



OPEN

Potential drug targets for tumors identified through Mendelian randomization analysis

Na Song^{1,2,4}, Pingyu Shi^{2,4}, Kai Cui^{2,4}, Liqun Zeng², Ziwei Wang², Wenyu Di¹, Jinsong Li¹, Yanwu Fan², Zhanjun Li³, Jinghang Zhang¹, Wei Su¹✉ & Haijun Wang^{1,2}✉

According to the latest cancer research data, there are a significant number of new cancer cases and a substantial mortality rate each year. Although a substantial number of clinical patients are treated with existing cancer drugs each year, the efficacy is unsatisfactory. The incidence is still high and the effectiveness of most cancer drugs remains unsatisfactory. Therefore, we evaluated the human proteins for their causal relationship to for cancer risk and therefore also their potential as drug targets. We used summary tumors data from the FinnGen and cis protein quantitative trait loci (cis-pQTL) data from a genome-wide association study, and employed Mendelian randomization (MR) to explore the association between potential drug targets and nine tumors, including breast, colorectal, lung, liver, bladder, prostate, kidney, head and neck, pancreatic caners. Furthermore, we conducted MR analysis on external cohort. Moreover, Bidirectional MR, Steiger filtering, and colocalization were employed to validate the main results. The DrugBank database was used to discover potential drugs of tumors. Under the threshold of False discovery rate (FDR) < 0.05, results showed that S100A16 was protective protein and S100A14 was risk protein for human epidermal growth factor receptor 2-positive (HER-positive) breast cancer, phosphodiesterase 5A (PDE5A) was risk protein for colorectal cancer, and melanoma inhibitory activity (MIA) was protective protein for non-small cell lung carcinoma (NSCLC). And there was no reverse causal association between them. Colocalization analysis showed that S100A14 (PP.H4.abf = 0.920) and S100A16 (PP.H4.abf = 0.932) shared causal variation with HER-positive breast cancer, and PDE5A (PP.H4.abf = 0.857) shared causal variation with colorectal cancer (CRC). The MR results of all pQTL of PDE5A and MIA were consistent with main results. In addition, the MR results of MIA and external outcome cohort were consistent with main results. In this study, genetic predictions indicate that circulating S100 calcium binding protein A14 (S100A14) and S100 calcium binding protein A16 (S100A16) are associated with increase and decrease in the risk of HER-positive breast cancer, respectively. Circulating PDE5A is associated with increased risk of CRC, while circulating MIA is associated with decreased risk of NSCLC. These findings suggest that four proteins may serve as biomarkers for cancer prevention and as potential drug targets that could be expected for approval.

Keywords Drug target, Biomarker, Plasma protein, Tumor, Mendelian randomization

According to the latest cancer survey data, there are a significant number of new cancer cases and substantial mortality rates each year. Malignant tumors such as lung cancer, colorectal cancer, and pancreatic cancer rank in the top ten in both incidence and mortality rates for both male and female¹. Furthermore, prostate cancer remains the leading cause of death among men. Despite a significant decline in prostate cancer mortality rates, it still stands as the second leading cause of cancer-related deaths². Similarly, although survival rates for breast cancer have improved significantly over the past few decades, it remains the most common cancer and the second

¹Department of Pathology, Xinxiang Key Laboratory of Precision Medicine, The First Affiliated Hospital of Xinxiang Medical University, Jiankang Road No.88, Xinxiang 453100, China. ²Department of Pathology, School of Basic Medical Sciences, Xinxiang Medical University, Jinsui Road No. 601, Xinxiang 453000, China. ³State Key Laboratory for Diagnosis and Treatment of Severe Zoonotic Infectious Diseases, Key Laboratory for Zoonosis Research of the Ministry of Education, Institute of Zoonosis, and College of Veterinary Medicine, Jilin University, Changchun 130062, China. ⁴These authors contributed equally: Na Song, Pingyu Shi and Kai Cui. ✉email: weisu_xxmu@163.com; 151044@xxmu.edu.cn

deadliest cancer among female³. Therefore, cancer prevention and treatment are important in the field of scientific research. However, the identification of biomarkers for cancer prevention is still insufficient, and the unique mechanisms within the tumor microenvironment pose challenges for drug treatments^{4–6}. Currently, commonly used anti-cancer drugs include 5-fluorouracil, capecitabine, paclitaxel, gemcitabine, and others. However, the effectiveness of these drugs remains limited to a small subset of patients^{7,8}. In recent years, studies have attempted to discover novel and efficient cancer therapeutic drugs^{9,10}. However, some findings encountered obstacles in the process of translating into clinical outcomes. For example, some anti-cancer drugs may encounter issues such as insufficient efficacy, ineffectiveness, or adverse reactions¹¹. Revealing precancerous markers and identifying potential therapeutic targets may offer new insights into the future prevention and treatment of cancer.

Proteins play crucial roles in various biological processes in the human body, as well as in the progression of cancer^{12,13}. Currently, due to the vital functions of proteins, multiple drug targets are primarily concentrated on proteins. In recent years, MR studies have been widely employed in drug target screening and drug repurposing^{14,15}. MR is a recently emerging research method that employs genetic instrumental variables for causal association analysis^{16,17}. Single nucleotide polymorphisms (SNPs) identified through genome-wide association studies are used as instrumental variables to assess the causal association between exposure and outcome. Here, the study of human plasma protein data using MR may help to uncover underlying genetic factors in tumors.

In our study, we first performed a two-sample MR analysis with plasma protein as exposure and 9 tumors in the FinnGen database as outcomes. Our results were then further explored by external queue validation, reverse MR, and co-localization analysis. Finally, four plasma proteins are being considered as potential drug targets for tumors, and Drugs that identify potential targets based on DrugBank database are considered potential therapeutics for tumors.

