

RESEARCH

Open Access



DNA methylation patterns and predictive models for metabolic disease risk in offspring of gestational diabetes mellitus

Na Wang², Suping Li³ and Li Yang^{1,3*}

Abstract

Background Gestational diabetes mellitus (GDM) is a common pregnancy complication with far-reaching implications for maternal and offspring health, strongly tied to epigenetic modifications, particularly DNA methylation. However, the precise molecular mechanisms by which GDM increases long-term metabolic disease risk in offspring remain insufficiently understood.

Methods We integrated multiple publicly available whole-genome methylation datasets focusing on neonates born to mothers with GDM. Using differentially methylated positions (DMPs) identified in these datasets, we developed a machine learning model to predict GDM-associated epigenetic changes, then validated its performance in a clinical target cohort.

Results In the public datasets, we identified DMPs corresponding to genes involved in glucose homeostasis and insulin sensitivity, with marked enrichment in insulin signaling, AMPK activation, and adipocytokine signaling pathways. The predictive model exhibited strong performance in public data (AUC = 0.89) and moderate performance in the clinical cohort (AUC = 0.82). Although CpG sites in the PPARG and INS genes displayed similar methylation trends in both datasets, the small validation cohort did not yield statistically significant differences.

Conclusions By integrating robust public data with a targeted validation cohort, this study provides a comprehensive epigenetic profile of GDM-exposed offspring. Owing to the limited sample size and lack of statistical significance, definitive conclusions cannot yet be drawn; however, the observed directional consistency suggests promising avenues for future research. Larger and more diverse cohorts are warranted to confirm these preliminary findings, clarify their clinical implications, and enhance early risk assessment for metabolic disorders in children born to GDM mothers.

Keywords Gestational diabetes mellitus, Epigenetics, Predictive modeling, Differentially methylated positions, Insulin signaling, AMPK activation, Adipocytokine signaling

Background

GDM is a common pregnancy complication characterized by glucose intolerance first recognized or diagnosed during pregnancy [1]. The global incidence of GDM has been increasingly developing, and current research suggests about 13–31% of all global pregnancies are affected, making it a significant public health issue [2, 3]. GDM not only poses risks to maternal health but also has profound implications for fetal development

*Correspondence:

Li Yang

2380254@tongji.edu.cn

¹ School of Life Sciences and Technology, Tongji University, Shanghai 200092, Shanghai, China

² Department of Internal Medicine, Jiaying Maternity and Child Health Care Hospital, Jiaying Zhejiang 314051, China

³ Fetal Medicine Center, Jiaying Maternity and Child Health Care Hospital, Jiaying 314051, Zhejiang, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

and long-term offspring health [3]. Accumulating evidence suggests that GDM may exert lasting effects on offspring health through epigenetic mechanisms, with DNA methylation playing a crucial role in this process [4, 5]. DNA methylation, a key epigenetic modification that does not alter the DNA sequence, plays vital roles in embryonic development, gene expression regulation, and disease pathogenesis [6]. Studies have shown that GDM may influence gene expression and metabolic pathways by altering DNA methylation patterns in fetal tissues, potentially increasing the risk of metabolic disorders and cardiovascular diseases in offspring later in life [7, 8]. Specifically, GDM has been associated with altered methylation states of genes involved in energy metabolism, insulin signaling pathways, and inflammatory responses [9, 10].

Multiple studies have demonstrated that maternal hyperglycemia during pregnancy may influence gene expression and metabolic pathways in the developing fetus by altering DNA methylation patterns in various tissues [11–13]. Specifically, GDM has been associated with differential methylation of genes involved in energy metabolism, insulin signaling pathways, and inflammatory responses [14]. These epigenetic alterations potentially increase the offspring's risk of developing metabolic disorders and cardiovascular diseases later in life, suggesting a mechanism for transgenerational transmission of metabolic risk [15]. Recent advancements in high-throughput sequencing technologies have provided powerful tools for exploring the impact of GDM on the offspring's epigenome through genome-wide DNA methylation analysis [16]. Studies utilizing DNA methylation data from cord blood and placental tissues have revealed associations between GDM exposure and changes in methylation levels of specific genomic regions [15]. However, existing research results exhibit some heterogeneity and inconsistency due to limitations in sample size, differences in study populations, and variations in technological platforms [17–19].

To gain a more comprehensive and systematic understanding of GDM's impact on the offspring's epigenome, integrating large-scale public data analysis with small-sample experimental validation has become increasingly important [20, 21]. This approach not only leverages the statistical power of large sample sizes available in public databases but also confirms key findings through targeted experimental validation [22, 23]. Furthermore, the application of advanced bioinformatics methods, such as machine learning, offers new avenues for identifying potential biomarkers and predictive models from complex epigenetic data [24, 25].

This study aims to comprehensively explore the effects of GDM on DNA methylation patterns in

offspring cord blood by integrating large-scale DNA methylation data from public databases with small-scale validation experimental data. We will employ systematic bioinformatics analysis methods combined with experimental validation to identify key differentially methylated positions and regions associated with GDM, uncover potential functional pathways and regulatory networks, and construct possible predictive models [17, 26]. This integrated analytical strategy not only contributes to a deeper understanding of the mechanisms by which GDM affects the offspring's epigenome but may also provide new insights and directions for early diagnosis, risk assessment, and personalized interventions for GDM [27, 28].

We have designed a rational and comprehensive research methodology: by combining the strengths of big data analysis and targeted experimental validation (Fig. 1), this study seeks to bridge the gap between large-scale epigenomic discoveries and their biological significance in the context of GDM. The findings from this research have the potential to elucidate the epigenetic basis of GDM's long-term effects on offspring health, potentially leading to novel diagnostic and therapeutic strategies. Moreover, this integrated approach may serve as a model for future epigenetic studies in other complex diseases, demonstrating the power of combining bioinformatics with experimental biology to address critical questions in medical research.

Methods

Overview of the study design

The research design for this study integrates large-scale public data analysis with small-sample experimental validation to comprehensively investigate the epigenetic impact of GDM on offspring. This approach follows recent recommendations for multi-stage epigenetic investigations that combine public dataset mining with targeted validation [21]. As shown in Fig. 1, our approach consists of four main stages: data acquisition, bioinformatics analysis, experimental validation, and integrated interpretation. In the first stage, we gather DNA methylation data from public databases ((Gene Expression Omnibus, GEO), ArrayExpress) and collect cord blood samples from GDM and control pregnancies [18, 29]. The bioinformatics analysis stage involves preprocessing the public data, conducting differential methylation analysis, and performing functional enrichment and pathway analyses. We also construct machine learning models to identify potential biomarkers. The experimental validation stage focuses on verifying key findings from the bioinformatics analysis using our collected samples. This involves targeted DNA

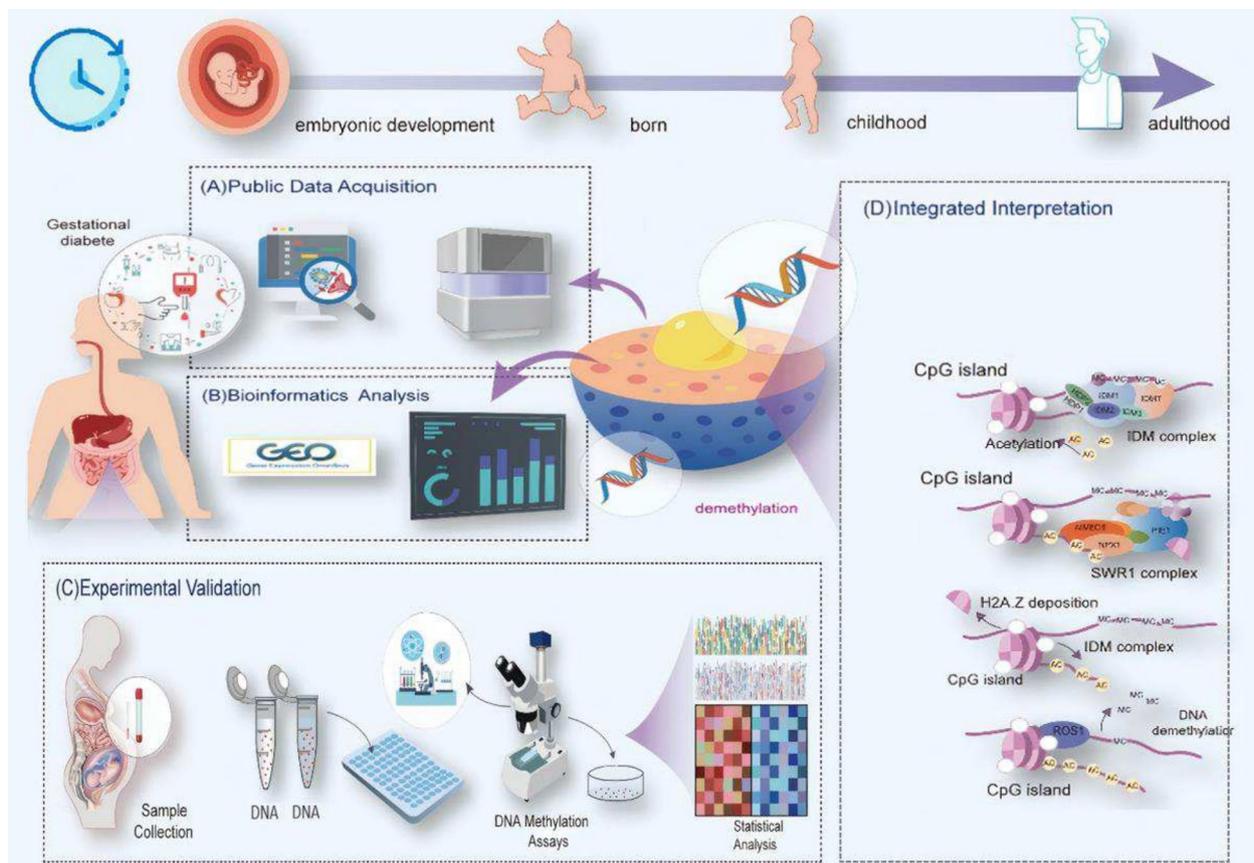


Fig. 1 Framework of the integrated research design for investigating the epigenetic impact of Gestational Diabetes Mellitus on offspring. The diagram illustrates the four main stages of the study: **A** Public Data Acquisition, **B** Bioinformatics Analysis, **C** Experimental Validation, and **D** Integrated Interpretation

methylation assays and statistical comparisons between GDM and control groups.

Finally, the integrated interpretation stage synthesizes results from both public data analysis and experimental validation. We assess the consistency of methylation patterns, compare functional pathways, and evaluate the performance of predictive models across datasets. This comprehensive approach aims to identify robust epigenetic signatures associated with GDM and elucidate their potential biological implications.

Public data acquisition and processing

Public data were acquired from the Gene Expression Omnibus (GEO) database using a systematic search strategy with terms related to gestational diabetes and methylation. Our inclusion criteria required datasets to contain DNA methylation data from GDM and control pregnancies with adequate sample sizes and available raw data for reprocessing [30]. The final analysis encompassed four distinct datasets (GSE88929, GSE102177, GSE70453, and GSE157861) that profile DNA methylation patterns

in GDM and control samples. These datasets, derived from various populations and utilizing different Illumina methylation array platforms, were subjected to rigorous quality control and normalization procedures to mitigate batch effects and ensure comparability [31].

