Journal of Cardiothoracic Surgery

RESEARCH

Open Access

# A causal association between chemokines and the risk of lung cancer: a univariate and multivariate mendelian randomization study

Mengmeng Wang[1,2†], Mingjun Gao[1,2†], Wenbo He[3†], Siding Zhou[3], Yusheng Shu[4*] and Xiaolin Wang[4*]

## Abstract

**Background** Observational studies and experimental evidence have shown that chemokines play important roles in lung cancer development, progression, and treatment. However, few studies have examined the causal association between them.

**Methods** Summary data of chemokines and lung cancer were obtained from genome-wide association studies. Mendelian randomization (MR) analyses were performed by five methods, Inverse variance weighted (IVW), Weighted median estimation, MR-Egger, Simple mode and Weighted, with IVW as the primary analysis method. Sensitivity analysis was used to assess the reliability of the MR results. Multivariate Mendelian randomization studies were used to infer whether causality was influenced by potential mediators. The expression levels of CCL21 were analyzed by quantitative real-time PCR.

**Results** We found that CCL21 was negatively associated with lung adenocarcinoma risk. CCL25 was positively associated with lung squamous cell carcinoma risk. CCL5 was negatively associated with small cell lung cancer risk. CCL21, CCL24, CCL27, and CCL28 was positively associated with small cell lung cancer risk. After multivariate Mendelian randomization adjustment for smoking behavior, it was found that the effect of CCL25 on lung squamous cell cancer disappeared, and the effect of CCL21 on small cell lung cancer was quite opposite to the univariate. The receiver operating characteristic curve indicated that chemokines had high accuracy in the diagnosis of lung cancer. CCL21 expression levels showed large differences in lung adenocarcinoma cells.

**Conclusion** These results highlighted the causal effects of chemokines on lung cancer and suggested a mediating role of smoking behavior in the association between chemokines and lung cancer.

†Mengmeng Wang, Mingjun Gao and Wenbo He contributed equally to this work.

*Correspondence:
Yusheng Shu
18051061999@yzu.edu.cn
Xiaolin Wang
18051063909@yzu.edu.cn

Full list of author information is available at the end of the article

**Keywords** Chemokines, Lung cancer, Genome-wide association studies, Causal association, Mendelian randomization study

## Introduction

Lung cancer (LC) is one of the most common malignancies in the world and has become one of the leading causes of malignant death annually worldwide [1–3]. Statistics show that in 2020, there were about 2.2 million new cases of LC and about 1.8 million deaths worldwide, and its incidence and mortality rates rank first in the world [4]. Smoking remains the major risk factor for LC, and other non-tobacco factors include occupational exposure, air pollution, chronic lung disease, and infection [5]. While the evidence for a causal effect of inflammatory carcinogenesis remains limited [6, 7]. To prevent the incidence of LC and find potential therapeutic targets, we need a better understanding of the role of inflammatory biomarkers in LC.

Chronic inflammation may play an important role in the occurrence, development, and prognosis of LC [8], in which chemokines may regulate the migration of immune cells to tumor cells. Studies have shown that some important chemokines can be used not only as prognostic biomarkers for LC, but also as modulators in cancer immunotherapy [9, 10]. However, traditional observational studies are susceptible to reverse causation and confounding factors [11]. Analysis of chemokine expression levels and prognosis of chemokines using bioinformatics tools may lead to inconsistent results. Therefore, a complete understanding of the causal association between chemokines and LC remains a challenge.

Mendelian randomization (MR) uses genetic variation as an instrumental variable to infer causal association between exposure and outcome, where genetic variation and outcome are unaffected by confounding factors, as a novel research method to infer causal association between chemokines and LC [12, 13]. In this study, we performed univariate and multivariate MR analyses using summary statistics from genome-wide association studies (GWAS) to investigate the risk effects of chemokines on LC and histological subtypes, further providing more robust evidence of the potential role of chemokines in LC development.

## Method

### Research design

In this study, we used GWAS-derived chemokines as instrumental variables (IVs) to investigate the causal effects of chemokines on LC using two-sample MR [14]. The positive results were validated with additional LC samples. Subsequently, the effect of chemokines mediated by smoking behavior (including ever smoker, current cigarette smokers, light smokers, cigarettes smoked per day) on LC was assessed. Finally, it predicted the accuracy of the discovered chemokines to diagnose LC. A graphical overview of the study design is shown in Fig. 1.

### Data sources and instrumental variables

IVs for chemokines were determined from public GWAS summary statistics and did not require individual-level data [15]. After literature search and GWAS database screening, we identified 38 chemokines. GWAS summary data for LC were obtained from UK Biobank, Transdisciplinary Research in Cancer of the Lung (TRICL) and International Lung Cancer Consortium (ILCCO) [16]. All participants were of European descent. Association analyses were performed for lung cancer (2671 cases, 372016 controls), lung adenocarcinoma (LUAD) (3442 cases, 14894 controls), lung squamous cell carcinoma (LUSC) (3275 cases, 15038 controls), and small cell lung cancer (SCLC) (2791 cases, 20580 controls). Validation analyses were performed in another dataset, including lung cancer (11348 cases, 15861 controls), lung adenocarcinoma (11245 cases, 54619 controls), lung squamous cell carcinoma (7704 cases, 54763 controls), and small cell lung cancer (179 cases, 174006 controls). The TCGA-LUAD and TCGA-LUSC datasets were downloaded from the UCSC Xena database (https://xena.ucsc.edu), and the expression matrix of the GSE149507 small cell lung cancer dataset was downloaded from the GEO website (https://www.ncbi.nlm.nih.gov/geo/).

