



# Potential of Immune-Related Genes as Biomarkers for Diagnosis and Subtype Classification of Preeclampsia

Ying Wang, Zhen Li, Guiyu Song and Jun Wang\*

Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, Shenyang, China

**Objective:** Preeclampsia is the main cause of maternal mortality due to a lack of diagnostic biomarkers and effective prevention and treatment. The immune system plays an important role in the occurrence and development of preeclampsia. This research aimed to identify significant immune-related genes to predict preeclampsia and possible prevention and control methods.

**Methods:** Differential expression analysis between normotensive and PE pregnancies was performed to identify significantly changed immune-related genes. Generalized linear model (GLM), random forest (RF), and support vector machine (SVM) models were established separately to screen the most suitable biomarkers for the diagnosis of PE among these significantly changed immune-related genes. The consensus clustering method was used to divide the PE cases into several subgroups to explore the function of the significantly changed immune-related genes in PE.

**Results:** Thirteen significantly changed immune-related genes were obtained by the differential expression analysis. RF was the best model and was used to select the four most important explanatory variables (CRH, PI3, CCL18, and CCL2) to diagnose PE. A nomogram model was constructed to predict PE based on these four variables. The decision curve analysis (DCA) and clinical impact curves revealed that PE patients could significantly benefit from this nomogram. Consensus clustering analysis of the 13 differentially expressed immune-related genes (DIRGs) was used to identify 3 subgroups of PE pregnancies with different clinical outcomes and immune cell infiltration.

**Conclusion:** Our study identified four immune-related genes to predict PE and three subgroups of PE with different clinical outcomes and immune cell infiltration. Future studies on the three subgroups may provide direction for individualized treatment of PE patients.

**Keywords:** preeclampsia, immunity, biomarker, nomogram, consensus clustering

## INTRODUCTION

Preeclampsia (PE) is an idiopathic disease occurring during pregnancy, which can affect the function of multiple organs (Grandi et al., 2019; Phipps et al., 2019). PE affects 3–5% of pregnant women worldwide (Zhang et al., 2020b). It is a hypertension disease that is one of the main causes of increased mortality for pregnant women and perinatal infants (Say et al., 2014). In the

## OPEN ACCESS

### Edited by:

Liqing Tian,  
St. Jude Children's Research  
Hospital, United States

### Reviewed by:

Yi Zhang,  
Dana-Farber Cancer Institute,  
United States  
Sicco Scherjon,  
University Medical Center Groningen,  
Netherlands

### \*Correspondence:

Jun Wang  
Wangj1@sj-hospital.org

### Specialty section:

This article was submitted to  
Genomic Medicine,  
a section of the journal  
Frontiers in Genetics

Received: 03 July 2020

Accepted: 30 October 2020

Published: 01 December 2020

### Citation:

Wang Y, Li Z, Song G and Wang J  
(2020) Potential of Immune-Related  
Genes as Biomarkers for Diagnosis  
and Subtype Classification  
of Preeclampsia.  
*Front. Genet.* 11:579709.  
doi: 10.3389/fgene.2020.579709

absence of effective medical intervention, PE may progress to eclampsia, so effective diagnosis and treatment are crucial.

The etiology and pathogenesis of PE are complex and have not been fully elucidated. In recent years, an imbalance in the immune system has been closely related to PE occurrence and development (Ma et al., 2019). A placental origin for PE is widely accepted. In early pregnancy, successful placental implantation depends on the precise regulation of the maternal immune system (Redman and Sargent, 2010; LaMarca et al., 2013). During this process, immune cells are needed for the invasive behavior observed in the decidual layer of the uterus. These immune cells gather around the trophoblasts and carry out different functions. They can control the migration of *in situ* cells of the spiral arteries and make trophoblasts migrate moderately to the intima by secreting cytokines and angiogenic factors (Perez-Sepulveda et al., 2014; Rahimzadeh et al., 2016). The balance between the immune cells and cytokines produced at the maternal-fetal interface is important for a normal pregnancy. Any local imbalance in the immune response may lead to abnormal placental structure or angiogenesis (Robillard et al., 2014; Sahay et al., 2014; LaMarca et al., 2016). Indeed, an insufficient invasion of trophoblasts during placental implantation leads to defective placental spiral artery remodeling, and both placenta and fetus are in a relatively ischemic state (Lai et al., 2020).

