Diabetologia (2021) 64:1613–1625
https://doi.org/10.1007/s00125-021-05449-9

ARTICLE

Multi-omics analysis identifies CpGs near *G6PC2* mediating the effects of genetic variants on fasting glucose

Ren-Hua Chung[1] · Yen-Feng Chiu[1] · Wen-Chang Wang[2] · Chii-Min Hwu[3,4] · Yi-Jen Hung[5,6] · I-Te Lee[4,7,8] · Lee-Ming Chuang[9,10] · Thomas Quertermous[11] · Jerome I. Rotter[12,13] · Yii-Der I. Chen[12,13] · I-Shou Chang[14] · Chao A. Hsiung[1]

Received: 2 September 2020 / Accepted: 8 February 2021 / Published online: 12 April 2021

Abstract

Aims/hypothesis An elevated fasting glucose level in non-diabetic individuals is a key predictor of type 2 diabetes. Genome-wide association studies (GWAS) have identified hundreds of SNPs for fasting glucose but most of their functional roles in influencing the trait are unclear. This study aimed to identify the mediation effects of DNA methylation between SNPs identified as significant from GWAS and fasting glucose using Mendelian randomisation (MR) analyses.

Methods We first performed GWAS analyses for three cohorts (Taiwan Biobank with 18,122 individuals, the Healthy Aging Longitudinal Study in Taiwan with 1989 individuals and the Stanford Asia-Pacific Program for Hypertension and Insulin Resistance with 416 individuals) with individuals of Han Chinese ancestry in Taiwan, followed by a meta-analysis for combining the three GWAS analysis results to identify significant and independent SNPs for fasting glucose. We determined whether these SNPs were methylation quantitative trait loci (meQTLs) by testing their associations with DNA methylation levels at nearby CpG sites using a subsample of 1775 individuals from the Taiwan Biobank. The MR analysis was performed to identify DNA methylation with causal effects on fasting glucose using meQTLs as instrumental variables based on the 1775 individuals. We also used a two-sample MR strategy to perform replication analysis for CpG sites with significant MR effects based on literature data.

Results Our meta-analysis identified 18 significant ($p < 5 \times 10^{-8}$) and independent SNPs for fasting glucose. Interestingly, all 18 SNPs were meQTLs. The MR analysis identified seven CpGs near the *G6PC2* gene that mediated the effects of a significant SNP (rs2232326) in the gene on fasting glucose. The MR effects for two CpGs were replicated using summary data based on the European population, using an exonic SNP rs2232328 in *G6PC2* as the instrument.

✉ Ren-Hua Chung
rchung@nhri.org.tw

✉ Chao A. Hsiung
hsiung@nhri.org.tw

[1] Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Taiwan

[2] The Ph.D. Program for Translational Medicine, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

[3] Section of Endocrinology and Metabolism, Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan

[4] School of Medicine, National Yang-Ming University, Taipei, Taiwan

[5] Division of Endocrine and Metabolism, Tri-Service General Hospital, Taipei, Taiwan

[6] Institute of Preventive Medicine, National Defense Medical Center, Taipei, Taiwan

[7] Division of Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan

[8] School of Medicine, Chung Shan Medical University, Taichung, Taiwan

[9] Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan

[10] Institutes of Molecular Medicine, Collage of Medicine, National Taiwan University, Taipei, Taiwan

[11] Division of Cardiovascular Medicine and Stanford Cardiovascular Institute, Falk Cardiovascular Research Center, Stanford University, Stanford, CA, USA

[12] Institute for Translational Genomics and Population Sciences, the Lundquist Institute, Harbor-UCLA Medical Center, Torrance, CA, USA

[13] Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA, USA

[14] National Institute of Cancer Research, National Health Research Institutes, Zhunan, Taiwan

Research in context

What is already known about this subject?

- An elevated fasting glucose level in non-diabetic individuals is a key predictor of type 2 diabetes
- Genome-wide association studies (GWAS) performed in individuals with predominantly European ancestry have identified more than 40 genetic loci for fasting glucose

What is the key question?

- Are the CpG sites near GWAS-significant SNPs mediators of the effects of the SNPs on fasting glucose?

What are the new findings?

- All 18 GWAS-significant SNPs identified by our study using East Asian samples were methylation quantitative trait loci
- GWAS-significant SNPs in *G6PC2* (rs2232326 and rs2232328) may affect DNA methylation at CpGs near the gene, and the methylation may have downstream effects on fasting glucose

How might this impact on clinical practice in the foreseeable future?

- Our analysis provides new insights into the functional roles of GWAS-significant SNPs for fasting glucose; the GWAS-significant SNPs identified by our study and the CpGs mediating the SNP effects will be informative for constructing a genetic risk model for predicting future type 2 diabetes

Conclusions/interpretation Our analysis results suggest that rs2232326 and rs2232328 in *G6PC2* may affect DNA methylation at CpGs near the gene and that the methylation may have downstream effects on fasting glucose. Therefore, SNPs in *G6PC2* and CpGs near *G6PC2* may reside along the pathway that influences fasting glucose levels. This is the first study to report CpGs near *G6PC2*, an important gene for regulating insulin secretion, mediating the effects of GWAS-significant SNPs on fasting glucose.

Keywords DNA methylation · Fasting glucose · GWAS · Han Chinese · Mendelian randomisation · Multi-omics analysis

Abbreviations

G6PC2	Glucose 6-phosphatase catalytic subunit 2
GRS	Genetic risk score
GWAS	Genome-wide association studies
HALST	Healthy Aging Longitudinal Study in Taiwan
LD	Linkage disequilibrium
MAF	Minor allele frequency
meQTL	Methylation quantitative trait locus
MR	Mendelian randomisation
SAPPHIRe	Stanford Asia-Pacific Program for Hypertension and Insulin Resistance
SMR	Summary data based on MR
TWB	Taiwan Biobank

Introduction

Fasting glucose level is a glycaemic trait that when elevated in non-diabetic individuals can be a strong predictor for type 2 diabetes [1]. Several genome-wide association studies (GWAS) have been performed for fasting glucose in non-diabetic individuals with predominantly European ancestry [2–7]. These studies have identified SNPs in over 40 genetic loci that have implications in fasting glucose and many of these SNPs are also risk variants for type 2 diabetes. Hence, understanding the genetic factors that influence fasting glucose can help identify risk loci for type 2 diabetes and pathways implicated in glucose homeostasis [4, 8].

Many genetic risk models have been created for type 2 diabetes. However, due to the small effect sizes of the currently identified type 2 diabetes variants, these genetic models add little prediction value to the clinical models considering variables such as age, sex, BMI and family history [9]. One possible improvement in the prediction accuracy of the genetic risk model is to include new genetic variants relevant to type 2 diabetes identified in non-European ancestry populations [9]. Moreover, differential DNA methylation at CpG sites associated with genetic variants have been shown to affect the risk of type 2 diabetes [10–12]. Therefore, an integrative analysis of GWAS and epigenetics data may provide further insight into the pathogenesis of type 2 diabetes and improve current genetic risk models [9]. To our knowledge, only a few GWAS for blood glucose have been performed in East Asians [13–16]. Furthermore, several epigenome-wide association studies (EWAS) have been performed to examine the associations between DNA methylation levels at genome-wide

CpGs and fasting glucose [17–20]. However, the causal relationships between the CpGs and fasting glucose have not been fully investigated.

Mendelian randomisation (MR) has become a popular approach to investigate whether DNA methylation acts as a mediator between genetic variants and the complex trait in a causal pathway [21]. For example, MR was used to identify the mediation effects of CpGs on type 2 diabetes and CVD using the complex trait as the outcome, DNA methylation as the risk factor, and SNPs as instrumental variables [22, 23]. Recently, Liu et al [24] performed a cross-omics analysis for fasting glucose and used MR to identify a CpG at the *SLAMF1* gene nominally associated with fasting glucose using the genetic risk score as the instrumental variable in samples of European ancestry. However, to our knowledge, no similar analysis that integrates GWAS and epigenetics data to investigate the causal roles of CpGs on fasting glucose has been reported for the East Asian population.

