Discover Oncology

Research

Blood metabolites, protein regulatory networks and their roles in pan-cancer: a mendelian randomisation study

Shenglong Xia[1,2] · Zhengyang Xu[2] · Cheng Cheng[1,3] · Rui An[1,3] · Wenci Chen[4] · Daopo Lin[5] · Yuzhen Gao[1,3] · Liangjing Wang[2] · Xinyou Xie[1,3] · Jun Zhang[1,3]

Received: 19 January 2025 / Accepted: 28 April 2025
Published online: 10 May 2025

Abstract

Background Metabolic dysregulation was closely associated with cancers. However, there is a lack of studies to explore the relationship between blood metabolites, related proteins, and different types of cancer.

Methods Two-sample Mendelian randomization (MR) analysis was used to assess the causal effects of genetically determined metabolites and metabolite ratios on solid cancers. we analyzed 1400 metabolites/metabolite ratios as exposures and 16 cancers from UK Biobank/FinnGen as outcomes. Protein-metabolite interactions were mapped via MR and visualized with Cytoscape, followed by Gene Ontology enrichment. Clinical validation included metabolomic profiling of 75 breast cancer patients and 20 controls.

Results MR analysis identified 11 metabolites or metabolite ratios causally associated with cancer risk. Moreover, 48 proteins were demonstrated to be involved in the regulation of these metabolites, which are predominantly enriched in 5 significant metabolic pathways in cancers. Clinically, elevated lignoceroylcarnitine (C24) reduced breast cancer risk, while high glucose-to-mannose and alanine-to-asparagine ratios increased risk.

Conclusions Our study revealed a causal effects of metabolites and its related proteins/pathways on various types of cancers.

Key messages

What is already known on this topic?

Metabolic dysregulation in tumors has been reported in numerous studies and is closely linked to cancer initiation and prognosis. However, significant heterogeneity exists across different tumors and even among patients with the same

Shenglong Xia and Zhengyang Xu contributed equally to this work.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-02522-2.

✉ Liangjing Wang, wangljzju@zju.edu.cn; ✉ Xinyou Xie, scottxie@zju.edu.cn; ✉ Jun Zhang, jameszhang2000@zju.edu.cn; Shenglong Xia, xsl89@zju.edu.cn; Zhengyang Xu, 22218372@zju.edu.cn; Cheng Cheng, 22218693@zju.edu.cn; Rui An, 12118552@zju.edu.cn; Wenci Chen, chenwenci1210@163.com; Daopo Lin, wmuldp@163.com; Yuzhen Gao, yuzhengao@zju.edu.cn | [1]Department of Clinical Laboratory, Sir Run Run Shaw Hospital of Zhejiang University School of Medicine, Hangzhou 310016, Zhejiang, China. [2]Department of Gastroenterology, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou 310009, Zhejiang, China. [3]Key Laboratory of Precision Medicine in Diagnosis and Monitoring Research of Zhejiang Province, Hangzhou 310013, Zhejiang, China. [4]Department of Rehabilitation, Wenzhou Hospital of Integrated Traditional Chinese and Western Medicine, Wenzhou 325000, China. [5]Department of Gastroenterology, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou 325000, China.

tumor type, making it challenging to identify the true drivers of cancer. Additionally, the molecular networks regulating these metabolites remain poorly understood amid the complex landscape of metabolic differences.

What this study adds?

This study, using two-sample Mendelian randomization (MR) analysis, identified eight metabolites and three metabolite ratios causally associated with cancer risk. These findings highlight shared factors in cancer development, providing critical molecules for cancer prevention, prognostic prediction, and therapeutic target screening. Moreover, 48 regulatory proteins were identified, predominantly enriched in five key metabolic pathways related to cancer: cell kinase apoptotic function, ERK1 migration and ERK2 signaling cascade, astrocyte amyloid fibril formation, cellular aldehyde catabolic processes, and mRNA export from the nucleus. These results elucidate common metabolic regulatory mechanisms across cancers, offering significant insights into the design of metabolic-targeted cancer therapies.

How this study might affect research, practice, or policy?

Our findings hold significant implications for cancer prognosis prediction. By measuring blood metabolite and regulatory protein levels in cancer patients and analyzing multiple metabolite/protein combinations, tumor prognosis can be more accurately assessed. Additionally, our study highlights both the commonalities and differences in blood metabolite profiles among cancer patients, providing valuable insights into the mechanisms of tumor initiation and progression. Finally, by unveiling, for the first time, the connections between metabolites and regulatory proteins in pan-cancer, this research lays the groundwork for subsequent mechanistic studies and therapeutic target development.

Keywords Mendelian randomization · Metabolites · Protein · Metabolic pathways · Pan-cancer

Abbreviations

AKT	Protein Kinase B
ALDHs	Aldehyde Dehydrogeases
ATP	Adenosine Triphosphate
BMI	Body Mass Index
CI	Confidence Interval
ctDNA	Circulating Tumor DNA
EDTA	Ethylenediaminetetraacetic Acid
ERK	Extracellular Signal-Regulated Kinase
FADH2	Flavin Adenine Dinucleotide
FAO	Fatty Acid Oxidation
FDR	False Discovery Rate
FinnGen	FinnGen Consortium
GO	Gene Ontology
GWAS	Genome-Wide Association Study
HER2	Human Epidermal Growth Factor Receptor 2
HPLC	High-Performance Liquid Chromatography
IVs	Instrumental Variables
IVW	Inverse-Variance Weighted
JNK	C-Jun N-terminal Kinase
LC–MS/MS	Liquid Chromatography-Tandem Mass Spectrometry
MAPKs	Mitogen-Activated Protein Kinases
MR	Mendelian Randomization
MR-PRESSO	Mendelian Randomization Pleiotropy RESidual Sum and Outlier
NADH	Nicotinamide Adenine Dinucleotide
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
OR	Odds Ratio
PCs	Principal Components
PI3 K	Phosphoinositide 3-Kinase
ROC	Receiver Operating Characteristic
SNPs	Single Nucleotide Polymorphisms
UKB	UK Biobank
UPLC	Ultra Performance Liquid Chromatography

1 Introduction

Cancer is a major global health challenge, responsible for nearly 10 million deaths annually [1]. Despite advances in cancer treatment, there has been slower progress in tumorigenesis, contributing to the high mortality rates across various cancer types. Biomarkers such as circulating tumor DNA (ctDNA), exosomes, and other emerging markers have been identified for cancer detection and treatment, but much of these findings remains at the experimental stage, far from clinical application [2]. Thus, identifying new biomarkers suitable for various types of cancer is crucial to deepen the understand of cancer development and advance therapeutic strategies.

