SCIENTIFIC REPORTS

Received: 16 November 2017 Accepted: 1 October 2018 Published online: 02 November 2018

OPEN Comprehensive Analysis of Gene **Expression Profiles and DNA** Methylome reveals Oas1, Ppie, Polr2g as Pathogenic Target Genes of Gestational Diabetes Mellitus

Yan Zhang¹, Tiancheng Zhang² & Yunyan Chen¹

Gestational Diabetes Mellitus (GDM) has a high incidence of pregnancy, which seriously affects the life quality of pregnant women and fetal health. DNA methylation is one of the most important epigenetic modification that can regulate the gene expression level, and thus affect the occurrence of various diseases. Increasing evidence has shown that gene expression changes caused by DNA methylation play an important role in metabolic diseases. Here we explored the mechanisms and biological processes that affect the occurrence and development of GDM through analyzing the gene expression profiles and DNA methylation data of GDM. We detected 24,577 differential CpG sites mapping to 9339 genes (DMGs, differential methylation gene) and 931 differential expressed genes (DEGs) between normal samples and GDM samples. GO (gene ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis of 326 overlapping genes between DMGs and DEGs showed obvious enrichment in terms related to metabolic disorders and immune responses. We identified Oas1, Ppie, Polr2g as possible pathogenic target genes of GDM by combining protein-protein interaction analysis. Our study provides possible targets for early diagnosis of GDM and information for clinical prevention of abnormal fetal development and type 2 diabetes.

Gestational Diabetes Mellitus (GDM) is caused by the failure of pancreatic β cells to meet the increased insulin requirements and metabolic demands of pregnancy¹. In addition to abnormal glucose tolerance and insulin resistance, GDM patients often show complications including chronic systemic inflammation², increased leptin concentration³, and reduced adiponectin content⁴. High risks of Type2 diabetes among women with GDM history and abnormal fetus development caused by GDM emphasized the importance of deciphering the molecular mechanisms governing the progression of GDM. GWAS⁵ and metabonomic studies these years shed light on the genomic propensity and metabolic stimulus causing GDM, but the transcriptional biomarkers of peripheral organ and the epigenetic regulatory mechanism of GDM were less studied.

As an important epigenetic modification, DNA methylation represses transcription by inhibiting the binding of specific transcription factors as well as by recruiting methyl-CpG-binding repressive chromatin remodelers⁶. DNA methylation involves in various biological processes including cell proliferation, cell fate transition, X chromosome inactivation, genomic imprinting and chromosome stability⁶. Benefiting from the advances of methylation sequencing technology and big data storage and analytics approaches, more and more clues on the mammalian development and diseases progression have been dug out. Information provided by systematic methylome analysis combining other omic studies is a huge booster for the development of the Precision Medicine⁷⁻¹⁰.

The methyl donor for DNA methylation is the intermediate product of one-carbon metabolism, and increasing evidence have shown that DNA methylation can response to nutritional and environmental influences and

¹Department of Obstetrics & Gynecology, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, P. R. China. ²Key Lab of Reproduction Regulation of NPFPC-Shanghai Institute of Planned Parenthood Research (SIPPR), Fudan University Reproduction and Development Institution, Shanghai, China. Correspondence and requests for materials should be addressed to Y.C. (email: yunycitu@126.com)





regulate gene expression pattern accordingly¹¹. For instance, DNA methylome changes were recognized as a faithful readout in tis-sues such as skeletal muscle and adipose in obese people with metabolic abnormalities^{12,13}. Furthermore, genome-wide methylome and transcriptome analysis in hepatocyte isolated from normal people and patients with type 2 diabetes revealed that genes involved in hepatic glycolysis and insulin re-sistance were abnormally hypomethylated and thus overexpressed in liver¹⁴. Therefore, it is reasonable to speculate that the placenta, as a peripheral tissue susceptible to metabolic disorders, its DNA methylation regulation plays an important role in the pathology.

