EPIGENETICS
2025, VOL. 20, NO. 1, 2599682
https://doi.org/10.1080/15592294.2025.2599682

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

# Exploring the association between DNA methylation and pancreatic cancer susceptibility through epigenome-wide Mendelian randomization and multi-omics data integration

Pengxu Wang [a,b], Feng Rong [b] and Fubao Liu [a]

[a]Department of General Surgery, The First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China; [b]Department of Oncology Center, Lu'an Affiliated Hospital of Anhui Medical University, Lu'an, Anhui, China

**ABSTRACT**

Investigating the role of DNA methylation in the development of pancreatic cancer (PC) may facilitate identification of potential targets for both diagnosis and treatment. We carried out a comprehensive epigenome-wide Mendelian randomization (EWMR) analysis to investigate the correlation of genetically predicted blood CpG sites with PC. Following this, we conducted various sensitivity analyses and repeated analyses using different selection criteria for instrumental variables and conditional Bayesian colocalization to guarantee the reliability of the results. External validation and a meta-analysis were then performed to further validate these results. Next, we conducted CpG site enrichment analysis, overlap with phenome-wide association studies (PheWAS) catalog analysis, overlap with epigenome-wide association studies (EWAS) Toolkit analysis, and drug target analysis to explore the enrichment, biological functions, and potential therapeutic targets associated with these sites. Finally, we used the SMR-IVW software to perform mediation analysis, aiming to uncover potential tumorigenesis pathways of PC at the transcriptional level from three distinct perspectives. Results showed 253 CpG sites passing sensitivity analysis were significantly associated with PC and 159 CpG sites were validated in at least one replication. After meta-analysis, 38 CpG sites were retained, and all 253 CpG sites were classified into three tiers. Among these, cg26373071 (CLPTM1L), cg14271713, cg11652496 (PSTPIP1), and cg20575191 (PSTPIP1) were placed in tier 1 with strong support. Finally, this study identified genetic susceptibility linked to 253 PC-related CpG sites. This study provides insights into the disease's origins and underscores potential targets for future research.

**ARTICLE HISTORY**
Received 20 August 2025
Revised 3 November 2025
Accepted 1 December 2025

**KEYWORDS**
Pancreatic cancer; DNA methylation; Mendelian randomization; risk factor

## Introduction

Pancreatic cancer (PC) is essentially one of the leading causes of tumor-related deaths, with an increasing incidence and a 5-year survival rate of approximately 13%, which drops to around 3% in metastatic cases [1]. The poor survival outcomes are primarily attributed to the tumor's aggressive behavior and the absence of early clinical symptoms, often resulting in delayed diagnosis [2]. Approximately 80% of patients receive a diagnosis at an advanced, inoperable stage, leading to a median overall survival (OS) of roughly 4 months regardless of the stage at diagnosis [3]. However, even in cases of resectable pancreatic cancer, outcomes remain poor as postoperative recurrence and chemotherapy resistance are frequently observed [4,5]. Therefore, uncovering biomarkers with a causal link to PC risk is crucial, as they may provide important insights into the disease's pathogenesis and contribute to improved risk assessment and the development of new treatment strategies.

DNA methylation (DNAm) primarily occurs at cytosine residues within dinucleotide sequences, especially 5-methylcytosine (5mC) in mammals [6,7]. In large-scale genetic studies, blood samples are frequently used to assess DNA methylation because they are easily accessible, minimally invasive to collect, and blood-based DNAm plays a significant role in inter-tissue communication. Methylation levels at specific

**CONTACT** Feng Rong wazhl1996@163.com Department of Oncology Center, Lu'an Affiliated Hospital of Anhui Medical University, 21 Wanxi West Road, Lu'an, Anhui 237005, P.R. China; Fubao Liu lancetlfb@126.com Department of General Surgery, The First Affiliated Hospital of Anhui Medical University, 218 Jixi Road, Hefei, Anhui 230022, P.R. China

Supplemental data for this article can be accessed online at https://doi.org/10.1080/15592294.2025.2599682

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

CpG sites have been strongly associated with various complex traits and phenotypic characteristics [8–11]. Furthermore, several case–control studies of PC have investigated differences in DNAm in blood leukocytes between patients and healthy individuals, identifying numerous CpG sites associated with the disease [12–14]. These studies suggest that DNAm biomarkers may enhance the understanding of the epigenetic processes and molecular mechanisms associated with PC. However, the observational nature of these studies weakens the robustness of their findings, as they may be influenced by possible reverse causality and confounding factors [15].

Recent studies on large-scale genetic effects on DNAm have broadened the understanding of DNAm quantitative trait loci (meQTL) [16–18], enabling the inference of causal links between blood DNAm sites and the risk of PC. This can be achieved using two-sample Mendelian Randomization (MR), wherein genetic variants or single-nucleotide polymorphisms (SNP) related to exposure were employed as instrumental variables (IVs) to mimic the design of randomized controlled trials [19–21]. MR has gained popularity as a method for investigating causal relationships in disease outcomes due to its reduced vulnerability to confounding factors and reverse causality. For instance, it can be used to evaluate the role of peripheral metabolites, circulating proteins, and gut microbiota in the development of PC [22–24]. Additionally, as an extension of the MR framework, EWMR studies explore the role of genetically predicted blood CpG sites in disease development. Earlier research has effectively utilized EWMR with meQTLs to identify CpG sites causally associated with various diseases or phenotypes, such as aging [25,26]. Therefore, integrating meQTL data with genome-wide association studies (GWAS) for PC allows evaluation of the causal effect of molecular changes on the development of PC by using two-sample MR.

In this study, we employed cis-meQTL data to perform an EWMR analysis on 420,509 CpG sites, aiming to identify blood-based CpG sites linked to PC. To verify our key findings, we conducted sensitivity analyses, conditional Bayesian colocalization, replication studies, and meta-analysis. We then carried out chromatin state, functional, and pathway enrichment analyses for the CpG sites which were identified through EWMR as being associated with PC and validated in at least one replication phase. Additionally, overlap with PheWAS catalog analysis and overlap with EWAS Toolkit analysis were employed to explore whether these CpG sites were linked to other diseases or phenotypes. We also evaluated functional and clinical significance of the identified CpG sites by using druggable gene target analysis. Finally, we used multivariable Mendelian Randomization (MVMR) to conduct mediation analysis, incorporating multiple dimensions of expression quantitative trait loci (eQTLs).

## Methods

### *Study overall design*

Figures 1 and 2 show the whole workflow a of this study. In summary, more than 420 thousand CpG sites data derived from Genetics of DNA Methylation Consortium (GoDMC) were employed for a two-sample EWMR in the primary discovery phase. Then, to refine and broaden our initial findings, we implemented a four-step analytical approach. In the first step, sensitivity analyses (e.g., heterogeneity, pleiotropy, and steiger directionality), along with repetitions with varying IVs selection criteria and conditional Bayesian colocalization analysis, were executed. In the second step, external validation was used to confirm the primary EWMR results, and meta-analysis was executed to integrate the experimental and validation results. In the third step, we performed CpG site enrichment analysis, overlap with PheWAS catalog analysis, overlap with EWAS Toolkit analysis, functional and pathway enrichment analysis, and drug target analysis to explore the enrichment, functions, and therapeutic relevance of the PC-associated CpG sites which were identified through EWMR and validated in at least one replication phase. In the final step, we applied mediation analysis using SMR-IVW software to infer potential tumorigenesis pathways at the transcriptional level for the PC-associated CpG sites identified through EWMR and validated in at least one replication phase. P-values were tuned by using the Bonferroni or Benjamini–Hochberg false discovery rate (FDR) method to correct for multiple tests in our study. All analyses followed the STROBE-MR guidelines