In recent years, growing evidence has demonstrated a strong link between cancer and metabolic dysregulation, with conditions such as diabetes, non-alcoholic fatty liver disease, and obesity [3–5]. Metabolites, representing the substrates and end-products of both the host's and microbiome's metabolic processes, provide a direct reflection of the organism's metabolic state. It has been shown that disturbances in glucose and amino acid metabolism are particularly pronounced in solid tumors [6]. Metabolomics, the comprehensive analysis of metabolites, enables the elucidation of biological phenotypes, disease progression, and adaptation to environmental factors. As a result, metabolomics offers a valuable tool for the identification and characterization of biomarkers [7]. Several studies have highlighted the diagnostic potential of metabolites, particularly those found in urine, plasma, and serum, in cancer detection [8, 9]. However, the causal relationships between these metabolic changes and their corresponding protein expressions remain insufficiently understood.

The recent development of high-throughput technologies has enabled large-scale Genome-Wide Association Studies (GWAS), which have identified numerous single nucleotide polymorphisms (SNPs) associated with metabolite levels. These advances have laid the groundwork for genetically determined metabolite databases [9]. Mendelian Randomization (MR) is a genetic epidemiology method that uses SNPs as instrumental variables (IVs) to assess causal relationships between exposures-such as metabolites or related proteins-and disease outcomes. By leveraging genetic variants, MR minimizes the potential biases of confounding and reverse causation present in traditional observational studies. In this study, we employ a two-sample MR design to investigate the causal impact of metabolites and their related proteins on various types of cancers. This approach aims to provide insight into the metabolic pathways that influence cancer risk and identify potential therapeutic targets.

2 Method

2.1 Study design

This study systematically investigates the causal relationship between human blood metabolites and cancer risk using a two-sample MR design. To ensure the robustness of MR findings, we adhered to three core assumptions: (1) genetic instruments must be directly associated with the exposure of interest (i.e., blood metabolites); (2) these instruments must be independent of the outcome (i.e., cancer types) and any known or unknown confounders; (3) the influence of instrumental variables (IVs) on the outcome should be mediated solely through the exposure of interest. We sourced genetic information for metabolites and cancer outcomes from distinct Genome-Wide Association Study (GWAS) datasets to eliminate sample overlap. The methods employed in this study adhere with the STROBE-MR checklist [10]. An overview of this MR study is presented in Fig. 1.

2.2 GWAS data for blood metabolites and cancers

We obtained GWAS summary datasets for 1091 blood metabolites and 309 metabolite ratios from Chen et al. [9], representing the most comprehensive metabolite-related GWAS data available to date. These summary statistics were deposited in the GWAS catalog (https://www.ebi.ac.uk/gwas/). Among the 1091 metabolites evaluated, 850 were identified across eight superpathways (lipid, amino acid, xenobiotics, nucleotide, cofactor and vitamins, carbohydrate, peptide, and energy), while the remaining 241 were classified as unknown or partially characterized.

To minimize biases related to population stratification, our analysis focused exclusively on the European population. Summary GWAS data for 16 types of cancers were acquired from UK Biobank (UKB) and FinnGen consortium R8 release data, including esophagus, stomach, small intestine, pancreas, bronchus, lung, melanoma, uterine cervix, bladder, brain, colorectal, bone, breast, prostate, kidney, ovary, and thyroid (Table 1). The UK Biobank is a large-scale biomedical

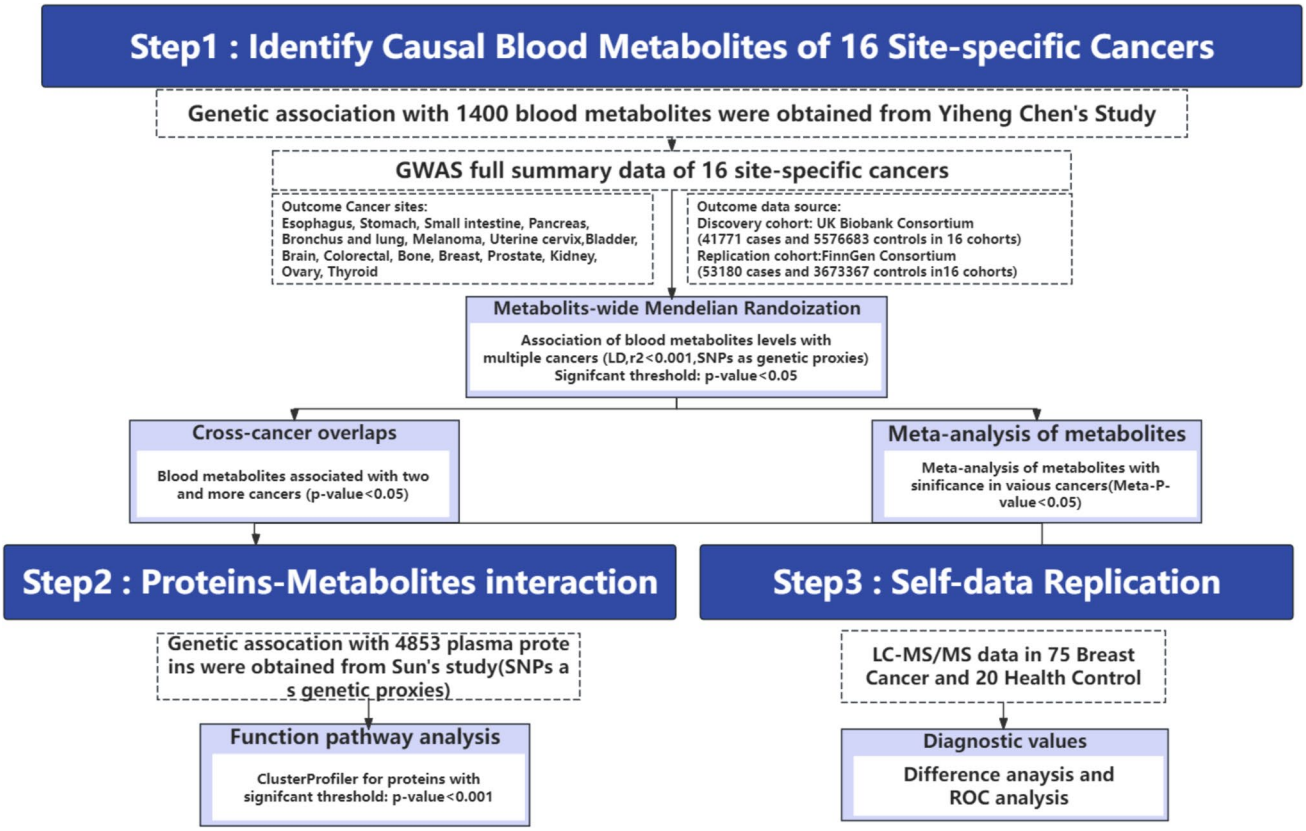

Fig. 1 Design of this Mendelian randomization study

database, providing comprehensive genetic and health information from over 500,000 participants aged 40–69, with extensive lifestyle, environmental, and medical history data [11]. The FinnGen project leverages the unique genetic characteristics of the Finnish population, along with its extensive national health registry, offering advantages in genetic research due to historical isolation and the founder effect, which allows for the identification of rare, high-impact genetic mutations [12]. Discovery analysis utilized UK Biobank data, while validation was performed in FinnGen R8. Phenotype definitions followed ICD-10 codes, and covariates included age, sex, and genetic principal components. The download links of all cohorts are provided in Tables S1–S2.

