



OPEN

## Study on the relationship between DNA methylation of target CpG sites in peripheral blood and gestational diabetes during early pregnancy

Xiaolei Wang<sup>1,4</sup>, Jin Huang<sup>1,4</sup>, Yixiang Zheng<sup>5</sup>, Sisi Long<sup>3</sup>, Huijun Lin<sup>1,4</sup>, Na Zhang<sup>1,4</sup>, Mengyuan Tian<sup>1,4</sup>, Xinrui Wu<sup>1,4</sup>, Rongjing An<sup>1,4</sup>, Shujuan Ma<sup>1,4</sup> & Hongzhuan Tan<sup>1,4</sup>

Genome-wide DNA methylation profiling have been used to find maternal CpG sites related to the occurrence of gestational diabetes mellitus (GDM). However, none of these differential sites found has been verified in a larger sample. Here, our aim was to evaluate whether first trimester changes in target CpG sites in the peripheral blood of pregnancy women predict subsequent development of GDM. This nested case–control study was based upon an early pregnancy follow-up cohort (ChiCTR1900020652). Target CpG sites were extracted from related published literature and bioinformatics analysis. The DNA methylation levels at 337 CpG sites of 80 GDM cases and 80 matched healthy controls during the early pregnancy (10–15 weeks) were assessed using MethylTarget sequencing. The best cut-off level for methylation of CpG site was determined using the generated ROC curve. The independent effect of CpG site methylation status on GDM was analyzed using conditional logistic regression. Methylation levels at 6 CpG sites were significantly higher in the GDM group than in controls, whereas those at another 6 CpG sites were significantly lower (FDR < 0.05). The area under the ROC curve at each methylation level of the significant CpG sites ranged between 0.593 and 0.650 for the occurrence of GDM. After adjusting for possible confounders, the hypermethylation status of CpG site 68167324 (OR = 3.168, 1.038–9.666) and 24837915 (OR = 5.232, 1.659–16.506) was identified as more strongly associated with GDM; meanwhile, the hypermethylation of CpG site 157130156 (OR = 0.361, 0.135–0.966) and 89438648 (OR = 0.206, 0.065–0.655) might indicate lower risk of GDM. The methylation status of target CpG sites in the peripheral blood of pregnant women during the first trimester may be associated with GDM pathogenesis, and has potential as a predictor of GDM.

### Abbreviations

GDM Gestational diabetes  
 ROC Receiver operating characteristic curves  
 OGTT Oral glucose tolerance test

Gestational diabetes mellitus (GDM) is defined as any degree of glucose intolerance with onset or first recognition during pregnancy<sup>1</sup>. Alongside the increasing prevalence of obesity and type 2 diabetes, the incidence of GDM has risen annually<sup>2</sup>. Specifically, in China, this prevalence has been estimated at 14.8% (12.8–16.7%)<sup>3</sup>. GDM in

<sup>1</sup>Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Xiangya Road, Kaifu District, Changsha City, Hunan Province 410078, China. <sup>2</sup>Reproductive and Genetic Hospital of CITIC-Xiangya, Clinical Research Center for Reproduction and Genetics in Hunan Province, Changsha City, Hunan Province 410008, China. <sup>3</sup>Hospital Infection Control Center, The Second Xiangya Hospital, Central South University, Changsha City, Hunan Province 410078, China. <sup>4</sup>Hunan Provincial Key Laboratory of Clinical Epidemiology, Changsha City, Hunan Province 410078, China. <sup>5</sup>Department of Infectious Diseases, Key Laboratory of Viral Hepatitis of Hunan, Xiangya Hospital, Central South University, Changsha City, Hunan Province 410078, China. ✉email: 790039940@qq.com; tanhz99@qq.com

pregnant women can result in severe adverse pregnancy outcomes complications, including macrosomia, premature birth, and fetal malformations<sup>4</sup>. GDM also has serious consequences for the offspring, such as childhood obesity, insulin resistance<sup>5</sup>, and impaired neurodevelopment<sup>6</sup>. In addition, GDM can recur in future pregnancies and also increases risk for postpartum type 2 diabetes<sup>7</sup>.

The mechanisms underlying GDM include but not limit to genetic background<sup>8</sup>, inflammatory factors<sup>9</sup>, and oxidative stress<sup>10</sup>. Wu et al. have shown that GDM has a genetic component, and the differences in GDM among ethnicity may be due to the interaction between genes and the environment<sup>11</sup>. Epigenetics bridges the gap between genes and the environment at the molecular level<sup>12,13</sup>. DNA methylation is one of the most commonly studied epigenetic modifications, which is catalyzed by DNA methyltransferases and uses S-adenosylmethionine as the methyl donor to convert the CpG site dinucleotide cytosine to the 5'-methylcytosine. The CpG site density and methylation degree of the upstream promoter region of the gene directly affect gene activity and expression<sup>13</sup>. DNA methylation therefore plays a key role in regulating genome transcription. Emerging studies from placental tissue and fetal cord blood suggest that in utero exposure to GDM impacts the placental and fetal epigenome<sup>8,14,15</sup>. However, these non-invasive samples collected at delivery during childbirth cannot reflect the level of maternal methylation before the onset of GDM, and could not be used for the exploration of the pathogenesis of GDM by methylation.

Peripheral blood samples of pregnant women can be used to reflect changes in DNA methylation levels during pregnancy. Current researches focused on GDM and DNA methylation are mainly based on the genome-wide DNA methylation analysis<sup>16,17</sup>. Kang et al. conducted genome-wide DNA methylation chip analysis of the peripheral blood samples in late pregnancy and revealed that the methylation levels of 200 CpG sites of 151 genes are different between the eight GDM patients and eight healthy controls<sup>18,19</sup>. Moreover, Wu et al. observed that the methylation levels of 100 CpG sites corresponding to 66 genes were different between the GDM group (n = 11) and the control group (n = 11), in the peripheral blood of pregnant women in their first trimester, further suggesting that the DNA methylation status of 5 CpG sites, in the *COPS8*, *PIK3R5*, *HAAO*, *CCDC124*, and *C5orf34* genes could be used as clinical biomarkers of GDM<sup>17</sup>. Similarly, Enquobahrie et al.<sup>20</sup> found 17 hypomethylated and 10 hypermethylated CpG sites in the GDM group with the help of genome-wide DNA methylation analysis.

