 on CD39+ CD4+ T cell: ebi-a-GCST90002061
CD45RA+ CD8+ T cell : ebi-a-GCST90001561
HLA DR on myeloid Dendritic Cell: ebi-a-GCST90002104
P-selectin glycoprotein ligand 1 levels: ebi-a-GCST90010165
C-C motif chemokine 3 levels: ebi-a-GCST90012055
C-X-C motif chemokine 16 levels: ebi-a-GCST90012059
DNA methylation Hannum age acceleration: ebi-a-GCST90014289

Study design

This study is based on three core assumptions: (1) Genetic variations should be associated with exposure; (2) Genetic variations should be independent of confounding factors; (3) Genetic variations should only affect the outcome through exposure. We used the endometriosis genome-wide association study (GWAS) dataset (finngen_R9_N14_ENDOMETRIOSIS) as the exposure factor. The population selection, gene genotyping, and relevant baseline data involving GWAS data have been previously reported in other studies. Data collection was approved by the original GWAS ethics committee, and this MR study follows the guidelines of the Strengthening the Reporting of Observational Studies in Epidemiology for Mendelian Randomization (STROBE-MR).

IBD-related GWAS meta-analysis

To investigate the relationship between endometriosis and the risk of developing IBD in the UK, we conducted a search for IBD-related GWAS data. The study included six GWAS datasets related to IBD: benign rectal tumor, finn-b-CD2_BENIGN_RECTUM; IBD, ieu-a-31; CD, ieu-a-12; CD, ieu-a-30; UC, ieu-a-32; UC, ieu-a-973. Among these, finn-b-CD2_BENIGN_RECTUM included 2,108 cases, 216,684 controls, and 16,380,466 SNP loci; ieu-a-31 included 12,882 cases, 21,770 controls, and 12,716,084 SNP loci; ieu-a-12 included 17,897 cases, 33,977 controls, and 1,248,880 SNP loci; ieu-a-30 included 5,956 cases, 14,927 controls, and 12,276,506 SNP loci; ieu-a-32 included 6,968 cases, 20,464 controls, and 12,255,197 SNP loci; ieu-a-973 included 6,687 cases, 19,718 controls, and 1,243,971 SNP loci.

Mendelian randomization estimation

Candidate instrumental variables for the two-sample MR study were selected from SNPs associated with endometriosis ($P<5e-06$). Subsequently, a clumping procedure was applied with $r^2=0.01$ and a clumping window of 10 Mb to remove linkage disequilibrium variants from the instrumental variables. The selection and quality control of instrumental variables were

https://dx.doi.org/10.21037/tcr-24-1211

computed using TwoSampleMR (v 0.5.6). Finally, causal relationships between endometriosis and IBD occurrence were analyzed using inverse variance weighting (IVW), MR-Egger, and weighted median (WM) methods. IVW estimates the causal effect of genes on the disease through weighted averaging; MR-Egger estimates the causal effect of genes on the disease by fitting a linear regression model and detects and corrects for genetic bias through Egger regression; WM provides robust estimates in the presence of genetic bias. Additionally, Cochrane's Q statistic was used to assess heterogeneity, and outliers were removed if detected, followed by re-evaluation of MR causal relationships. MR-PRESSO tested for horizontal pleiotropy and provided corrected estimates. Statistical analysis and data visualization were performed in R software version 4.1.3.