Genome-wide single-nucleotide polymorphisms (SNPs) associated with chemokines were generated from the full-site significance threshold $P<1.0\times10^{-8}$ [14, 17]. However, CCL1, CCL13, CCL20, CXCL9, CXCL10, CCL13, and CXCL14 had no associated SNPs, so the threshold $P<5.0\times10^{-6}$ was chosen. The following quality control steps were used to select SNPs to ensure the stability and accuracy of the causal association results between chemokines and LC risk. (1) Excluding SNPs with allelic inconsistency between the exposure and outcome samples (such as A / C); (2) Removing palindromic SNPs; (3) Eliminating SNPs causing linkage disequilibrium by employing the PLINK clumping method ($r^2<0.001$ and clump window = 10,000 kb); (4) Removing SNPs with minor allele frequency (MAF) < 0.01; (5) Using the PhenoScanners V2 database (phenoscanner.medschl.cam.ac.uk) [18], removal of mixed exposure and outcome related SNPs, such as two SNPs-rs508977 and rs1973612 of growth-regulated protein alpha levels were associated with exposure, thus excluding growth-regulated protein alpha levels. Abnormal SNPs were detected using MR-pleiotropy residual sum and outlier (MR-PRESSO) and

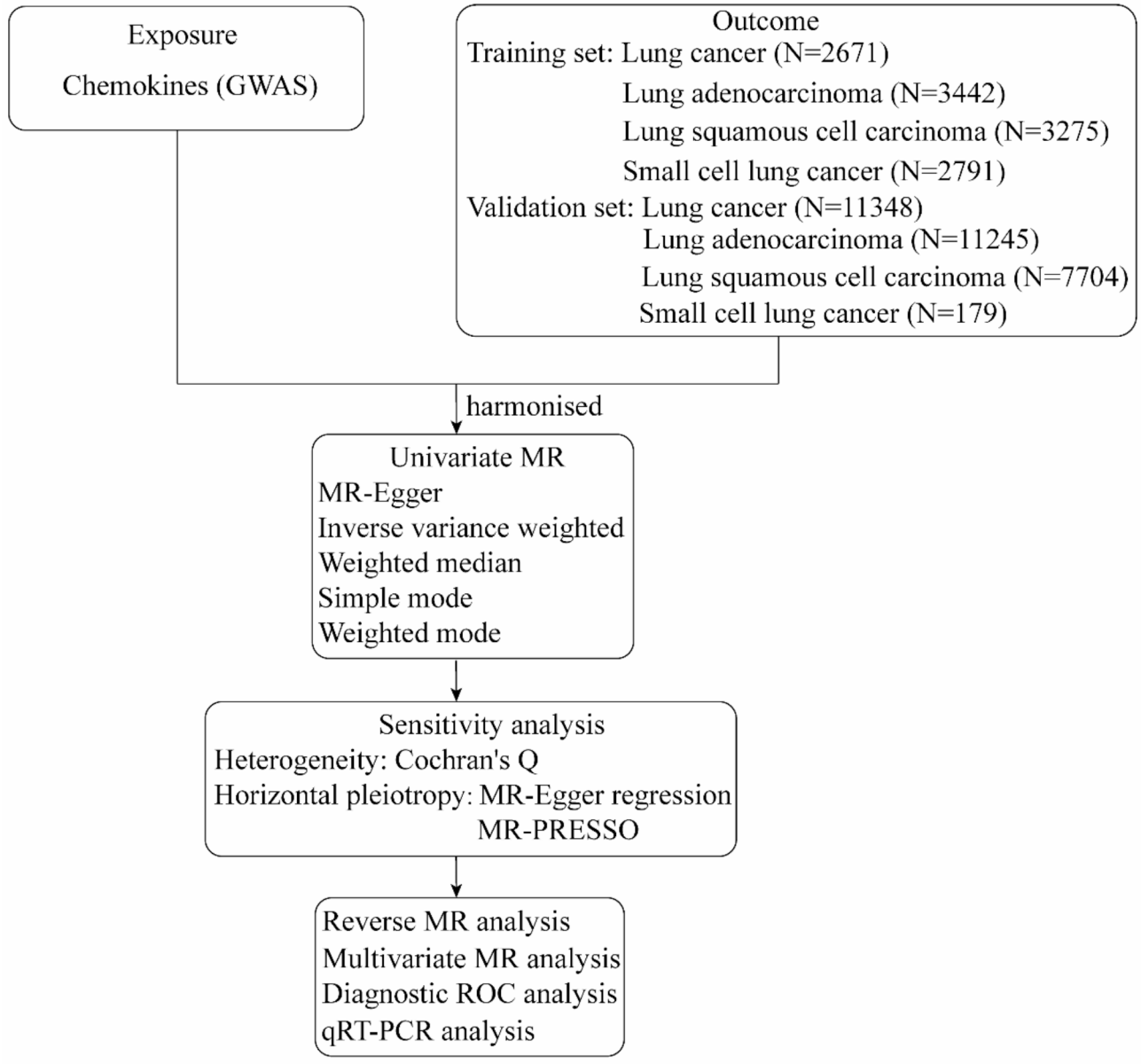


**Fig. 1** A graphical overview of the study design

pleiotropy was eliminated by removing outliers, followed by MR analysis after removing SNPs with pleiotropy. The $R^2$ and F statistics were used to assess weak instrumental deviation [19], and the F statistic was calculated using this formula: $R^2$ (n-k-1) / k (1-$R^2$). Where n represents the number of samples, k represents the number of IVs, and $R^2$ represents the variance explained by the IVs.

#### MR analysis

In this study, five methods, random or fixed effects inverse variance weighted (IVW) [20], weighted median estimation (WME) [20], MR-Egger [12], Simple mode (SM) [21], and Weighted mode (WM) [21], were used to verify whether there is a causal association between chemokines and lung cancer. IVW method, which uses the Wald ratio method for the association of individual SNP followed by meta-aggregation of multiple locus effects using a choice of fixed or random effects models, is able to provide the most accurate estimates of effects [20]. The premise of the WME method is to give an accurate assessment based on the assumption that at least 50% of IVs are valid [20]. MR-Egger method takes into account the presence of an intercept term that detects and adjusts for horizontal pleiotropy, and if horizontal pleiotropy is not present, then MR-Egger regression and IVW results are essentially identical [22]. SM and WM methods are also two important statistics in MR analysis.

In this study, the IVW method was used as the main method, and other methods were used as supplements.