Although genome-wide DNA methylation scanning can broadly mine differentially methylated CpG sites as potential candidate genes for disease diagnosis and prediction, prior studies on this topic have been limited by small sample sizes, ethnic differences, method of quantification and lack of verification<sup>16</sup>, the reproducibility of the test is poor, and the difference sites found among studies are less overlapping. On the contrary, the differential DNA methylation detection of candidate genes is a targeted identification based on existing discoveries and potential mechanisms. Here, we evaluated the DNA methylation levels in the peripheral blood of women in early pregnancy using specific target gene DNA methylation detection in order to verify the relationship between the methylation status of the targeted CpG sites and the onset of GDM.

## Material and methods

**Study design and population.** This was a nested case–control study based on an early pregnancy follow-up cohort. The cohort was established in Hunan Province Maternal and Child Health Hospital (ChiCTR1900020652) between March 2017 and December 2018, and a total of 890 pregnant women were enrolled. All the eligible participants agreed to participate in this study and provided written informed consent. The study protocol was approved by the Medical Ethical Committee of the Hunan Provincial Maternal and Child Health Hospital in South China (approval number: EC201624 on January 11, 2017) and all methods were performed in accordance with the relevant guidelines and regulations. Pregnant women were recruited in their first trimester (10–14 weeks) and followed up for 42 days post-partum. The inclusion criteria were: (1) singleton pregnancy and natural conception; (2) diabetes-free at recruitment; (3) had not received any antibiotic treatment throughout the current pregnancy; (4) no acute infection in the 2 weeks before sample collection; (5) planned to attend for all obstetric examinations and delivery at the above hospital. We collected questionnaire data and venous blood samples, while additional patient information and data concerning their clinical examinations were collected through the hospital's electronic recodes system. The venous blood samples (5 ml/person) were collected using blood collection tubes without anticoagulant during early pregnancy (10–15 gestational weeks), by certified nurses in the morning following a 10-h overnight fast. Serum and blood cells were separated by centrifugation at 3500 rpm for 15 min and stored at  $-80^{\circ}\text{C}$  until further use.

**Diagnostic criteria for GDM and selection of controls.** All subjects underwent a 2-h standard 75 g oral glucose tolerance (OGTT)<sup>21</sup> in the hospital outpatient department at 24–28 weeks of gestation. The oxidase method was used to estimate blood glucose levels, with measurement completed using an automatic biochemical analyzer (Hitachi 7600) at the hospital. GDM was defined according to the International Association of Diabetes and Pregnancy Study Groups (IADPSG) standard. That is, GDM was considered to be present when at least one of the following blood glucose concentrations was obtained:  $\geq 5.1$  mmol/L (fasting),  $\geq 10.0$  mmol/L (after 1 h), and  $\geq 8.5$  mmol/L (after 2 h)<sup>21</sup>. The controls were selected from women in the same cohort who had normal blood glucose levels throughout the pregnancy. A 1:1 pair match for each GDM patient was identified, based on the age of the pregnant ( $\pm 3$  years) and gestational week ( $\pm 1$  week) at the time of enrollment, resulting in a final study population of 80 eligible GDM patients and 80 healthy controls.

**Selection of candidate CpG sites.** Candidate CpG sites were mainly selected from published studies, and supplemented by bioinformatics analyses (Supplementary Table S3). Through a systematic review<sup>17–20,22</sup>, 21 target CpG sites were collected from the differential loci found in previous GDM-related whole-genome methylation sequencing analysis literatures, another 6 CpG islands were generated from the promoter regions of

four genes which were reported to be closely related to the pathogenesis of GDM<sup>17,23,24</sup>. Additionally, two target sites were identified from the methylation data of cervical cancer patient data sets in the The Cancer Genome Atlas (TCGA) and gene expression omnibus (GEO) databases by T-test method, with the Q value obtained from Benjamini and Hochberg modified *P* value controlling the false discovery rate of multiple hypothesis testing. Through the target site, the 50–100 bp upstream or downstream of its location was selected as the sequenced fragment, and all the CpG sites in the fragment were sequenced. Overall, this procedure therefore identified a total of 29 target fragments, containing 337 CpG sites. The primers and their sources are shown in Supplementary Tables S1, S2.

**DNA extraction.** Genomic DNA was extracted from frozen samples using Genomic Tip-500 columns (Qiagen, Valencia, CA, USA) and from bisulfite-converted samples using the EZ DNA Methylation™-GOLD Kit (Zymo Research, CA, USA) in accordance with the manufacturer's instructions. Genomic DNA integrity was measured using agarose gel electrophoresis and quality control was ensured using a NanoDrop 2000 (NanoDrop technologies, Wilmington, DE, USA), which requires that the DNA concentration  $\geq 20$  ng/ $\mu$ L, and that the total amount of DNA  $\geq 1$   $\mu$ g.

**DNA methylation analysis.** The DNA methylation level of the target CpG site is defined as the number of methylated reads at that site (i.e., the number of reads with base C detected) divided by the total number of reads at that site, and was obtained by MethylTarget sequencing (Genesky Biotechnologies Inc. Shanghai, China), a method based on next-generation sequencing-based multiple targeted CpG methylation analysis. Primer design and validation were performed using bisulfite-converted DNA samples on the Methylation Primer software. The primer sets were designed to flank each target CpG site by 100–300 nucleotides and are summarized in Supplementary Table S1. After PCR amplification (HotStarTaq polymerase kit, TAKARA, Tokyo, Japan) and library construction, paired-end sequencing was performed (Illumina HiSeq Benchtop Sequencer, CA, USA) in accordance with the manufacturer's protocol.