Materials and methods

Mendelian randomization (MR)

Mendelian Randomization (MR) is a novel analytical method that has emerged in recent years, utilizing statistical techniques to assess the impact of a specific factor on human diseases¹⁸. It utilizes instrumental variable to analysis, aiming to assess causal relationships between exposures and outcomes using non-experimental data. In MR analysis, genetic variation SNPs are considered instrumental variables. When DNA is passed from parents to offspring, alleles segregate independently. This method is akin to random assignment in randomized controlled trials (RCTs), with the purpose of simulating RCTs to minimize the risk of confounding¹⁹. However, before conducting MR analysis, three assumptions need to be met. First, the instrumental variables must be strongly correlated with the exposure factor. Second, the instrumental variables must be independent of confounding factors. Third, the instrumental variables should only affect the outcomes through the exposure factor²⁰. In accordance with the principles and assumptions of Mendelian randomization, we have crafted a flowchart outlining the analysis procedure for this study (Fig. 1).

pQTL and GWAS data

The study design was shown in Fig. 1. We conducted a MR analysis to explore the causal association between plasma protein concentration (n = 4917) and nine types of tumors, including breast, colorectal, lung, liver, bladder, prostate, kidney, head and neck, pancreatic caners. Summary-level data of plasma pQTL was obtained from a comprehensive genome-wide association study (GWAS) published by Ferkingstad et al.²¹. In addition, we performed the analysis with cis-pQTLs as they are known for their resistance to horizontal pleiotropy²². Previously, a study has identified several potential drug targets for acute pancreatitis through MR analysis of

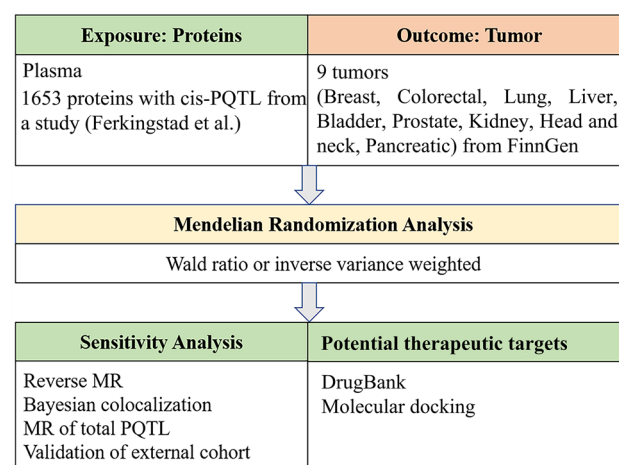


Figure 1. Study design for identification of plasma proteins causally associated with tumors. pQTL: protein quantitative trait loci.

cis-pQTLs for 4719 plasma proteins²³. Here, we employed the same plasma protein screening approach as follows: Firstly, a rigorous association screening was performed on 4917 plasma protein pQTL ($p < 5e-8$), which be advantageous for excluding weak instrumental variables. Secondly, SNPs with linkage disequilibrium (LD) were eliminated based on the criteria of $r^2 = 0.001$. Thirdly, to ensure that SNPs are not pleiotropy, the retained SNPs are limited to those with cis (Gene locus ± 1 Mb). In addition, due to algorithmic reasons, SNPs with a low minor allele frequency (MAF) exhibit higher error rates, hence it is necessary to set the MAF threshold to 0.01. Additionally, we utilized PhenoScanner, a database containing published literature on GWAS data information, to control for confounding factors²⁴. F-value is an indicator of the correlation between an instrumental variable and exposure, and the SNPs (F-value > 10) are strongly correlated with exposure²⁵.

The tumor GWAS data were obtained from the R9 version of the FinnGen database (https://www.finnngen.fi/en/access_results) (Supplement Table S1)²⁶, including malignant neoplasm of breast (HER-negative) (num_cases = 5965, num_controls = 167,017), malignant neoplasm of breast (HER-positive) (num_cases = 9698, num_controls = 167,017), colorectal cancer (num_cases = 6509, num_controls = 287,137), non-small cell lung cancer (num_cases = 4901, num_controls = 287,137), small cell lung cancer (num_cases = 676, num_controls = 167,017), hepatocellular carcinoma (num_cases = 453, num_controls = 287,137), clear cell adenocarcinoma of kidney (num_cases = 901, num_controls = 167,017), malignant neoplasm of bladder (num_cases = 2053, num_controls = 167,017), malignant neoplasm of head and neck (num_cases = 2131, num_controls = 287,137), adenocarcinoma and ductal carcinoma of pancreas (num_cases = 692, num_controls = 287,137), malignant neoplasm of prostate (num_cases = 13,216, num_controls = 119,948). All data were derived from public databases and approved by the ethics review committee.

Data for validation

The GWAS summary data for colorectal cancer and lung cancer were derived from the UK Biobank database and were obtained from the Open GWAS online database (<https://gwas.mrcieu.ac.uk>). The study included a total of 5,657 cases and 372,016 controls for colorectal cancer, as well as 2,671 cases and 372,016 controls for lung cancer. Moreover, the GWAS summary data of breast cancer were obtained from the Breast Cancer Association Consortium (BCAC)²⁷, encompassing a total of 69,501 cases and 105,974 controls. These data were used to validate the primary results.

MR analysis

We conducted a two-sample MR analysis using the retained SNPs to identify potential therapeutic targets for tumors. In this study, plasma protein concentration was employed as the exposure factor, while tumor served as the outcome factor. The methods of Inverse-variance weighted (IVW) and the Wald ratio were employed to assess the causal association between plasma protein concentration and tumors²⁸. The Wald ratio was employed for the analysis of a single SNP, while the IVW method was utilized to evaluate multiple SNPs. The main results were screened based on the threshold condition of false discovery rate (FDR) < 0.05 ²⁹, and the Benjamini–Hochberg method was employed for FDR calculation. In addition, we utilized the phenoscanner platform (<http://www.phenoscaner.medschl.cam.ac.uk/>) to fulfill the assumptions of independence and exclusivity in MR analysis³⁰. SNPs associated with tumor risk factors and those directly associated with tumors were eliminated. Afterwards, we utilized the ‘harmonise_data’ and ‘mr’ function in TwoSampleMR to extract the same SNPs in the outcome and obtained the results. And the accuracy of the direction was verified by steiger test. Furthermore, we validated the main results with additional tumor GWAS data. For external queue validation results, the threshold was adjusted to $p < 0.05$. Finally, MR Analysis of all pQTL data (cis-pQTL and trans-pQTL) was also used as a sensitivity analysis in order to verify the main results. The all methods used for the sensitivity analysis are shown in Fig. 1. Additionally, to verify the reliability of the results, we conducted a statistical power analysis for the main outcomes using the online website <https://sb452.shinyapps.io/power>.