In this study, we aimed to identify the mediation effects of CpGs between genetic variants and fasting glucose using samples of Han Chinese descent in Taiwan. We hypothesised that SNPs identified as significant from GWAS and nearby CpGs affected by the SNPs may reside along the same pathway to influence fasting glucose levels. Our analysis results can provide new biological insights into the functional roles of GWAS-significant SNPs for fasting glucose and mechanisms for type 2 diabetes.

Methods

Cohort descriptions and fasting glucose measurement The discovery dataset for the GWAS analyses consisted of three cohorts, from the Taiwan Biobank (TWB) [25], the Healthy Aging Longitudinal Study in Taiwan (HALST) [26] and the Stanford Asia-Pacific Program for Hypertension and Insulin Resistance (SAPPHIRe) [27] studies. Written informed consent for participation was obtained from all participants and the analysis was approved by the Institutional Review Board of the National Health Research Institutes in Taiwan. The TWB and HALST studies are both population-based longitudinal studies, while the SAPPHIRe study is comprised of concordant and discordant sib pairs for hypertension. The replication dataset (TWB2) for the GWAS results was another batch of cohorts obtained from the TWB. Detailed descriptions of the three cohorts are provided in electronic supplementary material (ESM) Methods.

Fasting glucose in mmol/l was measured from fasting serum in the TWB and HALST studies and from fasting plasma in the SAPPHIRe study. The fasting blood specimens (up to 30 ml) were collected in the morning after participants had fasted overnight for >8 h. Individuals were excluded if they had fasting glucose ≥7 mmol/l, HbA_{1c} ≥48 mmol/mol (6.5%), type 1 diabetes, type 2 diabetes or gestational diabetes (diagnosed, self-reported or on diabetes treatment). Fasting glucose values were inverse normal transformed for analyses.

Genotyping, imputation and quality control Samples from the TWB were genotyped using the TWB chip, which is a customised Affymetrix Axiom Genome-Wide Array consisting of approximately 653,000 SNPs [28]. The SNPs included a subset of SNPs on the Axiom Genome-Wide CHB 1 Array, SNPs that are reported to be ancestry informative, significant SNPs from other GWAS, and pharmacogenetics markers on the Affymetrix DMET Plus array. The HALST samples were genotyped based on the Metabochip [29], from which approximately 200,000 candidate SNPs for metabolic, cardiovascular and anthropometric traits were selected. Moreover, the SAPPHIRe samples were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0, which contains approximately 900,000 SNPs. The SNP genotypes and samples then underwent standard GWAS quality control procedures for each cohort study. Briefly, SNPs with call rates <95% and with Hardy–Weinberg equilibrium *p*-values $<10^{-4}$ were excluded [6]. Moreover, samples with call rates <95% [6], duplicated samples (the PLINK [30] pi_hat statistics >0.9 [31]), or samples who failed the PLINK sex check were excluded. For replicated samples, the sample with the highest call rate was kept. Furthermore, if the median of the PLINK pi_hat statistics of a sample with all other samples was >0.05, the sample was removed due to a possible sample contamination [32]. The aforementioned analyses were performed using PLINK v1.9 [33]. Principal components were calculated using the PC-AiR method, which accounted for sample relatedness, implemented in the R (v3.5.1) [34] package GENESIS (v2.14.3) [35]. The principal components were included as covariates in statistical analyses to adjust for population stratification. Samples in each cohort were then imputed against the reference panel consisting of haplotypes of 2504 samples from the 1000 Genomes Project [36] using the Michigan Imputation Server. SNPs with minor allele frequency (MAF) >1% and the imputation quality measure RSQ >0.9 [37] were used for the association analyses. After quality control, the TWB had 18,122 individuals, HALST had 1989 individuals, SAPPHIRe had 416 individuals (in 179 families) and TWB2 had 6955 individuals. There were 5,151,227, 374,794, 5,084,103 and 5,050,043 directly genotyped and imputed SNPs with MAFs >1% and imputation RSQ >0.9 in the TWB, HALST, SAPPHIRe and TWB2 datasets, respectively. Demographics, such as the proportions of sex and hypertension status, and the means of age, BMI and fasting glucose for the four datasets, are provided in ESM Table 1.

DNA methylation assays Blood samples were randomly selected from 2091 TWB participants who had been

genotyped by the GWAS arrays for genome-wide methylation profiling using the Illumina Infinium MethylationEPIC BeadChip [38]. The R package Bigmelon (v1.12.0) [39], one of the few tools that can process the MethylationEPIC chip data, was used for the sample and CpG-level quality control, and the quality control procedures as suggested in the Bigmelon paper [39] were followed. Briefly, outlier samples were first removed based on the principal component analysis [39]. Samples with <85% bisulphite conversion values, a conservative threshold as suggested in Bigmelon [39], were removed. Data were then normalised using the Dasen function in the wateRmelon R package (v1.30.0) [40]. The Qual function in Bigmelon was used to further remove samples whose normalised values changed significantly compared with the raw values. Samples with 1% of CpG sites with a detection p value >0.05 were also removed [39]. CpG sites with less than three bead counts in 5% of samples or with 1% of samples having a detection p value >0.05 were then removed [39, 41]. The data were then normalised again using the Dasen function. Outlier CpG values, defined as values beyond 4 × IQR, were set to missing using the Pwod function in WateRmelon, and samples or CpGs with missing rates >5% were removed [39]. Finally, CpG probes with common variants (e.g. SNPs, indels or structural variations with frequencies >1%) and probes with cross-hybridisation potential identified by McCartney et al [42], as well as CpGs on the sex chromosomes, were removed [41]. After quality control, 774,398 CpGs in 1775 samples remained for the analysis. Demographics for the 1775 samples are shown in ESM Table 2. The distributions of the variables were similar to those observed for TWB and TWB2 (ESM Table 1).

GWAS analyses Single SNP association analysis was performed for each cohort using the software EMMAX (v20120210) [43], which implements a linear mixed model-based approach accounting for sample relatedness, with covariates including age, age^2, BMI, sex, hypertension status, genotyping batch and the top 10 principal components. The fixed effects in the linear mixed model included the overall mean, covariates and the tested SNP, and the random effect was the polygenic background where its variance accounted for sample relationships using a kinship matrix. This linear mixed model framework has been used in other studies [44–46]. Wald tests based on the linear mixed model using the empirical kinship (i.e. the Balding–Nichols matrix estimated in EMMAX) were performed and genotype dosages were used in the model to account for imputation uncertainty. Then, the SNPs with RSQ >0.9 in each cohort were included in the meta-analysis. Therefore, meta-analysis for a SNP may include one, two or three cohort results, depending on the RSQ value for the SNP in each cohort. The meta-analysis was performed using META (v1.7) [47] with the inverse-variance method based on a fixed-effects model. SNPs were significant if their p values from the meta-analysis were $<5 \times 10^{-8}$. A replication analysis was performed for the significant SNPs using the TWB2 samples. We also performed a conditional analysis using the meta-analysis results to identify novel SNPs using GCTA (v1.91.1) [48]. A list of literature SNPs for fasting glucose was compiled based on the GWAS catalogue (https://www.ebi.ac.uk/gwas/; v1.0) [49]. We also extracted SNPs with p values $<10^{-6}$ for fasting glucose from the Type 2 Diabetes Knowledge Portal (https://t2d.hugeamp.org/, accessed 4 February 2020) and from the summary data of different studies including Dupuis et al [4], Scott et al [5], Manning et al [6], and Lagou et al [50] downloaded from the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) [51] website (https://magicinvestigators.org/downloads/, accessed 4 February 2020). A total of 1133 literature SNPs with results in our meta-analysis were extracted. Because many of the 1133 SNPs were in linkage disequilibrium (LD), there would be collinearity if all of them were included in a regression model. Therefore, PLINK was first used to prune for LD so that the LD r^2 was <0.98 for each pair of SNPs after pruning. LD was measured based on the TWB2 sample. A total of 426 SNPs remained and then the conditional analysis was performed for each significant SNP conditional on the 426 SNPs. The TWB2 was used as the reference sample for GCTA. Furthermore, to identify SNPs with independent signals in the significant SNPs, the stepwise model selection procedure implemented in GCTA was performed using the meta-analysis results. The phenotypic variance explained by the independent SNPs was also calculated by GCTA using the restricted maximum likelihood analysis [52]. We then constructed a genetic risk score (GRS), which was the weighted sum of the risk alleles with the weights proportional to the effect sizes of the risk alleles [4], using the significant and independent SNPs. The effect sizes for calculating the GRS were obtained from the meta-analysis and the GRS was calculated for each TWB2 sample.

