USCAP

ARTICLE

CpG-SNP site methylation regulates allele-specific expression of *MTHFD1* gene in type 2 diabetes

Manik Vohra[1] · Prabha Adhikari[2] · Sydney C.D'. Souza[3] · Shivashankar K. Nagri[4] · Shashikiran Umakanth[5] · Kapaettu Satyamoorthy[6] · Padmalatha S. Rai[1]

Abstract

The interaction of genetic and epigenetic mechanisms is one of the underlying causes of phenotypic variability in complex diseases such as type 2 diabetes (T2D). To explore the influence of genetic and epigenetic changes in T2D, we examined the effect of methylation of CpG-SNP sites on allele-specific expression (ASE) in one-carbon metabolism pathway genes in T2D. Case-control study was conducted on 860 individuals (430 T2D and 430 controls). CpG-SNPs shortlisted through in silico analysis were genotyped using tetra ARMS PCR and validated using Sanger DNA sequencing. Global DNA methylation was carried out using RP-HPLC. Promoter DNA methylation and CpG site-specific methylation were carried out using bisulfite sequencing. mRNA expression and ASE were examined by SYBR green and TaqMan assay, respectively. Four exonic CpG-SNPs of *MTHFD1*, *MTRR*, and *GGH* genes were identified in folate pathway genes. Among these, *MTHFD1* rs2236225 showed significant association with T2D independent of obesity, displayed ASE, and correlated with CpG-SNP site-specific methylation when compared with controls. Our results demonstrate that SNP rs2236225 in the CpG site of *MTHFD1*, which regulates allele-specific gene expression in PBMCs is methylation dependent and may perturb one-carbon metabolism pathway in T2D subjects.

✉ Padmalatha S. Rai
padmalatha.rai@manipal.edu

1 Department of Biotechnology, Manipal School of Life Sciences, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

2 Department of Internal Medicine, Yenepoya Medical College, Yenepoya Deemed to be University, Derlakatte, Mangalore, Karnataka 570018, India

3 Department of General Medicine, Yenepoya Medical College, Yenepoya Deemed to be University, Derlakatte, Mangalore, Karnataka 570018, India

4 Department of Medicine, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

5 Department of Medicine, Dr. T.M.A. Pai Rotary Hospital, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

6 Department of Cell and Molecular Biology, Manipal School of Life Sciences, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

Introduction

Phenotypic variations arising from dysregulated allele-specific expression (ASE) plays a vital role in various human diseases, including type 2 diabetes (T2D). Dysregulated ASE occurs due to the combined effect of environmental and genetic effects [1] and leads to altered expression of numerous disease-causing variants in different diseases, including T2D. T2D, a complex heterogeneous disorder with multiple complications, is a consequence of critical interaction between genetic, epigenetic, and environmental factors [2]. Over the decades, the search for cellular, biochemical, molecular, and functional aspects of this polygenic disease has provided insights for early prediction and treatment. The genome-wide association studies (GWAS) in different ethnic groups and successive waves of combined meta-analysis have identified susceptibility genes. Although there are more than 80 robustly identified loci [3–8], a clear genetic underpinning for risk prediction remains elusive. The realization that genetic factors alone are not sufficient to predict the risk for diabetes led to further studies on exploring epigenetic mechanisms in and around the identified susceptibility genes and other loci [9]. Many investigators have also attempted to explore the

environmental and genetic risk in T2D and T2D associated complications [10].

Recent studies have suggested a potential role of DNA methylation and histone modifications in the pathogenesis of T2D [11–13]. One of the intensely studied environmental factors that are known to influence DNA methylation is diet [14]. There is evidence to support the hypothesis that nutrients can alter DNA methylation patterns [15, 16]. Among these, folate provides the most compelling evidence of nutrient-induced DNA methylation changes due to its role in one-carbon metabolism [17]. From a genetics perspective, certain variations in one-carbon metabolism pathway genes can disrupt the methyl pool, as well as homocysteine levels [18]. Elevated homocysteine levels in T2D are linked to insulin resistance [19] and endothelial dysfunction [20].

Collectively, genetic and epigenetic factors directly influence the expression of genes. The clusters of DNA methylation sites under genetic control, also known as GeMes [21], can influence differential gene expression at allelic levels, commonly known as ASE. Previous studies by Ling et al. [22] have demonstrated the role of genetic and epigenetic factors in the control of gene expression of candidate T2D genes in skeletal muscle cells [22]. They show a negative correlation of DNA methylation with *NDUFB6* expression, which is associated with insulin sensitivity. Evidence for the single nucleotide polymorphisms (SNPs) in CpG sites that may alter DNA methylation and cause changes in gene expression in relation to insulin resistance in muscles of older individuals has been reported [22]. The SNPs at the CpG sites within the promoter region alters the binding of transcription factors such as Ta-BF [23]. Dayeh et al. showed the role of previously identified GWAS SNPs in CpG sites and their impact on gene expression in pancreatic islet cells [24]. Considering the importance of the one-carbon metabolism pathway in human physiology, the role of CpG-SNPs in folate pathway genes in T2D is, however, an unexplored area. In the present study, we have (1) identified high-frequency CpG-SNPs in key folate pathway genes, (2) analyzed the association of CpG-SNPs of folate pathway genes with T2D, (3) analyzed CpG-SNP sites specific DNA methylation in T2D individuals, (4) evaluated the relationship between the CpG-SNP, CpG site-specific DNA methylation, and global DNA methylation in T2D individuals, and (5) elucidated the impact of CpG-SNP on mRNA expression and ASE in T2D individuals.

Material and methods

Study population

The study participants were enrolled from outpatient departments of Kasturba Hospital, Manipal and Mangalore,

Dr. T.M.A. Pai Hospital, Udupi and Wenlock District Hospital, Mangalore, India. A total of 860 participants (430 T2D and 430 healthy controls) were included consecutively after clinical evaluations by the participating clinical collaborators. The participants who were recruited for the study were between the age of 25 and 70 years and south Indian origin. The inclusion of subjects in the T2D group was based on their diagnosis according to the world health organization criteria. Individuals with type 1 diabetes mellitus, gestational diabetes mellitus, maturity onset diabetes of young and secondary diabetes (e.g., hemochromatosis and pancreatitis), clinical hepatic or renal impairment, cardiac abnormalities (unstable angina pectoris and myocardial infarction), non-ketotic hyperosmolar coma or chronic diabetic complications, uncontrolled hypertension, use of medications that could affect glucose metabolism, patients infected with hepatitis virus, malignant diseases, psychosis, hematological diseases, autoimmune diseases, significant digestion and resorption disturbances acute cerebrovascular accidents within the last 6 months, use of anti-obesity medications were excluded from the study. For the control group, inclusion was based on <5.7% glycated hemoglobin (HbA1c) and fasting glucose levels (<126 mg/dl) and no history of diabetes in the first-degree relatives. The study protocol and procedures were reviewed and approved by the Institutional Ethics Committee, Kasturba Hospital, Manipal (Reference no. IEC/348/2014). All the research was carried out according to 'The Code of Ethics of the World Medical Association' (Declaration of Helsinki). The study procedure was explained to the participants, and written informed consent was obtained before recruitment. All methods were conducted in accordance with the relevant guidelines and regulations.

Clinical data collection

The clinical information of all the study participants was obtained from medical records. The blood chemistry, including HbA1c, fasting plasma glucose, total cholesterol, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, and triglycerides were documented (Table 1). The anthropometric evaluation of the participants was performed by trained examiners. Based on body mass index (BMI) values, the subjects were grouped as normal (18.5–22.9 kg/m²), overweight (23–25 kg/m²), or obese (>25 kg/m²) individuals. The information regarding lifestyle factors such as alcohol consumption, smoking, tobacco chewing, and diet was obtained by self-reporting.