In this study, we confirmed genes related to immune response and metabolic pathways were differentially expressed between healthy gestational women and women with GDM. Differences in expression profile were regulated at least in a great part by DNA methylation level according to our comprehensive DNA methylome studies. We also integrated protein-protein interaction network analysis in our study and find Oas1, Ppie, Polr2 are important GDM target genes. Our study will shed light on early diagnosis of GDM and contribute to the prevention of abnormal neonatal development caused by GDM.

Results

Differential methylation sites. Raw methylation data was preprocessed by filtering out high noise methylation sites and samples containing high noise regions. As a result, 473,281 out of 485,577 (97.47%) sites and all of the samples passed the quality control process, indicating high data quality. We firstly assessed differences in overall DNA methylation patterns between normal and DMG patients placental tissues as shown in Fig. 1A. Differential methylation analysis detected 24,577 differential CpG sites corresponding to 9339 genes (DMG, differential methylation gene) across samples. We then ordered all genes detected according to their expression level and divided them into four groups, namely 0–25%, 25–50%, 50–75%, 75–100%. As shown in Fig. 1B, methylation level negatively correlates to gene expression level which is in constant with their roles in repressing



Figure 2. (A) Box plot depicts the overall expression level across samples after normalization. (blue bar: normal samples, red bar: GDM samples). (B) Heatmap shows the differential expressed genes' expression profiles across samples under the standard of p < 0.05. Red in the scale bar represents high expression level, and green means relatively low expression level. c. Heatmap shows the 182 differential CpG methylation sites filtered by a more stringent criteria of p < 0.05 and |delta beta| > 0.2. Red in the scale bar represents sites with relatively higher methylation level and and green is the opposite. d. Cluster analysis of GO terms.

gene transcription. Besides, DNA methylation level decreased dramatically upstream the TSS, increased sharply toward the gene body regions, and remained at a low plateau until the 3'end of the gene body and showed the lowest point at TSS (Fig. 1B).

The function of DNA methylation relay on the local CpG content and position relative to genes¹⁵. To discover the regulatory effect of this DNA modification in our system, we mapped DMSs between two groups of samples to neighborhood location, including CpG islands(CGIs), their shores (± 2 kb of CGIs) and shelves (± 2 –4kb of shores), and to functional genomic regions, including distal promoters (± 1.5 kb), proximal promoters (± 0.2 kb), gene bodies, untranslated regions containing 3' UTR (define regions) and 5' UTR (define regions). The statistical results are detailed in the pie chart of Fig. 1C. Briefly, DMSs located most in CpG island regions in contrast to their upstream and downstream regions including the shores and shelves. When mapping into gene regions, gene body regions contained the most DMSs relative to their expression regulatory regions. We conducted region-level analysis to reveal the DMG distribution pattern across different genetic regions. Methylation levels in each gene element were compared among samples according to their beta value (see methods). Statistical result was shown in Fig. 1D. Most DMG screened out by methylation level changes in CpG island region enriched in promoters. Up-regulated DNA methylation on promoters of tumor suppressor genes is common in many kinds of malignant tumor. Taken together, our data imply that DNA methylation is highly differentially regulated and may function in regulating pathogenic gene expression in DMG patient placenta.

Differential expressed genes. Raw data were normalized and overall gene expression level was comparable across samples (Fig. 2A), indicating the high quality of the original data. We performed differential expression analysis with all genes detected and identified 931 differential expressed genes (DEGs) between normal samples and GDM patient samples. Figure 2B illustrates the heatmap of DEGs' expression profiles in normal and GDM patient samples with green and red color indicates low and high expression values respectively. What's more, we screened out 182 DMSs satisfied |delta beta| > 0.2 from the total DMSs, and Fig. 2C is the heatmap of 182 DMSs' beta value in normal and GDM patient samples.

There are 326 overlapping genes between DMG and DEG, and their expression are probably regulated by DNA methylation. To decipher the functional roles of these genes, we next conducted GO term enrichment analysis.

