# Profiling Susceptible Genes Associated with Gestational Diabetes Mellitus using Next Generation Sequencing: A Pilot Study

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>gen-2019-0136</td>
</tr>
<tr>
<td>Manuscript Type</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>02-Aug-2019</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Razif, Muhammad Fazril; University of Malaya Faculty of Medicine, Molecular Medicine Jamalpour, Sajad; Universiti Malaya, Pharmacology Sinniah, Ajantha; Universiti Malaya, Pharmacology Omar, Siti Zawiah; Universiti Malaya, Department of Obstetrics and Gynaecology Elias, Nurezwana; Universiti Malaya, Obstetrics and Gynaecology Yelumalai, Suseela; University of Malaya Faculty of Medicine, Obstetrics &amp; Gynaecology</td>
</tr>
<tr>
<td>Keyword:</td>
<td>Next Generation Sequencing, Gestational Diabetes Mellitus, Gene expression</td>
</tr>
<tr>
<td>Is the invited manuscript for consideration in a Special Issue?:</td>
<td>Not applicable (regular submission)</td>
</tr>
</tbody>
</table>

https://mc06.manuscriptcentral.com/genome-pubs
Profiling Susceptible Genes Associated with Gestational Diabetes Mellitus using Next Generation Sequencing: A Pilot Study

Muhammad Fazril Mohamad Razif¹#, Sajad Jamal Pour²#, Ajantha Sinniah², Siti Zawiah Omar³, Nurezwana Elias³, and Suseela Yelumalai³

¹Department of Molecular Medicine, University of Malaya, Kuala Lumpur, Malaysia
²Department of Pharmacology, University of Malaya, Kuala Lumpur, Malaysia
³Department of Obstetrics & Gynaecology, University of Malaya, Kuala Lumpur, Malaysia

#Equal contribution.

Correspondence to Suseela Yelumalai, Department of Obstetrics & Gynaecology, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.

E-mail: ysuseela@um.edu.my

Tel: +603-7967 7762
Fax: +603-7967 4193

This study was supported by the High Impact Research fund, University of Malaya (UM.C/625/1/HIR/ MOHE/ MED/28). Account No: H-20001-00-E000066.
Abstract

Comprehensive analysis of genetic markers associated with diabetes mellitus during and post-pregnancy have yet to be elucidated. Hence, this pilot study was designed to identify gene variants that are associated with gestational and postpartum diabetes using next-generation sequencing (NGS). NGS was employed to generate gene expression profiles of women with gestational diabetes mellitus (n=3), postpartum diabetes (n=3) and controls (n=3). Canonical pathways were generated via Ingenuity Pathway Analysis (IPA). Pathway enrichment analysis of all genes was performed using EnrichR (based on the KEGG and Wiki-Paths databases) and Gene Set Enrichment Analysis (GSEA). Our results revealed 117 and 66 genes were significantly expressed in patients with GDM and postpartum diabetes, respectively. 19 genes involved in the regulation of the immune system (FDR q-value: 1.32e⁻⁸) were identified to be associated with higher risk for GDM. Biological pathway analysis revealed that both angiogenesis and apoptosis were significantly enriched in patients with GDM. Our study has found a significant relationship involving IRS2 activating the PI3K-signalling pathway leading to insulin resistance. It is also the first study to observe that c-FOS expression is strongly correlated with gestational diabetes. Our findings provide an insight into understanding the influence of genetic factors in biochemical trajectories during pregnancy.

Keywords

Next Generation Sequencing; Gestational Diabetes Mellitus; Post-natal; Pregnancy; Gene Expression.
**Introduction**

Gestational diabetes mellitus (GDM) is a common, asymptomatic metabolic disorder of pregnancy caused by impaired glucose tolerance from the late second trimester of pregnancy onwards. The prevalence of GDM is estimated to be ~2–10% of pregnancies in USA, and ~3.5% in UK. Risk factors for GDM include ethnicity (South Asian, Afro-Caribbean and Middle-Eastern), obesity, previous history of GDM and first-degree family history of diabetes (Association, 2004). Glucose intolerance in GDM is the result of peripheral insulin resistance, failure of β-cell expansion, and insufficient maternal insulin production to cope with ensuing hyperglycaemia. Women with clinical presentations consistent with high risk for GDM can also be assessed for cardiovascular risk factors, requiring appropriate management to reduce the probability of coronary heart disease, cardiomyopathy and stroke.