### Sensitivity analysis

To further test the accuracy and stability of the findings, we performed sensitivity analysis using heterogeneity test, horizontal multiple validity test, and leave-one-out method. Cochrane's Q test was used to assess the heterogeneity of each chemokine-associated SNPs; if heterogeneity existed ($P<0.05$), a randomized IVW method was used; if heterogeneity did not exist ($P>0.05$), a fixed IVW method was used. The MR-Egger method means that when the intercept term is very different from 0, it indicates the presence of horizontal pleiotropy, in which case it is necessary to eliminate the horizontal pleiotropy by removing the abnormal SNPs with MR-PRESSO and re-performing the MR analysis [23, 24]. The leave-one-out method was used to assess whether MR causality was driven by a single SNP. To avoid false positive results, the false discovery rate (FDR) adjusted p-value in the main analysis was calculated using the fdrtool package [25]. Reverse MR analysis of chemokines causally related to LC in the forward MR analysis. Considering that smoking is a risk factor for LC, we performed a multivariate Mendelian randomization (MVMR) analysis to estimate the effect of each chemokine on LC after adjusting for smoking status. Further predicting the accuracy of the discovered chemokines to diagnose LC, we performed sensitivity and specificity analysis using expression matrices from the TCGA and GEO data.

### Lung adenocarcinoma cell lines

HBE, A549, H1299, H1975, and PC9 cell lines were acquired from the China Cell Resource Center (Shanghai, China). The cells were cultured in RPMI 1640 (Solarbio) supplemented with 10% fetal bovine serum (Procell). The cells were then incubated in a humidified incubator (Thermo Scientific, China) with 5% CO2 at 37 °C.

### RNA extraction and quantitative real-time PCR (qRT-PCR) assay

RNA extraction from cells was performed using the TRIzol reagent (Vazyme). The cDNA was synthesized using the Hifair® III 1st Strand cDNA Synthesis SuperMix for qPCR (gDNA digester plus) (Yeasen Biotechnology, Shanghai, China). The quantitative real time PCR was carried out with Hieff®qPCR SYBR Green Master Mix (High Rox Plus) (Yeasen Biotechnology, Shanghai, China) in the StepOne Plus Real-Time PCR System (Applied Biosystems). The relative expression levels of CCL21 mRNA were normalized to GAPDH as endogenous control respectively by using the $2-^{\triangle\triangle ct}$ method. The primer sequences were as follows: CCL21: F: 5′- GTTGCCTCA AGTACAGCCAAA-3′, R: 5′- AGAACAGGATAGCTGG GATGG-3′; GAPDH: F: 5′-TCATTTCCTGGTATGACA ACGA-3′, R: 5′-GTCTTACTCCTTGGAGGCC-3′.

### Statistical analysis

All MR statistical analyses were performed using the R software (version 4.2.0). MR analysis was performed using the TwosampleMR (version 0.5.7) and MR-PRESSO (version 3.6.0) R packages. The code used to perform all the analysis steps can be found in the online GitHub repository (https://github.com/jmzeng1314/GEO). The receiver operating characteristic (ROC) analysis of the data was performed using pROC [1.18.0], and the results were visualized with ggplot2 [3.3.6]. Dunnett's test in Ordinary one-way ANOVA to compare differences between multiple groups.

## Results

### Characteristics of the SNPs used for the analysis

Table 1 showed the source of the GWAS outcome data. Each SNP extracted from the different chemokines and its F statistic and $R^2$ are shown in Supplementary file Table S1. There were 173 SNPs in chemokine-LC, 149 chemokine-LUAD, 153 chemokine-LUSC, and 167 SNPs in chemokine-SCLC. The F-statistics ranged from 21 to 1484 and represent strong tools in MR analysis.

**Table 1** Characteristics of lung cancer in UK Biobank and lung cancer consortium

| test set | GWAS ID | Consortium | Cases/control | Population |
|---|---|---|---|---|
| Lung cancer | ieu-b-4954 | UK Biobank | 2671/372,016 | European |
| Lung adenocarcinoma | ieu-a-965 | ILCCO | 3442/14,894 | European |
| Lung squamous cell carcinoma | ieu-a-967 | ILCCO | 3275/15,038 | European |
| Small cell lung cancer | ieu-a-988 | TRICL | 2791/20,580 | European |
| **validation set** | | | | |
| Lung cancer | ieu-a-966 | ILCCO | 11,348/15,861 | European |
| Lung adenocarcinoma | ieu-a-984 | TRICL | 11,245/54,619 | European |
| Lung squamous cell carcinoma | ieu-a-989 | TRICL | 7704/54,763 | European |
| Small cell lung cancer | finn-b-C3_SCLC_EXALLC | NA | 179/174,006 | European |

ILCCO: International Lung Cancer Consortium, TRICL: Transdisciplinary Research in Cancer of the Lung

### Univariable MR analysis

The results of assessing the effect of chemokines on LC risk using methods such as Wald ratio or IVW analysis are shown in Fig. 2 and Supplementary file Table S2. After FDR correction, the test set showed CCL1[*OR*=1.001, 95%*CI* (1.000,1.002), *P*=0.015] was positively associated with LC risk, CCL15[*OR*=0.999, 95%*CI* (0.998,1.000), *P*=0.014] was negatively associated with LC risk, and CCL21[*OR*=0.999, 95%*CI* (0.998,1.000), *P*=0.014] was negatively associated with LC risk. However, the validation set did not find any statistical evidence that CCL1, CCL15, and CCL21 were associated with LC risk. In addition, the forest plot showed that CXCL13[*OR*=1.003, 95%*CI* (1.001, 1.005), *P*=0.008] was positively associated with LC risk. Although there was no heterogeneity by Cochran's Q test (Supplementary file Table S3), significant horizontal pleiotropy was found by MR-Egger regression intercept analysis (Supplementary file Table S4). Thus, there is insufficient evidence for a causal association.

IVW results showed that CCL21[*OR*=0.801, 95%*CI* (0.686, 0.935), *P*=0.005] was negatively associated with LUAD risk, but positively associated with SCLC risk [*OR*=5.346, 95%*CI* (1.005, 28.450), *P*=0.009].