Pathway names	Hits	P-value	Genes
Systemic lupus erythematosus	10	0.00194	HIST1H2AC, CD86, HIST1H2BD, HLA-DRB1, H2AFV, HIST1H2BG, HIST1H4F, HIST3H2A, C1R, HIST1H4D
Herpes simplex infection	11	0.00498	SP100, HLA-DRB1, TAP1, OAS1, HLA-C, NFKB1, CDC34, CCL5, HLA-E, STAT1, TBPL1
Transcriptional misregulation in cancer	10	0.00862	CD86, LMO2, FOXO1, BCL6, NFKB1, PBX3, MEIS1, ITGAM, WT1, CDK14
Ubiquitin mediated proteolysis	8	0.0248	RNF7, UBE2M, SIAH1, CDC34, UBE2QL1, UBE2Q2, BIRC3, UBE2B
Cell adhesion molecules (CAMs)	8	0.0294	CD86, HLA-DRB1, ITGB8, CLDN10, HLA-C, HLA-E, SELE, ITGAM
Graft-versus-host disease	4	0.0312	CD86, HLA-DRB1, HLA-C, HLA-E
Allograft rejection	4	0.0418	CD86, HLA-DRB1, HLA-C, HLA-E
Huntington's disease	9	0.0485	SDHA, POLR2G, NDUFS6, NDUFB10, COX7A2, NDUFA10, TBPL1, SOD2, AP2M1

Table 1. Significantly enriched KEGG pathways of overlaps between DEGs and DMGs.

.....

GO Terms showed obvious enrichment on metabolism indicating strong links to diabetes as a metabolic disorder (Fig. 2D). In line with previous research¹⁶, our data also enriched terms associated with immune reactions indicating connections between insulin resistance and immune pathways. Growing evidence in the literature suggests that insulin resistance is the result of an inflammatory milieu¹⁷. IL-1 and TNF- α induce synergistic pleiotropic responses that profoundly affect the production of extracellular matrix (ECM) proteins. Increased expression of fibronectin, laminin β -1, and metalloproteinases may induce a fibrotic response and disrupt the structural integrity of placental endothelial cells as in other cell types¹⁸. We next queried the KEGG (Kyoto Encyclopedia of Genes and Genomes) databases with DMGs that were differentially expressed between normal tissues and patient tissues, the core KEGG pathways were also closely related to immune systems including Systemic lupus erythematous and Herpes simplex infection (Table 1).

Network Analysis. To understand the interactive relationship of the overlapping genes between DMG and DEG, we performed network analyses using the STRING database. In general, genes with high STRING scores are often in a protein complex or parts of pathways. Besides, proteins involved in the same disease have an increased tendency to interact with each other¹⁹. It was reported that important target genes that related to the onset and development of a certain disease are always at the center of protein interaction network. To detect more efficient target genes, we performed modular analysis. As a result, we obtained a PPI network which contains 222 gene nodes and 539 edges (Fig. 3A), as well as 2 gene modules satisfied score >2 (Fig. 3B,C). *Oas1* has the highest node degree score in the PPI network and exists in our module (Fig. 3B). It was reported to be a downstream factor of IFN and involve in immune responses²⁰, conforming with our pathway enrichment analysis. The other two factors, *Ppie* and *Polr2g* also get high scores and appear in the module (Fig. 3C). Pipe was accepted as a factor involves in metabolism²¹. Additionally, expression changes of *PIPE*, *POLR2G* and *OAS1* were further confirmed through real time PCR in placenta tissue of three GDM samples and three normal samples as shown in Fig. 4. Taken together, our research finds 3 GDM target genes and contributes new insights to the research and development of targeted drugs.

Discussion

GDM patients are with higher risk to develop type 2 diabetes than pregnant women with normal blood glucose levels in the future²². Type 2 diabetes will be a major health burden without medical intervention properly. Besides, Pregnant women with gestational diabetes mellitus are prone to miscarriage during early pregnancy and the Impaired fetal development may lead to glucose intolerance and obesity of the baby²³. Therefore, the early diagnosis of GDM and personalized medical intervention are of significant importance.

