Original research

Polymorphic variants involved in methylation regulation: a strategy to discover risk loci for pancreatic ductal adenocarcinoma

Chiara Corradi ,[1] Giulia Lencioni,[1] Manuel Gentiluomo ,[1] Alessio Felici,[1] Anna Latiano,[2] Gediminas Kiudelis,[3] Casper H J van Eijck,[4] Katalin Marta,[5,6] Rita T Lawlor,[7] Francesca Tavano,[2] Ugo Boggi,[8] Frederike Dijk,[9] Giulia Martina Cavestro,[10] Roel C H Vermeulen,[11] Thilo Hackert,[12] Maria Chiara Petrone,[13] Faik Güntac Uzunoğlu,[14] Livia Archibugi,[13,15] Jakob R Izbicki,[14] Luca Morelli,[16] Alessandro Zerbi,[17,18] Stefano Landi,[1] Hannah Stocker,[19,20] Renata Talar-Wojnarowska,[21] Gregorio Di Franco,[16] Péter Hegyi,[5,6,22,23] Cosimo Sperti,[24] Silvia Carrara,[25] Gabriele Capurso,[13,15] Maria Gazouli ,[26] Hermann Brenner,[27,28] Stefania Bunduc,[5,6,29] Olivier Busch,[30] Francesco Perri,[2] Martin Oliverius,[31] Péter Jeno Hegyi,[5,6] Mara Goetz,[14] Pasquale Scognamiglio,[14] Andrea Mambrini,[32] Paolo Giorgio Arcidiacono,[13] Edita Kreivenaite,[3] Juozas Kupcinskas,[3] Tamas Hussein,[5,6] Stefano Ermini,[33] Anna Caterina Milanetto,[24] Pavel Vodicka,[34,35,36] Vytautas Kiudelis,[3] Viktor Hlaváč,[37] Pavel Soucek,[37] George E Theodoropoulos,[38] Daniela Basso,[39] John P Neoptolemos,[12] Mateus Nóbrega Aoki,[40] Raffaele Pezzilli,[41] Claudio Pasquali,[24] Roger Chammas,[40] Sabrina Gloria Giulia Testoni,[13] Beatrice Mohelnikova-Duchonova,[42] Maurizio Lucchesi,[32] Cosmeri Rizzato,[43] Federico Canzian,[44] Daniele Campa [1]

For numbered affiliations see end of article.

Correspondence to
Professor Daniele Campa, Department of Biology, University of Pisa, Pisa, Italy; daniele.campa@unipi.it

Received 1 September 2022
Accepted 4 April 2023
Published Online First 2 May 2023

To cite: Corradi C, Lencioni G, Gentiluomo M, *et al*. *J Med Genet* 2023;**60**:980–986.

ABSTRACTDANIELE CAMPA:

Introduction Only a small number of risk factors for pancreatic ductal adenocarcinoma (PDAC) has been established. Several studies identified a role of epigenetics and of deregulation of DNA methylation. DNA methylation is variable across a lifetime and in different tissues; nevertheless, its levels can be regulated by genetic variants like methylation quantitative trait loci (mQTLs), which can be used as a surrogate.

Materials and methods We scanned the whole genome for mQTLs and performed an association study in 14 705 PDAC cases and 246 921 controls. The methylation data were obtained from whole blood and pancreatic cancer tissue through online databases. We used the Pancreatic Cancer Cohort Consortium and the Pancreatic Cancer Case–Control Consortium genome-wide association study (GWAS) data as discovery phase and the Pancreatic Disease Research consortium, the FinnGen project and the Japan Pancreatic Cancer Research consortium GWAS as replication phase.

Results The C allele of 15q26.1-rs12905855 showed an association with a decreased risk of PDAC (OR=0.90, 95% CI 0.87 to 0.94, p=4.93×10^{-8} in the overall meta-analysis), reaching genome-level statistical significance. 15q26.1-rs12905855 decreases the methylation of a 'C-phosphate-G' (CpG) site located in the promoter region of the *RCCD1* antisense (*RCCD1-AS1*) gene which, when expressed, decreases the expression of the RCC1 domain-containing (*RCCD1*) gene (part of a histone demethylase complex). Thus, it is possible that the rs12905855 C-allele has a protective role in PDAC development through an increase of *RCCD1* gene expression, made possible by the inactivity of *RCCD1-AS1*.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The overall estimated heritability of pancreatic ductal adenocarcinoma (PDAC) is around 21.2%, but only 4.5% of it has been determined, suggesting that many additional loci remain to be identified. The majority of those loci lack a functional explanation of the statistical association.

WHAT THIS STUDY ADDS

⇒ A novel PDAC risk variant, 15q26.1-rs12905855, that regulates methylation in the pancreatic tissue has been identified with a genome-wide level of statistical significance (p=4.93×10^{-8}). This SNP modifies the methylation and the consequent expression in the pancreas of the RCC1 domain-containing (*RCCD1*) gene that is involved in chromosomal stability during mitosis.

HOW MIGHT THIS STUDY AFFECT RESEARCH, PRACTICE OR POLICY

⇒ These results improve the knowledge on PDAC genetic.

Conclusion We identified a novel PDAC risk locus which modulates cancer risk by controlling gene expression through DNA methylation.