https://dx.doi.org/10.21037/tcr-24-1211

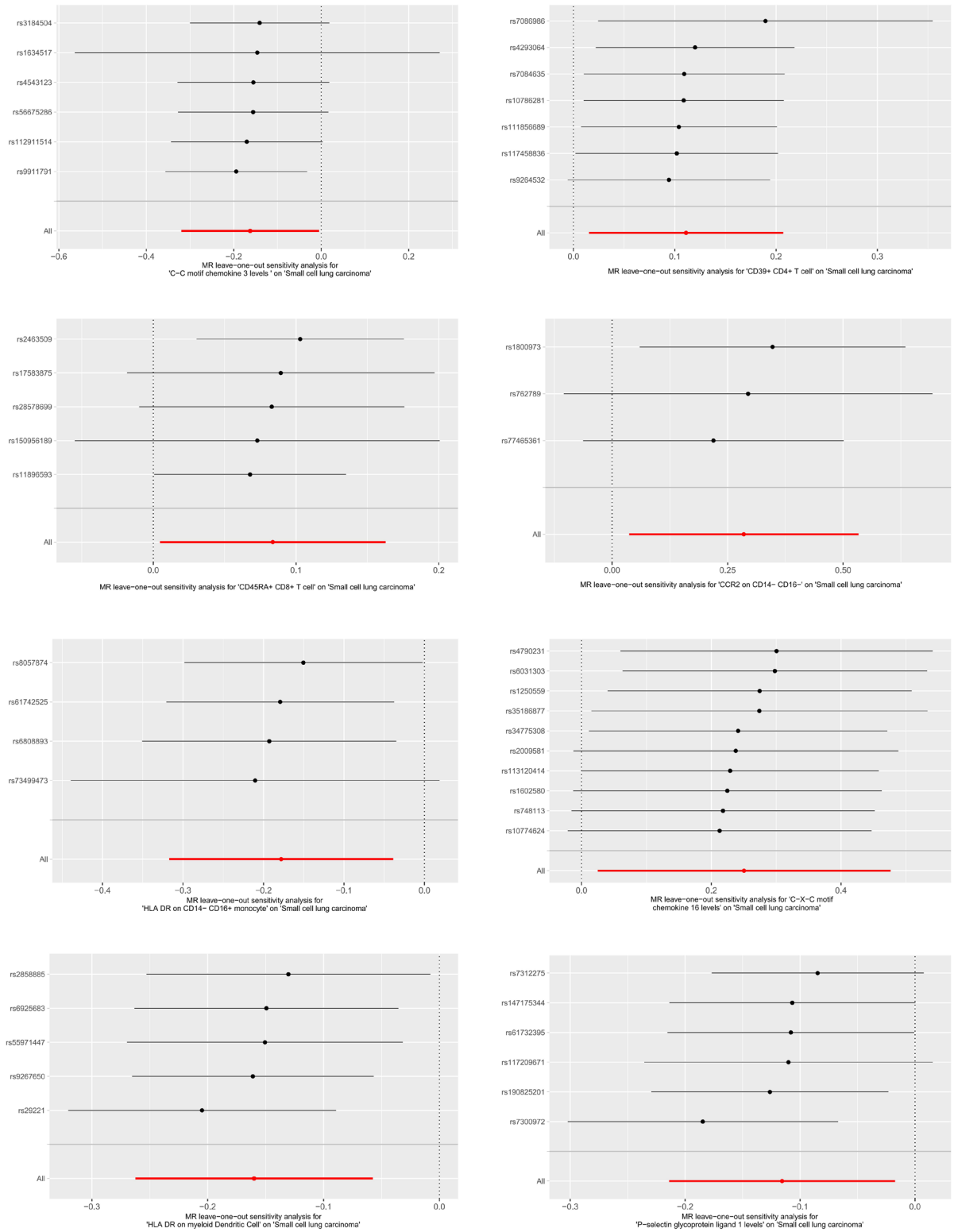

Figure S1 Forest plots for MR leave-one-out analysis between WBCs, chemokines and SCLC risk. Within each panel, the black points represent the causal estimate after discarding each SP in turn. Red points represent the pooled IVW estimates. Horizontal lines denote 95% confidence intervals. MR, Mendelian randomization; WBC, white blood cell; SCLC, small cell lung cancer.

