

## RESEARCH ARTICLE

# Genetically predicted 486 blood metabolites in relation to risk of colorectal cancer: A Mendelian randomization study

Zhangjun Yun<sup>1</sup>  | Ziwei Guo<sup>1</sup>  | Xiao Li<sup>2</sup> | Yang Shen<sup>2</sup> | Mengdie Nan<sup>1</sup> | Qing Dong<sup>2</sup> | Li Hou<sup>2</sup> 

<sup>1</sup>Dongzhimen Hospital, Beijing University of Chinese Medicine (BUCM), Beijing, China

<sup>2</sup>Department of Oncology and Hematology, Dongzhimen Hospital, Beijing University of Chinese Medicine (BUCM), Beijing, China

**Correspondence**

Qing Dong and Li Hou, Department of Oncology and Hematology, Dongzhimen Hospital, Beijing University of Chinese Medicine, 5 Haiyun Cang, Beijing 100700, People's Republic of China.  
Email: [drdong@tom.com](mailto:drdong@tom.com) and [houli1203@126.com](mailto:houli1203@126.com)

**Funding information**

National Natural Science Foundation of China Grant Program, Grant/Award Number: 81573959; the Capital Health Development Research Fund, Grant/Award Number: 2020-2-4193 and 2022-1-4171

**Abstract**

**Background:** Metabolic disorders are a hallmark feature of cancer. However, the evidence for the causality of circulating metabolites to promote or prevent colorectal cancer (CRC) is still lacking. We performed a two-sample Mendelian randomization (MR) analysis to assess the causality from genetically proxied 486 blood metabolites to CRC.

**Methods:** Genome-wide association study (GWAS) data for exposures were extracted from 7824 Europeans GWAS on metabolite levels. GWAS data for CRC from the GWAS catalog database GCST012879 were used for the preliminary analysis. The random inverse variance weighted (IVW) is the primary analysis for causality analysis while MR-Egger and weighted median as complementary analyses. Cochran Q test, MR-Egger intercept test, MR-PRESSO, Radial MR, and leave-one-out analysis were used for sensitivity analyses. For significant associations, additional independent CRC GWAS data GCST012880 were used for replication analysis and meta-analysis. For the final identification of metabolites, Steiger test, linkage disequilibrium score regression, and colocalization analysis were performed for further evaluation. Multivariable MR was performed to assess the direct effect of metabolites on CRC.

**Results:** The results of this study indicated significant associations between six metabolites pyruvate (odds ratio [OR]: 0.49, 95% confidence interval [CI]: 0.32–0.77,  $p=0.002$ ), 1,6-anhydroglucose (OR: 1.33, 95% CI: 1.11–1.59,  $p=0.002$ ), nonadecanoate (19:0) (OR: 0.40, 95% CI: 0.4–0.68,  $p=0.0008$ ), 1-linoleoylglycerophosphoethanolamine (OR: 0.47, 95% CI: 0.30–0.75,  $p=0.001$ ), 2-hydroxystearate (OR: 0.39, 95% CI: 0.23–0.67,  $p=0.0007$ ), gamma-glutamylthreonine (OR: 2.14, 95% CI: 1.02–4.50,  $p=0.040$ ) and CRC. MVMR analysis revealed that genetically predicted pyruvate, 1-linoleoylglycerophosphoethanolamine and gamma-glutamylthreonine can directly influence CRC independently of other metabolites.

**Conclusion:** The current work provides evidence to support the causality of the six circulating metabolites on CRC and a new perspective on the exploration of

Zhangjun Yun and Ziwei Guo contributed equally.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Cancer Medicine* published by John Wiley & Sons Ltd.

the biological mechanisms of CRC by combining genomics and metabolomics. These findings contribute to the screening, prevention and treatment of CRC.

#### KEYWORDS

blood metabolites, causality, colocalization analysis, colorectal cancer, Mendelian randomization

## 1 | INTRODUCTION

Colorectal cancer (CRC) is currently the most common form of digestive system cancer. According to updated cancer statistics, over 1.9 million new cases of CRC were diagnosed in 2020, resulting in 93,500 deaths.<sup>1</sup> This accounts for approximately one-tenth of all cancer cases and deaths. The incidence of CRC has increased from fifth to second place worldwide between 2018 and 2020.<sup>1,2</sup> Given the circumstances, enhancing the prevention, and screening of CRC is a crucial priority strategy. Previous studies have provided sufficient evidence that smoking,<sup>3</sup> alcohol consumption,<sup>4</sup> Type 2 diabetes (T2D),<sup>5</sup> body mass index (BMI),<sup>6</sup> waist-to-hip ratio (WHR),<sup>6</sup> and total cholesterol (TC)<sup>7</sup> are common risk factors for CRC. Nonetheless, limited research has been conducted on metabolic changes in CRC.

In recent years, the emergence of metabolomics as a component of systems biology has provided a novel approach to investigating the mechanisms underlying diseases. Specifically, metabolomics can provide insight into the biological mechanisms of diseases by identifying modified metabolites or metabolic pathways.<sup>8,9</sup> Over the past decades, there is a growing body of evidence that metabolic reprogramming and energy metabolism are crucial for the proliferation and metastasis of cancer cells.<sup>10,11</sup> For normal cells, while the altered metabolism provides support for cell proliferation or division, it can also affect cell differentiation making the cells predisposed to cancer.<sup>10</sup> In addition, targeted regulation of metabolites has potential applications in the treatment of cancer. The most noteworthy example of cancer treatment through targeted regulation of metabolism is dichloroacetate (DCA). Previous studies provide convincing evidence that DCA can cut off pyruvate dehydrogenase (PDH) phosphorylation<sup>12</sup> and reverse the Warburg effect to boost mitochondrial pyruvate oxidation to inhibit tumor proliferation.<sup>13,14</sup> Regulating cellular metabolism has been shown to increase the sensitivity of cancer cells to treatment.<sup>15</sup> The combination of cellular metabolism inhibitors is considered a promising strategy to overcome the chemotherapy resistance, presenting a potential avenue for future research.

Therefore, exploring the metabolites associated with the development of CRC not only contributes to early

screening and prevention of CRC, but also helps to understand the biological mechanisms of CRC for treatment. Unfortunately, the causality between metabolites and CRC is unclear, as there are no prospective studies of metabolites and CRC to date. The unavoidable design weaknesses of conventional observational studies, such as altered metabolites due to deliberate lifestyle changes in patients after the cancer diagnosis, chronic intake of certain drugs, and tumor cell-induced changes in metabolic substances, resulted in an ambiguous causal relationship between metabolites and CRC. Rigorous randomized controlled trials (RCTs) have the highest credibility in evidence-based medicine to prove causal effects, but they are difficult to implement due to ethical issues, observation time, costly money, and other constraints. Hence, metabolites that increase the risk of CRC cannot be identified based on the existing evidence.

Mendelian randomization (MR) studies have recently gained widespread usage in the investigation of disease etiology. In the absence of RCTs, MR is the most compelling strategy to explore the causality between the exposure of interest and outcome.<sup>16</sup> MR assessed the causal effects of genetically proxied exposure of interest on outcomes by selecting exposure-associated single nucleotide polymorphisms (SNPs) as instrumental variables (IVs).<sup>17</sup> Specifically, this IVs alternative approach mimics RCTs since SNPs are randomly assigned to offspring at the time of conception, which largely avoids confounding factors because sex and age are less likely to bias the causal effect.<sup>18</sup> Similarly, the reverse causality caused by MR studies is less likely since genotype formation is before the disease.

Since there is a lack of understanding related to the causality between blood metabolites and CRC, further research is required in this area. In this work, we performed MR analysis to comprehensively explore the causal effects of 486 blood metabolites on CRC via genome-wide association study (GWAS) summary data. Furthermore, we conducted colocalization analysis and metabolic pathway analysis to investigate the underlying biological processes of CRC at the gene and protein levels. This study aims to reveal the metabolism-related etiology of CRC and to provide insight into its biological processes.

## 2 | METHODS AND MATERIALS

### 2.1 | Study design

A valid MR study should comply with three assumptions: (1) IVs are strongly associated with the exposures of interest; (2) IVs are independent of confounding factors; (3) IVs are not associated with outcome and affect outcome only via exposures.<sup>19</sup> All MR analyses in this study were performed in R software (4.2.1) by Two Sample MR, MRPRESSO, and Radial MR packages. An overview of the study is shown in Figure 1. This study was designed with reference to the MR study by Cai et al.<sup>20</sup>