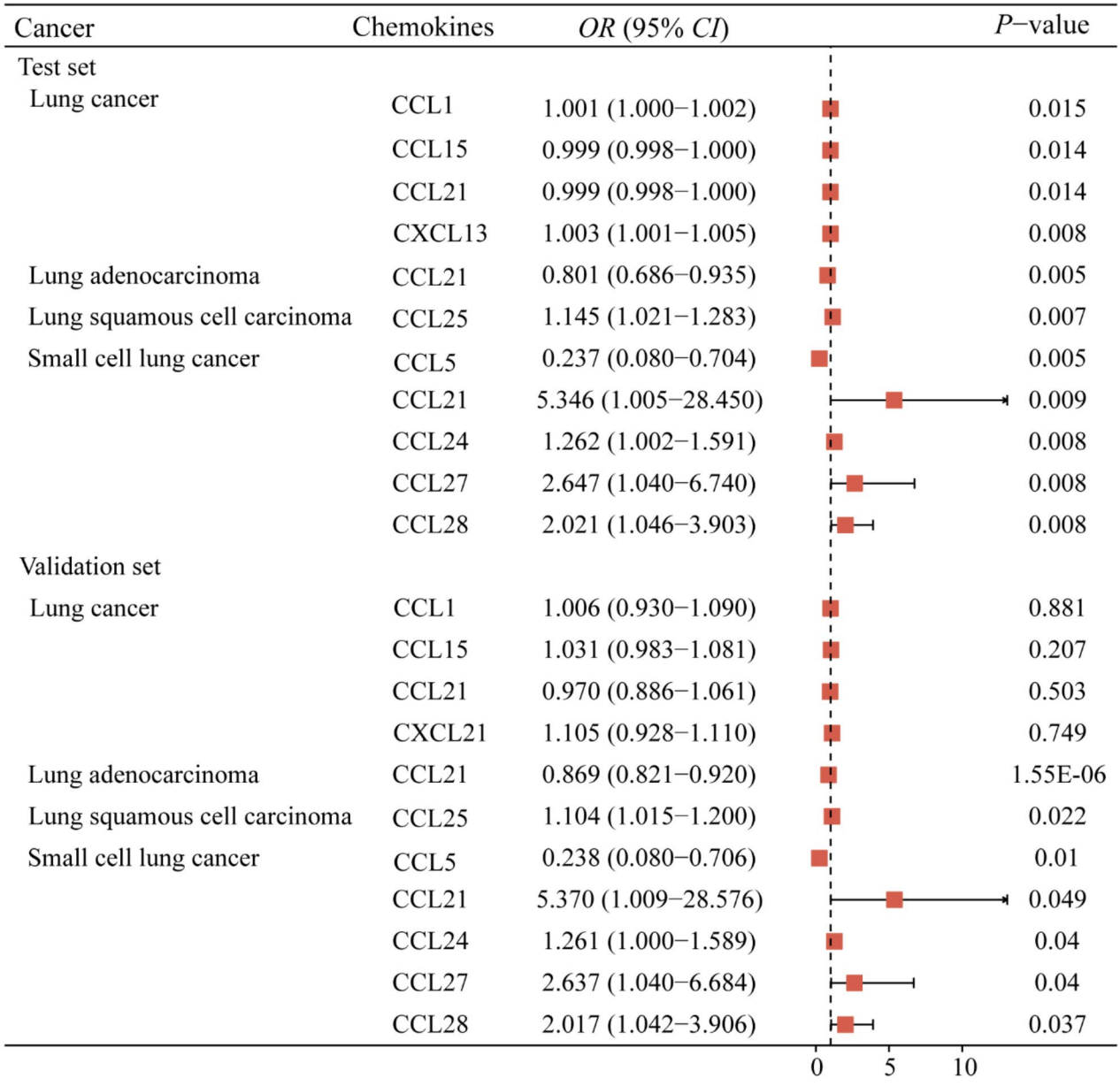


**Fig. 2** Mendelian analysis of chemokines on lung cancer risk
*OR*: Odds Ratio, *CI*: Confidence Interval

CCL25[*OR*=1.145, 95%*CI* (1.021, 1.283), *P*=0.007] was positively associated with LUSC risk. CCL5[*OR*=0.237, 95%*CI* (0.080, 0.704), *P*=0.005] was negatively associated with SCLC risk. CCL24[*OR*=1.262, 95%*CI* (1.002, 1.591), *P*=0.008], CCL27[*OR*=2.647, 95%*CI* (1.040, 6.740), *P*=0.008] and CCL28[*OR*=2.02, 95%*CI* (1.046, 3.903), *P*=0.008] were positively associated with SCLC risk. Similar results were obtained in the validation set. The Cochran's Q test results indicated that none of these chemokines were heterogeneous (Supplementary file Table S3). The results of MR-Egger regression intercept analysis indicated that there was no significant horizontal pleiotropy was present (Supplementary file Table S4). No horizontal pleiotropy was found by MR-PRESSO analysis (global test $p>0.05$, Supplementary file Table S4, S5).

### Reverse MR analysis

The reverse MR analysis showed that LC and histological subtypes had no significant causal association with the above six chemokines (Table 2). No heterogeneity was detected by the Cochran's Q test analysis (Supplementary file Table S6). The results of MR-Egger regression intercept and MR-PRESSO analysis indicated that there was no horizontal pleiotropy (Supplementary file Table S7, S8).