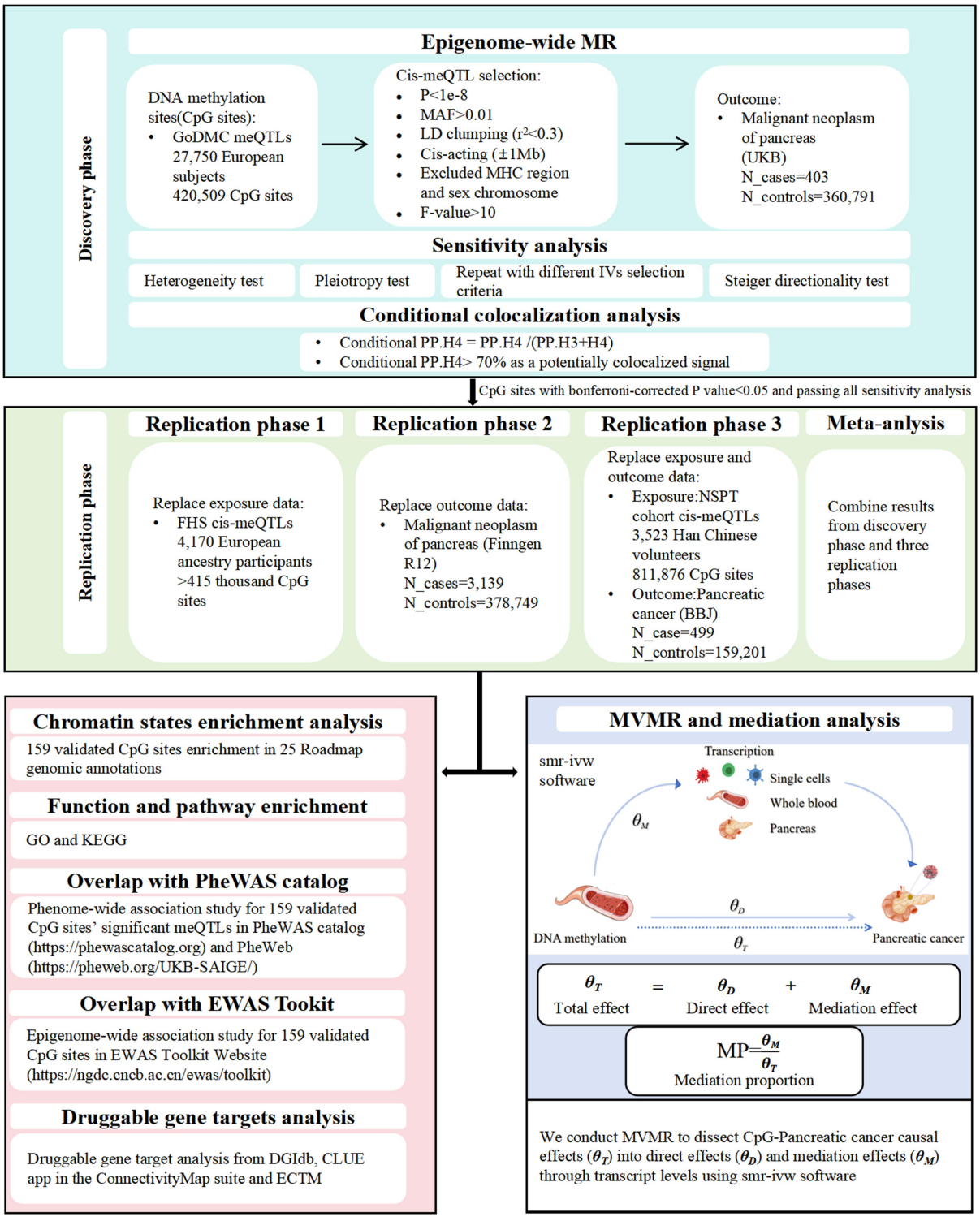



**Figure 1.** Study design flowchart. Flowchart of the study design. MR, Mendelian randomization; CpG sites, Cytosine-phosphate-Guanine sites; GoDMC, Genetics of DNA Methylation Consortium; meQTL, DNA methylation quantitative trait locus; MAF, Minor Allele Frequency; LD, Linkage Disequilibrium; MHC, Major Histocompatibility Complex; UKB, UK Biobank; IV, instrumental variable; N_cases, Number of cases; N_controls, Number of controls; FHS, Framingham Heart Study; NSPT, National Survey of Physical Traits; BBJ, Biobank Japan; EWAS, Epigenome-wide Association Study, PheWAS, Phenome-wide association study; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; DGIdb, Drug Gene Interaction Database; ECTM, Encyclopedia of Traditional Chinese Medicine; MVMR, Multivariable MR.

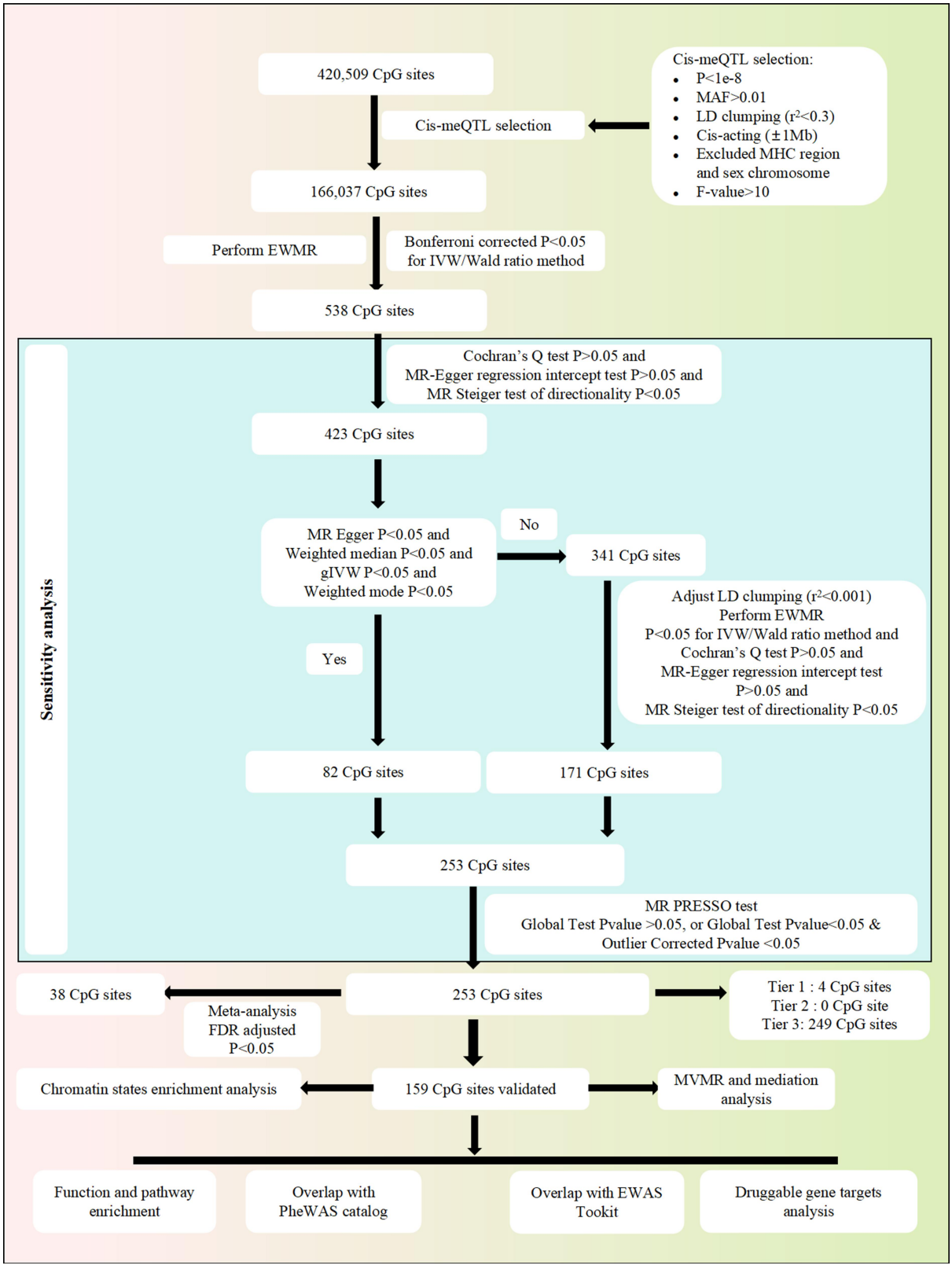



**Figure 2.** CpG sites selection flowchart. Flowchart of final putative CpG sites selection EWMR, Epigenome-wide Mendelian randomization; cis-meQTL, cis-acting DNA methylation quantitative trait loci; MAF, Minor allele frequency; LD, Linkage disequilibrium; MHC, Major histocompatibility complex; IVW, inverse variance-weighted.

[27]. The data of our analysis was derived from public GWAS summary statistics. This study involved no newly collected data, and no new ethical approval was required.

### Datasets and genetic instrument

In the primary discovery phase, summary statistics from the GWAS of PC were downloaded from the UK Biobank (https://www.nealelab.is/uk-biobank), wherein the data of over 360,000 participants were involved. In this study, PC was defined as malignant pancreas tumor, and pancreatic ductal adenocarcinoma is a typical one. Statistics of genetic associations with blood-based DNAm were sourced from the GoDMC, which integrates 36 cohorts and provides data on DNAm quantitative trait loci (meQTLs) associated with 420,509 CpG sites in whole blood from 27,750 European participants (http://mqtldb.godmc.org.uk) [17]. We applied the following criteria to filter the extracted meQTLs: a. meQTLs that exhibited a genome-wide significant association ($p < 1e\text{-}8$) with any CpG sites; b. meQTLs with a minor allele frequency (MAF) >0.01; c. a control linkage disequilibrium (LD) $r^2$ of <0.3 within a 10 Mb window; d. cis-acting meQTLs defined as those with an SNP-CpG distance of less than 1 MB; e. sex chromosomes and meQTLs located inside the major histocompatibility complex (MHC) region (Chromosome 6, 31–33 Mb) were excluded; f. to remove weak IVs, F-statistics were determined by $F=\frac{R^2\times(N-2)}{(1-R^2)}$ and $R^2=\frac{2\times\beta^2\times EAF\times(1-EAF)}{2\times\beta^2\times EAF\times(1-EAF)+2\times SE^2\times N\times EAF\times(1-EAF)}$, where $R^2$ denotes the proportion of CpG site variability explained by each instrument, SE denotes standard error, N denotes the quantity of participants, EAF denotes the effect allele frequency, and β denotes the estimated effect of the SNP, which allows us to assess whether the SNP can reliably predict the result [28, 29]. meQTLs with F-statistics below 10 were excluded. g. cis-meQTLs were not directly associated with PC, meaning that the P-value of putative SNPs in the GWAS of PC was greater than 1e-5. The participants in both GWAS studies were of European ancestry and all had received the appropriate ethical approval for their involvement.

### EWMR and sensitivity analysis

All data analysis was achieved by using R (version 4.4.1), and the EWMR analysis was achieved by using the 'TwoSampleMR (version 0.6.6)' package [30] and 'MendelianRandomization (version 0.10.0)' package [31]. Herein, the Wald ratio and inverse variance-weighted (IVW) methods were employed as the primary MR methods, as they are the most effective when all IVs are valid [32]. To explore the causal effects of genetic variants, we used the Wald ratio method when one IV was available only. For analyses involving multiple IVs, we employed the IVW method, which combines the individual SNP ratio estimates by meta-analysis. Before the MR analysis, the exposure and outcome datasets were aligned by utilizing the 'harmonise_data' in the 'TwoSampleMR' package. In this way, only SNPs common to both exposure and outcome datasets were included, while automatically excluding SNPs with strand ambiguity or palindromic sequences. Finally, we adjusted the P-values from the EWMR analysis with the Bonferroni correction, considering associations significant if their Bonferroni corrected $p < 0.05$. To assess heterogeneity across instruments, we applied Cochran's Q statistic, with $p < 0.05$ indicating significant heterogeneity. Additionally, to detect horizontal pleiotropy, we conducted the MR-Egger regression intercept analysis, applied the MR-PRESSO method for pleiotropy detection, and performed the MR-PRESSO global test [33,34]. As part of the sensitivity analysis, four additional MR methods, including weighted median, MR-Egger, weighted mode, and generalized inverse variance weighted (gIVW), were employed. The gIVW result was conducted using 'MendelianRandomization' R package. This method incorporated genetic correlation into IVW model. In order to account for potential bias in IV selection, we repeated the analysis using different IV inclusion thresholds, modifying the LD $r^2$ cutoffs to 0.001, 0.01, 0.1, and 0.2, respectively. To rule out potential MR signals arising from reverse causation, where changes in DNAm could be a result of the outcome phenotype (e.g., PC), we conducted the MR Steiger directionality test [35], which evaluates the inferred causal direction in MR analysis. Any signals suggesting a reversed causal relationship were excluded from further analysis. In general, we considered CpG sites with Bonferroni corrected $p < 0.05$ and those passing all sensitivity analysis in the discovery phase as the EWMR-identified PC-associated CpG sites. The detailed steps of

the sensitivity analyses and how they fed into the final set of selected CpG sites are shown in Figure 2. In addition, we performed a Pearson correlation analysis of the effect estimates (β) between the IVW method and each of four additional MR methods to re-examine the sensitivity of the EWMR-identified, PC-associated CpG sites.