### 2.2 | GWAS data for 486 blood metabolites and CRC

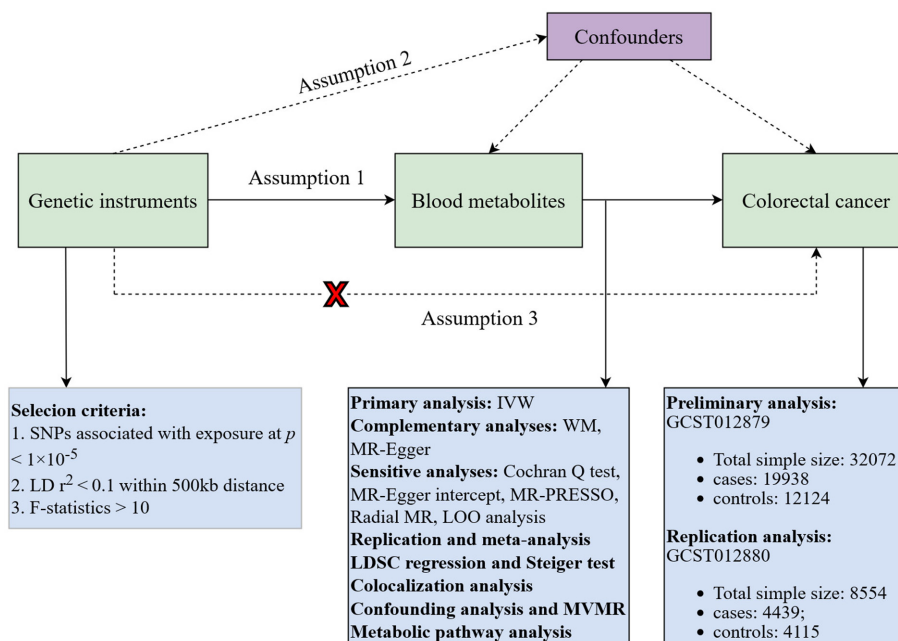
Genetic data for blood metabolites were accessed from the metabolomics GWAS server (<https://metabolomics.helmholtz-muenchen.de/gwas/>). Notably, this is the most comprehensive report to date on the genetic loci of blood metabolites, which eventually identified almost 2.1 million SNPs for 486 metabolites associated with human genetic variants by Genome-wide association scans with high-throughput metabolic profiling conducted by Shin et al.<sup>21</sup> The detailed names of 486 metabolites are presented in Table S1, where the chemical properties of the metabolites named with X- are

unknown. Specifically, the study included 7824 European, including 1768 from the KORA F4 study in Germany and 6056 from the UK Twin Study. Among the 486 metabolites, 107 were defined as unknown due to as yet poorly defined chemical properties. Another 309 metabolites were chemically authenticated and allocated to eight broad metabolic groups, including amino acid, carbohydrate, cofactors and vitamin, energy, lipid, nucleotide, peptide, and xenobiotic metabolism, as documented in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.<sup>22</sup>

GWAS summary data for CRC download from GWAS Catalog (<https://www.ebi.ac.uk/gwas/>)<sup>23</sup> on November 5, 2022 and the GWAS Catalog accession number is GCST012879. Specifically, the GWAS data containing 39,216,056 SNPs were derived from a meta-analysis of previous GWAS studies related to CRC conducted by Huyghe and colleagues, with a total sample size of 32,072 Europeans containing 19,938 cases and 12,124 controls.<sup>24</sup> More detailed information about the GWAS data can be obtained from the study of Huyghe et al.<sup>24</sup> The GWAS data for CRC mentioned above were used for preliminary analysis.

### 2.3 | IVs selection

To satisfy assumption (1), IVs associated with blood metabolites were identified by rigorous screening conditions from multiple perspectives. Given the modest number



**FIGURE 1** Overview of this Mendelian randomization (MR) analysis. Assumption 1, genetic instruments are strongly associated with the exposures of interest; Assumption 2, genetic instruments are independent of confounding factors; Assumption 3, genetic instruments are not associated with outcome and affect outcome only via exposures. IVW, inverse variance weighted; LD, linkage disequilibrium; LDSC, linkage disequilibrium score; LOO analysis, leave-one-out analysis; MR-PRESSO, MR-Pleiotropy RESidual sum and outlier; MVMR, multivariable Mendelian randomization analysis; SNPs, single nucleotide polymorphisms; WM, weighted median.

of metabolite-associated SNPs, we eased the significance threshold  $p < 1 \times 10^{-5}$  to select SNPs related to metabolites. Then, we clumped SNPs by removing linkage disequilibrium (LD,  $R^2 > 0.1$  and within 500 kb). This criterion has been widely applied in previous studies.<sup>25,26</sup> To eliminate the bias induced by poor instruments, we calculated  $R^2$  and  $F$  statistics for each SNP.  $R^2$  and  $F$  statistics are calculated as follows:

$$R^2 = \frac{2 * \beta^2 * EAF * (1 - EAF)}{[2 * \beta^2 * EAF * (1 - EAF) + 2 * (se(\beta))^2 * N * EAF * (1 - EAF)]};$$

$$F = \frac{N - k - 1}{k} * \frac{R^2}{1 - R^2}$$

where  $\beta$  is the effect size for the genetic variant of interest; EAF is the effect allele frequency for the genetic variant of interest;  $se(\beta)$  is the standard error of effect size for the genetic variant of interest;  $R^2$  is the instrumental variable that explains the degree of exposure (determinant coefficient of regression equation);  $N$  is the sample size of the exposure;  $k$  is the number of SNPs (instrumental variants). SNPs with  $F < 10$  were defined as poor genetic variants and were removed.<sup>18</sup> Next, we extracted metabolite-associated SNPs from the outcome and discarded SNPs associated with the outcome ( $p < 1 \times 10^{-5}$ ). We further harmonized SNPs for exposure and outcome, and palindromic effects and allelic inconsistent SNPs were removed (e.g. A/G vs. A/C). Then, to satisfy assumption (3), we removed outcome-related SNPs ( $p < 1 \times 10^{-5}$ ) in the IVs. Finally, we performed MR analysis on metabolites with more than two SNPs.<sup>27</sup>

## 2.4 | Statistical analysis and sensitivity analysis

The causal effect of blood metabolites and CRC was primarily assessed based on the results of random-effect inverse variance weighted (IVW). Because IVW estimates are derived from a summary analysis of Wald ratios for all genetic variants.<sup>28</sup> IVW is based on the assumption that there is no horizontal pleiotropy for all SNPs, under this premise IVW provides the most accurate assessment of causal effects.<sup>28</sup> Hence, we initially screened for blood metabolites with causal effects on CRC using IVW-based estimates. To acquire more reliable results, we applied other two methods to further evaluate the metabolites with significant estimates (IVW derived  $p < 0.05$ ). MR-Egger and weighted median (WM) methods were defined as complementary analyses. These two methods can provide more robust estimates under lenient conditions. WM allows less than 50% of SNPs to be invalid, while MR-Egger provides horizontal pleiotropy and heterogeneity detection in the presence of horizontal pleiotropy for all SNPs.<sup>28,29</sup> MR-Egger regression can provide unbiased estimates when

consistent with the InSIDE assumption (Strength of IVs independent of direct effects).<sup>29</sup>

Sensitivity analysis is essential because it examines the horizontal pleiotropy and heterogeneity that can severely violate MR estimates. Horizontal pleiotropy is observed when IVs influence the outcome through other pathways than the exposure of interest. As such, we performed several tests to ensure convincing estimates. In this study, we utilized four methods to detect and correct for heterogeneity and pleiotropy, including the Cochran Q test, MR-Egger intercept test, MR-Pleiotropy RESidual Sum and Outlier (MR-PRESSO), and Radial MR. Cochran Q test-derived  $p < 0.05$  was considered as heterogeneity of the results.<sup>30</sup> MR-Egger intercept was calculated to test for directional pleiotropy and bias due to invalid IVs.<sup>29</sup> Subsequently, Radial MR<sup>31</sup> was performed to identify outliers, and MR analysis was repeated after eliminating heterogeneous SNPs. Ultimately, we use MR-PRESSO<sup>32</sup> to check again for the presence of heterogeneous SNPs. In terms of the robustness of the results, we performed a leave-one-out (LOO) analysis which assesses whether the results are heavily influenced by a single SNP by discarding each SNP in turn and then performing MR analysis.<sup>29</sup>

In a word, we rigorously screened blood metabolites with potential causal effects on CRC by multiple criteria: (1)  $p$  value for the primary analysis was significant (IVW derived  $p < 0.05$ ). (2) Consistent direction and magnitude within the three MR methods. (3) No heterogeneity or horizontal pleiotropy in MR results. (4) MR estimates are not severely disturbed by a single SNP.

Moreover, to assess the statistical power of the estimates, we calculated the power with an online website (<https://shiny.cnsgenomics.com/mRnd/>).<sup>33</sup> Specifically, this tool calculates power values based on asymptotic theory to detect causal effects derived from IVs. We set the Type I error rate is 0.05 and calculated power using  $R^2$  of IVs, the proportion of cases with an outcome, and the odds ratio (OR) derived from IVW analysis.

## 2.5 | Replication and meta-analysis

To comprehensively estimate the robustness of the candidate metabolites identified based on the above criteria, we replicated the IVW analysis in an additional CRC cohort that included 4439 European ancestry cases and 4115 European ancestry controls.<sup>24</sup> The GWAS data for CRC in the replication analysis was also obtained from GWAS Catalog and the accession number is GCST012880. Specifically, this GWAS data is part of the results of a meta-analysis of 30 existing GWAS studies for CRC. The study included 8554 Europeans from the Germany, Netherlands, United States, and Nauru and identified 39,216,056 SNPs



for CRC. In brief, GWAS data with the accession number GCST012879 is used for preliminary analysis while GWAS data with accession number GCST012880 is used for replication analysis. We finally determined the blood metabolites with causal effects on CRC by the results of a meta-analysis of two MR analyses. Meta-analysis was implemented based on a random effects IVW model on Review Manager 5.4 software.

## 2.6 | Evaluation of genetic correlation and directionality

However, MR estimates can violate causal effects under the premise of genetic correlation between exposure and outcome of interest.<sup>34,35</sup> Although SNPs related to CRC were excluded in the selection of IVs, SNPs with no relation may also mediate the genetics of CRC. Linkage disequilibrium score (LDSC) regression can estimate coinheritance by performing Chi-squared statistics for two traits based on SNP. Hence, to ensure causal effects were not confused by the coheritability of exposure with the outcome, LDSC was implemented to check the genetic correlation between the screened metabolites and CRC.

In addition, we further performed the Steiger test to reject the resulting bias caused by reverse causality.<sup>36</sup> The direction of causal inference might be false in the presence of explained variance of IVs in CRC is stronger than blood metabolites.

