scientific reports

Check for updates

OPEN

Identification of biomarkers and potential drug targets for esophageal cancer: a Mendelian randomization study

Chengjun Li[1], Xiaomeng Cui[2], Mudan Ren[1], Yan Yin[1] & Shuixiang He[1]✉

Esophageal cancer (EC) is a common and deadly malignancy of the digestive system. Currently, effective treatments for EC are limited and patient prognosis remains poor. In this study, we utilized Mendelian Randomization (MR) to identify potential drug targets for EC by analyzing proteins linked to the disease risk. A total of 734 plasma proteins and 4,479 druggable genes were obtained from recent studies, and two-sample MR analyses were conducted to investigate causal relationships between these proteins and EC. The cis-pQTL data of the proteins was analyzed after filtering. The inverse variance weighted (IVW) method was the primary analytical approach in MR analysis. Steiger filtering, heterogeneity and pleiotropy tests, Summary-data-based Mendelian Randomization (SMR) analysis, and Bayesian co-localization analysis were implemented to consolidate the results further. Moreover, drugs corresponding to the identified proteins were found in the DrugBank database. Five proteins HPSE, ST3GAL1, CEL, KLK13, and GNRH2 were identified as highly associated with EC. HPSE and GNRH2 showed protective effects with odds ratios (OR) of 0.80 (95% confidence interval [CI], 0.70–0.92) and 0.73 (95% CI 0.54–0.98), respectively. In contrast, increased expression of ST3GAL1(OR, 1.37; 95% CI 1.04–1.82), CEL (OR, 1.27; 95% CI 1.08–1.49), and KLK13 (OR, 1.22; 95% CI 1.04–1.42) were all associated with a higher risk of EC. In addition, the HPSE protein showed moderate colocalization with EC [coloc.abf-posterior probability of hypothesis 4 (PPH4) = 0.637]. Furthermore, the sensitivity analyses indicated no heterogeneity or pleiotropy. Therefore, these findings present promising drug targets for EC and deserve further clinical investigation.

Keywords Esophageal cancer, Plasma protein, Druggable gene, Mendelian randomization, Drug target, Causality

Esophageal cancer (EC) ranks seventh in global incidence and is the sixth leading cause of cancer-related deaths worldwide[1]. Consequently, EC causes over 500,000 deaths annually, accounting for 5.3% of global cancer deaths and posing a serious threat to human health[2]. The two main histological subtypes of EC are esophageal squamous cell carcinoma (ESCC)and esophageal adenocarcinoma (EAC), with ESCC being the most common type worldwide. There are often no symptoms in the early stages of EC. However, progressive difficulty in swallowing, chest pain, and weight loss gradually develop over time. Furthermore, some patients may experience symptoms such as hoarseness and anemia[3]. Unfortunately, both EAC and ESCC have low survival rates and are prone to early metastasis.

The treatment strategies for EC can be conceptually divided into two categories: local treatment and systemic treatment. As such, each patient should be individually assessed to determine the appropriate treatment plan based on the type of cancer, the extent of local or regional involvement, and the patient's overall functional status[4]. According to the TNM staging of AJCC Cancer Staging, endoscopic mucosal resection (EMR) is recommended for T1a stage tumors (confined to the mucosa) that are smaller than 2 cm[5]. For T1b stage tumors (invasion into the superficial submucosa), endoscopic submucosal dissection is the therapeutic option. When the tumor is larger than 2 cm or at the T1b stage (invasion into the deep submucosa), a multidisciplinary and multimodal treatment approach is optimal[6].For the T2 and T3 stages, patients typically undergo neoadjuvant chemotherapy followed by surgery. Standard chemotherapy regimens include cisplatin, 5-fluorouracil, carboplatin, and

[1]Department of Gastroenterology, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, Shaanxi, China. [2]Department of Infectious Diseases, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710004, Shaanxi, China. ✉email: dyyyjxk@mail.xjtu.edu.cn

paclitaxel. Moreover, unresectable lesions are generally managed with palliative chemotherapy and radiotherapy. Immunotherapy and small molecule inhibitors are also often used in combination with chemotherapy, radiation therapy, and targeted therapy to enhance overall effectiveness[4]. However, long-term use of chemotherapy drugs and targeted therapies may lead to drug resistance, often accompanied by significant side effects that may impact patients' quality of life. In addition, some patients show poor response to current drug treatment regimens, resulting in limited therapeutic efficacy. Therefore, there is an urgent need to develop more effective chemical compounds to treat this malignant disease.

Accurately identifying candidate drugs for a specific disease is crucial for effective drug development targeting that disease. However, clinical trials for newly developed drugs are terminated due to small sample sizes, insufficient drug efficacy or safety data, and unsatisfactory results[7]. Human plasma proteins are vital in various biological processes and are primary drug targets. Studies have shown that if genetic associations support the link between a protein drug target and disease, its success rate in clinical development can double[8]. In recent years, Mendelian randomization (MR) studies have been widely applied in drug target development[9]. Notably, this approach uses single nucleotide polymorphisms (SNPs) from genome-wide association studies (GWAS) as genetic instrumental variables to investigate the causal effects of exposures on outcomes. In contrast to observational studies, MR can effectively avoid confounding factors and reverse causation, providing more reliable evidence for causal inference. With the advancement of high-throughput genomics and proteomics technologies in plasma, MR-based studies have made it feasible to identify potential therapeutic drug targets for many diseases including Parkinson's disease[10] and Alzheimer's disease[11].However, no studies have explored potential drug targets using MR analyses in EC. Furthermore, MR studies integrating GWAS data with protein quantitative trait loci (pQTL) data are rarely reported. Therefore, this study aimed to identify plasma proteins as potential therapeutic targets for EC.

This study is the first to identify potential drug targets for EC through MR analysis at the genetic level. First, we intersected the 734 plasma proteins identified by Zheng et al.[12] with the 4,479 druggable gene-encoded proteins from Finan et al.[13], and obtained 511 plasma proteins encoded by druggable genes. A two-sample MR analysis was conducted using the GWAS summary statistics for EC from the IEU Open GWAS database to identify the causal relationships between the druggable gene-encoded proteins and EC. After obtaining and filtering the pQTL data for these proteins from the deCODE database, we identified cis-pQTL data for 19 proteins. Subsequently, we performed the two-sample MR analysis to identify proteins with causal associations with EC. Heterogeneity and pleiotropy tests, Steiger filtering, SMR analysis, and Bayesian colocalization analysis validated our preliminary findings, ensuring the reliability and stability of the protein targets identified in EC. Importantly, these targets may become novel drug targets in the future. Finally, we retrieved the drugs for the identified proteins from the DrugBank database. Thus, the selected drugs may be promising for future treatment of EC.

Materials and methods

Druggable gene selection

We obtained a total of 4479 druggable genes from Finan's study[13] (Supplementary Data 1). The first group of 1427 genes included the efficacy targets of approved small molecule and biological therapeutic drugs and the clinical stage drug candidates. The second group of 682 genes encoded targets with known bioactive drug-like small molecule binding partners and ≥ 50% identity with approved drug targets (over ≥ 75% of the sequence). The third group of 2370 genes encoded secreted or extracellular proteins, or proteins with slightly lower similarity to approved drug targets, as well as key druggable gene family members not yet included in the first and second groups.

pQTL dataset

pQTL is utilized to study the correlation between genetic mutations and gene expression using protein expression as a quantifiable trait. Two sources of pQTL were used for the analysis. For the preliminary analysis, we used the cis-pQTL data from the study by Zheng et al.[12], including a total of 738 cis-SNPs (Single Nucleotide Polymorphism) of 734 proteins (Supplementary Data 2), to screen the druggable proteins. Then we obtained pQTL data of the screened proteins from 4674 proteins in the deCODE database from Ferkingstad's study[14]. These data were used as the main pQTL to identify potential drug targets for EC. Furthermore, the instrumental variable cis-pQTL was selected using the following criteria: (1) P value < 5e-08; (2) F statistic > 10; (3) SNPs in the human major histocompatibility complex (MHC) region (chr6, 26-34 MB) were excluded; (4) SNPs within 1 Mb upstream and downstream of the gene; (5) removal of linkage disequilibrium $r^2 < 0.1$; (6) *P*value > 5e-08 with outcome.

The selected datasets were all of European ethnic background.

Outcome dataset

We obtained the GWAS summary data of EC from the online IEU Open GWAS database (https://gwas.mrcieu.ac.uk/)[15] using the R package TwoSampleMR. A total of 998 samples with EC in the experimental group and 475,308 control samples were included. The outcome data were from European ethnic backgrounds.

Two-sample Mendelian randomization analysis

We conducted a two-sample Mendelian randomization analysis using the TwoSampleMR package, with the druggable protein studied by Zheng et al. as the exposure factor, and esophageal cancer as the outcome[12]. To ensure the independence assumption's validity, strict criteria were applied when selecting SNPs, including excluding SNPs associated with the MHC region and controlling for linkage disequilibrium, to minimize the impact of confounding bias. Moreover, we used the Wald ratio method to evaluate the results of Mendelian

randomization for exposures containing only one SNP. The IVW method was also used to assess the results of Mendelian randomization for exposures containing two or more SNPs. Finally, we used the TwoSampleMR package for the heterogeneity test and pleiotropy test, followed by the Steiger directional test to determine the correctness of the direction of causality.