### Conditional Bayesian colocalization

Genetic colocalization employs a Bayesian framework to assess the probability (PP.H4) that molecular traits and the outcome share a common causal variant, suggesting overlapping genetic signals [36]. This method is essential for reducing false positives in MR analysis by detecting and excluding associations driven only by LD or horizontal pleiotropy. All PC-associated CpG sites identified by EWMR were subsequently examined using Bayesian colocalization methods. In this step, the 'coloc' R package was used to investigate whether specific CpG sites and PC could share a common causal variant [37]. At each locus, we evaluated the posterior probabilities for five hypotheses: (PP.H0) no association with either the CpG site or PC; (PP.H1) association with the CpG site only; (PP.H2) association with PC only; (PP.H3) associations with both the CpG site and PC driven by distinct causal variants; and (PP.H4) associations with both the CpG site and PC attributable to a shared causal variant. A colocalization probability (PP.H4) greater than 70% was considered strong evidence for colocalization. Since GWAS data related to PC tend to be noisy, whereas cis-meQTLs generally display robust genetic signals, the overall colocalization probability (PP.H4) might be underestimated, and the probability that a signal is detected solely from the meQTL (PP.H1) could be overestimated. To mitigate this potential bias arising from differing statistical power between the exposure and outcome traits, we calculated a conditional colocalization probability defined as conditional PP.H4 = PP.H4/(PP.H3 + PP.H4), with the assumption that the PC contains a genetic signal in the region when a significant MR association is observed. CpG–PC pairs with a conditional PP.H4 above 70% were also considered indicative of potential colocalization [26].

### Validation analyses

We validated our findings from the primary discovery phase using a three-phase validation approach. In the first replication phase, we substituted the exposure data with cis-meQTLs from the Framingham Heart Study (FHS), which included over 415,000 DNAm CpG sites in whole blood samples from 4170 European ancestry individuals (https://ftp.ncbi.nlm.nih.gov/eqtl/original_submissions/FHS_meQTLs/) [16]. In the second replication phase, we used GWAS data for PC from the R12 release of the FinnGen consortium (https://www.finngen.fi/en) as the outcome replacement [38], which enrolled more than 500 thousand European individuals. These datasets had adequate sample sizes and broad SNP coverage. All studies above were conducted with European populations, and samples of exposure–outcome pairs were not overlapped, as each was derived from distinct research cohorts. As a result, the selection of these datasets adhered to the assumptions necessary for valid MR analysis and ensured their appropriateness for replication. The filtering process for cis-meQTLs was consistent with the discovery phase. Standard sensitivity analyses were also performed. Furthermore, in the third replication, the GWAS data for both exposure and outcome were replaced by GWAS data from East Asians. For this phase, cis-meQTLs were sourced from the National Survey of Physical Traits (NSPT) cohort (https://www.biosino.org/node/project/detail/OEP002902), which included 3523 Han Chinese participants and 811,876 CpG sites in whole blood [18]. Meanwhile, the GWAS data for PC were obtained from Biobank Japan (BBJ) (https://pheweb.jp/), with 499 cases and 159,201 controls [39]. Additionally, to expand the scope of our analysis, we explored CpG methylation sites that did not demonstrate significant associations with PC in the discovery phase by utilizing alternative GWAS datasets. Finally, a meta-analysis was conducted by utilizing the 'metagen' function from the 'meta' R package to integrate the MR estimates with the discovery phase and the three replication phases, considering the aggregated results as the final evidence for external validation. To determine whether to use a fixed-effect or random-effect model, we assessed heterogeneity across the studies using the $I^2$ statistic [40]. When $I^2 < 50\%$, we selected the fixed-effect model; when $I^2 \geq 50\%$, we chose the random-effect model [41]. Then, EWMR-identified PC-associated CpG sites were categorized into three tiers on the basis of the strength of evidence from colocalization and replication analyses: a. Tier 1 CpG sites: those with FDR-

adjusted $p < 0.05$ in the meta-analysis validated in at least one of the three replication phases, exhibiting consistent effect directions across both the discovery and validation phases and having a conditional colocalization PP.H4 greater than 0.7; b. Tier 2 CpG sites: those that met the same criteria as Tier 1 but had a conditional colocalization PP.H4 below 0.7; c. Tier 3 CpG sites: those that did not replicate in the external validation phase.

### Chromatin states enrichment analysis

To gain deeper insight into the characteristics of CpG sites identified as causally linked to PC, we conducted a chromatin state enrichment analysis for the identified CpG sites based on the 25 Roadmap epigenomic annotations [42]. We conducted enrichment analysis on the CpG sites identified by EWMR, which had P-values under 0.05 and passed all sensitivity tests during the discovery phase as well as the three replication phases. These were categorized into GoDMC-UKB pairs, GoDMC-Finngen pairs, FHS-UKB pairs, and NSPT-BBJ pairs. Furthermore, enrichment analysis was performed on validated EWMR-identified PC-associated CpG sites, referred to as GoDMC-UKB-Bonferroni-adjusted-validated pairs. The enrichment analysis was achieved by the Fisher's exact test, with results considered significant if the FDR-adjusted $p < 0.05$.

### Function and pathway enrichment

To explore the biological relevance of genes associated with the validated EWMR-identified PC-related CpG sites, we utilized the 'IlluminaHumanMethylation450kanno.ilmn12.hg19' R package for mapping of gene symbols to these CpG sites. Additionally, enrichment analysis was achieved by utilizing the 'clusterProfiler' R package, focusing on gene ontology (GO) terms (e.g., biological processes, molecular functions, and cellular components) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [43,44]. This enrichment analysis evaluated whether certain genes were disproportionately represented in specific pathways compared to the expected random distribution, based on the ratio of background genes related to each pathway. Pathways or functional categories with FDR-adjusted $p < 0.05$ were considered significantly enriched.

### Overlap with PheWAS catalog analysis

To investigate potential links between the validated EWMR-identified PC-associated CpG sites and different disease phenotypes, we analyzed the PheWAS catalog (https://phewascatalog.org), which includes data on 1358 phenotypes from the 13,835 European individuals participated in the eMERGE Network [45]. The PheWAS catalog was searched for meQTLs associated with validated EWMR-identified PC-related CpG sites, which were sourced from the GoDMC meQTL database. The analysis was limited to cis-acting meQTLs (with an SNP-CpG distance of <1 MB) that met the significance threshold of $p < 1e\text{-}8$ in the GoDMC database. Additionally, these meQTLs were further filtered to include only those present in the GWAS Catalog [46]. SNPs were considered to be significant in the PheWAS catalog if they had a P-value below 3.7e-5 (0.05/1,358), applying a Bonferroni correction for the 1358 phenotypes tested. For those SNPs deemed significant, we conducted an additional PheWAS search using the UKBiobank ICD PheWeb (https://pheweb.org/UKB-SAIGE/) [47] to validate the results from the initial PheWAS analysis.

### Overlap with EWAS Toolkit analysis

To further investigate whether the validated EWMR-identified PC-associated CpG sites are linked to major diseases or phenotypes, we mined the EWAS Toolkit, a collection of one-step online analysis tools, to identify CpG sites related to traits in the EWAS Toolkit (https://ngdc.cncb.ac.cn/ewas/toolkit) [48]. We overlapped the validated EWMR-identified PC-associated CpG sites with those reported in the EWAS Toolkit and used Fisher's exact test to determine whether the validated EWMR-identified CpG sites were enriched among those linked to disease phenotypes in the EWAS Toolkit. FDR-adjusted $p < 0.05$ was considered to be statistically significant.

### Druggable gene target analysis

To identify clinically relevant druggable targets, we examined both approved and investigational drugs that interact with genes uniquely mapped to the validated EWMR-identified PC-associated CpG sites (DGIdb, https://dgidb.org/) [49]. We also queried the CLUE app (https://clue.io/) in the Connectivity Map suite, developed by the Broad Institute, to cross-reference and validate our findings for the same gene set [50]. Additionally, as Traditional Chinese Medicine plays a significant role in modern drug research and development, we explored the herb targets and formula targets of those genes through the Encyclopedia of Traditional Chinese Medicine (ETCM, http://www.tcmip.cn/ETCM/) [51].

### MVMR and mediation analysis

Recognizing the complexity of malignant tumor development and the plausibility that DNAm exerts primarily indirect effects through transcriptional regulation, we applied MVMR to apportion the observed association ($\theta_T$) between validated EWMR-identified PC-associated CpG sites and PC into direct ($\theta_D$) and gene expression–mediated indirect ($\theta_M$) components ($\theta_T = \theta_D + \theta_M$). The mediation proportion (MP) was calculated using the formula: $MP = \frac{\theta_M}{\theta_T}$. This approach followed a three-sample MVMR framework, utilizing a SMR-IVW tool (https://github.com/masadler/smrivw) [52]. By combining molecular QTL data with GWAS statistics, the SMR-IVW tool implements an MVMR approach to quantify how DNAm influences complex traits through gene expression. The MeQTLs for whole blood were obtained from the GoDMC consortium, while the GWAS data for PC in whole blood were retrieved from the UK Biobank. Both datasets in summary statistics-based Mendelian randomisation (SMR) format were generated from the original GWAS summary data using SMR 1.3.1 [53]. To investigate the mediation of gene expression, we considered three distinct dimensions. In the first dimension (whole-blood DNAm → whole-blood transcripts in cis → PC), we sourced the full whole-blood cis-eQTL data (16,987 genes in 31,684 mostly healthy individuals) in SMR format from the eQTLGen Consortium (https://www.eqtlgen.org/cis-eqtls.html) [54]. In the second dimension (whole-blood DNAm → blood single-cell transcripts in cis → PC), we obtained summary statistics for single-cell cis-eQTLs (https://zenodo.org/records/14968281) generated by the jaxQTL model [55] from the OneK1K cohort. This dataset includes scRNA-seq profiles from 1.27 million PBMCs collected from 982 individuals [56]. In the third dimension (whole-blood DNAm → pancreas transcripts in cis → PC), we used cis-eQTL data in pancreas tissue from the V8 release of the Genotype-Tissue Expression (GTEx) Consortium (https://yanglab.westlake.edu.cn/software/smr/#DataResource). The GTEx project provides RNA sequencing and genotype data from whole-genome sequencing (WGS) across 15,201 samples representing 49 tissue types collected from 838 postmortem donors [57]. Transcript mediators were selected based on their proximity to the corresponding CpG site, within 500 kb (i.e., in cis), and their potential causal relationship with the CpG. A causal relationship was confirmed if univariable MR analysis showed a significant CpG effect on the transcript, with MR $p < 0.05$. LD was set to $r^2 = 0.3$, consistent with the EWMR process. Other parameters used in the mediation analysis with the SMR-IVW tool were kept at their default settings. Finally, we presented results for both the top mediator and all the selected mediators in cis for each CpG site.

