Peer Review File

Article information: https://dx.doi.org/10.21037/atm-21-3497

Reviewer A

Comment 1: The total number of proteins identified and the total number used in analysis after filtering are missing. How did the author remove proteins present in only a few samples, or did they impute the values of those proteins. It's not clear what steps authors performed after obtaining protein abundance data to obtain the final data. Since not all proteins are identified in all samples- did the author perform any data filtration or imputation is not cleared. Please expand the methods section.

Reply 1: Thank you for your comments. A total of 1783 proteins were identified in all samples. The data were preprocessed as follows: first, proteins identified in less than 70% of the samples were removed and then the missing values were imputed with the median value of each group. After the above processing, a total of 1328 proteins remained for further analysis.

We added these descriptions in line 218-219 in the result section and line 172-174 in the Method section (manuscript without edition).

Comment 2: Add the final protein relative abundance data or the final datasheet used in the analysis as a spreadsheet in supplementary.

Reply 2 and changes in the text: The raw datasheet and final datasheet have been added to Supplemental Table 1.

Comment 3: AUC using which algorithm? Again the method section is missing. Present a table showing the sensitivity, Specificity, Accuracy, ROC/AUC, and MCC for the important proteins (e.g., NOV, etc.) in the manuscript.

Reply 3: Thank you for your suggestion. The area under the receiver operating characteristic curve was calculated using MedCalc 15.8 (www.medcalc.org). The GDM diagnostic panel combining NOV, PRDX5, and HGFL was constructed using logistic regression algorithm using MedCalc 15.8.

The sensitivity, Specificity, Accuracy, ROC/AUC, and MCC for important proteins involved in GDM and SA diagnosis were added to revised manuscript as Table 3 as followings.

Table 3. The AUC, Sensitivity, Specificity, Accuracy, and MCC of biomarkers for GDM (A) and SA (B) diagnosis.

A

	NOV	HGF	PRDX5	combine
AUC	0.742	0.827	0.724	0.902

Sensitivity	66.7%	80.0%	60.0%	73.3%
Specificity	80.0%	93.3%	80.0%	100.0%
Accuracy	73.3%	86.7%	70.0%	86.7%
MCC	0.471	0.740	0.408	0.761

B

	GDF15	GNAQ	GBB1	GNAI3
AUC	0.812	0.778	0.801	0.751
Sensitivity	63.2%	63.2%	73.7%	68.4%
Specificity	94.7%	89.5%	84.2%	84.2%
Accuracy	79.0%	76.3%	78.9%	76.3%
MCC	0.610	0.546	0.582	0.533

Comment 4: Also, add a line plot showing the relative abundance of NOV, etc., from the 1st sample to the last sample for control and GDM/SA groups. That way, it would be easy to understand how or when those proteins' abundance starts changing.

Reply 4: Thank you for your suggestions. The line plot of the relative abundance of the three GDM biomarkers during the first, second, and third trimesters in the control and GDM groups has been added to the manuscript.

As shown in the figure below, compared to a normal pregnancy NOV was upregulated during the early trimester, and it recovered to normal levels during the second and third trimesters in patients with GDM. In previous reports, NOV was found to be highly expressed in smooth muscle cells of the arterial vessel wall [1], and to inhibit β-cell proliferation through different mechanisms, thus impairing β-cell insulin secretion [2]. Therefore, its upregulation might be related to β cells. However, NOV has also been reported as a proangiogenically secreted molecule involved in placental development and to be increasingly expressed during normal pregnancy in the placenta and serum [3]. Thus, NOV recovered to normal levels during normal pregnancy.

Compared to normal pregnancy, in GDM pregnancy, HGF was upregulated during the first and second trimesters but decreased to normal levels during the third trimester of pregnancy. HGF, a vital component of insulin resistance pathophysiology, plays roles in β-cell homeostasis, inflammatory response mediation, and the glucose metabolic flux in diverse insulin-sensitive cell types [4]. Increased levels of HGF have been linked to the development of insulin resistance, and HGF has been regarded as a serum biomarker of macroangiopathy in diabetes mellitus [5]. The decrease in HGF during the third trimester of patients with GDM might reflect the clinical treatment effect.