After selecting the proteins with significant causal relationships in EC, the pQTL data of the corresponding proteins were downloaded from the deCODE database as the exposure factors and esophageal cancer as the outcome. Using the same approach described above, a two-sample MR analysis was performed.

SMR analysis

SMR is a statistical method that utilizes GWAS and expression quantitative trait loci (eQTL) studies to investigate the pleiotropic associations between gene expression levels and complex traits of interest[16]. Additionally, the heterogeneity in dependent instruments (HEIDI) test is used to assess the presence of possible horizontal pleiotropy in colocalization signals. The null hypothesis of the HEIDI test is that there is no horizontal pleiotropy in the colocalization signals. SMR and HEIDI methods can be used to explain whether protein expression rather than other pathways mediate the effect of SNPs on phenotypes. We downloaded the SMR Linux version (1.3.1) from the website (https://yanglab.westlake.edu.cn/software/smr) and conducted the SMR analysis using the default parameters.

Co-localization analysis

We used the coloc package for colocalization analysis. The coloc package uses a Bayesian approach to assess the support for the following five exclusivity hypotheses: (1) SNPs were not associated with trait1 or trait2; (2) SNPs were associated with trait1; (3) SNPs were associated with trait2; (4) SNPs were associated with both trait1 and trait2 but were independent SNPs; (5) SNPs were associated with both trait1 and trait2 and they were shared. The posterior probabilities of each test were H0, H1, H2, H3 and H4. To estimate the posterior probability of shared variation, for each selected protein, all SNPs within 250 kb upstream and downstream of its top SNP were retrieved for colocalization analysis. We considered a posterior probability of hypothesis 4 (PPH4)>0.8 as evidence of colocalization between GWAS and pQTL.

Drug target analysis

We searched in DrugBank[17] to identify drugs corresponding to the protein and their modes of action. Notably, the selected drugs can be used to treat EC.

Statistical methods

All data calculations and statistical analyses were performed using R software (https://www.r-project.org/?2827). All p-values were considered statistically significant at $P<0.05$ if not specified. In GWAS studies, SNPs are typically selected as instrumental variables (IVs) for MR analysis. However, due to the multiple comparisons across the genome, a more stringent p-value < 5e-08 was used to reduce false positives and enhance statistical significance. This p-value threshold is a recognized standard in GWAS analyses for identifying significant SNPs associated with complex traits or diseases across the entire genome[18,19].

Results

Technology roadmap

The analysis flow of this study is shown in Fig. 1.

MR analysis of druggable proteins

Firstly, we intersected the 734 proteins studied by Zheng et al.[12] with 4479 proteins encoded by druggable genes, resulting in 511 proteins encoded by druggable genes. Subsequently, we performed a two-sample Mendelian randomization analysis on the 511 proteins and EC by TwoSampleMR, with p value < 0.05 as the significant causal screening condition. From the results in Table 1, a total of 24 proteins had a causal relationship with EC.

Since each of the 24 proteins contained only one SNP, sensitivity analysis could not be performed. Thus, we downloaded the pQTL files of these 24 proteins from the deCODE database for subsequent analysis. First, we screened the pQTL files according to the criteria of cis-pQTL and obtained the corresponding cis-pQTL of 19 proteins (Supplementary Data 3). Then, a two-sample Mendelian randomization analysis was performed on the 19 proteins and EC using TwoSampleMR. P- value < 0.05 was used as the significant causal screening criterion. Table 2; Fig. 2 showed that a total of 13 proteins had causal relationships with EC. Among these, ST3GAL1, C9, HDGF, PLA2R1, HSP90B1, MMP10, KLK13, CEL, CTSF, ADH4, and VIT were positively correlated with the risk of EC. In contrast, GNRH2 and HPSE were negatively correlated with the risk. Finally, the scatter plots of the effect estimates of the Mendelian randomization IVW model for the 12 proteins and EC are shown in Fig. 3A-L. The analysis of HDGF contained only one SNP and cannot show the scatter plots.

Sensitivity analysis of proteins for EC

First, we performed heterogeneity analysis on 12 proteins (ST3GAL1, C9, GNRH2, PLA2R1, HSP90B1, MMP10, KLK13, CEL, HPSE, CTSF, ADH4, VIT) and EC (Table 3). ST3GAL1, GNRH2, PLA2R1, HSP90B1, MMP10, KLK13, HPSE, CTSF, and VIT showed no heterogeneity in the MR results (Cochran Q p-value > 0.05). The MR results of proteins C9 and CEL showed moderate heterogeneity (Cochran Q p-value < 0.05, $25\% < I^2 < 50\%$). In contrast, the ADH4 showed high heterogeneity (Cochran Q p-value < 0.05, $50\% < I^2$). For proteins with heterogeneity, the IVW random-effects model (Table 4) was used to estimate causal associations. It showed that C9, CEL, and ADH4 had a significant positive correlation with EC.

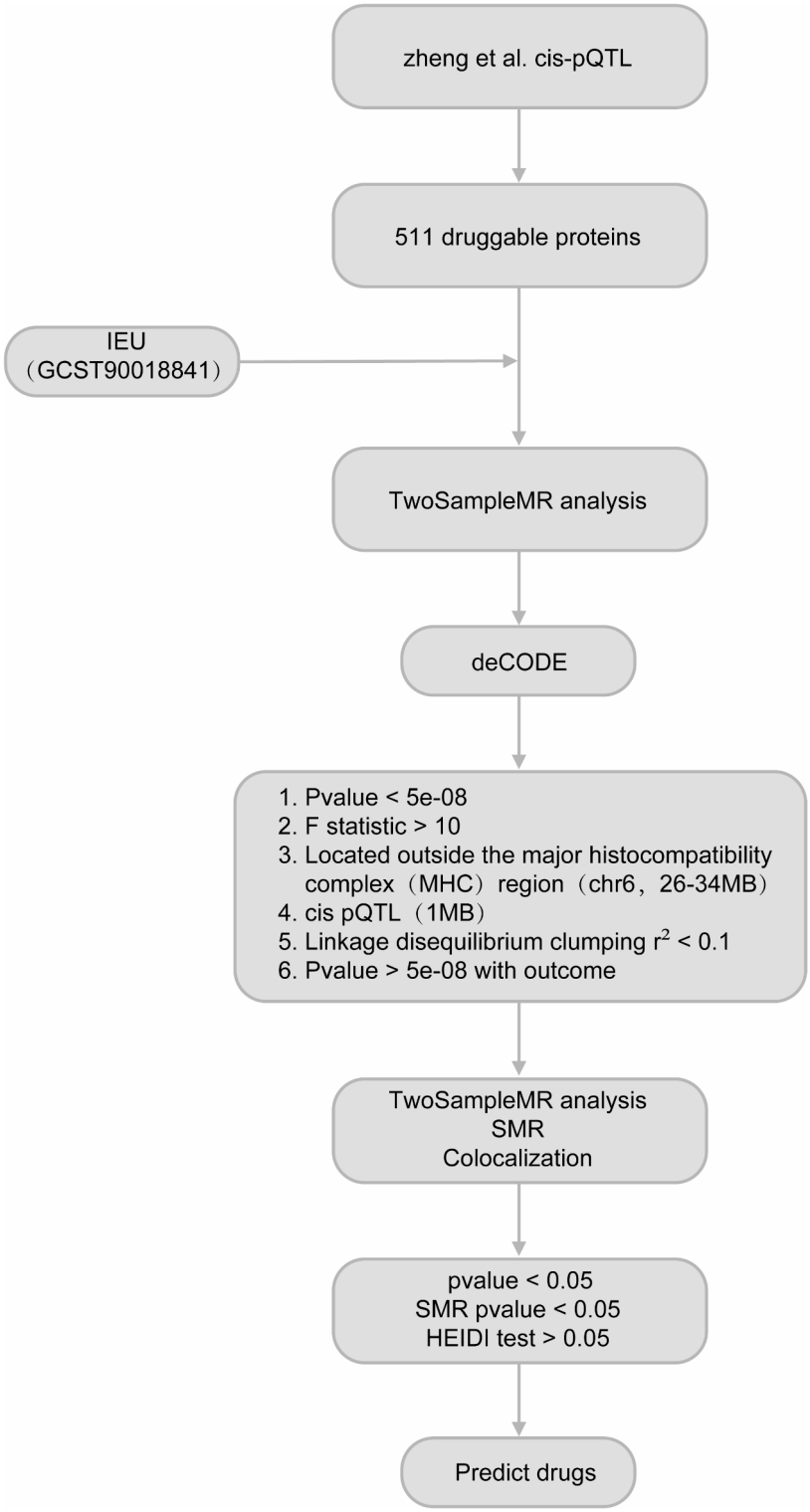

Fig. 1. Technology Roadmap. IEU: Integrative Epidemiology Unit; MR: Mendelian Randomization; SMR: Summary-data-based Mendelian Randomization; HEIDI: Heterogeneity in Dependent Instruments; pQTL: protein Quantitative Trait Locus.