Reverse MR analysis

In order to exclude the bias caused by reverse causality, we conducted a bidirectional MR analysis. According to the same screening conditions ($p < 5e-8$, kb = 10,000, $r^2 = 0.001$), SNPs were extracted from the GWAS data of breast tumor, colorectal cancer, and non-small cell lung cancer in the FinnGen database, with tumor as exposure and plasma protein concentration as outcome. IVW, MR-Egger, weighted median, simple mode, and weighted mode were used as the primary analysis methods^{31–33}. $p < 0.05$ was considered statistically significant.

Colocalization analysis

Colocalization analysis is a method of investigating whether a causal variation is shared between two traits and is used to enhance the results of MR Analysis³⁴. The Bayesian colocalization method provides posterior probabilities of five hypotheses³⁵. To assess the presence of shared causal variation in a specific genomic region between the two phenotypes, Bayesian colocalization analysis (pQTL-GWAS) was performed using SNPs located within 500 Mb from the lead SNP of all cis-pQTL with MAF > 0.01 ²³. There are four hypotheses for co-localization analysis. H0: Phenotype 1 and Phenotype 2 are not significantly associated with any SNP loci in a specific genomic region. H1/H2: Phenotype 1 or Phenotype 2 is significantly associated with SNP loci in a specific genomic region. H3: Both Phenotype 1 and Phenotype 2 are significantly associated with SNP loci in a specific genomic region, but driven by different causal variant positions. H4: Both Phenotype 1 and Phenotype 2 are significantly associated with SNP loci in a specific genomic region, and driven by the same causal variant position. If H4 has a higher probability in statistical terms, it suggests that the significant signal locus can influence the phenotype. In this study, based on the coloc package, we evaluated the posterior probability of hypothesis 4 (PP.H4). PP.H4 > 0.8 was identified

as being associated with a specific genomic region through shared causal variation between the protein and the tumor³⁶, which is strong evidence for co-localization. $0.4 < \text{PP.H4} < 0.8$ was regarded as moderate co-localization.

Potential target drugs

Plasma proteins with strong co-localization and favorable results in sensitivity analyses were selected as first-class targets. In addition, the STRING database (<https://string-db.org/>) and DrugBank (<https://go.drugbank.com/>) were utilized to investigate whether discovered cancer therapeutic drug targets interact with the first-class targets³⁷. Furthermore, the DrugBank database is being used separately to explore potential cancer therapy drugs for first-class targets. The molecular docking of potential target drugs was performed on MCULE online platform with default settings (<https://mcule.com/apps/1-click-docking/>)^{38–41}. The platform will generate top four ranked of molecular docking results according to the drug molecules and targets provided, and we selected Rank 1 to prepare the figures.

Results

Proteomic screening of tumor-causing proteins

In our results, 5365 cis-pQTL of a total of 1743 proteins were used for follow-up analysis after a series of screening (Supplement Table S2). However, after extracting SNPs in the outcome, some lacked SNPs, leaving SNPs with only 1653 plasma proteins in every tumor. A total of 1005 proteins were causally associated with tumors at the $p < 0.05$ threshold (Supplement Table S3), while there were 6 significant outcomes at the $\text{FDR} < 0.05$ threshold (NCAN, GREM1, S100A16, S100A14, PDE5A, MIA), and only the following 4 proteins (S100A16, S100A14, PDE5A, MIA) were causally associated with tumors after removing the SNPs associated with the outcomes by phenoscanner (Fig. 2). An instrumental variable of NCAN, rs2228603, was associated with nonalcoholic fatty liver disease. The instrumental variables of GREM1, rs4779584 and rs144674978, were associated with colorectal cancer. The statistical significance of NCAN and GREM1 did not reach the threshold of $\text{FDR} < 0.05$. In non-small cell lung cancer, rs2279011 was excluded because of its association with smoking. There were causal associations between S100A14 (Wald ratio, OR = 2.11, 95% CI, 1.50–2.96, $\text{FDR} = 0.0144$) and S100A16 (IVW, OR = 0.70, 95% CI, 0.60–0.82, $\text{FDR} = 0.0144$) and HER-positive breast tumors, between PDE5A (IVW, OR = 1.53, 95% CI, 1.25–1.87, $\text{FDR} = 0.0304$) and colorectal cancer, and between MIA (IVW, OR = 0.82, 95% CI, 0.76–0.88, $\text{FDR} = 0.0005$) and non-small cell lung cancer. And they had the correct causal direction ($p < 0.05$). There was no heterogeneity (SNP > 2) and pleiotropy (SNP > 3) in results (Supplement Table S3). These results suggest that S100A16 may be a risk protein for HER-positive breast cancer, S100A14 may be a protective protein for HER-positive breast cancer, PDE5A may be a risk protein for colorectal cancer, and MIA may be a risk protein for non-small cell lung cancer. Additionally, the results have been validated as reliable through statistical power analysis (Supplement Table S4).