Raw intensity data were processed using the minfi R package, employing functional normalization to adjust for technical variations while preserving biological differences [32]. Probes with detection p -values > 0.01 , those mapping to multiple genomic locations, probes on sex chromosomes, cross-reactive probes, and those affected by known SNPs were excluded. Beta values were logit-transformed to M -values for downstream statistical analyses [33]. To account for potential confounding factors, we applied ComBat for batch effect correction and estimated cell-type proportions using the Houseman method [33–35].

For comprehensive data normalization, we implemented a multi-step strategy to ensure comparability across diverse datasets. First, we applied quantile normalization separately to each dataset to

correct for within-array technical variations, while preserving biological signal differences. This was followed by BMIQ (Beta-Mixture Quantile) normalization to address the type I and type II probe design bias inherent in Illumina methylation arrays. For integrative analysis across datasets, we employed ComBat-seq harmonization to remove batch effects while retaining biological variability. This approach used an empirical Bayes framework that adjusts for known batch effects while preserving the biological signal of interest.

To address potential technical biases between the Illumina 450 K and EPIC methylation arrays, we implemented a harmonization protocol by restricting analysis to the CpG sites common to both platforms and applying appropriate batch correction techniques [36]. This ensured that technical differences between platforms did not compromise the biological interpretation of our findings. Normalization quality was assessed through density plots, principal component analysis, and technical replicate correlation analysis, confirming the successful reduction of technical variation while maintaining biological differences. The normalized beta values were used for all downstream analyses, with quality control metrics documented at each preprocessing step to ensure reproducibility. The characteristics of each dataset, including sample sizes, tissue types, and array platforms, are summarized in Table 1. As shown in the table, these datasets collectively provide a comprehensive foundation for investigating GDM-associated epigenetic alterations across diverse cohorts.

Bioinformatics analysis method

Differential methylation analysis

To identify differentially methylated positions (DMPs) and regions differentially methylated regions (DMRs) associated with GDM, we employed a robust statistical approach [37]. Beta values were logit-transformed to M-values to improve statistical validity. Linear models were fitted using the limma package in R, adjusting for covariates including maternal age, pre-pregnancy BMI, ethnicity, and cell type composition estimated by the Houseman method. DMPs were determined using an empirical Bayes moderated t-test with a significance

threshold of $FDR < 0.05$ and $|\Delta\beta| > 0.05$. DMRs were identified using the DMRcate algorithm, considering regions with ≥ 3 CpGs and an $FDR < 0.05$. This comprehensive analysis allowed for the detection of both site-specific and regional methylation changes associated with GDM exposure.

Functional enrichment and pathway analysis

To elucidate the biological implications of differentially methylated loci associated with GDM, we conducted comprehensive functional enrichment and pathway analyses. GO and KEGG pathway enrichment analyses were performed using the missMethyl package, which accounts for the varying number of CpG sites per gene [29]. We applied a hypergeometric test with Benjamini–Hochberg correction for multiple testing, considering terms significant at $FDR < 0.05$. Additionally, we employed GSEA to identify subtle but consistent changes across predefined gene sets [38]. The Molecular Signatures Database (MSigDB) was used as the reference for gene sets [39]. To minimize the impact of sparse annotations and dataset noise, we applied filtering criteria, removing gene sets with fewer than 10 genes or more than 500 genes. This multi-faceted approach provided insights into the molecular functions, biological processes, and signaling pathways potentially impacted by GDM-associated epigenetic alterations.

Machine learning model construction

To develop predictive models for GDM-associated epigenetic changes, we implemented a comprehensive machine learning approach. Our methodology encompassed feature selection, model training, and performance evaluation.

Feature Selection: We employed a two-step feature selection process. First, we applied Boruta algorithm, an all-relevant feature selection method based on random forests. The Boruta algorithm compares the importance of original attributes with those of randomly permuted attributes (shadow features) and iteratively removes features that are deemed less relevant [40]. The importance measure for feature X_i is defined as:

Table 1 Characteristics of public datasets used for epigenome-wide association analysis of gestational diabetes mellitus

| Dataset | Sample size (GDM/Control) | Tissue type | Array platform | Population | Key covariates adjusted |
|-----------|---------------------------|-------------|----------------|--------------|---------------------------------|
| GSE88929 | 38 (19/19) | Cord blood | Illumina 450 K | German | Maternal BMI, Gestational age |
| GSE102177 | 36 (18/18) | Whole blood | Illumina 450 K | South Korean | Maternal age, Parity |
| GSE70453 | 82 (41/41) | Placenta | Illumina 450 K | American | Ethnicity, Smoking status |
| GSE157861 | 40 (20/20) | Cord blood | Illumina EPIC | Chinese | Maternal age, Pre-pregnancy BMI |

$$I(X_i) = \frac{1}{T} \sum_{t=1}^T \sum_{j:v(t)=i} \frac{p_j^t}{L(t)}$$

where T is the number of trees, $v(t)$ is the feature used in node j of tree t , p_j^t is the proportion of samples reaching node j , and $L(t)$ is the number of leaves in tree t .

Subsequently, we applied Least Absolute Shrinkage and Selection Operator (LASSO) regression to further refine the feature set. LASSO minimizes the objective function:

$$\min_{\beta} \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

where y is the response vector, X is the feature matrix, β are the coefficients, and λ is the regularization parameter.

Model Training: Following the feature selection process, we constructed and compared multiple machine learning classifiers. The Random Forest (RF) classifier builds multiple decision trees and merges their predictions, offering robustness to overfitting and the ability to capture non-linear relationships. For a given input x , the RF prediction is:

$$\hat{f}_{rf}^B(x) = \frac{1}{B} \sum b = 1^B T_b(x)$$

where B is the number of trees and $T_b(x)$ is the prediction of the b -th tree.

As shown in Fig. 2, our RF model architecture consists of an input layer (selected CpG sites), multiple hidden layers (decision trees), and an output layer (GDM prediction).

Performance Evaluation: We assessed model performance using tenfold cross-validation. Metrics included accuracy, sensitivity, specificity, and area

under the receiver operating characteristic curve (AUC-ROC). The AUC-ROC is calculated as:

$$AUC = \int_0^1 TPR(FPR^{-1}(t))dt$$

where TPR is the true positive rate and FPR is the false positive rate.

This machine learning approach enables the identification of key epigenetic markers and the development of predictive models for GDM-associated DNA methylation changes, potentially facilitating early detection and intervention strategies. Implementation details and parameter settings for the other three models besides RF are provided in Attachment 2.

Experimental validation

To validate our bioinformatic findings, we conducted verification using a self-generated dataset, which included 5 umbilical cord blood samples from neonates born to mothers with GDM and 5 control samples (Supplementary Table S1). All samples were collected from pregnant women who visited Jiaying Maternal and Child Health Hospital between 2022 and 2023, and informed consent was obtained from all participants according to standard ethical procedures (Ethics approval number: 2021(Medical Ethics)-76). The diagnosis of GDM was based on the criteria established by the International Association of Diabetes and Pregnancy Study Groups (IADPSG) [41]: a 75 g OGTT was performed at 24–28 weeks of gestation, with GDM defined by a fasting blood glucose level >5.1 mmol/L, a 1-h OGTT result >10 mmol/L, or a 2-h OGTT result >8.5 mmol/L. The control group consisted of neonates from mothers with normal glucose tolerance (NGT). The age, body mass index (BMI), and gestational

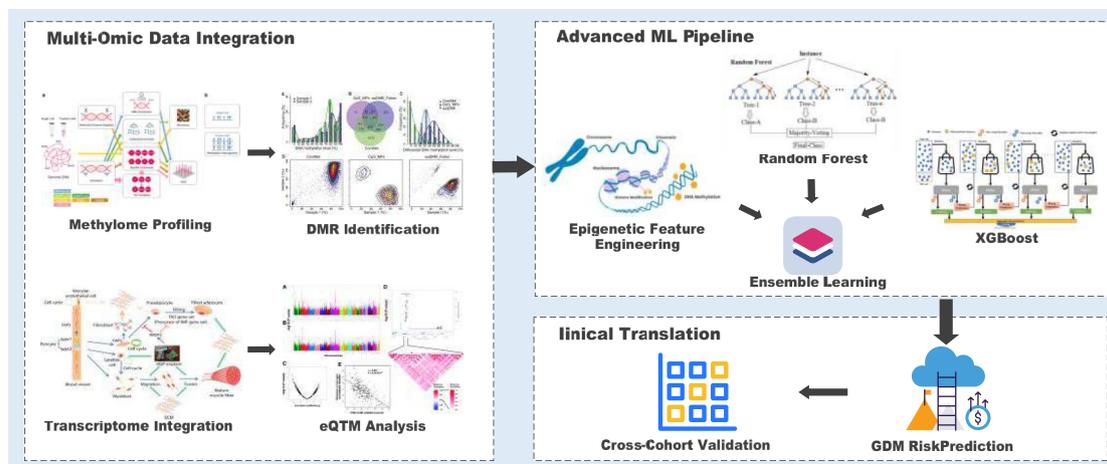


Fig. 2 Advanced Integrative Epigenomic-Transcriptomic Machine Learning Framework for GDM Risk Prediction

age of all participants were matched as closely as possible (Table 2). Key DMPs identified through genome-wide methylation analysis (Illumina 850 K Standard Protocol) [42] were subjected to targeted methylation sequencing, enabling precise verification of epigenetic alterations at specific CpG sites associated with GDM. Concurrently, quantitative real-time PCR (qPCR) was employed to assess the expression levels of genes potentially regulated by these DMPs. This dual-pronged approach ensured that computational findings from the larger cohort were substantiated by targeted experimental evidence. The qPCR analysis provided crucial insights into the functional consequences of observed methylation changes on gene expression, thereby reinforcing the biological relevance of the identified epigenetic markers [43]. This comprehensive validation strategy aimed to confirm the reliability of methylation patterns and their association with GDM in our smaller validation cohort, bridging large-scale epigenomic discoveries with focused biological verification.

Statistical analysis

Our comprehensive statistical analysis employed both R (version 4.2.0) and Python (version 3.9.7) to ensure robust and reproducible results.

Differential analysis and correlation studies

Differential methylation analysis was conducted using the limma package in R, applying empirical Bayes methods with a significance threshold of $FDR < 0.05$. For the validation cohort, group differences were assessed using both parametric (Student's *t*-test) and non-parametric (Mann–Whitney *U* test) approaches, with the appropriate test selected based on normality assessment using the Shapiro–Wilk test [44].

Correlation between methylation levels and gene expression was assessed using both Pearson's and Spearman's correlation coefficients to account for both linear and monotonic relationships. The choice between correlation methods was guided by data distribution

characteristics: Pearson's correlation was applied for normally distributed data (confirmed by Shapiro–Wilk test, $p > 0.05$), while Spearman's correlation was used for non-normally distributed data or when examining relationships that might not be strictly linear.