MR analysis for the causal effects of CpGs on fasting glucose We followed the procedures described by Richardson et al [22] to perform the MR analysis. These authors [22] suggest four possible explanations when observing the associations between methylation quantitative trait loci (meQTLs; SNPs influencing DNA methylation level) and the complex trait. Briefly, explanation 1 is our hypothesised relationship wherein the SNP affects fasting glucose, mediated by DNA methylation (Fig. 1a). For explanation 2, the SNP affects fasting glucose through other mechanisms, and fasting glucose affects the DNA methylation (Fig. 1b). In explanation 3, the SNP affecting DNA methylation is in LD with another SNP affecting fasting glucose (Fig. 1c). Finally, explanation 4 suggests that the SNP has separate effects on DNA

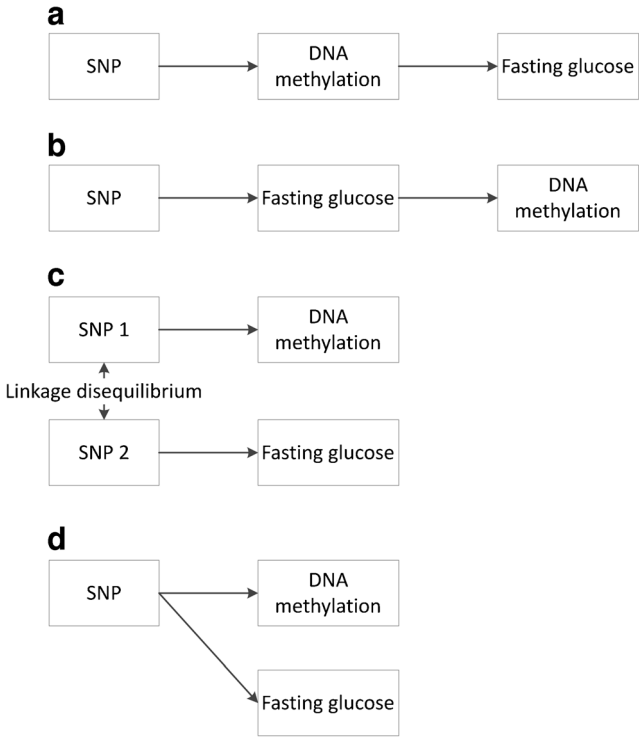

Fig. 1 The four possible explanations when observing associations between meQTLs and fasting glucose. (**a**) Explanation 1: the SNP affects fasting glucose through the mediation of the DNA methylation. (**b**) Explanation 2: the SNP affects fasting glucose through other mechanisms and fasting glucose affects the DNA methylation. (**c**) Explanation 3: the SNP affecting the DNA methylation is in LD with another SNP affecting fasting glucose. (**d**) Explanation 4: the SNP has separate effects on the DNA methylation and fasting glucose

methylation and fasting glucose through different pathways, also known as horizontal pleiotropy (Fig. 1d). The advantages of the procedures proposed by Richardson et al [22] are that they can distinguish explanation 1 from explanations 2 and 3. To be more specific, we first investigated whether the significant SNPs identified by the GWAS analyses can influence DNA methylation levels at nearby CpGs (i.e. whether the SNPs are *cis*-meQTLs). We extracted CpG sites that were within 300 kb of each significant SNP from the 1775 samples with both GWAS and methylation array data available, resulting in 3755 SNP–CpG pairs. A linear regression model accounting for age, age^2, BMI, sex, methylation assay batch and the top 10 principal components calculated by the methylation data was used to test the association between a SNP and the DNA methylation levels at its nearby CpGs. SNPs with significant effects on CpGs (i.e. meQTLs) passing the Bonferroni correction threshold (i.e. 1.33×10^{-5} considering 3755 tests) were extracted, resulting in 216 SNP–CpG pairs that were used for the following MR analysis. As the MR analysis assumes that the instrumental variable (i.e. a SNP) is independent of the factors that confound the association between the risk factor (i.e. methylation) and outcome (i.e. fasting glucose), we used generalised linear regression to evaluate the associations between SNP genotypes and potential confounders, such as age, sex and BMI. For the MR analysis, a two-stage least-squares regression model implemented in the R package Systemfit (v1.1.24) [53] was used, where the meQTL was used as the instrumental variable, DNA methylation influenced by the meQTL as the risk factor, and fasting glucose as the outcome. The two-stage least-squares regression analysis identified seven CpGs with significant MR effects passing the Bonferroni correction threshold (i.e. 2.31×10^{-4} for 216 MR tests). The seven CpGs were subsequently evaluated using the reverse MR analysis, by using the GRS as the instrumental variable, fasting glucose as the risk factor, and DNA methylation at the CpGs as the outcome in the 1775 samples, similar to the analysis described by Wahl et al [54]. The reverse MR analysis was used to investigate whether the CpGs with significant MR effects were actually affected by fasting glucose levels (Fig. 1b), rather than our hypothesis that fasting glucose was affected by the DNA methylation levels. CpGs with p value <0.05 for the reverse MR test were excluded from further analysis. We then used the bivariate fine-mapping analysis to evaluate whether the DNA methylation was influenced by a genetic variant that is simply in LD with another variant influencing fasting glucose (Fig. 1c). The software FINEMAP (v1.4) [55] was used to calculate the posterior probability that an meQTL or any SNP in LD with the meQTL is the causal SNP. LD measures were calculated using TWB2 and effect estimates of meQTLs were obtained from the meQTL analysis using the 1775 samples. To replicate the significant MR results, a two-sample MR analysis strategy using published data [56] was performed. The effect estimates of SNPs for fasting glucose were obtained from the literature SNPs as described in the GWAS analysis section. The effect estimates of SNPs for DNA methylation were downloaded from the meQTL database (summary data based on MR [SMR]) [41] based on results from 1175 UK samples, where DNA methylation levels were also profiled using the MethylationEPIC chip. The estimate of causal effect for a CpG on fasting glucose was calculated using the Wald ratio method, which is the effect estimate of a SNP for fasting glucose divided by the effect estimate of the SNP for the CpG [57]. The two-sample MR analysis results were considered as significant if their p values were $<7.14 \times 10^{-3}$ for replicating the seven CpGs identified using our sample.

Results

Figure 2 shows the flowchart of our analysis strategy, similar to that of Richardson et al [22].