Sensitivity analysis of tumor causal proteins

In our findings, four proteins were ultimately identified as potential drug targets for tumors, including S100A16, S100A14, PDE5A, and MIA. In addition, Reverse validation, Bayesian colocalization, MR analysis of the total pQTL, and validation of external queue were used as sensitivity analyses. In external queue validation, the results for S100A16 (IVW, OR = 0.99, 95% CI, 0.91–1.07, $p = 0.722$), S100A14 (Wald ratio, OR = 1.08, 95% CI, 0.90–1.30, $p = 0.406$) and PDE5A (Wald ratio, OR = 0.99, 95% CI, 0.98–1.00, $p = 0.105$) were not replicated, while the results for MIA (IVW, OR = 0.90, 95% CI, 0.83–0.97, $p = 0.009$) were replicated in non-small cell lung cancer (Fig. 3). And there was no heterogeneity and pleiotropy ($p > 0.05$) (Supplement Table S3). Moreover, in reverse MR Analysis, tumors were considered exposure, and plasma proteins were considered outcome. No causal association was found between tumors and plasma protein (Fig. 4). Meanwhile, to validate our main results, all pQTL data of four plasma proteins (cis-pQTL and trans-pQTL) were used for MR Analysis, the same findings were observed for plasma proteins PDE5A (IVW, OR = 1.39, 95% CI, 1.13–1.71, $p = 2e-3$) and MIA (IVW, OR = 0.84, 95% CI, 0.79–0.90, $p = 2e-7$) (Fig. 5). Interestingly, the results of colocalization analysis showed that S100A16 (coloc, $\text{PP.H4.abf} = 0.933$), S100A14 (coloc, $\text{PP.H4.abf} = 0.913$), and PDE5A (coloc, $\text{PP.H4.abf} = 0.855$) had the shared causal variation with HER-positive breast and colorectal cancers, respectively (Fig. 2). These results further validated the causal association between four plasma proteins and tumors.

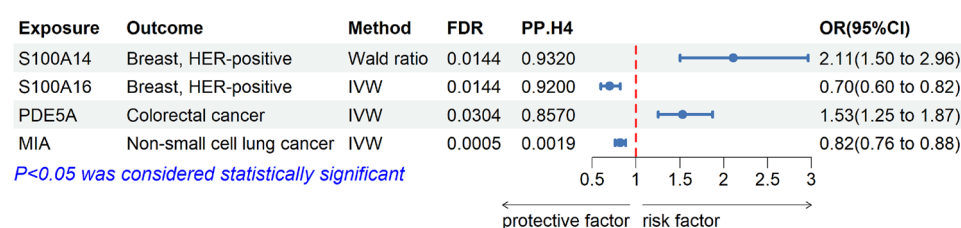


Figure 2. MR results for plasma proteins on the risk of tumors. Causal association of four plasma proteins (S100A14, S100A16, PDE5A, MIA) with tumors based on MR analysis with Wald ratio and Inverse Variance Weighted methods.

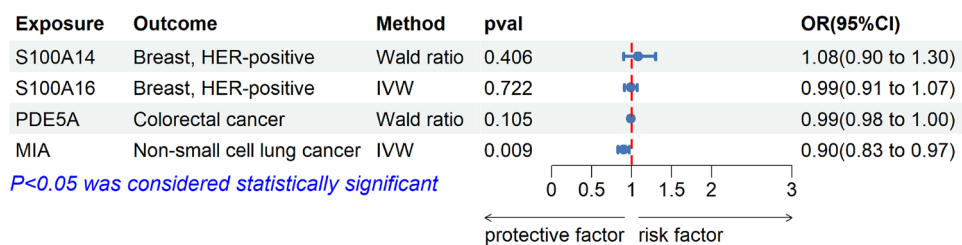


Figure 3. External queue validation of MR results for plasma proteins on the risk of tumors. External data for the four plasma proteins came from the UK Biobank and the Breast Cancer Association Consortium (BCAC).

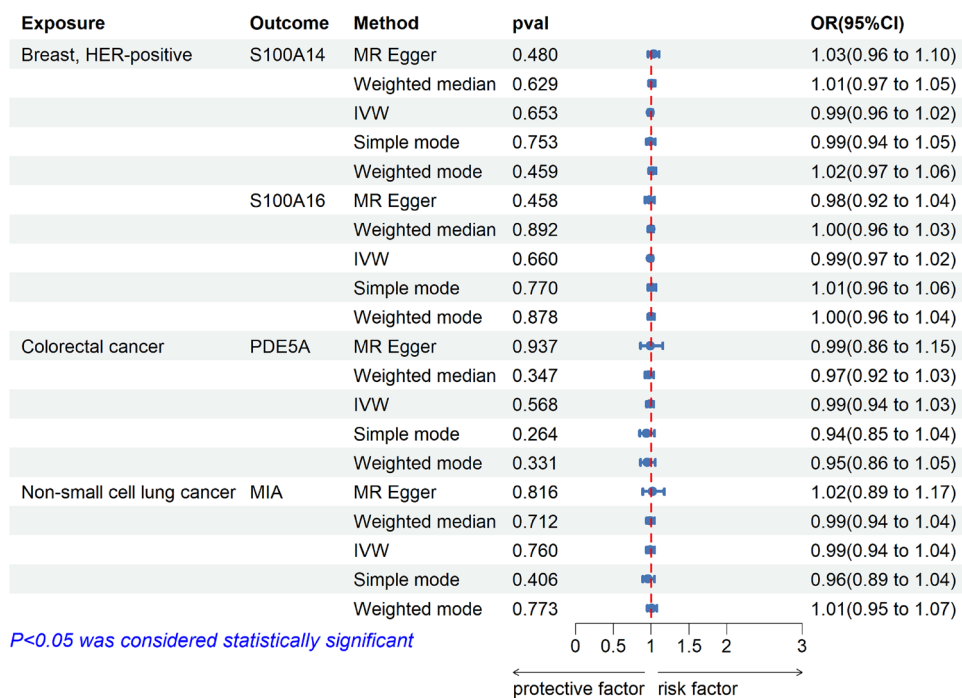


Figure 4. Reverse MR results for plasma proteins on the risk of tumors. There was no reverse causal effect on the primary MR results.