We sourced exposure and outcome GWAS datasets from independent cohorts. The metabolite GWAS [9] explicitly excluded participants from both UK Biobank and FinnGen. Cancer GWAS data were obtained separately from UK Biobank (UK population) and FinnGen (Finnish population), with no obvious geographical or genetic overlap between these cohorts. The GWAS summary statistics for metabolites and cancers were adjusted for age, sex, genetic principal components (PCs), and batch effects in the original studies. To minimize confounding by lifestyle factors (e.g., smoking, diet), we excluded SNPs associated with these traits ($P < 2.23 \times 10^{-6}$) using the PhenoScanner database.

2.3 Instrumental variables screening and MR analysis

The selection of IVs must satisfy the three assumptions [13]: (1) To account for multiple testing, we applied Bonferroni correction. The significance threshold was defined as $P_{threshold} = 0.05/N$, where N represents the total number of comparisons (1,400 metabolites × 16 cancer types = 22,400). Thus, the adjusted threshold was $P < 2.23 \times 10^{-6}$. SNPs associated with each metabolite at genome-wide significance ($P < 2.23 \times 10^{-6}$) were screened as IVs; (2) the SumStatsRehab software tool [14] was employed to supplement missing variant IDs for IVs, (3) Linkage disequilibrium analysis was performed using a clumping procedure with an r^2 threshold of < 0.001 and a 10,000 kb window, referencing the European 1000 Genomes Project Phase 3 panel [15]. Subsequently, we assessed the strength of all IVs by calculating the explained variance (R^2) and F-statistical parameters. IVs with F > 10 were considered strong instruments and were included in subsequent MR analyses. The "TwoSampleMR" package was used to perform MR analysis. For metabolites with a single instrument, the

Table 1 Characteristics of cancer GWAS datasets and associated Metabolites

Phenotypes	Discovery cohort: UK Biobank			Validation cohort: finngen_R8		
	Cases (N)	Controls (N)	Metabolites ($P<0.05$)	Cases (N)	Controls (N)	Metabolites ($P<0.05$)
Esophagus	975	419,556	61	503	259,583	10
Stomach	764	419,767	60	1227	259,583	17
Small intestine	244	420,287	61	455	259,583	10
Pancreas	933	419,598	52	1249	259,583	13
Bronchus and lung	3048	417,483	71	5118	259,583	16
Malignant melanoma	2598	417,933	81	2705	259,583	11
Uterine cervix	273	420,258	51	2913	149,394	18
Bladder	2576	417,955	56	2380	259,583	14
Brain	666	419,865	53	774	259,583	20
Colorectal	5693	386,740	79	5458	259,583	13
Bone	172	419,977	58	189	259,583	13
Breast	13,257	205,913	65	14,000	149,394	19
Prostate	7691	169,762	78	11,590	110,189	16
Kidney	1307	415,431	58	1830	259,583	14
Ovary	1147	207,089	64	1264	149,394	14
Thyroid	427	418,625	79	1525	259,583	6
Total numbers	41,771	5,576,683	255	53,180	3,673,367	187

Wald ratio method was utilized to estimate the log odds change in cancer risk for each standard deviation (SD) increase in circulating metabolite levels. For metabolites with multiple instruments, the inverse-variance weighted (IVW) method was applied to obtain MR effect estimates. Heterogeneity testing was conducted using the Q statistic to evaluate variability among genetic instruments. Finally, the estimates from the 16-site pan-cancer analysis in FinnGen and UKBB were combined using a random-effects meta-analysis approach.

2.4 Protein-metabolite-wide MR analysis

Genetic associations for 4,853 plasma proteins were sourced from Sun's study [16], with SNPs defined as genetic proxies. A protein-metabolite-wide MR analysis was performed to explore interactions between proteins and metabolites.Genetic instruments for proteins were selected at $P < 2.23 \times 10^{-6}$, with clumping parameters of $r^2 < 0.001$ and a 10,000 kb window. We applied inverse-variance weighted (IVW) as the primary method and conducted colocalization analysis to confirm shared causal variants. The outcomes of this analysis were visualized using Cytoscape v3.9.1 software [17]. Gene Ontology (GO) functional analysis [18] was subsequently performed based on the proteins associated with the metabolites, using Metascape (http://metascape.org) with a Benjamini–Hochberg adjusted $P < 0.05$.

2.5 Sample collection and UPLC-MS/MS analysis

To validate the results from the MR analysis, metabolomic profiling was conducted on clinical samples from 75 patients with breast cancer and 20 healthy controls. The samples were collected from Sir Run-Run Shaw Hospital between 2023.03 to 2024.09. Cases and controls were matched in 1:2 by age. Plasma samples were drawn in the early morning using EDTA-treated vacutainer tubes to minimize potential metabolic alterations. Within 30 min of collection, the samples were processed by centrifugation at 3000 rpm for 10 min to separate plasma from cellular components. The resulting plasma was carefully aliquoted into 1.5 mL cryogenic tubes and immediately stored at − 80 °C for future analysis. Logistic regression models were used to assess metabolite differences, adjusted for BMI, menopausal status (for breast cancer), and chemotherapy history.

Sample preparation was automated using the MicroLab STAR® system (Hamilton Company). Proteins were precipitated by adding methanol, followed by agitation for 2 min using a GenoGrinder 2000 (Glen Mills). Metabolomic profiling was performed by Metabolon, Inc., using a Waters ACQUITY UPLC coupled with a Thermo Scientific Q Exactive high-resolution mass spectrometer. The system employed a heated electrospray ionization (HESI–II) source and Orbitrap mass analyzer.