### Multivariable MR analysis

Since smoking was a high risk factor for LC, we further analyzed the effects of ever smoker, current cigarette smokers, light smokers and cigarettes smoked per day on LC by MVMR (Table 3, Supplementary file Table S9). After adjustment for cigarettes smoked per day [*OR*=2.412, 95%*CI* (1.743, 3.337), *P*=1.094E-07], CCL21 remained protective against this LUAD with the smoking status adjustment [*OR*=0.799, 95%*CI* (0.657, 0.971), *P*=0.024]. After adjusting for four smoking statuses, the risk of LUSC by CCL25 was found to disappear, suggesting that the increased risk of LUSC by CCL25 is likely to be mediated by smoking. CCL5, CCL24, CCL27, and CCL28 for the risk of SCLC had similar results to

**Table 2** Full result of MR estimates for the association between lung cancer and chemokines

| Exposure | Outcome | Method | No.of snp | OR (95% CI) | *P*-value |
|---|---|---|---|---|---|
| Lung adenocarcinoma | CCL21 | IVW | 2 | 0.976 (0.792–1.204) | 0.822 |
| Lung squamous cell carcinoma | CCL25 | MR Egger | 4 | 1.024 (0.620–1.692) | 0.933 |
| | | WME | | 0.950 (0.823–1.095) | 0.478 |
| | | IVW | | 0.995 (0.858–1.153) | 0.947 |
| | | SM | | 0.944 (0.771–1.156) | 0.614 |
| | | WM | | 0.946 (0.801–1.117) | 0.560 |
| Small cell lung cancer | CCL5 | MR Egger | 10 | 1.054 (0.986–1.126) | 0.162 |
| | | WME | | 1.019 (0.986–1.054) | 0.265 |
| | | IVW | | 1.020 (0.993–1.049) | 0.154 |
| | | SM | | 1.009 (0.957–1.064) | 0.740 |
| | | WM | | 1.015 (0.963–1.069) | 0.596 |
| | CCL21 | MR Egger | 10 | 0.991 (0.917–1.071) | 0.830 |
| | | WME | | 0.964 (0.924–1.005) | 0.081 |
| | | IVW | | 0.973 (0.944–1.004) | 0.086 |
| | | SM | | 0.939 (0.869–1.013) | 0.140 |
| | | WM | | 0.940 (0.865–1.021) | 0.176 |
| | CCL24 | MR Egger | 9 | 1.064 (0.908–1.246) | 0.467 |
| | | WME | | 1.033 (0.966–1.104) | 0.343 |
| | | IVW | | 1.020 (0.968–1.075) | 0.455 |
| | | SM | | 1.048 (0.954–1.151) | 0.358 |
| | | WM | | 1.042 (0.961–1.130) | 0.344 |
| | CCL27 | MR Egger | 10 | 1.066 (0.997–1.139) | 0.097 |
| | | WME | | 1.039 (0.998–1.081) | 0.066 |
| | | IVW | | 1.021 (0.993–1.050) | 0.143 |
| | | SM | | 1.052 (0.988–1.120) | 0.146 |
| | | WM | | 1.051 (0.992–1.112) | 0.124 |
| | CCL28 | MR Egger | 10 | 0.972 (0.910–1.039) | 0.428 |
| | | WME | | 0.988 (0.955–1.023) | 0.504 |
| | | IVW | | 0.988 (0.961–1.016) | 0.392 |
| | | SM | | 1.005 (0.952–1.061) | 0.856 |
| | | WM | | 0.989 (0.943–1.037) | 0.668 |

snp: single-nucleotide polymorphism, *OR*: Odds Ratio, *CI*: Confidence Interval

**Table 3** The effect of smoking status on chemokines

| Outcome | Exposure | method | OR (95%CI) | *P*-value |
|---|---|---|---|---|
| Lung adenocarcinoma | CCL21 | MVMR adjusted for ever smoker | 0.818 (0.679–0.985) | 0.034 |
| | | MVMR adjusted for current cigarette smokers | 0.795 (0.542–1.166) | 0.240 |
| | | MVMR adjusted for light smokers | 0.829 (0.478–1.437) | 0.504 |
| | | MVMR adjusted for cigarettes smoked per day | 0.799 (0.657–0.971) | 0.024 |
| Lung squamous cell carcinoma | CCL25 | MVMR adjusted for ever smoker | 1.127 (0.996–1.274) | 0.057 |
| | | MVMR adjusted for current cigarette smokers | 1.093 (0.893–1.339) | 0.387 |
| | | MVMR adjusted for light smokers | 1.159 (0.819–1.639) | 0.405 |
| | | MVMR adjusted for cigarettes smoked per day | 1.187 (0.890–1.582) | 0.244 |
| Small cell lung cancer | CCL5 | MVMR adjusted for ever smoker | 0.612 (0.306–1.226) | 0.166 |
| | | MVMR adjusted for current cigarette smokers | 0.225 (0.137–0.370) | 0.000 |
| | | MVMR adjusted for light smokers | 0.244 (0.081–0.731) | 0.012 |
| | | MVMR adjusted for cigarettes smoked per day | 0.212 (0.073–0.618) | 0.004 |
| | CCL21 | MVMR adjusted for ever smoker | 0.755 (0.503–1.135) | 0.177 |
| | | MVMR adjusted for current cigarette smokers | 0.807 (0.681–0.957) | 0.013 |
| | | MVMR adjusted for light smokers | 0.858 (0.570–1.291) | 0.462 |
| | | MVMR adjusted for cigarettes smoked per day | 0.903 (0.476–1.712) | 0.754 |
| | CCL24 | MVMR adjusted for ever smoker | 1.197 (0.977–1.466) | 0.082 |
| | | MVMR adjusted for current cigarette smokers | 1.250 (1.052–1.485) | 0.011 |
| | | MVMR adjusted for light smokers | 1.213 (1.009–1.457) | 0.040 |
| | | MVMR adjusted for cigarettes smoked per day | 1.209 (0.960–1.522) | 0.107 |
| | CCL27 | MVMR adjusted for ever smoker | 1.617 (0.849–3.078) | 0.144 |
| | | MVMR adjusted for current cigarette smokers | 2.875 (1.982–4.170) | 0.000 |
| | | MVMR adjusted for light smokers | 2.799 (1.590–4.927) | 0.000 |
| | | MVMR adjusted for cigarettes smoked per day | 2.311 (0.731–7.308) | 0.154 |
| | CCL28 | MVMR adjusted for ever smoker | 1.700 (1.027–2.815) | 0.039 |
| | | MVMR adjusted for current cigarette smokers | 1.687 (1.188–2.395) | 0.003 |
| | | MVMR adjusted for light smokers | 2.029 (1.205–3.416) | 0.008 |
| | | MVMR adjusted for cigarettes smoked per day | 1.588 (0.715–3.528) | 0.256 |

*OR*: Odds Ratio, *CI*: Confidence Interval, MVMR: Multivariate Mendelian Randomization

the univariate MR analysis. After adjusting for current cigarette smokers, CCL21[*OR*=34.858, 95%*CI* (17.093, 71.085), *P*=1.550E-22] on the risk of SCLC [*OR*=0.807, 95%*CI* (0.681, 0.957), *P*=0.013] was contrary to the univariate results [*OR*=34.858, 95%*CI* (17.093, 71.085), *P*=1.550E-22] in contrast, suggesting that smoking most likely as a mediator to regulate the effect of CCL21 on SCLC.