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC) constitutes the fourth cause of cancer-related deaths in Europe.[1 2] The early stages of PDAC are characterised by few or unspecific symptoms, making diagnosis very difficult. Surgery remains the only curative treatment but is possible in only a small portion of patients as most PDAC cases are diagnosed at an advanced stage.[3 4]

Several genome-wide association studies (GWASs) and gene candidate studies have identified around 30 common loci associated with PDAC susceptibility in the European population,[5–19] with a smaller number in non-European ethnicities.[20–22] The overall estimated heritability of PDAC is around 21.2%, but only 4.5% of it has been determined, suggesting that many additional loci remain to be identified.[23] Additionally, GWASs are prone to false-negative results since only SNPs that reach a very restrictive p-value threshold ($p<5\times10^{-8}$) are generally reported. Moreover, GWASs suffer, as an additional limitation, from the inherent difficulty in linking the identified variants with a function that explains their effect. A possible solution is offered by secondary analysis of large GWAS data using only functional SNPs and a replication in a large cohort for validation. This strategy has been useful in identifying regulatory SNPs associated with PDAC susceptibility that have been overlooked by the original GWASs.[11 24 25] In addition to genetics, several studies have also identified a relevant role of epigenetics in PDAC aetiology, and in particular in the deregulation of DNA methylation, with hypermethylation of tumour suppressor genes and hypomethylation of oncogenes.[26–28] However, DNA methylation shows variability across a lifetime, in response to environmental stimuli or ageing, and shows tissue specificity,[29] making the studies that rely on blood as a proxy tissue difficult to interpret with respect to the methylation in a specific organ.[30] Therefore, the epidemiological investigation of the effects of DNA methylation remains a complex challenge due to its fluid feature. However, DNA methylation has a genetic component. Methylation quantitative trait loci (mQTLs) are germline variants associated with DNA methylation level and are, by their nature, stable during time and disease.[31] The genetic component of methylation accounts for a small fraction of its variability; nevertheless, mQTLs could represent a good surrogate to study DNA methylation in epidemiological settings. Recently, several studies highlighted the association between mQTLs and risk of breast, bladder and prostate cancers.[32–35] Heyn and colleagues conducted a small study on 49 PDAC cases and seven controls and identified rs401681, an mQTL in the *TERT* locus, to be associated with PDAC development.[34] With these premises, in this study, we identified all human pancreatic mQTLs and analysed their involvement in PDAC susceptibility in a large case–control association study.

MATERIALS AND METHODS

Discovery phase

The discovery phase consisted of the data of the Pancreatic Cancer Cohort Consortium (PanScan) (PanScan I–III) and the Pancreatic Cancer Case–Control Consortium (PanC4). The data were downloaded from the National Center for Biotechnology Information Database of Genotypes and Phenotypes (study accession numbers phs000206.v5.p3 and phs000648.v1.p1, project reference #12644). Detailed information on the study participants, genotyping arrays used and analysis is described in the original papers.[5–8] The datasets were imputed separately using the Michigan Imputation Server (https://imputationserver.sph.umich.edu) and the Haplotype Reference Consortium V.r1.1 as reference panels, and subsequently the imputed datasets were merged. Before imputation, quality control procedures were performed and individuals with gender mismatches, call rate of <0.98, minimal or excessive heterozygosity (>3 SD from the mean), or cryptic relatedness (PI_HAT >0.2) were excluded from the dataset to be imputed. SNPs with low imputation quality (INFO score $r^2<0.7$), minor allele frequency of <0.01 or call rate of <0.9, and evidence for violation of the Hardy-Weinberg equilibrium (HWE) ($p<10^{-6}$) were excluded. Principal component analysis (PCA) was carried out with PLINK V.2.0 (www.cog-genomics.org/plink/2.0/), including genotypes from all the populations of the phase III of the 1000 Genomes Project. Individuals not clustering in the PCA with the 1000 Genomes subjects of European descent were excluded from further analysis. The final dataset comprised 15 772 individuals (8738 cases and 7034 controls).

Replication phase

For the replication phase, three consortia were analysed: the Pancreatic Disease Research (PANDoRA) consortium, the FinnGen project and the Japan Pancreatic Cancer Research (JaPAN) consortium.

The PANDoRA consortium consists of a multicentric study based mainly on European countries (Italy, Germany, Hungary, Czech Republic, Poland, Lithuania, the Netherlands and Greece), and it has been extensively described elsewhere.[36] For this study 3047 PDAC cases and 3225 controls were used. The controls were collected in the same geographical regions as the cases. Additional German controls from 'Epidemiologische Studie zu Chancen der Verhütung,Früherkennung und optimierten THerapie chronischerERkrankungen in der älteren Bevölkerung' (ESTHER) and Dutch controls from the European Prospective Investigation on Cancer (http://epic.iarc.fr/), two prospective cohorts with available GWAS data, have been included in the study.[2 37] PANDoRA also includes a subgroup of 69 PDAC cases and 258 controls from Brazil, which were analysed separately from the Europeans in this study.

The FinnGen project GWAS on 881 PDAC and 204 070 controls with Finnish ancestry was used for replication.[38] Subjects affected by other cancer types were excluded from the controls. Summary statistics were downloaded from the FinnGen website (FinnGen Release R6). More details on genotypes, data and statistical analysis are available at the FinnGen website (https://www.finngen.fi).

The third validation consisted of individuals of East Asian ancestry. The summary statistics of a meta-analysis including three GWASs (JaPAN consortium, National Cancer Centre and BioBank Japan) were downloaded from the JaPAN consortium website (http://www.aichi-med-u.ac.jp/JaPAN/index-e.html). This study contains 34 631 individuals of East Asian origin (2039 PDAC cases and 32 592 controls). Detailed information regarding the JaPAN study is given elsewhere.[22]

The total number of individuals analysed in the study summed up to 14 705 cases and 246 921 controls (table 1).

SNP selection

The SNP selection was made using five different studies/databases in which mQTLs are annotated. One, PanCan meQTL,[39] lists mQTLs identified on pancreatic cancer tissue samples, while the

Table 1 Description of study subjects

Colonna1	PanScan I–III and PanC4	PANDoRA	FinnGen	JaPAN	Total
Diagnosis					
Cases	8738	3047	881	2039	14705
Controls	7034	3225	204070	32592	246921
Total	15772	6272	204951	34631	261626
Median age (years)					
Cases	65	65	–	65	
Controls	65	57	–	51	
Sex (%)					
Male	53	49	–	57	
Female	47	51	–	43	

Note: In the data of the FinnGen project, information about age and sex is not present.
JaPAN, Japan Pancreatic Cancer Research; PanC4, Pancreatic Cancer Case–Control Consortium; PANDoRA, Pancreatic Disease Research; PanScan, Pancreatic Cancer Cohort Consortium.

other four, namely, mQTLdb,[40] Bios mQTL,[41] Go DMC[42] and the data produced by Hawe and colleagues (from now onward HJBlood),[43] identified mQTLs using blood samples. mQTLs were identified through PanCan meQTL, and to increase the chances that the annotation as mQTL of the identified variants was not due to chance, we analysed only those variants that have been reported as mQTLs in at least three of the other four databases. To prioritise SNPs based on their function, these mQTLs were investigated through GTEx (https://gtexportal.org/home/) to select those that are also expression quantitative trait loci (eQTLs) for the same gene in the pancreatic tissue. All mQTLs that showed a statistically significant association with PDAC risk ($p<0.05$) in all datasets of the discovery phase (PanScan I–III, PanC4 and PanScan I–III+PanC4) were selected to be validated in the replication phase using the summary statistics of FinnGen, JaPAN and de novo genotyping in PANDoRA. The details of the SNP selection and the workflow of the study are shown in figure 1.