As can be seen from the funnel plot of Fig. 4A-L, the instrumental variables of 11 proteins (ST3GAL1, C9, GNRH2, PLA2R1, HSP90B1, MMP10, KLK13, CEL, HPSE, CTSF, VIT) were evenly distributed on the left and right sides of the IVW line without apparent heterogeneity. This suggested that the effect estimates of these proteins were relatively stable and had good reliability. However, the instrumental variable for ADH4 was

exposure	Outcome	nSNP	beta	se	OR (95% CI)	*P* value	Method
ST3GAL1	Esophageal cancer	1	0.31739	0.14391	1.37 (1.04, 1.82)	0.02743	Wald ratio
CEL	Esophageal cancer	1	0.23694	0.08287	1.27 (1.08, 1.49)	0.00425	Wald ratio
ADAM12	Esophageal cancer	1	0.311	0.14354	1.36 (1.03, 1.81)	0.03026	Wald ratio
GNRH2	Esophageal cancer	1	−0.32043	0.15189	0.73 (0.54, 0.98)	0.03489	Wald ratio
TNFRSF6B	Esophageal cancer	1	−0.5692	0.29018	0.57 (0.32, 1.00)	0.04982	Wald ratio
C3	Esophageal cancer	1	0.52841	0.20644	1.70 (1.13, 2.54)	0.01048	Wald ratio
TGFB1	Esophageal cancer	1	0.28185	0.11583	1.33 (1.06, 1.66)	0.01496	Wald ratio
C9	Esophageal cancer	1	0.8672	0.3528	2.38 (1.19, 4.75)	0.01397	Wald ratio
CD55	Esophageal cancer	1	0.15083	0.07638	1.16 (1.00, 1.35)	0.04828	Wald ratio
ADH4	Esophageal cancer	1	−0.92229	0.22229	0.40 (0.26, 0.61)	0.00003	Wald ratio
MMP10	Esophageal cancer	1	0.38537	0.17922	1.47 (1.03, 2.09)	0.03154	Wald ratio
HSP90B1	Esophageal cancer	1	0.0565	0.0277	1.06 (1.00, 1.12)	0.0414	Wald ratio
VEGFA	Esophageal cancer	1	−0.08415	0.04257	0.92 (0.85, 1.00)	0.04809	Wald ratio
PDIA3	Esophageal cancer	1	0.61856	0.30361	1.86 (1.02, 3.37)	0.04161	Wald ratio
HDGF	Esophageal cancer	1	0.27175	0.13271	1.31 (1.01, 1.70)	0.0406	Wald ratio
PLA2R1	Esophageal cancer	1	0.08913	0.03281	1.09 (1.03, 1.17)	0.0066	Wald ratio
VIT	Esophageal cancer	1	0.23509	0.11927	1.27 (1.00, 1.60)	0.04872	Wald ratio
SIGLEC12	Esophageal cancer	1	0.05856	0.02902	1.06 (1.00, 1.12)	0.04355	Wald ratio
RTN4R	Esophageal cancer	1	0.19611	0.08862	1.22 (1.02, 1.45)	0.0269	Wald ratio
SLAMF7	Esophageal cancer	1	0.13821	0.05358	1.15 (1.03, 1.28)	0.0099	Wald ratio
CTSF	Esophageal cancer	1	0.29502	0.14347	1.34 (1.01, 1.78)	0.03975	Wald ratio
KLK13	Esophageal cancer	1	0.19665	0.07809	1.22 (1.04, 1.42)	0.0118	Wald ratio
HPSE	Esophageal cancer	1	−0.21719	0.06765	0.80 (0.70, 0.92)	0.00133	Wald ratio
TNFRSF11A	Esophageal cancer	1	−0.36197	0.15526	0.70 (0.51, 0.94)	0.01973	Wald ratio

Table 1. MR causal effect estimates of druggable proteins on EC (Zheng et al.). nSNP, number of single nucleotide polymorphism; beta, Mendelian randomization analysis effect coefficient; se, standard error; OR, odds ratio; CI, confidence interval.

exposure	Outcome	nSNP	beta	se	OR (95% CI)	*P* value	Method
ST3GAL1	Esophageal cancer	26	0.20798	0.0824	1.23 (1.05, 1.45)	0.0116	Inverse variance weighted
C9	Esophageal cancer	18	0.28535	0.11844	1.33 (1.05, 1.68)	0.01599	Inverse variance weighted
GNRH2	Esophageal cancer	8	−0.45188	0.14753	0.64 (0.48, 0.85)	0.00219	Inverse variance weighted
HDGF	Esophageal cancer	1	0.91839	0.46114	2.51 (1.01, 6.19)	0.04642	Wald ratio
PLA2R1	Esophageal cancer	107	0.04509	0.02035	1.05 (1.01, 1.09)	0.02673	Inverse variance weighted
HSP90B1	Esophageal cancer	137	0.06667	0.02473	1.07 (1.02, 1.12)	0.00703	Inverse variance weighted
MMP10	Esophageal cancer	87	0.10875	0.04354	1.11 (1.02, 1.21)	0.0125	Inverse variance weighted
KLK13	Esophageal cancer	40	0.15004	0.05292	1.16 (1.05, 1.29)	0.00458	Inverse variance weighted
CEL	Esophageal cancer	64	0.09378	0.03989	1.10 (1.02, 1.19)	0.01873	Inverse variance weighted
HPSE	Esophageal cancer	6	−0.78632	0.22178	0.46 (0.29, 0.70)	0.00039	Inverse variance weighted
CTSF	Esophageal cancer	16	0.34421	0.17227	1.41 (1.01, 1.98)	0.04571	Inverse variance weighted
ADH4	Esophageal cancer	4	2.22524	0.80026	9.26 (1.93, 44.42)	0.00542	Inverse variance weighted
VIT	Esophageal cancer	79	0.08259	0.04038	1.09 (1.00, 1.18)	0.04081	Inverse variance weighted

Table 2. MR causal effect estimates of druggable proteins on EC from DeCODE. nSNP, number of single nucleotide polymorphism; beta, Mendelian randomization analysis effect coefficient; se, standard error; OR, odds ratio; CI, confidence interval.

significantly unevenly distributed on both sides of the IVW line, showing evident heterogeneity, which suggests that other factors may influence its effect estimate and should be interpreted cautiously.

We then performed the pleiotropy test for the 12 proteins (ST3GAL1, C9, GNRH2, PLA2R1, HSP90B1, MMP10, KLK13, CEL, HPSE, CTSF, ADH4, VIT) and EC (Table 5). From Table 5, the p-value of the pleiotropy test for all proteins was greater than 0.05 ($P>0.05$), and the intercept was close to 0. This indicates that horizontal pleiotropy did not affect these proteins' causal inferences. This result further validates the stability and reliability of the causal relationships in this study.

Fig. 2. The causal effect forest plot of MR analysis of proteins on EC. SNP, Single Nucleotide Polymorphism; OR, Odds Ratio; CI, Confidence Interval.