## 2.7 | Colocalization analysis

To further investigate whether the associations of the identified metabolites between CRC were driven by a locus within a genomic region, we performed colocalization analysis using the coloc R package.<sup>37</sup> Genes with important regulatory effects on metabolites were identified from previous studies. The cis-expression quantitative trait loci (eQTL) and protein quantitative trait loci (pQTL) summary-level data for genes expression in blood were obtained from the eQTLGen Consortium including 31,684 blood samples (<https://www.eqtlgen.org/>)<sup>38</sup> and a large-scale pQTL study in 35,559 Icelanders,<sup>39</sup> respectively. This colocalization analysis can pinpoint the causal association between metabolites and CRC to the same causal variant locus within a specific genomic region. For each locus of variation, the coloc method assessed the posterior probability (H0, H1, H2, H3, and H4) of the following five hypotheses in the Bayesian framework: (1) no association with either trait; (2) association with Trait 1 only; (3) association with Trait 2 only; (4) both traits are associated, but distinct

causal variants were for two traits; and (5) both traits are associated, and the same shares causal variant for both traits.<sup>40</sup> The colocalization analysis was performed using the default priors ( $p1 = 1 \times 10^{-4}$ ,  $p2 = 1 \times 10^{-4}$ , and  $p12 = 1 \times 10^{-5}$ ). PP.H4 > 80% colocalization analysis (H4) results provide strong evidence supporting the existence of shared causal variants within specific genomic regions affecting gene expression and CRC risk.

## 2.8 | Confounding analysis and multivariable MR analysis

Although we assessed the horizontal pleiotropy of the MR results through a series of sensitivity analyses to detect any SNPs that violated the MR assumptions,<sup>2</sup> there may also be little residual confounding SNPs. We checked IVs for metabolites at the Phenoscanner V2 website (<http://www.phenoscanner.medschl.cam.ac.uk/>) to evaluate whether each SNP was associated with known risk factors for CRC, such as smoking,<sup>3</sup> alcohol consumption,<sup>4</sup> T2D<sup>5</sup> BMI,<sup>6</sup> WHR,<sup>6</sup> and TC.<sup>7</sup> If any SNPs were observed to be associated with the above confounding factors ( $p < 1 \times 10^{-5}$ ), MR analysis would be re-performed after removing these SNPs to verify the reliability of the results.

To avoid IVs violating Assumptions 2 and 3 of MR, MR analysis is required to ensure that genetic variants are associated with a single risk factor. In practice, however, some genetic variants are associated with multiple risk factors, which is described as pleiotropy.<sup>41</sup> In this case, multivariable MR (MVMR) can correct for interactions of genetic variation between exposures by incorporating multiple exposures that may interact with each other. In brief, univariable MR assessed the total effect of exposure on outcome, and MVMR assessed the direct effect of each exposure (independent of any other exposures) on outcome.<sup>41</sup> In this study, we performed MVMR on the identified metabolites to adjust for their interactions. MVMR was performed using IVW<sup>42</sup> and MR-PRESSO.<sup>32</sup> The IVW method of multivariate MR is to regress all exposed SNPs with the outcome, weighting for the inverse variance of the outcome. MRPRESSO can remove outliers to correct for the pleiotropy of IVs.

## 2.9 | Metabolic pathway analysis

To specify the biological mechanisms involved in blood metabolites that have causal effects on CRC, we further performed metabolic pathway analyses using the MetaboAnalyst 5.0 (<https://www.metaboanalyst.ca/>)<sup>43</sup> to explore the potential pathogenesis of CRC.

### 3 | RESULTS

#### 3.1 | Preliminary analysis

After tightly controlling the quality of IVs, 486 metabolites were eventually captured in the MR study. The filtered IVs contained SNPs ranging from 3 to 501 (genetically proxied for glutamate consisted of 3 SNPs; 2-methoxyacetaminophen sulfate generated the most genetically proxied: 501 SNPs). The F statistics of all SNPs associated with metabolites were greater than 10, which indicated a strong power of IVs. The detailed data of IVs are presented in Table S2. All outliers are identified and removed by Radial MR prior to MR analysis (Table S3). Then IVW analysis was performed to preliminarily identify 27 metabolites with potential causal effects on CRC, including 20 metabolites whose chemical identity is known and seven metabolites whose chemical identity is unknown (Figure 2). As shown in Figure 2, the 20 known metabolites were chemically assigned to the amino acid, carbohydrate, dipeptide, lipid, nucleotide, peptide, and xenobiotics. After combining complementary and sensitivity analyses, nine eligible metabolites that met the rigorous screening criteria were identified as candidates (Table 1), including pyruvate (OR 0.47, 95% confidence interval (CI): 0.28–0.79,  $p=0.005$ ), 1,6-anhydroglucose (OR 1.29, 95% CI: 1.06–1.58,  $p=0.012$ ), nonadecanoate (19:0) (OR 0.39, 95% CI: 0.21–0.73,  $p=0.003$ ), taurochenodeoxycholate (OR 0.64, 95% CI: 0.49–0.83,  $p=0.0008$ ), 1-linoleoylglycerophosphoethanolamine (OR 0.42, 95% CI: 0.2–0.73,  $p=0.002$ ), 2-hydroxystearate (OR 0.39, 95% CI: 0.21–0.72,  $p=0.002$ ), gamma-glutamylthreonine (OR 2.84, 95% CI: 1.37–5.90,  $p=0.005$ ), X-11470 (OR 0.43, 95% CI: 0.32–0.58,  $p=2.98 \times 10^{-08}$ ), X-14205 (OR 0.71, 95% CI: 0.54–0.93,  $p=0.012$ ). In summary, the estimates derived from IVW are significant ( $p < 0.05$ ) while the direction and magnitude of IVW, MR-Egger, and WM estimates were consistent (Figure 3). After removing the outliers, the MR-PRESSO results also did not favor the existence of heterogeneous SNPs (Table S4). The Cochran Q test ( $p > 0.05$ ) and MR-Egger intercept test ( $p > 0.05$ ) provided strong evidence for the absence of heterogeneity and pleiotropy (Table 1). LOO analysis results supported that a single SNP did not cause bias in MR estimation (Figure S1). The statistical power of all estimates is  $>0.8$  (Table 1). These nine blood metabolites are considered to be candidates for further analysis.

#### 3.2 | Replication and meta-analysis

To enhance the persuasiveness of the estimates, we replicated the MR analysis with another GWAS data for CRC. As

would be expected, similar trends were observed for candidate metabolites in another GWAS data for CRC, although the results were not significant due to the enormous difference in sample size. The results of the meta-analysis further determined that eight blood metabolites (six known and two unknown) can affect CRC (Figure 4). In detail, genetic susceptibility for higher levels of pyruvate (OR 0.49, 95% CI: 0.32–0.77,  $p=0.002$ ), nonadecanoate (19:0) (OR 0.40, 95% CI: 0.24–0.68,  $p=0.0008$ ), 1-linoleoylglycerophosphoethanolamine (OR 0.47, 95% CI: 0.30–0.75,  $p=0.001$ ), 2-hydroxystearate (OR 0.39, 95% CI: 0.23–0.67,  $p=0.0007$ ), X-11470 (OR 0.52, 95% CI: 0.32–0.83,  $p=0.006$ ), X-14205 (OR 0.73, 95% CI: 0.58–0.92,  $p=0.008$ ) decreased risk of CRC, while genetic liability for higher levels of 1,6-anhydroglucose (OR 1.33, 95% CI: 1.11–1.59,  $p=0.002$ ), gamma-glutamylthreonine (OR 2.14, 95% CI: 1.02–4.50,  $p=0.040$ ) promoted susceptibility to CRC. Even though it showed a consistent direction in both MR analyses, taurochenodeoxycholate was discarded due to the meta-analysis revealed non-significant estimates (OR 0.75, 95% CI: 0.52–1.08,  $p=0.120$ ).

#### 3.3 | Evaluation of genetic correlation and directionality

LDSC-based estimates noted little genetic correlation observed between CRC and pyruvate ( $R_g=0.259$ ,  $Se=0.218$ ,  $p=0.236$ ), 1,6-anhydroglucose ( $R_g=0.285$ ,  $Se=0.304$ ,  $p=0.348$ ) nonadecanoate (19:0) ( $R_g=0.169$ ,  $Se=0.186$ ,  $p=0.363$ ), 1-linoleoylglycerophosphoethanolamine ( $R_g=-0.003$ ,  $Se=1.636$ ,  $p=0.999$ ), 2-hydroxystearate ( $R_g=0.098$ ,  $Se=0.134$ ,  $p=0.466$ ), gamma-glutamylthreonine ( $R_g=-0.017$ ,  $Se=0.268$ ,  $p=0.950$ ), X-11470 ( $R_g=0.273$ ,  $Se=0.286$ ,  $p=0.340$ ), X-14205 ( $R_g=0.109$ ,  $Se=0.380$ ,  $p=0.774$ ) suggesting that MR estimates are not confounded by shared genetic components. Subsequently, we estimated the SNP-heritability of eight metabolites based on LDSC, and the SNP-heritability (proportion of variance attributed to genome-wide SNPs) of metabolites ranged from 0.0318 (1-linoleoylglycerophosphoethanolamine) to 0.1601 (gamma-glutamylthreonine) (Table S5). In addition, the Steiger test revealed that the causality between genetically proxied metabolites and CRC was not violated by reverse causal effects (Table S6).