Sample preparation

Peripheral blood sample (5 ml) was obtained from study participants through venipuncture. Density gradient

Table 1 Clinical characteristics of study subjects.

	Control (n = 430)	Cases (n = 430)	p -value
Demographs			
Age (years)	52.8 ± 11.12	53.37 ± 9.12	0.4
Sex (male/female)	289/141	254/176	0.01*
Anthropometric measurements			
Waist circumference (cm)	84.03 ± 9.45	94.44 ± 14.62	<0.0001*
BMI (kg/m^2)	23.78 ± 3.24	25.34 ± 3.83	<0.0001*
Glycemic index			
HbA1c (%)	–	8.4 ± 2.24	–
Fasting glucose (mg/dl)	92.82 ± 4.8	179.09 ± 65.4	<0.0001*
Postprandial glucose (mg/dl)	138 ± 42.6	250.2 ± 92.41	<0.0001*
Lipid profile			
Total cholesterol (mg/dl)	156.92 ± 27.13	179.19 ± 44.87	<0.0001*
HDL cholesterol (mg/dl)	50.85 ± 13.72	46.61 ± 31.21	0.01*
LDL cholesterol (mg/dl)	83.98 ± 21.97	97.73 ± 38.01	<0.0001*
Triglycerides (mg/dl)	113.73 ± 29.09	158.37 ± 69.25	<0.0001*
Lifestyle factors			
Alcohol consumption (Yes/No)	27/397	42/383	0.07
Smoker (Yes/No)	29/394	46/380	0.05
Tobacco chewer (Yes/No)	5/418	19/407	0.005
Diet (Veg/Non-Veg)	222/208	284/146	<0.0001*

All continuous variables expressed as mean ± standard deviation. Students t test and chi-square test was performed using online statistical tools.

BMI body mass index, *HbA1c* glycated hemoglobin, *HDL* high-density lipoproteins, *LDL* low-density lipoproteins.

*p value < 0.05 was considered statistically significant.

centrifugation was performed using Ficoll-Paque PLUS (GE healthcare, UK) to separate granulocytes and peripheral blood mononuclear cells (PBMCs). Granulocytes and 50% of PBMCs fractions were subjected to DNA isolation using the standard phenol-chloroform extraction method. The remaining 50% of PBMCs fraction were subjected to isolation of RNA using the TRIzol method. The isolated plasma, DNA, and RNA were stored at −80 °C until used.

CpG-SNP selection and genotyping

The selection of SNPs in key folate pathway genes was carried out using previously described in silico strategy with slight modifications [25]. The selection criteria were (1) SNPs reported in 1000 Genomes project, (2) SNPs with known functions in 'All SNP 142 UCSC' tracks, (3) allele frequency > 0.2, (4) the nucleotide change should include addition or removal of cytosine residue adjacent to guanine residue, and (5) the genes should be expressed in blood. The selection criteria are explained in Fig. S1. Using DNA extracted from granulocytes, SNP genotyping was performed by tetra primer ARMS PCR (T-ARMS PCR), as described before [26]. The details of the primers designed for T-ARMS PCR are provided in the Table S1. Samples were chosen randomly to confirm the genotype using direct

DNA sequencing method performed on 3130 genetic analyzer (Applied Biosystems, USA).

Bisulfite DNA sequencing

DNA isolated from PBMCs fraction was subjected to bisulfite conversion by EZ DNA Methylation-Gold kit (Zymo Research, Orange, CA, USA) using the manufacturer's protocol. Briefly, sodium bisulfite treatment and PCR were performed on 1.5 μg of genomic DNA using bisulfite specific primers (Table S1). The PCR fragments were purified and sequenced on 3130 genetic analyzer (Applied Biosystems, USA) using BigDye terminator cycle sequencing kit (Applied Biosystems, USA) [27]. The results obtained from DNA sequencing were checked for quality, and absolute methylation of CpG sites in the PCR amplicons was analyzed using ESME software [28] and cross-checked manually.

Global DNA methylation

The overall 5-methylcytosine content of the DNA was estimated by reversed-phase high-performance liquid chromatography (RP-HPLC) as described previously [29, 30]. Purified genomic DNA (1–2 μg) from PBMCs was digested with amplification grade DNase1 (Merck KGaA, Darmstadt, Germany). This was followed by digestion with Nuclease P1 enzyme in the presence of 10 mM zinc sulfate and 30 mM sodium acetate. Further, samples were treated with calf intestinal alkaline phosphatase (CIP). The digested products were used for RP-HPLC (Waters 2695). The separation was performed on C18 HPLC column (Waters symmetry 250 × 4.6 mm, 5 μm) with 0.1 M potassium dihydrogen phosphate (pH 4.1) buffer consisting of 10% methanol as the mobile phase. The elution of each deoxynucleoside was detected at 260 and 280 nm. The area under the chromatographic peaks was considered to calculate the percentage of 5-methylcytosine using following formula [29, 30]:

$$\%5 - mdCMP = [5 - mdCMP/(dCMP + 5 - mdCMP)] \times 100,$$

where methylated deoxycytosine monophosphate is 5-mdCMP and deoxycytosine monophosphate is dCMP.

Quantification of total gene expression and ASE

A total of 1 μg of isolated RNA was used to prepare cDNA using Reverse transcription kit (Applied Biosystems, USA) as per the manufacturer's protocol. Primers specific for cDNAs of *MTHFD1*, *MTRR*, and *GGH* genes were designed for measuring the gene expression (Table S1). Total gene expression levels of cDNA samples were estimated using PowerUp™ SYBR™ Green Master Mix

(Applied Biosystems, USA) on 7500 Fast Real-Time PCR system (Applied Biosystems, USA). The results were normalized with the total expression levels of GAPDH. A Ct value of 35 was considered as the cutoff for the expression of genes, i.e., PCR, which yielded a Ct value lower than 35 cycles were considered for expression analysis. ASE was determined in selected heterozygous cDNA samples using custom-designed Taqman technology (Applied Biosystems, USA). The standard curve was generated using a heterozygous DNA sample, which served as reference of 50:50 allele ratio, as described previously [31]. The Ct values for ASE were calculated using 7500 Software v2.0.6 (Applied Biosystems, USA). Differential allelic expression was calculated based on the $\log_2$ of the ratio of VIC (C-allele) and FAM (T-allele).