### The diagnostic performance of the chemokines and expression levels of CCL21

We further explored the accuracy of the discovered chemokines in diagnosing LC using ROC curves. The results are shown in Fig. 3, The area under curve (AUC) of CCL21 predicted LUAD is 0.754 (CI: 0.694–0.813) (Fig. 3a), CCL25 predicted that the AUC of LUSC is 0.777 (CI: 0.720–0.835) (Fig. 3b), the AUC of CCL5 predicted SCLC is 0.769 (CI: 0.584–0.953) (Fig. 3c), the AUC of SCLC predicted by CC21 is 0.836 (CI: 0.709–0.964) (Fig. 3d), CCL24 predicted that the AUC of SCLC is 0.596 (CI: 0.404–0.787) (Fig. 3e), that the AUC of SCLC for CC27 is 0.682 (CI: 0.494–0.870) (Fig. 3f), the AUC of SCLC predicted by CC28 is 0.722 (CI: 0.547–0.898) (Fig. 3g). The above results indicated the high diagnostic performance of the discovered chemokines for the corresponding outcome. The qRT-PCR indicated that CCL21 expression was lower in A549 and PC9 cells and higher in H1299 and H1975 cells (Fig. 3h), demonstrating the specificity of CCL21 expression levels in lung adenocarcinoma cells.

## Discussion

As investigated, this is the first MR study to explore whether chemokines are causally associated with LC at the gene level. While randomized controlled trials (RCTs) can provide the most convincing evidence, they involve many ethical issues and are costly. For observational studies, although these results are observational studies freely adjusted for other relevant variables, some insignificant bias will be overlooked. Thus, the results provided by MR are the most convincing. To minimize bias such as reverse and confounders resulting from the MR hypothesis, we performed a series of sensitivity analyses. And multivariate MR was performed to adjust for smoking. In the study, the relevant SNPs exposed to growth-regulated protein alpha levels were all related to

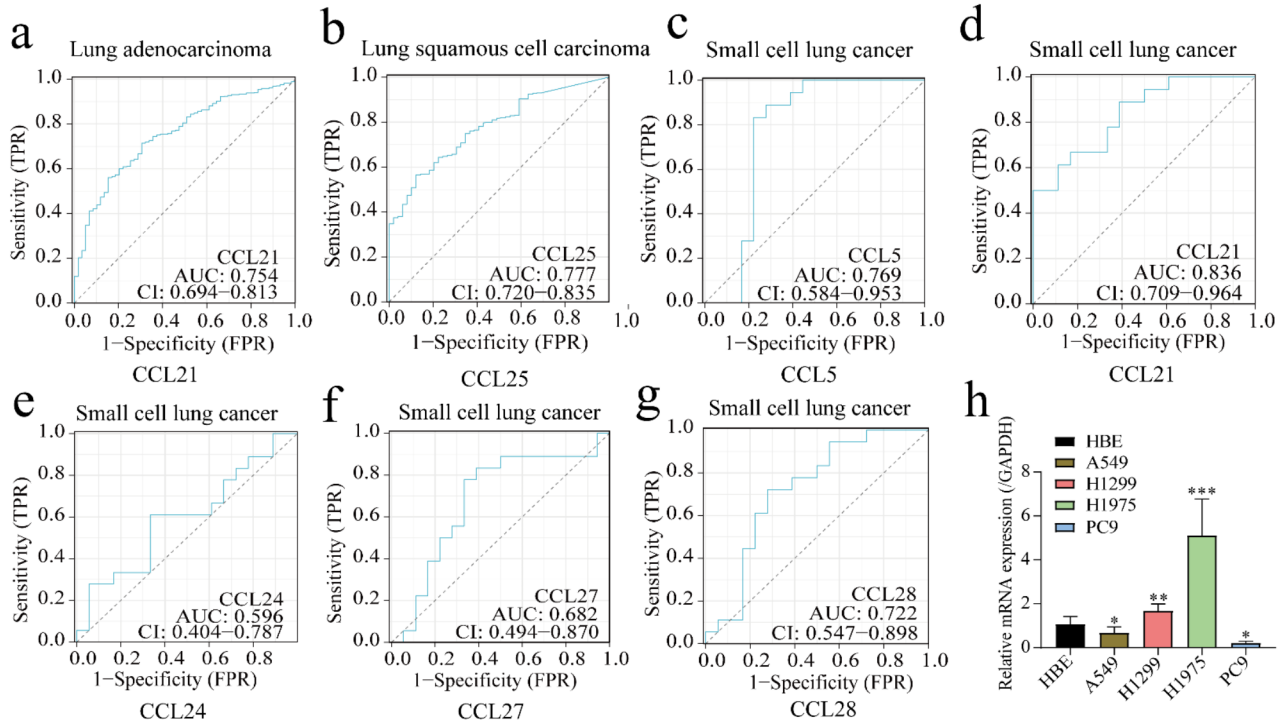


**Fig. 3** The ROC curve predicted the diagnostic performance of chemokines and the expression level of CCL21 as measured by qRT-PCR
ROC: Receiver Operating Characteristic, FPR: False Positive Rate, TPR: True Positive Rate, AUC: Area Under Curve, CI: Confidence Interval. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

the exposure itself, which did not meet the independence hypothesis in the MR hypothesis, so it was not explored in depth. Analysis of 38 chemokines identified CCL21 as a protective factor for LUAD and SCLC. CCL25 was a risk factor for LUSC. CCL5 was a protective factor for SCLC. CCL24, CCL27, and CCL28 were risk factors for SCLC. Meanwhile, these chemokines had high diagnostic performance in predicting the corresponding outcome. These results may be helpful for interventions conducted to reduce cancer risk.