Genotyping

DNA of PANDoRA samples was isolated from whole blood using QIAamp DNA extraction kit (Qiagen, California, USA) and distributed in 384-well plates for genotyping. For quality control, 8% of the samples were randomly replicated throughout the plates, and no-template controls were included in each plate. Genotyping was performed using TaqMan (ABI, Applied Biosystems, Foster City, California, USA) probes. A QuantStudio 5 instrument and QuantStudio software (Applied Biosystems) were used to detect the genotypes. After genotyping, deviation from HWE distribution was assessed in controls, considering the overall population and dividing by the country of origin of the samples. The concordance rate between the duplicated samples was 99.65%, and all the genotyped SNPs were in HWE.

Statistical analysis

The association of the SNPs with PDAC risk was assessed through unconditional logistic regression, adjusting by age, sex and the eight best principal components for PanScan and PanC4 and by age, sex and country of origin for PANDoRA (PANDoRA lacks GWAS data, therefore PCA cannot be performed). For FinnGen and JaPAN, summary statistics were instead used. A meta-analysis was performed using all the subjects (PanScan, PanC4, PANDoRA, FinnGen and JaPAN), using the fixed-effect or random-effect models, depending on evidence of heterogeneity. To account for multiple testing, we considered Linkage Disequilibrium (LD) ($r^2 > 0.6$) among the SNPs used in the discovery phase to obtain a list of independent variants (n=702), and the resulting Bonferroni-corrected threshold was $0.05/702=7.12\times10^{-5}$.

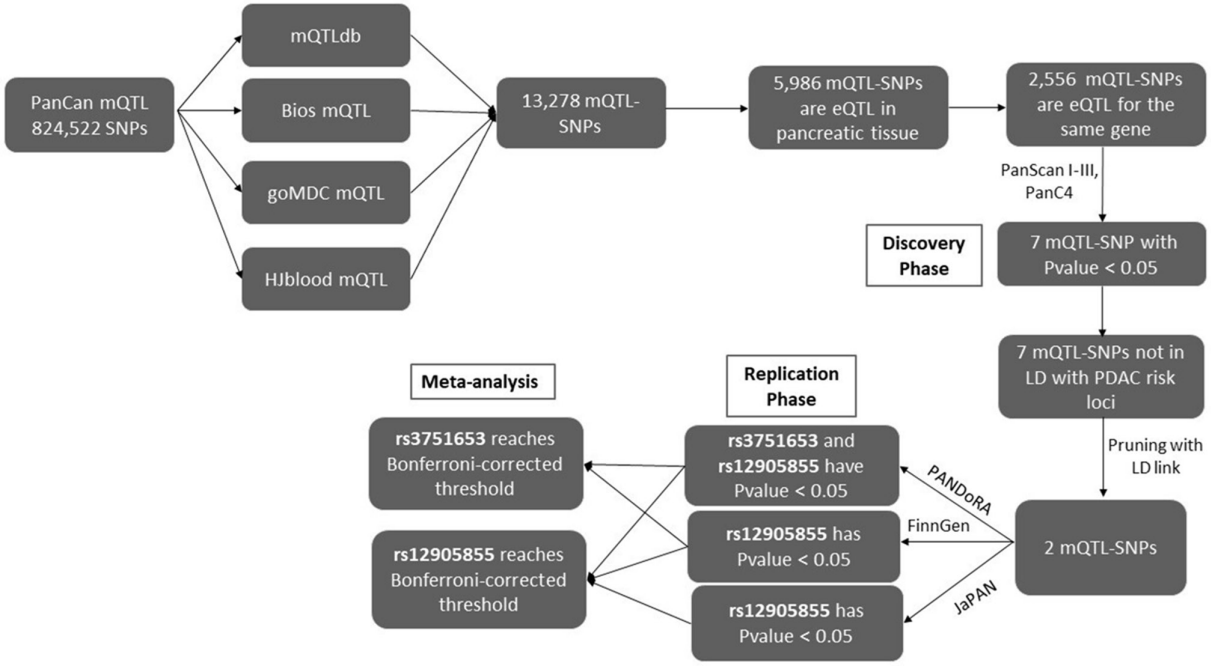

Figure 1 Workflow of the study. eQTL, expression quantitative trait locus; mQTL, methylation quantitative trait locus.

Functional evaluation

We used the Ensembl genome browser (https://www.ensembl.org/index.html) to identify the positions of CpG sites in the genome and to evaluate if their locations were inside of regulatory elements as promoter or enhancer. In addition, we used PanCan meQTL (http://bioinfo.life.hust.edu.cn/PancanmeQTL/) to evaluate the possible effect of the SNPs on gene methylation and GTEX portal to identify the potential SNPs' association with different gene expressions.[39]

RESULTS

A total of 13 278 variants were identified as mQTLs, among which 2556 are eQTLs for the same gene in the pancreatic tissue. Seven of those mQTLs showed a statistically significant association with PDAC risk in the discovery phase ($p<0.05$). After pruning for residual LD ($r^2>0.6$), two independent mQTLs (15q26.1-rs12905855 and 15q15.1-rs3751653) were selected to be replicated. The C allele of 15q26.1-rs12905855 showed a statistically significant association in PDAC risk in PANDoRA, FinnGen and JaPAN, with the effect direction consistent with the discovery phase. Instead, the C allele of 15q15.1-rs3751653 showed a statistically significant association in PANDoRA with an increase of PDAC risk but did not show an association in FinnGen and JaPAN. The two mQTLs were also tested in the Brazilian individuals of PANDoRA but did not show any association in this subgroup.