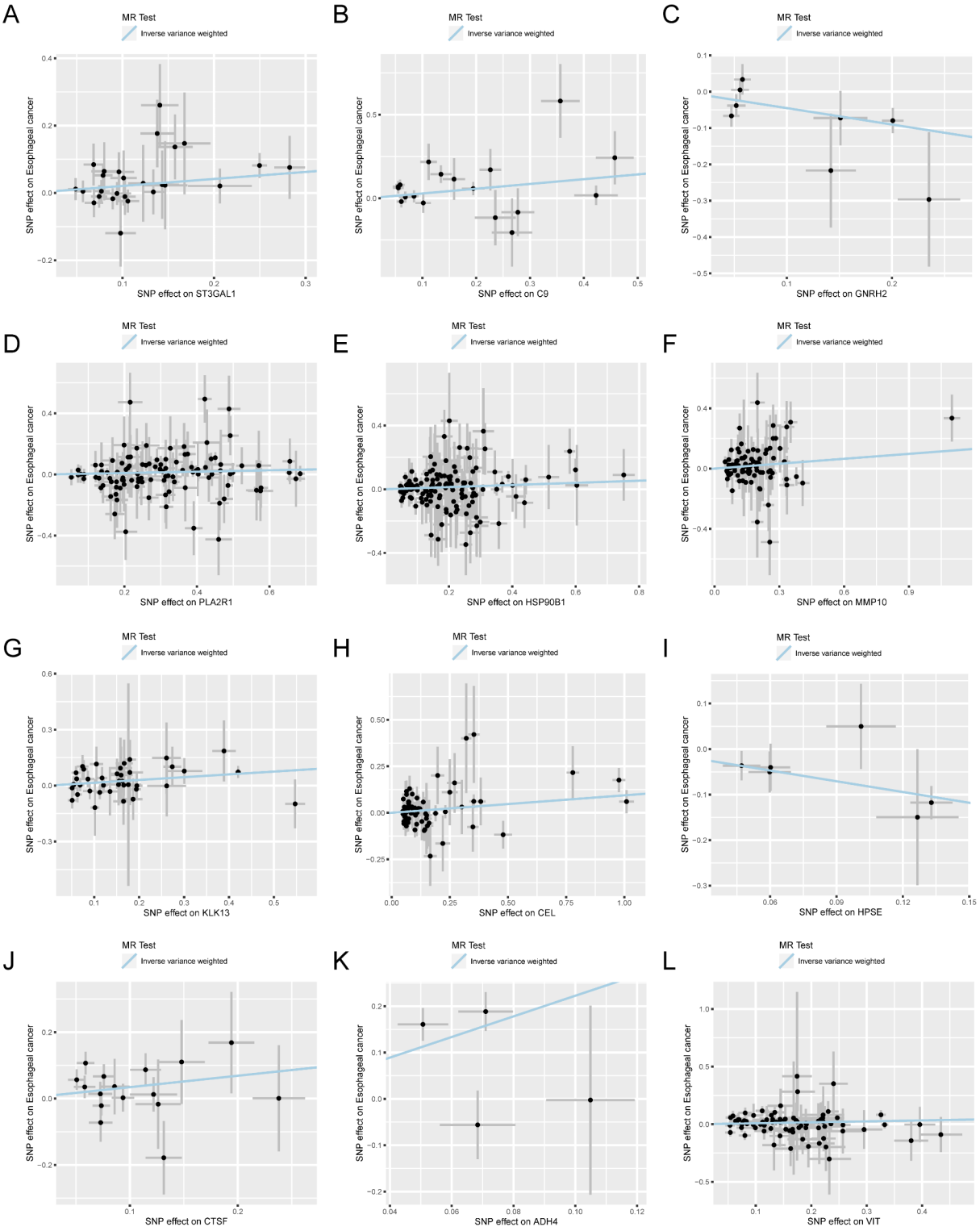

Fig. 3. Scatter plots of effect estimates from the IVW model of MR analysis for EC proteinsA-L. (**A**) ST3GAL1, (**B**) C9, (**C**) GNRH2, (**D**) PLA2R1, (**E**) HSP90B1, (**F**) MMP10, (**G**) KLK13, (**H**) CEL, (**I**) HPSE, (**J**) CTSF, (**K**) ADH4, (**L**) VIT. Scatter plot of Mendelian randomization IVW model effect estimation of (**A**–**L**) on EC. IVW, inverse variance weighted. Black dots: Represent SNP effect estimates, with position showing the size and direction. Error bars: The vertical lines above and below each black dot represent the effect estimate's standard error (SE). Blue line: The IVW regression line shows the overall effect. An upward slope suggests that higher protein levels may increase EC risk, while a downward slope indicates a decrease.

exposure	Outcome	Q	Q_df	Q_pval	I^2(%)
ST3GAL1	Esophageal cancer	19.65825269	25	0.764514547	0
C9	Esophageal cancer	31.58126231	17	0.016954247	46.17
GNRH2	Esophageal cancer	7.612199896	7	0.36802749	8.04
PLA2R1	Esophageal cancer	116.3377009	106	0.231598612	8.89
HSP90B1	Esophageal cancer	121.9865514	136	0.799537326	0
MMP10	Esophageal cancer	97.94522112	86	0.178248487	12.2
KLK13	Esophageal cancer	31.07901443	39	0.813013844	0
CEL	Esophageal cancer	92.34185956	63	0.009402932	31.78
HPSE	Esophageal cancer	2.18279018	5	0.823317898	0
CTSF	Esophageal cancer	21.85649356	15	0.111622394	31.37
ADH4	Esophageal cancer	11.67949749	3	0.00856573	74.31
VIT	Esophageal cancer	76.11144636	78	0.539401636	0

Table 3. The heterogeneity test for MR analysis of proteins on EC. Q, Cochran Q test statistic; Q_df, Q test degree of freedom; Q_pval, p-value of Q test; The I^2 statistic reflected the proportion of the heterogeneity part of the instrumental variables in the total variations: if $I^2 \leq 0$, it was set to 0, indicating no heterogeneity. $I^2 = 0–25\%$, indicating mild heterogeneity; $I^2 = 25\text{-}50\%$, indicating moderate heterogeneity; $I^2 > 50\%$ indicated high heterogeneity. The specific calculation formula was $I^2=(Q\text{-}df)/Q\times100\%$.

exposure	Outcome	nSNP	beta	se	OR (95% CI)	*P* value	Method
C9	Esophageal cancer	18	0.28535	0.11844	1.33 (1.05, 1.68)	0.01599	Inverse variance weighted (multiplicative random effects)
CEL	Esophageal cancer	64	0.09378	0.03989	1.10 (1.02, 1.19)	0.01873	Inverse variance weighted (multiplicative random effects)
ADH4	Esophageal cancer	4	2.22524	0.80026	9.26 (1.93, 44.42)	0.00542	Inverse variance weighted (multiplicative random effects)

Table 4. Causal effect estimates from the MR IVW model of druggable proteins on EC from DeCODE. nSNP, number of single nucleotide polymorphism; beta, Mendelian randomization analysis effect coefficient; se, standard error; OR, odds ratio; CI, confidence interval.

We used the Steiger directional test for analysis to ensure that the causal direction of proteins affecting the incidence of EC was correct. We found that the p-value was far less than 0.05 ($P<0.05$), indicating the correct direction (Table 6).

SMR and colocalization analysis

A HEIDI test in SMR analysis was used to test the pleiotropy further. From Table 7, the p_HEIDI of HPSE, ST3GAL1, CEL, KLK13, and GNRH2 was greater than 0.05 ($P>0.05$). This indicated that there was no pleiotropy in the SNPs of these 5 proteins, further supporting the reliability of the results. Meanwhile, the SMR analysis results for these proteins showed p_SMR all less than 0.05 ($P<0.05$). Thus, this indicated a statistically significant causal relationship between these proteins and EC. Furthermore, these findings collectively reinforce the importance of their potential impact on the development of EC.

Based on the coloc analysis (Table 8), we found a moderate colocalization relationship between HPSE and EC ($0.5 < PP.H4 < 0.8$). This indicated that genetic variations in the levels of HPSE protein and the genetic variation in EC might be shared. Moreover, this suggests that changes in HPSE protein levels influence the risk of EC through a specific mechanism. Therefore, this finding provides novel insights for further investigating the role of HPSE in the development of EC and lays the groundwork for future treatment strategies.

Drug target analysis

We used the 5 druggable targets to analyze the potential drugs in the DURGBANK database. As shown in Table 9, GNRH2, KLK13 and ST3GAL1-encoded proteins had no corresponding available drugs, which could be further explored. In addition, the protein encoded by CEL corresponded to the drug DB04348, and the protein encoded by HPSE corresponded to the drugs DB01109 and DB06779.

Discussion

Despite the emergence of new therapies in recent years, the current treatment options for EC are primarily confined to surgery. However, there are no specific targeted drug therapies proven to be highly effective for the majority of EC patients. The human proteome is currently a major therapeutic target. To our knowledge, this study is the first to combine plasma proteomics data with EC research, using two-sample MR analysis and Bayesian colocalization to assess causal proteins for EC. In our analysis, we assessed the pleiotropy of the selected proteins and evaluated heterogeneity using Cochran's Q test. Additionally, we verified the correct causal direction using the Steiger directional test. The results showed that horizontal pleiotropy did not significantly influence the causal relationship between these proteins and EC. We also used cis-pQTLs as instruments and