Statistical analysis

Sample size calculation was performed on Quanto version 1.2.4 (http://biostats.usc.edu/Quanto.html) using minor allele frequency obtained from the dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/). The differences in the continuous variables (clinical variables between T2D and controls) were evaluated using students *t* test, and data is represented in the form of mean ± SD. The association analysis of SNPs with T2D was performed using SNPstat software [32]. The results were assessed on the basis of odds ratio (OR) and 95% confidence interval (CI). Hardy–Weinberg equilibrium test was performed to assess and compare genotype frequencies. The association of CpG-SNPs with T2D was also analyzed after categorizing the individuals based on the BMI (normal, overweight, and obese). The association was tested with normal BMI healthy individuals as a reference, and a threshold of significance $p < 0.05$ was considered. The difference in the methylation levels of CpG-SNP site and nearby CpG sites was evaluated in three different genotypes (i.e., wild type, heterozygous, and mutant) separately by comparing T2D subjects with controls using *t*-test a threshold significance of $p < 0.05$ (Fig. 1c–e). Pearson' correlation was also performed to detect the influence of CpG-SNP site methylation on methylation levels of neighboring CpGs for three different genotypes separately using the coMET package in R [33] (Fig. 1c–e). The mean methylation levels of the CpG sites in gene promoter were compared between T2D subjects and controls using the unpaired *t* test. The global DNA methylation was expressed as mean ± standard deviation and analyzed for differences using ANOVA at $p < 0.05$ significance level. The mRNA expression of genes were categorized on the basis of genotypes in T2D and control subjects, and the difference in expression was evaluated using ANOVA at $p < 0.05$ significance level. The differential allelic expression was analyzed to test any

deviations from $\log_2(1.20)$, i.e., $\log_2$(ratio of FAM/VIC) considered as a null hypothesis based on the derivations from the heterozygous DNA sample. One-way ANOVA was used to analyze the extent of variance in the deviation of allelic ratios in T2D and control individuals. Pearson's correlation was performed between the methylation at CpG-SNP site and differential allelic expression considering p value < 0.05 as statistically significant.

Results

Characteristics of study subjects

The present study was conducted on 860 clinically confirmed participants (430 T2D and 430 healthy controls). Table 1 summarizes the physiological and clinical characteristics of all the study subjects. The T2D subjects showed significantly higher glycemic index, anthropometric measurements, and lipid profile, with the exception of HDL-cholesterol.

CpG-SNPs in folate pathway genes is associated with T2D under statistical models

We selected four CpG-SNPs of *MTHFD1* (rs2236225), *MTRR* (rs1532268), *GGH* (rs11545077), and *GGH* (rs1800909) in the folate pathway based on the in silico analysis [34]. The association of selected CpG-SNPs of folate pathway genes was tested in T2D individuals after adjusting for age and sex (Table 2). *MTHFD1* rs2236225 was found to be significantly associated with diabetes across various statistical models. The *MTHFD1* CT and TT genotypes were significantly associated with T2D across a co-dominant model with OR 1.89 (1.36–2.61) and 2.43 (1.66–3.55), p value < 0.0001. The dominant and recessive models also showed significant risk association with OR 2.05 (1.51–2.79), p value < 0.0001 and OR 1.62 (1.18–2.22), p value = 0.0025, respectively. Interestingly, *GGH* rs1800909 showed protective association with T2D across various models [CT genotype OR 0.73 (0.54–98), p value = 0.017; CC-genotype OR 0.58 (0.37–92), p value = 0.017; dominant model OR 0.69 (0.52–90), p value = 0.0069]. The association of CpG-SNPs with T2D in the background of obesity was tested (Table S2). The allelic OR of *MTHFD1* rs2236225 CpG-SNP was found to be significantly associated with T2D subjects with normal BMI [1.5(1.1–2.1), p value = 0.009] and T2D obese individuals [1.4(1.09–1.9), p value = 0.009] when compared with healthy controls with normal BMI (Table S2). The allele frequency distribution of studied polymorphisms in T2D and control subjects is presented in Table S2.
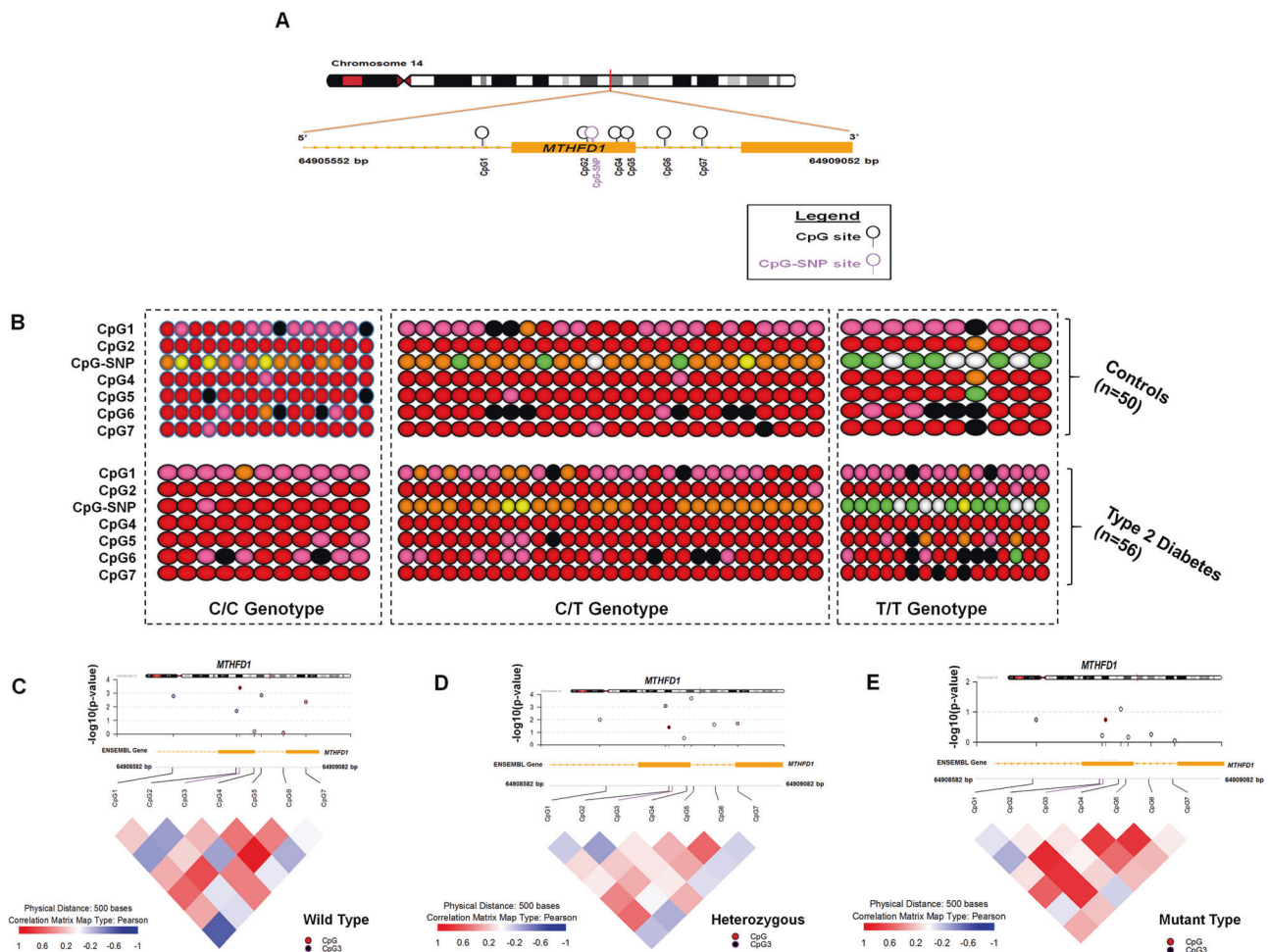
Fig. 1 The methylation analysis of target locus in the exon of **MTHFD1** gene identified in our study groups. **a** The genomic organization and location of CpG sites in the target locus in *MTHFD1* gene. **b** The methylation levels of each CpG site of individual samples grouped on the basis of CpG-SNP genotypes. **c** Correlation of DNA methylation levels of neighboring CpG sites in rs2236225 CC-genotype individuals. **d** Correlation of DNA methylation levels of neighboring CpG sites in rs2236225 CT genotype individuals. **e** Correlation of DNA methylation levels of neighboring CpG sites in rs2236225 TT genotype individuals. The top panel in (**c**)–(**e**) shows the log-transformed *p* value for the difference in methylation levels at each CpG site between T2D and control samples. The bottom panel in (**c**)–(**e**) shows the correlation matrix of methylation levels at each CpG site with each other calculated using Pearson's correlation method.