Studies on diabetes focus mostly on the pancreatic islets. The cell fate decision, maturation, the development of beta cell heterogeneity as well as aging and apoptosis were comprehensively studied. More and more omic data such as RNA-seq, DNA methylome, transcription factor and histone modification ChIP-seq are deciphering these cellular events. However, early stage prenatal diagnosis does not rely on the molecular released from islets of the pregnant women. So in-depth study of the molecular signatures on placenta is an urgent need when diagnosing GDM. Placenta is an important link between maternal and fetal nutrient exchange, and the mRNA of placental origin is readily detectable in maternal plasma²⁴. Furthermore, the placenta are peripheral tissues that can be considered as a readout of the insulin resistance. Therefore, deciphering the transcriptome changes and corresponding epigenetic regulatory mechanism of planceta tissue in GDM patients is an ideal way to understand the pathology of the GDM and will shed light on GDM treatment and T2D prevention.

Our work deeply interprets the DNA methylome and gene expression data with normal placenta tissue and GDM patient placenta tissue, and find a list of genes that may be regulated by the epigenetic way. The development of dcas9 (deficient CRISPR-associated 9 nuclease) technique provides the possibility for the site-specific edition of epigenetic modifications^{25–27}. Our research has dug out three genes, namely *Oas1, Ppie, Polr2g*. Further studies will explore the regulatory strategies for the expression level of these genes and try to reveal the roles of these proteins in their interaction network. We can edit important target genes by dcas9 or other epigenome editing method to alleviate the harmful effect brought to fetal development by GDM. Moreover, women with GDM





Figure 3. PPI network analysis of the overlaps between DEGs and DMGs. (A) overall network. (B,C) is the subnetwork module with module score >2.



Figure 4. Relative mRNA level of *OAS1*, *PPIE* and *POLR2G* in placenta tissues of GDM patients and normal controls quantified by real time PCR.

can take specific small-molecule inhibitors targeting pathogenic genes during pregnancy to alleviate symptoms and reduce the chances for their children to get metabolic disorders in the future.

In conclusions, our study identified several biological processes and genes that might contribute the development of GDM, this should be helpful for its diagnosis and targeted therapy.

Materials and Methods

Microarray datasets. The microarray datasets in this study were downloaded from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). The gene expression (GSE70493) and DNA methylation profiles (GSE70453) were deposited by Binder *et al.* and composed with the same samples: 32 GDM patient placenta tissue and 31 normal samples²⁸. The genome-wide expression profiles were quantified by the commercial gene microarray GeneChip[®] Human Transcriptome Array 2.0 (Affymetrix). Methylation profiling data was obtained through the Illumina Infinium HumanMethylation450 Beadchips.

Microarray Analysis. Methylation data was preprocessed for the reliability of the following analyzes, which mainly involve the filtering of methylation sites with detection P-value > 0.05 in more than 20% samples and samples contains more than 75% methylation sites with detection P-value > 0.05. Methylation level of CpG site is represented by beta value, which calculated as (methylated signal intensity)/(100 + methylated signal intensity + unmethylated signal intensity). Methylation level for specific region, such as the promoter of a gene, is the average beta value of all CpG sites contained in this region. Differential methylation analysis at site-level and region-level analysis were performed using IMA package with the criteria of P-value < 0.05 for the identification of differential methylation sites (DMS) and regions (DMRs).

We then conducted differential expression analysis between normal person and patients with Gestational diabetes Mellitus. Briefly, the raw data were imported into R and normalized via the oligo package. Then the limma package was used for the screening of differential expression genes (DEGs) with the criteria of P-value < 0.05.

Functional clustering analysis. We conducted the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of DEGs and differential methylation genes (DMGs, i.e. genes contain at least one DMS) based on the Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.abcc.ncifcrf.gov/). The thresholds used in this study is P-value < 0.05.