The C allele of the 15q26.1-rs12905855 showed an association at genome-wide level in the meta-analysis (OR=0.90, 95% CI 0.87 to 0.94, $p=4.93\times10^{-8}$), while the C allele of the 15q15.1-rs3751653 showed evidence for heterogeneity ($p_{heterogeneity}=0.04$ and $I^2=63\%$) and no association with risk ($p=0.17$). Considering the high heterogeneity for this SNP, Brazilian and East Asian subjects were removed, and a meta-analysis of PanScan, PanC4, PANDoRA and FinnGen was performed. This subgroup analysis showed a statistically significant increase in risk for carriers of the C allele (OR=1.09, 95% CI 1.05 to 1.13, $p=8.13\times10^{-6}$), which remained significant after correction for multiple testing. All the results are summarised in table 2.

According to PanCan meQTL, the C allele of 15q26.1-rs12905855 increases the methylation of the CpG site cg23684204 that is located in the promoter of the antisense of the RCC1 domain-containing (*RCCD1*) gene and increases the expression of the *RCCD1* gene. The C allele of 15q15.1-rs3751653 increases the methylation of CpG site cg13045913 that maps in the promoter of the zinc finger FYVE-type containing 19 (*ZFYVE19*) gene and, according to GTEx, decreases its expression (figure 2). The boxplots from PanCan meQTL and the violin plots from GTEx of two mQTLs are shown in figure 3A,B.

DISCUSSION

DNA methylation is variable throughout human life and across tissues; therefore, studies on DNA methylation are challenging in the context of an epidemiological setting. However, DNA methylation is at least partially regulated by SNPs called mQTLs, which can be used as surrogates of a direct measure. The role of mQTLs in cancer development has already been identified for bladder, breast, colorectal and pancreatic cancers.[32–34]

To investigate the association between mQTLs and PDAC, we conducted a two-phase study in 14 705 cases and 246 921 controls. The C allele of rs12905855 showed a clear association with decreasing risk of developing PDAC. The association was statistically significant in all studies with consistent ORs and reached genome-wide significance ($p=4.93\times10^{-8}$) in the meta-analysis. This mQTL regulates the methylation levels of the CpG site cg23684204, which is located in the promoter of the *RCCD1* antisense (*RCCD1-AS1*) gene. The C allele of this mQTL increases the methylation of cg23684204 and increases the expression of *RCCD1*. This gene encodes a protein that acts in a histone demethylase complex involved in chromosomal stability during mitosis.[44] Therefore, the molecular data obtained by

Table 2 Associations of 15q26.1-rs12905855 and 15q15.1-rs3751653 with PDAC risk

Chr	SNP	Position	M/m	Phase	OR (95% CI)	P value	I* (%)	P value Het
15	rs12905855	15q26.1	G/C	PanScan I–III/PanC4	0.93 (0.88 to 0.98)	2.40×10^{-3}		
				PANDoRA†§	0.89 (0.89 to 0.98)	1.80×10^{-2}		
				PANDoRA*§	0.91 (0.83 to 0.99)	4.00×10^{-2}		
				FinnGen	0.85 (0.76 to 0.94)	2.40×10^{-3}		
				JaPAN	0.89 (0.83 to 0.96)	3.40×10^{-3}		
				Meta-analysis†§	0.90 (0.87 to 0.94)	4.93×10^{-8}	0	4.50×10^{-1}
				Meta-analysis*§	0.91 (0.87 to 0.94)	1.42×10^{-7}	0	4.70×10^{-1}
15	rs3751653	15q15.1	T/C	PanScan I–III/PanC4	1.08 (1.04 to 1.13)	4.25×10^{-4}		
				PANDoRA†§	1.11 (1.03 to 1.20)	8.00×10^{-3}		
				PANDoRA*§	1.10 (1.02 to 1.18)	1.90×10^{-2}		
				FinnGen	1.00 (0.91 to 1.10)	9.60×10^{-1}		
				JaPAN	0.97 (0.95 to 1.12)	6.40×10^{-1}		
				Meta-analysis†§	1.04 (0.98 to 1.11)	1.70×10^{-1}	68	4.00×10^{-2}
				Meta-analysis*§	1.04 (0.98 to 1.11)	1.70×10^{-1}	61	5.00×10^{-2}
				Meta-analysis‡§	1.09 (1.05 to 1.13)	8.13×10^{-6}	0	7.90×10^{-1}

The analysis in PANDoRA was adjusted by age, sex and country of origin. Meta-analysis was performed by applying the fixed-effect model (rs12905855) or the random-effect model for the SNP showing heterogeneity (rs3751653).
*Brazilians were included in the analysis.
†Brazilians (69 cases and 258 controls) were excluded from the analysis.
‡Brazilians and JaPAN (2108 cases and 32 850 controls) were excluded from the analysis.
§The analysis included the German and Dutch controls from ESTHER and European Prospective Investigation on Cancer consortia.
Het, heterogeneity; JaPAN, Japan Pancreatic Cancer Research; m, minor allele; M, major allele; PanC4, Pancreatic Cancer Case–Control Consortium; PANDoRA, Pancreatic Disease Research; PanScan, Pancreatic Cancer Cohort Consortium; PDAC, pancreatic ductal adenocarcinoma.