https://dx.doi.org/10.21037/tcr-24-1211

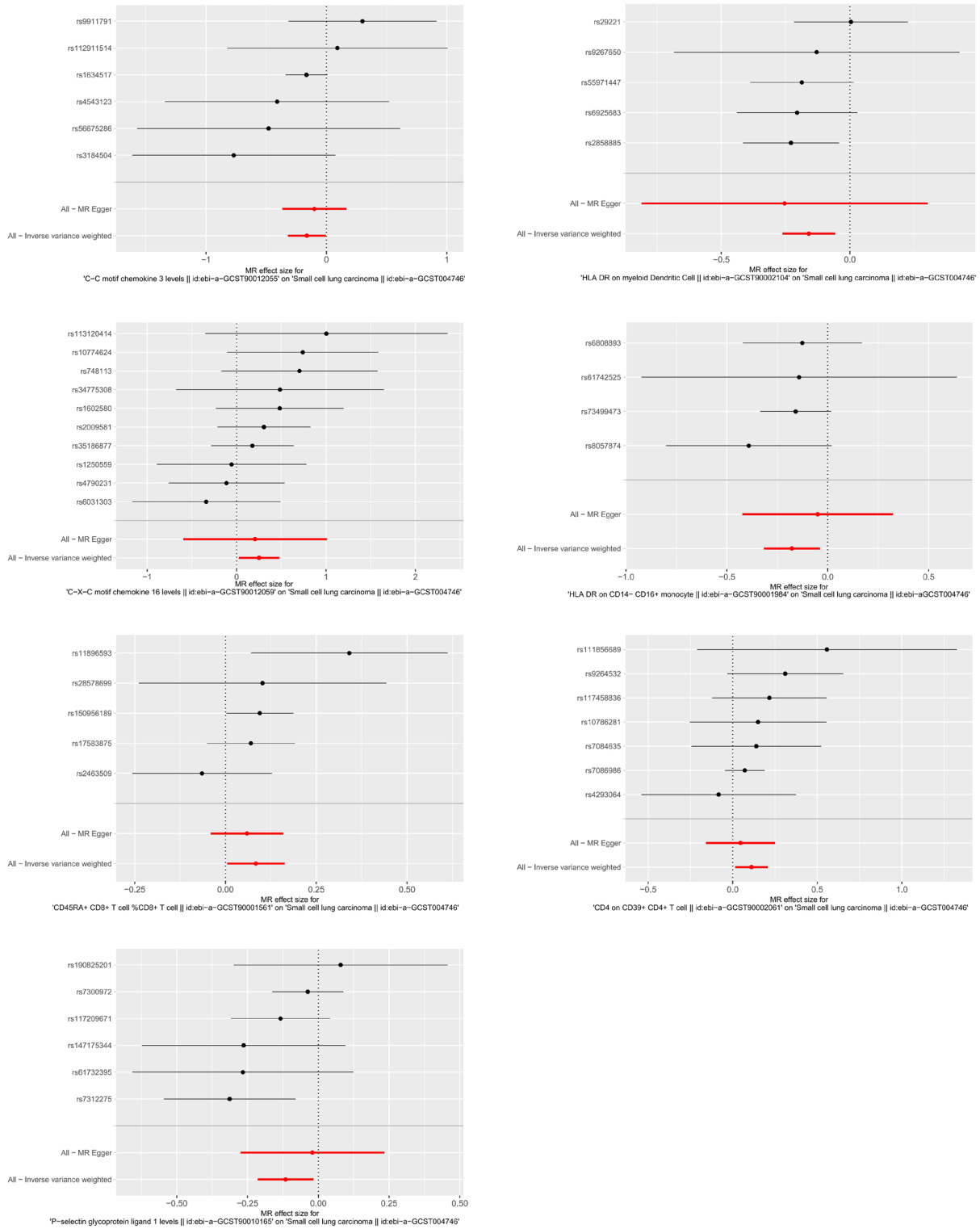

Figure S2 Forest plots for the exposure of circulating WBCs and chemokines on SCLC. WBC, white blood cell; SCLC, small cell lung cancer.