Potential therapeutic drugs for tumors

Based on the above results, we considered PDE5A as a first-class therapeutic target, S100A16, S100A14 and MIA as second-class targets, and the other proteins with $p < 0.05$ threshold as third-class targets. While indirect potential drug targets were not identified through STRING, we discovered drugs in the medication database that may interact with the targets. Through the search of DrugBank database, we found that target of the marketed drugs Sildenafil, Vardenafil, Dipyridamole, Theophylline, Tadalafil and Avanafil was PDE5A, and they were used as inhibitors of PDE5A (Fig. 6). These results suggest that PDE5A inhibitor drugs may have a preventive effect as well as a potential therapeutic effect on colorectal cancer. However, there were no suitable drugs targeting S100A16, S100A14 and MIA. These findings suggest that targeting PDE5A holds significant potential as a therapeutic target for colorectal cancer in future clinical investigations.

Discussion

In this study, we primarily used the two-sample MR method to investigate the causal association between human plasma protein concentration and tumors, aiming to discover potential drug therapeutic targets and cancer prevention biomarkers. Based on our research, we set a criterion: Proteins with strong colocalization and robust sensitivity were identified as potential first-class targets, while proteins that met only one condition were identified as second-class potential targets. Ultimately, based on our analysis, S100A16 and S100A14 were validated as potential second-class drug targets for HER-positive breast cancer, PDE5A as potential first-class drug targets for colorectal cancer, and MIA as potential second-class drug targets for non-small cell lung cancer. In these findings, the causal relationship between MIA and lung cancer was further confirmed in external queue (ieu-b-4954), providing additional credibility to the potential drug targets. Our analysis also revealed that MR Analyses of trans-pQTL and cis-pQTL for PDE5A and MIA were consistent with the main results.

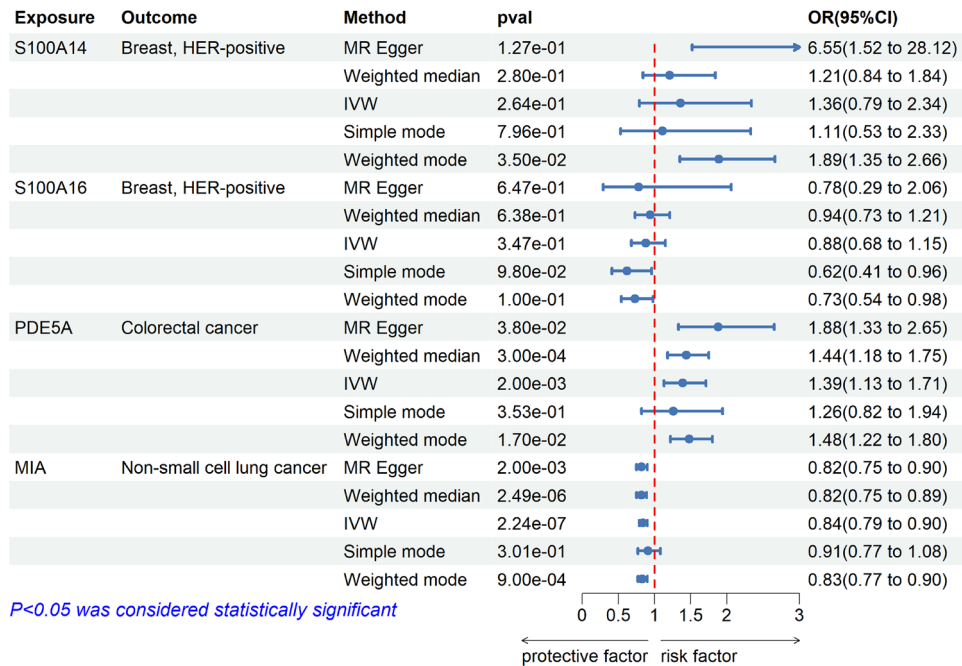


Figure 5. MR results of all pQTL for plasma proteins on the risk of tumors. The results of PDE5A and MIA were replicated in all PQTL (cis-pQTL and Trans-pQTL) data as exposed.

Firstly, in order to eliminate the effects of confounding factors and horizontal pleiotropy, PhenoScanner was used to remove SNPs associated with outcomes and confounding factors before main analysis. Secondly, evaluation results based on data from previous drug development programs showed that target indication pairings identified by MR and co-localization were more likely to be approved⁴². To assess the potential of these findings as targets for tumor drug therapy, we further evaluated the causal association between plasma proteins and tumors through colocalization analysis⁴³. $HH.P4 > 0.8$ was considered evidence of strong colocalization. Three proteins identified by MR, S100A16, S100A14 and PDE5A, might share causal variables with HER-positive breast cancer and colorectal cancer. In addition, Reverse causation MR analysis was performed to remove the effects of reverse causation. Results showed no reverse causal association between plasma proteins and tumors.