#### 3.4 | Colocalization analysis

For the six known metabolites that have been identified in this study that are associated with CRC risk, we searched for relevant targets of metabolites in previously published studies. We found that lactate dehydrogenase A (LDHA) regulated pyruvate metabolism and was significantly

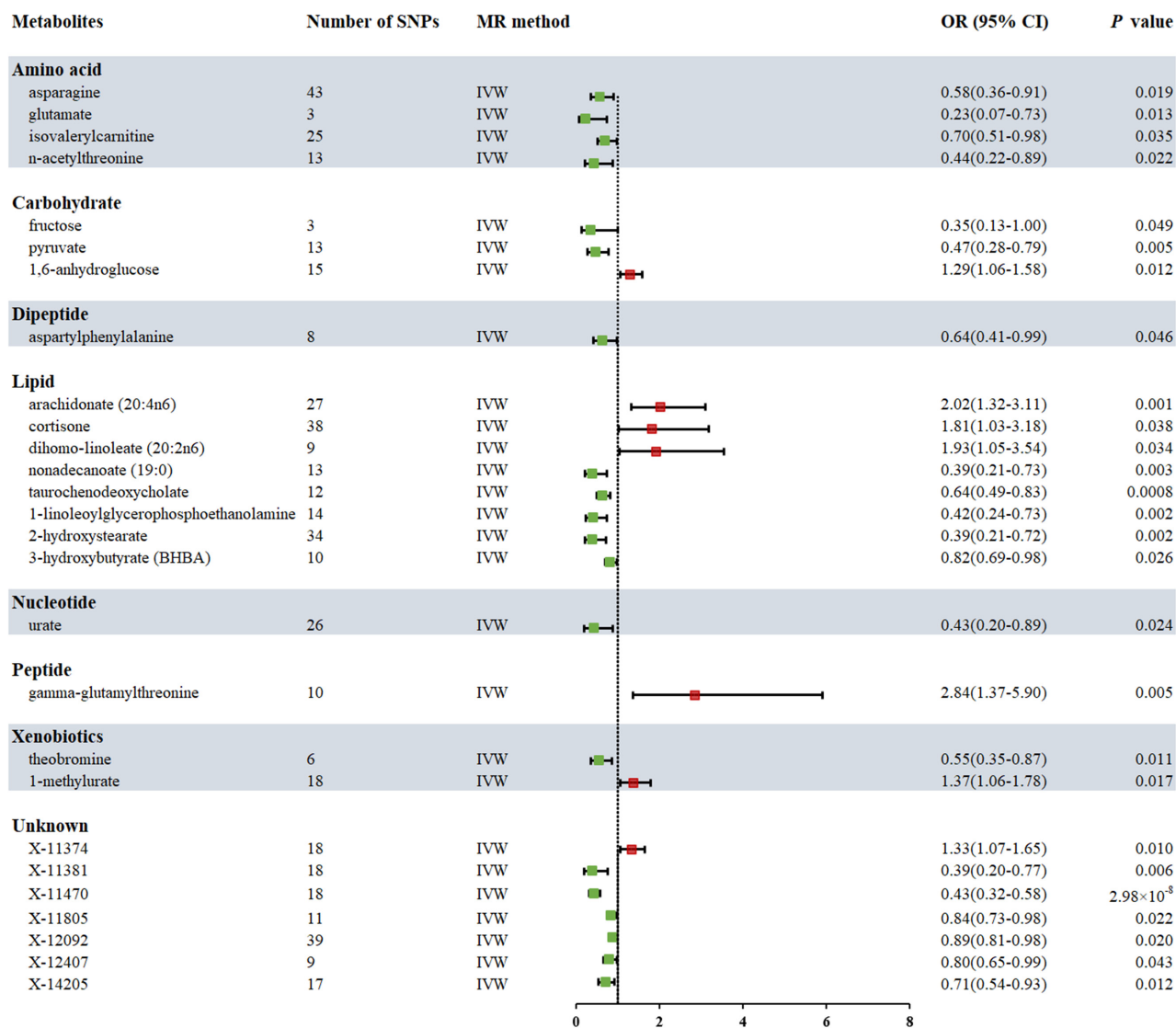


FIGURE 2 Forest plot for the causality of blood metabolites on colorectal cancer derived from inverse variance weighted (IVW) analysis. CI, confidence interval; IVW, inverse variance weighted; OR, odds ratio; SNPs, single nucleotide polymorphisms.

associated with the development of cancer,<sup>44,45</sup> while the eQTL and pQTL of LDHA in blood were also available in publicly available databases. Few targets with important regulatory effects on other metabolites have been reported. Therefore, we only performed colocalization analysis of LDHA from the gene and protein level. We accessed whole blood LDHA eQTL and pQTL signals at genome-wide significant levels ( $p < 5 \times 10^{-8}$ ) from the eQTL consortium and a large-scale GWAS on blood proteome (Table S7,S8). Colocalization analysis results from eQTL (PP.H4 = 99.59%, SNP = rs6486426) and pQTL (PP.H4 = 99.85%, SNP = rs116841148) strongly support that LDHA expression and CRC risk are driven by the shared causal variant loci. Specifically, significant loci on GWAS signaling can increase CRC risk by influencing the biological process of LDHA expression.

### 3.5 | Confounding analysis and MVMR

Although SNPs that violated the estimates have been excluded by sensitivity analysis in this work, to satisfy assumption<sup>2</sup> (IVs are independent of confounding factors), we checked whether all SNPs associated with eight metabolites were independent of common risk factors (smoking, alcohol consumption, T2D, BMI, and WHR) for CRC in the Phenoscanner one by one. We found gamma-glutamylthreonine to be independent of any confounding factors. Among the other IVs of seven metabolites, a total of 13 SNPs were observed to be associated with common risk factors for CRC (Table S9). As expected, after removing these SNPs, the estimates remained significant: pyruvate (OR 0.53, 95% CI: 0.30–0.93,  $p = 0.027$ ), 1,6-anhydroglucose (OR

TABLE 1 Supplementary and sensitivity analyses for causality from blood metabolites on colorectal cancer.

Metabolites	N	MR analysis			Heterogeneity		Pleiotropy		Power
		Methods	OR (95% CI)	p	Q	p	Intercept p		
Amino acid									
Asparagine	43	ME	0.40(0.17–0.96)	0.047	35.19	0.76	0.005	0.35	1.00
		WM	0.50(0.25–1.00)	0.053					
Glutamate	3	ME	0.0003(4 × 10 <sup>-11</sup> -2600.78)	0.503	0.67	0.72	0.116	0.57	1.00
		WM	0.18(0.04–0.79)	0.023					
Isovalerylcarnitine	25	ME	0.47(0.25–0.89)	0.030	16.19	0.88	0.011	0.17	1.00
		WM	0.60(0.37–0.98)	0.041					
N-acetylthreonine	13	ME	0.17(0.02–1.38)	0.127	11.24	0.51	0.017	0.37	1.00
		WM	0.39(0.14–1.10)	0.075					
Carbohydrate									
Fructose	3	ME	0.05(0.001–1.77)	0.347	1.33	0.51	0.052	0.46	1.00
		WM	0.26(0.07–0.94)	0.041					
Pyruvate	13	ME	0.53(0.12–2.42)	0.429	13.17	0.36	-0.003	0.88	1.00
		WM	0.38(0.18–0.79)	0.010					
1,6-anhydroglucose	15	ME	1.22(0.82–1.82)	0.335	10.41	0.73	0.004	0.76	0.99
		WM	1.28(0.95–1.71)	0.105					
Dipeptide									
Aspartylphenylalanine	8	ME	0.87(0.27–2.77)	0.822	8.07	0.33	-0.012	0.59	1.00
		WM	0.63(0.36–1.12)	0.118					
Lipid									
Arachidonate (20:4n6)	27	ME	2.58(1.18–5.60)	0.025	20.09	0.79	-0.006	0.47	1.00
		WM	3.07(1.65–5.74)	0.0004					
Cortisone	38	ME	3.18(0.91–11.06)	0.077	24.87	0.94	-0.007	0.33	1.00
		WM	2.42(1.05–5.53)	0.037					
Dihomo-linoleate (20:2n6)	9	ME	0.85(0.12–5.94)	0.878	7.47	0.49	0.019	0.42	1.00
		WM	1.57(0.67–3.68)	0.304					
Nonadecanoate (19:0)	13	ME	0.48(0.14–1.68)	0.276	15.09	0.24	-0.006	0.70	1.00
		WM	0.33(0.14–0.75)	0.008					
Taurochenodeoxycholate	12	ME	0.52(0.33–0.80)	0.015	12.41	0.33	0.014	0.27	1.00
		WM	0.60(0.42–0.85)	0.004					
1-linoleoylglycerophosphoethanolamine	14	ME	0.42(0.05–3.56)	0.442	9.95	0.70	-9.47E-06	1.00	1.00
		WM	0.50(0.24–1.07)	0.076					
2-hydroxystearate	34	ME	0.84(0.19–3.76)	0.817	22.71	0.91	-0.009	0.28	1.00
		WM	0.31(0.13–0.75)	0.009					
3-hydroxybutyrate (BHBA)	10	ME	0.90(0.71–1.16)	0.451	3.56	0.94	-0.012	0.33	0.84
		WM	0.82(0.66–1.02)	0.081					
Nucleotide									
Urate	26	ME	0.26(0.05–1.28)	0.111	29.37	0.25	0.007	0.50	1.00
		WM	0.25(0.09–0.68)	0.007					
Peptide									
Gamma-glutamylthreonine	10	ME	5.18(0.67–39.87)	0.153	12.93	0.17	-0.016	0.55	1.00
		WM	2.02(0.83–4.89)	0.120					

(Continues)



TABLE 1 (Continued)