Two distinct gradient elution methods were used: one for hydrophilic compounds under acidic, positive ionization conditions (water and methanol with 0.05% PFPeA and 0.1% FA) and one for basic compounds under negative ionization (methanol, water, and 6.5 mM ammonium bicarbonate at pH 8). Hydrophobic compounds were analyzed using a C18 column under similar gradient conditions. Additional separation was achieved using hydrophilic interaction liquid chromatography (HILIC) for negative ionization with 10 mM ammonium formate at pH 10.8. Mass spectrometry data were collected in MS and data-dependent MS^n modes, with dynamic exclusion, within a 70–1000 m/z scan range. Raw data were processed for peak extraction and identification using Metabolon's proprietary library of purified standards. Data normalization and scaling were performed in MetaboAnalyst 5.0, applying "Normalization by sum" and "auto scaling" for statistical analysis and interpretation.

3 Results

3.1 Causal effects of metabolites and metabolite ratios on cancer

Using the IVW method, we assessed the causal effects of 1400 metabolites and metabolite ratios on various types of cancer. In the discovery cohort (UK Biobank), we identified 255 potential causal associations ($P < 0.05$), with the number of metabolites associated with individual type of cancer ranging from 52 to 81 (Table 1). In the replication cohort (FinnGen_R8), a total of 187 metabolites were associated with cancers, with the number of associated metabolites per cancer type ranging from 6 to 20. We also examined the causal relationships between individual metabolites and different cancer types. As shown in Fig. 2, the UK Biobank cohort revealed 21 metabolites and 7 metabolite ratios with potential causal associations, including 10 lipids, 5 amino acids, 1 xenobiotic, 1 carbohydrate, 1 peptide, 2 energy-related metabolites, and 1 unknown metabolite. A similar pattern of associations was observed for these 28 metabolites and ratios in the FinnGen_R8 cohort. The detailed statistical results are provided in Supplementary Tables S3–S6, Table S9

3.2 Meta-analysis and validation of causal relationships

To mitigate heterogeneity across datasets, we performed a meta-analysis to further elucidate the relationship between these metabolites and cancer risk. The results, depicted in Fig. 3, show that 8 metabolites and 3 metabolite ratios had significant causal relationships with cancer risk. These included 2-linoleoylglycerol (18:2), X-25433, glycodeoxycholate 3-sulfate, lignoceroylcarnitine (C24), sulfate from the piperine metabolite C18H21NO3 (3), trans-urocanate, cystine, androstenediol (3beta17beta) monosulfate (1), as well as the glucose-to-mannose ratio, alanine-to-asparagine ratio, and 3-phosphoglycerate-to-adenosine 5′-diphosphate (ADP) ratio. Notably, 2-linoleoylglycerol (18:2), X-25433, the glucose-to-mannose ratio, and the alanine-to-asparagine ratio were associated with an increased risk of cancer. In contrast, glycodeoxycholate 3-sulfate, lignoceroylcarnitine (C24), sulfate from the piperine metabolite C18H21NO3 (3), trans-urocanate, cystine, androstenediol (3beta17beta) monosulfate (1), and the 3-phosphoglycerate-to-ADP ratio were linked to a decreased cancer risk. Detailed statistical results can be found in Supplementary Table S7.

3.3 Analysis of protein-metabolite interactions

We further explored the interactions between proteins and the 11 metabolites or metabolite ratios identified above through protein-metabolite-wide MR analysis and multiple testing corrections. As shown in Fig. 4, 48 proteins were found to be associated with these metabolites, including 16 proteins related to metabolites that increase cancer risk, 31 proteins linked to metabolites that decrease cancer risk, and 1 protein associated with metabolites that both positively and negatively influence cancer risk. Among the 16 proteins linked to increased cancer risk, 9 (C5orf38, LPA, ADH4, CCL24, LOXL3, LDLR, KITLG, BPTF, and COTL1) positively regulate these metabolites, while 7 (CLC, FAM20A, MPI, PRG3, PCOLCE2, LTA, PFKFB2) negatively regulate these metabolites. Of the 31 proteins associated with decreased cancer risk, 11 proteins (INHBA, APOC3, SEPW1, MANF, DLK1, TYRO3, EPCAM, BLMH, SULT2A1, UBLCP1, and FAIM3) positively regulated, whereas 20 (SKP1, CXCL9, ETS2, RP2, NPPA, HAGH, ANGPT1, DAB2, APP, PDGFA, CGB2, LAT, SPARC, MPIG6B, THBS1, CD40LG, DDX19B, DDX19A, KLK11 and KCNA10) negatively regulated these metabolites. Interestingly, the protein F9 was found to positively regulate both lignoceroylcarnitine (C24) levels and the alanine-to-asparagine ratio, thus influencing both increased and decreased cancer risk. Further details are available in Supplementary Table S8.