In our analysis, S100A14, S100A16 and MIA were regarded as second-class target of HER-positive breast cancer. S100A14 and S100A16 both belong to calcium-binding protein. These proteins were involved in various biological functions of the human body, such as cell signal transduction, cell proliferation, and apoptosis^{44,45}. Studies had shown that S100A14 might be related to the breast cancer metastasis by promoting expression of chemokines CCL2 and CXCL5⁴⁶. Mechanistically, S100A14 activated NF- κ B signaling to upregulate chemokine expression. This was consistent with our findings that S100A14 was considered a risk protein for HER-positive breast cancer. Similarly, another study showed that S100A14 was an independent prognostic factor in triple-negative breast cancer⁴⁷. In addition, expression of S100A16 activated epithelial-mesenchymal transition and promoted breast cancer progression⁴⁸, which was contrary to our findings. Interestingly, a study had shown that co-expression of S100A16 and S100A14 in breast cancer promoted the invasive ability of cancer cells; overexpression and knockdown of S100A14 could affect the expression of S100A16⁴⁹. Therefore, we speculated that S100A14 might play a leading role in breast cancer. In addition, MIA was regarded as second-class target of non-small cell lung cancer in our analysis. MIA was a small secreted protein that was primarily secreted by but not limited to melanoma⁵⁰. Overexpression of MIA not only promoted melanoma metastasis but also activated the invasive ability of pancreatic cancer cells. However, the association between MIA and lung cancer had not been studied so far. Moreover, PDE5A were regarded as first-class target of colorectal cancer. So far, multiple studies had shown PDE5A as a potential therapeutic target for tumors. PDE5A was overexpressed in breast cancer stroma and affected the growth of breast cancer cells by inducing the expression of chemokine CXCL16⁵¹. In addition, the use of the PDE5A inhibitors sildenafil or vardenafil enhanced apoptosis of human castration-resistant prostate cancer cells⁵². In colorectal cancer, the use of PDE5A inhibitors had been found to be associated with a favorable prognosis and lower metastasis rate in male with colorectal cancer⁵³. To our knowledge, no previous studies had used MR combined with colocalization to explore the causal association between PDE5A and colorectal cancer. Our results provide evidence for the clinical translation of PDE5 inhibitors in treatment of colorectal cancer.

Although the number of potential therapeutic targets identified had increased in recent years, that had not translated into clinical outcomes. Based on the phenomenon, we used cis-pQTL for MR and colocalization analysis to identify potential therapeutic targets for tumors, which will be more likely to be approved.

Despite our study identifying potential targets and drugs advantageous for clinical cancer treatment, however, there are some limitations in this analysis. The limitations of this study are as follows: First, we extracted 4917

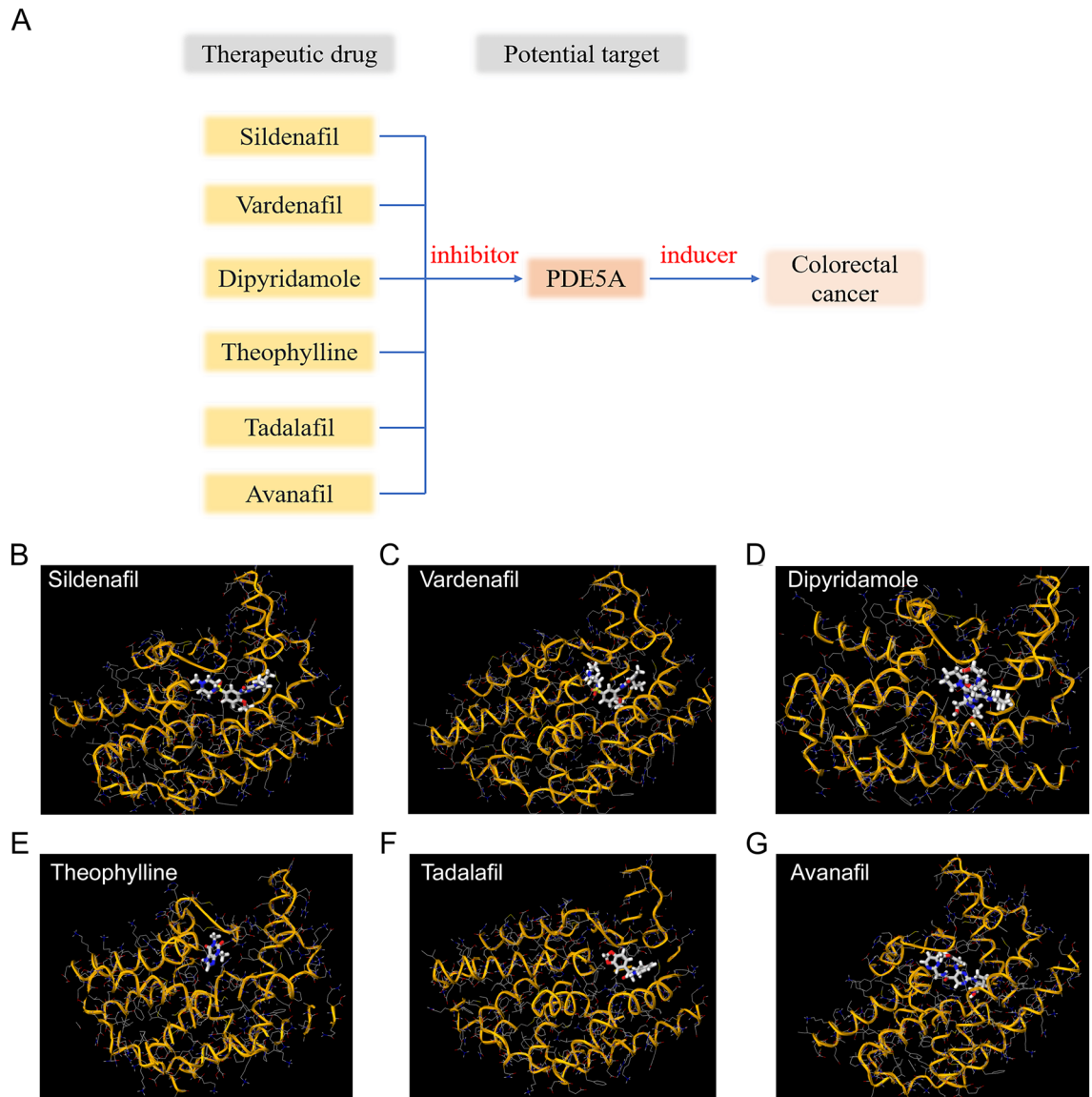


Figure 6. Potential target drugs for colorectal cancer based on MR analysis. Potential target drugs came from DrugBank database. (A) Potential drugs for the target PDE5A in colorectal cancer; (B–G) Molecular docking results of Sildenafil, Vardenafil, Dipyridamole, Theophylline, Tadalafil and Avanafil based on MCULE online platform.