Metabolites	N	MR analysis			Heterogeneity		Pleiotropy		Power
		Methods	OR (95% CI)	p	Q	p	Intercept p		
Xenobiotics									
Theobromine	6	ME	0.30(0.009–9.59)	0.534	4.50	0.48	0.021	0.75	1.00
		WM	0.62(0.34–1.14)	0.125					
1-methylurate	18	ME	1.48(0.95–2.32)	0.103	9.06	0.94	−0.004	0.68	1.00
		WM	1.45(1.02–2.07)	0.040					
Unknown									
X-11374	18	ME	1.67(1.04–2.69)	0.050	9.61	0.92	−0.013	0.30	1.00
		WM	1.44(1.05–1.96)	0.023					
X-11381	18	ME	1.02(0.15–6.97)	0.987	11.23	0.84	−0.016	0.32	1.00
		WM	0.46(0.19–1.22)	0.087					
X-11470	18	ME	0.31(0.19–0.52)	0.0003	14.08	0.66	0.015	0.14	1.00
		WM	0.38(0.23–0.63)	0.0002					
X-11805	11	ME	0.83(0.66–1.04)	0.148	10.57	0.39	0.002	0.88	1.00
		WM	0.82(0.67–1.01)	0.060					
X-12092	39	ME	0.89(0.78–1.01)	0.089	22.12	0.98	0.0007	0.89	1.00
		WM	0.91(0.81–1.03)	0.129					
X-12407	9	ME	0.79(0.45–1.37)	0.434	3.30	0.91	0.001	0.95	0.89
		WM	0.82(0.63–1.07)	0.152					
X-14205	17	ME	0.33(0.14–0.81)	0.029	9.49	0.89	0.032	0.10	1.00
		WM	0.66(0.46–0.95)	0.025					

Abbreviations: CI, confidence interval; ME, MR-Egger; N, number of single nucleotide polymorphisms; OR, odds ratio; WM, weighted median.

1.31, 95% CI: 1.06–1.61,  $p = 0.011$ ), nonadecanoate (19:0) (OR 0.35, 95% CI: 0.19–0.66,  $p = 0.001$ ), 1-linoleoylglycerophosphoethanolamine (OR 0.49, 95% CI: 0.26–0.91,  $p = 0.025$ ), 2-hydroxystearate (OR 0.39, 95% CI: 0.21–0.72,  $p = 0.003$ ), X-11470 (OR 0.44, 95% CI: 0.32–0.60,  $p = 3.00 \times 10^{-7}$ ), X-14205 (OR 0.75, 95% CI: 0.56–1.00,  $p = 0.047$ ).

After adjusting for metabolite interactions, MVMR estimates based on IVW (Figure 5A) and MR-PRESSO (Figure 5B) methods both showed that genetically predicted pyruvate, 1-linoleoylglycerophosphoethanolamine and gamma-glutamylthreonine can directly influence CRC independently of other metabolites.

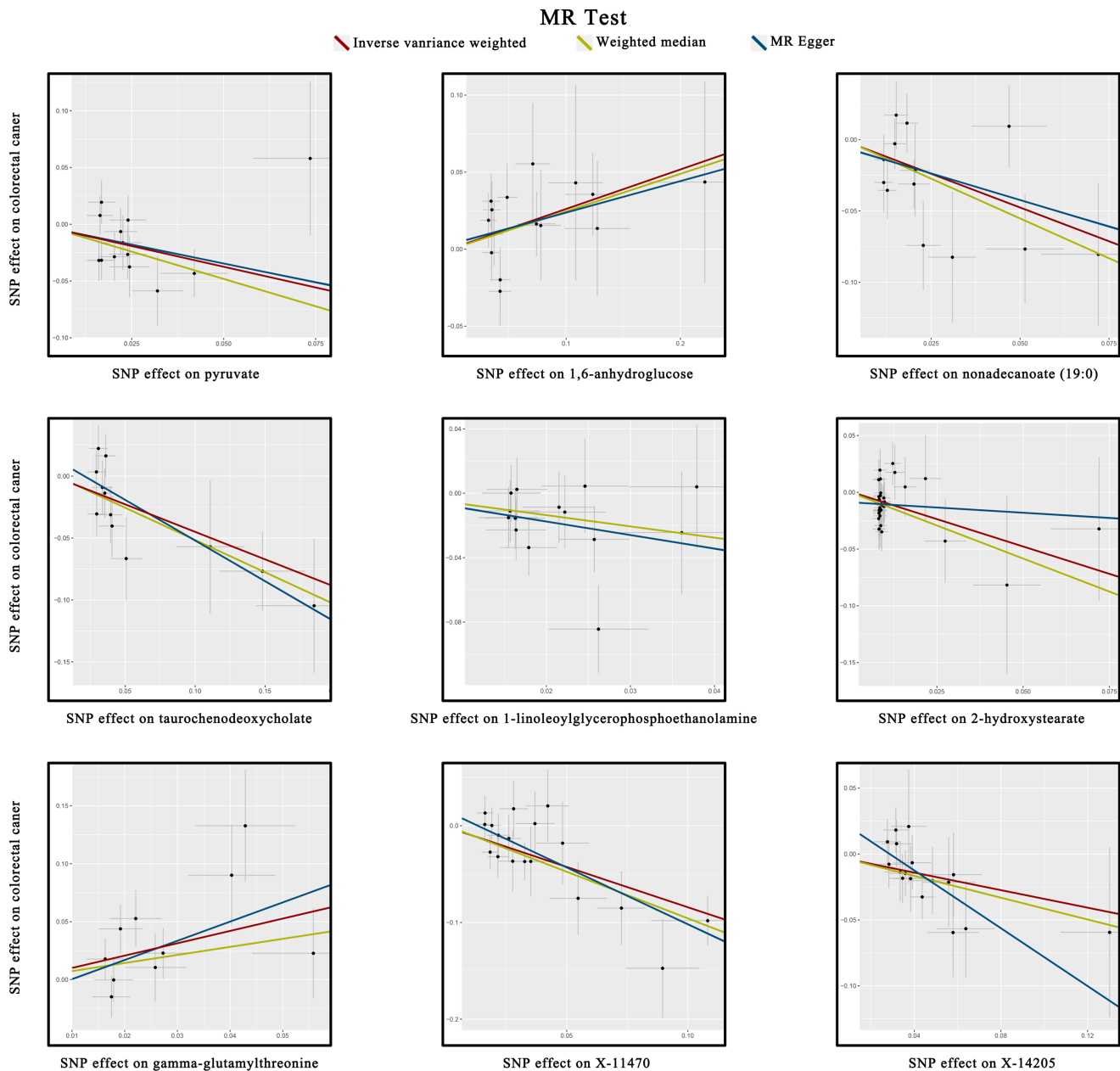
### 3.6 | Metabolic pathway analysis

Based on six known metabolites, we identified nine metabolic pathways that may be involved in the mechanism of CRC pathogenesis (Table S10). Citrate cycle (TCA cycle), pyruvate metabolism, glycolysis/gluconeogenesis, alanine, aspartate and glutamate metabolism, glyoxylate and dicarboxylate metabolism, glycine, serine, and threonine metabolism, cysteine and methionine metabolism,

arginine and proline metabolism, and tyrosine metabolism may be the potential biological mechanism for the development of CRC. Notably, pyruvate is involved in all metabolic pathways. This means that pyruvate and related metabolic pathways take an important role in the pathogenesis of CRC.

## 4 | DISCUSSION

In the current work, we integrated two large-scale GWAS data to explore the causal effects of genetically proxied 486 blood metabolites on CRC via a rigorous MR design. We finally determined that genetically determined high levels of pyruvate, nonadecanoate (19:0), 1-linoleoylglycerophosphoethanolamine, 2-hydroxystearate, X-11470, and X-14205 associated with lower CRC risk while genetically predisposition to high levels of 1,6-anhydroglucose and gamma-glutamylthreonine increased risk of CRC. MVMR estimates suggested that pyruvate, 1-linoleoylglycerophosphoethanolamine and gamma-glutamylthreonine can directly affect CRC independently of other metabolites. Colocalization analysis provided strong evidence that significant loci on GWAS



**FIGURE 3** Scatterplot of significantly associated (IVW derived  $p < 0.05$ ) and directionally consistent estimates. SNP, single nucleotide polymorphisms.

signaling can increase CRC risk by influencing the expression of LDHA. Subsequently, we identified 9 metabolic pathways that may be involved in the biological mechanisms of CRC. To our knowledge, this is the first MR study to date that applied the most comprehensive blood metabolite GWAS data to explore the causality with CRC and incorporates metabolic pathways and colocalization analysis.