**Quality control.** In order to ensure the consistency of the DNA methylation level detection results of all CpG sites, all DNA samples were sent in the same batch and tested using the same test method. For the data results after sequencing, we evaluated the quality of the original data through Fast QC software. The main evaluation index was the basic quality index (Q value). During sequencing, Q20 indicates that the Q value is greater than or equal to 20, that is, the sequencing error rate (P) during sequencing is less than or equal to 1%; Q30 indicates that the Q value is greater than or equal to 30, that is, the sequencing error rate (P) during sequencing is less than or equal to 0.1%. Generally, Q20  $\geq 90\%$  (ie 90% base sequencing error rate  $\leq 1\%$ ), and Q30  $\geq 85\%$  (ie 85% base sequencing error rate  $\leq 0.1\%$ ) are considered as qualified sequencing results. At the same time, the lowest conversion rate of bisulfite in the DNA methylation level detection process was 98.88%. For details, see Supplementary Data QC (excel file).

**Covariates.** In this study, we collected information on maternal demographics, lifestyle, and pregnancy history through structured questionnaires during each follow-up. This included factors with the potential to confound the exposure-outcome relationship, including pre-pregnancy body mass index (BMI) (continuous), history of drinking (yes/no, defined as drinking alcohol one or more times for 6 consecutive months), history of smoking (yes/no), parity (continuous), pregnancy order (continuous), polycystic ovary syndrome (PCOS) (yes/no), and waist circumference at enrolment (continuous). Many included participants were primiparas, so history of GDM was not considered in the analysis.

**Statistical analysis.** Continuous data and categorical data were represented by the mean  $\pm$  standard deviation (SD) and frequency (percentage), respectively. Paired-samples T test were used to compare normally distributed continuous data, whereas Wilcoxon signed rank test were used to analyze non-normally distributed continuous data. Meanwhile, FDR (False positive rate) correction analysis was performed for CpG sites with differences in univariate analysis. Dichotomous variables were analyzed using McNemar  $\chi^2$  test.  $P < 0.05$  was considered statistically significant, and all statistical tests were two-sided. ROC curve analysis was used to assess the possible predictive value of the methylation level of individual CpG site for the occurrence of GDM. When the level of DNA methylation was positively correlated with GDM, GDM would be used as the value of the state variable for ROC curve analysis, conversely, the control would be used. Through the ROC curve, the methylation status (high or low) of the target CpG site was classified based on the best cut-off value, defined as the DNA methylation level with the highest Youden index. Conditional logistic regression analysis was used to determine the independent influence of target CpG site methylation status on GDM. The model variable selection criterion was  $\alpha_{in} = 0.05$ ; the variable elimination criterion was  $\alpha_{out} = 0.10$ ; the Wald forward method was used to establish a conditional logistic regression model to screen CpG sites with independent effects. All the statistical analyses were performed using SPSS software v25.0 (IBM Corporation, Armonk, NY, USA).

## Results

**Patient characteristics.** The participant characteristics are summarized in Table 1. The age of the GDM patients ranged between 23 and 43 years (mean: 31.6 years), whereas that of the healthy controls ranged between 24 and 45 years (mean: 32.0 years). No significant difference ( $P > 0.05$ ) was observed between the two groups in terms of gravidity, parity, PCOS, smoking history, alcohol intake history, age, or gestational age. The GDM group

Variable	Control (N = 80)		GDM (N = 80)		P value
<b>Participant</b>					
Gravidity, M (Q1–Q3)	2 (1–3)		2 (1–3)		0.750
Parity, M (Q1–Q3)	1 (0–1)		1 (0–1)		0.950
PCOS, n (%)	2 (2.5)		7 (8.8)		0.086
Smoking history, n (%)	2 (2.5)		1 (1.3)		0.560
Drink history, n (%)	3 (3.8)		3 (3.8)		1
<b>Demographic characteristics, Mean ± SD, range</b>					
Maternal age (years)	31.6 ± 4.3	23.0–43.0	32.0 ± 4.5	24.0–45.0	0.617
Gestational age (weeks)	12.7 ± 0.7	10.0–14.4	12.6 ± 0.8	9.9–14.3	0.634
Pre-pregnant BMI (kg/m <sup>2</sup> )	20.4 ± 2.6	16.2–30.9	22.4 ± 3.2	16.7–31.9	<0.001
Waist circumference (cm)	77.3 ± 7.5	63.0–108.0	82.0 ± 9.2	58.0–107.5	<0.001
<b>75 g OGTT in second trimester, Mean ± SD, range</b>					
Fasting glucose (mmol/L)	4.4 ± 0.3	3.8–5.0	4.9 ± 0.5	3.9–5.9	<0.001
1-h post-OGTT (mmol/L)	6.4 ± 0.9	3.5–7.7	9.3 ± 1.6	5.0–13.0	<0.001
2-h post-OGTT (mmol/L)	5.7 ± 0.6	3.8–6.6	8.2 ± 1.4	4.0–11.3	<0.001

**Table 1.** Study participant's characteristics. Abbreviation: *M* median; (Q1–Q3), interquartile ranges, *PCOS* polycystic ovary syndrome.