Although GDM is considered to be one of the most common pregnancy complications, there is no consensus on the preferred early screening tool. Therefore, conflicting arguments regarding the timing of screening, diagnostic thresholds, optimal pharmacological management, and post-partum follow-ups remains accrued (Johns *et al*., 2018). Hence, according to the World Health Organization (WHO), the diagnostic criteria should be identified to enable appropriate management upon confirmation of GDM. The diagnostic criteria associated with signs and symptoms of diabetes are: fasting plasma glucose (FPG) ≥7.0 mmol/L or 126 mg/dL, and/or 2-hour 75-g oral glucose tolerance test (OGTT) ≥11.1 mmol/L or 200 mg/dL, or random plasma glucose (RPG) ≥11.1 mmol/L or 200 mg/dL. The primary intervention for GDM patients remains to be lifestyle management such as nutritional counselling and level of physical activity (Guo *et al*., 2019). Therefore repeated reinforcements of choosing the right quantity and quality of food and appropriate exercise could manage their blood glucose level within the normal range. However, if lifestyle modifications fail to optimise glycaemic control, other pharmacological treatment options need to be considered (Patti *et al*.,...
Draft

The 2017 American Diabetes Association Standards of Medical Care (ADA) stated that insulin has been the mainstay therapy for patients with GDM to improve glycaemic control when lifestyle modifications are inadequate (Corcoy et al., 2018) (Corcoy et al., 2018) (Corcoy et al., 2018). Oral anti-hyperglycaemic drugs (OADs) are also considered as alternative treatment options for GDM, however it remains controversial for use as first-line treatment in comparison to insulin by many professional bodies due of its ability to cross the placental barrier (Patti et al., 2018). Metformin has good efficacy and short-term safety data; however the long-term safety profile in pregnant women is yet to be elucidated. Glyburide is efficacious for the treatment of GDM, but may be associated with increased rates of large-for-gestational-age (LGA) infants and neonatal hypoglycaemia compared to insulin (Martis et al., 2018). Since the exact multifactorial pathogenesis of GDM is not well understood, many studies have been devoted to understanding the underlying mechanism of this condition.

Recent advancements have identified that single-nucleotide polymorphisms (SNPs), DNA methylation and microRNAs can be explored as biomarkers to accurately predict the onset and monitor the status of GDM in women (Dias et al., 2018) (Dias et al., 2018) (Dias et al., 2018). Interestingly, several studies evaluated the association of microRNAs with the regulation of β-cell mass and function in the pathogenesis of chronic metabolic diseases (Guay and Regazzi, 2013; Sebastiani et al., 2018; Guarino et al., 2018). Several studies have also showed evidence for a genetic predisposition to GDM, whilst gene–environment interactions lead to population-specific variations in GDM occurrence (Dias et al., 2018; Wu et al., 2016). The progression of GDM is undeniably the manifestation of aberrant expression of multiple genes, hence exploiting gene expression profiles via gene microarray and high-throughput sequencing technology could lead to identifying valuable diagnostic and treatment biomarkers (Zhao and Li, 2018).
High-throughput RNA-seq provides portraits of the transcriptomic landscape at an unprecedented resolution. RNA-seq typically produces millions of sequencing reads, each of which provides pieces of information regarding genomic events in the cell. Thus, unlike microarray, RNA-seq has diverse applications for genomic analysis including quantification of gene expression, discovery of new transcripts, detection of single nucleotide polymorphisms and RNA editing. Of interest in identifying genetic markers of GDM is quantifying gene expression, which is performed by simply counting the reads aligned to each gene or exon region. As RNA-seq detects both coding and multiple forms of non-coding RNA, it can accurately measure gene abundance and identify known and novel features of the transcriptome.