GWAS analyses For the GWAS analyses (step 1 in Fig. 2), the discovery dataset consisted of three cohorts from the TWB,

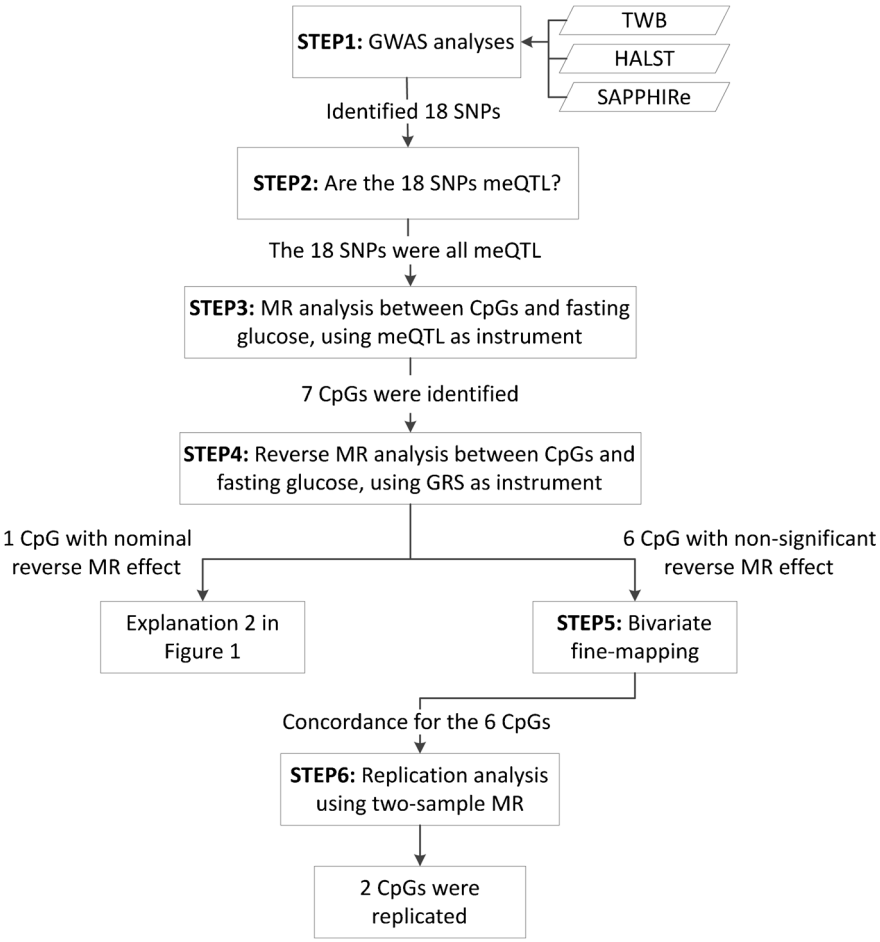

Fig. 2 Flowchart of the analysis procedures

HALST and SAPPHIRe studies, and the replication dataset is from the TWB2. The quantile–quantile (Q-Q) plots for the association p values from individual cohorts are shown in ESM Fig. 1. In the plots, the p values from the HALST and SAPPHIRe analyses generally followed the expected lines under the null, while an excess of small p values was observed from the TWB analysis. The results suggest that the TWB study with a larger sample size showed higher power than the HALST and SAPPHIRe studies, which had smaller sample sizes. The genomic inflation factor (λ) for TWB, HALST and SAPPHIRe was 0.984, 0.999, and 1.002, respectively, suggesting that confounding factors such as population substructures and batch effects were controlled. The p values from individual cohorts were adjusted by λ for the meta-analysis.

The significant SNPs from the meta-analysis were annotated using ANNOVAR [58] based on the hg19 coordinates. Then SNPs located within 1 MB to a previously reported locus of fasting glucose were annotated to the locus. For a Manhattan plot of the meta-analysis results, showing the lead SNPs (i.e. the most significant SNP passing the genome-wide threshold of $p < 5 \times 10^{-8}$ in each loci), see ESM Fig. 2. To be more specific, there were 827 SNPs with p values $<5 \times 10^{-8}$ in the meta-analysis results. They were all located in previously reported loci. The stepwise model selection suggests that there were 18 independent SNPs among these 827 SNPs. The 18 SNPs all had p values $<6.04 \times 10^{-5}$ (accounting for the multiple testing of 827 SNPs) in the joint model identified by the stepwise model selection. Fifteen of the 18 SNPs were replicated in TWB2 with p values $<2.78 \times 10^{-3}$ (accounting for the multiple testing of 18 SNPs). The conditional analysis based on the 426 literature SNPs showed that one of the 18 SNPs had a p value of 0.039 but the SNP would not be significant using the Bonferroni correction threshold of 2.78×10^{-3}. The meta-analysis, replication analysis, conditional analysis and joint analysis results for the 18 SNPs are shown in Table 1. The correlation structures of the 18 SNPs are shown in ESM Table 3. For SNPs that are in the same locus such as *G6PC2*, the pairwise LD (measured by r^2 based on the TWB2 sample) among them was low (<0.05). The 18 independent SNPs explained approximately 7.1% (SE 2.27%) of the phenotypic variance calculated by GCTA based on the TWB2 replication sample. This estimate was slightly higher than the estimate of 3.2–4.4% using 14 SNPs in Dupuis et al [4]. ESM Fig. 3 shows the mean fasting glucose levels categorised by GRS calculated based on the 18 SNPs in the TWB2 sample.

Table 1 Meta-, replication, conditional and joint analysis results for the 18 independent SNPs

Chr	Position	SNP	Locus[a]	Allele (ref/effect)	Freq[b]	Meta analysis[c]	I_2 (p value)[d]	Replication analysis[c]	Conditional analysis[c]	Joint analysis[c]
2	27741237	rs780094	*GCKR*	T/C	0.52	0.080 (0.010) 1.39×10^{-15}	41.80 (0.18)	0.053 (0.017) 1.94×10^{-3}	NA	0.080 (0.010) 1.53×10^{-15}
2	45188353	rs895636	*SIX2-SIX3*	C/T	0.42	0.110 (0.011) 5.14×10^{-25}	0.00 (0.82)	0.109 (0.017) 2.73×10^{-10}	0.009 (0.011) 4.16×10^{-1}	0.110 (0.011) 6.96×10^{-25}
2	169756930	rs540524	*G6PC2*	A/G	0.40	−0.061 (0.010) 3.02×10^{-9}	54.82 (0.11)	−0.033 (0.018) 6.51×10^{-2}	−0.004 (0.010) 6.79×10^{-1}	−0.113 (0.011) 8.95×10^{-27}
2	169763148	rs560887	*G6PC2*	T/C	0.97	0.246 (0.027) 1.66×10^{-19}	8.89 (0.33)	0.212 (0.046) 3.49×10^{-6}	NA	0.319 (0.028) 1.04×10^{-30}
2	169764491	rs2232326	*G6PC2*	T/C	0.05	−0.399 (0.024) 3.99×10^{-62}	0.00 (0.85)	−0.416 (0.043) 1.12×10^{-21}	NA	−0.457 (0.025) 4.22×10^{-77}
2	173591412	rs56960326	*RAPGEF4-AS1*	T/G	0.53	0.059 (0.011) 2.48×10^{-8}	65.44 (0.09)	0.062 (0.017) 2.70×10^{-4}	0.004 (0.011) 7.08×10^{-1}	0.057 (0.011) 7.87×10^{-8}
5	95630705	rs17085593	*LOC101929710*	C/G	0.33	−0.068 (0.011) 6.08×10^{-10}	0.00 (0.53)	−0.072 (0.018) 6.46×10^{-5}	−0.004 (0.011) 7.30×10^{-1}	−0.068 (0.011) 6.31×10^{-10}
7	14898282	rs17168486	*DGKB*	C/T	0.48	0.065 (0.010) 9.71×10^{-11}	0.00 (0.77)	0.074 (0.017) 1.69×10^{-5}	NA	0.059 (0.010) 4.40×10^{-9}
7	15066045	rs1990379	*DGKB*	T/A	0.69	0.089 (0.011) 2.02×10^{-15}	0.00 (0.69)	0.090 (0.019) 3.49×10^{-6}	0.006 (0.011) 5.68×10^{-1}	0.084 (0.011) 9.43×10^{-14}
7	44178626	rs1008384	*MYL7*	A/G	0.24	−0.090 (0.013) 2.27×10^{-12}	57.07 (0.13)	−0.063 (0.021) 2.75×10^{-3}	−0.026 (0.013) 3.90×10^{-2}	−0.076 (0.013) 3.11×10^{-9}
7	44229068	rs1799884	*GCK*	C/T	0.19	0.144 (0.013) 4.71×10^{-30}	0.00 (0.97)	0.180 (0.021) 4.2310^{-17}	0.002 (0.013) 8.59×10^{-1}	0.136 (0.013) 8.13×10^{-27}
8	118185025	rs3802177	*SLC30A8*	G/A	0.46	−0.066 (0.010) 5.49×10^{-11}	0.00 (0.97)	−0.093 (0.017) 5.86×10^{-8}	NA	−0.066 (0.010) 5.75×10^{-11}
9	571582	rs12684488	*KANK1*	T/A	0.26	0.089 (0.012) 1.20×10^{-13}	0.00 (0.69)	0.067 (0.020) 7.09×10^{-4}	0.016 (0.012) 1.86×10^{-1}	0.089 (0.012) 1.66×10^{-13}
9	4285119	rs3934283	*GLIS3*	C/G	0.46	0.061 (0.010) 1.08×10^{-9}	35.59 (0.21)	0.065 (0.017) 1.51×10^{-4}	0.005 (0.010) 5.87×10^{-1}	0.061 (0.010) 1.43×10^{-9}
9	22134253	rs10811662	*CDKN2B-AS1*	G/A	0.41	−0.067 (0.010) 4.58×10^{-11}	24.55 (0.27)	−0.046 (0.017) 7.76×10^{-3}	NA	−0.067 (0.010) 4.79×10^{-11}
11	92708710	rs10830963	*MTNR1B*	C/G	0.44	0.105 (0.010) 5.34×10^{-26}	0.00 (0.71)	0.153 (0.017) 4.11×10^{-19}	NA	0.105 (0.010) 7.15×10^{-26}
13	28492405	rs61944004	*PLUT*	G/A	0.41	0.066 (0.011) 7.97×10^{-10}	NA	0.041 (0.017) 1.68×10^{-2}	0.009 (0.011) 3.95×10^{-1}	0.066 (0.011) 8.28×10^{-10}
20	22560932	rs72470563	*FOXA2*	T/A	0.14	−0.091 (0.014) 2.55×10^{-10}	0.00 (0.93)	−0.130 (0.025) 1.40×10^{-7}	NA	−0.091 (0.014) 2.65×10^{-10}