CpG site <sup>a</sup>	Gene Symbol	Chr	Distance to TSS	Methylation level (Mean ± SD)		difference <sup>b</sup>	P value	FDR
				GDM (80)	Control (80)			
chr20: 37274257	ARHGAP40	20	(+) 43,681	0.601 ± 0.046	0.584 ± 0.048	0.017	0.015	0.033
chr5: 43487508	C5orf34	5	(+) 27,765	0.456 ± 0.078	0.414 ± 0.105	0.042	0.012	0.044
chr7: 157130156	DNAJB6	7	(+) 446	0.010 ± 0.006	0.012 ± 0.005	0.002	0.013	0.033
chr7: 157130085	DNAJB6	7	(+) 375	0.006 ± 0.003	0.007 ± 0.004	0.001	0.018	0.031
chr11: 101980999	YAP1	11	(-) 192	0.009 ± 0.003	0.007 ± 0.003	0.002	0.001	0.013
chr11: 101981060	YAP1	11	(-) 131	0.011 ± 0.004	0.012 ± 0.004	0.001	0.028	0.036
chr21: 34775358	IFNGR2	21	(+) 157	0.024 ± 0.038	0.036 ± 0.039	0.012	0.012	0.044
chr22: 42466321	NAGA	22	(+) 525	0.049 ± 0.014	0.055 ± 0.016	0.006	0.008	0.052
chr14: 68167324	RDH12	14	(-) 1278	0.273 ± 0.009	0.270 ± 0.009	0.003	0.019	0.027
chr14: 68167386	RDH12	14	(-) 1216	0.314 ± 0.009	0.311 ± 0.013	0.003	0.042	0.045
chr15: 89438648	HAPLN3	15	(+) 209	0.011 ± 0.004	0.013 ± 0.006	0.002	0.048	0.048
chr14: 24837915	NFATC4	14	(-) 288	0.015 ± 0.003	0.014 ± 0.003	0.001	0.018	0.031
chr2: 191879104	STAT1	2	(-) 128	0.008 ± 0.003	0.009 ± 0.005	0.001	0.037	0.043

**Table 2.** Varying DNA methylation target CpG site and related genes. Abbreviation: *Chr* chromosome, *TSS* transcription start site and strand; CpG site<sup>a</sup>: position on chromosome; difference<sup>b</sup>: the difference of the mean methylation level in the GDM group to that in the Control group.

had higher fasting glucose, 1-h post-OGTT glucose, 2-h post-OGTT glucose, pre-pregnant BMI, and waist circumference than the control group.

**Varying DNA methylation levels at target CpG sites.** We assessed the DNA methylation levels at 337 CpG sites, and the results of the paired-samples T test or Wilcoxon signed rank test analysis for all the CpG sites are summarized in Supplementary Table S4. The mean imputation method was used to deal with missing data (Supplementary Table S4). There were 13 CpG sites with results differing significantly between the group (Table 2; scatter diagram in Supplementary Figure S1). Through FDR correction analysis, the methylation levels at 6 CpG sites within the *ARHGAP40*, *STAT1*, *C5orf34*, *RDH12*, and *YAP1* genes were higher in the GDM group than in the control group, whereas those at 6 CpG sites within the *HAPLN3*, *IFNGR2*, *YAP1*, *NFATC4*, and *DNAJB6* genes were lower in the GDM group than in the control group. Brief introduction to the function of the genes where those differential CpG sites are located can be found in Supplementary Table S5.

**Roc curve analysis of different sites.** For the 12 significantly different DNA methylation CpG sites, we further estimated the possible predictive value of the methylation level of individual CpG site for the occurrence of GDM using the ROC curve. The ROC curve parameters and the cut-off value are summarized in Table 3. The largest area under the curve (AUC) reached 0.650.

CpG site	Gene Symbol	Cut-off value	Sensitivity	Specificity	AUC	95% CI	P value
<b>For GDM</b>							
chr20: 37274257	ARHGAP40	0.600	0.550	0.638	0.594	0.506–0.682	0.040*
chr5: 43487508	C5orf34	0.420	0.775	0.475	0.602	0.514–0.691	0.026*
chr14: 24837915	NFATC4	0.014	0.713	0.525	0.628	0.541–0.715	0.005*
chr14: 68167386	RDH12	0.315	0.475	0.750	0.601	0.513–0.689	0.028*
chr14: 68167324	RDH 12	0.278	0.550	0.563	0.589	0.501–0.677	0.051
chr11: 101980999	YAP1	0.009	0.525	0.713	0.650	0.566–0.735	0.001*
<b>For Control</b>							
chr15: 89438648	HAPLN3	0.013	0.688	0.532	0.597	0.509–0.688	0.034*
chr7: 157130156	DNAJB6	0.013	0.747	0.463	0.608	0.521–0.696	0.019*
chr2: 191879104	STAT1	0.008	0.663	0.525	0.583	0.495–0.672	0.069
chr7: 157130085	DNAJB6	0.004	0.506	0.688	0.593	0.505–0.681	0.043*
chr11: 101981060	YAP1	0.014	0.550	0.575	0.582	0.494–0.670	0.074
chr21: 34775358	IFNGR2	0.028	0.671	0.566	0.614	0.533–0.708	0.014*

**Table 3.** The AUC and cut-off value of CpG site DNA methylation. \* $P < 0.05$ .

CpG site	Gene symbol	GDM (n)		Control (n)		$\chi^2$	P value
		hypo-	hyper-	hypo-	hyper-		
chr20: 37274257	ARHGAP40	38	42	51	29	1.21	0.328
chr15: 89438648	HAPLN3	61	19	42	38	5.34	0.027*
chr14: 24837915	NFATC4	23	57	40	40	7.57	0.043*
chr5: 43487508	C5orf34	16	64	36	44	4.59	0.000*
chr21: 34775358	IFNGR2	55	25	33	47	0.63	0.488
chr14: 68167324	RDH12	55	25	69	11	29.33	0.000*
chr14: 68167386	RDH12	42	38	59	21	7.00	0.011*
chr11: 101980999	YAP1	38	42	57	23	3.69	0.072
chr11: 101981060	YAP1	65	15	56	24	18.89	0.000*
chr7: 157130156	DNAJB6	46	34	62	18	12.25	0.001*
chr2: 191879104	STAT1	42	38	27	53	1.27	0.305
chr7: 157130085	DNAJB6	25	55	16	64	17.09	0.000*

**Table 4.** Distribution of DNA methylation at CpG sites between GDM group and control group. \* $P < 0.05$ ; hypo-, hypomethylation; hyper-, hypermethylation.