Thus, this pilot study was designed to identify gene variants that are shared and/or unique between women with gestational and postpartum diabetes using next-generation sequencing (NGS) in order to examine their potential as a predictive tool for GDM. We identified several genes that are closely linked to GDM and mapped out their related pathways. This study can provide better understanding of the genes and genomic aberrations leading to the development of GDM.

**Materials and methods**

**Participants**

The protocol was approved by the University Malaya Medical Centre (UMMC), Medical Ethics Committee (MEC) and written informed consent was obtained from all participants. A total of 9 pregnant women were recruited for this study. They were referred to the Prenatal Diagnostic Centre, UMMC, Kuala Lumpur, Malaysia. The study population was categorized
as normal pregnant women (SSC; n=3), patients with gestational diabetes mellitus (SSGP; n=3), and patients with post-partum diabetes mellitus (SSPN; n=3). Inclusion criteria: (i) aged between 18-45 years, (ii) singleton pregnancy, and (iii) planned to carry the pregnancy to term and delivery at UMMC. Exclusion criteria: Subject has (i) other maternal-related disorders, (ii) a history of drug abuse, smoking and depression, (ii) carrier of an infectious disease, and (iv) multi-fetal pregnancy. Controls were age-matched (between 18 to 45 years old) with normal pregnancy, and delivered their babies between 38-41 weeks. Gestational age was calculated from the first day of the women’ last menstrual cycle, confirmed or modified by ultrasound.

General screening was done for pregnant women between 24-28 weeks of gestation who had (i) BMI >27 kg/m², (ii) previous macrosomic baby weighing ≥4 kg, (iii) previous GDM, (iv) first-degree relative with diabetes, (iv) history of unexpected ante-partum fetal death, (v) history of congenital abnormalities, (vi) glycosuria on the first or any prenatal visit, and (vii) existing obstetric complications (critical hypertension, pregnancy-induced hypertension, polyhydramnios, and recent consumption of steroids). GDM was diagnosed by a trained physician based on fasting glucose ≥ 5.1 or impaired glucose tolerance with a 120 minute blood sugar value of ≥7.8 mmol/l using a modified Oral Glucose Tolerance Test (mOGTT).

Women at high risk of developing GDM but presented normal initial screening results were subjected to another mOGTT at 4–6 weeks later.

**Sample collection**

Maternal peripheral blood samples were collected upon admission to UMMC (~1 day to 1 week before delivery). Blood samples were placed in EDTA-treated tubes. Plasma was obtained by centrifugation at 3000 rpm for 10 mins at 4°C. Protease inhibitors (Sigma; 1:100) were added to samples prior to storage at -80°C and minimized freeze/thaw cycles.
Total RNA extraction and cDNA synthesis

Total RNA was extracted using QIAzol Lysis Reagent (Qiagen, USA) and RNeasy Mini Kit (Qiagen, USA) according to manufacturer's instructions. Purity of the RNA samples was assessed using NanoDrop ND-1000 spectrophotometer (Thermo Scientific, USA). RNA concentration was accessed via the Qubit Fluorometer (ThermoFisher Scientific, UK). RNA integrity was assessed via the Agilent 2100 Bioanalyzer (RNA 6000 Pico Chip; Agilent Technologies, Denmark). Subsequently, RNA was digested using Baseline-ZERO DNase (Epicenter Biotechnologies, USA) to remove any remaining genomic DNA from the samples.

RNA-Seq library construction and sequencing

Baseline-ZERO-treated RNA from both patients and controls were used to generate an RNA-Seq library using ScriptSeq Complete Gold Kits (Blood) based on low input protocol (Illumina, USA). Ribosomal RNA was removed from Baseline-ZERO-treated RNA prior to the library construction. Cytoplasmic and mitochondrial rRNA, as well as globin mRNA was depleted using the Ribo-Zero Magnetic Gold Kit (Human/Mouse/Rat; Illumina, USA). Briefly, 4 µL of rRNA Removal Solution was added to 100 ng total-RNA, and the mixture was incubated for 10 mins at 68°C and 15 mins at room temperature. Probes with hybridized rRNA were removed by adding the RNA-probe mixture to 30 µL of washed magnetic beads, followed by magnetic separation and transfer of the supernatant to an RNase-free tube. The rRNA-depleted RNA was purified using Agencourt RNAClean XP Kit (Beckman Coulter, USA) according to manufacturer’s instructions.