[a] The nearest gene to the SNP

[b] Frequency of the effect allele

[c] Effect size, SE and p value. NA for a SNP in the conditional analysis means that the SNP was from the literature

[d] The heterogeneity index and the p value for heterogeneity

I_2, heterogeneity index

Similar to the observation made by Dupuis et al [4], higher fasting glucose levels were generally observed in individuals with higher GRS.

meQTL evaluation We then evaluated whether the 18 SNPs were meQTLs (step 2 in Fig. 2). Table 2 shows the CpGs with the smallest *p* values that were associated with the 18 significant SNPs. Interestingly, all the 18 SNPs had a significant *p* value, with at least one nearby CpG, suggesting that they were all meQTLs. One study suggested that meQTLs are enriched in GWAS SNPs with low *p* values for metabolic traits such as LDL-cholesterol and BP [59]. In addition, approximately 60% of SNPs in the GWAS catalogue were found to be *cis*-meQTLs [60]. The results of our meQTL analysis supported the findings in the literature, implying that the 18 significant SNPs may be involved in pathways affecting fasting glucose with effects on DNA methylation levels at local CpGs.

Moreover, 16 of the 18 SNPs were reported as meQTLs based on our search in public meQTL databases (ESM Table 4). The two SNPs not reported in the databases were rare in the European population (MAF <0.5%). Therefore, the power to detect the effects of the two SNPs on DNA methylation in the European population may be low or the effects may be specific to the East Asian population.

MR analysis The significant SNP–CpG pairs were then used for the MR analysis (step 3 in Fig. 2). Table 3 shows the MR analysis results for seven CpGs with significant causal effects on fasting glucose using the meQTL as the instrumental variable. The seven significant CpGs were generally upstream or downstream of the *G6PC2* gene and were all associated with rs2232326 located in an exon of the gene. It has been found that multiple CpGs can share the same genetic effect [41]. Genotypes at rs2232326 were not associated with potential confounders of age, age^2, sex and BMI (ESM Table 5). The DNA methylation rates and the correlations of the DNA methylation levels for the seven CpGs are shown in ESM Table 6. The correlation structures suggest that there are four methylation blocks. The seven CpGs were mostly partially methylated (methylation rates between 0.3 and 0.8; one had a methylation rate of 0.88) and three CpGs were highly correlated (correlation coefficients >0.8).

Reverse MR analysis To further investigate whether there was a reverse causal relationship between the DNA methylation and fasting glucose (i.e. explanation 2 in Fig. 1b), a reverse MR analysis was performed using the DNA methylation at the seven CpGs as the outcome, fasting glucose as the risk factor, and GRS as the instrumental variable (step 4 in Fig. 2). The results are shown in ESM Table 7. Only one CpG, cg00689835, showed a nominal reverse MR effect ($p = 0.017$), while the effects for the other six CpGs were not significant. The associations between GRS and potential confounders are also shown in ESM Table 5. There was a

Table 2 Association of the 18 significant SNPs with methylation levels at nearby CpGs

SNP	SNP position	Locus	No. of CpGs[a]	Best CpG[b]	Best CpG position	Best CpG *p* value
rs780094	2:27741237	*GCKR*	29 / 501	cg12648201	2:27665141	4.60×10^{-53}
rs895636	2:45188353	*SIX2-SIX3*	20 / 328	cg16198908	2:45192207	1.14×10^{-68}
rs540524	2:169756930	*G6PC2*	11 / 178	cg06269299	2:169756942	4.03×10^{-103}
rs560887	2:169763148	*G6PC2*	1 / 178	cg25044149	2:169777650	1.08×10^{-6}
rs2232326	2:169764491	*G6PC2*	7 / 178	cg25467166	2:169770254	1.10×10^{-55}
rs56960326	2:173591412	*RAPGEF4-AS1*	27 / 176	cg14982212	2:173591687	3.04×10^{-201}
rs17085593	5:95630705	*LOC101929710*	6 / 151	cg04095375	5:95617251	1.17×10^{-19}
rs17168486	7:14898282	*DGKB*	6 / 64	cg27295253	7:14774665	6.42×10^{-32}
rs1990379	7:15066045	*DGKB*	1 / 64	cg19272540	7:15055459	1.51×10^{-222}
rs1008384	7:44178626	*MYL7*	31 / 418	cg09354581	7:44150992	2.93×10^{-298}
rs1799884	7:44229068	*GCK*	11 / 418	cg22208713	7:44194466	8.51×10^{-27}
rs3802177	8:118185025	*SLC30A8*	4 / 109	cg22200611	8:117950296	1.96×10^{-9}
rs12684488	9:571582	*KANK1*	15 / 139	cg24849973	9:567490	7.37×10^{-160}
rs3934283	9:4285119	*GLIS3*	11 / 133	cg05406233	9:4290574	1.67×10^{-76}
rs10811662	9:22134253	*CDKN2B-AS1*	1 / 102	cg23900144	9:22142053	3.87×10^{-6}
rs10830963	11:92708710	*MTNR1B*	3 / 139	cg14824936	11:92675685	8.88×10^{-14}
rs61944004	13:28492405	*PLUT*	12 / 329	cg21101465	13:28493404	8.52×10^{-22}
rs72470563	20:22560932	*FOXA2*	20 / 150	cg19575804	20:22559456	2.68×10^{-190}

[a] No. of significant CpGs ($p<1.33\times10^{-5}$) and the number of the total tested CpGs near the SNP

[b] The most significant CpG in the region

Table 3 MR analysis results between DNA methylation and fasting glucose

SNP	SNP position	Locus	CpG	CpG position	Location[a]	meQTL		MR	
						Effect (SE)	*p* value	Effect (SE)	*p* value
rs2232326	2:169764491	*G6PC2*	cg00689835	2:169757618	Upstream	−0.976 (0.075)	3.06×10^{-37}	0.492 (0.089)	3.68×10^{-8}
rs2232326	2:169764491	*G6PC2*	cg11152513	2:169757751	UTR5	−0.438 (0.077)	1.70×10^{-8}	1.119 (0.265)	2.59×10^{-5}
rs2232326	2:169764491	*G6PC2*	cg07024094	2:169769071	Downstream	0.812 (0.076)	5.11×10^{-26}	−0.603 (0.112)	7.74×10^{-8}
rs2232326	2:169764491	*G6PC2*	cg15742848	2:169769501	Downstream	1.078 (0.074)	1.09×10^{-45}	−0.455 (0.078)	5.54×10^{-9}
rs2232326	2:169764491	*G6PC2*	cg05703053	2:169769616	Downstream	1.108 (0.074)	2.68×10^{-48}	−0.443 (0.076)	5.82×10^{-9}
rs2232326	2:169764491	*G6PC2*	cg22447106	2:169769804	Downstream	0.851 (0.075)	1.51×10^{-28}	−0.577 (0.104)	3.34×10^{-8}
rs2232326	2:169764491	*G6PC2*	cg25467166	2:169770254	Downstream	−1.164 (0.073)	1.73×10^{-53}	0.422 (0.072)	5.39×10^{-9}