In recent years, the high incidence and mortality rates of CRC have put a heavy burden on people around the world, making early screening and prevention of CRC an urgent strategy. The emergence of metabolomics

technologies has led to an increasing interest in exploring the perceived value of metabolites in CRC. Notably, blood metabolites visually provide a snapshot of biological mechanisms since they simultaneously capture endogenous and exogenous processes.<sup>46</sup> For example, recent evidence suggests that tryptophan is metabolized by tumor cell-secreted indoleamine-2,3-dioxygenase (IDO) and tryptophan-2,3-dioxygenase (TDO) catabolism to generate metabolites, particularly kynurenine.<sup>47</sup> Depletion of tryptophan and accumulation of kynurenine induced T cell dysfunction and apoptosis resulting in immunosuppression.<sup>47,48</sup> Emerging IDO inhibitors are under clinical

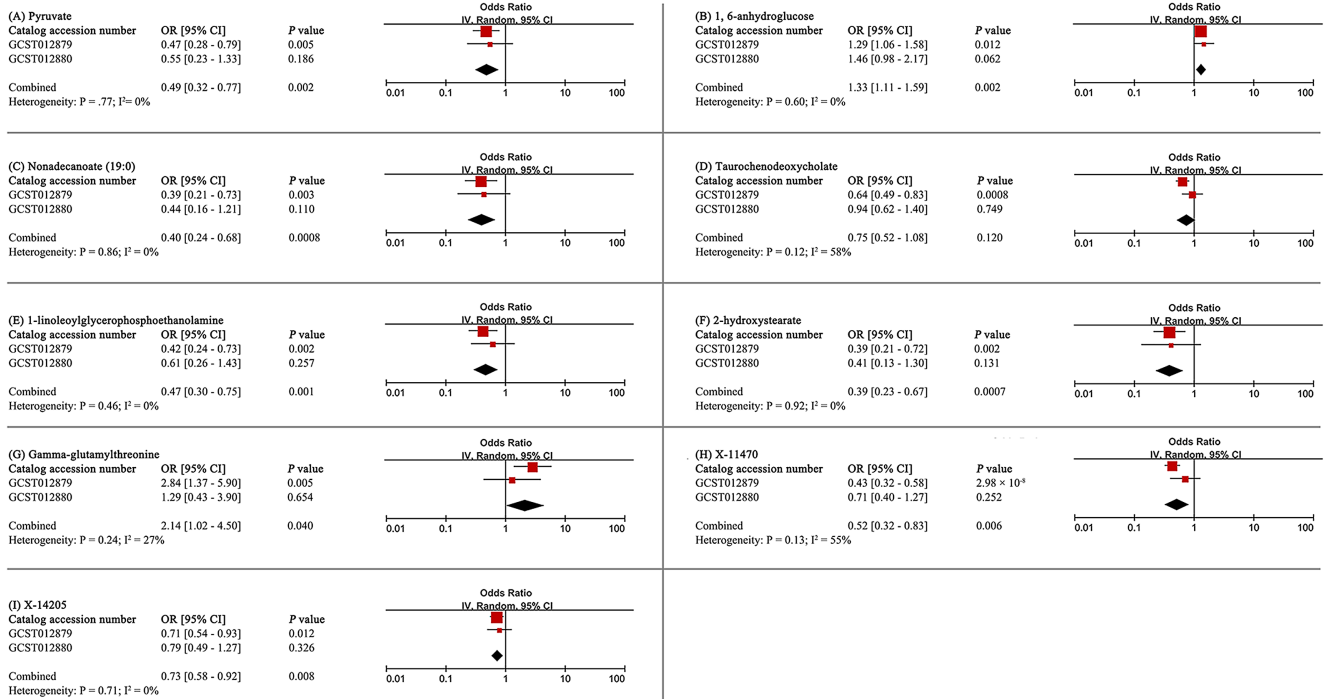


FIGURE 4 Meta-analysis of significantly associated (IVW derived  $p < 0.05$ ) between metabolites and colorectal cancer. 95% CI, 95% confidence interval; OR, odds ratio.

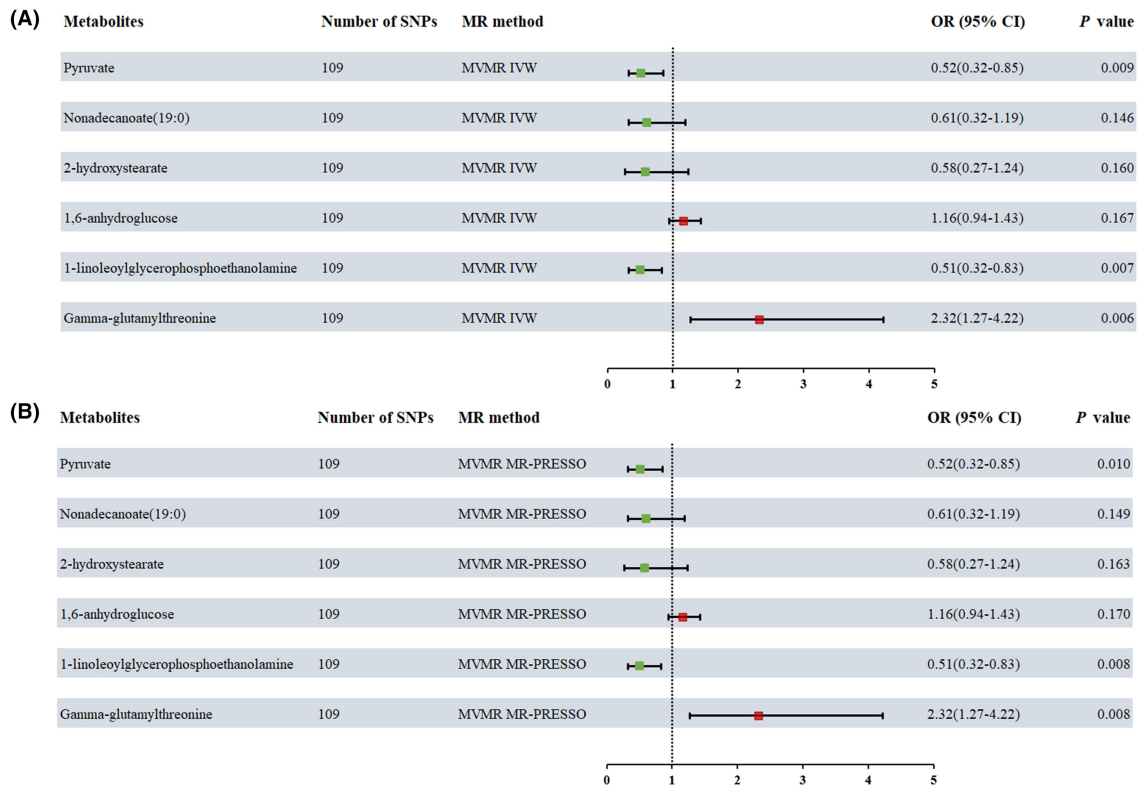


FIGURE 5 Multivariable MR analysis of the final identified blood metabolites. 95% CI, 95% confidence interval; IVW, inverse variance weighted; MVMR, Multivariable Mendelian randomization; MR-PRESSO, MR-Pleiotropy RESidual Sum and Outlier; OR, odds ratio.

trials in a variety of cancers and are promising immunotherapeutic approaches applied to CRC.<sup>48,49</sup> Although previous studies have provided convincing evidence that metabolites are involved in the biological mechanisms of CRC, which are beneficial for CRC treatment, the contribution to early screening and prevention of CRC is limited due to the ambiguous causal relationship between the both. Therefore, we implemented a critical MR study in the hope that the causal relationship between blood metabolites and CRC and the metabolic pathways involved in them can be clarified, thus providing a reference direction for the screening and treatment of CRC.

In this work, we found that genetic sensitivity to high levels of pyruvate can keep the body away from CRC, while pyruvate is involved in nine significantly enriched metabolic pathways including the citrate cycle and pyruvate metabolism. Several previous studies are consistent with our results that pyruvate shows great opportunities for the prevention and treatment of cancer. Pyruvate bridges cytoplasmic and mitochondrial metabolism as a hub of cellular metabolism, and it mainly originates from the final process of glycolysis that pyruvate kinase encoding the conversion of phosphoenolpyruvate to pyruvate.<sup>50</sup> Pyruvate in the cytosol is converted to lactate by lactate dehydrogenase (LDH), and a portion of pyruvate enters the mitochondrial matrix via the mitochondrial pyruvate carrier (MPC) and is converted by the PDH complex to produce acetyl coenzyme A and other substances to carry out the citrate cycle or to regulate cholesterol and lipid metabolism.<sup>51</sup> Returning to the evidence from previous studies, we speculate that high levels of pyruvate may reduce the risk of CRC through the following pathways. The first and most important pathway is to reverse the Warburg Effect. The Warburg Effect was first reported by Warburg O et al. in 1927 and widely accepted that tumor cells rely more on glycolysis than mitochondrial phosphorylation for energy supply which is different from normal cells.<sup>52</sup> One of the culprits of the Warburg Effect is the weakening of MPC activity.<sup>53</sup> CRC is more glycolysis-dependent in the early stages, while MPC activity is downregulated.<sup>54</sup> High levels of pyruvate can increase MPC activity and thus enhance mitochondrial oxidative phosphorylation to reduce glycolysis and lactate production, which reverses the Warburg effect to some extent.<sup>55</sup> Second, pyruvate is an endogenous anti-inflammatory and anti-oxidant molecule. Ramos-Ibeas P et al. reported that pyruvate presented the best antioxidant stress effect on fibroblasts and embryonic stem cells, superior to traditional well-known antioxidants such as selenium, N-acetylcysteine, and Trolox.<sup>56</sup> Pyruvate also inhibited the expression of various inflammatory factors such as tumor necrosis factor, NF- $\kappa$ B pathway, and interleukin-6 and promoted insulin secretion.<sup>57</sup> As described, the anti-oxidative and anti-inflammatory

effects of pyruvate support a microenvironment that is not conducive to cancer development. Ultimately, pyruvate can also reduce fat and alleviate insulin resistance. Ultimately, pyruvate can also reduce fat and alleviate insulin resistance. High levels of pyruvate promote the tricarboxylic acid cycle and reduce lactate synthesis. Lactate is a key mediator in PDH kinase-induced adipogenesis and is associated with obesity and insulin resistance.<sup>58,59</sup> Therefore, adequate amounts of pyruvate may indirectly decrease the synthesis of fat and ameliorate obesity to lower the risk of CRC. To summarize, pyruvate appears to be a metabolite of great opportunity in the prevention and treatment of CRC.