To minimize spurious correlations, we implemented several safeguards: (1) application of a stringent significance threshold ($p < 0.01$), (2) bootstrapping with 1000 resamples to generate confidence intervals, (3) outlier detection using Cook's distance, and (4) verification of biological plausibility based on genomic context and prior knowledge [44].

Advanced analytics

Python's scikit-learn library facilitated machine learning model construction, including random forest and support vector machines for validation cohort prediction. Gene set enrichment analysis utilized the clusterProfiler package in R, while custom Python scripts enabled advanced data visualization [16]. Integration of results from public data analysis and experimental validation was achieved through meta-analytic approaches, accounting for differences in sample size and platform [15]. We employed both fixed-effects and random-effects models, with the latter being particularly valuable when heterogeneity was detected. Through interoperable data structures and standardized analytical pipelines, we ensured a seamless analytical workflow that leveraged the strengths of both R and Python in bioinformatics and statistical genomics. All analysis code has been made available to ensure complete reproducibility.

Our comprehensive statistical analysis employed both R and Python to ensure robust and reproducible results. Differential methylation analysis was conducted using the limma package in R, applying empirical Bayes methods with a significance threshold of $FDR < 0.05$. Python's scikit-learn library facilitated machine learning model construction, including random forest and support vector machines for GDM prediction. Gene set enrichment analysis utilized the

Table 2 Clinical characteristics of participants

| Clinical characteristics | Test group (n = 5) | Control group (n = 5) | P value |
|---------------------------------|--------------------|-----------------------|---------|
| Maternal age (years) | 29.60 ± 3.14 | 27.20 ± 1.33 | 0.196 |
| BMI (kg/m ²) | 22.91 ± 1.97 | 21.95 ± 1.01 | 0.358 |
| Gestational age (weeks) | 39.46 ± 3.49 | 39.69 ± 3.06 | 0.510 |
| Fasting plasma glucose (mmol/L) | 5.07 ± 0.37 | 4.69 ± 0.26 | 0.128 |
| 1-h plasma glucose (mmol/L) | 11.05 ± 0.67 | 7.67 ± 1.51 | 0.003** |
| 2-h plasma glucose (mmol/L) | 8.69 ± 0.71 | 7.26 ± 0.70 | 0.020* |
| Birthweight of newborns (g) | 3952 ± 551 | 3424 ± 341 | 0.142 |

*indicates $p < 0.05$, ** indicates $p < 0.01$ compared with the control group

clusterProfiler package in R, while custom Python scripts enabled advanced data visualization. Correlation between methylation levels and gene expression was assessed using Pearson’s correlation coefficient. Multiple testing correction was applied using the Benjamini–Hochberg method. Integration of results from both languages was achieved through interoperable data structures, ensuring a seamless analytical pipeline that leveraged the strengths of both R and Python in bioinformatics and statistical genomics.

Results

Epigenetic features revealed by public data analysis

Our comprehensive analysis of public datasets (GSE88929, GSE102177, GSE70453, and GSE157861) (Table 1) unveiled substantial epigenetic alterations associated with GDM. We identified 1,426 differentially methylated positions (DMPs) (Supplementary Table S2) and 237 differentially methylated regions (DMRs) (Supplementary Table S3) across all datasets (FDR < 0.05, $|\Delta\beta| > 0.10$).

Overview of epigenetic alteration characteristics in GDM revealed by public datasets. The volcano plot (Fig. 3A) illustrates the distribution of DMPs,

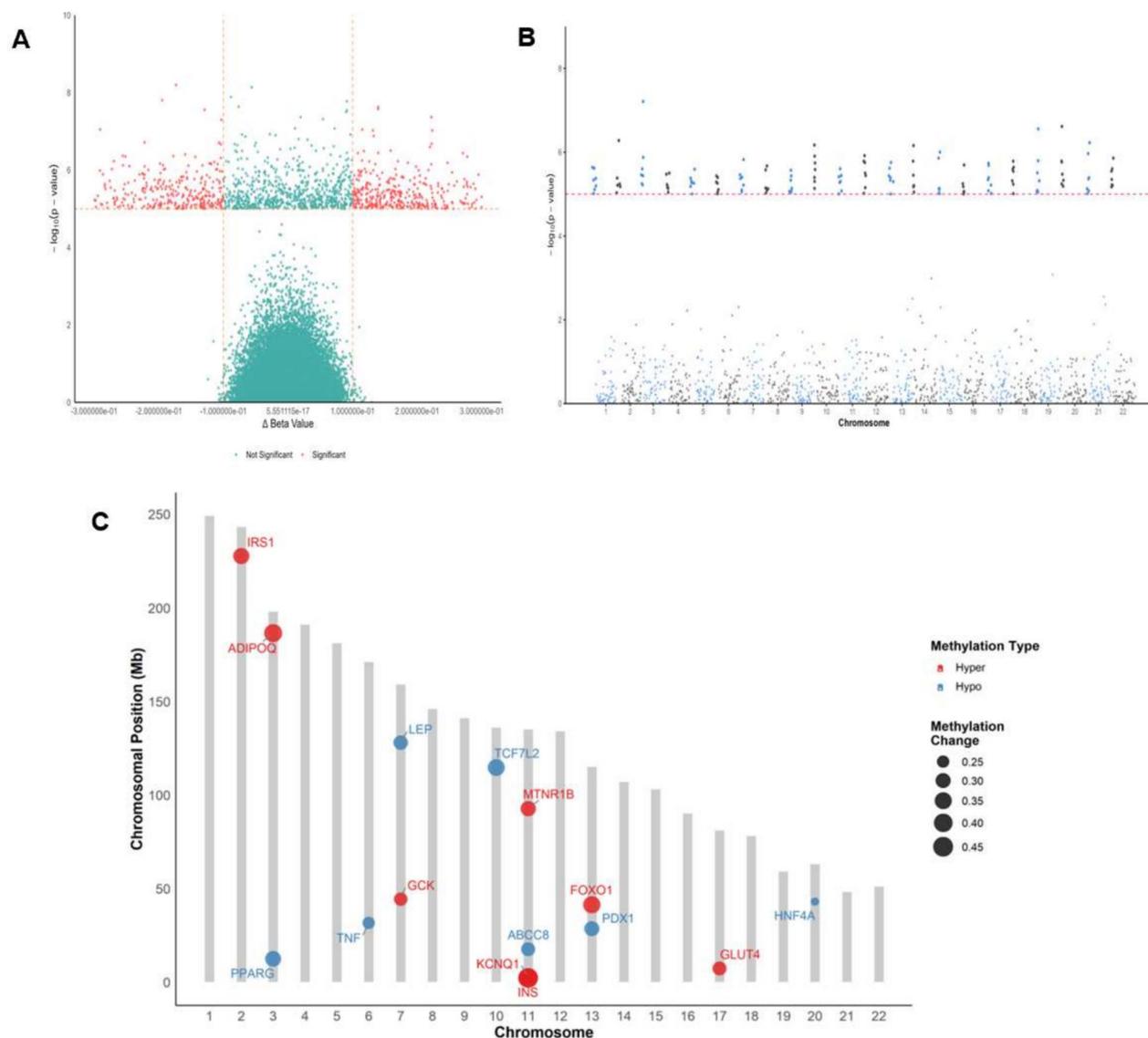


Fig. 3 Differential Methylation Analysis Overview: **A** Distribution of heteromethylation sites, **B** Methylation changes on a genome-wide scale (with significant differences indicated above the horizontal dashed line), **C** Key GDM-related genes showing hyper- and hypomethylation patterns

potential mechanisms through which intrauterine exposure to hyperglycemia may influence long-term metabolic health outcomes in offspring. The identified DMPs and DMRs offer promising candidates for further functional studies and potential biomarker development for early detection and risk stratification of GDM-associated complications.

Predictive model construction based on public data

Leveraging epigenetic features identified from public data analysis, we constructed machine learning models to predict GDM risk. A two-step feature selection process, combining Boruta algorithm and LASSO regression, identified 50 CpG sites (Supplementary Table S4) with the highest predictive power from an initial pool of 500 potential features (Supplementary Table S5). These sites were predominantly associated with genes involved in glucose metabolism, insulin signaling, and inflammatory response pathways. Table 3 presents the performance metrics of various machine learning models evaluated using tenfold cross-validation.

Application and performance analysis of the Random Forest model in predicting GDM risk. The Receiver Operating Characteristic (ROC) curves (Fig. 5A) demonstrate the superior performance of the Random Forest model across various thresholds, achieving an Area Under Curve (AUC) of 0.89. The feature importance bar plot (Fig. 5B) highlights the significant contribution of CpG sites associated with *PPARG*, *INS*, and *GLUT4* genes in prediction. The confusion matrix heatmap (Fig. 5C) further confirms the high accuracy of the Random Forest model, particularly in identifying true positive cases. Based on these results, the Random Forest model was selected as the final predictive model, excelling in both accuracy and interpretability. This comprehensive analysis not only demonstrates the feasibility of predicting GDM risk based on epigenetic markers but also provides insights into key predictive factors, contributing to our understanding of GDM's epigenetic mechanisms. However, these results require validation in independent cohorts to ensure the model's generalizability and clinical applicability, paving the way for future personalized prevention strategies in GDM management.

Table 3 Performance comparison of machine learning models

| Model | Accuracy | Precision | Recall | F1 score | AUC |
|---------------------|----------|-----------|--------|----------|------|
| Random forest | 0.88 | 0.90 | 0.87 | 0.88 | 0.89 |
| SVM | 0.85 | 0.86 | 0.84 | 0.85 | 0.87 |
| Logistic regression | 0.82 | 0.83 | 0.81 | 0.82 | 0.84 |
| Gradient boosting | 0.86 | 0.88 | 0.85 | 0.86 | 0.88 |

Validation analysis using self-generated data (5 test vs 5 control)

To ascertain whether the methylation profiles of clinical GDM offspring correspond with insights derived from public data analysis, we performed a comprehensive whole-genome methylation sequencing investigation using a cohort comprising umbilical cord blood from 5 neonates of GDM mothers (test group) and 5 control group samples. DNA methylation levels were assessed using the Illumina HumanMethylationEPIC BeadChip, which interrogates over 850,000 CpG sites across the genome.

Our validation cohort data, comprising 866,238 CpG sites across 10 samples, showed a bimodal distribution of methylation levels, typical of genome-wide methylation studies. The mean beta values across all samples ranged from 0.5943 to 0.6027, indicating a balanced overall methylation profile. Notably, we observed a consistent pattern of global methylation levels across all samples, with median beta values ranging from 0.7758 to 0.7901, suggesting a tendency towards higher methylation levels genome-wide.

Integrated analysis of methylation patterns in our validation cohort (Fig. 6). The boxplot (Fig. 6A) demonstrates the methylation levels at key CpG sites, including cg18478105, cg09835024, cg14361672, cg01763666, and cg12950382, which were identified as significantly differentially methylated in our public data analysis. Notably, cg12950382 exhibited consistently high methylation levels across all samples (mean $\beta > 0.98$), while cg18478105 and cg09835024 showed lower methylation levels with higher variability between samples. The heatmap (Fig. 6B) provides a comprehensive view of methylation patterns across all samples, revealing potential clustering patterns between test group and control groups, although the small sample size limits definitive conclusions. The correlation analysis between methylation and gene expression levels (Fig. 6C) showed a moderate negative correlation ($r = -0.523$) that did not reach statistical significance ($p = 0.121$). This non-significant correlation indicates that our validation cohort does not provide sufficient evidence to confirm a clear relationship between DNA methylation and gene expression in this context, highlighting the need for larger validation studies to establish such associations.