**Comparison of the DNA methylation status of target CpG sites.** To clearly show the effect of DNA methylation at the target CpG sites on GDM occurrence, we classified the DNA methylation levels into hypermethylation and hypomethylation statuses based on the best cut-off value (Table 3). Table 4 presents the differences in the DNA methylation statuses of the CpG sites between the GDM and control groups. Significant differences were observed in 8 CpG sites based on the McNemar  $\chi^2$  test ( $P < 0.05$ ).

**Conditional logistic regression analysis for DNA methylation status and GDM.** Conditional logistic regression analysis was used to analyze the independent effect of the methylation status of the individual site on GDM occurrence. The independent variables included the methylation status of the eight significantly different CpG sites listed in Table 4 (0 = “hypomethylation”; 1 = “hypermethylation”). The confounding variable included waist circumference and pre-pregnancy BMI.

We found that the methylation status of four CpG sites influenced GDM occurrence. Specifically, the hypermethylation of CpG site 68167324 (OR = 3.168; 95% CI 1.038–9.666), and CpG site 24837915 (OR = 5.232; 95% CI 1.659–116.506) may indicate increased risk of GDM occurrence. In contrast, the hypermethylation of CpG site 157130156 (OR = 0.361; 95% CI 0.135–0.966) and CpG site 89438648 (OR = 0.206; 95% CI 0.065–0.655) may indicate decreased risk of GDM (Table 5).

## Discussion

An increasing number of studies have explored the pathogenesis of GDM from the perspective of epigenetics. However, most of these were small (< 30 GDM cases), and they mainly observed the associations between GDM occurrence and the DNA methylation level of cord blood or placental tissue<sup>16,25–27</sup>. In this study, we evaluated the DNA methylation status of GDM-related CpG sites in the peripheral blood of women in early pregnancy using MethylTarget sequencing. In addition, we verified the associations between target CpG sites and GDM using

CpG site	Gene symbol	$\beta$	Wals 2	P value	OR	95% CI
chr15: 89438648	HAPLN3	-1.581	7.164	0.007	0.206	0.065–0.655
chr14: 24837915	NFATC4	1.655	7.970	0.005	5.232	1.659–16.506
chr14: 68167324	RDH12	1.153	4.104	0.043	3.168	1.038–9.666
chr7: 157130156	DNAJB6	-1.020	4.116	0.042	0.361	0.135–0.966

**Table 5.** Conditional logistic analysis of methylation status of CpG sites on GDM. OR odds ratio,  $\beta$  regression coefficients, 95% CI 95% confidence interval of OR.

relatively large sample size (80 GDM cases and 80 matched controls). Overall, we identified 13 CpG sites with significant differences in DNA methylation levels between the GDM and control groups based on quantitative analysis. The AUCs of the ROC curve for each methylation level of the significant CpG sites ranged from 0.593 to 0.650 predictive utility in relation to GDM. The methylation status of eight individual CpG sites were identified as differing significantly between GDM and control groups by qualitative analysis, and these were located in the promoter regions of *RDH12*, *HAPLN3*, *NFATC4*, *YAP1*, and *DNAJB6*, and the intron region of *C5orf34*. Importantly, we found that the methylation statuses of four CpG sites were significantly associated with GDM occurrence, namely CpG site 89438648 (*HAPLN3*), 68167324 (*RDH12*), 157130156 (*DNAJB6*), and 24837915 (*NFATC4*), using conditional logistic regression analysis.

In this study, hypermethylation of the CpG site 89438648, located in the promoter region of *HAPLN3*, was found to suggest a lower risk of GDM (OR = 0.206; 95% CI 0.065–0.655). *HAPLN3* codes for hyaluronan and proteoglycan link protein 3 (*HAPLN3*), and the connexin 3 belong to the hyaluronic acid and proteoglycan connexin (*HAPLN*) family, which plays roles in the aggregation of proteoglycans and hyaluronic acid, and in cell adhesion<sup>28</sup>. *HAPLN3* is involved in the organization and stability of the hyaluronic acid (HA)-dependent extracellular matrix (ECM) in many tissues. HA is one component of the ECM within the islet tissue of humans and mice<sup>29</sup>. It can cause islet amyloid deposition, which is associated with decreased  $\beta$ -cell area and an increase in  $\beta$  cell apoptosis<sup>30</sup>. Hull et al. suggested that islet amyloid deposition could reduce the number of  $\beta$ -cells<sup>30,31</sup>. Hypermethylation of the CpG site 89438648 located in the *HAPLN3* promoter region, could reduce the level of *HAPLN3*, in turn reducing the stability of the HA-ECM, and consequently reducing the impact amyloid deposition on  $\beta$  cells.

We found that the hypermethylation status of CpG site 68167324 located in *RDH12*, can increase the risk of GDM (OR = 3.168; 95% CI 1.038–9.666). *RDH12* encodes retinol dehydrogenase 12 (*RDH12*), a member of the short-chain dehydrogenases/reductases (SDRs) family<sup>32</sup>, which participates in steroid and retinol metabolism<sup>33</sup>. *RDH12*, a NADPH-dependent all-trans retinol dehydrogenase, is the key enzyme in the metabolism of retinoids<sup>34</sup>. Two oxidation products of retinoids, 9-cis-retinoic acid and all-trans retinoic acid, function to stimulate insulin secretion<sup>35</sup>. In adipocytes, retinoic acid induces the expression of the insulin signaling gene *PDK-1* and that of the glucose transporter *GLUT4*. Activating retinoic acid induces the expression of genes involved in lipid and glucose metabolism, thereby improving insulin action<sup>36</sup>. Thus, hypermethylation of the CpG 68167386 located upstream of the promoter region of *RDH12* may inhibit its transcriptional activity and reduce *RDH12* levels in peripheral blood. Subsequently, the retinoic acid metabolic pathway would be inhibited, affecting insulin secretion, and reducing its effectiveness.