The rRNA-depleted RNA was checked by on-chip electrophoresis (Agilent RNA 6000 Pico Chip; Agilent Technologies, Denmark) of a 1 µL sample using the Agilent 2100 Bioanalyzer.
(Agilent Technologies, Denmark) to ensure that all rRNA was depleted. Directional, paired-end and indexed RNA-Seq libraries was synthesized according to manufacturer’s instructions. In brief, to fragment the RNA, 5 µL of Globin-Zero-treated RNA was added with 1 µL of RNA Fragmentation Solution and 2 µL cDNA Synthesis Primer. The fragmented RNA was converted to cDNA by adding 3.0 µL cDNA Synthesis PreMix, 0.5 µL 100 mM DTT and 0.5 µL StarScript Reverse Transcriptase. cDNA was subsequently tagged and purified using Agencourt AMPure XP Kit (Beckman Coulter, USA). Unique index was added to each of the purified tag cDNA to generate the sequenced ready library. The RNA-Seq libraries were checked by using Agilent High Sensitivity DNA chip on Agilent 2100 Bioanalyzer (Agilent Technologies, Denmark). The RNA-Seq library concentration was checked using Qubit Fluorometer (ThermoFisher Scientific, UK). Furthermore, the KAPA Library Amplification Kit (KK2611; KapaBiosystems, USA) was used to quantify the amount of the library fragment that can be amplified. The RNA-Seq libraries were denaturated and diluted to 1.5 pM with pre-chilled hybridization buffer and loaded onto the NextSeq 500 High Output Kit (150 cycles) (Illumina, USA). This was followed by indexed paired-end sequencing (76+6+76 bp) on a NestSeq 500 (Illumina, USA).

**Read Mapping**

Sequencing data for all samples were obtained in FASTQ format text file. The reads were subjected to sequencing adapter trimming and base quality (Q≥20) trimming to ensure that only good quality bases derived from mRNAs were further analysed. Trimmed reads of less than 35 base pairs (bp) were discarded with its pair. High quality sequencing reads were mapped to pre-indexed human genome reference sequence (CRGh37/hg19) with corresponding genome annotations (Zerbino *et al.*, 2018) using TopHat2 (version 2.0.11; Kim *et al.*, 2013). Two samples (1 from SSC; 1 from SSGP) were excluded from further
analysis due to low data quality yield. Hence, a total of seven samples (2 SSC, 2 SSGP and 3 SSPN) were used for downstream analysis. Expression quantification and profiling analysis was performed using featureCounts (Subread v.1.4.6; Liao et al., 2014). Separately, to obtain transcript-level expression, good quality reads were mapped to human reference transcript sequences (Ensembl GRCh37). Sequencing read mapping, expression quantification and profiling analysis were performed using Salmon version 0.9.1 (Patro et al., 2017); available at https://github.com/COMBINElab/Salmon). Gene set enrichment analysis were expressed as SSGP vs, SSC and SSPN vs. SSC.

**Annotation for gene expression interactions**

Based on gene expression profiles, various plots were produced to assess the consistency of biological replicates within each group using DESeq 2 [version 3.8, BiocManager 1.30.4; R (3.5.2 GUI 1.70 El Capitan build (7612)]. Pair-wise differential expression and related statistical testing for significance were also analysed using DESeq2 (version 3.8, BiocManager 1.30.4; R (3.5.2 GUI 1.70 El Capitan build (7612)). CummeRbund R package (version 2.24.0) was used to generate plots to illustrate differential expression results between groups. Genes showing altered expression with a p-value <0.05 were considered significant.