## Results

### EWMR identified 253 PC-associated CpG sites in the discovery phase

SNPs with $p > 1e\text{-}8$ and/or meQTLs on the sex chromosomes in the MHC region were excluded in the discovery phase. After applying this filter, 167,823 CpG sites and 51,637,120 cis-meQTLs remained. We then performed LD clumping and removed IVs with F-statistic <10 or MAF < 0.01, resulting in 167,760 CpG sites with 1,939,516 SNPs for harmonization. After harmonizing and excluding SNPs directly linked to PC, we retained 1,912,674 SNPs as valid IVs for 166,037 CpG sites in the primary MR analysis. 1–254 cis-meQTLs served as instruments for each CpG site. Comprehensive results can be found in Table S2.

After correcting for multiple testing with Bonferroni-corrected $p < 0.05$, 538 CpG sites showed a statistically significant association with PC. Upon conducting heterogeneity, horizontal pleiotropy analysis and Steiger directionality test, 423 of these CpG sites remained significant. Of the 423 CpG sites, 82 showed significant in all the four additional MR methods, including weighted median, gIVW, MR Egger, and weighted mode, 341 CpG sites hadn't show significant in either one of the four methods. After setting the LD $r^2$ threshold to 0.001 for the 341 CpG sites and performing EWMR, we identified 171 significant CpG sites. When combined with the 82 CpG sites that were significant across all four additional MR methods, this yielded a final set of 253 EWMR-identified, PC-associated CpG sites. Results from these analyses are shown in Figures 2-4 and Table S1.

To confirm our primary findings, the MR analysis was repeated on the 253 EWMR-identified PC-associated CpG sites using other four different LD clumping thresholds ($r^2 = 0.001$, 0.01, 0.1, and 0.2). In 1788 replicated tests, none showed an effect direction that contradicted the initial results. The number of confirmed CpG sites varied with the LD threshold, ranging from 196 (77.47%) at $r^2 = 0.01$ to 237 (93.68%) at $r^2 = 0.2$. Notably, 162 CpG sites (64.03%) were consistently replicated across all four LD clumping settings. These findings are depicted in Figure 5, with complete details provided in Table S9. In addition, significant correlations of the effect estimates between the IVW/Wald ratio and each of four additional MR methods confirm the stability of the putative EWMR-identified, PC-associated CpG sites (Supplementary Figure S1).

### Conditional colocalization analysis

To determine whether the detected association of genetically predicted CpG sites with PC is caused by a shared causal variant rather than by LD, we performed colocalization analysis. Among the 253 EWMR-identified PC-associated CpG sites, 21 (8.30%) CpG sites had PP.H4 exceeding 0.7; however, 195 (77.06%) CpG sites had conditional PP.H4 exceeding 0.7. The detailed results are presented in Figure 6 and Table S10. Notably, null results from colocalization do not necessarily weaken the conclusions drawn from MR analyses [58].

### Validation analysis

To verify the stability of the key associations found in the discovery phase, we performed replication analysis on the basis of additional datasets. As shown in Figure 5, when the cis-meQTL data were obtained from FHS and the selection criteria were applied (see methods section above) in Phase 1, we ultimately retained 119,806 CpG site-PC pairs for MR analysis (Table S4). After performing a sensitivity test, 113 (44.66% of the 253 EWMR-identified PC-associated CpG sites) CpG sites were successfully replicated in phase 1 (detailed results in Table S3).

When the GWAS data for PC were retrieved from FinnGen R12 and the selection criteria were applied in Phase 2, we obtained 164,957 CpG site-PC pairs for MR analysis (Table S6). After performing a sensitivity test, 74 (29.25% of the 253 EWMR-identified PC-associated CpG sites) CpG sites were successfully replicated in phase 2 (detailed results in Table S5).

When both cis-meQTL and GWAS data for PC were obtained from an East Asian ancestry cohort and the criteria were applied to select cis-meQTLs in Phase 3, we ultimately obtained 249,405 CpG site-PC pairs for MR analysis (Table S8). After performing a sensitivity test, 24 (9.49% of the 253 EWMR-identified PC-associated CpG sites) CpG sites were successfully replicated in phase 3 (detailed results in Table S7).

Notably, 159 CpG sites (62.85% of the EWMR-identified PC-associated CpG sites) were successfully replicated in at least one of the replication phases. A subsequent meta-analysis was performed to synthesize the MR estimates derived in the discovery phase and the replication phase. After aggregating effect sizes and adjusting P-values using the FDR approach, the results indicated that 38 CpG sites and PC had significant associations (Figure 2 and Table S11).

Based on the collective findings, the circulating CpG sites were categorized into three distinct tiers (refer to methods section for details). Of these, four CpG sites ranked in the top tier, as evidenced by both colocalization and replication results. Among the 4 tier 1 CpG sites, cg14271713 (intergenic, chr9:136150031-136151365, S shelf), cg11652496 (PSTPIP1, chr15:77324520-77324759, gene body, island), and cg20575191 (PSTPIP1, chr15:77324520-77324759, gene body, island) showed consistent causation

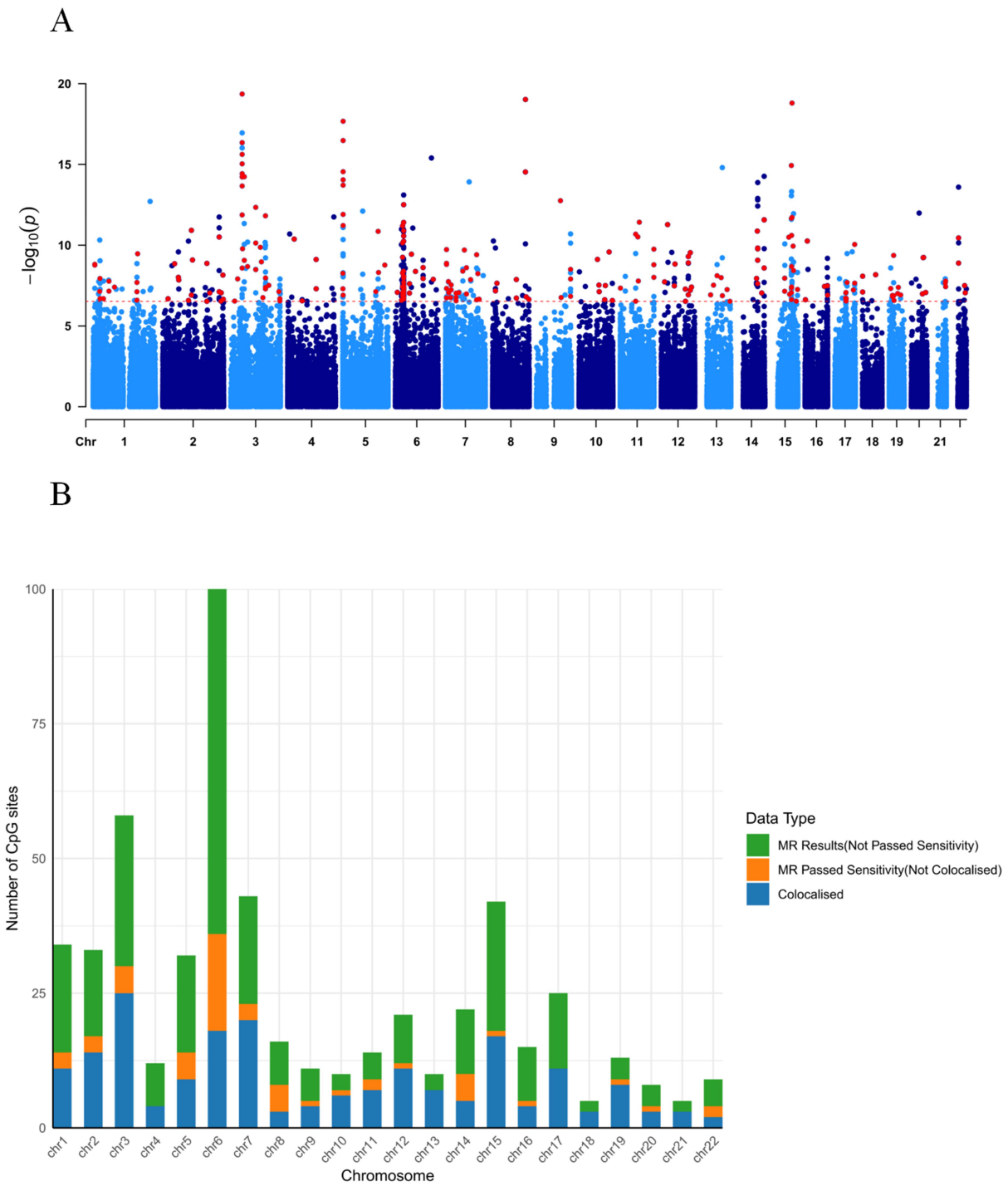



**Figure 3.** Main results of EWMR identified 538 PC-associated CpG sites in the discovery phase. Presentation of MR results for causal relationship between CpG sites and PC in discovery phase. A Manhattan plot showing the -log10(P) of IVW/Wald ratio method for each CpG site on PC plotted on the y-axis against genomic positionon on the x-axis. Red dots represent the 253 CpG sites which had passed the sensitivity analysis. Red dotted line corresponds to the false discovery rate threshold of 5% after Bonferroni correction. PC, Pancreatic cancer; MR, Mendelian randomization; IVW, Inverse variance-weighted; B Stacked colunm chart illustrating the number of MR-identified 538 CpG sites.

direction across the discovery phase, replication phase 1, and replication phase 2. Cg26373071 (CLPTM1L, gene body) showed consistent causation direction between the discovery phase and replication phase 1.