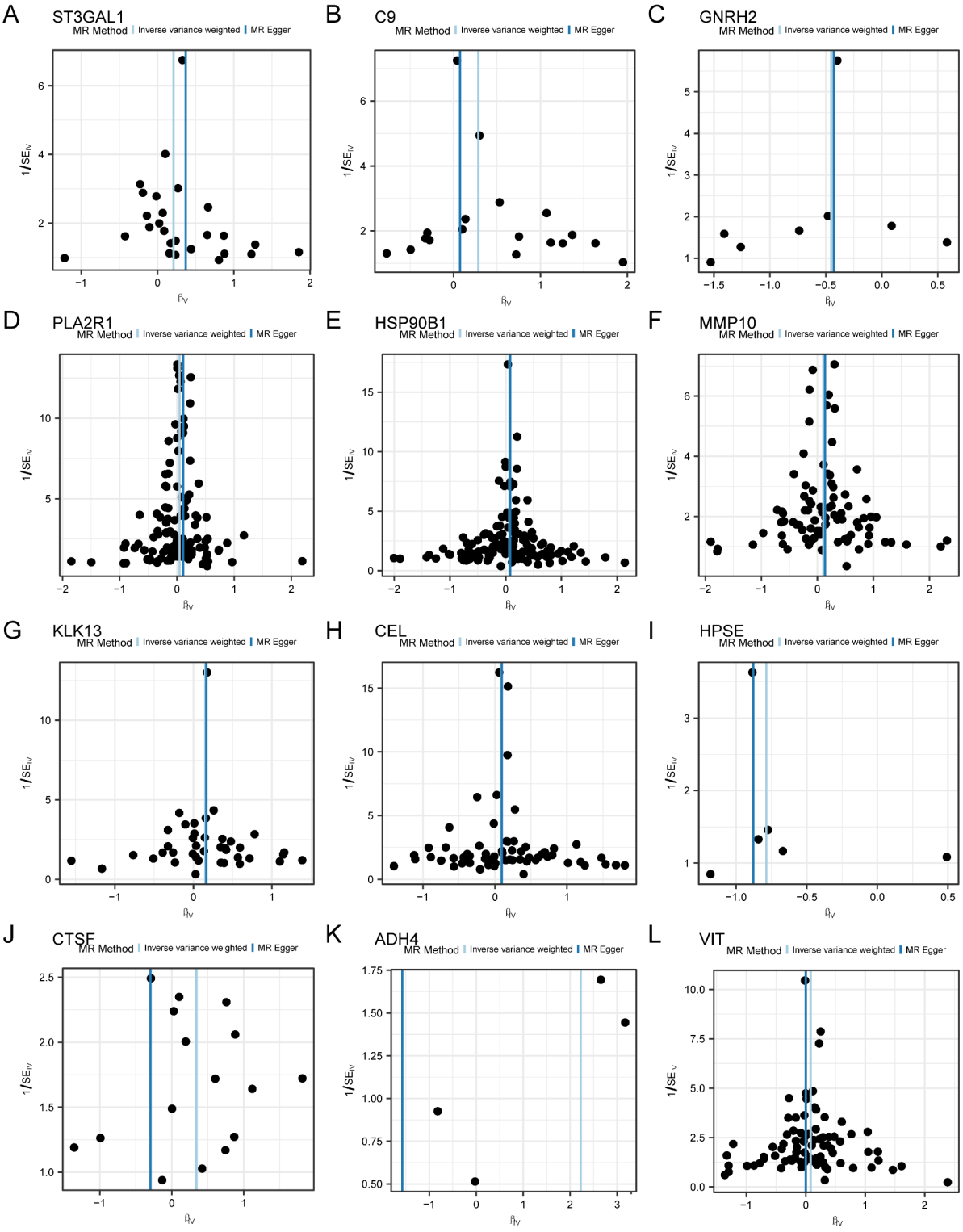

Fig. 4. Funnel plot of heterogeneity test in MR analysis of proteins on EC. A-L. (**A**) ST3GAL1, (**B**) C9, (**C**) GNRH2, (**D**) PLA2R1, (**E**) HSP90B1, (**F**) MMP10, (**G**) KLK13, (**H**) CEL, (**I**) HPSE, (**J**) CTSF, (**K**) ADH4, (**L**) VIT; Funnel plot of (**A**–**L**) and EC in Mendelian randomization analysis.

conducted sensitivity analyses, limiting bias from heterogeneity and horizontal pleiotropy. Additionally, SMR and colocalization analyses were used to exclude bias introduced by linkage disequilibrium (LD). Ultimately, we identified five proteins (HPSE, ST3GAL1, CEL, KLK13, and GNRH2) as the potential drug targets for EC and demonstrated the reliability of the results.

exposure	Outcome	egger_intercept	se	*p* value
ST3GAL1	Esophageal cancer	−0.020890252	0.020003478	0.306737663
C9	Esophageal cancer	0.03729088	0.023887437	0.138057888
GNRH2	Esophageal cancer	−0.002896005	0.028112892	0.921309108
PLA2R1	Esophageal cancer	−0.021426352	0.013095323	0.104795277
HSP90B1	Esophageal cancer	−0.004071644	0.007637045	0.59481051
MMP10	Esophageal cancer	−0.004607224	0.012119404	0.704780348
KLK13	Esophageal cancer	−0.002969012	0.013131347	0.82233454
CEL	Esophageal cancer	−0.00041571	0.009067866	0.963581667
HPSE	Esophageal cancer	0.008854063	0.045355452	0.854739674
CTSF	Esophageal cancer	0.054665419	0.042147769	0.215601578
ADH4	Esophageal cancer	0.23889267	0.28903415	0.495417362
VIT	Esophageal cancer	0.013635606	0.011576131	0.242461386

Table 5. Horizontal Pleiotropy tests of MR analysis for proteins on EC. Se, standard error.

exposure	Outcome	snp_r^2 exposure	snp_r^2 outcome	correct_causal_direction	steiger_pval
ST3GAL1	Esophageal cancer	0.069460863	5.46E−05	TRUE	0
C9	Esophageal cancer	0.035113436	8.89E−05	TRUE	5.67E−238
GNRH2	Esophageal cancer	0.018974958	3.74E−05	TRUE	1.16E−129
HDGF	Esophageal cancer	0.002255322	8.32E−06	TRUE	3.31E−16
PLA2R1	Esophageal cancer	0.978757248	0.000255608	TRUE	0
HSP90B1	Esophageal cancer	0.627485495	0.00027138	TRUE	0
MMP10	Esophageal cancer	0.281739638	0.00022058	TRUE	0
KLK13	Esophageal cancer	0.134981147	8.21E−05	TRUE	0
CEL	Esophageal cancer	0.215972601	0.000210956	TRUE	0
HPSE	Esophageal cancer	0.011077739	3.09E−05	TRUE	9.35E−75
CTSF	Esophageal cancer	0.024844294	5.81E−05	TRUE	2.15E−168
ADH4	Esophageal cancer	0.005250419	8.77E−05	TRUE	7.01E−31
VIT	Esophageal cancer	0.233781284	0.000168628	TRUE	0

Table 6. Steiger directional test for MR analysis of proteins on EC. SNP, single nucleotide polymorphism; r^2, variance explained rate.

exposure	Outcome	Top SNP	nSNP_HEIDI	p_SMR	p_HEIDI
HDGF	Esophageal cancer	rs11264535	NA	0.05182031	NA
VIT	Esophageal cancer	rs12053542	7	0.8924382	0.3228587
PLA2R1	Esophageal cancer	rs35605786	9	0.9319964	0.09590785
HPSE	Esophageal cancer	rs11732892	3	0.001741115	0.9320437
ADH4	Esophageal cancer	rs7673231	NA	8.57E-05	NA
C9	Esophageal cancer	rs62358178	NA	0.1292766	NA
ST3GAL1	Esophageal cancer	rs9643300	6	0.02781427	0.1175009
CEL	Esophageal cancer	rs592267	3	0.006659349	0.2405136
CTSF	Esophageal cancer	rs185445693	NA	0.9960081	NA
MMP10	Esophageal cancer	rs486055	8	0.3832902	0.3306923
HSP90B1	Esophageal cancer	rs1165593	6	0.4850232	0.6925374
KLK13	Esophageal cancer	rs11669237	5	0.02410394	0.8332331
GNRH2	Esophageal cancer	rs6051545	5	0.02330921	0.2287843

Table 7. The SMR analysis results of proteins on EC. SMR, Summary-data-based Mendelian Randomization; HEIDI, Heterogeneity in dependent instruments; SNP, single nucleotide polymorphism.

exposure	Outcome	PP.H0.abf	PP.H1.abf	PP.H2.abf	PP.H3.abf	PP.H4.abf
CEL	Esophageal cancer	2.1259E−263	0.453155438	6.9911E−264	0.149	0.398
GNRH2	Esophageal cancer	1.9188E−79	0.700297054	4.23121E−80	0.154	0.145
HPSE	Esophageal cancer	7.34216E−39	0.281720122	2.12213E−39	0.081	0.637
KLK13	Esophageal cancer	0	0.666989169	0	0.138	0.195
ST3GAL1	Esophageal cancer	1.0644E−205	0.644067992	3.8511E−206	0.233	0.123

Table 8. The co-localization analysis for the proteins and EC. PP.H0, Posterior Probability of Hypothesis 0; PP.H1, Posterior Probability of Hypothesis 1; PP.H2, Posterior Probability of Hypothesis 2 ; PP.H3, Posterior Probability of Hypothesis 3; PP.H4, Posterior Probability of Hypothesis 4; abf, approximate Bayes factor.

Target	Uniprot	Drugbank ID	Name	Drug group	Pharmaco- logical action	Actions
CEL	P19835	DB04348	Taurocholic acid	Approved, experimental	Unknown	NA
GNRH2	O43555	NA	NA	NA	NA	NA
HPSE	Q9Y251	DB01109	Heparin	Approved, investigational	Unknown	Substrate
		DB06779	Dalteparin	Approved	Unknown	Substrate
KLK13	Q9UKR3	NA	NA	NA	NA	NA
ST3GAL1	Q11201	NA	NA	NA	NA	NA

Table 9. Drug information of druggable targets in drugbank.

By combining large-scale plasma proteomics data with available druggable genomes, we investigated 511 potential gene targets and identified 12 proteins with strong causal relationships to EC. These proteins were ST3GAL1, C9, PLA2R1, HSP90B1, MMP10, KLK13, CEL, CTSF, ADH4, VIT, GNRH2, and HPSE. Subsequently, HPSE, ST3GAL1, CEL, KLK13, and GNRH2 were validated through colocalization analysis. Among these proteins, we found drugs corresponding to the CEL and HPSE proteins, indicating existing drugs available for these targets. However, the GNRH2, KLK13, and ST3GAL1 proteins currently have no known corresponding drugs, suggesting these may be under-researched or underutilized drug targets. Moreover, this data provides new possibilities and directions for future drug target development and design. If these drugs demonstrate safety and efficacy in future clinical trials, they could be repurposed or further optimized for EC treatment.