The *DNAJB6* (DnaJ homolog, subfamily B, member 6) protein is a member of the heat shock protein 40 (HSP40) family<sup>37</sup> and acts as a molecular chaperone for various cellular processes. While observing insulin resistant and diabetic patients, Kurucz et al.<sup>38</sup> found that *HSP* expression was significantly changed without diabetes, and that the mRNA level of HSP72-inducible subtypes was significantly reduced in patients with type 2 diabetes. Additionally, the expression of *HSP70* in the skeletal muscle of patients with type 2 diabetes is reduced and has been shown to correlate with the degree of insulin resistance<sup>39</sup>. These HSP molecular chaperones are related to diabetes<sup>40</sup>. However, the exact association between *DNAJB6* and type 2 diabetes needs further study. In this study, hypermethylation of CpG sites 157130156, located in the promoter region of *DNAJB6*, was observed in the GDM group. This might result in increased *DNAJB6* levels via the up-regulation of *DNAJB6* transcription, thereby reducing the risk of GDM (OR = 0.361; 95% CI 0.135–0.966).

*NFATC4* codes the nuclear factor of activated T cells 4 (*NFATC4*), which is a member of the transcription factor family under the control of calcineurin (a  $\text{Ca}^{2+}$ -dependent phosphatase)<sup>41</sup>. In adipose tissue, *NFATC4* has been shown to promote the secretion of inflammatory factors<sup>42</sup>, and to act as a transcriptional repressor in regulating adiponectin gene expression, suggesting that adiponectin expression is down-regulated in obesity and type 2 diabetes<sup>43</sup>. In this study, hypermethylation of the CpG site 24837915 located in the promoter region of *NFATC4*, was associated with the presence of GDM (OR = 5.232; 95% CI 1.659–16.506).

During pregnancy, early anabolism increases and mild insulin resistance occurs<sup>44</sup>. When insulin secretion fails to balance insulin resistance, impaired glucose tolerance develops, which might subsequently lead to GDM<sup>45</sup>. Therefore, impaired secretion by  $\beta$  cells is also a key factor in GDM pathogenesis. Here, we explored the pathogenesis of GDM from an epigenetic perspective and identified 13 CpG sites that had methylation levels showing associations with GDM pathogenesis. Furthermore, conditional logistic regression analysis showed that the methylation status of four CpG sites located in the promoter regions of four genes was associated with GDM pathogenesis. These CpG sites are located in genes that could contribute to the development of GDM. Of these four CpG sites, hypermethylation of CpG site 24837915 and CpG site 68167324 was shown to be associated with GDM, whereas that of CpG site 89438648 and CpG site 157130156 could indicate reduced risk of GDM.

Thus, the methylation status of these genes may function as predictors of GDM. No publications reporting on the relationship between methylation of these four CpG sites and GDM have been found, so our suggestion of such a relationship is based on the known modes of action of the genes concerned.

However, the study also had some limitations. First, the selection of our target CpG sites was based on published literature, and we did not screen for differential CpG sites in the same population in this study, so there may be other CpG sites related to the pathogenesis of GDM that have not been verified. Second, probably due to the large total number of detected CpG sites in the background, as well as the limited significant different sites, no significant differences were found anymore after the FDR correction was used for the 337 tests. Third, the blood samples were centrifuged to separate serum and blood cells within 24 h after collection, but the cell type composition was not further separated. Which prevented us from considering DNA methylation on cell type composition or the adjustment for cell proportions. Finally, since that our sample size was relatively limited, these findings need to be further verified using a larger and broader population.

## Conclusions

In summary, by determining the DNA methylation of target CpG sites in the peripheral blood of women in early pregnancy, we found that the methylation levels of 13 CpG sites were related to GDM by quantitative analysis. After adjusting for possible confounding factors by conditional logistic regression, four CpG sites showed independent effects on GDM. These findings indicate that the methylation status of these CpG sites in the peripheral blood of pregnant women during the first trimester might be associated with was related to the pathogenesis of GDM. But the exact relationship still needs further research and verification.

## Data availability

Raw data of patient characteristics and the data of target fragment DNA methylation sequencing generated during the current study is available from the corresponding author upon reasonable request.