PRDX5 was downregulated during the first trimester but recovered to a normal level during the second and third trimesters of pregnancy compared to the normal controls. PRDX5, a member of the family of antioxidant enzyme, participated in eliminating

hydrogen peroxide and neutralizing other reactive oxygen species [6]. PRDX5 significantly prevents high glucose-induced apoptotic cell death, and it has been reported as an early sign of retinal pathology in diabetic patients [7]. During the first trimester of pregnancy, when maternal-fetal blood circulation is newly established, the oxygen concentration increases rapidly in the placenta, leading to an increase in oxidative stress and the activation of the antioxidant enzymes to protect against oxidative stress [8]. The generation of reactive oxygen species causes a decrease in endogenous antioxidants [9]. Therefore, the early downregulation of PRDX5 in patients with GDM might be related to its overconsumption in the process of defending against oxidative stress. During the second and third trimesters, the consumption of PRDX5 of GDM and normal pregnancy was similar and showed the similar expression level.

We added above results in Figure 3F, in the result section (line 303-306), and discussion section (line 402-425) in the revised manuscript.

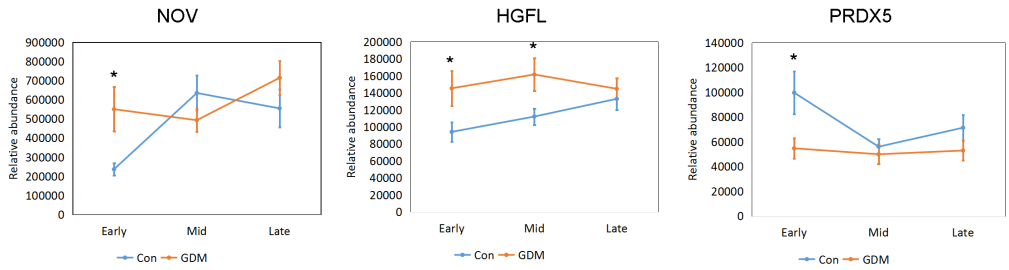

Figure: The line plot of the relative abundance of 3 GDM biomarkers in the early, mid, late stage of pregnancy for control and GDM group. The average relative abundance and standard error of each group was shown. *: $p<0.05$.

Reference

[1]. Lin CG, Leu SJ, Chen N, et al. Ccn3 (nov) is a novel angiogenic regulator of the ccn protein family. J Biol Chem 2003;278(26):24200-24208.

[2]. Martinerie C, Garcia M, Do TT, et al. Nov/ccn3: A new adipocytokine involved in obesity-associated insulin resistance. Diabetes 2016;65(9):2502-2515.

[3] Alexandra G, Markus S, Caroline D, et al. The circulating proangiogenic factors CYR61 (CCN1) and NOV (CCN3) are significantly decreased in placentae and sera of preeclamptic patients. Multicenter Study Reprod Sci. 2007;14(8 Suppl):46-52

[4]. Oliveira AG, Araújo TG, Carvalho BM, et al. The role of hepatocyte growth factor (hgf) in insulin resistance and diabetes. Front Endocrinol (Lausanne) 2018;9:503.

[5]. Konya H, Miuchi M, Satani K, et al. Hepatocyte growth factor, a biomarker of macroangiopathy in diabetes mellitus. World J Diabetes 2014;5(5):678-688.

[6]. Kropotov A, Usmanova N, Serikov V, et al. Mitochondrial targeting of human peroxiredoxin v protein and regulation of prdx5 gene expression by nuclear transcription factors controlling biogenesis of mitochondria. FEBS J 2007;274(22):5804-5814.

[7]. Kubo E, Singh DP, Fatma N, et al. Tat-mediated peroxiredoxin 5 and 6 protein transduction protects against high-glucose-induced cytotoxicity in retinal pericytes. Life Sci 2009;84(23-24):857-864.
[8]Jauniaux E，WatsonA L，Hempstock J，et al．Onset of matemal arterial blood flow and placental oxidative stress．A possible factor in human early pregnancy failure．Am J Pathol，2000，157(6)：2111-2122．
[9]Myan L，Cuix．oxidative stress in the placenta．Histochem Cell Bio. 2004，122(4)：369--382．
[10]Houstis N，Rosen ED，Lander ES．Reactive oxygen species have a causal role in
ltipie forms of insulin resistance. Nature,2006, 4:944-948

Reviewer B

Comment 1: In the manuscript, the authors mentioned that they pooled the samples followed by HPLC separation and MS analysis. I wonder how to distinguish the sample source of detected proteins? In many similar studies, labeling peptide samples, e.g. with iTRAQ, TMT, silac and so on, guarantees that the sample source can be determined so that protein/peptide samples from different sources can be pooled, analyzed and quantified simultaneously.