Protein-protein interaction network analysis. We performed protein-protein interaction (PPI) network analysis for the overlaps between DEGs and DMGs based on the STRING database (STRING, https://string-db.org/) with the criteria of comprehensive score >0.4. Protein-protein interaction was visualized with Cytoscape (http://cytoscape.org/). For the interpretation of PPI network, we conducted modular analysis with MCODE plugin, which could identify protein clusters that might represent protein families and parts of pathways through seeking dense regions of network, of Cytoscape and identified network modules with score >2.

Real time PCR. Real time PCR was performed for validating expressions of OAS1, PPIE and POLR2G. Placenta tissues from three GDM patients and three normal samples were collected and subjected to RNA extraction. The concentration and purity of the RNA were determined by spectrophotometer (Spectrophotometry GE). The total RNA obtained was stored in a -80 °C refrigerator, and then reverse-transcribed into cDNA using the PrimeScript RT kit according to the instructions. The obtained cDNA could be used as a template to detect the amount of mRNA of the corresponding gene. Primers of those three genes were as follows: *POLR2G* forward: GACCTGCACAGGGAAGTATGGC, reverse: CTGGATAAAGGACAAAGCCTCGG; OAS1 forward: GGATTCTGCTGGCTGAAAGCAAC, reverse: GGAGTGTGCTGGGTCTATGAGAG; *PPIE* forward: GGGCAAGTCCATCTATGGGAAG, reverse: GGATAGTAGACCTGGTCCCGTA. This study was approved by the Ethics Committee of the Shanghai family planning hospital and written informed consent was obtained from all participants.

Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

References

- 1. Nolan, C. J. Lipotoxicity, beta cell dysfunction, and gestational diabetes. Cell metabolism 19, 553-554 (2014).
- Qiu, C., Sorensen, T. K., Luthy, D. A. & Williams, M. A. A prospective study of maternal serum C-reactive protein (CRP) concentrations and risk of gestational diabetes mellitus. *Paediatric and perinatal epidemiology* 18, 377–384 (2004a).
- Qiu, C., Williams, M. A., Vadachkoria, S., Frederick, I. O. & Luthy, D. A. Increased maternal plasma leptin in early pregnancy and risk of gestational diabetes mellitus. Obstetrics and gynecology 103, 519–525 (2004b).
- Lappas, M., Yee, K., Permezel, M. & Rice, G. E. Release and regulation of leptin, resistin and adiponectin from human placenta, fetal membranes, and maternal adipose tissue and skeletal muscle from normal and gestational diabetes mellitus-complicated pregnancies. *The Journal of endocrinology* 186, 457–465 (2005).
- 5. Lowe, W. L. Jr., Scholtens, D. M., Sandler, V. & Hayes, M. G. Genetics of Gestational Diabetes Mellitus and Maternal Metabolism. *Current diabetes reports* 16, 15 (2016).
- 6. Bender, J. DNA methylation and epigenetics. Annual review of plant biology 55, 41-68 (2004).
- Angulo, J. C., Lopez, J. I. & Ropero, S. DNA Methylation and Urological Cancer, a Step Towards Personalized Medicine: Current and Future Prospects. *Molecular diagnosis & therapy* 20, 531–549 (2016).
- Robertson g, J. H., Cheng, W. S., Hood, L. & Tian, Q. An epigenetic biomarker panel for glioblastoma multiforme personalized medicine through DNA methylation analysis of human embryonic stem cell-like signature. *Omics: a journal of integrative biology* 18, 310–323 (2014).
- 9. Robertson, K. D. DNA methylation and human disease. Nature reviews Genetics 6, 597-610 (2005).