**Identification of gene expression**

A two-sided Mann-Whitney U-test was performed as implemented by the wilcox.test function in cummeRbund R package (version 2.24.0)(Vargha and Delaney, 2000). In order to account for multiple testing, Benjamini-Hochberg correction was applied; p-values <0.05 were considered significant. Expression data for each gene was analyzed using a linear model framework via the limma package (version 3.28.17;(Law et al., 2014). All further analysis and visualization of data were performed using custom R software (version 2.24.0) (Ihaka
and Gentleman, 1996, Hammoud and Kramer, 2018). Significantly expressed genes were
analysed using the DAVID online analysis tool (http://david.abcc.ncifcrf.gov/). Biological
processes were determined via Gene Ontology (Ashburner et al., 2000; The Gene Ontology
Consortium, 2019) and gene expressions were screened.

**Pathway analysis**

Pathway and network analysis were performed using IPA which computes a score for each
network according to the fit of the set of supplied focus genes. These scores indicate the
likelihood of focus genes to belong to a network versus those obtained by chance. A score >2
indicates a ≥ 99% confidence that a focus gene network was not generated by chance alone.
Fischer’s exact test with Fast Discovery Rate (FDR) option was used to calculate the
significance of the canonical pathway. Gene Set Enrichment Analysis (GSEA) version 6.3
(Subramanian et al., 2005; Mootha et al., 2003) normalized RNA-seq expression data; they
were pre-ranked based on the fold change between case and control samples. The Hallmark
curated gene sets in MSigDB database (version 5.0) were used for GSEA analysis. Gene sets
were tested for enrichment in rank ordered lists via the GSEA software (Subramanian et al.,
2005). Permutation number was set to 1000, and p-value <0.05 was considered statistically
significant.

**Statistical analysis**

Statistical analysis were performed using SPSS (IBM Corp. Released 2017. IBM SPSS
Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.) and GraphPad prism
(version 7.0 for Windows, GraphPad Software, USA, www.graphpad.com). Data was
presented as mean ± standard deviation (S.D) for triplicate independent measurements.
Student's t-test was used to assess the differences between experimental groups. Differences
with p-values < 0.05 were considered statistically significant. Due to the similarities between
the groups at baseline, multivariate analysis was not performed. For enrichment of the
targeted regions, we used the Agilent SureSelect Solution hybridization method. For each
gene, exon coordinates were obtained from the RefSeq database to identify coding and
untranslated regions. Sequencing was performed on Illumina HiSeq2000 Sequencing System
(Illumina, USA). Data quality check was performed on Illumina Sequencing Analysis Viewer
(SAV; Illumina, USA). Demultiplexing was performed using the Illumina CASAVA program
(version 1.8.2; Illumina, USA). Gene expression levels were estimated using RSEM v1.2.15
(Li & Dewey, 2011). Trimmed mean of M-values (TMM) was used to normalize gene
expression.

Results

Gene expression analysis

Demographics and clinical data for SSGP, SSPN and SSC patients and controls are presented
in Table 1. A total of 117 significantly expressed genes were identified following analysis
with 63 genes up-regulated and 54 genes down-regulated in SSGP vs. SSC (Fig. 1a). In
SSPN vs. SSC, 66 genes were significantly expressed, with 29 genes up-regulated and 37
genes down-regulated (Fig. 1a). Samples were mapped to the reference genome sequence
with >90% overall mapping rate (Fig. 2). Comparative analysis between SSGP vs. SSC (Fig.
2a) and SSPN vs. SSC (Fig. 2b) revealed 1943 and 1641 significantly expressed transcripts,
respectively.

Gene sets were tested for enrichment in rank ordered lists via GSEA and genes mediating
apoptosis by activation of caspases. Our analysis identified genes encoding cell
differentiation markers (CD180, MSR1 and IL18RAP from SSGP; CD83 and CD69 from SSPN) and transcription factors (FOS, HIF1A, CEBPE, ZNF117 and MLLT4 from SSGP; PER1 and KLF11 from SSPN). In addition, genes associated with protein kinases (MAP2K2 and SIK1) and those that play significant roles in the oestrogen-signalling pathway were highlighted (GNAI3, FOS, MA2K2 and DUSP1). Overlapping genes in SSGP vs. SSC and SSPN vs. SSC (Fig. 1b,c) highlights links between transcriptomic variants involved in the pathways presented.