CpG-SNP of *MTHFD1* is a significant predictor of site-specific DNA methylation in T2D

The impact of four CpG-SNPs, namely *MTHFD1* (rs2236225), *MTRR* (rs1532268), *GGH* (rs11545077), and *GGH* (rs1800909) on DNA methylation in PBMCs of T2D individuals, was examined. DNA methylation data were successfully generated for *MTHFD1* rs2236225 and *MTRR* rs1532268. DNA methylation of CpG sites surrounding *MTHFD1* rs2236225 (7 CpG sites) and *MTRR* rs1532268 (4 CpG sites) was also analyzed. However, DNA methylation data could not be obtained for *GGH* rs11545077 and *GGH* rs1800909 due to PCR failure. *MTHFD1* rs2236225 (Figs. 1a, b and S2) showed significant differential DNA methylation at CpG-SNP site ($p < 0.0001$) but, *MTRR*

rs1532268 CpG-SNP did not show variation in DNA methylation at the CpG-SNP site in PBMCs of study participants.

CpG-SNPs can serve as differentially methylated positions (DMPs)

Using a coMET package in R, a correlation analysis of CpG-SNP site methylation for three different genotypes of *MTHFD1* rs2236225 along with the neighboring CpG sites was performed in T2D and controls subjects (Fig. 1c–e). In Fig. 1c–e, the top panel represents the strength of association signal of individual CpG-site methylation with T2D represented through log-transformed *p* value for difference between T2D and controls at each CpG site. The middle

Table 2 Association analysis of *MTHFD1* rs2236225, *MTRR* rs1532268, *GGH* rs11545077, and *GGH* rs1800909 with type 2 diabetes.

Gene/SNP ID	Model	Genotype	Type 2 diabetes ($N = 430$)	Healthy control ($N = 430$)	OR (95% CI)	*p* value	AIC
MTHFD1 rs2236225	Codominant	C/C	89 (20.7%)	150 (34.9%)	1 (Reference)	**<0.0001***	1174.3
		C/T	217 (50.5%)	194 (45.1%)	**1.89 (1.36–2.61)**		
		T/T	124 (28.8%)	86 (20%)	**2.43 (1.66–3.55)**		
	Dominant	C/C	89 (20.7%)	150 (34.9%)	1 (Reference)	**<0.0001***	1174.5
		C/T–T/T	341 (79.3%)	280 (65.1%)	**2.05 (1.51–2.79)**		
	Recessive	C/C–C/T	306 (71.2%)	344 (80%)	1 (Reference)	**0.0025***	1187.1
		T/T	124 (28.8%)	86 (20%)	**1.62 (1.18–2.22)**		
	Allele	C vs. T			**1.57 (1.30–1.90)**	**<0.0001***	1174.2
MTRR rs1532268	Codominant	C/C	199 (46.3%)	198 (46%)	1 (Reference)	0.6	1197.2
		C/T	173 (40.2%)	164 (38.1%)	1.05 (0.79–1.40)		
		T/T	58 (13.5%)	68 (15.8%)	0.85 (0.57–1.27)		
	Dominant	C/C	199 (46.3%)	198 (46%)	1 (Reference)	0.95	1196.2
		C/T–T/T	231 (53.7%)	232 (54%)	0.99 (0.76–1.30)		
	Recessive	C/C–C/T	372 (86.5%)	362 (84.2%)	1 (Reference)	0.33	1195.3
		T/T	58 (13.5%)	68 (15.8%)	0.83 (0.57–1.21)		
	Allele	C vs. T			0.95 (0.79–1.15)	0.6	1195.9
GGH rs11545077	Codominant	C/C	229 (53.3%)	257 (59.8%)	1 (Reference)	0.16	1194.5
		C/T	171 (39.8%)	148 (34.4%)	1.30 (0.98–1.72)		
		T/T	30 (7%)	25 (5.8%)	1.35 (0.77–2.36)		
	Dominant	C/C	229 (53.3%)	257 (59.8%)	1 (Reference)	0.054	1192.5
		C/T–T/T	201 (46.7%)	173 (40.2%)	1.30 (1.00–1.71)		
	Recessive	C/C–C/T	400 (93%)	405 (94.2%)	1 (Reference)	0.49	1195.7
		T/T	30 (7%)	25 (5.8%)	1.21 (0.70–2.10)		
	Allele	C vs. T			1.23 (0.99–1.53)	0.067	1192.9
GGH rs1800909	Codominant	A/A	272 (63.3%)	233 (54.2%)	1 (Reference)	**0.017***	1190.1
		G/A	122 (28.4%)	144 (33.5%)	**0.73 (0.54–0.98)**		
		G/G	36 (8.4%)	53 (12.3%)	**0.58 (0.37–0.92)**		
	Dominant	A/A	272 (63.3%)	233 (54.2%)	1 (Reference)	**0.0069***	1188.9
		G/A-G/G	158 (36.7%)	197 (45.8%)	**0.69 (0.52–0.90)**		
	Recessive	A/A-G/A	394 (91.6%)	377 (87.7%)	1 (Reference)	0.056	1192.6
		G/G	36 (8.4%)	53 (12.3%)	0.65 (0.42–1.02)		
	Allele	G vs. A			**0.75 (0.61–0.92)**	**0.0046***	1188.2

OR odds ratio, *CI* confidence interval, *AIC* akaike information criterion.

*p value < 0.05 was considered significant.

The significant values are in bold.

panel provides the annotation tracks extracted from genome assembly hg19, and the bottom panel depicts the correlation between CpG sites in the selected genomic region. Although no significant correlation of DNA methylation was found between CpG-SNP with the neighboring CpG sites within the groups, there was a significant difference between DNA methylation at CpG-SNP site between T2D and controls across *MTHFD1* rs2236225 CC and CT genotypes (*p* value = 0.0004 and 0.04, respectively) (Fig. 1c, d). Thus, the CpG-SNP can substitute DMPs as opposed to differentially methylated regions, which are traditionally examined in epigenetic studies.

CpG-SNP of *MTHFD1* rs2236225 does not affect promoter methylation and is associated with Global DNA methylation status in PBMCs