- 10. Tang, J., Xiong, Y., Zhou, H. H. & Chen, X. P. DNA methylation and personalized medicine. *Journal of clinical pharmacy and therapeutics* **39**, 621–627 (2014).
- 11. Anderson, O. S., Sant, K. E. & Dolinoy, D. C. Nutrition and epigenetics: an interplay of dietary methyl donors, one-carbon metabolism and DNA methylation. *The Journal of nutritional biochemistry* 23, 853–859 (2012).
- 12. Barres, R. *et al.* Weight loss after gastric bypass surgery in human obesity remodels promoter methylation. *Cell reports* **3**, 1020–1027 (2013).
- 13. Benton, M. C. et al. An analysis of DNA methylation in human adipose tissue reveals differential modification of obesity genes before and after gastric bypass and weight loss. Genome biology 16, 8 (2015).
- Kirchner, H. et al. Altered DNA methylation of glycolytic and lipogenic genes in liver from obese and type 2 diabetic patients. Molecular metabolism 5, 171-183 (2016).
- 15. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature reviews Genetics* **13**, 484–492 (2012).
- Wojcik, M., Zieleniak, A., Zurawska-Klis, M., Cypryk, K. & Wozniak, L. A. Increased expression of immune-related genes in leukocytes of patients with diagnosed gestational diabetes mellitus (GDM). *Experimental biology and medicine* 241, 457–465 (2016).
 Samuel, V. T. *et al.* Mechanism of hepatic insulin resistance in non-alcoholic fatty liver disease. *The Journal of biological chemistry*
- 279, 32345–32353 (2004).
 18. Ruan, H. *et al.* Profiling gene transcription *in vivo* reveals adipose tissue as an immediate target of tumor necrosis factor-alpha:
- implications for insulin resistance. *Diabetes* 51, 3176–3188 (2002).
 19. Barabasi, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature reviews Genetics* 12, 56–68 (2011).
- Karaulov, A. V., Shulzhenko, A. E. & Karsonova, A. V. Expression of IFN-Inducible Genes with Antiviral Function OAS1 and MX1 in Health and under Conditions of Recurrent Herpes Simplex Infection. *Bulletin of experimental biology and medicine* 163, 370–373 (2017).
- Kakar, S. S., Winters, S. J., Zacharias, W., Miller, D. M. & Flynn, S. Identification of distinct gene expression profiles associated with treatment of LbetaT2 cells with gonadotropin-releasing hormone agonist using microarray analysis. *Gene* 308, 67–77 (2003).
- 22. Jarvela, I. Y. et al. Gestational diabetes identifies women at risk for permanent type 1 and type 2 diabetes in fertile age: predictive role of autoantibodies. Diabetes care 29, 607–612 (2006).
- Radaelli, T., Varastehpour, A., Catalano, P. & Hauguel-de Mouzon, S. Gestational diabetes induces placental genes for chronic stress and inflammatory pathways. *Diabetes* 52, 2951–2958 (2003).
- Ng, E. K. et al. mRNA of placental origin is readily detectable in maternal plasma. Proceedings of the National Academy of Sciences of the United States of America 100, 4748–4753 (2003).
- 25. Lei, Y. *et al.* Targeted DNA methylation *in vivo* using an engineered dCas9-MQ1 fusion protein. *Nature communications* **8**, 16026 (2017).
- 26. Stepper, P. et al. Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. Nucleic acids research 45, 1703–1713 (2017).
- 27. Xiong, T. *et al.* Targeted DNA methylation in human cells using engineered dCas9-methyltransferases. *Scientific reports* 7, 6732 (2017).
- 28. Binder, A. M., LaRocca, J., Lesseur, C., Marsit, C. J. & Michels, K. B. Epigenome-wide and transcriptome-wide analyses reveal gestational diabetes is associated with alterations in the human leukocyte antigen complex. *Clinical epigenetics* 7, 79 (2015).

Author Contributions

Yan Zhang contributed to the data collection and analysis, wrote the manuscript. Tiancheng Zhang contributed to the concept, data collection and analysis, Yunyan Chen contribute to study concept and coordination, and manuscript writing.

Additional Information

Competing Interests: The authors declare no competing interests.

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018