Oncogenes, such as MLLT4, PER1 and ACSL6, are translocated cancer genes while TNFAIP3 is a tumour-suppressor gene. Nineteen genes were identified to be involved in regulation of the immune system (FDR q-value: 1.32 e-8): IGHM, IGHG3, PELI1, PPP2R3C, SAMSN1, LGALS3, PDE4B, CD180, FOS, KRT1, SKP1, CLC, UBASH3A, TREML4, CHRN2, IRS2, HIF1A, SNCA and ORM1. Seven out of these 19 genes (IGHM, IGHG3, PELI1, LGALS3, FOS, IRS2 and HIF1A) play significant roles in the regulation of B-cell activation (FDR q-value: 3.6e-6).

Volcano plots were used to illustrate the negative log of the p-value against the log of the fold change between two conditions, SSC vs. SSGP (Fig. 3a.i) and SSC vs. SSPN (Fig. 3a.ii). Significantly expressed genes and transcripts are highlighted in green and red in the volcano plots, respectively (Fig. 3a). In order to assess the quality of replicates, sample distances and clustering matrices were used to show similarities of the expression profiles between samples; SSC vs. SSGP (Fig. 3c.i) and SSC vs. SSPN (Fig. 3c.ii). Discretely distributed expression values were transformed into a regularised log transformation model (rlog). Then, the distances between samples were calculated and clustered. Distance clustering can be used to assess sample heteroprofiles or profile similarities between intergroup and intragroup
replicates. A dendrogram of consistent biological replicates clusters intragroup samples together while exhibiting clear separation between intergroup samples. Principle Components Analysis (PCA) plots were used to visualize the overall influence of experimental covariates and batch effects (Fig. 3d). The closer the sample coordinates are on the plot, the higher the similarities between profiles are. Sample groups with consistent replicates, SSC vs. SSGP (Fig. 3d.i) and SSC vs. SSPN (Fig. 3d.ii), are expected to have their replicates clustering close to each other.

**Prediction of gene interactions**

Ingenuity Pathway Analysis (IPA; Qiagen, USA) was used to explore the context of deregulated genes in both SSGP and SSPN. Upon enrichment and pathway analysis of gene expression levels, the tool for upstream regulator prediction established that both *LGALS3* and *IRS2* were inhibited in the presence of *HIF1A* in gestational diabetes mellitus in apoptosis-related networks (Fig. 4a). Interestingly, activation of *FOS, PELP1* and *DUSP1* were also predicted to decrease activation in that combination. Analysis of genes involved in post-natal diabetes mellitus showed down-regulation of *CD69, TNFAIP3* and *RABGEF1* and up-regulation of *TMSB10/TMSB4X* in chronic inflammation and lymphatic tumour-related networks (Fig. 4b).

**Discussion**

GDM is common in pregnancy and poses a serious threat to both mothers and infants. GDM is increasing in global prevalence and is associated with increased risk of obesity and diabetes in offspring, affecting health of current and future generations. Such high prevalence highlights the urgent need to identify biological markers that predict early onset of GDM. With regards to the overall genetic susceptibility for GDM, it should be recognized that gene
variants such as IRS2, FOS, DUSP1 and many others, as shown in this study, were reported
to be associated with reduced first phase insulin secretion, while second phase insulin
secretion remained intact. Previous studies have reported only few gene variants in GDM as
compared to our present study (Łabno et al., 2016, Zhu et al., 2015, Gjesing et al., 2017)
Thus far no study has investigated the association of these variants during pregnancy and
post-partum among women with GDM. Current results also reveal a high concordance in
both SSGP and SSPN on how these genes regulate innate immunity, MAPK, apoptosis,
insulin and cancer-related pathways.

The phosphatidylinositol 3'-kinase (PI3K)-AKT signalling pathway is activated by many
types of cellular stimuli or toxic insults, and regulates fundamental cellular functions such as
transcription, translation, proliferation, growth, and survival. The present study evidently
shows the significant relationship of IRS2 in activating the PI3K-signalling pathway leading
to insulin resistance in SSGP (Fig. 5a). IRS2 may also mediate various cellular processes via
insulin. IRS2 promotes Wnt/β-catenin signalling which is critical for cell growth (Geng et al.,
2014). It has been demonstrated that dysregulation of this signalling pathway leads to cancer,
obesity and diabetes.