The promoter methylation analysis was performed using ESME software [28] and the heat map was created using the BISMA tool [35]. It was observed that clusters of methylated sites were present in the control group as opposed to the T2D group (Fig. 2a). There was no significant difference observed between T2D and control groups in the mean methylation levels at individuals CpG sites in the promoter region of the *MTHFD1* gene (Fig. 2b). We also observed
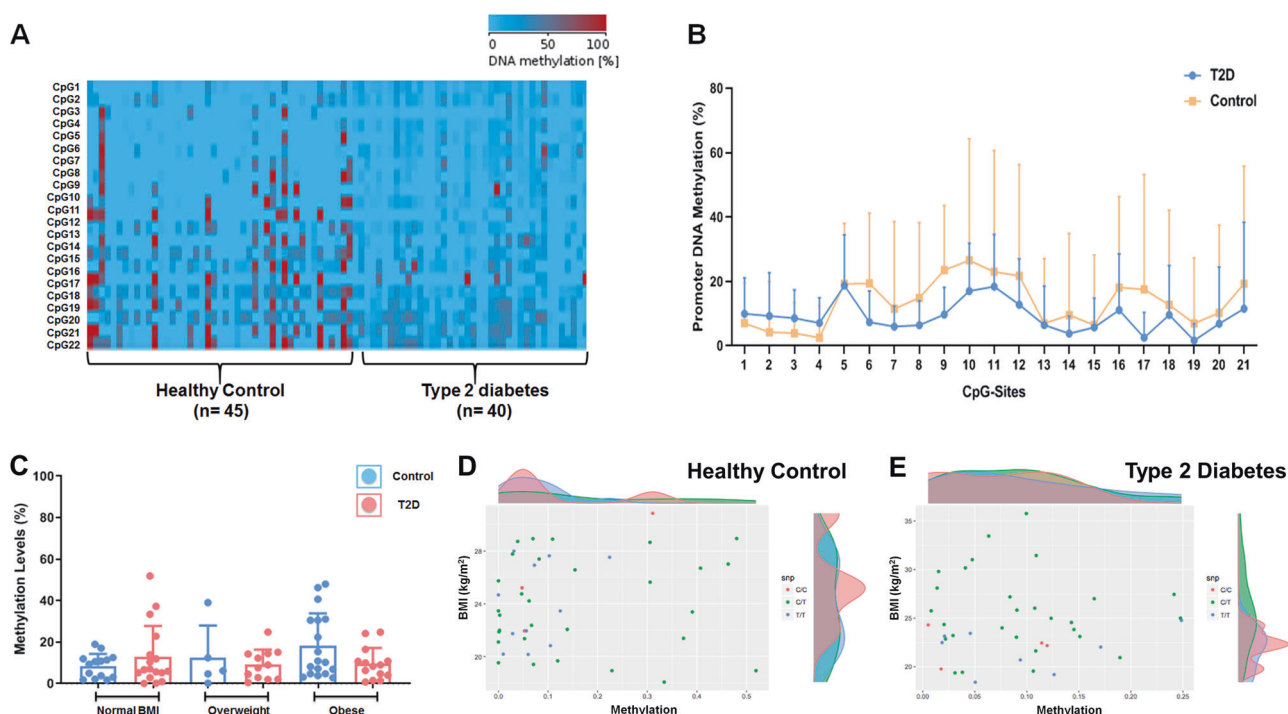
Fig. 2 The methylation analysis of *MTHFD1* gene promoter.
a *MTHFD1* promoter methylation levels in T2D and controls. **b** Mean DNA methylation levels of individual CpG sites in the *MTHFD1* promoter region in T2D and controls. **c** *MTHFD1* promoter methylation levels in T2D and controls grouped on the basis of normal BMI, overweight, and obese.

that BMI did not show any impact on the mean methylation of the promoter region in T2D and control subjects (Fig. 2c). The clustering analysis for BMI and promoter methylation levels based on rs2236225 genotypes did not show distinct clusters which shows that CpG-SNP of *MTHFD1* rs2236225 does not affect promoter methylation (Fig. 2d, e). Further, the global DNA methylation data obtained from the RP-HPLC method was subjected to categorization based on *MTHFD1* rs2236225 genotypes and methylation states (Fig. 3). The five categories analyzed were (1) methylated C/C genotype (mC/C), (2) unmethylated C/C genotype (C/C), (3) methylated C/T genotype (mC/T), (4) unmethylated C/T (C/T), and (5) T/T genotype (T/T). The mean global DNA methylation of the mC/T group was significantly lower than C/C and C/T groups (Fig. 3). Also, there was a significant difference in mean global DNA methylation between C/C and T/T groups (Fig. 3).

CpG-SNPs in folate pathway genes may influence mRNA expression in T2D PBMCs

The DNA cytosine methylation at discrete regions of a gene is known to influence the gene expression [36, 37]. We, therefore, analyzed the expression of *MTHFD1*, *MTRR*, and *GGH* genes. *MTHFD1* gene showed a 3.4-fold lower expression in CC (wild) genotype in T2D individuals when

compared with the controls (p value = 0.0014). However, we observed 3.3-fold higher expression in TT (mutant) genotype control individuals (p value < 0.0001). In T2D individuals, the *MTHFD1* expression was 2.9-fold higher in TT genotype when compared with CC-genotype individuals (p value = 0.015). In control individuals, the *MTHFD1* expression was fivefold lower in TT genotype when compared with CC-genotype individuals (p value < 0.0001). *MTRR* gene expression, on the other hand, did not show any significant difference between the different genotypes in T2D and control groups (Fig. 4). Out of the total samples screened, we were able to detect *GGH* gene expression in only four samples with Ct value of over 35. Thus, we excluded the *GGH* gene from the analysis of gene expression data.

CpG-SNP of *MTHFD1* rs2236225 is associated with ASE

The heterozygous samples were selected to analyze the ASE of *MTHFD1* rs2236225. We categorized individuals based on BMI to evaluate the differences in allelic ratios presented by each group. We found considerable variations in the magnitude of ASE in T2D individuals with normal BMI with the ratio in favor of mutant T-allele (Fig. 5a, b). However, T2D individuals with high BMI and controls with both high and normal BMI showed the normal distribution
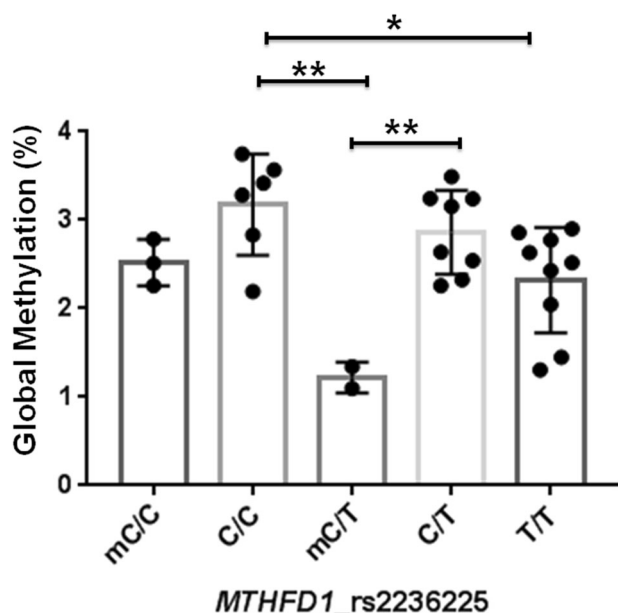
Fig. 3 Global DNA methylation levels in PBMCs. Difference in global DNA methylation levels in PBMCs of T2D individuals between genotypes of *MTHFD1* rs2236225.

of allelic ratios (50:50). The mean allelic ratio in T2D individuals with normal BMI was statistically significant in comparison with obese and overweight T2D individuals ($p = 0.001$), obese controls ($p = 0.001$), overweight controls ($p = 0.002$), and controls with normal BMI ($p = 0.007$). Since the mean allelic ratio in T2D individuals with normal BMI was statistically significant, correlation analysis with the CpG-SNP site methylation was conducted, which showed a positive correlation of site-specific methylation with mean allelic ratio ($R^2 = 0.76$, $p < 0.05$) (Fig. 5c).