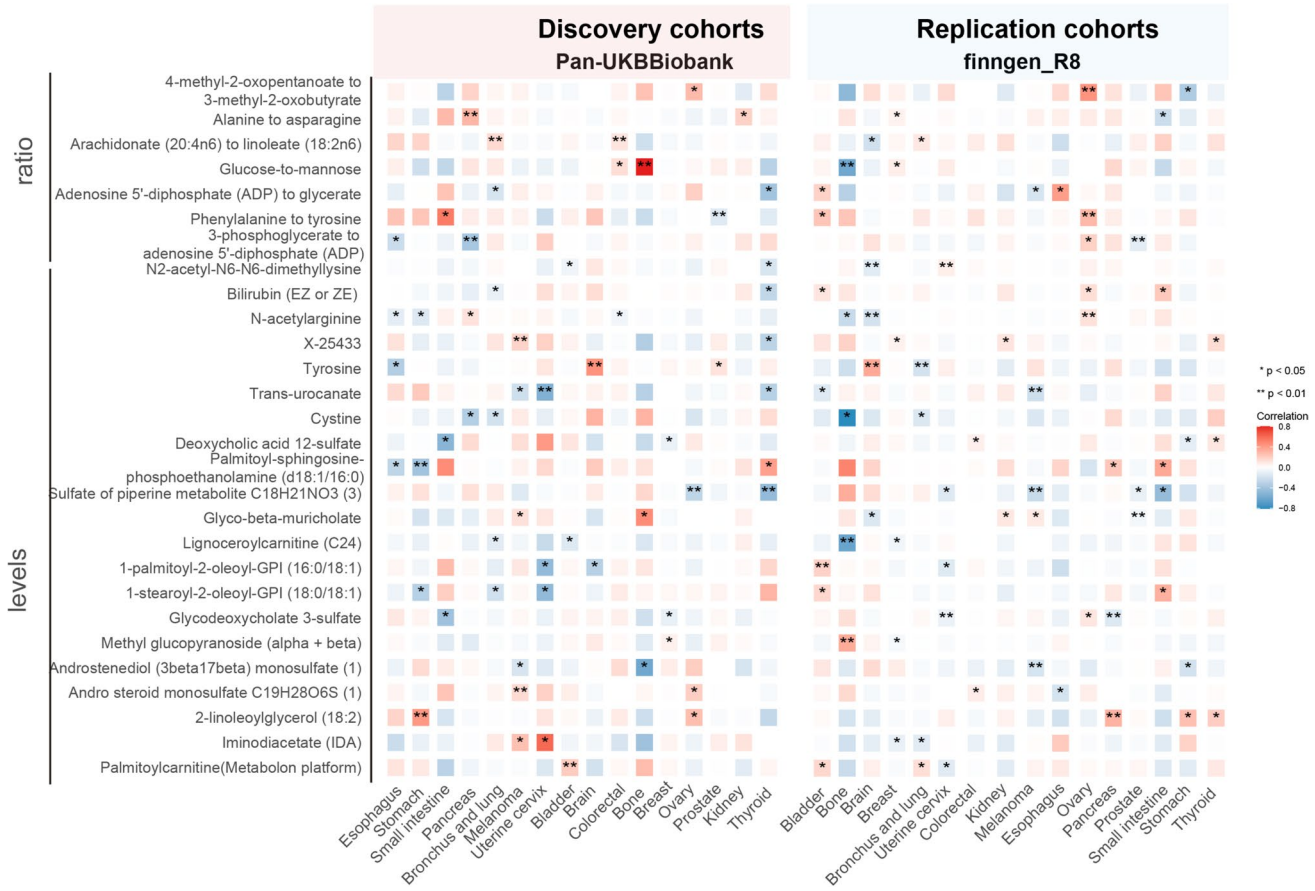

Fig. 2 Heatmap of metabolites related to different cancers in the two cohorts. $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$

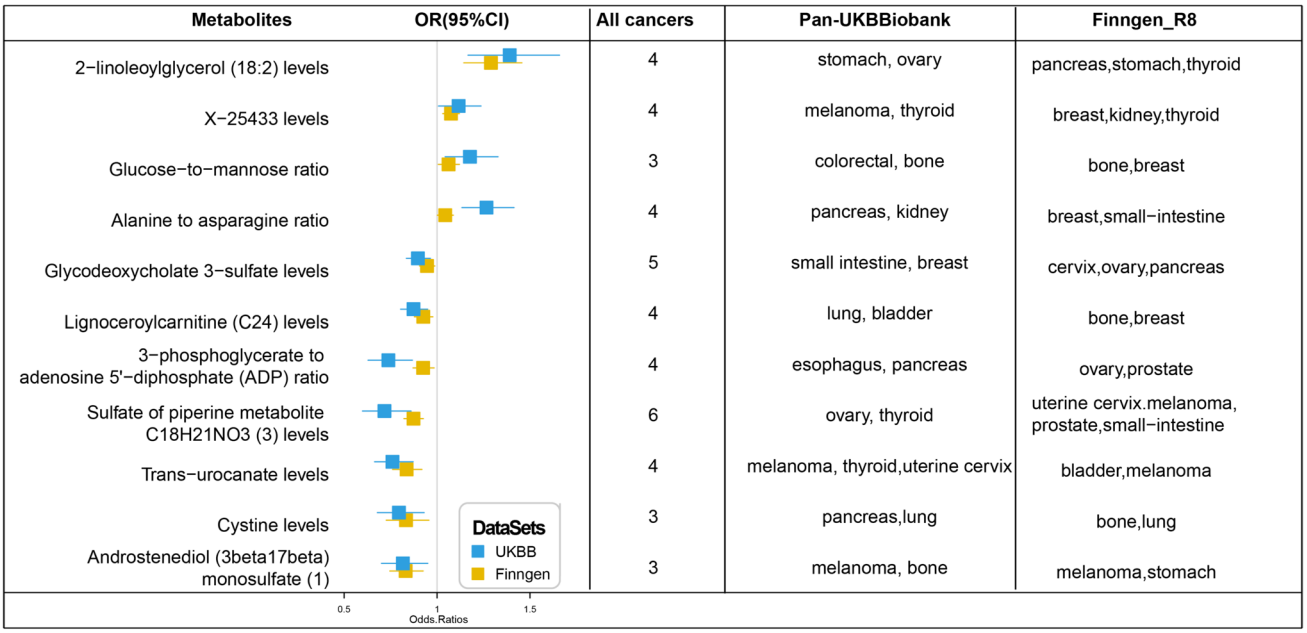

Fig. 3 Forest plots of the associations between the selected metabolites and cancers. The results from the UK Biobank database are presented in blue, while the results from Finngen are presented in yellow. OR, odds ratio; CI confidence interval