Among the identified five proteins, heparanase (HPSE) showed a moderate colocalization relationship with EC ($0.5 < PP.H4 < 0.8$). This indicates that genetic variation in HPSE protein levels has a strong causal relationship with the risk of EC. Additionally, changes in HPSE protein levels may be an essential marker for increased risk of EC. HPSE is a key enzyme that degrades the extracellular matrix and promotes invasion and metastasis in various cancers[20]. Moreover, HPSE degrades heparan sulfate proteoglycans (HSPGs), leading to ECM degradation, remodeling, and angiogenesis, making it easier for tumor cells to invade and metastasize[21]. Notably, it is extensively involved in biological and pathological processes such as tissue repair, inflammation, tumor angiogenesis, invasion, and metastasis[22]. HPSE can also activate several pro-cancer signaling pathways, including ERK, AKT, and Wnt/β-catenin pathways, to enhance tumor cell proliferation, migration, and drug resistance[23]. In gastric and pancreatic cancers, overexpression of HPSE is associated with poor prognosis, and it potentiates the invasion and migration of pancreatic cancer cells via epithelial-to-mesenchymal transition through the Wnt/β-catenin pathway, while its specific mechanism in esophageal cancer requires further investigation[24,25]. Our study found a significant causal relationship between HPSE and EC, with no pleiotropy detected in the SMR analysis. We also observed that HPSE was inversely correlated with the risk of EC. This finding is consistent with the study of Wang et al.[26], in which they suggested that HPSE expression was downregulated in ESCC tissues, and low levels of HPSE expression in tumor tissues were significantly associated with an increased risk of cancer-related death in ESCC patients. In addition, analysis of complete clinical and pathological information from ESCC tissues also indicates that HPSE expression in cancer tissues negatively correlates with the clinical pathological classification N stage and clinical stage[26]. However, HPSE is preferentially upregulated in the carcinogenesis of Barrett's esophagus and intestinal-type gastric cancer[27], suggesting it plays a key role in the development of inflammation-related upper gastrointestinal adenocarcinomas. Therefore, HPSE may serve as a novel prognostic marker in EC tissues and suggests the potential effectiveness of HPSE-targeted therapies in treating EC. Some experimental evidence indicated that HPSE could release HS-bound growth factors by cleaving the heparan sulfate proteoglycan (HSPG) side chains in hepatocellular carcinoma (HCC). These growth factors inhibited the proliferation and signal activation of tumor cells, such as melanoma and HCC[28,29]. Therefore, we hypothesized that low HPSE expression would enhance the ability of growth factors to bind to EC cells, leading to a poor prognosis. Our study revealed that Heparin and Dalteparin are the corresponding drugs targeting HPSE, as confirmed by the studies of Vlodavsky et al.[30]and Fairbrother et al.[31] Additionally, heparin inhibits the enzymatic activity of HPSE and blocks the production of VEGF. While heparin can inhibit its extracellular functions, it can't prevent the regulatory effects of HPSE at the gene level[32]. Thus, targeting HPSE is considered a potential treatment strategy for cancer[33]. At present, there are no studies combining Heparin or Dalteparin targeting HPSE for the treatment of EC. This may open up a new direction for future research.

In this study, we also identified GNRH2, gonadotropin-releasing hormone 2, as a potential target negatively associated with the risk of EC. GNRH2 is widely distributed in the central nervous system and expressed in peripheral tissues of mammals[34]. Studies have indicated that GnRH2 can reduce the migration and proliferation capacity of prostate cancer cells. Moreover, GnRH2 can inhibit the expression of ribosomal phosphorylation proteins in human cancer cells, which are key regulators of protein elongation[35]. GNRH2 has been proven to exhibit strong anti-proliferative and pro-apoptotic effects on various cancers, including breast, endometrial, ovarian, and prostate, through its receptors expressed on cancer cells[36]. Importantly, its role in inhibiting cancer invasiveness is consistent with our conclusion. Although there is no direct evidence linking GNRH2 with EC, its similar role in tumor biology may offer insights into its potential clinical application. The most studied genotypes of GNRH2 in tumors included SNP rs3761243 and SNP rs6051545. Notably, SNP rs3761243 is significantly associated with the survival status of osteosarcoma patients, providing crucial insights for the screening, diagnosis, and prognosis of osteosarcoma[37,38]. SNP rs6051545 of GNRH2 is expressed in both normal prostate and prostate cancer, impacting prostate biology and the treatment of prostate cancer[34]. Our study indicated that the SNP rs6051545 of GNRH2 had a negative regulatory effect on EC. Regardless of whether it may be a direct or indirect target, it may inhibit the proliferation and survival of EC cells. Currently, there are no related studies in EC, and our research highlights the potential of this field. The underlying mechanisms behind this phenomenon remain unclear and should be further elucidated in future studies.

In our study, KLK13, ST3GAL1, and CEL were risk-associated proteins for EC. The human kallikrein-related peptidase (KLK) family encodes a series of secreted serine proteases involved in many biological functions, including skin desquamation, immune response, and enamel formation[39]. KLK is also involved in tumorigenesis, activating proteolytic processes, and is associated with tumor phenotypes. It is also known as KLK-L4, which has a DNA sequence of 10 kb and consists of five exons and four introns. Some studies suggested that KLK13, either alone or in combination with other biomarkers, could serve as a prognostic factor for ovarian cancer[40], gastric cancer[41], colorectal cancer[42], oral cancer[43], and bladder cancer[44]. However, the role of KLK13 in ESCC remains to be investigated. Research by Lin et al.[45] indicated that KLK13 was significantly downregulated in ESCC tissues. Lower KLK13 mRNA levels were associated with higher tumor grade, TNM (UICC, 2009) stage, and poorer survival rates. Additionally, overexpression of KLK13 was found to inhibit cell invasion and migration. However, Nohara et al.[46] found that patients with positive KLK13 expression in ESCC samples had significantly poorer prognoses than those with negative expressions. High levels of KLK13 expression were also associated with tumor progression, advanced tumor stage, and adverse prognosis, which was consistent with the findings of Shimomura et al.[47] Although their findings complemented those of Lin et al.[45], some discrepancies were likely due to variations in tumor stages, preoperative treatments affecting KLK13 expression levels, and differences in experimental methods. Compared to the studies by Nohara[46] and Shimomura et al.[47], Lin's[45]study only detected mRNA levels in ESCC tissues and conducted cell functional experiments, without performing immunohistochemical pathological diagnosis. From this perspective, their research is insufficient. Moreover, Shimomura's[47] research found that anticancer drugs could promote KLK13 demethylation through the upregulation of TET2/3, leading to the positive expression of KLK13 after preoperative treatment. This is also a likely reason for the upregulation of KLK13 in ESCC samples. Our findings are consistent with the conclusions of Nohara[46] and Shimomura[47], indicating that the transition of KLK13 expression from negative to positive predicted poor prognosis in EC. Thus, KLK13 may be a unique novel molecule for monitoring ESCC progression, and elucidating the role of KLK13 in ESCC may provide new insights into the function of this molecule as a target for novel therapeutic drugs.

Sialyltransferase ST3GAL1 is a key enzyme involved in glycosylation modifications. It catalyzes the α-2,3-sialylation of glycoproteins and glycolipids, influencing tumor progression[48]. Consequently, glycosylation plays a critical role in tumorigenesis, and abnormal sialylation can alter cell adhesion properties, enhance immune evasion, and promote cancer cell invasion and metastasis. For example, in breast cancer and pancreatic cancer, ST3GAL1 promotes cancer cell survival and immune evasion by regulating adhesion molecules (such as MUC1 and integrins)[49] and immune checkpoint molecules (such as PD-L1)[50]. In addition, overexpression of ST3GAL1 can promote proliferation, migration, and invasion of intrahepatic cholangiocarcinoma (iCCA) cells, while inhibiting apoptosis[51]. Clinically, high expression of ST3GAL1 was associated with poor outcomes in patients with breast cancer[52] and glioblastoma[53], and it might increase tumor immune evasion through enhanced sialylation of CD55[54]. Previous studies have indicated that abnormal glycosylation in esophageal squamous cell carcinoma may impact tumor growth and invasion[55]. As a key sialyltransferase, ST3GAL1 may promote EC progression by altering the structure of tumor-associated glycoproteins. Carboxyl ester lipase (CEL) is an important lipase that is primarily involved in lipid metabolism[56]. It could hydrolyze dietary fats, cholesterol esters, and fat-soluble vitamins in the duodenum[57]. Beyond its digestive function, growing evidence indicated that this enzyme may have important extraintestinal roles. It could also cross the intestinal mucosa into the bloodstream and has been detected in the urine of healthy individuals[58]. Cui et al. demonstrated that high expression of CEL was an independent prognostic factor for poor survival rates in breast cancer[59]. Additionally, aberrant expression of CEL gene variants has been found in the pancreas of patients with chronic pancreatitis and pancreatic cancer[57,60]. Recent studies have found that CEL may be regulated by metabolism in tumorigenesis[57]. By modulating fatty acid metabolism, CEL may drive alterations in the tumor microenvironment and inflammatory responses. Furthermore, lipid metabolism imbalance is closely associated with the development of gastrointestinal cancers, including EC[61]. CEL also influences lipid metabolism and related inflammatory processes, which may indirectly contribute to the development of EC[62]. The co-localization of CEL in EC was relatively high in our study, suggesting a higher possibility of CEL acting as a pathogenic protein in EC. Furthermore, our research found that taurocholic acid could interact with the CEL gene, consistent with Fontbonne et al.'s report[63] that taurocholate could serve as a reference CEL activator. As such, chronic exposure to bile acids, such as taurocholic acid, may stimulate esophageal epithelial cells, leading to cell damage and carcinogenesis, with specific bile acid

metabolites acting as carcinogenic factors in EC[64]. CEL may regulate their metabolism, leading to adaptive cell changes in response to the bile acid environment, thereby promoting tumorigenesis[57]. The accumulation of bile acids may also lead to increased inflammation, and the role of the CEL gene in inflammation or metabolic pathways may further affect the progression of EC. Although the specific mechanism of CEL in EC requires further investigation, its association with metabolic reprogramming and inflammation makes it a potential research focus.