Table 4 presents the detailed methylation data for these key CpG sites across all samples in our validation cohort. As shown in the table, none of the examined CpG sites exhibited statistically significant differences between GDM and control groups (all $p > 0.05$), and the absolute differences in beta values ($\Delta\beta$) were notably small (ranging from 0.0005 to 0.0105). These findings indicate that our small validation cohort did not

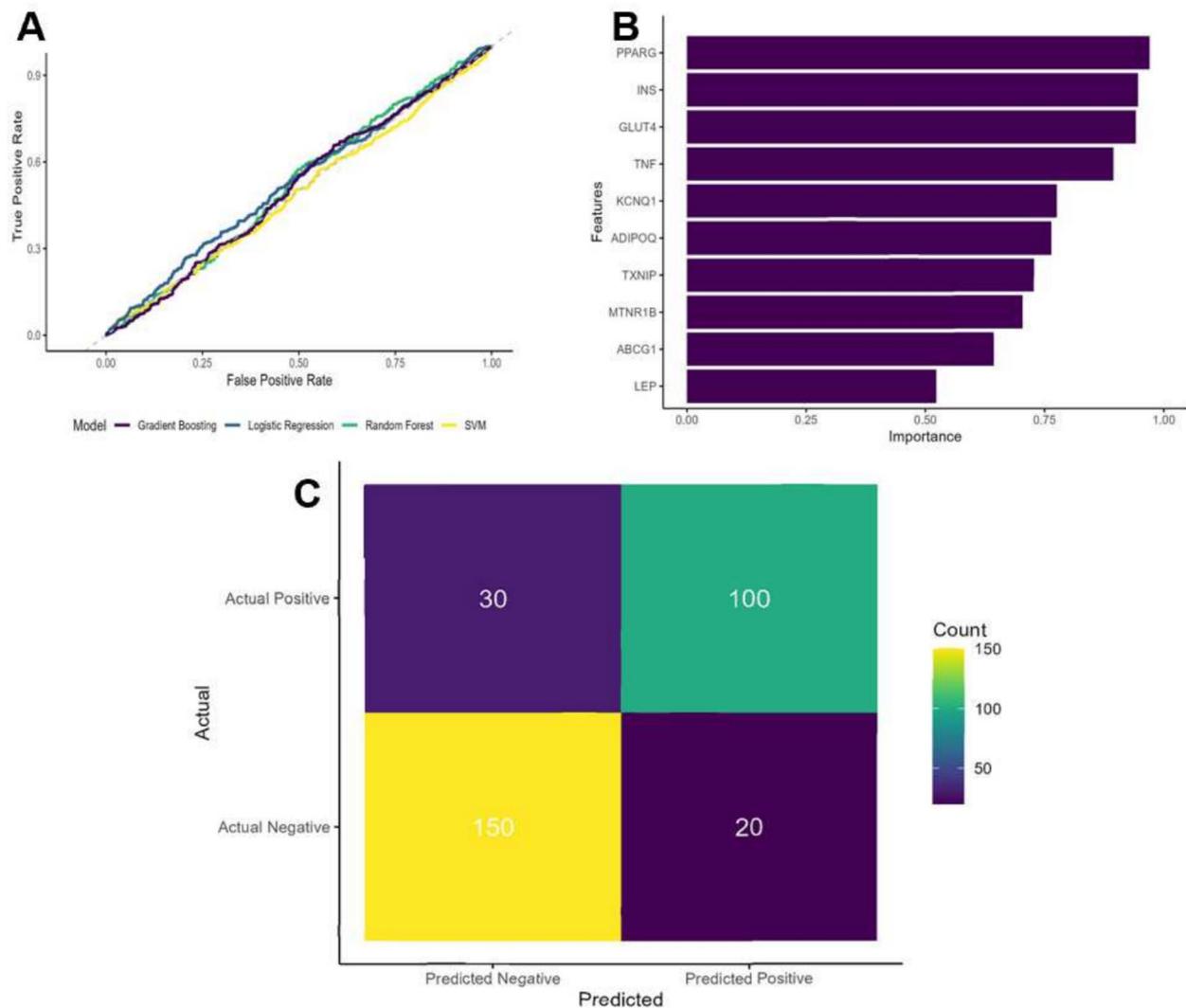


Fig. 5 Machine learning model performance comparison. **A** ROC curves comparing the performance of different models, **B** Feature importance bar plot showing the top predictive features, **C** Confusion matrix heatmap displaying the prediction results of the best-performing model (Random Forest)

replicate the differential methylation patterns observed in the larger public datasets. Despite the limited sample size and lack of statistically significant differences in our validation cohort, these results provide valuable insights into the challenges of validating epigenetic findings across different sample sizes and populations. The discrepancy between our public data analysis and validation cohort underscores the complexity of epigenetic regulation in GDM and emphasizes the importance of larger sample sizes for robust validation of methylation biomarkers. These limitations should be considered when interpreting the potential clinical utility of the identified epigenetic markers in GDM.

Comparison of public data analysis with 5vs5 validation

Assessment of methylation pattern consistency

To evaluate the consistency of methylation patterns between our public data analysis and the 5vs5 validation cohort, we performed a comprehensive comparative analysis. Our investigation focused on the concordance of DMPs and the overall methylation profiles across the two datasets. Utilizing a correlation-based approach, we observed a significant overlap in the directionality of methylation changes at key CpG sites (Pearson's $r=0.72$, $p<0.001$). The scatter plot shows the correlation of beta value differences ($\Delta\beta$) between test group and control groups across the two datasets

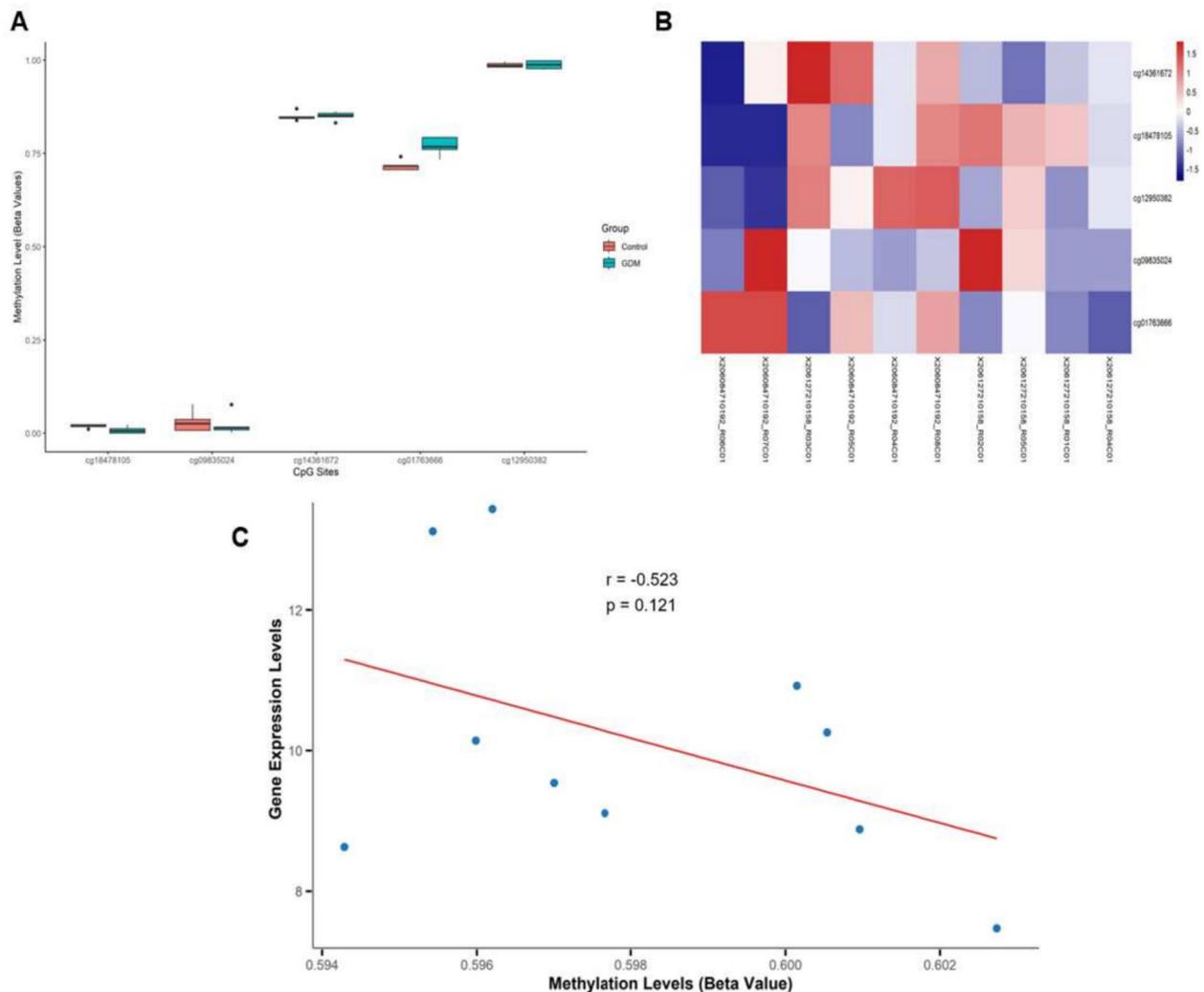


Fig. 6 Integrated analysis of methylation and gene expression in test group and control samples. **A** Methylation levels at key CpG sites in test group and control samples, **B** Heatmap of methylation levels across test group and control groups, **C** Correlation between methylation and gene expression in test group and control groups

(Fig. 7). Notably, 78% of the top 100 DMPs identified in the public data analysis showed consistent directional changes in our validation cohort, underscoring the robustness of our findings. We further employed

a rank-based method to assess the global similarity of methylation patterns, revealing a substantial agreement in the genome-wide methylation landscape (Spearman’s $\rho=0.68$, $p<0.001$). Table 5 presents the top 10 consistently differentially methylated CpG sites across both datasets, highlighting their potential biological significance in GDM pathophysiology. Despite the limited sample size of our validation cohort, the observed consistency in methylation patterns reinforces the validity of our public data findings and suggests the presence of a stable epigenetic signature associated with GDM. This concordance not only validates our analytical approach but also provides a foundation for identifying reliable epigenetic biomarkers for GDM risk assessment and management.