Notably, cg26373071 (CLPTM1L) exhibited not only a conditional colocalization PP.H4 above 0.7 but also a colocalization PP.H4 above 0.7, providing strong evidence for its association with PC. For CpG sites




**Figure 4.** Volcano plot of EWMR identified 538 PC-associated CpG sites in the discovery phase. Volcano plot showing MR results for causal relationship between CpG sites and PC in discovery phase. The y-axis represents the -log10(P) of IVW/Wald ratio method, and the x-axis indicates the effect size (beta) of the MR estimates. Red dots denote significant CpG sites that passed the sensitivity analysis with positive effect estimates (beta >0), while green dots represent significant CpG sites that passed the sensitivity analysis with negative effect estimates (beta <0). The red dotted line marks the false discovery rate threshold of 5% after Bonferroni correction. The top 10 CpG-gene pairs with the smallest (most negative) and largest (most positive) beta values are annotated on the plot.

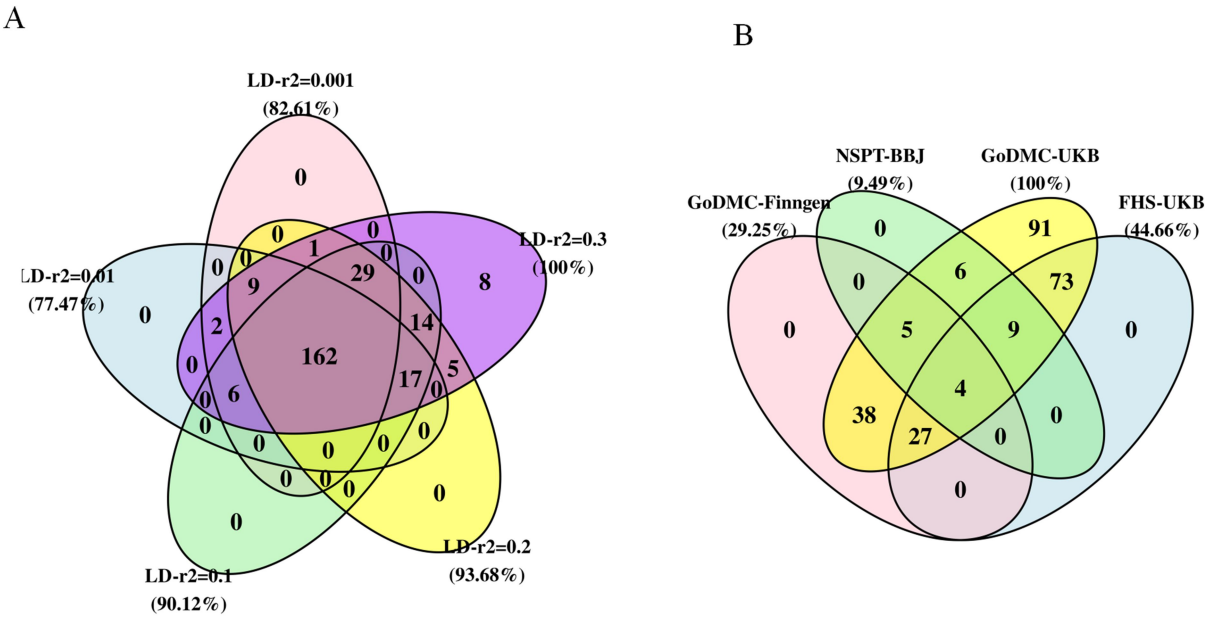



**Figure 5.** Validation results. Venn plots showing proportion of 253 CpG sites that EWMR identified and passed sensitivity analysis. A validated with different LD criteria and B validated in different replication phases.

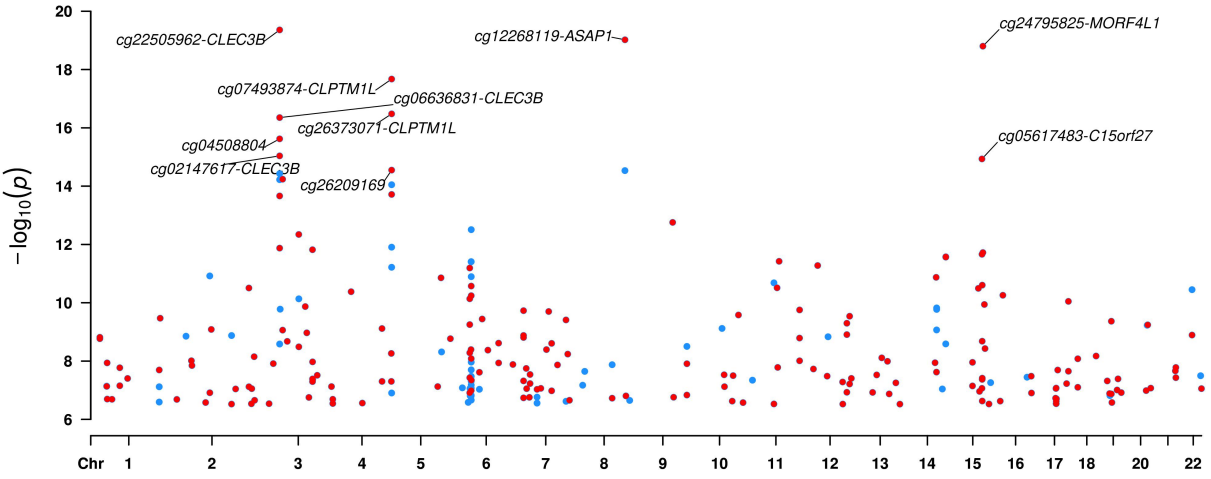



**Figure 6.** Conditional colocalization analysis results. Manhattan plot showing conditional Bayesian colocalization signals for the 253 CpG sites with colored dots that EWMR identified and passed the sensitivity analysis. Red dots correspond to 195 CpG sites that showed colocalized signals. The top 10 colocalized CpG sites with smallest *p* value were annotated with the CpG site names and corresponding UCSC RefGene Name.

with tier 2 evidence, there were no evidence supporting that. The remaining 249 CpG sites were classified as tier 3 (Figure 2 and Table S11).

The meta-analysis results and categorization details for the EWMR-identified PC-associated CpG sites are outlined in Table S11.

### *Chromatin states, function, and pathway enrichment*

To explore the characteristics of the 159 validated EWMR-identified CpG sites associated with PC, we executed an enrichment analysis by utilizing online datasets (https://egg2.wustl.edu/roadmap/web_portal/) [41]. The analysis results revealed that the likely causal CpG sites for PC were primarily found in bivalent promoter (PromBiv), promoter upstream of TSS (PromU), and repressed polycomb (ReprPC) regions. In contrast, these sites were underrepresented in the promoter downstream of TSS 1 (PromD1) quiescent genomic regions (Supplementary Figure S2). To investigate the potential involvement of genes related to these identified CpG sites in specific biological pathways, we executed functional enrichment analysis by utilizing GO terms and KEGG pathways. The most significantly enriched terms included calcium channel activity and calcium ion transmembrane transporter activity (molecular function) (Supplementary Figure S3).

### *Overlap with PheWAS catalog analysis*

Among the 159 validated CpG sites identified by EWMR as being associated with PC, we identified 210,827 significant cis-meQTL SNPs ($p < 1e\text{-}8$) using data from the GoDMC database [17]. Associations of these SNPs with 1358 traits included in the PheWAS Catalog were explored [44]. Following the application of a Bonferroni correction (threshold = $P < 0.05/1{,}358$), we discovered 78 significant SNP-phenotype associations, involving 26 unique cis-meQTL SNPs associated with 16 unique CpG sites (see Table S12). Out of these 26 SNPs, 18 were found in the MHC region (chromosome 6, 31–33 Mb; Table S12). The remaining eight SNPs, located outside the MHC region, were primarily associated with lipid metabolism disorders, such as hyperlipidemia, hypercholesterolemia, and disorders of lipoid metabolism. Conversely, the 18 SNPs in the MHC region were mostly associated with autoimmune disorders, including rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis, celiac disease, and diabetes and its complications. All SNPs were compared with the GWAS Catalog (https://www.ebi.ac.uk/gwas/downloads). A secondary PheWAS analysis, using UK Biobank ICD PheWeb [46], was conducted for these SNPs (Supplementary Figure S4). The results

confirmed the associations with lipid metabolism disorders and immune-related diseases. For example, consistent with the PheWAS Catalog, rs2647044 (mapped to HLA-DRB1) showed strong associations with type 1 diabetes ($p = 6.10e\text{-}57$) and celiac disease ($p = 0$), while rs693 (annotated with APOB) was associated with hypercholesterolemia ($p = 1.20e\text{-}12$) and hyperlipidemia ($p = 1.90e\text{-}12$) in large-scale GWAS datasets.

### Overlap with EWAS Toolkit analysis

From the 159 validated CpG sites identified through EWMR as being associated with PC, we identified associations with 41 distinct phenotypes in the EWAS Atlas (see Table S13). These phenotypes encompass a wide range of conditions, including aging, multiple sclerosis, B acute lymphoblastic leukemia with dic (9;20)(P11;Q13), smoking, and Down syndrome, as well as Primary sjögren's syndrome (Pss), colorectal laterally spreading tumor.

Enrichment analysis revealed that these validated PC-associated CpG sites were significantly over-represented in the phenotypes mentioned above (see Supplementary Figure S5 for the top 20 based on adjusted P-values). Significant enrichments included a 4-fold increase for aging ($p = 1.46e\text{-}15$), 4-fold increase for smoking ($p = 9.78e\text{-}11$), and a 8-fold increase for multiple sclerosis ($p = 1.18e\text{-}11$), as well as a 86-fold increase for educational attainment. Of all phenotypes, aging and smoking were associated with the highest number of overlapping CpG sites, 21 and 16, respectively, while educational attainment exhibited the greatest fold enrichment, 86-fold ($p = 0.000279$).