This study has multiple critical advantages as follows. To our knowledge, this is the first study to explore the causal relationship between potential protein targets and EC by the two-sample MR method. This approach addressed the inherent flaws of previous traditional observational studies, such as reverse causality and confounding factors. This study also provided new directions and theories for developing potential drug targets for EC. Firstly, by selecting plasma proteins and druggable gene-encoded proteins from different studies and taking their intersection, we further narrowed the scope of the study, making it more targeted. The results are more reliable as they are supported by the literature evidence. Secondly, since each of the prioritized proteins had only one SNP, we downloaded their pQTL data for supplementation, thereby broadening the application of this analysis. The SNPs selected for protein pQTL were strong instruments, meeting the F-statistic greater than 10. PQTL dataset analysis focuses on genetic variation at the protein level. Specifically, cis-pQTLs directly influence the expression levels of target proteins, providing clear insights into causal relationships. Therefore, the effects of common confounding factors in observational data do not need to be directly considered. Additionally, we have ensured the independence of genetic instruments through linkage disequilibrium (LD r^2 threshold < 0.1) to help reduce confounding bias due to correlations between the instruments. Compared to the study by Zhu et al.[65]., which used proteomics and Mendelian randomization to identify therapeutic targets for hepatocellular carcinoma, our study offers a deeper analysis of the results. In addition to performing heterogeneity and pleiotropy tests, we also used the Steiger directional test to ensure the accuracy of the causal direction. This multi-layered analytical approach effectively eliminates potential biases and provides stronger support for the reliability of the conclusions. Recent studies have focused on single GWAS or pQTL analyses[66], lacking a comprehensive and systematic approach. This study introduces a novel interdisciplinary methodology by combining two-sample MR and SMR analyses to examine the causal relationship between gene expression and EC. We further enhanced the reliability of causal inference by performing cross-validation with HEIDI and colocalization, a relatively uncommon practice in the current literature.

However, our study also has several limitations. The datasets were from European populations, which may limit the broader applicability of the causal inferences. Genetic variation frequencies differ across populations, and some pQTLs that are significant in the European population may have lower frequencies in East Asian or African populations, potentially affecting the robustness of gene-phenotype associations. Additionally, gene-environment interactions vary between ethnic groups. Effect attenuation may occur since the genetic instrumental variables in this study were selected based on European GWAS data. This means that SNP effects could be weaker in other populations, potentially impacting the generalizability of the results. Therefore, we plan to incorporate data from a more diverse population in future research to strengthen the generalizability and external validity of the results. Although we carefully selected pQTLs associated with the proteins as instrument variables, Mendelian randomization analysis may still be influenced by non-genetic effects (e.g., environmental factors), affecting the validity of causal inferences. Furthermore, despite performing heterogeneity tests, some proteins (ADH4, C9, CEL) exhibited moderate to high heterogeneity, which may be attributable to differences in the effects of SNPs, gene-gene interactions, or environmental factors. To minimize the impact of heterogeneity on our results, we applied the IVW random-effects model for proteins with higher heterogeneity and quantified heterogeneity levels using Cochran's Q statistic to ensure the robustness of causal estimates further. Meanwhile, we assessed the distribution of SNP effects using funnel plots. The results indicated that the SNP effects for most proteins were balanced, but ADH4 showed noticeable heterogeneity. We acknowledge that this result should be interpreted with caution. Since MR analysis relies on genetic data, we cannot completely exclude the possibility of unmeasured confounding factors such as environmental factors and other non-genetic effects. Therefore, future research should employ experimental biology techniques to further elucidate the specific roles of these proteins and their potential clinical applications in EC development. Nevertheless, this study provides novel insights into the molecular mechanisms underlying EC and offers valuable clues for identifying potential therapeutic targets.

Conclusions

In summary, our comprehensive analysis indicates that HPSE, ST3GAL1, CEL, KLK13, and GNRH2 proteins have a causal relationship with the risk of EC. The identified proteins may serve as attractive drug targets for EC, especially HPSE. Further research is needed to explore the roles of these candidate proteins in EC.

Data availability

Data is provided within the manuscript or supplementary information files.

Received: 29 November 2024; Accepted: 4 March 2025
Published online: 10 March 2025

References

1. Morgan, E. et al. The global landscape of esophageal squamous cell carcinoma and esophageal adenocarcinoma incidence and mortality in 2020 and projections to 2040: New estimates from GLOBOCAN 2020. *Gastroenterology* **163**(3), 649–658e2 (2022).
2. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**(6), 394–424 (2018).

3. Cavallin, F. et al. Esophageal Cancer clinical presentation: Trends in the last 3 decades in a large Italian series. *Ann. Surg.* **267**(1), 99–104 (2018).
4. Sohda, M. & Kuwano, H. Current status and future prospects for esophageal Cancer treatment. *Ann. Thorac. Cardiovasc. Surg.* **23**(1), 1–11 (2017).
5. Ell, C. et al. Curative endoscopic resection of early esophageal adenocarcinomas (Barrett's cancer). *Gastrointest. Endosc.* **65**(1), 3–10 (2007).
6. Ono, S. et al. Long-term outcomes of endoscopic submucosal dissection for superficial esophageal squamous cell neoplasms. *Gastrointest. Endosc.* **70**(5), 860–866 (2009).
7. Bates, S. E. et al. Advancing clinical trials to streamline drug development. *Clin. Cancer Res.* **21**(20), 4527–4535 (2015).
8. Nelson, M. R. et al. The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**(8), 856–860 (2015).
9. Reay, W. R. & Cairns, M. J. Advancing the use of genome-wide association studies for drug repurposing. *Nat. Rev. Genet.* **22**(10), 658–671 (2021).
10. Storm, C. S., Kia, D. A., Almramhi, M. M., Bandres-Ciga, S. & Finan, C. Finding genetically-supported drug targets for Parkinson's disease using Mendelian randomization of the druggable genome. *Nat. Commun.* **12**(1), 7342 (2021).
11. Su, W. et al. Systematic druggable genome-wide Mendelian randomisation identifies therapeutic targets for Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* **94**(11), 954–961 (2023).
12. Zheng, J. et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* **52**(10), 1122–1131 (2020).
13. Finan, C. et al. The druggable genome and support for target identification and validation in drug development. *Sci. Transl Med.* **9**(383), eaag1166 (2017).
14. Ferkingstad, E. et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* **53**(12), 1712–1721 (2021).
15. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, e34408 (2018).
16. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**(5), 481–487 (2016).
17. Wishart, D. S. et al. DrugBank 5.0: a major update to the drugbank database for 2018. *Nucleic Acids Res.* **46**(D1), D1074–D1082 (2018).
18. Zhang, Y. et al. Causal effects of gut microbiota on erectile dysfunction: A two-sample Mendelian randomization study. *Front. Microbiol.* **14**, 1257114 (2023).
19. Ning, P., Mu, X., Guo, X. & Li, R. Hearing loss is not associated with risk of Parkinson's disease: A Mendelian randomization study. *Heliyon* **10**(11), e32533 (2024).
20. Hassan, N., Greve, B., Espinoza-Sánchez, N. A. & Götte, M. Cell-surface Heparan sulfate proteoglycans as multifunctional integrators of signaling in cancer. *Cell. Signal.* **77** 109822 (2021).
21. Zahavi, T. et al. Heparanase: A potential marker of worse prognosis in Estrogen receptor-positive breast cancer. *NPJ Breast cancer* **7**(1), 67 (2021).
22. Jayatilleke, K. M. & Hulett, M. D. Heparanase and the hallmarks of cancer. *J. Transl Med.* **18**(1), 453 (2020).
23. Koganti, R., Suryawanshi, R. & Shukla, D. Heparanase, cell signaling, and viral infections. *Cell. Mol. Life Sci.* **77**(24), 5059–5077 (2020).
24. Zheng, L. et al. miRNA-558 promotes gastric cancer progression through attenuating Smad4-mediated repression of heparanase expression. *Cell. Death Dis.* **7**(9), e2382 (2016).
25. Wang, C. et al. Heparanase potentiates the invasion and migration of pancreatic cancer cells via epithelial–to–mesenchymal transition through the Wnt/β–catenin pathway. *Oncol. Rep.* **44**(2), 711–721 (2020).
26. Wang, Y. et al. Heparanase is a prognostic biomarker independent of tumor purity and hypoxia based on bioinformatics and immunohistochemistry analysis of esophageal squamous cell carcinoma. *World J. Surg. Oncol.* **20**(1), 236 (2022).
27. Sonoda, R. et al. Preferential up-regulation of heparanase and cyclooxygenase-2 in carcinogenesis of Barrett's oesophagus and intestinal-type gastric carcinoma. *Histopathology* **57**(1), 90–100 (2010).
28. Ogasawara, S., Yano, H., Iemura, A., Hisaka, T. & Kojiro, M. Expressions of basic fibroblast growth factor and its receptors and their relationship to proliferation of human hepatocellular carcinoma cell lines. *Hepatology* **24**(1), 198–205 (1996).
29. Pang, R. & Poon, R. T. Angiogenesis and antiangiogenic therapy in hepatocellular carcinoma. *Cancer Lett.* **242**(2), 151–167 (2006).
30. Vlodavsky, I. et al. The impact of heparanese and heparin on cancer metastasis and angiogenesis. *Pathophysiol Haemost Thromb.* **35**(1–2), 116–127 (2006).
31. Fairbrother, W. J., Champe, M. A., Christinger, H. W., Keyt, B. A. & Starovasnik, M. A. Solution structure of the heparin-binding domain of vascular endothelial growth factor. *Structure* **6**(5), 637–648 (1998).
32. Yang, Y. et al. Nuclear heparanase-1 activity suppresses melanoma progression via its DNA-binding affinity. *Oncogene* **34**(47), 5832–5842 (2015).
33. Yang, Y. et al. Potential roles of heparanase in cancer therapy: Current trends and future direction. *J. Cell. Physiol.* **238**(5), 896–917 (2023).
34. Sissung, T. M. et al. GNRH2 polymorphism in men with prostate Cancer treated with androgen deprivation therapy. *Anticancer Res.* **43**(9), 4023–4030 (2023).
35. Darby, S. et al. Expression of GnRH type II is regulated by the androgen receptor in prostate cancer. *Endocr. Relat. Cancer* **14**(3), 613–624 (2007).
36. Limonta, P. & Manea, M. Gonadotropin-releasing hormone receptors as molecular therapeutic targets in prostate cancer: Current options and emerging strategies. *Cancer Treat. Rev.* **39**(6), 647–663 (2013).
37. Feng, W. et al. The effects of common variants in MDM2 and GNRH2 genes on the risk and survival of osteosarcoma in Han populations from Northwest China. *Sci. Rep.* **10**(1), 15939 (2020).
38. Kushlinskii, N. E. et al. Associations of single nucleotide polymorphisms with malignant and borderline bone tumors. *Klin. Lab. Diagn.* **10**, 58–60 (2013). 22 – 4.
39. Stefanini, A. C., da Cunha, B. R., Henrique, T. & Tajara, E. H. Involvement of Kallikrein-Related Peptidases in Normal and Pathologic Processes. *Dis Markers*. 946572 (2015). (2015).
40. White, N. M. et al. J. KLK6 and KLK13 predict tumor recurrence in epithelial ovarian carcinoma. *Br. J. Cancer*. **101**(7), 1107–1113 (2009).
41. Konstantoudakis, G., Florou, D., Mavridis, K., Papadopoulos, I. N. & Scorilas, A. Kallikrein-related peptidase 13 (KLK13) gene expressional status contributes significantly in the prognosis of primary gastric carcinomas. *Clin. Biochem.* **43**(15), 1205–1211 (2010).
42. Talieri, M. et al. The use of kallikrein-related peptidases as adjuvant prognostic markers in colorectal cancer. *Br. J. Cancer* **100**(10), 1659–1665 (2009).
43. Ishige, S. et al. Decreased expression of kallikrein-related peptidase 13: Possible contribution to metastasis of human oral cancer. *Mol. Carcinog.* **53**(7), 557–565 (2014).
44. Tokas, T. et al. Downregulated KLK13 expression in bladder cancer highlights tumor aggressiveness and unfavorable patients' prognosis. *J. Cancer Res. Clin. Oncol.* **143**(3), 521–532 (2017).
45. Lin, Q. et al. Downregulation of KLK13 promotes the invasiveness and metastasis of oesophageal squamous cell carcinoma. *Biomed. Pharmacother* **96**, 1008–1015 (2017).