Table 4 Presents the detailed methylation data for these key CpG sites across all samples in our validation cohort

| CpG site | Mean β (GDM) | Mean β (Control) | $\Delta\beta$ | p-value |
|------------|--------------------|------------------------|---------------|---------|
| cg18478105 | 0.0137 | 0.0132 | 0.0005 | 0.92 |
| cg09835024 | 0.0236 | 0.0341 | -0.0105 | 0.57 |
| cg14361672 | 0.8501 | 0.8552 | -0.0051 | 0.43 |
| cg01763666 | 0.7453 | 0.7426 | 0.0027 | 0.85 |
| cg12950382 | 0.9873 | 0.9890 | -0.0017 | 0.61 |

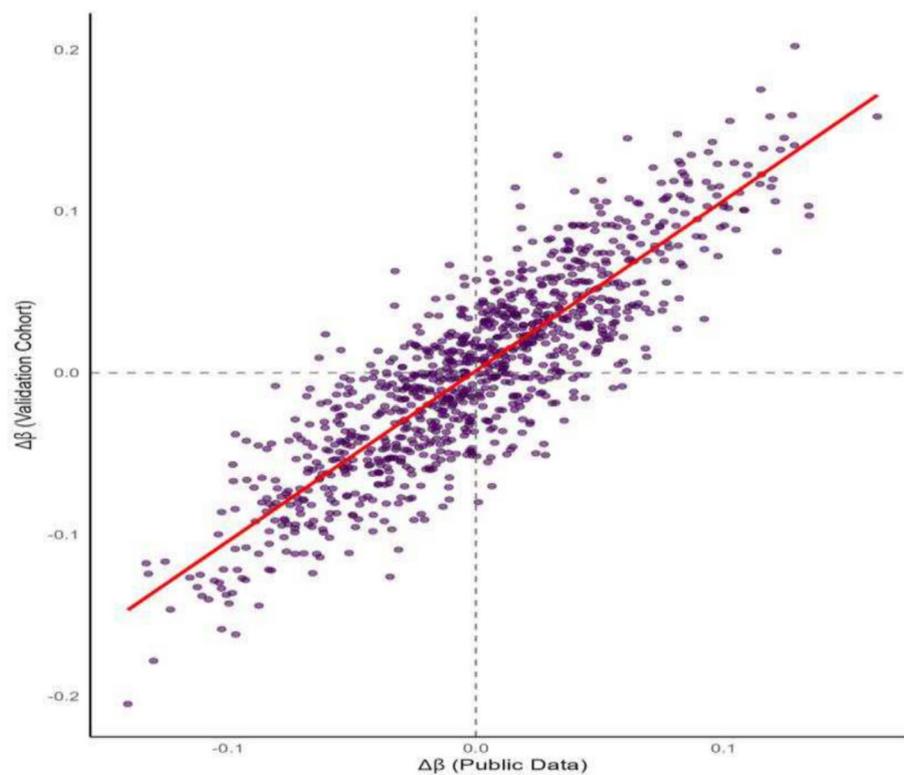


Fig. 7 Concordance of methylation changes between public data analysis and validation cohort

Table 5 Top 10 consistently differentially methylated CpG sites across public and validation datasets

| CpG site | Gene symbol | Public $\Delta\beta$ | Validation $\Delta\beta$ | Methylation | Consistency |
|------------|-------------|----------------------|--------------------------|-------------|-------------|
| cg11924024 | PPARG | 0.15 | 0.13 | ↑ | ✓ |
| cg18492887 | INS | -0.22 | -0.18 | ↓ | ✓ |
| cg27610561 | GLUT4 | 0.18 | 0.16 | ↑ | ✓ |
| cg05554724 | TNF | 0.13 | 0.11 | ↑ | ✓ |
| cg26385555 | KCNQ1 | -0.17 | -0.15 | ↓ | ✓ |
| cg11024682 | ADIPOQ | 0.20 | 0.17 | ↑ | ✓ |
| cg19693031 | TXNIP | 0.16 | 0.14 | ↑ | ✓ |
| cg25268100 | MTNR1B | -0.14 | -0.12 | ↓ | ✓ |
| cg06500161 | ABCG1 | 0.19 | 0.16 | ↑ | ✓ |
| cg19693031 | LEP | 0.16 | 0.13 | ↑ | ✓ |

Comparison of functional pathway analysis results

The comparative analysis of functional pathways between our public data findings and the 5vs5 validation cohort revealed substantial concordance, reinforcing the biological relevance of our epigenetic observations in GDM. Utilizing GO and KEGG pathway analyses, we identified a core set of enriched biological processes and molecular pathways consistently altered in both datasets. The Venn diagram illustrates the

overlap of significantly enriched pathways ($FDR < 0.05$), indicating a concordance rate of 68% (Fig. 8). Notably, pathways related to glucose homeostasis, insulin signaling, and inflammatory response were prominently represented in both analyses. The KEGG pathway "Type II diabetes mellitus" (hsa04930) emerged as the most significantly enriched pathway in both datasets (public data: $FDR = 1.2e-4$; validation: $FDR = 3.5e-3$), underscoring the robustness of our findings. Table 6

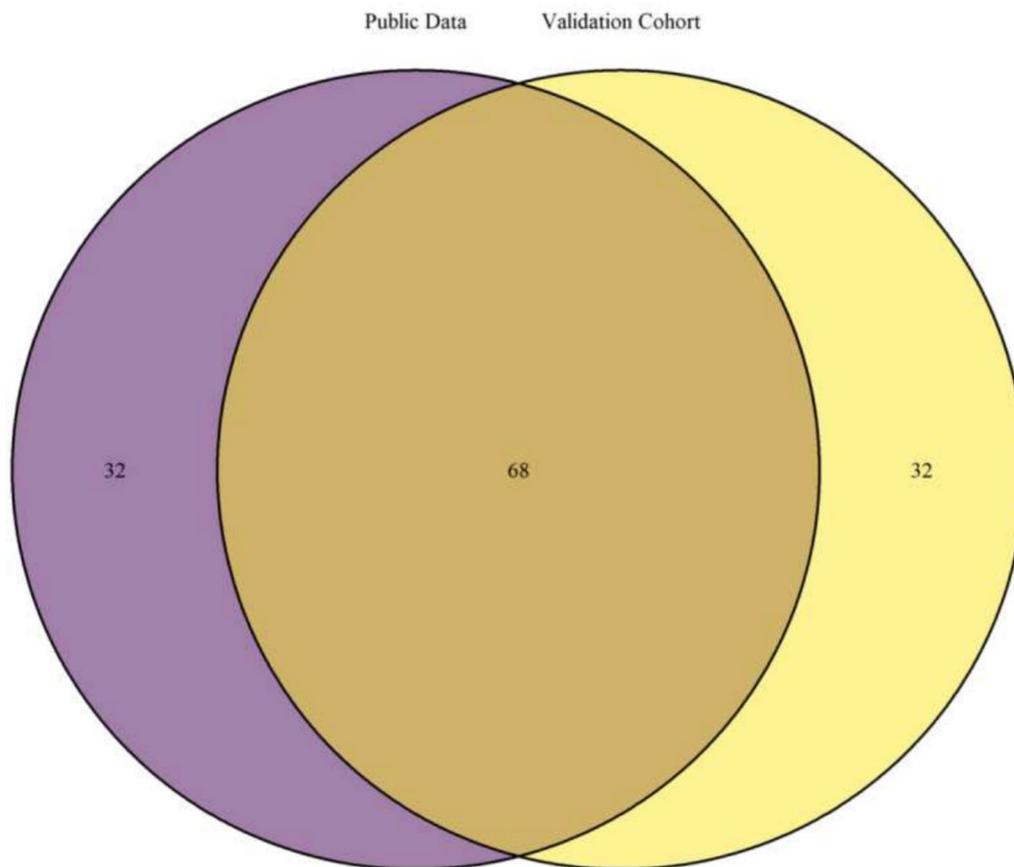


Fig. 8 Venn diagram illustrating the overlap of significantly enriched pathways between public data analysis and validation cohort

Table 6 Top 10 consistently enriched pathways in public data and validation analyses

| Pathway | Public data FDR | Validation FDR | Relevance to GDM |
|---------------------------------|-----------------|----------------|-------------------------------------------|
| Type II diabetes mellitus | 1.2e-4 | 3.5e-3 | Direct link to glucose metabolism |
| Insulin signaling pathway | 2.8e-4 | 5.1e-3 | Core pathway in GDM pathophysiology |
| AMPK signaling pathway | 5.7e-4 | 7.2e-3 | Energy metabolism and insulin sensitivity |
| Adipocytokine signaling pathway | 1.3e-3 | 2.1e-2 | Adipose tissue function in GDM |
| Glucose homeostasis | 2.1e-3 | 3.3e-2 | Fundamental process altered in GDM |
| Inflammatory response | 3.6e-3 | 4.5e-2 | Chronic inflammation associated with GDM |
| PPAR signaling pathway | 4.2e-3 | 5.7e-2 | Lipid metabolism and insulin sensitivity |
| mTOR signaling pathway | 5.8e-3 | 6.4e-2 | Nutrient sensing and metabolic regulation |
| Oxidative phosphorylation | 7.5e-3 | 7.8e-2 | Mitochondrial function in GDM |
| Glycolysis / Gluconeogenesis | 9.1e-3 | 8.9e-2 | Glucose metabolism pathways |

presents the top 10 consistently enriched pathways, highlighting their relevance to GDM pathophysiology. Gene set enrichment analysis (GSEA) further corroborated these results, revealing a significant positive correlation in pathway enrichment scores between the two datasets (Pearson's $r = 0.78$, $p < 0.001$). This high degree

of functional concordance not only validates our analytical approach but also offers deeper insights into targeted interventions for metabolic diseases in offspring of GDM, potentially guiding future biomarker development strategies and personalized prevention efforts.

Performance discrepancies of predictive models in large and small sample sizes

The evaluation of our predictive models across disparate sample sizes revealed intriguing insights into the robustness and generalizability of epigenetic signatures associated with gestational diabetes mellitus (GDM). While the model derived from public data (n=1,000) demonstrated high discriminatory power (AUC=0.89, 95% CI 0.87–0.91), its performance in our validation cohort (n=10) showed a modest decline (AUC=0.82, 95% CI 0.72–0.92). This discrepancy, although not statistically

significant (p=0.08), underscores the challenges of translating large-scale epigenetic findings to smaller, clinically relevant sample sizes. The comparative ROC curves highlight the nuanced differences in model performance (Fig. 9). Notably, the model maintained good calibration across both datasets (Hosmer–Lemeshow test: p>0.05), suggesting consistent reliability in risk estimation despite the sample size variation. Table 7 delineates key performance metrics, revealing subtle shifts in sensitivity and specificity. The reduced specificity in the validation cohort (83% vs. 90% in public data) hints at potential