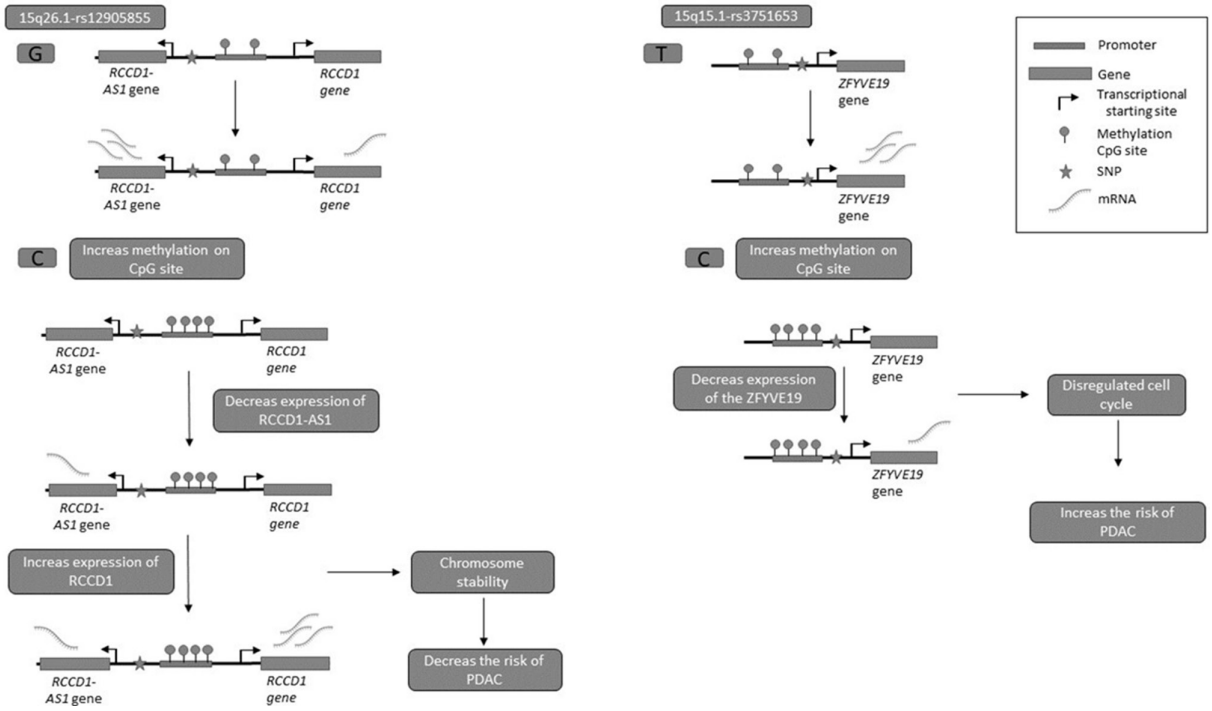

Figure 2 Schematic representation of the functional explanation of 15q26.1-rs12905855 and 15q15.1-rs3751653. PDAC, pancreatic ductal adenocarcinoma.

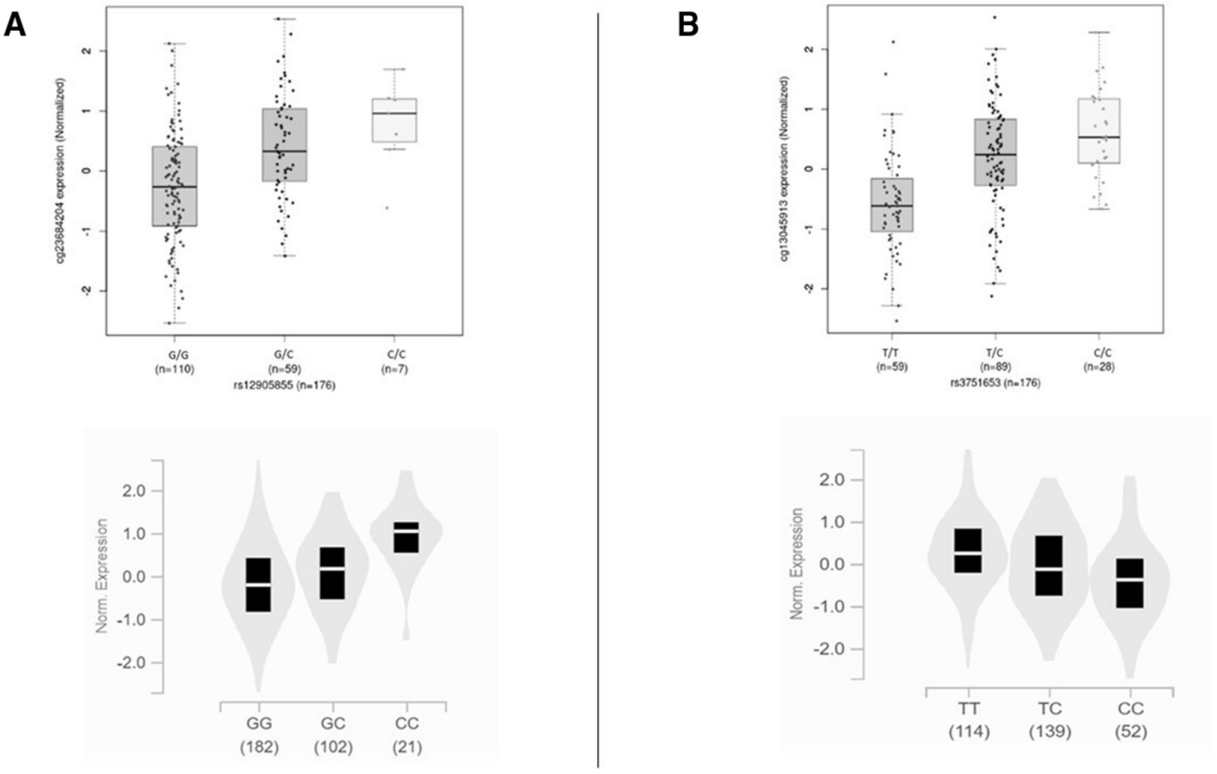

Figure 3 Relation between genotypes of SNPs and methylation of CpG site/gene expression. (A) Relation between genotypes of 15q26.1-rs12905855 and methylation of CpG site cg23684204/*RCCD1* expression. (B) Relation between genotypes of 15q15.1-rs3751653 and methylation of CpG site cg13045913/*ZFYVE19* expression. Note: PAAD=PDAC. The data used for the analysis described here were obtained from the PanCan meQTL database, downloaded on 11 April 2022, and GTEx analysis release V.8, accessed on 11 April 2022. meQTL, methylation expression quantitative trait locus; PAAD, pancreatic adenocarcinoma; PDAC, pancreatic ductal adenocarcinoma.