46. Nohara, K. et al. Expression of kallikrein-related peptidase 13 is associated with poor prognosis in esophageal squamous cell carcinoma. *Gen. Thorac. Cardiovasc. Surg.* **66**(6), 351–357 (2018).
47. Shimomura, A. et al. Induction of Kallikrein-Related peptidase 13 and TET2/3 by anticancer drugs and poor prognosis of patients with esophageal squamous cell carcinoma after preoperative treatment. *Ann. Surg. Oncol.* **31**(1), 251–261 (2024).
48. Zhang, N., Lin, S., Cui, W. & Newman, P. J. Overlapping and unique substrate specificities of ST3GAL1 and 2 during hematopoietic and megakaryocytic differentiation. *Blood Adv.* **6**(13), 3945–3955 (2022).
49. Burchell, J. et al. An alpha2,3 sialyltransferase (ST3Gal I) is elevated in primary breast carcinomas. *Glycobiology* **9**(12), 1307–1311 (1999).
50. Rodriguez, E. et al. Sialic acids in pancreatic cancer cells drive tumour-associated macrophage differentiation via the Siglec receptors Siglec-7 and Siglec-9. *Nat. Commun.* **12**(1), 1270 (2021).
51. Chen, F. et al. ST3GAL1 promotes malignant phenotypes in intrahepatic cholangiocarcinoma. *Mol. Cell. Proteom.* 100821 (2024).
52. Fan, T. C. et al. Reciprocal feedback regulation of ST3GAL1 and GFRA1 signaling in breast cancer cells. *Cancer Lett.* **434**, 184–195 (2018).
53. Chong, Y. K. et al. ST3GAL1-Associated transcriptomic program in glioblastoma tumor growth, invasion, and prognosis. *J. Natl. Cancer Inst.* **108**(2), djv326 (2015).
54. Lin, W. D. et al. Sialylation of CD55 by ST3GAL1 facilitates immune evasion in Cancer. *Cancer Immunol. Res.* **9**(1), 113–122 (2021).
55. Liu, S. et al. Identification of serum N-glycans signatures in three major Gastrointestinal cancers by high-throughput N-glycome profiling. *Clin. Proteom.* **21**(1), 64 (2024).
56. Torsvik, J. et al. Endocytosis of secreted carboxyl ester lipase in a syndrome of diabetes and pancreatic exocrine dysfunction. *J. Biol. Chem.* **289**(42), 29097–29111 (2014).
57. Johansson, B. B. et al. The role of the carboxyl ester lipase (CEL) gene in pancreatic disease. *Pancreatology* **18**(1), 12–19 (2018).
58. Comte, B. et al. Detection of bile salt-dependent lipase, a 110 kda pancreatic protein, in urines of healthy subjects. *Kidney Int.* **69**(6), 1048–1055 (2006).
59. Cui, Y., Jiao, Y., Wang, K., He, M. & Yang, Z. A new prognostic factor of breast cancer: High carboxyl ester lipase expression related to poor survival. *Cancer Genet.* **239**, 54–61 (2019).
60. Lombardo, D. et al. Pancreatic adenocarcinoma, chronic pancreatitis, and MODY-8 diabetes: Is bile salt-dependent lipase (or carboxyl ester lipase) at the crossroads of pancreatic pathologies? *Oncotarget* **9**(15), 12513–12533 (2017).
61. Shen, G. Y. et al. Identification of a prognostic gene signature based on lipid metabolism-related genes in esophageal squamous cell carcinoma. *Pharmgenomics Pers. Med.* **16**, 959–972 (2023).
62. Qiu, Y. et al. Carboxyl ester lipase is highly conserved in utilizing maternal supplied lipids during early development of zebrafish and human. *Biochim. Biophys. Acta Mol. Cell. Biol. Lipids.* **1865**(6), 158663 (2020).
63. Fontbonne, H., Puigserver, A., Bouza, B., Lombardo, D. & Ajandouz, H. Activation of bile salt dependent lipase by (lyso) phosphatidic acid and platelet activating factor. *FEBS Lett.* **587**(18), 3002–3007 (2013).
64. Kumar, A. et al. Shifts in serum bile acid profiles associated with barrett's esophagus and stages of progression to esophageal adenocarcinoma. *Clin. Transl Gastroenterol.* **15**(10), e1 (2024).
65. Zhu, W. et al. Therapeutic targets for hepatocellular carcinoma identified using proteomics and Mendelian randomization. *J. Gastroenterol. Hepatol.* https://doi.org/10.1111/jgh.16785 (2024). Advance online publication.
66. Yao, T. et al. Identifying therapeutic targets for breast cancer: Insights from systematic Mendelian randomization analysis. *Front. Oncol.* **14**, 1407795 (2024).

Acknowledgements

We thank all contributors to the IEU open GWAS database and acknowledge all the investigators and participants of this study.

Author contributions

Conceptualization: S.H. and C.L.; Methodology: C.L.; Formal analysis: C.L. and X.C.; Validation: Y.Y.; Investigation: M.R.; Resources: S.H.; Data curation: C.L. and X.C.; writing—original draft, review and editing: C.L.; Supervision: S.H.; Funding: S.H.. All authors have read and agreed to the published version of the manuscript.

Funding

The research is supported by the Key Research and Development Program of Shaanxi (No. 2021ZDLSF02-06).

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

The data and analysis of this study are publicly available and do not require ethical approval.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-93068-4.

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