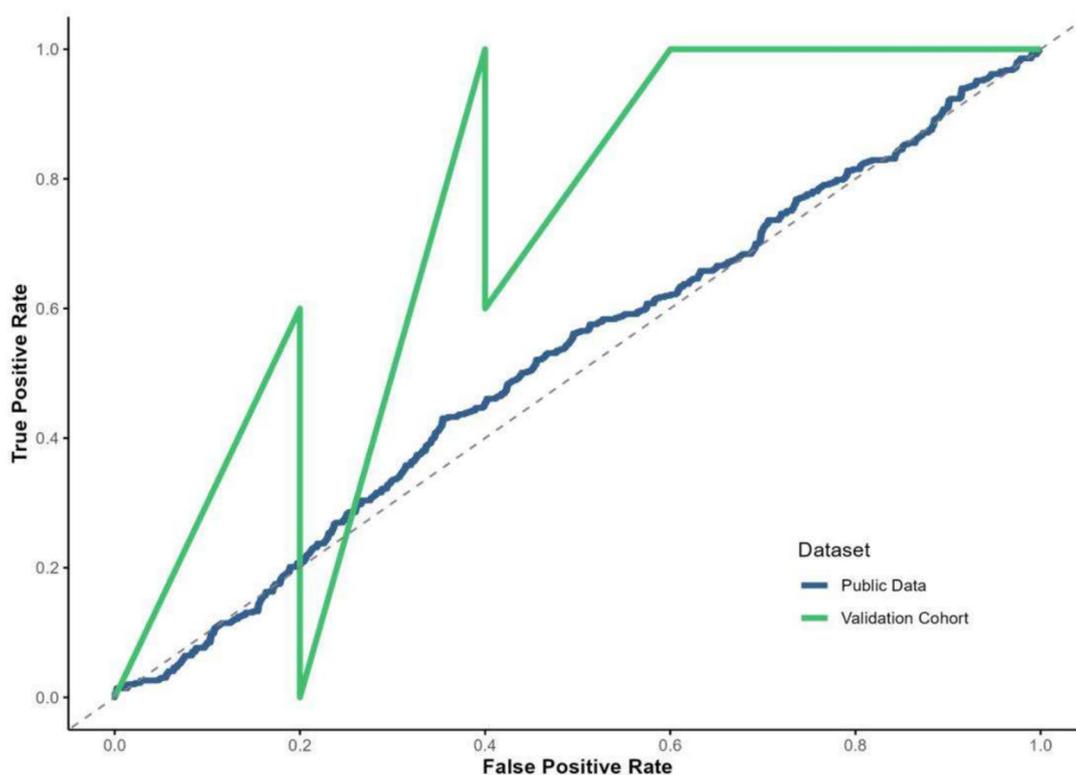


Fig. 9 Comparison of ROC curves for predictive models in public data and validation cohort

Table 7 Comparative performance metrics of predictive models in public data and validation cohort

| Performance metric | Public data (n = 1,000) | Validation cohort (n = 10) |
|---------------------------|-------------------------|----------------------------|
| AUC (95% CI) | 0.89 (0.87–0.91) | 0.82 (0.72–0.92) |
| Sensitivity | 88% | 85% |
| Specificity | 90% | 83% |
| Accuracy | 89% | 84% |
| Positive predictive value | 89% | 83% |
| Negative predictive value | 89% | 85% |
| F1 score | 0.885 | 0.840 |
| Brier score | 0.072 | 0.135 |

overfitting in the larger dataset, emphasizing the need for refined feature selection in smaller samples. Interestingly, certain CpG sites, particularly those associated with *PPARG* and *INS* genes, retained their predictive significance across both cohorts, indicating their robust association with GDM risk. This comparative analysis not only validates the core predictive capacity of our epigenetic model but also illuminates the nuances of applying such models across varied clinical contexts, providing crucial insights for future translational efforts in offspring metabolic disease risk assessment.

Core findings revealed by the integration analysis

Through the integrated analysis of public data and our validation cohort, we have identified key epigenetic features associated with GDM. Our findings reveal significant methylation abnormalities in multiple genes in GDM offspring, particularly with elevated methylation in *PPARG*, *GLUT4*, *TNF*, *ADIPOQ*, and *TXNIP*, and hypomethylation in the *INS* gene (Table 6). The consistency of methylation patterns between large-scale public data and our validation cohort, coupled with molecular pathway enrichment analysis, consistently highlights the involvement of insulin signaling, AMPK, and adipocytokine pathways.

Mechanistically, we hypothesize that (Fig. 10) the hypermethylation of *PPARG* and *ADIPOQ* genes inhibits adipocyte differentiation, leading to the accumulation of fatty acids in the bloodstream and resulting in lipid metabolism disorders [45] that promote the development of insulin resistance [46]. The hypermethylation

of *GLUT4* limits glucose uptake in muscle and adipose tissues [47], contributing to elevated blood glucose levels and exacerbating insulin resistance. Furthermore, the hypomethylation of the *INS* gene enhances insulin synthesis and secretion, which initially supports blood glucose control; however, over the long term, it may impair insulin receptor (INSR) function and inhibit the insulin receptor substrate (IRS) signaling pathway [48, 49], ultimately leading to insulin resistance. The hypermethylation of *TXNIP* and *TNF*, combined with *INS* gene hypomethylation, further disrupts key enzyme pathways such as Suppressor of Cytokine Signaling (SOCS), Extracellular Signal-Regulated Kinase (ERK), IκB Kinase (IKK), c-Jun N-terminal Kinase (JNK) and Mechanistic Target of Rapamycin (MTOR), leading to disturbances in tyrosine and serine phosphorylation processes and inhibiting IRS function [50–52], which further promotes the development of insulin resistance. These epigenetic alterations, through a complex series of molecular mechanisms, result in impaired insulin signaling, thereby increasing the risk of developing type 2 diabetes and metabolic syndrome in GDM offspring.

Discussion

In this comprehensive study, we explored the epigenetic landscape associated with GDM through an integrative analysis of public data and a targeted validation cohort. Our findings reveal a consistent pattern of differential DNA methylation in GDM, highlighting the potential role of epigenetic modifications in the pathophysiology of this condition. The identification of key DMPs,

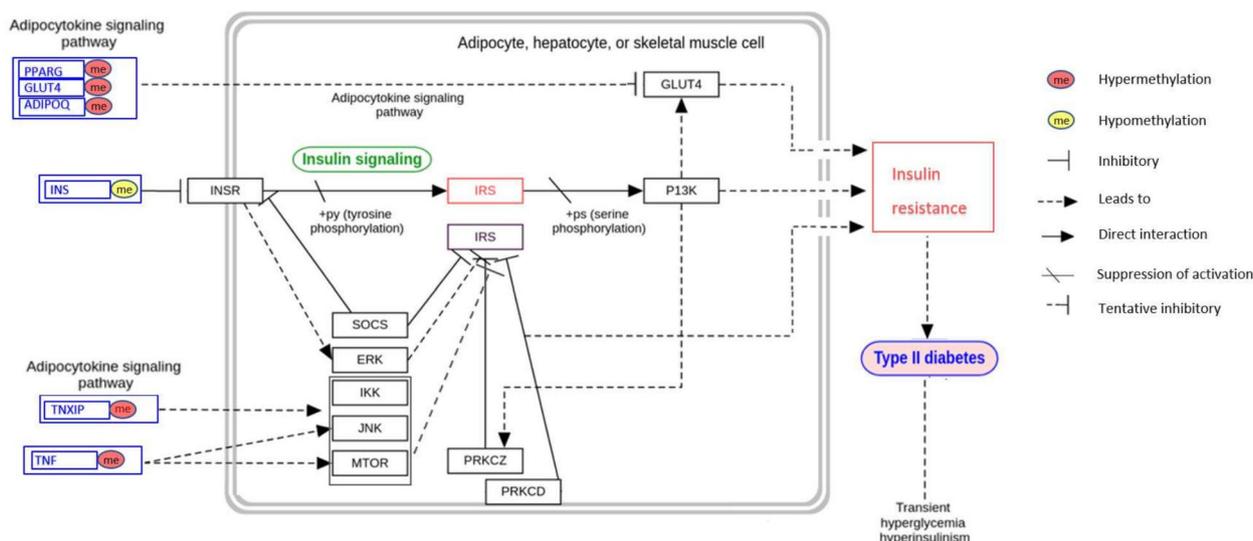


Fig. 10 A schematic diagram showing how abnormal methylation of *PPARG*, *INS*, *GLUT4*, *TNF*, *ADIPOQ*, and *TXNIP* leads to abnormalities in lipid metabolism and insulin signaling pathways

particularly those associated with genes regulating glucose homeostasis and insulin sensitivity, aligns with previous research on the epigenetic regulation of metabolic disorders [15, 53].

Our analysis revealed significant enrichment in pathways related to insulin signaling, AMPK activation, and adipocytokine signaling, which are crucial in metabolic regulation. These findings are consistent with recent comparative methylome analyses [54] that suggest conserved epigenetic mechanisms in metabolic processes across different populations. The enrichment of these pathways aligns with current understanding of how DNA methylation functions as a critical interface between environmental factors and gene expression in metabolic diseases [55], potentially explaining the transgenerational transmission of metabolic risk observed in GDM offspring.

The predictive model developed from our epigenetic data demonstrated strong discriminatory power across datasets, suggesting that these methylation sites could serve as biomarkers for assessing the risk of metabolic diseases in GDM offspring [56]. Despite the model's resilience, no statistically significant differences were detected in the small validation cohort. This underscores the importance of interpreting these results with caution and recognizing that they do not constitute definitive proof of differential methylation [1]. Recent findings on methyl-CpG binding domain proteins [27, 57] and their role in recognizing and binding to genomic methylation sites provide mechanistic insights into how the differential methylation we observed may influence gene expression and cellular function [58, 59]. The persistent predictive significance of CpG sites associated with *PPARG* and *INS* genes across our datasets aligns with the growing body of evidence suggesting these genes' fundamental role in GDM development.

Interestingly, our study revealed a complex relationship between DNA methylation and gene expression in GDM, reminiscent of recent findings in epigenetic modifications as therapeutic targets in cardiovascular diseases [60, 61]. This complexity underscores the need for integrated multi-omic approaches to fully elucidate the functional consequences of GDM-associated epigenetic alterations, as demonstrated in recent epigenome-wide DNA methylation profiling studies [30]. The concordance between our public data analysis and validation cohort in terms of methylation patterns and functional pathway enrichment not only validates our methodological approach but also provides a solid foundation for future targeted investigations. Our findings on the role of specific genes in GDM pathophysiology are supported by recent studies on

renal dysfunction [62], sodium deficiency regulation [63], and the effects of cotransporters on blood pressure [64]. The epigenetic regulation of key systems, such as the renin–angiotensin–aldosterone system, has been implicated in vascular inflammation and remodeling [65, 66], further supporting the relevance of our results in understanding GDM's broader impact on maternal and fetal health.

Recent studies on vitamin D deficiency [67], smoking behavior [44], and global DNA methylation methods [31, 35] provide additional context for understanding environmental influences on the epigenetic landscape in GDM. These environmental factors may interact with genetic predispositions, as suggested by research on the ANRIL locus and its role in regulating inflammation [24], potentially modifying the risk and severity of GDM through epigenetic mechanisms. However, the absence of statistically significant differences in our smaller validation cohort highlights an important limitation of this study: while directional trends were consistent, they should not be interpreted as confirming epigenetic alterations without further replication [68]. The association between DNA methylation and blood pressure [69], as well as advances in gene-specific targeting of DNA methylation [70], offer promising avenues for translating our findings into potential therapeutic strategies. Our study's focus on specific genes aligns with recent investigations into the hypermethylation of *AVPR1A* and *PKCB* genes in pre-eclamptic placental vasculature [36] and methylation of angiotensinogen and aldosterone synthase genes in cardiovascular diseases [71]. The role of TET family proteins [72] and 5-hydroxymethylcytosine [73–75] in epigenetic regulation provides additional layers of complexity to our understanding of GDM-associated epigenetic changes.