protein pQTLs from Ferkingstad et al.²¹ for exposure in MR analysis. After preliminary processing, only 1653 plasma proteins remained for analysis, and many potential plasma proteins still await validation. In addition, external data of exposure were not further verified. The results may be subject to some bias. Second, cis-pQTL was reserved for analysis, and some pQTL had only one instrumental variable, so it could not be analyzed for heterogeneity or pleiotropy. Although we subsequently added trans-pQTL, there may be some bias in the results. However, we subsequently enhanced the accuracy of the results by employing colocalization analysis. Thirdly, our study data originated from the decode data and the FinnGen cohort, both of which consist of European populations. Therefore, our results may be applicable primarily to European populations. In addition, it is well known that different tumor subtypes have different therapeutic strategies based on different information such as tumor marker expression, driver gene mutation, and tumor stage and grade, etc. In this study, because FinnGen database did not provide further clinical classification based on clinicopathological information such as gene alteration and tumor stages/grade, etc., we could not analyze the correlation between clinical subtypes of related tumors. Fourth, as we focused on exploring potential drug targets for cancer, the conclusions drawn have not been validated through clinical trials. Therefore, these results require further investigation.

Conclusions

Overall, our study employed MR analysis to investigate causal relationships between plasma proteins and nine types of cancer. We found that elevated S100A14 in circulation increased the risk of HER-positive breast cancer, while S100A16 decreased the risk of HER-positive breast cancer. PDE5A was associated with an increased risk of CRC, and MIA was linked to a reduced risk of SCLC. Additionally, we identified PDE5A, a target for erectile dysfunction drugs, as a potential therapeutic target for CRC, suggesting that such medications may serve as novel treatments by inhibiting PDE5A. However, further research is needed to validate these findings and understand their specific roles in tumor diagnosis and drug target development.

Data availability

All data for this study is available from the corresponding author upon reasonable request.

Received: 4 February 2024; Accepted: 14 May 2024

Published online: 18 May 2024

References

- Siegel, R. L. *et al.* Cancer statistics, 2023. *CA Cancer J. Clin.* **73**, 17–48 (2023).
- Pinsky, P. F. *et al.* Screening for prostate cancer. *N. Engl. J. Med.* **388**, 1405–1414 (2023).
- Nolan, E. *et al.* Deciphering breast cancer: from biology to the clinic. *Cell.* **186**, 1708–1728 (2023).
- Sherman, M. H. *et al.* Tumor microenvironment in pancreatic cancer pathogenesis and therapeutic resistance. *Annu. Rev. Pathol.* **18**, 123–148 (2023).
- Cai, M. *et al.* Current therapy and drug resistance in metastatic castration-resistant prostate cancer. *Drug Resist. Updat.* **68**, 100962 (2023).
- Li, F. *et al.* Regulation of cisplatin resistance in bladder cancer by epigenetic mechanisms. *Drug Resist Updat.* **68**, 100938 (2023).
- Vodenkova, S. *et al.* 5-fluorouracil and other fluoropyrimidines in colorectal cancer: Past, present and future. *Pharmacol. Ther.* **206**, 107447 (2020).
- Ashrafzadeh, M. *et al.* New insight towards development of paclitaxel and docetaxel resistance in cancer cells: EMT as a novel molecular mechanism and therapeutic possibilities. *Biomed. Pharmacother.* **141**, 111824 (2021).
- Verma, S. *et al.* “Targeting” improved outcomes with antibody-drug conjugates in non-small cell lung cancer—an updated review. *Curr. Oncol.* **30**, 4329–4350 (2023).
- Desai, A. *et al.* Antibody-drug conjugates: A promising novel therapeutic approach in lung cancer. *Lung Cancer.* **163**, 96–106 (2022).
- Moreau Bachelard, C. *et al.* Risks and benefits of anticancer drugs in advanced cancer patients: A systematic review and meta-analysis. *EClinicalMedicine* **40**, 101130 (2021).
- Gong, Y. *et al.* PUMILIO proteins promote colorectal cancer growth via suppressing p21. *Nat. Commun.* **13**, 1627 (2022).
- Wu, J. *et al.* CMTM family proteins 1–8: Roles in cancer biological processes and potential clinical value. *Cancer Biol. Med.* **17**, 528–542 (2020).
- Lin, J. *et al.* Potential drug targets for multiple sclerosis identified through Mendelian randomization analysis. *Brain* **146**, 3364–3372 (2023).
- Rosoff, D. B. *et al.* Mendelian randomization study of PCSK9 and HMG-CoA reductase inhibition and cognitive function. *J. Am. Coll. Cardiol.* **80**, 653–662 (2022).
- Zhu, G. *et al.* Chickenpox and multiple sclerosis: A Mendelian randomization study. *J. Med. Virol.* **95**, e28315 (2023).
- Sun, J. X. *et al.* The association between human papillomavirus and bladder cancer: Evidence from meta-analysis and two-sample mendelian randomization. *J. Med. Virol.* **95**, e28208 (2023).
- Birney, E. Mendelian Randomization. *Cold Spring Harb Perspect Med.* **12**, (2022).
- Larsson, S. C. *et al.* Mendelian randomization for cardiovascular diseases: Principles and applications. *Eur. Heart J.* **44**, 4913–4924 (2023).
- Choi, K. W. *et al.* Assessment of bidirectional relationships between physical activity and depression among adults: A 2-sample Mendelian randomization study. *JAMA Psychiatry.* **76**, 399–408 (2019).
- Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* **53**, 1712–1721 (2021).
- Yang, C. *et al.* Mendelian randomization and genetic colocalization infer the effects of the multi-tissue proteome on 211 complex disease-related phenotypes. *Genome. Med.* **14**, 140 (2022).
- Bourgault, J. *et al.* Proteome-wide Mendelian randomization identifies causal links between blood proteins and acute pancreatitis. *Gastroenterology* **164**, 953–965.e953 (2023).
- Staley, J. R. *et al.* PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics* **32**, 3207–3209 (2016).
- Bottigliengo, D. *et al.* A Mendelian randomization study investigating the causal role of inflammation on Parkinson's disease. *Brain* **145**, 3444–3453 (2022).
- Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
- Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92–94 (2017).
- Chen, D. *et al.* Assessing causality between osteoarthritis with urate levels and gout: a bidirectional Mendelian randomization study. *Osteoarthritis Cartilage* **30**, 551–558 (2022).
- Gossmann, A. *et al.* FDR-corrected sparse canonical correlation analysis with applications to imaging genomics. *IEEE Trans. Med. Imaging.* **37**, 1761–1774 (2018).
- Freuer, D. *et al.* Association between inflammatory bowel disease and both psoriasis and psoriatic arthritis: A bidirectional 2-sample mendelian randomization study. *JAMA Dermatol.* **158**, 1262–1268 (2022).
- Huang, D. *et al.* Association between COVID-19 and telomere length: A bidirectional Mendelian randomization study. *J. Med. Virol.* **94**, 5345–5353 (2022).
- Wu, F. *et al.* Mendelian randomization study of inflammatory bowel disease and bone mineral density. *BMC Med.* **18**, 312 (2020).
- Li, P. *et al.* Association between gut microbiota and preeclampsia-eclampsia: A two-sample Mendelian randomization study. *BMC Med.* **20**, 443 (2022).
- Bouras, E. *et al.* Circulating inflammatory cytokines and risk of five cancers: A Mendelian randomization analysis. *BMC Med.* **20**, 3 (2022).
- Han, X. *et al.* Integrating genetics and metabolomics from multi-ethnic and multi-fluid data reveals putative mechanisms for age-related macular degeneration. *Cell Rep. Med.* **4**, 101085 (2023).
- Chen, Y. *et al.* Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. *Nat. Genet.* **55**, 44–53 (2023).
- Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–d1082 (2018).