### Druggable gene target analysis

Of the 89 genes annotated to the 159 validated EWMR-identified PC-associated CpG sites, 27 of these genes were mapped to a total of 476 drug compounds, identified through integration with various drug databases via DGIdb. Additionally, analysis using the ConnectivityMap (CMap) resource revealed 34 compounds corresponding to 10 of these genes. Notably, eight genes were found to have drug targets in both datasets, including CACNA1D, TERT, IL6, APOB, CACNA1A, CASR, ABAT, RXRB, all of which are targeted by at least one approved therapeutic agent for various medical conditions. Furthermore, data from the ECTM herbal medicine database identified nine herbal compounds targeting genes within this set of 12 genes. Complete details are summarized in Table S14.

### MVMR and mediation analysis

To explore how gene expression might mediate the relationship between the 159 validated CpG sites (identified through EWMR as being associated to PC and validated in at least one of the three replication phases) and disease progression, we conducted MVMR and mediation analysis. This analysis incorporated three mediator layers and applied MVMR models to assess both exposure and mediator effects on the outcome. Our findings identified 118 genes as potential mediators based on the direct MR causal effect estimates when considering only the primary mediator (top mediator). For example, in the relationship between the whole-blood CpG cg12403698 and PC risk, CACUL1 accounted for almost 40% MP when restricted to the top mediator. When considering all cis mediators, cg12403698 showed a similar MP estimate for PC risk. Another example is in CD8+ effector memory T cells within PBMCs: in the relationship between cg03698956 and PC risk, RP11-77H9.2 accounted for about 17% MP for the protective association with PC risk when restricted to the top mediator; when considering all cis mediators, cg03698956 showed a similar MP for this protective association.

Detailed mediation results, including those based on models incorporating all selected mediators, are provided in Supplementary Table S15. Regarding the direct and total effect estimates, Supplementary Figure S6 shows that they are highly correlated both when only the top mediator is considered and when all cis mediators are included.

## Discussion

Extensive research of DNA methylation has been devoted to identifying new biomarkers that could improve risk evaluation and treatment strategies for PC [12,59,60]. However, the majority of earlier studies relied on EWAS, which lacked the statistical power due to limited sample sizes to clarify the causal relationship between CpG sites and PC. Although recent large-scale genetic studies have demonstrated that certain CpG sites exert causal effects across multiple diseases or traits in both Asian and European populations, PC has not yet been included in these analyses [17,18]. Considering these points, it is valuable to clarify the causal relation of blood CpG sites and PC, with the goal of efficiently identifying potential diagnostic biomarkers and therapeutic targets through MR that offer strong supporting evidence and reduced susceptibility to confounding.

In this study, we identified eligible cis-meQTLs meeting core MR assumptions by following rigorous quality control procedures. These variants were subsequently applied in an EWMR analysis, and the results indicated that genetically predicted blood 253 CpG sites were significantly related to the risk of PC that passed sensitivity analysis. The subsequent analysis was divided into three sections according to the analytical objectives. In the first section, focused on external validation and meta-analysis, 38 CpG sites retained and all the 253 significant CpG sites were categorized into three tiers. Cg14271713 (intergenic, chr9:136150031-136151365, S shelf), cg11652496 (PSTPIP1, chr15:77324520-77324759, gene body, island), cg20575191 (PSTPIP1, chr15:77324520-77324759, gene body, island), and Cg26373071 (CLPTM1L, gene body) lay on the tier 1 level with high support evidence. Subsequently, in second section for exploration the biological function and relation to various diseases or phenotypes of the validated EWMR identified PC-associated CpG sites, we executed various enrichment analysis, overlap with PheWAS catalog analysis, overlap with EWAS Toolkit analysis, and druggable gene target analysis. Additionally, MVMR and mediation analysis were employed to dissect the total effect of identified CpG sites to PC into direct effect and indirect effect mediated by gene expression. Notably, to make our results more robust, we selected 159 validated EWMR identified PC-associated CpG sites to perform downstream analysis.

As shown in the present study, cg26373071, located in the gene body of CLPTM1L (Cleft Lip And Palate Transmembrane Protein 1-Like Protein), was a risk factor of PC. CLPTM1L encodes a membrane-bound protein, and its elevated expression in ovarian cancer cells resistant to cisplatin can induce resistance to antitumor drugs (e.g., cisplatin) [61]. Genetic variants within this gene are related to a heightened risk of various cancers [62–65]. Previous studies have indicated that CLPTM1L could enhance the proliferation of pancreatic cancer cells both in vivo and in vitro, potentially exerting its effects by disrupting normal cell division processes [66]. Furthermore, anti-CLPTM1L antibodies suppressed the membrane localization of CLPTM1L, reduced Akt phosphorylation, and diminished both anchorage-independent proliferation and resistance to chemotherapy in lung and pancreatic cancer cells [67]. Therefore, we speculated that elevated methylation levels of cg26373071, increasing CLPTM1L expression, may contribute to increasing the risk of PC through above process. Additionally, cg07493874, mapped to the CLPTM1L gene, which was calssified as tier 3, also exhibited association with PC (see Table S11 for details). Although tier 3 findings are supported by weaker evidence, they still highlight the intricate role of CLPTM1L methylation in PC progression. Interestingly, MVMR and mediation analysis suggested that cg26373071 may influence PC risk via the MRPL36 gene in B Men (Memory B cell) single cells within PBMCs, with an MP of −15.82%. Likewise, this CpG site also appeared to affect PC risk through NDUFS6 in CD8 ET (CD8+ Effector memory T cell) single cells, showing an MP of 1.49%. In addition, mediation analysis also showed that cg07380026 and cg27619457 exerted effect on PC generation through CLPTM1L gene expression (see Table S 15 for details). Overall, cg26373071 (CLPTM1L) may be a key element despite that the causal effect of DNA methylation on PC may be based on multiple CpG sites and genes.

Cg20575191 and cg11652496, both located in PSTPIP1 (Proline-Serine-Threonine Phosphatase Interacting Protein 1) gene body, also exerted as a risk factor to PC. It is essentially a cytoskeleton adaptor protein involved in the regulation of various cellular processes, including interleukin-1β secretion, cytoskeletal organization, cell migration, and the activation of T cells [68,69]. Mutations at the PSTPIP1 are related to inflammatory conditions collectively referred to as the PSTPIP1-associated myeloid-related proteinemia inflammatory syndrome (PAMI syndrome) [70]. Previous studies have demonstrated that elevated PSTPIP1 expression serves as a potential indicator of reduced OS of clear cell renal cell carcinoma

[71]. Nonetheless, the involvement of PSTPIP1 in PC remains unexplored in current studies, underscoring the necessity for additional investigation into this potential connection.

Subsequently, although our GO and KEGG enrichment revealed no significance in biological process, cellular component, and KEGG which may due to the number of genes mapped to selected CpG sites, the overlap with PheWAS catalog analysis explicated that the autoimmune disorders, especially type 1 diabetes mellitus, were related to PC. A study based on large population using the US Surveillance, Epidemiology, and End Results Program (SEER)-Medicare linked data also revealed the close association of many autoimmune disorders and PC [72]. Diabetes mellitus is a risk factor for PC and a result of the disease [73,74]. Overlap with EWAS Toolkit analysis also revealed that several phenotypes or diseases were related to PC through epigenetic influences. For example, the 159 validated EWMR identified PC-associated CpG sites significantly overlapped with those driven by smoking, showing close link between smoking and PC. Smoking has been demonstrated to be closely related to the development of PC [75,76]. Interestingly, overlap with EWAS Toolkit analysis showed that epithelial ovarian cancer risk was linked to PC with large odd ratio. However, recent research demonstrated that injecting ovarian cancer cells can markedly suppress pancreatic cancer xenograft growth by altering the expressions of genes related to pancreatic function and phenotype [77]. In summary, the underlying relationship of ovarian cancer and PC remains to be elucidated in the future.

This study offers several notable strengths. To the best of our knowledge, it represents the most comprehensive and large-scale EWMR analysis conducted for PC to date. Its broad scope, analytical depth, and methodological rigor enabled reliable identification of potential targets for screening and therapeutic interventions of PC. First, multiple strategies were employed to minimize MR assumption violations and relieve confounding bias. meQTLs in the MHC region or positioned far from their associated CpG sites (i.e., trans-meQTLs) were excluded. Given the sensitivity of MR results to the selection of IVs, we further validated our findings using alternative IV inclusion criteria adopted in previous multi-omic MR research [23,78]. To strengthen the reliability of associations identified in this study, we conducted a series of additional sensitivity analysis, directionality tests, and colocalization analysis. Furthermore, consistent findings from replicated MR analysis using alternative datasets from both European and Asian populations, followed by meta-analysis, were used to avoid false-positive results. Second, findings from enrichment analysis, overlap with PheWAS catalog analysis, overlap with EWAS Toolkit analysis, and druggable gene target analysis offer insights into the mechanisms by which the candidate targets affect PC. Third, we conducted MVMR and mediation analysis across three different mediator dimensions, single-cell data, whole blood, and pancreatic tissue, to further explore the underlying pathogenic pathways. This approach enabled us to differentiate the total effect of each CpG site on PC into direct and indirect effect size.

Nevertheless, limitations of the present study shall be acknowledged. First, although individuals of East Asian ancestry were included in the validation phase, most participants were European, resulting in low generalizability of our findings to populations of African or other ancestries. Second, while using cis-meQTLs rather than trans-meQTLs helps to avoid horizontal pleiotropy, this approach may restrict the number of analyzable CpG sites due to the absence of suitable SNPs for some loci. Third, our analysis primarily focused on CpG sites identified in plasma samples, and their potential roles in pancreatic tissue could not be examined due to limited data availability. Similarly, the lack of corresponding protein QTL data for genes linked to the identified CpG sites prevented us from assessing associations at the plasma protein level. Additionally, the overlap with EWAS Toolkit results was not validated using independent datasets, warranting cautious interpretation. Finally, although MR is a powerful tool to infer causal relationships, it cannot completely rule out confounding or reverse causation. Therefore, further experimental studies and epidemiological investigations are needed to validate our findings. It is particularly worth noting that for the part of MVMR and mediation analysis, the calculated MP is an estimate of the underlying quantity; accordingly, estimates outside the [0,1] interval can occur. Conceptually, MPs may depend on the magnitude of the total causal effect; when the direct and indirect effects have opposite signs, the total effect can be weak, yielding MP estimates outside [0,1] [52]. In the present study, we report MP for DNAm–PC pairs under two specifications: (i) 'top mediation,' restricted to pairs in which the top mediator is significantly associated with the exposure, and (ii) 'overall mediation,' aggregating across all cis mediators.