https://dx.doi.org/10.21037/tcr-24-1211

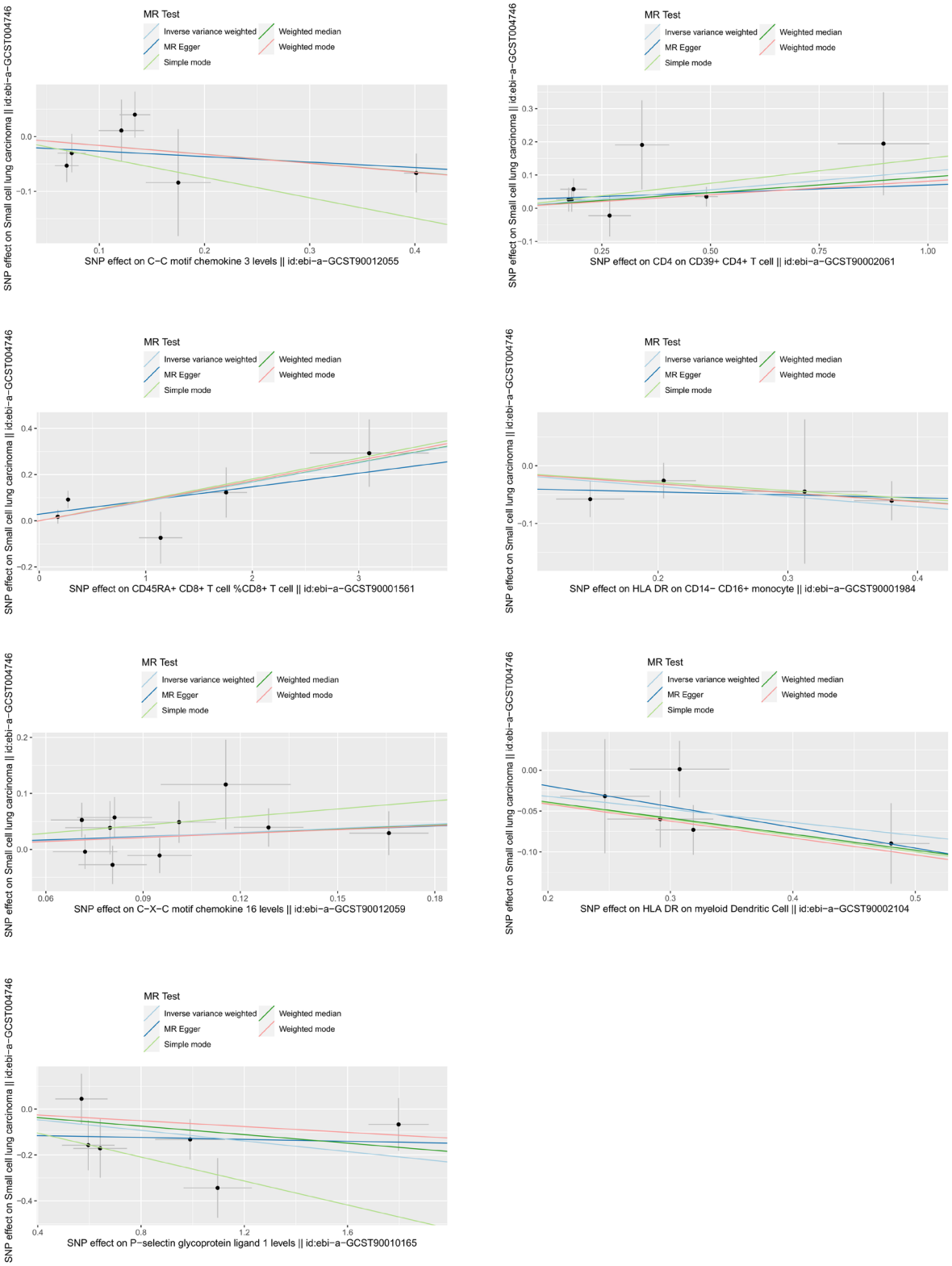

Figure S3 SNP effect on circulating WBCs and chemokines with SCLC. Each point on the scatter plot represents an SNP, with the horizontal coordinate being the SNP effect on exposure and the vertical coordinate being the SNP effect on SCLC. SNP, single nucleotide polymorphism; WBC, white blood cell; SCLC, small cell lung cancer.

https://dx.doi.org/10.21037/tcr-24-1211

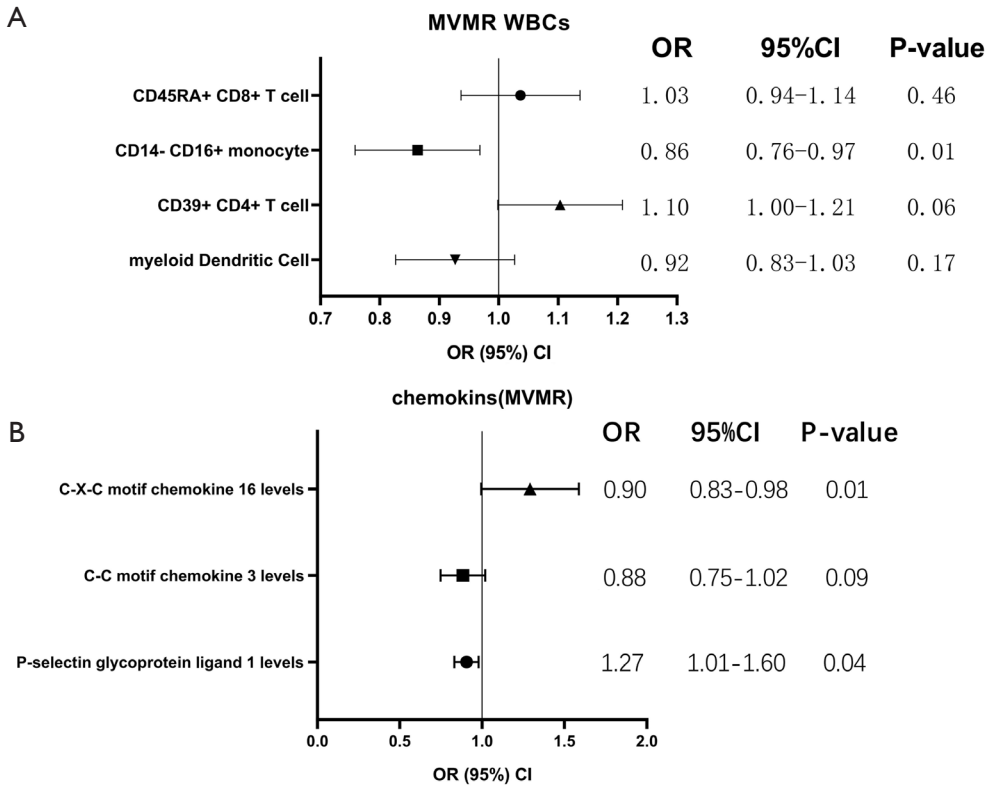

Figure S4 MVMR results for the potential relationships of circulating WBCs and chemokines with SCLC. MVMR, multivariable Mendelian randomization; WBC, white blood cell; SCLC, small cell lung cancer; OR, odds ratio; CI, confidence interval.

https://dx.doi.org/10.21037/tcr-24-1211