[a] Location relative to the *G6PC2* gene

UTR5, 5′ untranslated region

nominal association between GRS and age but the association would not be significant considering multiple testing using four tests. To eliminate the possibility that the SNP affecting the DNA methylation is in LD with another SNP affecting fasting glucose (explanation 3 in Fig. 1c), the six CpGs with non-significant reverse MR effects were tested for the concordance of top SNPs based on rs2232326 and the other two SNPs (rs139014876 and rs192788039) in LD with rs2232326, using the bivariate fine-mapping analysis (step 5 in Fig. 2). The LD r^2 for rs139014876 and rs192788039 with rs2232326 were 0.78 and 0.89, respectively, calculated using the TWB2 sample. The results are shown in ESM Table 8. Concordance was observed for all of the six CpGs that rs2232326 had the highest posterior probability and Bayes factor to be causal to each CpG. The posterior probabilities for four CpGs also suggested that there was at most one variant (i.e. rs2232326) affecting the DNA methylation. The stepwise model selection in GCTA also selected rs2232326 as the representative SNP in the GWAS analyses. Therefore, the relationship between the SNP, DNA methylation and fasting glucose was not likely due to explanation 3 in Fig. 1. This finding supports our hypothesis that DNA methylation is part of the pathway between genetic variants and fasting glucose (i.e. explanation 1 in Fig. 1).

Replication analysis for the MR analysis results The SNP rs2232326 had a MAF of 0.048 in our data but is rare (MAF =0.3%) in the European population in the 1000 genomes data, which made the replication analysis for the MR results based on the same SNP–CpG pairs difficult. We further hypothesised that CpGs may mediate the effects of different SNPs in *G6PC2* on fasting glucose in different populations. Therefore, we extracted SNPs within 300 kb of *G6PC2* from the literature SNPs, resulting in 148 SNPs. Among the 148 literature SNPs, two (rs12475700 and rs2232328) were suggested as meQTLs influencing three CpGs from the SMR database. The two-sample MR analysis was performed based on the effect estimates of the two SNPs for fasting glucose obtained from Manning et al [6] and the effect estimates of the two SNPs for the three CpGs downloaded from the SMR database (step 6 in Fig. 2). The study samples for Manning et al [6] and the SMR database are both of European origin. Table 4 shows the two-sample MR analysis results. Compared with the results presented in Table 3, two of the same CpGs, cg15742848 and cg05703053, with the same direction of effects, were replicated in blood, although their instrumental SNPs were different. The results suggest that CpGs in *G6PC2* may mediate the effects of different SNPs on fasting glucose in different populations.

Discussion

In this study, we identified two CpGs near *G6PC2* that were putative mediators between a SNP in *G6PC2* and fasting glucose. The *G6PC2* gene encodes glucose 6-phosphatase catalytic subunit 2 (G6PC2); members of this enzyme family are part of a multicomponent system for catalysing the hydrolysis of glucose 6-phosphate to glucose and inorganic phosphate [61]. *G6PC2* is primarily expressed in pancreatic beta cells [62] and the knockout of *G6PC2* resulted in reduced fasting glucose levels in mice [63, 64]. It is hypothesised that *G6PC2* functions as a negative regulator of basal glucose-stimulated insulin secretion that hydrolyses glucose 6-phosphate so that the action of the glucose sensor is opposed [65]. Furthermore, knockdown of *G6PC2* in human EndoC-βH1 cell line increased insulin secretion at submaximal glucose levels, suggesting that basal glucose sensitivity is enhanced by promoting glycolytic flux at substimulatory glucose concentrations [66]. Multiple genetic variants in *G6PC2* have been found to be associated with fasting glucose in several GWAS [16, 65, 67], studies that focused on exome regions via exome chips [15, 66, 68, 69] and studies using whole-exome and whole-genome sequencing [44, 70].

Table 4 Replication results using the two-sample MR analysis between DNA methylation and fasting glucose

SNP	SNP position	CpG	CpG position	meQTL effect (SE)	Trait effect (SE)	MR	
						Effect (SE)	*p* value
rs12475700	2:169757780	cg07338205	2:169757780	−0.011 (0.002)	0.049 (0.003)	−4.451 (0.803)	2.96×10^{-8}
rs2232328	2:169764546	cg15742848[a]	2:169769501	0.062 (0.008)	−0.039 (0.005)	−0.623 (0.117)	1.00×10^{-7}
rs2232328	2:169764546	cg05703053[a]	2:169769616	0.055 (0.006)	−0.039 (0.005)	−0.703 (0.125)	2.10×10^{-8}

[a] CpGs replicated in Table 3

In vitro studies have been performed to further explore functional roles of the significant genetic variants in *G6PC2* [66, 68]. Reduced levels of G6PC2 protein were observed by Mahajan et al [68] when mutations of three genetic variants were introduced in the cell lines of human embryonic kidney and rat pancreatic beta cells, and the loss of *G6PC2* function reduced the fasting glucose levels. Ng et al [66] investigated functional roles of ten variants, including the three variants studied by Mahajan et al [68], and also found that mutations resulting in seven *G6PC2* variants reduced the levels of G6PC2 protein in human and rat cell lines. Both studies found that proteins resulting from *G6PC2* mutations were mainly degraded through the ubiquitin–proteasome pathway. Ng et al [66] further found that G6PC2 proteins resulting from loss-of-function variants in *G6PC2* may influence beta cell endoplasmic reticulum homeostasis. Moreover, a variant at the non-coding literature SNP rs560887 (also found significant in our study) may alter splicing, increasing the expression of *G6PC2* and subsequently elevating fasting glucose levels [66].

The association of the exonic SNP rs2232326 in *G6PC2* with fasting glucose was first reported in a study using Chinese samples [15]. It is a missense mutation and the increase of the alternative allele lowered the fasting glucose level. The mutation was predicted to be 'probably damaging' to the function of the G6PC2 protein by PolyPhen-2 [71]. Similarly, rs2232328 is also an exonic variant in *G6PC2* and a missense mutation. It was significantly associated with fasting glucose levels in the European population with the same direction of effect as rs2232326 [6]. The functional effects of the mutation, predicted to be 'benign' in PolyPhen-2, is not clear [72]. Our analysis results may provide new biological insights into how the two variants in *G6PC2* affect fasting glucose levels. That is, mutations in rs2232326 or rs2232328 in *G6PC2* may increase the DNA methylation levels at CpGs near the gene, as our analysis results showed a positive association between the SNP and the CpGs. The elevated DNA methylation levels may affect the gene expression levels for the nearby genes in pancreatic islets, which influence fasting glucose levels.

An instrumental variable in MR analysis must satisfy three assumptions: (1) the instrumental variable is associated with the risk factor; (2) the instrumental variable is independent of the confounders that confound the association between the risk factor and outcome; and (3) the instrumental variable is independent of the outcome given the risk factor and confounders [57]. As seen in Table 2, the meQTLs had strong effects on methylation, thus satisfying the first assumption. In addition, as seen in ESM Table 5, the meQTLs were not associated with potential confounders (age, sex and BMI) between methylation and fasting glucose. Finally, horizontal pleiotropy (explanation 4 in Fig. 1d) can cause violation of the third assumption. As discussed in Richardson et al [22], using single instrumental variable in MR, it is difficult to distinguish between horizontal pleiotropy and mediation. Alternatively, MR with multiple instrumental variables can be used to assess pleiotropy and provides consistent estimates of the causal effects [73]. In our analysis, this requires multiple meQTLs associated with the same CpG, which were not identified based on our analysis framework.