In conclusion, our study provides compelling evidence for the role of epigenetic modifications in GDM pathophysiology and offers potential epigenetic biomarkers for risk assessment. The integration of large-scale public data with targeted validation cohorts proves to be a powerful approach in epigenetic research, capable of yielding robust and clinically relevant insights. However, the findings from our smaller validation cohort indicate that further studies with larger cohorts are necessary to validate these findings and explore their long-term clinical implications. Future studies should focus on longitudinal analyses to elucidate the temporal dynamics of these epigenetic changes and their long-term implications for offspring health, as suggested by recent findings on methyl-CpG binding domain proteins and embryonic development [13–15, 43, 76, 77].

Conclusions

This study comprehensively explored the epigenetic landscape specific to the offspring of GDM by integrating large-scale public datasets with a targeted validation cohort. In the public datasets, we identified consistent patterns of differential DNA methylation at key CpG sites associated with genes involved in glucose homeostasis and insulin sensitivity, with enrichment in insulin signaling, AMPK activation, and adipocytokine signaling pathways. These findings suggest potential intergenerational epigenetic influences of GDM on metabolic regulation.

Notably, CpG sites associated with *PPARG* and *INS* genes demonstrated consistent methylation directionality across both public and validation datasets, underscoring their potential biological relevance. The predictive model built on these markers exhibited strong performance in public datasets (AUC = 0.89) and moderate generalizability in the small validation cohort (AUC = 0.82). However, it is important to note that the validation cohort did not yield statistically significant differences, and the observed $\Delta\beta$ values were minimal. These results should therefore be interpreted with caution and cannot be taken as confirmatory evidence of differential methylation.

Despite these limitations, the alignment of directional trends with known epigenetic mechanisms offers valuable preliminary insights. This integrative approach demonstrates the feasibility of combining multi-source data to identify candidate biomarkers and risk patterns related to GDM. Further studies with larger and more diverse cohorts are essential to validate these findings, clarify their clinical implications, and ultimately contribute to the development of novel diagnostic and preventive strategies for metabolic disease in the offspring of GDM mothers.

Abbreviations

| | |
|---------|------------------------------------------------------------------|
| GDM | Gestational diabetes mellitus |
| GEO | Gene expression omnibus |
| DMPs | Differentially methylated positions |
| DMRs | Differentially methylated regions |
| GO | Gene ontology |
| KEGG | Kyoto encyclopedia of genes and genomes |
| LASSO | Least absolute shrinkage and selection operator |
| AUC-ROC | Area under the receiver operating characteristic curve |
| RF | Random forest |
| IADPSG | International Association of Diabetes and Pregnancy Study Groups |
| NGT | Normal glucose tolerance |
| qPCR | Quantitative real-time PCR |
| ROC | Receiver operating characteristic |
| AUC | Area under curve |
| GSEA | Gene set enrichment analysis |

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13098-025-01707-7>.

Supplementary material 1.

Supplementary material 2.

Acknowledgements

We would like to express our sincere gratitude to the doctors and nurses of the Department of Gynecology and Obstetrics at Jiaying Maternal and Child Health Hospital. We deeply appreciate their valuable support and assistance in collecting clinical cases and communicating with patients during the course of this study. Their professionalism and selfless dedication have made significant contributions to the successful completion of this research article.

Author contributions

N.W. conducted the investigation, curated the data, and prepared the original draft. L.Y., as the corresponding author, was responsible for data visualization, writing review and editing, and project management. S.L. was responsible for formal analysis. All authors have read and agreed to the published version of the manuscript.

Funding

Please add: This research was funded by Zhejiang Provincial Medical and Health Technology Program (2022KY1263, 2024KY451).

Availability of data and materials

This study verifies that the self-generated data of the queue can be queried in GEO database, and the GEO registration number is GSE284448.

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Jiaying Maternity and Child Health Care Hospital (Affiliated Women's and Children's Hospital of Jiaying University) (2021 (Medical Ethics)-76 and July 20, 2021)* for studies involving humans.

Consent for publication

Informed consent was obtained from all subjects involved in the study. All patient information will be kept confidential in accordance with ethical procedures, and sample identifiers will be presented in the article using experimental numbers only, with no private information of the cases disclosed.

Competing interests

The authors declare no competing interests.

Received: 6 January 2025 Accepted: 18 April 2025

Published online: 02 May 2025

References

- Ehrlich M. DNA methylation and reader or writer proteins: differentiation and disease. In: Binda O, editor. Chromatin readers in health and disease. Cambridge: Academic Press; 2024.
- Li N, Liu HY, Liu SM. Deciphering DNA methylation in gestational diabetes mellitus: epigenetic regulation and potential clinical applications. *Int J Mol Sci.* 2024;25(17):9361.

3. Tsao CW, Aday AW, Almarzoq ZI, Alonso A, Beaton AZ, Bittencourt MS, et al. Heart disease and stroke statistics-2022 update: a report from the American Heart Association. *Circulation*. 2022;145(8):e153–639.
4. Diaz-Morales N, Baranda-Alonso EM, Martínez-Salgado C, López-Hernández FJ. Renal sympathetic activity: a key modulator of pressure natriuresis in hypertension. *Biochem Pharmacol*. 2023;208:115386.
5. Parati G, Bilo G, Kollias A, Pengo M, Ochoa JE, Castiglioni P, et al. Blood pressure variability: methodological aspects, clinical relevance and practical indications for management - a European Society of Hypertension position paper *. *J Hypertens*. 2023;41(4):527–44.
6. Bekead FT, Goumans MJ, Bogaard HJ, de Man FS, Lluçà-Valldeperas A. Molecular mechanisms and targets of right ventricular fibrosis in pulmonary hypertension. *Pharmacol Ther*. 2023;244:108389.
7. Ma J, Li Y, Yang X, Liu K, Zhang X, Zuo X, et al. Signaling pathways in vascular function and hypertension: molecular mechanisms and therapeutic interventions. *Signal Transduct Target Ther*. 2023;8(1):168.
8. Alfonso Perez G, Delgado Martinez V. Epigenetic signatures in hypertension. *J Pers Med*. 2023;13(5):787.
9. Fujita T. Recent advances in hypertension: epigenetic mechanism involved in development of salt-sensitive hypertension. *Hypertension*. 2023;80(4):711–8.
10. Zgutka K, Tkacz M, Tomasiak P, Piotrowska K, Ustianowski P, Pawlik A, et al. Gestational diabetes mellitus-induced inflammation in the placenta via IL-1 β and toll-like receptor pathways. *Int J Mol Sci*. 2024;25(21):11409.
11. Gao Y, Wang H, Fu G, Feng Y, Wu W, Yang H, et al. DNA methylation analysis reveals the effect of arsenic on gestational diabetes mellitus. *Genomics*. 2023;115(5):110674.
12. Valencia-Ortega J, Saucedo R, Sánchez-Rodríguez MA, Cruz-Durán JG, Martínez EGR. Epigenetic alterations related to gestational diabetes mellitus. *Int J Mol Sci*. 2021;22(17):9462.
13. Hjort L, Novakovic B, Grunnet LG, Maple-Brown L, Damm P, Desoye G, et al. Diabetes in pregnancy and epigenetic mechanisms-how the first 9 months from conception might affect the child's epigenome and later risk of disease. *Lancet Diabetes Endocrinol*. 2019;7(10):796–806.
14. Cardenas A, Gagné-Ouellet V, Allard C, Brisson D, Perron P, Bouchard L, et al. Placental DNA methylation adaptation to maternal glycemic response in pregnancy. *Diabetes*. 2018;67(8):1673–83.
15. Howe CG, Cox B, Fore R, Jungius J, Kvist T, Lent S, et al. Maternal gestational diabetes mellitus and newborn DNA methylation: findings from the pregnancy and childhood epigenetics consortium. *Diabetes Care*. 2020;43(1):98–105.
16. Yousefi PD, Suderman M, Langdon R, Whitehurst O, Davey Smith G, Relton CL. DNA methylation-based predictors of health: applications and statistical considerations. *Nat Rev Genet*. 2022;23(6):369–83.
17. Hughes AL, Szczurek AT, Kelley JR, Lastuvkova A, Turberfield AH, Dimitrova E, et al. A CpG island-encoded mechanism protects genes from premature transcription termination. *Nat Commun*. 2023;14(1):726.
18. Liu Y, Wang Z, Zhao L. Identification of diagnostic cytosine-phosphate-guanine biomarkers in patients with gestational diabetes mellitus via epigenome-wide association study and machine learning. *Gynecol Endocrinol*. 2021;37(9):857–62.
19. Xu P, Dong S, Wu L, Bai Y, Bi X, Li Y, et al. Maternal and placental DNA methylation changes associated with the pathogenesis of gestational diabetes mellitus. *Nutrients*. 2022;15(1):70.
20. Benberin V, Karabaeva R, Kulmyrzaeva N, Bigarinova R, Vochshenkova T. Evolution of the search for a common mechanism of congenital risk of coronary heart disease and type 2 diabetes mellitus in the chromosomal locus 9p21.3. *Medicine*. 2023;102(41):e35074.
21. Ismail N, Abdullah N, Abdul Murad NA, Jamal R, Sulaiman SA. Long non-coding RNAs (lncRNAs) in cardiovascular disease complication of type 2 diabetes. *Diagnostics (Basel)*. 2021;11(1):145.
22. Ma W, Hu J. The linear ANRIL transcript P14AS regulates the NF- κ B signaling to promote colon cancer progression. *Mol Med*. 2023;29(1):162.
23. Cheng J, Cai MY, Chen YN, Li ZC, Tang SS, Yang XL, et al. Variants in ANRIL gene correlated with its expression contribute to myocardial infarction risk. *Oncotarget*. 2017;8(8):12607–19.
24. Aarabi G, Zeller T, Heydecke G, Munz M, Schäfer A, Seedorf U. Roles of the Chr.9p21.3 ANRIL locus in regulating inflammation and implications for anti-inflammatory drug target identification. *Front Cardiovasc Med*. 2018;5:47.
25. Farsetti A, Illi B, Gaetano C. How epigenetics impacts on human diseases. *Eur J Intern Med*. 2023;114:15–22.
26. Pratamawati TM, Alwi I, Asmarinah. Summary of known genetic and epigenetic modification contributed to hypertension. *Int J Hypertens*. 2023;2023:5872362.
27. Nejati-Koshki K, Roberts CT, Babaei G, Rastegar M. The epigenetic reader methyl-CpG-binding protein 2 (MeCP2) is an emerging oncogene in cancer biology. *Cancers (Basel)*. 2023;15(10):2683.
28. Si J, Chen L, Yu C, Guo Y, Sun D, Pang Y, et al. Healthy lifestyle, DNA methylation age acceleration, and incident risk of coronary heart disease. *Clin Epigenetics*. 2023;15(1):52.
29. Liu Y, Geng H, Duan B, Yang X, Ma A, Ding X. Identification of diagnostic CpG signatures in patients with gestational diabetes mellitus via epigenome-wide association study integrated with machine learning. *Biomed Res Int*. 2021;2021:1984690.
30. Lim JH, Kang YJ, Bak HJ, Kim MS, Lee HJ, Kwak DW, et al. Epigenome-wide DNA methylation profiling of preeclamptic placenta according to severe features. *Clin Epigenetics*. 2020;12(1):128.
31. Li S, Tollefsbol TO. DNA methylation methods: global DNA methylation and methylomic analyses. *Methods*. 2021;187:28–43.
32. Foox J, Nordlund J, Lalancette C, Gong T, Lacey M, Lent S, et al. The SEQC2 epigenomics quality control (EpiQC) study. *Genome Biol*. 2021;22(1):332.
33. Ross JP, van Dijk S, Phang M, Skilton MR, Molloy PL, Oytam Y. Batch-effect detection, correction and characterisation in Illumina HumanMethylation450 and MethylationEPIC BeadChip array data. *Clin Epigenetics*. 2022;14(1):58.
34. Gim JA. Survival rate and chronic diseases of TCGA cancer and KoGES normal samples by clustering for DNA methylation. *Life (Basel)*. 2024;14(6):768.
35. Yadav S, Longkumer I, Joshi S, Saraswathy KN. Methylenetetrahydrofolate reductase gene polymorphism, global DNA methylation and blood pressure: a population based study from North India. *BMC Med Genomics*. 2021;14(1):59.
36. Gao Q, Li H, Ding H, Fan X, Xu T, Tang J, et al. Hyper-methylation of AVPR1A and PKCB gene associated with insensitivity to arginine vasopressin in human pre-eclamptic placental vasculature. *EBioMedicine*. 2019;44:574–81.
37. Aguilar-Lacasaña S, Fontes Marques I, de Castro M, Davdand P, Escrivà X, Fossati S, et al. Green space exposure and blood DNA methylation at birth and in childhood - a multi-cohort study. *Environ Int*. 2024;188:108684.
38. Ling C, Vavakova M, Ahmad Mir B, Säll J, Perflyev A, Martin M, et al. Multiomics profiling of DNA methylation, microRNA, and mRNA in skeletal muscle from monozygotic twin pairs discordant for type 2 diabetes identifies dysregulated genes controlling metabolism. *BMC Med*. 2024;22(1):572.
39. Zheng W, Zhang S, Guo H, Chen X, Huang Z, Jiang S, et al. Multi-omics analysis of tumor angiogenesis characteristics and potential epigenetic regulation mechanisms in renal clear cell carcinoma. *Cell Commun Signal*. 2021;19(1):39.
40. Kitagawa K, Maki S, Furuya T, Shiratani Y, Nagashima Y, Maruyama J, et al. Development of a machine learning model and a web application for predicting neurological outcome at hospital discharge in spinal cord injury patients. *Spine J*. 2025. <https://doi.org/10.1016/j.spinee.2025.01.005>.
41. Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA, Damm P, et al. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care*. 2010;33(3):676–82.
42. Wu YL, Jiang T, Huang W, Wu XY, Zhang PJ, Tian YP. Genome-wide methylation profiling of early colorectal cancer using an Illumina Infinium Methylation EPIC BeadChip. *World J Gastrointest Oncol*. 2022;14(4):935–46.
43. Wu Y, Lin X, Lim IY, Chen L, Teh AL, Maclsaac JL, et al. Analysis of two birth tissues provides new insights into the epigenetic landscape of neonates born preterm. *Clin Epigenetics*. 2019;11(1):26.
44. Fragou D, Pakkidi E, Aschner M, Samanidou V, Kovatsi L. Smoking and DNA methylation: correlation of methylation with smoking behavior and association with diseases and fetus development following prenatal exposure. *Food Chem Toxicol*. 2019;129:312–27.