Comment 2: If the authors use pooled sample to generate the overall protein profiles, followed by non-labeling individual MS. Please also clearly explain the procedures.

Reply 1 and 2: Thank you for your comments. The data-independent acquisition (DIA) approach is a widely used quantitative proteomics technology for proteomic analysis. The DIA method consecutively acquires high-resolution and accurate mass fragment ion spectra throughout chromatographic elution (retention time) by repeatedly cycling through the sequential isolation windows [1]. Thus, the DIA method acquires fragment ions for all precursors and yields high reproducibility, high validity, and deep proteome coverage [1, 2]. The DIA strategy usually uses the following workflow [3,5]. Firstly, a pooled sample is analyzed in data-dependent acquisition (DDA) tandem mass spectrometry (MS) mode to generate a spectral library. Secondly, each individual sample is analyzed in the DIA mode. Thirdly, the DIA-MS data are analyzed using Spectronaut Pulsar software 14.10 (Biognosys, Schlieren, Switzerland) by searching the DDA spectral library. This DIA approach has been widely used in various proteomic studies [3-5].

In this study, we used a similar strategy. First, a pooled sample was obtained by mixing the peptides from each sample, and an extensive DDA spectral library of the pooled samples was generated using two-dimensional liquid chromatography–tandem mass spectrometry in DDA mode. At this stage, the overall protein profiles were from the

pooled sample. Second, the individual samples were sequentially analyzed in DIA mode. Third, the individual sample data were imported into the Spectronaut Pulsar software and searched the DDA spectral library, and each sample was analyzed individually.

The overall procedures were added in the method section in line 136-140.

Reference

[1]L. C. Gillet, P. Navarro, S. Tate, H. Rost, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, Mol Cell Proteomics 2012, 11, O111 016717

[2]R. Bruderer, O. M. Bernhardt, T. Gandhi, Y. Xuan, J. Sondermann, M. Schmidt, D. Gomez-Varela, L. Reiter, Mol Cell Proteomics 2017

[3]Y. Liu, A. Buil, B. C. Collins, L. C. Gillet, L. C. Blum, L. Y. Cheng, O. Vitek, J. Mouritsen, G.Lachance, T. D. Spector, E. T. Dermitzakis, R. Aebersold, Mol Syst Biol 2015, 11, 786.

[4]J. Muntel, Y. Xuan, S. T. Berger, L. Reiter, R. Bachur, A. Kentsis, H. Steen, J Proteome Res 2015, 14, 4752.

[5] Mindi Zhao , Xiang Liu , Haidan Sun , Zhengguang Guo , Xiaoyan Liu 2, Wei Sun. Proteomics Clin Appl. 2019, 13(5),e1800152

Comment 3: Control-1 and Control-2 could be confused with the CON-1, CON-2 and CON-3 in Figure 2 and 4. Maybe change other names.

Reply 3 and changes in the text: The names CON-1, CON-2, and CON-3 have been revised to Control-1, Control-2, and Control-3, respectively, in Figures 2 and 4.

Comment 4: The abbreviations should be consistent within the manuscript. For example, GDM-1 vs. G1, GDM-2 vs. G2 and GDM-3 vs. G3.

Reply 4 and changes in the text: The names G1, G2, and G3 have been revised to GDM-1, GDM-2, and GDM-3, respectively, in the manuscript.

Comment 5: In Figure 3, the data of subfigures C, D and E was derived from which trimester?

Reply 5 and changes in the text: The data of subfigures C, D, and E were derived from the first trimester.

This description has been added to lines 714-716 of the revised manuscript.

Comment 6: The figure should have clear labels, e.g. y-axis of F3d, F3e and x-axis of F3c.

Reply 6 and changes in the text: We have added axis labels in the revised manuscript.

Comment 7: In Figure 4, no auROC analysis was conducted like Figure 3?

Reply 7 and changes in the text: The area under the receiver operating characteristic curve analysis of the four SA biomarkers has been included as Figure 4F.

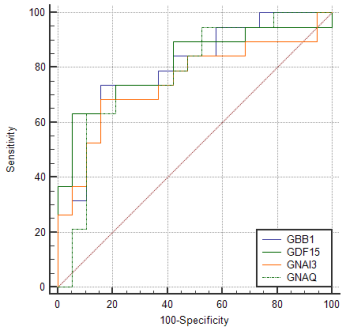

Figure: The AUC curve of the four SA differential proteins