This MR study also identified three other blood metabolites (nonadecanoate (19:0), 1-linoleoylglycerophosphoethanolamine, and 2-hydroxystearate) that have protective effects on CRC. Unfortunately, there are few previous reports of nonadecanoate-related effects, and the relationship between nonadecanoate and cancer deserves further study. For 1-linoleoylglycerophosphoethanolamine, Mika A et al. noted that phosphoethanolamine was highly expressed in CRC tissues but its contribution to CRC was not clear.<sup>60</sup> Another prospective cohort study suggested that serum higher levels of linoleoylglycerophosphocholine decreased the risk of atherosclerosis and kidney failure.<sup>61</sup> In addition, 1-linoleoylglycerophosphoethanolamine is an important part of the phosphatidylethanolamine (PE), which also consists of ethanolamine, phosphoric acid, and glycerol.<sup>62</sup> PE as a major component of cell membrane phospholipids is important for maintaining the stability of cell structure. The precursors of PE, ethanolamine and ethanolamine phosphate, have been shown to inhibit the proliferation and metastasis of a variety of cancer cells.<sup>63</sup> Significant tumor shrinkage was observed after 2 weeks of ethanolamine intervention in colon xenograft mice.<sup>63,64</sup> However, the gap in the specific biological mechanism of PE in CRC remains to be filled. For 2-hydroxystearate, the study conducted by Hongping Xia et al. pointed out that 2-hydroxystearate is highly expressed in hepatocellular carcinoma tissue with diabetes. This seems to imply that pyruvate can affect cancer through diabetes as a mediator.<sup>65</sup>

We also confirmed genetic predisposition to higher levels of gamma-glutamylthreonine and 1,6-anhydroglucose were detrimental to CRC. To date, there are no reports of gamma-glutamylthreonine and CRC, with only one study indicating an increased risk of prostate cancer at high gamma-glutamylthreonine concentrations.<sup>66</sup> We attempted to speculate on the association of  $\gamma$ -glutamine amino acids and CRC from the  $\gamma$ -glutamine cycle due to  $\gamma$ -glutamylcyclotransferase catalyzing the generation of various  $\gamma$ -glutamyl amino acids from



glutathione.  $\gamma$ -glutamylcyclotransferase is highly expressed in a variety of cancers including CRC and is associated with poor prognosis in CRC.<sup>67,68</sup> High levels of  $\gamma$ -glutamylcyclotransferase caused the accumulation of  $\gamma$ -glutamine amino acids. This seems to imply an association between  $\gamma$ -glutamine amino acids and the risk of CRC although further studies are needed to confirm this. For 1,6-anhydroglucose, it is involved in glycolysis and pyruvate metabolism. 1,5-anhydroglucitol, which has a similar structure to 1,6-anhydroglucose, reflects blood glucose levels for 1–2 weeks and is associated with type 1 diabetes while the reports on 1,6-anhydroglucose are extremely limited.<sup>69</sup> However, the specific effects of these metabolites on CRC need to be explored in detail under experimental conditions.

In addition, the high support of colocalization evidence from gene and protein wide was observed between LDHA and CRC. LDH catalyzes pyruvate to lactate, the last step of glycolysis. The elevation of LDH increases the rate of glycolysis, which in turn provides an energy source for cancer cells. Moreover, the acidic microenvironment resulting from the accumulation of lactate also promotes cancer cell invasion and angiogenesis.<sup>44</sup> LDHA, a subtype of LDH, has been demonstrated to have excellent anticancer activities. Inhibition of LDHA enhances the sensitivity of cancer cells to chemotherapy, while suppressing cell proliferation, invasion, and epithelial-to-mesenchymal transition (EMT).<sup>44,70</sup> Transcription factors like c-myc, HIF-1, and p53 can inhibit LDHA activity, exerting anti-tumor effects.<sup>44</sup> Our colocalization analysis results also support that significant sites on GWAS signaling can affect CRC by influencing LDHA expression. Nevertheless, further experimental validation is needed to determine if LDHA can be a viable treatment target for CRC.

This MR analysis has several advantages. First and most critically this is the most complete and systematic study to date on exploring the causality between blood metabolites and CRC because in this work we analyzed 486 blood metabolites. Second, rigorous MR analysis was applied to reject the inevitable defects of previous studies such as reverse causality and confounding interference. In detail, to generate convincing estimates a series of methods are implemented to ensure that any factors that violate the MR assumptions are removed. The consistency of the three MR estimates in the direction and the sensitivity analysis demonstrated the robustness of the results. Third, the reliability of the results was further verified by additional GWAS data for replication analysis and meta-analysis. Although the results of the replication analysis were not significantly attributable to differences in sample size, it showed consistent directionality with the primary analysis, which does not appear to have occurred by accident. Fourth, we assessed the heritability of IVs and the

genetic correlation between metabolites and CRC using LDSC, which made the MR estimates more convincing. We also performed colocalization analysis to demonstrate, at both the gene and protein levels, that alterations in significant loci on GWAS signaling can influence CRC by regulating LDHA expression.

The current study also has several limitations. One limitation of our study is the limited number of SNPs available for the exposure of interest at the genome-wide level. To address this, we set slightly relaxed thresholds for our MR analysis, a practice commonly adopted in other studies. However, the F-statistic value for all selected SNPs exceeded 10, suggesting that our IVs were sufficiently robust. Moreover, the consistent causal direction supported by the results of the Steiger test lends credibility to our relaxed threshold setting. Second, to minimize the impact of ethnic variability, we only used GWAS data from individuals of European ancestry for this MR analysis. Hence, the generalizability of our findings to other populations warrants further exploration and validation. The third limitation of our study is that the precision of MR estimation partly relies on sample size. Therefore, expanding the sample size is necessary to confirm the reliability of our results. Additionally, while MR analysis provides valuable insights into etiology, it is important to note that our findings should be validated through rigorous RCTs and basic research before application in the clinic.

## 5 | CONCLUSION

In summary, this MR study revealed that six genetically proxied blood metabolites have causal effects on CRC, and we identified nine metabolic pathways that could be implicated in the development of the CRC. LDHA deserve further investigation as potential therapeutic targets for CRC. The discovery of these serum metabolites provides valuable insights for early screening, prevention, and treatment of CRC, as well as for the design of future clinical studies. Moreover, this MR analysis combining genomics and metabolomics provides a reference direction for the exploration of the etiology and pathogenesis of CRC.

### AUTHOR CONTRIBUTIONS

**Zhangjun Yun:** Formal analysis (lead); methodology (lead); software (lead); validation (lead); writing – original draft (lead). **Ziwei Guo:** Resources (lead); software (lead); validation (lead); writing – original draft (lead). **Xiao Li:** Formal analysis (equal); validation (equal); visualization (equal). **Yang Shen:** Data curation (equal); investigation (equal). **Mengdie Nan:** Investigation (equal). **Qing Dong:** Conceptualization (equal); project administration

(equal); supervision (equal); writing – review and editing (equal). **Li Hou:** Conceptualization (lead); funding acquisition (lead); methodology (lead); project administration (lead); supervision (lead); writing – review and editing (lead).

## ACKNOWLEDGMENTS

We first appreciate the original GWAS data provided by Shin et al. and Huyghe et al. and the GWAS Catalog database. We are very grateful to smart-servier (<https://smart.servier.com/>) for the medical photo material.

## FUNDING INFORMATION

This study was supported by the National Natural Science Foundation of China Grant Program (Grant/Award Number: 81573959), Capital Health Development Research Fund (Grant/Award Number: 2020–2-4193 and 2022–1-4171).

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

We have annotated the article with the source of all original data, please contact the original authors for access if needed. The results of this study can be obtained by contacting the corresponding author.

## ETHICS STATEMENT

All data in this study are available in publicly available databases. No additional ethical approval was required.

## ORCID

Zhangjun Yun  <https://orcid.org/0000-0002-3422-9140>

Ziwei Guo  <https://orcid.org/0000-0001-7246-147X>

Li Hou  <https://orcid.org/0000-0003-2921-0821>

## REFERENCES

- 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):209-249.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394-424.
- Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. Smoking and colorectal cancer: a meta-analysis. *Jama.* 2008;300(23):2765-2778.
- Vieira AR, Abar L, Chan DSM, et al. Foods and beverages and colorectal cancer risk: a systematic review and meta-analysis of cohort studies, an update of the evidence of the WCRF-AICR continuous update project. *Ann Oncol.* 2017;28(8):1788-1802.
- Guraya SY. Association of type 2 diabetes mellitus and the risk of colorectal cancer: a meta-analysis and systematic review. *World J Gastroenterol.* 2015;21(19):6026-6031.
- Bull CJ, Bell JA, Murphy N, et al. Adiposity, metabolites, and colorectal cancer risk: Mendelian randomization study. *BMC Med.* 2020;18(1):396.
- Rodriguez-Broadbent H, Law PJ, Sud A, et al. Mendelian randomization implicates hyperlipidemia as a risk factor for colorectal cancer. *Int J Cancer.* 2017;140(12):2701-2708.
- Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol.* 2016;17(7):451-459.
- Lains I, Gantner M, Murinello S, et al. Metabolomics in the study of retinal health and disease. *Prog Retin Eye Res.* 2019;69:57-79.
- Ward PS, Thompson CB. Metabolic reprogramming: a cancer hallmark even Warburg did not anticipate. *Cancer Cell.* 2012;21(3):297-308.
- Dang CV. Links between metabolism and cancer. *Genes Dev.* 2012;26(9):877-890.
- Whitehouse S, Cooper RH, Randle PJ. Mechanism of activation of pyruvate dehydrogenase by dichloroacetate and other halogenated carboxylic acids. *Biochem J.* 1974;141(3):761-774.
- Christofk HR, Vander Heiden MG, Harris MH, et al. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature.* 2008;452(7184):230-233.
- Christofk HR, Vander Heiden MG, Wu N, Asara JM, Cantley LC. Pyruvate kinase M2 is a phosphotyrosine-binding protein. *Nature.* 2008;452(7184):181-186.
- Zhao Y, Butler EB, Tan M. Targeting cellular metabolism to improve cancer therapeutics. *Cell Death Dis.* 2013;4(3):e532.
- Zuccolo L, Holmes MV. Commentary: Mendelian randomization-inspired causal inference in the absence of genetic data. *Int J Epidemiol.* 2017;46(3):962-965.
- Richmond RC, Davey SG. Mendelian randomization: concepts and scope. *Cold Spring Harb Perspect Med.* 2022;12(1):a040501.
- Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol.* 2013;37(7):658-665.
- Boef AG, Dekkers OM, le Cessie S. Mendelian randomization studies: a review of the approaches used and the quality of reporting. *Int J Epidemiol.* 2015;44(2):496-511.
- Cai J, Li X, Wu S, et al. Assessing the causal association between human blood metabolites and the risk of epilepsy. *J Transl Med.* 2022;20(1):437.
- Shin SY, Fauman EB, Petersen AK, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet.* 2014;46(6):543-550.
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 2012;40(D1):D109-D114.
- Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47(D1):D1005-D1012.
- Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet.* 2019;51(1):76-87.