As key proteins of immune cell migration, chemokines have been demonstrated for their effects on inflammation-related cancers. Chemokines act in an autocrine or paracrine manner to attract other cytokines that promote growth and inflammation and mediate processes of angiogenesis and tumor growth, including pro-or anti-angiogenesis and activation or inhibition of tumor growth and survival [26, 27]. This is the first time to genetically predict the association between chemokines and LC, and thus requires detailed analysis and exploration of the results of this study and previous studies, which will help us to further the understanding of the mechanism of chemokines in LC.

Our analysis found a negative association between CCL21 and both LUAD and SCLC. CCL21 may exert an antitumor effect, and the main mechanism was that the activated CCL21 / CCR7 axis increased CD4a expression in $CD11^+$ T cells and promoted the interaction with $CD20^+$ B cells, which can induce the formation of tertiary lymphoid structures [28]. Tertiary lymphoid structures can inhibit the development of LUAD and improve patient survival [29, 30]. CCL21 was associated with immune cells [31], which may be an important reason for its protective effect against LUAD. However, how CCL21 influences the development of SCLC is still not studied and needs to be further explored in the future.

CCL25 was positively associated with LUSC risk. The study found that CCL25 levels in cancer tissues and serum of LUSC patients were higher than normal controls, its high expression promoted invasion of LUSC cells and metastasis, reduced patient survival [32, 33], and CCL25 promoted migration and invasion of several other types of cancer [34]. Although our univariate results were consistent with them, multivariate analysis found a mediating role of smoking behavior, so whether CCL25 levels can serve as a biomarker for noninvasive diagnostic or prognostic purposes still needs to be deeply explored.

Biological information found that high expression of CCL5 was positively associated with survival in SCLC patients, and CCL5 was a potential biomarker for predicting the response of SCLC to immune checkpoint blockade [35]. Currently, there are few studies on the development and treatment of CCL24, CCL27, and

CCL28 affect the development of SCLC patients, but several studies [36–38] have found the malignant biological behavior of CCL27 and CCL28 on non-small cell lung cancer cells, including proliferation, migration and invasion. The mechanism of how CCL24, CCL27, and CCL28 affect SCLC is still in large gaps, which deserve our further exploration in the future. Moreover, the results of the diagnostic ROC curve indicate that these chemokines not only play a role in the development of LC, but also have a high accuracy performance in predicting LC.

However, this study also has several limitations. Firstly, subgroup analysis of differences including sex was impossible because original summary statistics were used in the analysis, and potential bias caused could not be excluded. Secondly, for sensitivity analysis, we set a threshold of SNPs for genetic variants of individual chemokines to $p<5\times10^{-6}$, which may introduce a weak instrumental bias. Thirdly, although the data for chemokines and LC from the GWAS meta-analysis were of European ancestry, there may still be population stratification interference, which could lead to bias estimation and influence generalizability. Future studies of MR between chemokines and LC need to be considered in populations of different ethnic groups in different countries to achieve better generalizability.

## Conclusion

In conclusion, this study comprehensively analyzed the causal association of chemokines with LC by two-sample bidirectional MR methods. CCL21 decreased the risk of LUAD and SCLC. CCL25 increased the risk of LUSC. CCL5 reduced the risk of SCLC, and CCL24, CCL27, and CCL28 increased the risk of SCLC. The results suggested that these findings may provide new insights into the mechanisms of chemokine-mediated cancer development, while also facilitating the future discovery of more harmful and beneficial chemokines affecting LC progression, which was the most important clinical significance of this study.

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13019-024-03128-5.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Supplementary Material 7

Supplementary Material 8

Supplementary Material 9

**Acknowledgements**
Not applicable.

**Author contributions**
MMW analyzed the patient data of lung cancer and chemokines, and wrote the manuscript. MJG determined the study design and wrote the manuscript. WBH conducted data collection and analysis. SDZ contributed to the data interpretation. YSS and XLW revised the manuscript. All authors read and approved the final manuscript.

**Funding**
This work was supported by Jiangsu Provincial Health Commission Elderly Health Research Project (No. LKZ2022019) and Yangzhou Social Development and Clinical Frontier Technology Project (No. YZ2023084).

**Data availability**
No datasets were generated or analysed during the current study.

## Declarations

**Ethics approval and consent to participate**
Not applicable.

**Consent for publication**
Not applicable.

**Competing interests**
The authors declare no competing interests.

**Author details**
[1]Dalian Medical University, Dalian 116000, China
[2]The Yangzhou School of Clinical Medicine of Dalian Medical University, Yangzhou 225000, China
[3]Northern Jiangsu People's Hospital Affiliated to Yangzhou University, Yangzhou 225000, China
[4]Department of Thoracic Surgery, Northern Jiangsu People's Hospital, No. 98 Nantong West Road, Yangzhou 225000, China

Received: 28 March 2024 / Accepted: 2 November 2024
Published online: 08 November 2024