Discussion

Using extensive database, bioinformatics, and in silico analysis, we have previously mapped the distribution of 12,473 SNPs across 48 genes of folate metabolism pathway [25]. These also included high-frequency exonic SNPs in folate pathway genes that introduced or destroyed the CpG sites. Among these CpG-SNPs, *MTHFD1* rs2236225, *MTRR* rs1532268, *GGH* rs11545077, and rs1800909 were selected for the present study. *MTHFD1* rs2236225 has been extensively studied for a variety of diseases including neural tube defects [38], congenital heart defects [39], and Alzheimer's disease [40]. Similarly, *MTRR* rs1532268 has been reported to be associated with different types of cancer such as lung cancer [41], childhood acute leukemia [42], and colorectal cancer [43]. Another folate pathway gene, *GGH* with SNPs (rs11545077 and rs1800909) significantly alters plasma homocysteine levels and may influence

methotrexate toxicity [44]. However, molecular mechanisms through which these polymorphisms exert their effect on folate pathway genes are not completely understood.

DNA methylation is one of the major mechanisms of epigenetic modulation of chromatin structure and function, including transcription, DNA replication, and repair [45]. Methylation of cytosine is facilitated by the enzymatic transfer of methyl group of S-adenosylmethionine (SAM) mediated through DNA methyltransferases. SAM is directly generated from methionine, which is either obtained from the diet or from remethylation of homocysteine. The overall methyl flux generated from folate metabolic reactions is directed toward remethylation of homocysteine in methionine synthase reaction. In turn, folate pathway genes are regulated by two interdependent DNA methylation effects: (1) the CpG islands that influence gene expression and (2) SNPs at CpG sites within the coding, noncoding and regulatory regions. Dayeh et al. [24] categorized GWAS SNPs and evaluated CpG-SNPs for their association with T2D, influence on gene expression, methylation, function, and disease pathology in pancreatic islets [24]. In the present study, we evaluated the role of CpG-SNPs localized in folate metabolism pathway genes that influence homocysteine levels and methylation pool in the cells. Of the four CpG-SNPs, only *MTHFD1* rs2236225 showed an association with T2D (Table 2). Since obesity is one of the major risk factors for T2D, we used BMI as a measure of obesity and categorized T2D and control groups into normal weight (18.5–22.9 kg/m^2), overweight (23–25 kg/m^2), and obese (>25 kg/m^2). Interestingly, the association of *MTHFD1* was retained after categorization on the basis of BMI, and the frequency of mutant genotype for *MTHFD1* (rs2236225 TT) was higher in obese T2D when compared with controls. Previous studies have shown that obesity is influenced by folate metabolism, and the serum folate levels were lower in obese subjects [46]. *MTHFD1* CpG-SNP may be one of the factors which influence folate metabolism in obese subjects. However, DNA methylation is another mechanism that can alter the expression of genes and cause a metabolic delay. The methylation analysis confirmed the genotype-specific DNA methylation in the *MTHFD1* CpG-SNP site in T2D and healthy controls (Fig. 1). Since promoter DNA methylation is known to influence the gene expression, we also evaluated promoter DNA methylation for the *MTHFD1* gene. The promoter DNA methylation, however, did not show a significant difference in comparison between T2D and control individuals. Moreover, the promoter methylation of the *MTHFD1* gene was not associated with T2D in the background of obesity (Fig. 2c). CpG-SNPs, as suggested earlier, may control the molecular mechanism to influence the effect of SNPs on a phenotype. They may act as DMPs under the direct influence of SNPs. This effect of CpG-SNPs on local DNA methylation may affect the
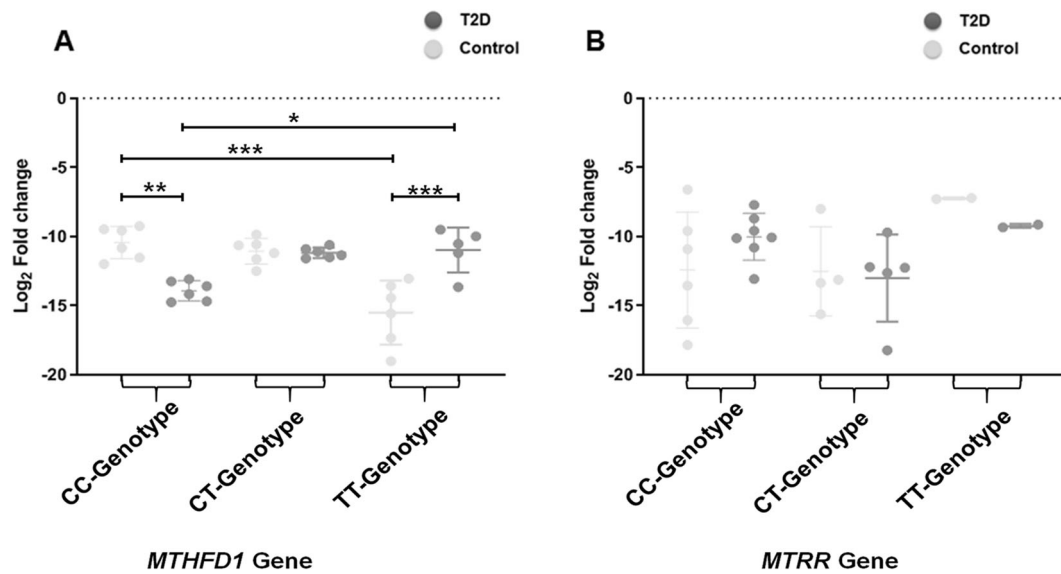
Fig. 4 Gene expression of *MTHFD1* and *MTRR* in study participants. Expression of mRNA in T2D PBMCs of individuals with **a** *MTHFD1* rs2236225 and **b** *MTRR* rs1532268.
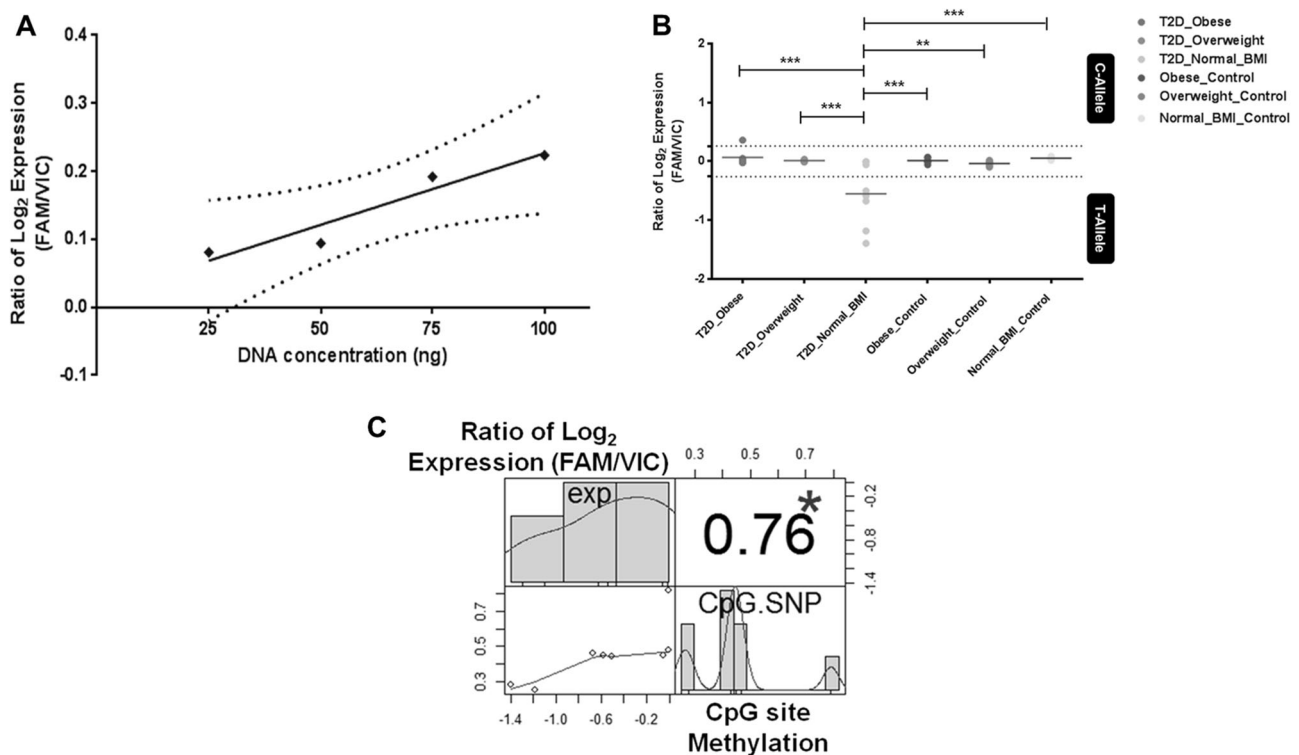