The major strength of this study is that replication analyses were performed for the significant results and most of the major findings were replicated. Moreover, a comprehensive and rigorous pipeline (shown in Fig. 2) was performed to support our findings. One weakness of the study, as mentioned previously, is that the MR analysis with single instrumental variable may not distinguish between horizontal pleiotropy and mediation. Another limitation in this study is that DNA methylation levels were measured in blood and the conclusions may not be generalised to other tissues such as pancreatic beta cells. To further study the functional roles of the identified CpGs in gene expression, DNA methylation and gene expression should be measured in disease-relevant tissues, as described by Dayeh et al [74].

In conclusion, our analysis results demonstrate that the integrative analysis using GWAS and DNA methylation data can provide new biological insights into the significant GWAS findings for glucose homeostasis. Further studies will be warranted to investigate the pathway involving the SNPs in *G6PC2* and CpGs near *G6PC2* implicated in our study, and the potential role of the CpGs in influencing gene expression at nearby genes and their effects on fasting glucose.

Supplementary Information The online version contains peer-reviewed but unedited supplementary material available at https://doi.org/10.1007/s00125-021-05449-9.

Acknowledgements We thank the participants from the TWB, HALST and SAPPHIRe studies.

Data availability The datasets generated during and/or analysed during the current study are not publicly available due to the data containing sensitive information about patients but are available from the corresponding author upon reasonable request.

Funding This study was supported by grants BS-090-PP-01, BS-091-PP-01, BS-092-PP-01, BS-093-PP-01, BS-094-PP-01, PH-098-SP-02, PH-099-SP-01, PH-100-SP-01, PH-101-SP-01, PH-102-SP-01, PH-105-PP-03, PH-106-PP-03, PH-107-PP-03, PH-108-PP-03 and PH-109-PP-03 from the National Health Research Institutes in Taiwan. This study was also supported in part by the National Center for Advancing Translational Sciences (CTSI grant UL1TR001881) and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) (grant DK063491) to the Southern California Diabetes Endocrinology Research Center.

Author's relationships and activities The authors declare that there are no relationships or activities that might bias, or be perceived to bias, their work.

Contribution statement RHC and CAH contributed to the study concept and design. CMH, YJH, ITL, LMC, TQ and YIC collected the data and also contributed to the conception of the study. RHC, YFC and WCW performed the statistical analyses. All authors contributed to the interpretation of the analysis results and reviewing and editing of the manuscript. All authors read and approved the final manuscript. CAH is the guarantor of this work and, as such, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Brambilla P, La Valle E, Falbo R et al (2011) Normal fasting plasma glucose and risk of type 2 diabetes. Diabetes Care 34:1372–1374. https://doi.org/10.2337/dc10-2263
2. Chen WM, Erdos MR, Jackson AU et al (2008) Variations in the G6PC2/ABCB11 genomic region are associated with fasting glucose levels. J Clin Invest 118:2620–2628
3. Bouatia-Naji N, Bonnefond A, Cavalcanti-Proenca C et al (2009) A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. Nat Genet 41:89–94. https://doi.org/10.1038/ng.277
4. Dupuis J, Langenberg C, Prokopenko I et al (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 42:105–116. https://doi.org/10.1038/ng.520
5. Scott RA, Lagou V, Welch RP et al (2012) Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet 44:991–1005. https://doi.org/10.1038/ng.2385
6. Manning AK, Hivert MF, Scott RA et al (2012) A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet 44:659–669. https://doi.org/10.1038/ng.2274
7. Nagy R, Boutin TS, Marten J et al (2017) Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants. Genome Med 9:23
8. Ingelsson E, Langenberg C, Hivert MF et al (2010) Detailed physiologic characterization reveals diverse mechanisms for novel genetic Loci regulating glucose and insulin metabolism in humans. Diabetes 59:1266–1275. https://doi.org/10.2337/db09-1568
9. Lyssenko V, Laakso M (2013) Genetic screening for the risk of type 2 diabetes: worthless or valuable? Diabetes Care 36(Suppl 2):S120–S126
10. Kong A, Steinthorsdottir V, Masson G et al (2009) Parental origin of sequence variants associated with complex diseases. Nature 462:868–874. https://doi.org/10.1038/nature08625
11. Olsson AH, Volkov P, Bacos K et al (2014) Genome-wide associations between genetic and epigenetic variation influence mRNA expression and insulin secretion in human pancreatic islets. PLoS Genet 10:e1004735. https://doi.org/10.1371/journal.pgen.1004735
12. Xue A, Wu Y, Zhu Z et al (2018) Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nat Commun 9:2941
13. Kim YJ, Go MJ, Hu C et al (2011) Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. Nat Genet 43:990–995. https://doi.org/10.1038/ng.939
14. Hwang JY, Sim X, Wu Y et al (2015) Genome-wide association meta-analysis identifies novel variants associated with fasting plasma glucose in East Asians. Diabetes 64:291–298
15. Spracklen CN, Shi J, Vadlamudi S et al (2018) Identification and functional analysis of glycemic trait loci in the China Health and Nutrition Survey. PLoS Genet 14:e1007275. https://doi.org/10.1371/journal.pgen.1007275
16. Kanai M, Akiyama M, Takahashi A et al (2018) Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. Nat Genet 50:390–400. https://doi.org/10.1038/s41588-018-0047-6
17. Hidalgo B, Irvin MR, Sha J et al (2014) Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network study. Diabetes 63:801–807. https://doi.org/10.2337/db13-1100
18. Kriebel J, Herder C, Rathmann W et al (2016) Association between DNA Methylation in Whole Blood and Measures of Glucose Metabolism: KORA F4 Study. PLoS One 11:e0152314. https://doi.org/10.1371/journal.pone.0152314
19. Kulkarni H, Kos MZ, Neary J et al (2015) Novel epigenetic determinants of type 2 diabetes in Mexican-American families. Hum Mol Genet 24:5330–5344. https://doi.org/10.1093/hmg/ddv232
20. Walaszczyk E, Luijten M, Spijkerman AMW et al (2018) DNA methylation markers associated with type 2 diabetes, fasting glucose and HbA1c levels: a systematic review and replication in a case-control sample of the Lifelines study. Diabetologia 61:354–368. https://doi.org/10.1007/s00125-017-4497-7
21. Relton CL, Davey Smith G (2012) Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int J Epidemiol 41:161–176. https://doi.org/10.1093/ije/dyr233
22. Richardson TG, Zheng J, Davey Smith G et al (2017) Mendelian Randomization Analysis Identifies CpG Sites as Putative Mediators for Genetic Influences on Cardiovascular Disease Risk. Am J Hum Genet 101:590–602. https://doi.org/10.1016/j.ajhg.2017.09.003
23. Juvinao-Quintero DL, Hivert MF, Sharp GC, Relton CL, Elliott HR (2019) DNA Methylation and Type 2 Diabetes: the Use of Mendelian Randomization to Assess Causality. Curr Genet Med Rep 7:191–207. https://doi.org/10.1007/s40142-019-00176-5
24. Liu J, Carnero-Montoro E, van Dongen J et al (2019) An integrative cross-omics analysis of DNA methylation sites of glucose and insulin homeostasis. Nat Commun 10:2581
25. Fan CT, Lin JC, Lee CH (2008) Taiwan Biobank: a project aiming to aid Taiwan's transition into a biomedical island. Pharmacogenomics 9:235–246. https://doi.org/10.2217/14622416.9.2.235
26. Hsu CC, Chang HY, Wu IC et al (2017) Cohort Profile: The Healthy Aging Longitudinal Study in Taiwan (HALST). Int J Epidemiol 46:1106–1106j. https://doi.org/10.1093/ije/dyw331