analysing together methylation (PanCan meQTL) and gene expression (GTEx) suggest that the protecting role of the rs12905855 C allele results from an increased methylation of *RCCD1-AS1*, which in turn leads to *RCCD1* increased expression, which contributes to chromosomal stabilising activity. Supporting this hypothesis, the results showed that higher expression of *RCCD1* gene has been associated with a decreased risk of breast and ovarian cancer.[45 46] Furthermore, a recent transcriptome-wide association study for PDAC has identified *RCCD1* as a possible risk locus.[47]

The other mQTL, 15q15.1-rs3751653, did not show a statistically significant association in the meta-analysis when all studies were analysed together. However, removing individuals with non-European ancestry from the analysis (East Asians and Brazilians), we found that the significance level of the association improved ($p=8.13\times10^{-6}$) and reached the Bonferroni-corrected threshold. This result could be explained by the possibility that 15q15.1-rs3751653 is not directly responsible for association but in LD with a causative SNP. Therefore, different LD architectures between populations could dilute the association when considering different ethnicities together.[48] According to PanCan meQTL, 15q15.1-rs3751653 regulates the methylation of the promoter of the *ZFYVE19* gene and its expression in the pancreatic tissue. *ZFYVE19* encodes a protein that regulates the abscission checkpoint of cytokinesis, delaying cell division in the presence of chromosome damage.[49] According to Thoresen and colleagues, depletion of the *ZFYVE19* gene is associated with accelerated timing on the cellular abscission, with a consequent increment of cell cytokinesis defects that may result in cancer development.[49] Therefore, the association of the C allele of 15q15.1-rs3751653 with increased risk of developing PDAC can be explained by a higher methylation level on the *ZFYVE19* gene promoter, resulting in lower expression of the gene in pancreatic tissue, with a consequent deregulation of the abscission checkpoint of cytokinesis in the pancreas.

A strength of this work is the study design, which included a discovery phase and a replication phase performed using 261626 individuals, making this study the most extensive mQTL scan performed on PDAC. Additionally, the molecular data obtained by PanCan meQTL and GTEx link the alleles both to methylation and gene expression changes that are consistent with the observed associations in the cellular context of the genes they regulate. It is also worth noting that the association of the C allele of rs12905855 with decreased risk is consistent in populations of different ancestries, decreasing the possibility of a spurious association.

The results of this study clearly highlight the importance of secondary analysis to discover new susceptibility loci of complex diseases. Neither SNPs were reported in the original studies because they focused on p value as the sole selection criteria to validate and report associations. Here, we show that combining GWAS data with functional data is an effective approach to discover new risk loci and to further our knowledge of disease biology.

In conclusion, our results point towards the regulation of DNA methylation through mQTL as a significant factor affecting the risk of developing in PDAC.

Author affiliations
[1]Department of Biology, University of Pisa, Pisa, Italy
[2]Division of Gastroenterology and Research Laboratory, IRCCS Ospedale Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy
[3]Department of Gastroenterology, Institute for Digestive Research, Medical Academy, Lithuanian University of Health Sciences, Kaunas, Lithuania
[4]Department of Surgery, Erasmus Medical Center, Erasmus University, Rotterdam, Netherlands
[5]Center for Traslational Medicine, Semmelweis University, Budapest, Hungary
[6]Division of Pancreatic Disease, Heart and Vascular Center, Semmelweis University, Budapest, Hungary
[7]ARC-NET, Centre for Applied Research on Cancer, University and Hospital Trust of Verona, Verona, Italy
[8]Division of General and Transplant Surgery, Pisa University Hospital, Pisa, Italy
[9]Department of Pathology, Amsterdam UMC, University of Amsterdam, Cancer Center Amsterdam, Amsterdam, Netherlands
[10]Division of Experimental Oncology, Gastroenterology and Gastrointestinal Endoscopy Unit, Vita-Salute San Raffaele University, IRCCS San Raffaele Scientific Institute, Milano, Italy
[11]University of Utrect, Utrecht, Netherlands
[12]Department of General Surgery, University of Heidelberg, Heidelberg, Germany
[13]Pancreato-Biliary Endoscopy and Endoscopic Ultrasound, Pancreas Translational and Clinical Research Center, IRSSC San Raffaele Scientific Institute, Milan, Italy
[14]Department of General, Visceral and Thoracic Surgery, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[15]Digestive and Liver Disease Unit, Sant'Andrea Hospital, Roma, Italy
[16]General Surgery, Department of Translational Research and New Technologies in Medicine and Surgery, Università di Pisa, Pisa, Italy
[17]Pancreatic Unit, IRCCS Humanitas Research Hospital, Rozzano, Italy
[18]Department of Biomedical Sciences, Humanitas University, Milan, Italy
[19]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
[20]Network Aging Research (NAR), Heidelberg University, Heidelberg, Germany
[21]Department of Digestive Tract Diseases, Medical University of Lodz, Lodz, Poland
[22]Institute for Translational Medicine, Medical School, University of Pécs, Pecs, Hungary
[23]Janos Szentagothai Research Center, University of Pecs, Pecs, Hungary
[24]Department of Surgery-DiSCOG, Padua University Hospital, Padova, Italy
[25]Endoscopic Unit, Department of Gastroenterology, IRCCS Humanitas Research Hospital, Rozzano, Italy
[26]Laboratory of Biology, Medical School, National and Kapodistrian University of Athens, Athens, Greece
[27]Division of Clinical Epidemiology and Aging Research, Cancer Research Center (DKFZ), Heidelberg, Germany
[28]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
[29]Carol Davila University of Medicine and Pharmacy, Bucarest, Romania
[30]Department of Surgery, Amsterdam UMC, University of Amsterdam, Cancer Center Amsterdam, Amsterdam, Netherlands
[31]Department of Surgery, Third Faculty of Medicine, University Hospital Kralovske Vinohrady, Charles University, Prague, Czech Republic
[32]Oncology of Massa Carrara, Oncological Department, Azienda USL Toscana Nord Ovest, Pisa, Italy
[33]Blood Transfusion Service, Azienda Ospedaliero Universitaria Meyer, Firenze, Italy
[34]Department of Molecular Biology of Cancer, Institute of Experimental Medicine Czech Academy of Sciences, Prague, Czech Republic
[35]Biomedical Centre and Department of Surgery, Faculty of Medicine in Pilsen, Charles University, Pilsen, Czech Republic
[36]First Faculty of Medicine, Institute of Biology and Medical Genetics, Charles University, Prague, Czech Republic
[37]Biomedical Center, Faculty of Medicine in Pilsen, Charles University, Pilsen, Czech Republic
[38]First Propaedeutic University Surgery Clinic, Hippocratio General Hospital, Medical School, National and Kapodistrian University of Athens, Athens, Greece
[39]Department of Medicine-DIMED, Padua University Hospital, Padova, Italy
[40]Laboratory for Applied Science and Technology in Health, Carlos Chagas Institute, Oswaldo Cruz Foundation (Fiocruz), Curitiba, Brazil
[41]County Medical Association of Potenza, Potenza, Italy
[42]Department of Oncology, Faculty of Medicine and Dentistry, Palacky University Olomouc, Olomouc, Czech Republic
[43]Department of Translational Research and new Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy
[44]Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany

Acknowledgements This research used genotyping data provided to the Pancreatic Disease Research consortium by the European Prospective Investigation on Cancer (EPIC) cohort, for which we would like to thank the contributors from EPIC NL. We also acknowledge the contributions of the late Dr Bas Bueno-de-Mesquita.

Contributors CC and GL are joint first authors, contributed equally, and performed the lab work and the analysis of the data. FC and DC are joint last authors. DCacts as guarantor. DC conceived and designed the study. CC, DC, MG, CR and FC wrote the first draft of the manuscript. All authors contributed to the writing and approved the final version of the manuscript.

Funding This work was supported by intramural funding of DKFZ (to FC); Fondazione Tizzi (www.fondazionetizzi.it); Fondazione Arpa (www.fondazionearpa.it, to DC); and Associazione Italiana per la Ricerca sul Cancro (AIRC IG 2021-26201, to GC). This work was supported by Italian Ministry of Health grants

(Ricerca Corrente 2022-2024) to Fondazione 'Casa Sollievo della Sofferenza' IRCCS Hospital, San Giovanni Rotondo (FGU), Italy and by the '5x1000' voluntary contribution.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants. The Pancreatic Cancer Cohort Consortium and Pancreatic Cancer Case–Control Consortium genotyping data are available from the Database of Genotypes and Phenotypes (study accession numbers phs000206.v5.p3 and phs000648.v1.p1).The PANDoRA primary data for this work will be made available to researchers who may submit a reasonable request to the corresponding author, conditional to approval by the PANDoRA Steering Committee and Ethics Committee of the Medical Faculty of Heidelberg University, Germany. All subjects provided written informed consent, and the ethical approval for the PANDoRA study protocol (including for controls from ESTHER and EPIC cohorts) was obtained from the Ethics Commission of the Medical Faculty of Heidelberg University.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iDs
Chiara Corradi http://orcid.org/0000-0001-6216-9803
Manuel Gentiluomo http://orcid.org/0000-0002-0366-9653
Maria Gazouli http://orcid.org/0000-0002-3295-6811
Daniele Campa http://orcid.org/0000-0003-3220-9944