Fig. 5 Allele-specific gene expression of *MTHFD1* rs2236225. a Standard curve derived from ratio of $\log_2$ expression (FAM/VIC) of heterozygous DNA sample. **b** Allele-specific gene expression of *MTHFD1* rs2236225 in cDNA of each categories of sample. **c** Correlation between CpG-SNP site methylation and allele-specific gene expression of *MTHFD1* rs2236225 in cDNA of T2D individuals with normal BMI.

expression of the genes either by influencing the binding of CpG methyl binding proteins [23] or through transcriptional splicing [47]. Irrespective of both of these mechanisms, we were able to identify variation in *MTHFD1* gene expression in different genotypes of rs2236225 in T2D individuals. Recent reports have demonstrated the impact of *MTHFD1* rs2236225 on global DNA methylation in human PBMCs [48]. Our integrated genetic and epigenetic analysis

confirms the impact of methylation at CpG-SNP site on global DNA methylation (Fig. 3). Previously, while evaluating the acetyl CoA network in adipose tissue, *MTHFD1* was reported as a novel biomarker for T2D [49]. However, the differences in the gene expression of *MTHFD1* in our study may be due to competing mechanisms of rs2236225 or T2D pathology. To identify the underlying cause, we first analyzed *MTHFD1* gene expression and found higher expression in TT genotypes of the T2D group and CC genotype of the control group. This was followed by the suggestion that both *MTHFD1* rs2236225 frequency (T-allele) and gene expression are higher in the T2D individuals. The influence of C-allele and T-allele over one another could only be resolved through the assessment of ASE. The ASE is usually measured in heterozygous individuals as an expression of two alleles under the influence of identical conditions within the individual. The study of ASE has an advantage as it nullifies the impact of confounding factors on the gene expression and provides a comparative expression of two alleles in a given state. The evaluation of ASE of *MTHFD1* rs2236225 in our study showed the difference in the allelic ratios and the expression of T-allele was higher in T2D. Interestingly, the allele skewing toward T-allele was observed in T2D individuals with normal BMI which correlated with methylation of CpG sites in heterozygous individuals. This distribution of differential allelic expression can be further exploited to study the nature of regulatory *cis* elements that causes an imbalance in allelic ratio.

Wang et al. [50] have recently proposed a working model termed as an 'SNP intensifier' according to which SNPs can impart disease risk either by altering protein function or by allele-specific binding of the transcription factor, ASM and ASE [50]. Therefore, SNPs can intensify the ASE through creation or disruption of transcription factor binding site or state of DNA methylation at the CpG site. A major challenge is to provide evidence for the susceptibility and impact of genetic variants on common human diseases and their phenotypic variability. Our analysis provides the evidence for complex crosstalk between the genetic variant and epigenetic events such as DNA methylation in *MTHFD1* gene. The genetic association analysis with T2D highlights the disease risk attributed by *MTHFD1* rs2236225 variant and the disruption of CpG site by rs2236225 implicates changes in DNA methylation. Moreover, rs2236225 and its CpG site methylation showed changes in global DNA methylation. This highlights the importance of *MTHFD1* gene variant and site-specific methylation on the overall methyl pool generated by the one-carbon metabolism pathway. In addition, the *MTHFD1* gene also showed genotype-dependent changes in the gene expression with lower and higher expression in CC and TT genotypes, respectively in T2D individuals when compared with controls. Finally, the association of CpG site-specific methylation and ASE in heterozygous individuals suggests the critical role of site-specific methylation on ASE.

In summary, we demonstrate the association of *MTHFD1* CpG-SNP (rs2236225) with T2D. We have also identified the interaction of genetic (SNPs) and epigenetic (DNA methylation) variations in the *MTHFD1* gene in T2D. Our data provide evidence for ASE through which CpG-SNPs may affect gene function in T2D. However, further replication of the observed levels in target organs from T2DM and healthy individuals is required before the present information can be utilized for designing treatment regimens and disease management strategies.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Knowles DA, Davis JR, Edgington H, Raj A, Favé MJ, Zhu X, et al. Allele-specific expression reveals interactions between genetic variation and environment. Nat Methods. 2017;14: 699–702.
2. Prasad RB, Groop L. Genetics of type 2 diabetes—pitfalls and possibilities. Genes. 2015;6:87–123.
3. Morris A, Voight B, Teslovich T. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44:981–90.
4. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010; 42:579–89.
5. Kooner JS, Saleheen D, Sim X, Sehmi J, Zhang W, Frossard P, et al. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet. 2011;43:984–9.
6. Cho YS, Chen CH, Hu C, Long J, Ong RTH, Sim X, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet. 2011;44:67–72.
7. Steinthorsdottir V, Thorleifsson G, Sulem P, Helgason H, Grarup N, Sigurdsson A, et al. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. Nat Genet. 2014;46:294–8.
8. Ma RC, Hu C, Tam CH, Zhang R, Kwan P, Leung TF, et al. Genome-wide association study in a Chinese population identifies a susceptibility locus for type 2 diabetes at 7q32 near PAX4. Diabetologia. 2013;56:1291–305.

9. Kwak SH, Park KS. Recent progress in genetic and epigenetic research on type 2 diabetes. Exp Mol Med. 2016;48:e220–8.

10. Murea M, Ma L, Freedman BI. Genetic and environmental factors associated with type 2 diabetes and diabetic vascular complications. Rev Diabet Stud. 2012;9:6–22.

11. Ling C, Del Guerra S, Lupi R, Rönn T, Granhall C, Luthman H, et al. Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion. Diabetologia. 2008;51:615–22.

12. Yang BT, Dayeh TA, Kirkpatrick CL, Taneera J, Kumar R, Groop L, et al. Insulin promoter DNA methylation correlates negatively with insulin gene expression and positively with HbA(1c) levels in human pancreatic islets. Diabetologia. 2011;54:360–7.

13. Yang BT, Dayeh TA, Kirkpatrick CL, Taneera J, Kumar R, Groop L, et al. Increased DNA methylation and decreased expression of PDX-1 in pancreatic islets from patients with type 2 diabetes. Mol Endocrinol. 2012;26:1203–12.

14. Altmann S, Murani E, Schwerin M, Metges CC, Wimmers K, Ponsuksili S. Somatic cytochrome c (CYCS) gene expression and promoter-specific DNA methylation in a porcine model of prenatal exposure to maternal dietary protein excess and restriction. Br J Nutr. 2012;107:791–9.

15. Anderson OS, Sant KE, Dolinoy DC. Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation. J Nutr Biochem. 2012;23:853–9.