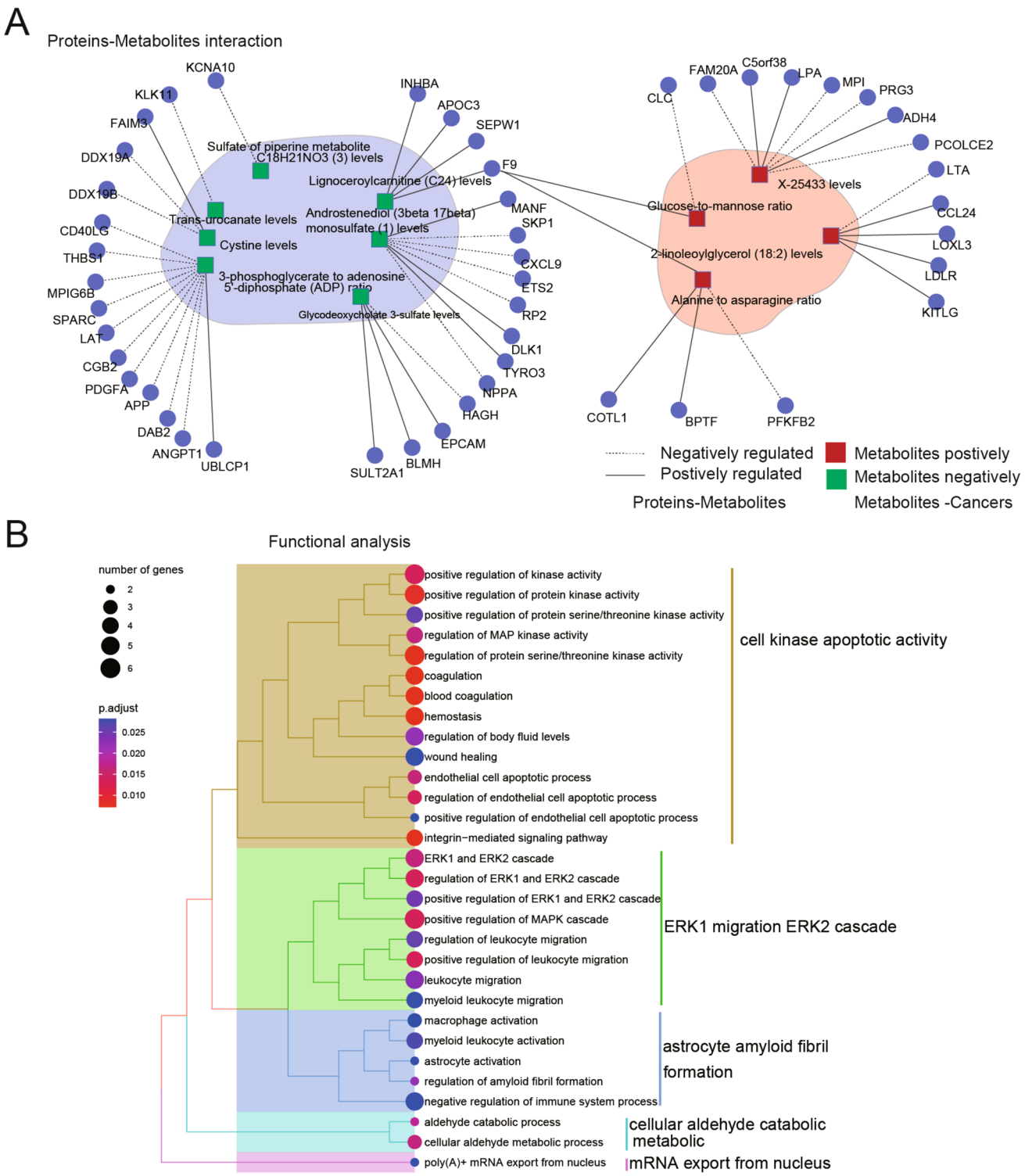

Fig. 4 Protein-metabolite interactions and functional analysis. **A** Interaction of 48 metabolite-related proteins with 11 metabolites or metabolite ratios. **B** Functional analysis of metabolite-related proteins

Functional enrichment analysis revealed that these proteins are implicated in 5 key metabolic pathways associated with cancer, including cell kinase apoptotic function, ERK1 migration and ERK2 cascade, astrocyte amyloid fibril formation, cellular aldehyde catabolic metabolic processes, and mRNA export from the nucleus (Fig. 4).

3.4 Validation with self-reported data

To validate the findings of MR, LC−MS/MS data from 40 patients with breast cancer and 20 healthy controls were analyzed. As shown in Fig. 5, six metabolites or ratios exhibited significant differences between breast cancer patients and controls. Specifically, glucose, mannose, the glucose-to-mannose ratio, and the alanine-to-asparagine ratio were elevated, while asparagine and lignoceroylcarnitine (C24) were reduced in breast cancer patients (all $P < 0.05$). ROC curve analysis indicated that the glu/mannose ratio, alanine/asparagine ratio, and level of lignoceroylcarnitine (C24) performed well in terms of diagnostic value, whereas the diagnostic value of glycodeoxycholate 3-sulfate was relatively poor (Fig. 5).

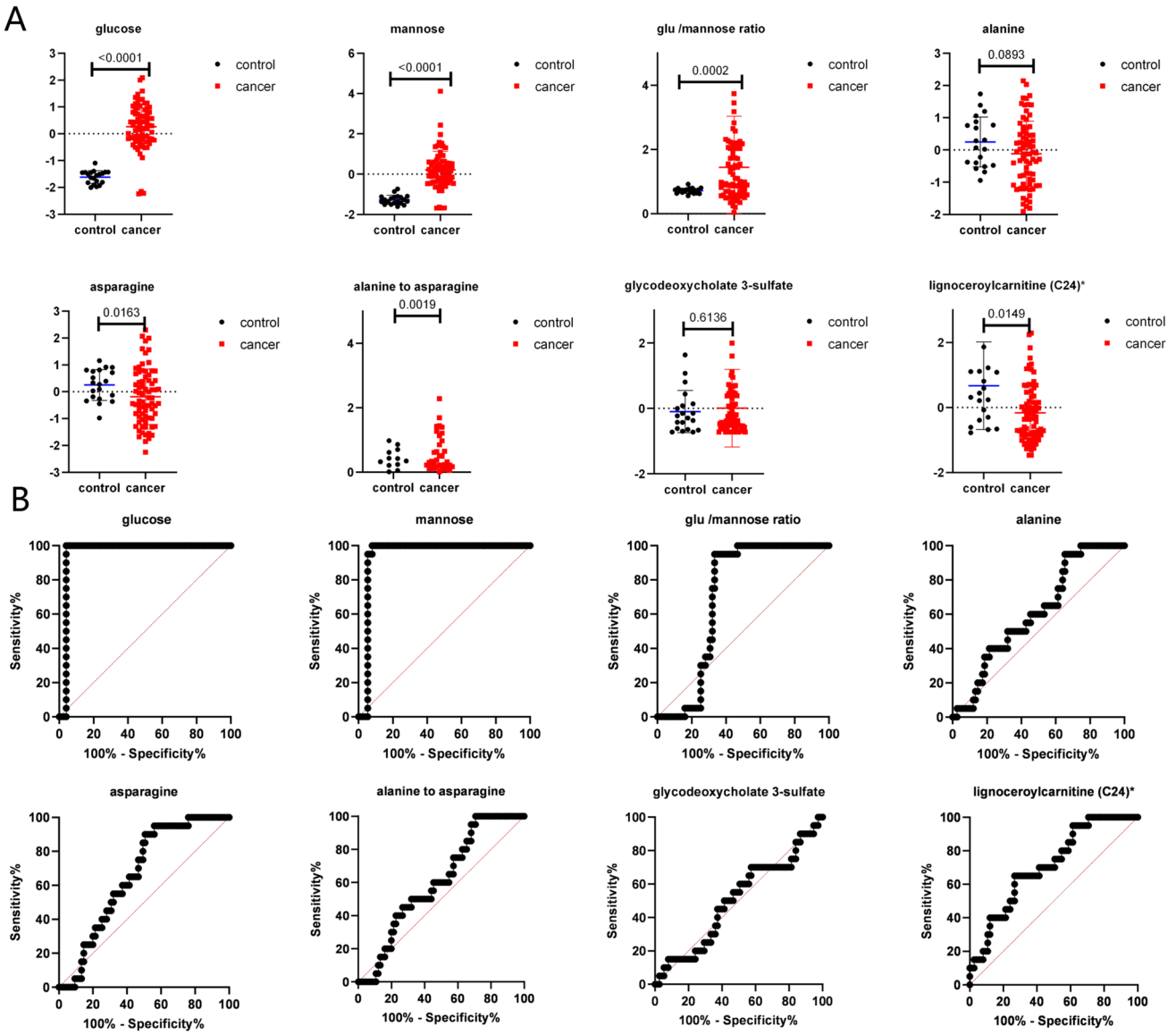

Fig. 5 Self-data replication in breast cancer. **A** Levels or ratios of different metabolites in breast cancer patients versus controls. **B** ROC curve of metabolite levels or ratios. Sen, sensitivity; Sep, specificity