The latter provides a more comprehensive quantification of the contribution of cis transcripts to DNAm effects, as restricting analyses to pairs with a detected mediator may bias estimates upward. To assess internal consistency, we computed the Pearson correlation between the direct and total effects, and we incorporated multiple layers of eQTL data – including single-cell datasets from peripheral blood mononuclear cells and pancreas – to mitigate artefacts arising from bulk whole blood. Notwithstanding these steps, the mediation results should be interpreted with caution and warrant further experimental validation. Plans are already in progress to conduct follow-up cell-based and animal studies to provide deeper insights into the pathogenic roles of selected CpG sites.

## Conclusion

In summary, 253 CpG sites were identified as having potential etiological relevance to pancreatic cancer. Among them, four were supported by the strongest evidence. With further validation, these candidate CpG sites and their corresponding genes are potentially biomarkers for early detection and represent potential therapeutic targets for pancreatic cancer.

## Ethics approval

This study involved no newly collected data, and no new ethical approval was required.

## Author contributions

CRediT: **Pengxu Wang:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Feng Rong:** Funding acquisition, Supervision, Writing – review & editing; **Fubao Liu:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the Key Project of the Anhui Provincial Sci-Tech Innovation Breakthrough Plan under Grant [202423k09020009], Key Project of the Scientific Research Program of Anhui Provincial Health Commission under Grant [AHWJ2023A10028] and Wu Jieping Medical Foundation Special Funding for Clinical Research under Grant [320.6750.2023–18-70].

## ORCID

Pengxu Wang http://orcid.org/0009-0002-1786-0851
Fubao Liu http://orcid.org/0000-0003-3431-2187

## Data availability statement

The data of GWAS summary statistics of PC from UK Biobank can be accessed at https://www.nealelab.is/uk-biobank. The data of meQTL from GoDMC can be accessed at http://mqtldb.godmc.org.uk. The meQTL data from FHS can be accessed at https://ftp.ncbi.nlm.nih.gov/eqtl/original_submissions/FHS_meQTLs/. GWAS data for PC from the R12 release of the FinnGen consortium can be accessed at https://www.finngen.fi/en. GWAS data for East Asian meQTL can be accessed at https://www.biosino.org/node/project/detail/OEP002902. GWAS data for PC from BBJ can be accessed at https://pheweb.jp/. The PheWAS catalog can be accessed at https://phewascatalog.org. PheWAS search using the UKBiobank ICD PheWeb can be accessed at https://pheweb.org/UKB-SAIGE/. The EWAS Atlas database can be accessed at https://ngdc.cncb.ac.cn/ewas/toolkit. The DGIdb database can be accessed at https://dgidb.org/. The CLUE app can be accessed at https://clue.io/. The ETCM database can be accessed at http://www.tcmip.cn/ETCM/. SMR-IVW tool can be accessed at https://github.com/masadler/smrivw. SMR format whole-blood cis-eQTL data from

the eQTLGen Consortium can be accessed at https://www.eqtlgen.org/cis-eqtls.html. Single-cell cis-eQTLs generated by the jaxQTL model from the OneK1K cohort can be accessed at https://zenodo.org/records/14968281. Cis-eQTL data in pancreas tissue from the V8 release of GTEx in SMR format can be acessed at https://yanglab.westlake.edu.cn/software/smr/#DataResource. Datesets from Road Epigenomics Project can be accessed at https://egg2.wustl.edu/roadmap/web_portal/. GWAS Catalog can be accessed at https://www.ebi.ac.uk/gwas/downloads. The data that support the findings of this study are openly available in figshare at http://doi.org/10.6084/m9.figshare.29965814, reference number 10.6084/m9.figshare.29965814.

## Abbreviations

| | |
|---|---|
| PC | Pancreatic cancer |
| DNAm | DNA methylation |
| meQTL | DNA methylation quantitative trait loci |
| MR | Mendelian randomization |
| Cis-meQTL | Cis-acting DNA methylation quantitative trait loci |
| IV | Instrumental variable |
| SNP | Single-nucleotide polymorphisms |
| EWMR | Epigenome-wide Mendelian randomization |
| GWAS | Genome-wide association study |
| EWAS | Epigenome-wide association study |
| PheWAS | Phenome-wide association study |
| MVMR | Multivariable Mendelian Randomization |
| GoDMC | Genetics of DNA Methylation Consortium |
| FDR | False discovery rate |
| MAF | Minor allele frequency |
| LD | Linkage disequilibrium |
| MHC | Major histocompatibility complex |
| IVW | Inverse variance-weighted |
| FHS | Framingham Heart Study |
| NSPT | National Survey of Physical Traits cohort |
| BBJ | Biobank Japan |
| DGIdb | Drug–Gene Interaction Database |
| ETCM | Encyclopedia of Traditional Chinese Medicine |
| MP | Mediation proportion |
| eQTL | Expression quantitative trait loci |
| scRNA-seq | Single-cell RNA sequencing |
| PBMC | Peripheral blood mononuclear cell |
| GTEx | Genotype-Tissue Expression |

## References

[1] Siegel RL, Kratzer TB, Giaquinto AN, et al. Cancer statistics, 2025. CA Cancer J Clin. 2025;75(1):10–45. doi: 10.3322/caac.21871
[2] Connor AA, Gallinger S. Pancreatic cancer evolution and heterogeneity: integrating omics and clinical data. Nat Rev Cancer. 2022;22(3):131–142. doi: 10.1038/s41568-021-00418-1
[3] Stoop TF, Javed AA, Oba A, et al. Pancreatic cancer. The Lancet. 2025;405(10485):1182–1202. doi: 10.1016/S0140-6736(25)00261-2
[4] Klein AP. Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors. Nat Rev Gastroenterol Hepatol. 2021;18(7):493–502. doi: 10.1038/s41575-021-00457-x
[5] Rahib L, Wehner MR, Matrisian LM, et al. Estimated projection of US cancer incidence and death to 2040. JAMA Netw Open. 2021;4(4):e214708. doi: 10.1001/jamanetworkopen.2021.4708
[6] Wu H, Zhang Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. Cell. 2014;156(1–2):45–68. doi: 10.1016/j.cell.2013.12.019
[7] Smith ZD, Hetzel S, Meissner A. DNA methylation in mammalian development and disease. Nat Rev Genet. 2024;26(1):7–30. doi: 10.1038/s41576-024-00760-8
[8] Gaiti F, Chaligne R, Gu H, et al. Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. Nature. 2019;569(7757):576–580. doi: 10.1038/s41586-019-1198-z
[9] Vidal E, Sayols S, Moran S, et al. A DNA methylation map of human cancer at single base-pair resolution. Oncogene. 2017;36(40):5648–5657. doi: 10.1038/onc.2017.176

[10] Xie W, Kagiampakis I, Pan L, et al. DNA methylation patterns separate senescence from transformation potential and indicate cancer risk. DNA Methylation Patterns Sep Senescence From Transform Potential And Indicate Cancer Risk Cancer Cell. 2018;33(2):309–321.e305. doi: 10.1016/j.ccell.2018.01.008
[11] De Cecco M, Criscione SW, Peckham EJ, et al. Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. Aging Cell. 2013;12(2):247–256. doi: 10.1111/acel.12047
[12] Michaud DS, Ruan M, Koestler DC, et al. Epigenome-wide association study using prediagnostic bloods identifies new genomic regions associated with pancreatic cancer risk. JNCI Cancer Spectr. 2020;4(5):kaa041. doi: 10.1093/jncics/pkaa041
[13] Neale RE, Clark PJ, Fawcett J, et al. Association between hypermethylation of DNA repetitive elements in white blood cell DNA and pancreatic cancer. Cancer Epidemiol. 2014;38(5):576–582. doi: 10.1016/j.canep.2014.08.006
[14] Dauksa A, Gulbinas A, Barauskas G, et al. Whole blood DNA aberrant methylation in pancreatic adenocarcinoma shows association with the course of the disease: a pilot study. PLOS ONE. 2012;7(5):e37509. doi: 10.1371/journal.pone.0037509
[15] Fewell Z, Davey Smith G, Sterne JA. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. Am J Epidemiol. 2007;166(6):646–655. doi: 10.1093/aje/kwm165
[16] Huan T, Joehanes R, Song C, et al. Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. Nat Commun. 2019;10(1):4267. doi: 10.1038/s41467-019-12228-z
[17] Min JL, Hemani G, Hannon E, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. Nat Genet. 2021;53(9):1311–1321. doi: 10.1038/s41588-021-00923-x
[18] Peng Q, Liu X, Li W, et al. Analysis of blood methylation quantitative trait loci in East Asians reveals ancestry-specific impacts on complex traits. Nat Genet. 2024;56(5):846–860. doi: 10.1038/s41588-023-01494-9
[19] Sekula P, Del Greco MF, Pattaro C, et al. Mendelian randomization as an approach to assess causality using observational data. JASN. 2016;27(11):3253–3265. doi: 10.1681/ASN.2016010098
[20] Emdin CA, Khera AV, Kathiresan S. Mendelian randomization. Jama. 2017;318(19):1925–1926. doi: 10.1001/jama.2017.17219
[21] Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014;23(R1):R89–98. doi: 10.1093/hmg/ddu328
[22] Zhong H, Liu S, Zhu J, et al. Elucidating the role of blood metabolites on pancreatic cancer risk using two-sample Mendelian randomization analysis. Int J Cancer. 2024;154(5):852–862. doi: 10.1002/ijc.34771
[23] Zhou S, Tao B, Guo Y, et al. Integrating plasma protein-centric multi-omics to identify potential therapeutic targets for pancreatic cancer. J Transl Med. 2024;22(1):557. doi: 10.1186/s12967-024-05363-9
[24] Jiang Z, Mou Y, Wang H, et al. Causal effect between gut microbiota and pancreatic cancer: a two-sample Mendelian randomization study. BMC Cancer. 2023;23(1):1091. doi: 10.1186/s12885-023-11493-y
[25] Richardson TG, Haycock PC, Zheng J, et al. Systematic Mendelian randomization framework elucidates hundreds of CpG sites which may mediate the influence of genetic variants on disease. Hum Mol Genet. 2018;27(18):3293–3304. doi: 10.1093/hmg/ddy210
[26] Ying K, Liu H, Tarkhov AE, et al. Causality-enriched epigenetic age uncouples damage and adaptation. Nat Aging. 2024;4(2):231–246. doi: 10.1038/s43587-023-00557-0
[27] Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomization: the STROBE-MR statement. JAMA. 2021;326(16):1614–1621. doi: 10.1001/jama.2021.18236
[28] Papadimitriou N, Dimou N, Tsilidis KK, et al. Physical activity and risks of breast and colorectal cancer: a Mendelian randomisation analysis. Nat Commun. 2020;11(1):597. doi: 10.1038/s41467-020-14389-8
[29] Shim H, Chasman DI, Smith JD, et al. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. PLOS ONE. 2015;10(4):e0120758. doi: 10.1371/journal.pone.0120758
[30] Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7:7. doi: 10.7554/eLife.34408
[31] Yavorska OO, Burgess S. MendelianRandomization an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol. 2017;46(6):1734–1739. doi: 10.1093/ije/dyx034
[32] Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. Statist Med. 2016;35(11):1880–1906. doi: 10.1002/sim.6835
[33] Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512–525. doi: 10.1093/ije/dyv080
[34] Verbanck M, Chen CY, Neale B, et al. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50(5):693–698. doi: 10.1038/s41588-018-0099-7
[35] Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLOS Genet. 2017;13(11):e1007081. doi: 10.1371/journal.pgen.1007081