Our study is the first to observe that c-FOS expression is strongly correlated with SSGP (Fig.
5a). As seen in the activation of the PI3K-AKT signalling pathway by IRS2, a similar
mechanism directly activates extracellular signal-regulated-kinases (ERK1 and 2) to inhibit c-
FOS, resulting in insulin resistance. These findings may suggest a novel relationship between
IRS2 activation in association with c-FOS inhibition leading to the development of GDM.
Interestingly, these results further indicate direct auto-regulation between IRS2 and c-FOS.
The FOS proteins have been implicated as regulators of cell proliferation, differentiation, and
transformation. In some cases, expression of c-FOS has been associated with apoptotic cell death. The exact mechanism by which c-FOS contributes to apoptosis is not clearly understood, but observations in human hepatocellular carcinoma cells indicate that c-FOS is a mediator of c-myc-induced cell death and may induce apoptosis via the p38 MAP kinase pathway (Kalra and Kumar, 2004).

Interestingly, our study revealed that DUSP1 plays a novel role in SSGP and SSPN. DUSP1 and HIF1A appear to have an indirect interaction, which results in the regulation of the former in both SSGP and SSPN, leading to glucose intolerance (Fig. 5b). This finding suggests that DUSP1 may play a key role in the development of GDM. DUSP1 was originally identified as a growth factor and stress-inducible gene. DUSP1 encodes for a dual-specificity protein phosphatase that dephosphorylates both threonine and tyrosine residues on members of all 3 major MAPK subfamilies (MAPK/JNK, MAPK/p38, and MAPK/ERK).

DUSP1 is also involved in the regulation of cell cycle and apoptosis. Thus, finding strategies to balance the interaction between DUSP1-HIF1A is key in regulating the apoptosis pathway amongst GDM patients. In summary, our study concludes that genes involved in both angiogenesis and apoptosis, such as IRS2, c-FOS and DUSP1, lead to an increased risk of GDM. Although these genes may be used as a potential screening tool for the early detection of GDM, further in-depth investigations focused on their biological roles and regulations are necessary prior to future therapeutic interventions.

Acknowledgements

We would like to thank all the women, medical and nursing staff for their participation in this study. This work was supported by the High Impact Research fund, University of Malaya (UM.C/625/1/HIR/ MOHE/ MED/28). Account No: H-20001-00-E000066.

Conflict of interest
The authors have no conflict of interest to disclose.

References


Kalra, N. & Kumar, V. 2004. c-Fos is a mediator of the c-myc-induced apoptotic signaling in serum-deprived hepatoma cells via the p38 mitogen-activated protein kinase pathway. J. Biol. Chem. 279: 25313-25319. doi: 10.1074/jbc.M400932200.


421 and bias-aware quantification of transcript expression. Nat Methods. 14: 417. doi:
422 10.1038/nmeth.4197.
423
427
428 Vargha, A. & Delaney, H. D. 2000. A critique and improvement of the CL common language
431
432 Zhao, X. & Li, W. 2018. Gene coexpression network analysis identified potential biomarkers
435
437 microRNA expression in early pregnancy to predict gestational diabetes mellitus. Int.
### Table

Table 1: Demographic and clinical data for SSGP, SSPN and SSC.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>SSGP (n=3)</th>
<th>SSPN (n=3)</th>
<th>SSC (n=3)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal age (years)</td>
<td>23.67 ± 3.51</td>
<td>35.33 ± 4.73</td>
<td>28.00 ± 2.65</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>62.97 ± 2.28</td>
<td>61.23 ± 3.86</td>
<td>53.90 ± 1.61</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>159.67 ± 4.04</td>
<td>168.00 ± 3.61</td>
<td>159.34 ± 4.58</td>
<td>&gt; 0.05</td>
</tr>
<tr>
<td>Median maternal BMI (kg/M²)</td>
<td>24.82 ± 1.95</td>
<td>21.63 ± 1.86</td>
<td>21.56 ± 1.84</td>
<td>&gt; 0.05</td>
</tr>
<tr>
<td>Fasting plasma glucose/mmol/L</td>
<td>6.33 ± 0.59</td>
<td>5.70 ± 0.72</td>
<td>4.53 ± 0.35</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>2 hours plasma glucose/mmol/L</td>
<td>8.37 ± 0.84</td>
<td>6.83 ± 0.31</td>
<td>5.03 ± 0.42</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