4 Discussion

In this study, we explored the causal relationships between serum metabolites, their regulatory proteins, and cancers. We identified 8 metabolites and 3 metabolite ratios that were causally associated with cancer risk. Furthermore, 48 proteins were found to regulate these metabolites, predominantly enriched in 5 key metabolic pathways linked to cancer: cell kinase apoptotic function, ERK1 migration and ERK2 cascade, astrocyte amyloid fibril formation, cellular aldehyde catabolic metabolic processes, and mRNA export from the nucleus. Clinical data from breast cancer patients validated that high ratios of glucose-to-mannose and alanine-to-asparagine were associated with increased cancer risk, while elevated lignoceroylcarnitine (C24) levels were identified as protective factors, supporting the findings of MR analysis in the pan-cancer study. These findings not only enhance our understanding of cancer pathogenesis but also provide potential avenues for future preventative and therapeutic strategies.

Previous studies have mainly focused on the association of single metabolites (such as lactate) with tumors, but have failed to explain the overall disruption of the metabolic network. This study, through ratio analysis, reveals the universality of metabolite ratios in pan-cancer. Single metabolite levels are easily influenced by transient factors such as sample collection time and dietary status, whereas ratios can reflect the dynamic balance between two metabolites (such as the competitive absorption or conversion of glucose and mannose), providing a closer representation of the true metabolic flux in the body. Additionally, combinatorial biomarkers integrate multidimensional metabolic information, overcoming the diagnostic limitations of single metabolites. By standardizing, they eliminate systematic errors from sample processing or instrument detection, enhancing the reproducibility of results. Specifically, both the glucose-to-mannose and alanine-to-asparagine ratios were shown to be causally associated with increased cancer risk in our MR analysis and were validated in breast cancer patients, suggesting that these ratios are cancer risk factors. Previous research by Yue et al. also identified the glucose-to-mannose ratio as a risk factor for HER2-positive breast cancer [19]. This highlights the possibility that elevated glucose levels or reduced mannose levels increase the ratio. Glucose metabolism is known to promote tumor cell growth and proliferation [20], while while mannose (C6H12O6) has demonstrated tumor growth inhibition in both in vitro and in vivo studies [20]. The glucose-to-mannose ratio in serum shows promise as a potential biomarker for ovarian cancer [21]. Noteworthy findings have demonstrated that higher mannose levels reduce the recurrence and mortality risks in esophageal adenocarcinoma (EAC) patients, underscoring mannose's potential diagnostic and prognostic utility for various cancers. [22, 23]. However, limited research exists on the associations between cancer and other metabolites such as 2-linoleoylglycerol (18:2), X-25433, and the alanine-to-asparagine ratio, warranting further investigation into the mechanisms underlying these associations.

On the other hand, 7 metabolites or metabolites ratio, especially lignoceroylcarnitine (C24), were found to be causally associated with a reduced risk of mutiple cancers. Lignoceroylcarnitine (C24), a member of the acylcarnitines class, plays a crucial role in the energy supply pathway via long-chain fatty acid β-oxidation [24]. Impaired fatty acid oxidation (FAO) and lipid accumulation contribute to conditions such as obesity, insulin resistance, and cardiomyopathy. The long-chain fatty acid β-oxidation of FAO occurs in the mitochondrial matrix. In recent years, research has highlighted that FAO plays a crucial role in the production of several important molecules, including nicotinamide adenine dinucleotide (NADH), flavin adenine dinucleotide (FADH2), nicotinamide adenine dinucleotide phosphate (NADPH), and adenosine triphosphate (ATP), providing a survival advantage to cancer cells and enhance ATP production when needed, helping to inhibit pro-apoptotic pathways in cancer[25, 26]. Due to the lake of evidence, the exact role of lignoceroylcarnitine (C24) in cancer remains unclear. Its protective effect may be linked to fatty acid oxidation metabolism in cancer cells, indicating a potential area for future research.

Our analysis of protein-metabolite interactions revealed that 48 proteins regulate these metabolites, closely linked to several pathways involved in cancer: kinase apoptotic function, ERK1/2 pathway, astrocyte amyloid fibril formation, cellular aldehyde catabolic processes, and mRNA export from the nucleus. Cancer is characterized by dysregulation of key cellular processes, including kinase-mediated apoptotic pathways such as MAPKs (p38, JNK), PI3K, and AKT, which are critical for regulating cell death and metabolic activities [27, 28]. The ERK cascade, by promoting metabolic flexibility, allows cancer cells to switch between various substrates such as glucose, lipids, and amino acids depending on nutrient availability [29]. Amyloid fibrils, typically associated with neurodegenerative diseases, have been observed in the tumor microenvironment, especially in brain tumors like gliomas [30]. Amyloid fibril formation in astrocytes may disrupt protein metabolism and redox homeostasis, facilitating uncontrolled growth and resistance to therapies. Aldehydes, which are byproducts of lipid peroxidation and amino acid metabolism, are highly reactive

and cytotoxic if not efficiently metabolized. In cancer cells, the upregulation of aldehyde dehydrogenases (ALDHs) detoxifies aldehydes, reducing oxidative damage and maintaining redox balance [31]. ALDHs also play a dual role by promoting the oxidation of aldehydes into less toxic acids, thereby supporting cancer cell survival under oxidative stress [31]. Additionally, the export of mRNA from the nucleus, a critical step in gene expression, is dysregulated in several cancers. The export of mRNA, especially those encoding metabolic enzymes and signaling molecules, ensures that tumor cells can adapt their metabolic programs in response to environmental changes.

Our study has several strengths. First, this is the first MR study to integrate metabolomics, proteomics, and genomics to examine the causal relationships between serum metabolites, their related proteins, and cancer. Second, the large sample sizes used in our MR analyses, drawn from different databases, enhance the reliability of our findings. Third, self-reported data were used to replicate and validate our MR results. Our study also has several limitations. First, MR relies on valid genetic instruments; weak instrument bias may affect estimates for metabolites with low heritability. Second, horizontal pleiotropy cannot be fully excluded, though sensitivity analyses showed no significant bias (Supplementary Tables). Besides, MR reflects only lifelong genetic effects and is unable to capture the dynamic changes in metabolites over time and their impact on cancer. Thus, further research is needed to validate our results.