REFERENCES

1 Sung H, Ferlay J, Siegel RL, *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;71:209–49.
2 Guo F, Chen C, Holleczek B, *et al*. Strong reduction of colorectal cancer incidence and mortality after screening colonoscopy: prospective cohort study from germany. *Am J Gastroenterol* 2021;116:967–75.
3 Kleeff J, Korc M, Apte M, *et al*. Pancreatic cancer. *Nat Rev Dis Primers* 2016;2:16022.
4 Huang L, Jansen L, Balavarca Y, *et al*. Resection of pancreatic cancer in Europe and USA: an international large-scale study highlighting large variations. *Gut* 2019;68:130–9.
5 Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, *et al*. Genome-Wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet* 2009;41:986–90.
6 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;42:224–8.
7 Childs EJ, Mocci E, Campa D, *et al*. Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. *Nat Genet* 2015;47:911–6.
8 Wolpin BM, Rizzato C, Kraft P, *et al*. Genome-Wide association study identifies multiple susceptibility loci for pancreatic cancer. *Nat Genet* 2014;46:994–1000.
9 Klein AP, Wolpin BM, Risch HA, *et al*. Genome-wide meta-analysis identifies five new susceptibility loci for pancreatic cancer. *Nat Commun* 2018;9:556.
10 Zhang M, Wang Z, Obazee O, *et al*. Three new pancreatic cancer susceptibility signals identified on chromosomes 1q32.1, 5p15.33 and 8q24.21. *Oncotarget* 2016;7:66328–43. 10.18632/oncotarget.11041 Available: https://www.oncotarget.com/lookup/doi/10.18632/oncotarget.v7i41
11 Corradi C, Gentiluomo M, Gajdán L, *et al*. Genome-wide scan of long noncoding RNA single nucleotide polymorphisms and pancreatic cancer susceptibility. *Int J Cancer* 2021;148:2779–88.
12 Campa D, Pastore M, Gentiluomo M, *et al*. Functional single nucleotide polymorphisms within the cyclin-dependent kinase inhibitor 2A/2B region affect pancreatic cancer risk. *Oncotarget* 2016;7:57011–20.
13 Campa D, Rizzato C, Stolzenberg-Solomon R, *et al*. TERT gene harbors multiple variants associated with pancreatic cancer susceptibility. *Int J Cancer* 2015;137:2175–83.
14 Gentiluomo M, Peduzzi G, Lu Y, *et al*. Genetic polymorphisms in inflammatory genes and pancreatic cancer risk: a two-phase study on more than 14 000 individuals. *Mutagenesis* 2019;34:395–401.
15 Gentiluomo M, Canzian F, Nicolini A, *et al*. Germline genetic variability in pancreatic cancer risk and prognosis. *Semin Cancer Biol* 2022;79:105–31.
16 Giaccherini M, Farinella R, Gentiluomo M, *et al*. Association between a polymorphic variant in the CDKN2B-AS1/ANRIL gene and pancreatic cancer risk. *Int J Cancer* 30, 2022.
17 Peduzzi G, Archibugi L, Katzke V, *et al*. Common variability in oestrogen-related genes and pancreatic ductal adenocarcinoma risk in women. *Sci Rep* 2022;12:18100.
18 Galeotti AA, Gentiluomo M, Rizzato C, *et al*. Polygenic and multifactorial scores for pancreatic ductal adenocarcinoma risk prediction. *J Med Genet* 2021;58:369–77.
19 Campa D, Gentiluomo M, Obazee O, *et al*. Genome-wide association study identifies an early onset pancreatic cancer risk locus. *Int J Cancer* 2020;147:2065–74. 10.1002/ijc.33004 Available: https://onlinelibrary.wiley.com/toc/10970215/147/8
20 Low S-K, Kuchiba A, Zembutsu H, *et al*. Genome-wide association study of pancreatic cancer in japanese population. *PLoS One* 2010;5:e11824.
21 Wu C, Miao X, Huang L, *et al*. Genome-Wide association study identifies five loci associated with susceptibility to pancreatic cancer in Chinese populations. *Nat Genet* 2012;44:62–6.
22 Lin Y, Nakatochi M, Hosono Y, *et al*. Genome-Wide association meta-analysis identifies GP2 gene risk variants for pancreatic cancer. *Nat Commun* 2020;11:3175.
23 Chen F, Childs EJ, Mocci E, *et al*. Analysis of heritability and genetic architecture of pancreatic cancer: a panc4 study. *Cancer Epidemiol Biomarkers Prev* 2019;28:1238–45.
24 Lu Y, Corradi C, Gentiluomo M, *et al*. Association of genetic variants affecting micrornas and pancreatic cancer risk. *Front Genet* 2021;12:693933.
25 Pistoni L, Gentiluomo M, Lu Y, *et al*. Associations between pancreatic expression quantitative traits and risk of pancreatic ductal adenocarcinoma. *Carcinogenesis* 2021;42:1037–45.
26 Natale F, Vivo M, Falco G, *et al*. Deciphering dna methylation signatures of pancreatic cancer and pancreatitis. *Clin Epigenetics* 2019;11:132.
27 González-Borja I, Alors-Pérez E, Amat I, *et al*. Deciphering Chfr role in pancreatic ductal adenocarcinoma. *Front Med* 2021;8:720128.
28 Lukosiute-Urboniene A, Mazeike A, Kazokaite M, *et al*. Epigenetic regulation of APAF-1 through DNA methylation in pancreatic cancer. *Anticancer Res* 2020;40:3765–79.
29 Joehanes R, Just AC, Marioni RE, *et al*. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet* 2016;9:436–47.
30 Gunasekara CJ, Waterland RA. A new era for epigenetic epidemiology. *Epigenomics* 2019;11:1647–9.
31 Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. *Nat Rev Genet* 2013;14:585–94.
32 Ho PJ, Dorajoo R, Ivanković I, *et al*. DNA methylation and breast cancer-associated variants. *Breast Cancer Res Treat* 2021;188:713–27.
33 Jordahl KM, Phipps AI, Randolph TW, *et al*. Mediation by differential dna methylation of known associations between single nucleotide polymorphisms and bladder cancer risk. *BMC Med Genet* 2020;21:228.
34 Heyn H, Sayols S, Moutinho C, *et al*. Linkage of dna methylation quantitative trait loci to human cancer risk. *Cell Rep* 2014;7:331–8.
35 Dai JY, Wang X, Wang B, *et al*. DNA methylation and cis-regulation of gene expression by prostate cancer risk snps. *PLoS Genet* 2020;16:e1008667.
36 Campa D, Rizzato C, Capurso G, *et al*. Genetic susceptibility to pancreatic cancer and its functional characterisation: the pancreatic disease research (pandora) consortium. *Dig Liver Dis* 2013;45:95–9.
37 Riboli E, Hunt K, Slimani N, *et al*. European prospective investigation into cancer and nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002;5:1113–24.
38 Kurki MI, Karjalainen J, Palta P, *et al*. FinnGen: unique genetic insights from combining isolated population and national health register data. *Genetic and Genomic Medicine* [Preprint] 2022.
39 Gong J, Wan H, Mei S, *et al*. Pancan-meqtl: a database to systematically evaluate the effects of genetic variants on methylation in human cancer. *Nucleic Acids Res* 2019;47:D1066–72.
40 the BIOS Consortium, Bonder MJ, Luijk R, *et al*. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* 2017;49:131–8.
41 Gaunt TR, Shihab HA, Hemani G, *et al*. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol* 2016;17:61.
42 Min JL, Hemani G, Hannon E, *et al*. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet* 2021;53:1311–21.
43 Hawe JS, Wilson R, Schmid KT, *et al*. Genetic variation influencing DNA methylation provides insights into molecular mechanisms regulating genomic function. *Nat Genet* 2022;54:18–29.
44 Marcon E, Ni Z, Pu S, *et al*. Human-chromatin-related protein interactions identify a demethylase complex required for chromosome segregation. *Cell Rep* 2014;8:297–310.
45 Hoffman JD, Graff RE, Emami NC, *et al*. Cis-eqtl-based trans-ethnic meta-analysis reveals novel genes associated with breast cancer risk. *PLoS Genet* 2017;13:e1006690.
46 Lu Y, Beeghly-Fadiel A, Wu L, *et al*. A transcriptome-wide association study among 97,898 women to identify candidate susceptibility genes for epithelial ovarian cancer risk. *Cancer Res* 2018;78:5419–30.
47 Zhong J, Jermusyk A, Wu L, *et al*. A transcriptome-wide association study identifies novel candidate susceptibility genes for pancreatic cancer. *J Natl Cancer Inst* 2020;112:1003–12.
48 Lonjou C, Zhang W, Collins A, *et al*. Linkage disequilibrium in human populations. *Proc Natl Acad Sci U S A* 2003;100:6069–74.
49 Thoresen SB, Campsteijn C, Vietri M, *et al*. ANCHR mediates aurora-B-dependent abscission checkpoint control through retention of VPS4. *Nat Cell Biol* 2014;16:550–60.