Received: 10 May 2021; Accepted: 28 September 2021

Published online: 14 October 2021

## References

- Abell, S. K. *et al.* Inflammatory and other biomarkers: Role in pathophysiology and prediction of gestational diabetes mellitus. *Int. J. Mol. Sci.* **16**(6), 13442–13473. <https://doi.org/10.3390/ijms160613442> (2015).
- Anna, V. *et al.* Sociodemographic correlates of the increasing trend in prevalence of gestational diabetes mellitus in a large population of women between 1995 and 2005. *Diabetes Care* **31**(12), 2288–2293. <https://doi.org/10.2337/dc08-1038> (2008).
- Gao, C. *et al.* Prevalence of gestational diabetes mellitus in mainland China: A systematic review and meta-analysis. *J. Diabetes Investig.* **10**(1), 154–162. <https://doi.org/10.1111/jdi.12854> (2019).
- Chiefari, E. *et al.* Gestational diabetes mellitus: An updated overview. *J. Endocrinol. Invest* **40**(9), 899–909. <https://doi.org/10.1007/s40618-016-0607-5> (2017).
- Dabelea, D., Hanson, R. L., Lindsay, R. S. & Pettitt, D. J. Intrauterine exposure to diabetes conveys risks for type 2 diabetes and obesity—Study of discordant sibships. *Diabetes* **49**, 2208–2211 (2000).
- Fraser, A. *et al.* Associations of existing diabetes, gestational diabetes, and glycosuria with offspring IQ and educational attainment: The Avon longitudinal study of parents and children. *Exp. Diabetes Res.* **2012**, 963735. <https://doi.org/10.1155/2012/963735> (2012).
- Coustan, D. R. Gestational diabetes mellitus. *Clin. Chem.* **59**(9), 1310–1321. <https://doi.org/10.1373/clinchem.2013.203331> (2013).
- Ruchat, S. M. *et al.* Gestational diabetes mellitus epigenetically affects genes predominantly involved in metabolic diseases. *Epi-genetics* **8**(9), 935–943. <https://doi.org/10.4161/epi.25578> (2013).
- Halvatsiotis, P. *et al.* Methylation profile of genes involved in inflammation, in the blood from pregnancies with maternal preeclampsia due to untreated gestational diabetes mellitus. *Hormones (Athens)* **18**(2), 173–178. <https://doi.org/10.1007/s42000-019-00111-x> (2019).
- Plows, J. F. *et al.* The pathophysiology of gestational diabetes mellitus. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms19113342> (2018).
- Wu, L. *et al.* Genetic variants associated with gestational diabetes mellitus: A meta-analysis and subgroup analysis. *Sci. Rep.* **6**, 30539. <https://doi.org/10.1038/srep30539> (2016).
- Moore, L. D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* **38**(1), 23–38. <https://doi.org/10.1038/npp.2012.112> (2013).
- Kim, J. K., Samaranyake, M. & Pradhan, S. Epigenetic mechanisms in mammals. *Cell Mol. Life Sci.* **66**(4), 596–612. <https://doi.org/10.1007/s00018-008-8432-4> (2009).
- Quilter, C. R. *et al.* Impact on offspring methylation patterns of maternal gestational diabetes mellitus and intrauterine growth restraint suggest common genes and pathways linked to subsequent type 2 diabetes risk. *FASEB J.* **28**(11), 4868–4879. <https://doi.org/10.1096/fj.14-255240> (2014).
- Finer, S. *et al.* Maternal gestational diabetes is associated with genome-wide DNA methylation variation in placenta and cord blood of exposed offspring. *Hum. Mol. Genet.* **24**(11), 3021–3029. <https://doi.org/10.1093/hmg/ddv013> (2015).
- Dias, S. *et al.* Molecular biomarkers for gestational diabetes mellitus. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms19102926> (2018).
- Wu, P. *et al.* Maternal genome-wide DNA methylation profiling in gestational diabetes shows distinctive disease-associated changes relative to matched healthy pregnancies. *Epi-genetics* **13**(2), 122–128. <https://doi.org/10.1080/15592294.2016.1166321> (2018).
- Kang, J. *et al.* Genome-wide DNA methylation variation in maternal and cord blood of gestational diabetes population. *Diabetes Res. Clin. Pract.* **132**, 127–136. <https://doi.org/10.1016/j.diabres.2017.07.034> (2017).
- Kang, J. *et al.* Association of interleukin-10 methylation levels with gestational diabetes in a Taiwanese population. *Front. Genet.* **9**, 222. <https://doi.org/10.3389/fgene.2018.00222> [publishedOnlineFirst:2018/07/11] (2018).
- Enquobahrie, D. A. *et al.* Early pregnancy maternal blood DNA methylation in repeat pregnancies and change in gestational diabetes mellitus status—a pilot study. *Reprod. Sci.* **22**(7), 904–910. <https://doi.org/10.1177/1933719115570903> [publishedOnlineFirst:2015/02/14] (2015).
- International Association of D, Pregnancy Study Groups Consensus, P., Metzger, B.E. *et al.* International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care* **33**(3), 676–682. <https://doi.org/10.2337/dc09-1848> (2010).
- Li, E., Luo, T. & Wang, Y. Identification of diagnostic biomarkers in patients with gestational diabetes mellitus based on transcriptome gene expression and methylation correlation analysis. *Reprod. Biol. Endocrinol.* **17**(1), 112. <https://doi.org/10.1186/s12958-019-0556-x> (2019).