In conclusion, our MR study systematically analyzed the causal effects of metabolites and their related proteins on cancer risk. We identified 11 metabolites or metabolite ratios and 48 associated proteins linked to the risk of multiple cancers, along with 5 key metabolic pathways implicated in cancer development. Further research is needed to fully elucidate the pathogenic mechanisms of these metabolites and proteins in cancer.

Author contributions SLX: Data curation, Writing-Original draft preparation. SLX, ZYX: Conceptualization, Methodology, Software. CC, RA, WCC, DPL: Visualization, Investigation. YZG: Software, Validation. LJW, XYX, JZ:Supervision. LJW, XYX, JZ: Writing-Reviewing and Editing. All authors approve the final version of the manuscript and take responsibility for the integrity of the work.

Funding This study was funded by the Natural Science Foundation of Zhejiang Province (no. LQN25H160028, LQ24H270001), the Key Research and Development Program of Zhejiang Province (no. 2019 C03021), the National Natural Science Foundation of China (no. 82172362), the Key Laboratory of Precision Medicine in Diagnosis and Monitoring Research of Zhejiang Province (2022E10018).

Data availability The data are available from the UK Biobank (www.ukbiobank.ac.uk/) and finnGen (https://www.finngen.fi/en).

Declarations

Ethics approval and consent to participate This study was performed in compliance with all relevant ethical regulations, and all participants provided informed consent. The protocols for the collection of human blood samples in the study were approved by the Ethical Committee of Sir Run-Run Shaw Hospital (no. 20180226–51). All participants provided written informed consent to participate in this study.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49.
2. Wu L, Qu X. Cancer biomarker detection: recent achievements and challenges. Chem Soc Rev. 2015;44(10):2963–97.
3. Giovannucci E, Harlan DM, Archer MC, et al. Diabetes and cancer: a consensus report. Diabetes Care. 2010;33(7):1674–85.
4. Mantovani A, Petracca G, Beatrice G, et al. Non-alcoholic fatty liver disease and increased risk of incident extrahepatic cancers: a meta-analysis of observational cohort studies. Gut. 2022;71(4):778–88.

5. Avgerinos KI, Spyrou N, Mantzoros CS, et al. Obesity and cancer risk: emerging biological mechanisms and perspectives. Metabolism. 2019;92:121–35.
6. Pavlova NN, Thompson CB. The emerging hallmarks of cancer metabolism. Cell Metab. 2016;23(1):27–47.
7. Johnson CH, Patterson AD, Idle JR, et al. Xenobiotic metabolomics: major impact on the metabolome. Annu Rev Pharmacol Toxicol. 2012;52:37–56.
8. Ikeda A, Nishiumi S, Shinohara M, et al. Serum metabolomics as a novel diagnostic approach for gastrointestinal cancer. Biomed Chromatograp BMC. 2012;26(5):548–58.
9. Chen Y, Lu T, Pettersson-Kymmer U, et al. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. Nat Genet. 2023;55(1):44–53.
10. Markozannes G, Kanellopoulou A, Dimopoulou O, et al. Systematic review of mendelian randomization studies on risk of cancer. BMC Med. 2022;20(1):41.
11. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3): e1001779.
12. Kurki MI, Karjalainen J, Palta P, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023;613(7944):508–18.
13. Yang H, Wang Y, Zhao Y, et al. Causal effects of genetically determined metabolites and metabolite ratios on esophageal diseases: a two-sample Mendelian randomization study. BMC Gastroenterol. 2024;24(1):310.
14. Matushyn M, Bose M, Mahmoud AA, et al. SumStatsRehab: an efficient algorithm for GWAS summary statistics assessment and restoration. BMC Bioinformat. 2022;23(1):443.
15. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7: e34408.
16. Sun J, Zhao J, Jiang F, et al. Identification of novel protein biomarkers and drug targets for colorectal cancer by integrating human plasma proteome with genome. Genome Med. 2023;15(1):75.
17. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.
18. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res. 2017;45(D1):D331–8.
19. Yue J, Fang H, Chen S, et al. Causal role of blood metabolites in HER-positive and HER-negative breast cancer: a Mendelian randomization (MR) study. Aging. 2024;16(15):11626–55.
20. Tang M, Ren X, Fu C, et al. Regulating glucose metabolism using nanomedicines for cancer therapy. J Mater Chem B. 2021;9:5749–64.
21. Chen Y, Yao Q, Zhang L, et al. HPLC for simultaneous quantification of free mannose and glucose concentrations in serum: use in detection of ovarian cancer. Front Chem. 2023;11:1289211.
22. Gu J, Liang D, Pierzynski JA, et al. D-mannose: a novel prognostic biomarker for patients with esophageal adenocarcinoma. Carcinogenesis. 2017;38:162–7.
23. Jin H, Liu X, Liu HX. Biological function, regulatory mechanism, and clinical application of mannose in cancer. Biochim Biophys Acta Rev Cancer. 2023;1878(6): 188970.
24. Guo P, Furnary T, Vasiliou V, et al. Non-targeted metabolomics and associations with per- and polyfluoroalkyl substances (PFAS) exposure in humans: a scoping review. Environ Int. 2022;162: 107159.
25. Choupani E, Mahmoudi Gomari M, Zanganeh S, Nasseri S, Haji-Allahverdipoor K, Rostami N, Hernandez Y, Najafi S, Saraygord-Afshari N, Hosseini A. Newly developed targeted therapies against the androgen receptor in triple-negative breast cancer: a review. Pharmacol Rev. 2023;75(2):309–27.
26. Li YJ, Fahrmann JF, Aftabizadeh M, Zhao Q, Tripathi SC, Zhang C, Yuan Y, Ann D, Hanash S, Yu H. Fatty acid oxidation protects cancer cells from apoptosis by increasing mitochondrial membrane lipids. Cell Rep. 2022;39(9): 110870.
27. Ullah R, Yin Q, Snell AH, et al. RAF-MEK-ERK pathway in cancer evolution and treatment. Semin Cancer Biol. 2022;85:123–54.
28. Wagner EF, Nebreda AR. Signal integration by JNK and p38 MAPK pathways in cancer development. Nat Rev Cancer. 2009;9(8):537–49.
29. Papa S, Choy PM, Bubici C. The ERK and JNK pathways in the regulation of metabolic reprogramming. Oncogene. 2019;38(13):2223–40.
30. Brown MR, Radford SE, Hewitt EW. Modulation of β-amyloid fibril formation in Alzheimer's Disease by microglia and infection. Front Mol Neurosci. 2020;13: 609073.
31. Xia J, Li S, Liu S, Zhang L. Aldehyde dehydrogenase in solid tumors and other diseases: potential biomarkers and therapeutic targets. MedComm. 2023;4(1):195.