p-value ≤ 0.05 indicates significant differences between groups as per obtained using one-way analysis of variance (ANOVA; GraphPad Prism 5.0). Data are shown as mean ± standard deviation (SD).
Figure 1: Gene expression analysis for SSGP and SSPN vs. SSC. (a) Core analysis of expressed genes (p \leq 0.05; R (3.5.2 GUI 1.70 El Capitan build (7612); (Ihaka and Gentleman, 1996, Hammoud and Kramer, 2018). (b) Top analysis Log Ratio of up-regulated and down-regulated genes in both groups. Red arrow indicates up-regulation; green arrow signifies down-regulation. (c) Venn diagram analysis of differentially expressed genes. Numbers of genes expressed differentially are shown in the diagram. The sum of numbers in each circle is the total number of differentially expressed genes in each comparison group, and the overlapping section is the number of common differentially expressed genes among the comparison groups; U/D indicates total number of up-regulated and down-regulated genes.
Figure 2

(a) SSGP
(b) **SSPN**

[Image of a detailed diagram representing SSPN]
**Figure 2:** Heat maps for SSGP and SSPN. Patterns of expression in (a) SSGP and (b) SSPN. Each heat map includes dendrograms and clustering gene expression according to genes and samples. A gene family is described as any collection of transcription factors that share a common feature, eg. biochemical activity.
Figure 3: (a) Volcano plots showing significantly expressed genes in (i) SSC vs SSGP and (ii) SSC vs SSPN via DESeq2 (BiocManager 1.30.4). Colour differences show relative abundance of at least 2-fold, or log2 equal to or less than 1 (black) versus greater than 1 (gold); FDR < 0.05 (red); Differentially expressed genes with $|\text{Log FC}| > 1$ and FDR < 0.05 (Green). (b) Volcano plots showing significantly expressed transcripts in (i) SSC vs SSGP and (ii) SSC vs SSPN via DESeq2 (BiocManager 1.30.4). Stably expressed transcripts across both groups are indicated in black. Differentially expressed transcripts (p ≤ 0.05) are indicated in red; dots on the left denote down-regulation; dots on the right denote up-regulation. (c) Heatmap showing the distribution of expressed genes and transcripts [R (3.5.2 GUI 1.70 El Capitan build (7612); heatmap.2: g-plots package]. (d) Principle Components Analysis (PCA) plots showing the overall influence of experimental covariates and batch effects for (i) SSC vs SSGP and (ii) SSC vs SSPN. The closer the sample coordinates are on the plot, the higher the similarities between profiles.
**Figure 4:** Pathway analysis. (a) Upstream regulators of gestational diabetes mellitus via apoptosis-related network using Ingenuity Pathway Analysis (IPA; Qiagen, USA) illustrating the relationship between differentially expressed genes (SSC vs. SSGP). (b) Upstream regulators of post-natal diabetes mellitus via chronic inflammation and lymphatic tumor-related networks using IPA (SSC vs. SSPN). **Blue:** predicted inactivation; orange: predicted activation; **red:** up-regulation; **green:** down-regulation. The color gradient in the network indicates the strength of expression denoted by fold changes. Arrows (\(\rightarrow\)) indicates a direct interaction between the transcript products; (\(\ldots\ldots\ldots\rightarrow\)) indicates an indirect interaction between the transcript products; (\(\circ\)) indicates auto regulation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).
**Figure 5**: Pathway analysis generated using Ingenuity Pathway Analysis (IPA; Qiagen, USA). (a) PI3K/AKT Signalling Pathway. (b) Insulin Resistance Pathway.