23. Naidoo, V., Naidoo, M. & Ghai, M. Cell- and tissue-specific epigenetic changes associated with chronic inflammation in insulin resistance and type 2 diabetes mellitus. *Scand. J. Immunol.* **88**(6), e12723. <https://doi.org/10.1111/sji.12723> [publishedOnlineFirst: 2018/12/28] (2018).
24. de la Garza, A. L. *et al.* Modulation of hyperglycemia and TNF $\alpha$ -mediated inflammation by helichrysum and grapefruit extracts in diabetic db/db mice. *Food Funct.* **5**(9), 2120–2128. <https://doi.org/10.1039/c4fo00154k> (2014).
25. Yang, I. V. *et al.* Epigenetic marks of in utero exposure to gestational diabetes and childhood adiposity outcomes: The EPOCH study. *Diabet. Med.* **35**(5), 612–620. <https://doi.org/10.1111/dme.13604> (2018).
26. Haertle, L. *et al.* Epigenetic signatures of gestational diabetes mellitus on cord blood methylation. *Clin. Epigenet.* **9**, 28. <https://doi.org/10.1186/s13148-017-0329-3> (2017).
27. Dalfra, M. G. *et al.* Genetics and epigenetics: New insight on gestational diabetes mellitus. *Front. Endocrinol. (Lausanne)* **11**, 602477. <https://doi.org/10.3389/fendo.2020.602477> (2020).
28. Spicer, A. P., Joo, A. & Bowling, R. A. Jr. A hyaluronan binding link protein gene family whose members are physically linked adjacent to chondroitin sulfate proteoglycan core protein genes: the missing links. *J. Biol. Chem.* **278**(23), 21083–21091. <https://doi.org/10.1074/jbc.M213100200> (2003).
29. Bogdani, M. *et al.* Extracellular matrix components in the pathogenesis of type 1 diabetes. *Curr. Diab. Rep.* **14**(12), 552. <https://doi.org/10.1007/s11892-014-0552-7> (2014).
30. Jurgens, C. A. *et al.* beta-cell loss and beta-cell apoptosis in human type 2 diabetes are related to islet amyloid deposition. *Am. J. Pathol.* **178**(6), 2632–2640. <https://doi.org/10.1016/j.ajpath.2011.02.036> (2011).
31. Hull, R. L. *et al.* Hyaluronan: A mediator of islet dysfunction and destruction in diabetes?. *J. Histochem. Cytochem.* **63**(8), 592–603. <https://doi.org/10.1369/0022155415576542> (2015).
32. Keller, B. & Adamski, J. RDH12, a retinol dehydrogenase causing Leber's congenital amaurosis, is also involved in steroid metabolism. *J. Steroid Biochem. Mol. Biol.* **104**(3–5), 190–194. <https://doi.org/10.1016/j.jsbmb.2007.03.015> (2007).
33. Jörnvall, H. *et al.* Short-chain dehydrogenases/reductases (SDR). *Biochemistry* **34**(18), 6003–6013. <https://doi.org/10.1021/bi00018a001> (1995).
34. Duester, G. Families of retinoid dehydrogenases regulating vitamin A function: Production of visual pigment and retinoic acid. *Eur. J. Biochem.* **267**(14), 4315–4324. <https://doi.org/10.1046/j.1432-1327.2000.01497.x> (2000).
35. Blumentrath, J., Neye, H. & Verspohl, E. J. Effects of retinoids and thiazolidinediones on proliferation, insulin release, insulin mRNA, GLUT 2 transporter protein and mRNA of INS-1 cells. *Cell Biochem. Funct.* **19**(3), 159–169. <https://doi.org/10.1002/cbf.907> (2001).
36. Berry, D. C. & Noy, N. All-trans-retinoic acid represses obesity and insulin resistance by activating both peroxisome proliferation-activated receptor beta/delta and retinoic acid receptor. *Mol. Cell Biol.* **29**(12), 3286–3296. <https://doi.org/10.1128/MCB.01742-08> (2009).
37. Elistratova, I. V. *et al.* Heat shock protein HSP40 family chaperone DNAJB6/MRJ expression analysis in blood cells obtained from patients with atopic dermatitis in different phases. *Patologicheskaja Fiziologija Eksperimental'naja Terapija* **60**(3), 23–30 (2016).
38. Kurucz, I. *et al.* Decreased expression of heat shock protein 72 in skeletal muscle of patients with type 2 diabetes correlates with insulin resistance. *Diabetes* **51**(4), 1102–1109. <https://doi.org/10.2337/diabetes.51.4.1102> (2002).
39. Stokov, I. A. *et al.* The function of endogenous protective systems in patients with insulin-dependent diabetes mellitus and polyneuropathy: Effect of antioxidant therapy. *Bull. Exp. Biol. Med.* **130**(10), 986–990 (2000).
40. Zilae, M. & Shirali, S. Heat shock proteins and diabetes. *Can. J. Diabetes* **40**(6), 594–602. <https://doi.org/10.1016/j.cjcd.2016.05.016> (2016).
41. Rao, A., Luo, C., Hogan, P.G. TRANSCRIPTION FACTORS OF THE NFAT FAMILY: Regulation and function. **15**(1), 707–747 <https://doi.org/10.1146/annurev.immunol.15.1.707> (1997).
42. Kim, H. B. *et al.* Lipin 1 represses NFATc4 transcriptional activity in adipocytes to inhibit secretion of inflammatory factors. *Mol. Cell Biol.* **30**(12), 3126–3139. <https://doi.org/10.1128/MCB.01671-09> (2010).
43. Kim, H. B. *et al.* NFATc4 and ATF3 negatively regulate adiponectin gene expression in 3T3-L1 adipocytes. *Diabetes* **55**(5), 1342–1352. <https://doi.org/10.2337/db05-1507> (2006).
44. Galtier-Dereure, F., Boegner, C. & Bringer, J. Obesity and pregnancy: complications and cost. *Am. J. Clin. Nutr.* **71**(5 Suppl), 1242s–s1248. <https://doi.org/10.1093/ajcn/71.5.1242s> (2000).
45. Di Cianni, G. *et al.* Intermediate metabolism in normal pregnancy and in gestational diabetes. *Diabetes Metab. Res. Rev.* **19**(4), 259–270. <https://doi.org/10.1002/dmrr.390> (2003).

## Acknowledgements

The authors would like to thank all research participants for their contributions and time.

## Author contributions

X.W. wrote this article, H.T. and S.M. designed the study, and H.T. supervised the study and guaranteed the work. X.W., M.S., J.H., Y.Z., S.L. and H.L. analyzed the data. X.W., S.M., H.L., J.H., N.Z., M.T. and X.W. participated in the collection of basic data and samples. X.W., S.M., S.L., H.L., J.H., N.Z., M.T. and R.A. participated in the discussion and reviewed/edited the manuscript. X.W., Y.Z., H.T. and S.M. Participate in the revision of the revised draft. The final manuscript read and approved by all authors.

## Funding

This work was supported by Natural Science Foundation of Hunan Province (Grant #2020JJ5789 and #2020JJ5954), National Natural Science Foundation of China (Grant #81773535), Hunan Provincial Grant for Innovative Province Construction (2019SK4012).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-99836-2>.

**Correspondence** and requests for materials should be addressed to S.M. or H.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021