**References**

1. The L. Lung cancer: some progress, but still a lot more to do. Lancet. 2019;394(10212).
2. Gao S, Li N, Wang S, et al. Lung Cancer in people's Republic of China. J Thorac Oncol. 2020;15(10):1567–76.
3. Thai AA, Solomon BJ, Sequist LV, et al. Lung cancer. Lancet. 2021;398(10299):535–54.
4. Sung H, Ferlay J, Siegel RL et al. Global Cancer Statistics. 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021;71(3):209 – 49.
5. Bade BC, Dela Cruz CS. Lung Cancer 2020. Clin Chest Med. 2020;41(1):1–24.
6. Patel B, Priefer R. Impact of chronic obstructive pulmonary disease, lung infection, and/or inhaled corticosteroids use on potential risk of lung cancer. Life Sci. 2022;294:120374.
7. Rayens NT, Rayens EA, Tighe RM. Co-occurrence of pneumoconiosis with COPD, pneumonia and lung cancer. Occup Med (Lond). 2022;72(8):527–33.
8. Budisan L, Zanoaga O, Braicu C et al. Links between infections, Lung Cancer, and the Immune System. Int J Mol Sci. 2021;22(17).
9. Srivastava S, Mohanty A, Nam A, et al. Chemokines and NSCLC: emerging role in prognosis, heterogeneity, and therapeutics. Semin Cancer Biol. 2022;86(Pt 2):233–46.
10. Unver N. Identification of the dominant angiogenic CXCL class chemokines associated with non-small cell lung cancer via bioinformatics tools. Med Oncol. 2021;38(6).
11. Davey Smith G, Ebrahim S. Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*. Int J Epidemiol. 2003;32(1):1–22.

12. Bowden J, Holmes MV. Meta-analysis and mendelian randomization: a review. Res Synthesis Methods. 2019;10(4):486–96.
13. Dan Y-L, Wang P, Cheng Z, et al. Circulating adiponectin levels and systemic lupus erythematosus: a two-sample mendelian randomization study. Rheumatology. 2021;60(2):940–6.
14. Yu X, Zhang Y, Lin Y et al. The association between plasma chemokines and breast cancer risk and prognosis: a mendelian randomization study. Front Genet. 2023;13.
15. Pierce BL, Burgess S. Efficient design for mendelian randomization studies: Subsample and 2-Sample instrumental variable estimators. Am J Epidemiol. 2013;178(7):1177–84.
16. McKay JD, Hung RJ, Han Y, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nat Genet. 2017;49(7):1126–32.
17. Bouras E, Karhunen V, Gill D et al. Circulating inflammatory cytokines and risk of five cancers: a mendelian randomization analysis. BMC Med. 2022;20(1).
18. Staley JR, Blackshaw J, Kamat MA, et al. PhenoScanner: a database of human genotype–phenotype associations. Bioinformatics. 2016;32(20):3207–9.
19. Burgess S, Thompson SG. Avoiding bias from weak instruments in mendelian randomization studies. Int J Epidemiol. 2011;40(3):755–64.
20. Boehm FJ, Zhou X. Statistical methods for mendelian randomization in genome-wide association studies: a review. Comput Struct Biotechnol J. 2022;20:2338–51.
21. Zhang D, Hu Y, Guo W et al. Mendelian randomization study reveals a causal relationship between rheumatoid arthritis and risk for pre-eclampsia. Front Immunol. 2022;13.
22. Burgess S, Thompson SG. Interpreting findings from mendelian randomization using the MR-Egger method. Eur J Epidemiol. 2017;32(5):377–89.
23. Li C, Liu C, Li N. Causal associations between gut microbiota and adverse pregnancy outcomes: a two-sample mendelian randomization study. Front Microbiol. 2022;13.
24. Gao N, Kong M, Li X et al. Systemic Lupus Erythematosus and Cardiovascular Disease: a mendelian randomization study. Front Immunol. 2022;13.
25. Zhu M, Ma Z, Zhang X et al. C-reactive protein and cancer risk: a pan-cancer study of prospective cohort and mendelian randomization analysis. BMC Med. 2022;20(1).
26. Zulfiqar B, Farooq A, Kanwal S et al. Immunotherapy and targeted therapy for lung cancer: current status and future perspectives. Front Pharmacol. 2022;13.
27. Märkl F, Huynh D, Endres S, et al. Utilizing chemokines in cancer immunotherapy. Trends Cancer. 2022;8(8):670–82.
28. Yin X, Wang H, Li R et al. Tobacco exposure primes the secretion of CCL21 positively associated with tertiary lymphoid structure and response to immunotherapy. J Immunother Cancer. 2023;11(6).
29. Sun K, Zhang Z, Wang D et al. B cell-related tertiary lymphoid structure may exert inhibitory effects on lung adenocarcinoma and SARS-COV-2. Heliyon. 2023;9(3).
30. Zhao H, Wang H, Zhao Y et al. Tumor-Resident T cells, Associated with Tertiary Lymphoid structure maturity, improve survival in patients with Stage III Lung Adenocarcinoma. Front Immunol. 2022;13.
31. Han L, Zhang L. CCL21/CCR7 axis as a therapeutic target for autoimmune diseases. Int Immunopharmacol. 2023;121:110431.
32. Gupta P, Sharma PK, Mir H, et al. CCR9/CCL25 expression in non-small cell lung cancer correlates with aggressive disease and mediates key steps of metastasis. Oncotarget. 2014;5(20):10170–9.
33. Niu Y, Tang D, Fan L et al. CCL25 promotes the migration and invasion of non–small cell lung cancer cells by regulating VEGF and MMPs in a CCR9–dependent manner. Experimental Therapeutic Med. 2020.
34. Xu B, Deng C, Wu X, et al. CCR9 and CCL25: a review of their roles in tumor promotion. J Cell Physiol. 2020;235(12):9121–32.
35. Tang Y, Hu Y, Niu Y et al. CCL5 as a prognostic marker for Survival and an Indicator for Immune Checkpoint therapies in Small Cell Lung Cancer. Front Med. 2022;9.
36. Li B, Wei C, Zhong Y, et al. The CCL27-CCR10 axis contributes to promoting proliferation, migration, and invasion of lung squamous cell carcinoma. Histol Histopathol. 2023;38(3):349–57.
37. Liu Y, Xiao A, Zhang B. CCR10/CCL27 crosstalk regulates cell metastasis via PI3K-Akt signaling axis in non-small-cell lung cancer. Am J Transl Res. 2021;13(11):13135–46.
38. Huang G, Tao L, Shen S, et al. Hypoxia induced CCL28 promotes angiogenesis in lung adenocarcinoma by targeting CCR3 on endothelial cells. Sci Rep. 2016;6:27152.

## Publisher's note