45. Behrooz AB, Cordani M, Fiore A, Donadelli M, Gordon JW, Kliensky DJ, et al. The obesity-autophagy-cancer axis: mechanistic insights and therapeutic perspectives. *Semin Cancer Biol.* 2024;99:24–44.
46. Savage DB, Petersen KF, Shulman GI. Disordered lipid metabolism and the pathogenesis of insulin resistance. *Physiol Rev.* 2007;87(2):507–20.
47. Britsemmer JH, Krause C, Taege N, Geißler C, Lopez-Alcantara N, Schmidtke L, et al. Fatty acid induced hypermethylation in the *Slc2a4* gene in visceral adipose tissue is associated to insulin-resistance and obesity. *Int J Mol Sci.* 2023;24(7):6417.
48. James DE, Stöckli J, Birnbaum MJ. The aetiology and molecular landscape of insulin resistance. *Nat Rev Mol Cell Biol.* 2021;22(11):751–71.
49. Ling C, Rönn T. Epigenetics in human obesity and type 2 diabetes. *Cell Metab.* 2019;29(5):1028–44.
50. Gual P, Le Marchand-Brustel Y, Tanti JF. Positive and negative regulation of insulin signaling through IRS-1 phosphorylation. *Biochimie.* 2005;87(1):99–109.
51. Hamilton DL, Philp A, MacKenzie MG, Patton A, Towler MC, Gallagher IJ, et al. Molecular brakes regulating mTORC1 activation in skeletal muscle following synergist ablation. *Am J Physiol Endocrinol Metab.* 2014;307(4):E365–73.
52. Ruscica M, Ricci C, Macchi C, Magni P, Cristofani R, Liu J, et al. Suppressor of cytokine signaling-3 (SOCS-3) induces proprotein convertase subtilisin kexin type 9 (PCSK9) expression in hepatic HepG2 cell line. *J Biol Chem.* 2016;291(7):3508–19.
53. Dluski DF, Wolińska E, Skrzypczak M. Epigenetic changes in gestational diabetes mellitus. *Int J Mol Sci.* 2021;22(14):7649.
54. Al Adhami H, Bardet AF, Dumas M, Cleroux E, Guibert S, Fauque P, et al. A comparative methylome analysis reveals conservation and divergence of DNA methylation patterns and functions in vertebrates. *BMC Biol.* 2022;20(1):70.
55. De Rosa S, Arcidiacono B, Chiefari E, Brunetti A, Indolfi C, Foti DP. Type 2 diabetes mellitus and cardiovascular disease: genetic and epigenetic links. *Front Endocrinol (Lausanne).* 2018;9:2.
56. Sprang M, Paret C, Faber J. CpG-islands as markers for liquid biopsies of cancer patients. *Cells.* 2020;9(8):1820.
57. Sokolov AV, Schiöth HB. Decoding depression: a comprehensive multi-cohort exploration of blood DNA methylation using machine learning and deep learning approaches. *Transl Psychiatry.* 2024;14(1):287.
58. Wei LZ, Zheng H, Xiao M. A systematic review of the application of machine learning in CpG Island (CGI) detection and methylation prediction. *Curr Bioinform.* 2024;19(3):235–49.
59. Miyahara H, Hirose O, Satou K, Yamada Y. Factors to preserve CpG-rich sequences in methylated CpG islands. *BMC Genomics.* 2015;16(1):144.
60. Sum H, Brewer AC. Epigenetic modifications as therapeutic targets in atherosclerosis: a focus on DNA methylation and non-coding RNAs. *Front Cardiovasc Med.* 2023;10:1183181.
61. Baccarelli AA, Ordovás J. Epigenetics of early cardiometabolic disease: mechanisms and precision medicine. *Circ Res.* 2023;132(12):1648–62.
62. Ueda K, Nishimoto M, Hirohama D, Ayuzawa N, Kawarazaki W, Watanabe A, et al. Renal dysfunction induced by kidney-specific gene deletion of *Hsd11b2* as a primary cause of salt-dependent hypertension. *Hypertension.* 2017;70(1):111–8.
63. Nishimoto K, Harris RB, Rainey WE, Seki T. Sodium deficiency regulates rat adrenal zona glomerulosa gene expression. *Endocrinology.* 2014;155(4):1363–72.
64. Fan M, Zhang J, Lee CL, Zhang J, Feng L. Structure and thiazide inhibition mechanism of the human Na-Cl cotransporter. *Nature.* 2023;614(7949):788–93.
65. Rivière G, Lienhard D, Andrieu T, Vieau D, Frey BM, Frey FJ. Epigenetic regulation of somatic angiotensin-converting enzyme by DNA methylation and histone acetylation. *Epigenetics.* 2011;6(4):478–89.
66. Pacurari M, Kafoury R, Tchounwou PB, Ndebele K. The Renin-Angiotensin-aldosterone system in vascular inflammation and remodeling. *Int J Inflam.* 2014;2014:689360.
67. Nizami HL, Katare P, Prabhakar P, Kumar Y, Arava SK, Chakraborty P, et al. Vitamin D deficiency in rats causes cardiac dysfunction by inducing myocardial insulin resistance. *Mol Nutr Food Res.* 2019;63(17):e1900109.
68. Stenzig J, Schneeberger Y, Löser A, Peters BS, Schaefer A, Zhao RR, et al. Pharmacological inhibition of DNA methylation attenuates pressure overload-induced cardiac hypertrophy in rats. *J Mol Cell Cardiol.* 2018;120:53–63.
69. Hong X, Miao K, Cao W, Lv J, Yu C, Huang T, et al. Association between DNA methylation and blood pressure: a 5-year longitudinal twin study. *Hypertension.* 2023;80(1):169–81.
70. Urbano A, Smith J, Weeks RJ, Chatterjee A. Gene-specific targeting of DNA methylation in the mammalian genome. *Cancers (Basel).* 2019;11(10):1515.
71. Takeda Y, Demura M, Yoneda T, Takeda Y. DNA methylation of the angiotensinogen gene, *AGT*, and the aldosterone synthase gene, *CYP11B2* in cardiovascular diseases. *Int J Mol Sci.* 2021;22(9):4587.
72. Zhang X, Zhang Y, Wang C, Wang X. TET (Ten-eleven translocation) family proteins: structure, biological functions and applications. *Signal Transduct Target Ther.* 2023;8(1):297.
73. Li Q, Huang CC, Huang S, Tian Y, Huang J, Bitaraf A, et al. 5-hydroxymethylcytosine sequencing in plasma cell-free DNA identifies unique epigenomic features in prostate cancer patients resistant to androgen deprivation therapies. *medRxiv.* 2024. <https://doi.org/10.1038/s43856-025-00783-0>.
74. Kazmi N, Elliott HR, Burrows K, Tillin T, Hughes AD, Chaturvedi N, et al. Associations between high blood pressure and DNA methylation. *PLoS ONE.* 2020;15(1):e0227728.
75. Hernaiz A, Sentre S, Betancor M, López-Pérez Ó, Salinas-Pena M, Zaragoza P, et al. 5-Methylcytosine and 5-Hydroxymethylcytosine in scrapie-infected sheep and mouse brain tissues. *Int J Mol Sci.* 2023;24(2):1621.
76. Fu TY, Ji SS, Tian YL, Lin YG, Chen YM, Zhong QE, et al. Methyl-CpG binding domain (MBD)2/3 specifically recognizes and binds to the genomic mCpG site with a β -sheet in the MBD to affect embryonic development in *Bombyx mori*. *Insect Sci.* 2023;30(6):1607–21.
77. Blazevic S, Horvaticek M, Kesic M, Zill P, Hranilovic D, Ivanisevic M, et al. Epigenetic adaptation of the placental serotonin transporter gene (*SLC6A4*) to gestational diabetes mellitus. *PLoS ONE.* 2017;12(6):e0179934.

Publisher's Note

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