25. Yang J, Yan B, Zhao B, et al. Assessing the causal effects of human serum metabolites on 5 major psychiatric disorders. *Schizophr Bull.* 2020;46(4):804-813.
26. Choi KW, Chen CY, Stein MB, et al. Assessment of bidirectional relationships between physical activity and depression among adults: a 2-sample Mendelian randomization study. *JAMA Psychiat.* 2019;76(4):399-408.
27. Gill D, Brewer CF, Monori G, et al. Effects of genetically determined iron status on risk of venous thromboembolism and carotid atherosclerotic disease: a Mendelian randomization study. *J Am Heart Assoc.* 2019;8(15):e012994.
28. Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol.* 2013;178(7):1177-1184.
29. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-egger method. *Eur J Epidemiol.* 2017;32(5):377-389.
30. Cohen JF, Chalumeau M, Cohen R, Korevaar DA, Khoshnood B, Bossuyt PM. Cochran's Q test was useful to assess heterogeneity in likelihood ratios in studies of diagnostic accuracy. *J Clin Epidemiol.* 2015;68(3):299-306.
31. Bowden J, Spiller W, Del Greco MF, et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the radial plot and radial regression. *Int J Epidemiol.* 2018;47(4):1264-1278.
32. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet.* 2018;50(5):693-698.
33. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol.* 2013;42(5):1497-1501.
34. O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet.* 2018;50(12):1728-1734.
35. Reay WR, Kiltschewskij DJ, Geaghan MP, et al. Genetic estimates of correlation and causality between blood-based biomarkers and psychiatric disorders. *Sci Adv.* 2022;8(14):8969.
36. Hemani G, Tilling K, Davey SG. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* 2017;13(11):e1007081.
37. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014;10(5):e1004383.
38. Vosa U, Claringbould A, Westra HJ, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet.* 2021;53(9):1300-1310.
39. Ferkingstad E, Sulem P, Atlason BA, et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet.* 2021;53(12):1712-1721.
40. Foley CN, Staley JR, Breen PG, et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat Commun.* 2021;12(1):764.
41. Sanderson E. Multivariable Mendelian randomization and mediation. *Cold Spring Harb Perspect Med.* 2021;11(2):a038984.
42. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol.* 2015;181(4):251-260.
43. Jewison T, Su Y, Disfany FM, et al. SMPDB 2.0: big improvements to the small molecule pathway database. *Nucleic Acids Res.* 2014;42(D1):D478-D484.
44. Sharma D, Singh M, Rani R. Role of LDH in tumor glycolysis: regulation of LDHA by small molecules for cancer therapeutics. *Semin Cancer Biol.* 2022;87:184-195.
45. Paul SK, Dutta Chowdhury K, Dey SR, Paul A, Haldar R. Exploring the possibility of drug repurposing for cancer therapy targeting human lactate dehydrogenase a: a computational approach. *J Biomol Struct Dyn.* 2022;1-10:1-10.
46. Rappaport SM, Barupal DK, Wishart D, Vineis P, Scalbert A. The blood exposome and its role in discovering causes of disease. *Environ Health Perspect.* 2014;122(8):769-774.
47. Platten M, Wick W, Van den Eynde BJ. Tryptophan catabolism in cancer: beyond IDO and tryptophan depletion. *Cancer Res.* 2012;72(21):5435-5440.
48. Zhai L, Spranger S, Binder DC, et al. Molecular pathways: targeting IDO1 and other tryptophan dioxygenases for cancer immunotherapy. *Clin Cancer Res.* 2015;21(24):5427-5433.
49. Santhanam S, Alvarado DM, Ciorba MA. Therapeutic targeting of inflammation and tryptophan metabolism in colon and gastrointestinal cancer. *Transl Res.* 2016;167(1):67-79.
50. Gray LR, Tompkins SC, Taylor EB. Regulation of pyruvate metabolism and human disease. *Cell Mol Life Sci.* 2014;71(14):2577-2604.
51. Bricker DK, Taylor EB, Schell JC, et al. A mitochondrial pyruvate carrier required for pyruvate uptake in yeast, drosophila, and humans. *Science.* 2012;337(6090):96-100.
52. Warburg O, Wind F, Negelein E. The metabolism of tumors in the body. *J Gen Physiol.* 1927;8(6):519-530.
53. Schell JC, Rutter J. The long and winding road to the mitochondrial pyruvate carrier. *Cancer Metab.* 2013;1(1):6.
54. Bensard CL, Wisidagama DR, Olson KA, et al. Regulation of tumor initiation by the mitochondrial pyruvate carrier. *Cell Metab.* 2020;31(2):284-300 e287.
55. Compan V, Pierredon S, Vanderperre B, et al. Monitoring mitochondrial pyruvate carrier activity in real time using a BRET-based biosensor: investigation of the Warburg effect. *Mol Cell.* 2015;59(3):491-501.
56. Ramos-Ibeas P, Barandalla M, Colleoni S, Lazzari G. Pyruvate antioxidant roles in human fibroblasts and embryonic stem cells. *Mol Cell Biochem.* 2017;429(1-2):137-150.
57. Das UN. Pyruvate is an endogenous anti-inflammatory and anti-oxidant molecule. *Med Sci Monit.* 2006;12(5):RA79-RA84.
58. Kang HJ, Min BK, Choi WI, et al. Pyruvate dehydrogenase kinase 1 and 2 deficiency reduces high-fat diet-induced hypertrophic obesity and inhibits the differentiation of preadipocytes into mature adipocytes. *Exp Mol Med.* 2021;53(9):1390-1401.
59. Nellesmann B, Gormsen LC, Sorensen LP, Christiansen JS, Nielsen S. Impaired insulin-mediated antilipolysis and lactate release in adipose tissue of upper-body obese women. *Obesity (Silver Spring).* 2012;20(1):57-64.
60. Mika A, Pakiet A, Czumaj A, et al. Decreased triacylglycerol content and elevated contents of cell membrane lipids in colorectal cancer tissue: a lipidomic study. *J Clin Med.* 2020;9(4):1095.
61. Bernard L, Zhou L, Surapaneni A, et al. Serum metabolites and kidney outcomes: the Atherosclerosis Risk in Communities study. *Kidney Med.* 2022;4(9):100522.

62. Calzada E, Onguka O, Claypool SM. Phosphatidylethanolamine metabolism in health and disease. *Int Rev Cell Mol Biol.* 2016;321:29-88.
63. Stoica C, Ferreira AK, Hannan K, Bakovic M. Bilayer forming phospholipids as targets for cancer therapy. *Int J Mol Sci.* 2022;23(9):5266.
64. Saxena R, Yang C, Rao M, et al. Preclinical development of a nontoxic oral formulation of monoethanolamine, a lipid precursor, for prostate cancer treatment. *Clin Cancer Res.* 2017;23(14):3781-3793.
65. Xia H, Chen J, Sekar K, Shi M, Xie T, Hui KM. Clinical and metabolomics analysis of hepatocellular carcinoma patients with diabetes mellitus. *Metabolomics.* 2019;15(12):156.
66. Huang J, Mondul AM, Weinstein SJ, et al. Prospective serum metabolomic profiling of lethal prostate cancer. *Int J Cancer.* 2019;145(12):3231-3243.
67. Gromov P, Gromova I, Friis E, et al. Proteomic profiling of mammary carcinomas identifies C7orf24, a gamma-glutamyl cyclotransferase, as a potential cancer biomarker. *J Proteome Res.* 2010;9(8):3941-3953.
68. Huang Q, Zhou Y, Li Y, Liao Z. GGCT promotes colorectal cancer migration and invasion via epithelial-mesenchymal transition. *Oncol Lett.* 2020;20(2):1063-1070.
69. Ying L, Ma X, Shen Y, et al. Serum 1,5-anhydroglucitol to glycated albumin ratio can help early distinguish fulminant type 1 diabetes mellitus from newly onset type 1A diabetes mellitus. *J Diabetes Res.* 2020;2020:1243630-1243638.
70. Cai H, Li J, Zhang Y, et al. LDHA promotes oral squamous cell carcinoma progression through facilitating glycolysis and epithelial-mesenchymal transition. *Front Oncol.* 2019;9:1446.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Yun Z, Guo Z, Li X, et al. Genetically predicted 486 blood metabolites in relation to risk of colorectal cancer: A Mendelian randomization study. *Cancer Med.* 2023;12:13784-13799. doi:[10.1002/cam4.6022](https://doi.org/10.1002/cam4.6022)