27. Wu KD, Hsiao CF, Ho LT et al (2002) Clustering and heritability of insulin resistance in Chinese and Japanese hypertensive families: a Stanford-Asian Pacific Program in Hypertension and Insulin Resistance sibling study. Hypertens Res 25:529–536. https://doi.org/10.1291/hypres.25.529
28. Chen CH, Yang JH, Chiang CWK et al (2016) Population structure of Han Chinese in the modern Taiwanese population based on 10,000 participants in the Taiwan Biobank project. Hum Mol Genet 25:5321–5331. https://doi.org/10.1093/hmg/ddw346
29. Voight BF, Kang HM, Ding J et al (2012) The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS Genet 8:e1002793. https://doi.org/10.1371/journal.pgen.1002793
30. Purcell S, Neale B, Todd-Brown K et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. https://doi.org/10.1086/519795
31. Staples J, Maxwell EK, Gosalia N et al (2018) Profiling and Leveraging Relatedness in a Precision Medicine Cohort of 92,455 Exomes. Am J Hum Genet 102:874–889. https://doi.org/10.1016/j.ajhg.2018.03.012
32. Chung RH, Chiu YF, Hung YJ et al (2017) Genome-wide copy number variation analysis identified deletions in SFMBT1 associated with fasting plasma glucose in a Han Chinese population. BMC Genomics 18:591
33. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7
34. R Core Team (2018) R: A Language and Environment for Statistical Computing. In. R Foundation for Statistical Computing, Vienna, Austria
35. Conomos MP, Miller MB, Thornton TA (2015) Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet Epidemiol 39:276–293. https://doi.org/10.1002/gepi.21896
36. Genomes Project C, Auton A, Brooks LD et al (2015) A global reference for human genetic variation. Nature 526:68–74. https://doi.org/10.1038/nature15393
37. Verma SS, de Andrade M, Tromp G et al (2014) Imputation and quality control steps for combining multiple genome-wide datasets. Front Genet 5:370. https://doi.org/10.3389/fgene.2014.00370
38. Pidsley R, Zotenko E, Peters TJ et al (2016) Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol 17:208
39. Gorrie-Stone TJ, Smart MC, Saffari A et al (2019) Bigmelon: tools for analysing large DNA methylation datasets. Bioinformatics 35:981–986. https://doi.org/10.1093/bioinformatics/bty713
40. Pidsley R, Wong CCY, Volta M, Lunnon K, Mill J, Schalkwyk LC (2013) A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics 14:293. https://doi.org/10.1186/1471-2164-14-293
41. Hannon E, Gorrie-Stone TJ, Smart MC et al (2018) Leveraging DNA-Methylation Quantitative-Trait Loci to Characterize the Relationship between Methylomic Variation, Gene Expression, and Complex Traits. Am J Hum Genet 103:654–665. https://doi.org/10.1016/j.ajhg.2018.09.007
42. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL (2016) Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. Genom Data 9:22–24. https://doi.org/10.1016/j.gdata.2016.05.012
43. Kang HM, Sul JH, Service SK et al (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42:348–354. https://doi.org/10.1038/ng.548
44. Fuchsberger C, Flannick J, Teslovich TM et al (2016) The genetic architecture of type 2 diabetes. Nature 536:41–47. https://doi.org/10.1038/nature18642
45. Mahajan A, Taliun D, Thurner M et al (2018) Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 50:1505–1513. https://doi.org/10.1038/s41588-018-0241-6
46. Takeuchi F, Yokota M, Yamamoto K et al (2012) Genome-wide association study of coronary artery disease in the Japanese. Eur J Hum Genet 20:333–340. https://doi.org/10.1038/ejhg.2011.184
47. Liu JZ, Tozzi F, Waterworth DM et al (2010) Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nat Genet 42:436–440. https://doi.org/10.1038/ng.572
48. Yang J, Ferreira T, Morris AP et al (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet 44:369–375, S361-363. https://doi.org/10.1038/ng.2213
49. MacArthur J, Bowler E, Cerezo M et al (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res 45:D896–D901. https://doi.org/10.1093/nar/gkw1133
50. Lagou V, Magi R, Hottenga JJ et al (2021) Sex-dimorphic genetic effects and novel loci for fasting glucose and insulin variability. Nat Commun 12:24
51. Prokopenko I, Langenberg C, Florez JC et al (2009) Variants in MTNR1B influence fasting glucose levels. Nat Genet 41:77–81. https://doi.org/10.1038/ng.290
52. Yang J, Benyamin B, McEvoy BP et al (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42:565–569. https://doi.org/10.1038/ng.608
53. Henningsen A, Hamann JD (2007) systemfit: A package for estimating systems of simultaneous equations in R. J Stat Softw 23:1–40
54. Wahl S, Drong A, Lehne B et al (2017) Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature 541:81–86. https://doi.org/10.1038/nature20784
55. Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M (2016) FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics 32:1493–1501. https://doi.org/10.1093/bioinformatics/btw018
56. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, Consortium E-I (2015) Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J Epidemiol 30:543–552. https://doi.org/10.1007/s10654-015-0011-z
57. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G (2008) Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 27:1133–1163. https://doi.org/10.1002/sim.3034
58. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38:e164. https://doi.org/10.1093/nar/gkq603
59. Gaunt TR, Shihab HA, Hemani G et al (2016) Systematic identification of genetic influences on methylation across the human life course. Genome Biol 17:61
60. Huan T, Joehanes R, Song C et al (2019) Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. Nat Commun 10:4267
61. Marcolongo P, Fulceri R, Gamberucci A, Czegle I, Banhegyi G, Benedetti A (2013) Multiple roles of glucose-6-phosphatases in pathophysiology: state of the art and future trends. Biochim Biophys Acta 1830:2608–2618
62. Arden SD, Zahn T, Steegers S et al (1999) Molecular cloning of a pancreatic islet-specific glucose-6-phosphatase catalytic subunit-related protein. Diabetes 48:531–542. https://doi.org/10.2337/diabetes.48.3.531

63. Wang Y, Martin CC, Oeser JK et al (2007) Deletion of the gene encoding the islet-specific glucose-6-phosphatase catalytic subunit-related protein autoantigen results in a mild metabolic phenotype. Diabetologia 50:774–778. https://doi.org/10.1007/s00125-006-0564-1
64. Pound LD, Oeser JK, O'Brien TP et al (2013) G6PC2: a negative regulator of basal glucose-stimulated insulin secretion. Diabetes 62: 1547–1556. https://doi.org/10.2337/db12-1067
65. O'Brien RM (2013) Moving on from GWAS: functional studies on the G6PC2 gene implicated in the regulation of fasting blood glucose. Curr Diab Rep 13:768–777. https://doi.org/10.1007/s11892-013-0422-8
66. Ng NHJ, Willems SM, Fernandez J et al (2019) Tissue-specific alteration of metabolic pathways influences glycemic regulation. bioRxiv: 790618
67. Wheeler E, Marenne G, Barroso I (2017) Genetic aetiology of glycaemic traits: approaches and insights. Hum Mol Genet 26: R172–R184. https://doi.org/10.1093/hmg/ddx293
68. Mahajan A, Sim X, Ng HJ et al (2015) Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. PLoS Genet 11:e1004876. https://doi.org/10.1371/journal.pgen.1004876
69. Wessel J, Chu AY, Willems SM et al (2015) Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. Nat Commun 6:5897
70. Flannick J, Fuchsberger C, Mahajan A et al (2017) Sequence data and association statistics from 12,940 type 2 diabetes cases and controls. Sci Data 4:170179
71. Adzhubei IA, Schmidt S, Peshkin L et al (2010) A method and server for predicting damaging missense mutations. Nat Methods 7:248–249. https://doi.org/10.1038/nmeth0410-248
72. Al-Daghri NM, Pontremoli C, Cagliani R et al (2017) Susceptibility to type 2 diabetes may be modulated by haplotypes in G6PC2, a target of positive selection. BMC Evol Biol 17:43
73. Burgess S, Thompson SG (2017) Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol 32:377–389
74. Dayeh T, Tuomi T, Almgren P et al (2016) DNA methylation of loci within ABCG1 and PHOSPHO1 in blood DNA is associated with future type 2 diabetes risk. Epigenetics 11:482–488. https://doi.org/10.1080/15592294.2016.1178418