38. Muteeb, G. *et al.* Identification of a Potential Inhibitor (MCULE-8777613195-0-12) of New Delhi Metallo-beta-Lactamase-1 (NDM-1) Using In Silico and In Vitro Approaches. *Molecules*. **27**, (2022).
39. Odhar, H. A. *et al.* Molecular docking and dynamics simulation analysis of the human FXIIa with compounds from the Mcule database. *Bioinformatics* **19**, 160–166 (2023).
40. Odhar, H. A. *et al.* Molecular docking enabled updated screening of the matrix protein VP40 from Ebola virus with millions of compounds in the MCULE database for potential inhibitors. *Bioinformatics* **15**, 627–632 (2019).
41. Onawole, A. T. *et al.* Identification of potential inhibitors against the Zika virus using consensus scoring. *J. Mol. Graph. Model.* **73**, 54–61 (2017).
42. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* **52**, 1122–1131 (2020).
43. Lin, Y. *et al.* Circulating inflammatory cytokines and female reproductive diseases: a Mendelian randomization analysis. *J Clin Endocrinol Metab.* (2023).
44. Zhu, M. *et al.* Calcium-binding protein S100A14 induces differentiation and suppresses metastasis in gastric cancer. *Cell Death Dis.* **8**, e2938 (2017).
45. Zhang, W. S. *et al.* S100a16 deficiency prevents hepatic stellate cells activation and liver fibrosis via inhibiting CXCR4 expression. *Metabolism* **135**, 155271 (2022).
46. Li, X. *et al.* A S100A14-CCL2/CXCL5 signaling axis drives breast cancer metastasis. *Theranostics* **10**, 5687–5703 (2020).
47. Ehmsen, S. *et al.* S100A14 is a novel independent prognostic biomarker in the triple-negative breast cancer subtype. *Int. J. Cancer* **137**, 2093–2103 (2015).
48. Zhou, W. *et al.* Up-regulation of S100A16 expression promotes epithelial-mesenchymal transition via Notch1 pathway in breast cancer. *J. Biomed. Sci.* **21**, 97 (2014).
49. Tanaka, M. *et al.* Co-expression of S100A14 and S100A16 correlates with a poor prognosis in human breast cancer and promotes cancer cell invasion. *BMC Cancer* **15**, 53 (2015).
50. El Fitori, J. *et al.* Melanoma Inhibitory Activity (MIA) increases the invasiveness of pancreatic cancer cells. *Cancer Cell. Int.* **5**, 3 (2005).
51. Catalano, S. *et al.* Phosphodiesterase 5 (PDE5) Is Highly Expressed in Cancer-Associated Fibroblasts and Enhances Breast Tumor Progression. *Cancers*. **11**, (2019).
52. Chang, J. F. *et al.* Phosphodiesterase type 5 (PDE5) inhibitors sensitize topoisomerase II inhibitors in killing prostate cancer through PDE5-independent impairment of HR and NHEJ DNA repair systems. *Front Oncol.* **8**, 681 (2018).
53. Huang, W. *et al.* Phosphodiesterase-5 inhibitors use and risk for mortality and metastases among male patients with colorectal cancer. *Nat. Commun.* **11**, 3191 (2020).

Acknowledgements

The authors are grateful to the participants and investigators of the FinnGen and decode study, as well as the open GWAS for providing the research information. We also thanks to Dr. Zhanjun Li for the help and guidance in molecular docking.

Author contributions

H.W and N.S conceived and coordinated the study. K.C and N.S wrote the paper. K.C, L.Z, J.Z and Y.F analyzed the data and Figures & Tables, P.S, Z.W, W.D, J.L and W.S for the data processing, N.S, K.C and H.W reviewed and revised the article. H.W and Z.L for the molecular docking work. All authors contributed to interpretation of data, manuscript revision, and critical discussion. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Program for Young Key Teachers in Colleges and Universities in Henan Province (2020GGJS150, 2021GGJS104), and Key Medical Science and Technology Research Program Project of Henan Province (No. 20232028).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-62178-w>.

Correspondence and requests for materials should be addressed to W.S. or H.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024