[36] Wallace C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. PLOS Genet. 2020;16(4):e1008720. doi: 10.1371/journal.pgen.1008720
[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. doi: 10.1371/journal.pgen.1004383
[38] Kurki MI, Karjalainen J, Palta P, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023;613(7944):508–518. doi: 10.1038/s41586-022-05473-8
[39] Sakaue S, Kanai M, Tanigawa Y, et al. A cross-population atlas of genetic associations for 220 human phenotypes. Nat Genet. 2021;53(10):1415–1424. doi: 10.1038/s41588-021-00931-x
[40] Thorlund K, Imberger G, Johnston BC, et al. Evolution of heterogeneity (I2) estimates and their 95% confidence intervals in large meta-analyses. PLOS ONE. 2012;7(7):e39471. doi: 10.1371/journal.pone.0039471
[41] Higgins JP. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557–560. doi: 10.1136/bmj.327.7414.557
[42] Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317–330. doi: 10.1038/nature14248
[43] Dennis G, Sherman BT, Hosack DA, et al. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol. 2003;4(5):3. doi: 10.1186/gb-2003-4-5-p3
[44] Yu G, Wang LG, Han Y, et al. Clusterprofiler: an R package for comparing biological themes among gene clusters. Omics: A J Intgr Biol. 2012;16(5):284–287. doi: 10.1089/omi.2011.0118
[45] Denny JC, Bastarache L, Ritchie MD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol. 2013;31(12):1102–1111. doi: 10.1038/nbt.2749
[46] Sollis E, Mosaku A, Abid A, et al. The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. NHGRI-EBI Nucleic Acids Res. 2023;51(D1):D977–d985. doi: 10.1093/nar/gkac1010
[47] Gagliano Taliun SA, VandeHaar P, Boughton AP, et al. Exploring and visualizing large-scale genetic associations by using PheWeb. Nat Genet. 2020;52(6):550–552. doi: 10.1038/s41588-020-0622-5
[48] Xiong Z, Yang F, Li M, et al. EWAS open platform: integrated data, knowledge and toolkit for epigenome-wide association study. Nucleic Acids Res. 2022;50(D1):D1004–d1009. doi: 10.1093/nar/gkab972
[49] Cannon M, Stevenson J, Stahl K, et al. DGIdb 5.0: rebuilding the drug-gene interaction database for precision medicine and drug discovery platforms. Nucleic Acids Res. 2024;52(D1):D1227–d1235. doi: 10.1093/nar/gkad1040
[50] Subramanian A, Narayan R, Corsello SM, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell. 2017;171(6):1437–1452.e17. doi: 10.1016/j.cell.2017.10.049
[51] Xu HY, Zhang YQ, Liu ZM, et al. ETCM: an encyclopaedia of traditional Chinese medicine. Nucleic Acids Res. 2019;47(D1):D976–d982. doi: 10.1093/nar/gky987
[52] Sadler MC, Auwerx C, Lepik K, et al. Quantifying the role of transcript levels in mediating DNA methylation effects on complex traits and diseases. Nat Commun. 2022;13(1):7559. doi: 10.1038/s41467-022-35196-3
[53] Wu Y, Zeng J, Zhang F, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. Nat Commun. 2018;9(1):918. doi: 10.1038/s41467-018-03371-0
[54] Võsa 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. doi: 10.1038/s41588-021-00913-z
[55] Zhang ZE, Kim A, Suboc N, et al. Efficient count-based models improve power and robustness for large-scale single-cell eQTL mapping. medRxiv [Preprint]. 2025;25320755. doi: 10.1101/2025.01.18.25320755
[56] Yazar S, Alquicira-Hernandez J, Wing K, et al. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. Science. 2022;376(6589):eabf3041. doi: 10.1126/science.abf3041
[57] Consortium TG, Anand S, Ardlie KG. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369(6509):1318–1330. doi: 10.1126/science.aaz1776
[58] Zuber V, Grinberg NF, Gill D, et al. Combining evidence from Mendelian randomization and colocalization: review and comparison of approaches. Am J Hum Genet. 2022;109(5):767–782. doi: 10.1016/j.ajhg.2022.04.001
[59] Zhu J, Yang Y, Kisiel JB, et al. Integrating genome and methylome data to identify candidate DNA methylation biomarkers for pancreatic cancer risk. Cancer Epidemiol Biomarker Prev. 2021;30(11):2079–2087. doi: 10.1158/1055-9965.EPI-21-0400
[60] Wu Y, Seufert I, Al-Shaheri FN, et al. Dna-methylation signature accurately differentiates pancreatic cancer from chronic pancreatitis in tissue and plasma. Gut. 2023;72(12):2344–2353. doi: 10.1136/gutjnl-2023-330155
[61] Yamamoto K, Okamoto A, Isonishi S, et al. A novel gene, CRR9, which was up-regulated in CDDP-resistant ovarian tumor cell line, was associated with apoptosis. Biochem Biophys Res Commun. 2001;280(4):1148–1154. doi: 10.1006/bbrc.2001.4250
[62] Jeong MS, Mun JY, Yang GE, et al. Exploring the relationship between CLPTM1L-MS2 variants and susceptibility to bladder cancer. Genes (Basel). 2023;15(1):15. doi: 10.3390/genes15010050
[63] Ni Z, Tao K, Chen G, et al. Clptm1l is overexpressed in lung cancer and associated with apoptosis. PLOS ONE. 2012;7(12):e52598. doi: 10.1371/journal.pone.0052598

[64] Petersen GM, Amundadottir L, Fuchs CS, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet. 2010;42(3):224–228. doi: 10.1038/ng.522
[65] Haiman CA, Chen GK, Vachon CM, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. Nat Genet. 2011;43(12):1210–1214. doi: 10.1038/ng.985
[66] Jia J, Bosley AD, Thompson A, et al. Clptm1l promotes growth and enhances aneuploidy in pancreatic cancer cells. *Cancer Res*. 2014;74(10):2785–2795. doi: 10.1158/0008-5472.CAN-13-3176
[67] Puskás LG, Mán I, Szebeni G, et al. Novel Anti-CRR9/CLPTM1L Antibodies With Antitumorigenic Activity Inhibit Cell Surf Accumulation, PI3K Interact, And Survival Signaling Mol Cancer Ther. 2016;15(5):985–997. doi: 10.1158/1535-7163.MCT-15-0717
[68] Cugno M, Borghi A, Marzano AV. Papa, PASH and PAPASH syndromes: pathophysiology, presentation and treatment. Am J Clin Dermatol. 2017;18(4):555–562. doi: 10.1007/s40257-017-0265-1
[69] Sood AK, Db M, Googe PB, et al. Successful treatment of papa syndrome with dual adalimumab and tacrolimus therapy. J Clin Immunol. 2019;39(8):832–835. doi: 10.1007/s10875-019-00685-6
[70] Marzano AV, Damiani G, Ceccherini I, et al. Autoinflammation in pyoderma gangrenosum and its syndromic form (pyoderma gangrenosum, acne and suppurative hidradenitis). Br J Dermatol. 2017;176(6):1588–1598. doi: 10.1111/bjd.15226
[71] Chen Y, Qin X, Sun L, et al. Expression Of PSTPIP1 In Renal Cell Carcinoma And Its Prognostic Value. Biochem Genet. 2025. doi: 10.1007/s10528-025-11163-7
[72] Yuan F, Pfeiffer RM, Julián-Serrano S, et al. Autoimmune conditions and pancreatic cancer risk in older American adults. Int J Cancer. 2023;152(2):172–182. doi: 10.1002/ijc.34235
[73] Elena JW, Steplowski E, Yu K, et al. Diabetes and risk of pancreatic cancer: a pooled analysis from the Pancreatic Cancer Cohort Consortium. Cancer Causes Control. 2013;24:13–25.
[74] Chari ST, Leibson CL, Rabe KG, et al. Probability of pancreatic cancer following diabetes: a population-based study. Gastroenterology. 2005;129(2):504–511. doi: 10.1016/j.gastro.2005.05.007
[75] Iodice S, Gandini S, Maisonneuve P, et al. Tobacco and the risk of pancreatic cancer: a review and meta-analysis. Langenbecks Arch Surg. 2008;393(4):535–545. doi: 10.1007/s00423-007-0266-2
[76] Bosetti C, Lucenteforte E, Silverman DT, et al. Cigarette smoking and pancreatic cancer: an analysis from the International Pancreatic Cancer Case-control Consortium (PanC4). Ann Oncol. 2012;23(7):1880–1888. doi: 10.1093/annonc/mdr541
[77] Lin X, Cui C, Cui Q. Combating pancreatic cancer with ovarian cancer cells. Aging (Albany NY). 2023;15(6):2189–2207. doi: 10.18632/aging.204608
[78] Bourgault J, Abner E, Manikpurage HD, et al. Proteome-wide Mendelian randomization identifies causal links between blood proteins and acute pancreatitis. Gastroenterology. 2023;164(6):953–965.e3. doi: 10.1053/j.gastro.2023.01.028