16. Zhang N. Epigenetic modulation of DNA methylation by nutrition and its mechanisms in animals. Anim Nutr. 2015;1:144–51.

17. Milagro FI, Mansego ML, De Miguel C, Martínez JA. Dietary factors, epigenetic modifications and obesity outcomes: progresses and perspectives. Mol Asp Med. 2013;34:782–812.

18. Stover PJ, Pathologies F, Stover PJ. One-carbon metabolism—genome interactions. J Nutr. 2009;139:2402–5.

19. Buysschaert M, Dramais AS, Wallemacq PE, Hermans MP. Hyperhomocysteinemia in type 2 diabetes: relationship to macroangiopathy, nephropathy, and insulin resistance. Diabetes Care. 2000;23:1816–22.

20. Doupis J, Eleftheriadou I, Kokkinos A, Perrea D, Pavlatos S, Gonis A, et al. Acute hyperhomocysteinemia impairs endothelium function in subjects with type 2 diabetes mellitus. Exp Clin Endocrinol Diabetes. 2010;118:453–8.

21. Feinberg AP. The key role of epigenetics in human disease prevention and mitigation. N Engl J Med. 2018;378:1323–34.

22. Ling C, Poulsen P, Simonsson S, Rönn T, Holmkvist J, Almgren P, et al. Genetic and epigenetic factors are associated with expression of respiratory chain component NDUFB6 in human skeletal muscle. J Clin Investig. 2007;117:3427–35.

23. Taqi MM, Bazov I, Watanabe H, Sheedy D, Harper C, Alkass K, et al. Prodynorphin CpG-SNPs associated with alcohol dependence: elevated methylation in the brain of human alcoholics. Addict Biol. 2011;16:499–509.

24. Dayeh TA, Olsson AH, Volkov P, Almgren P, Rönn T, Ling C. Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. Diabetologia. 2013;56:1036–46.

25. Vohra M, Sharma AR, Paul B, Bhat MK, Satyamoorthy K, Rai PS. In silico characterization of functional single nucleotide polymorphisms of folate pathway genes. Ann Hum Genet. 2018;82:186–99.

26. Ye S, Dhillon S, Ke X, Collins AR, Day IN. An efficient procedure for genotyping single nucleotide polymorphisms. Nucleic Acids Res. 2001;29:e88.

27. Engemann S, El-Maarri O, Hajkova P, Oswald J, Walter J. Bisulfite-based methylation analysis of imprinted genes. In: Ward A, editor. Genomic imprinting. Methods in molecular biology. Humana Press; Totowa, NJ, 2002. p. 217–28.

28. Lewin J, Schmitt AO, Adorján P, Hildmann T, Piepenbrock C. Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplificates. Bioinformatics. 2004;20:3005–12.

29. Guerrero-Preston R, Santella RM, Blanco A, Desai M, Berdasco M, Fraga M. Global DNA hypomethylation in liver cancer cases and controls: a phase I preclinical biomarker development study. Epigenetics. 2007;2:223–6.

30. Armstrong KM, Bermingham EN, Bassett SA, Treloar BP, Roy NC, Barnett MP. Global DNA methylation measurement by HPLC using low amounts of DNA. Biotechnol J. 2011;6:113–7.

31. Maia AT, Spiteri I, Lee AJ, O'Reilly M, Jones L, Caldas C, et al. Extent of differential allelic expression of candidate breast cancer genes is similar in blood and breast. Breast Cancer Res. 2009;11: R88.

32. Solé X, Guinó E, Valls J, Iniesta R, Moreno V. SNPStats: a web tool for the analysis of association studies. Bioinformatics. 2006;22:1928–9.

33. Martin TC, Yet I, Tsai PC, Bell JT. coMET: visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns. BMC Bioinform. 2015;16:131.

34. Hiraoka M, Kagawa Y. Genetic polymorphisms and folate status. Congenit. Anom. 2017;57:142–9.

35. Rohde C, Zhang Y, Reinhardt R, Jeltsch A. BISMA–fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences. BMC Bioinform. 2010;11:230.

36. Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, Sandoval J, et al. DNA methylation contributes to natural human variation. Genome Res. 2013;23:1363–72.

37. Lee CJ, Evans J, Kim K, Chae H, Kim S. Determining the effect of DNA methylation on gene expression in cancer cells. Methods Mol Biol. 2014;1101:161–78.

38. Carroll N, Pangilinan F, Molloy AM, Troendle J, Mills JL, Kirke PN, et al. Analysis of the MTHFD1 promoter and risk of neural tube defects. Hum Genet. 2009;125:247–56.

39. Khatami M, Ratki FM, Tajfar S, Akrami F. Relationship of the MTHFD1 (rs2236225), eNOS (rs1799983), CBS (rs2850144) and ACE (rs4343) gene polymorphisms in a population of Iranian pediatric patients with congenital heart defects. Kaohsiung J Med Sci. 2017;33:442–8.

40. Bi XH, Zhao HL, Zhang ZX, Liu Q, Zhang JW. Association analysis of CbetaS 844ins68 and MTHFD1 G1958A polymorphisms with Alzheimer's disease in Chinese. J Neural Transm. 2010;117:499–503.

41. Amigou A, Rudant J, Orsi L, Goujon-Bellec S, Leverger G, Baruchel A, et al. Folic acid supplementation, MTHFR and MTRR polymorphisms, and the risk of childhood leukemia: the ESCALE study (SFCE). Cancer Causes Control. 2012;23:1265–77.

42. Lautner-Csorba O, Gézsi A, Erdélyi DJ, Hullám G, Antal P, Semsei ÁF, et al. Roles of genetic polymorphisms in the folate pathway in childhood acute lymphoblastic leukemia evaluated by bayesian relevance and effect size analysis. PLoS ONE. 2013;8: e69843.

43. Pardini B, Kumar R, Naccarati A, Prasad RB, Forsti A, Polakova V, et al. MTHFR and MTRR genotype and haplotype analysis and colorectal cancer susceptibility in a case-control study from the Czech Republic. Mutat Res. 2011;721:74–80.

44. Oppeneer SJ, Ross JA, Koh WP, Yuan JM, Robien K. Genetic variation in folylpolyglutamate synthase and gamma-glutamyl hydrolase and plasma homocysteine levels in the Singapore Chinese Health Study. Mol Genet Metab. 2012;105:73–8.

45. Handy D, Castro R, Loscalzo J. Epigenetic modifications basic mechanisms and role in cardiovascular disease. Circulation. 2011;123:2145–56.

46. Da Silva VR, Hausman DB, Kauwell GP, Sokolow A, Tackett RL, Rathbun SL, et al. Obesity affects short-term folate pharmacokinetics in women of childbearing age. Int J Obes. 2013;37:1608–10.

47. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature. 2011;479:74–9.

48. Udali S, Guarini P, Moruzzi S, Ruzzenente A, Tammen SA, Guglielmi A, et al. Global DNA methylation and hydroxymethylation differ in hepatocellular carcinoma and cholangiocarcinoma and relate to survival rate. Hepatology. 2015; 62:496–504.

49. Dharuri H, 't Hoen PA, van Klinken JB, Henneman P, Laros JF, Lips MA, et al. Downregulation of the acetyl-CoA metabolic network in adipose tissue of obese diabetic individuals and recovery after weight loss. Diabetologia. 2014;57:2384–92.

50. Wang H, Lou D, Wang Z. Crosstalk of genetic variants, allele-specific DNA methylation, and environmental factors for complex disease risk. Front Genet. 2019;9:1–15.