In our research, we screened for differentially expressed immune-related genes (DIRGs) between normotensive and PE pregnancies in the GSE60438 dataset. A stochastic forest model was constructed to predict the potential value of the DIRGs in diagnosing PE, and the important DIRGs were selected to establish a nomogram model to predict PE. In addition, we used a consensus clustering algorithm to classify PE pregnancies into three subgroups based on DIRG expression and explored the immune phenotype of these three subgroups.

## MATERIALS AND METHODS

### Data Source

The GSE60438 dataset containing two batches was obtained from the GEO database<sup>1</sup>. The second batch including 42 normotensive and 35 PE pregnancies was selected as training dataset (Yong et al., 2015). The first batch including 23 normotensive and 25 PE pregnancies was selected as validation dataset. We downloaded the gene expression profile and clinical data for our research. The probe number of the expression profile data was converted to gene symbols based on the annotation files. We obtained 2,499 immune-related genes (IRGs) from the ImmPort database<sup>2</sup>. The IRGs contained genes related to antigen processing and presentation, antimicrobials, the BCR and TCR signaling pathways, chemokines, cytokines, interleukins, and their respective receptors, natural killer cell cytotoxicity, and TGF $\beta$ , TGF $\beta$  receptor, TNF, and TNF receptor family members (Bhattacharya et al., 2014).

<sup>1</sup><http://www.ncbi.nlm.nih.gov/geo>

<sup>2</sup><https://immport.niaid.nih.gov>

### Screening for Differentially Expressed Immune-Related Genes

The “limma” package in R was used to identify differentially expressed genes (DEGs) between normotensive and PE pregnancies. We then crossed the DEGs with the IRGs to obtain the DIRGs for further investigation.

### Construction and RF, GLM, and SVM Models

We created generalized linear (GLM), random forest (RF), and support vector machine (SVM) models based on the training set. Positive-negative class balancing has been done by Combination/integration approach. The occurrence or absence of PE was used as the response variable, and the DIRGs were used as the explanatory variables. Next, we used the explain function of the “DALEX” package in R to analyze the three models and plotted the residual distribution to select the best model based on the validation dataset. Finally, we analyzed the importance of the variables and selected the four most important explanatory variables for further study.

### Construction and Evaluation of the Nomogram Model

A nomogram model was constructed using the “rms” package to facilitate the clinical application. The “Points” indicate the score of each factor under different conditions, while the “Total Points” refer to the total score of all factors. We measured the predictive accuracy of the nomogram by calibration curves and generated a clinical impact curve. The decision curve analysis (DCA) data were plotted to evaluate the clinical value of the nomogram.

### Consensus Clustering for DEGs

We performed consensus clustering to divide the PE cases into several subgroups based on the expression profiles of the identified DEGs using the “ConsensusClusterPlus” package in R. We used 1,000 iterations to ensure the stability of the clustering. The consensus k number, which was used to select the number of subgroups to divide the PE samples into, was determined by the cumulative distribution function (CDF) curves, delta area score of CDF, and consensus matrix heat maps (Wilkerson and Hayes, 2010; Zhang et al., 2020a).

### Evaluation of the Proportion of 22 Immune Cell Types in PE Pregnancies

Currently, the algorithms used to estimate the cell components in tissue based on gene expression profiles can be divided into two categories: Gene Set Enrichment Analysis (GSEA) and deconvolution. CIBERSORT is a deconvolution algorithm that can combine the labeled genomes of different immune cell subpopulations to calculate the proportion of 22 immune cell types in tissues. These immune cell types are presented in **Supplementary Table 1**. In this study, the CIBERSORT online platform<sup>3</sup> was used for this analysis, and each sample obtained a

<sup>3</sup><http://cibersort.stanford.edu/>

$p$ -value. Samples with a CIBERSORT output value of  $p < 0.05$  were considered statistically significant and analyzed further (Zhou et al., 2018; Zhang et al., 2019).

## Statistical Analysis

We used Wilcox.test or Kruskal-Wallis tests to compare the differences between the groups. The “RCircos” package in R was used to map the chromosomal positions of the DIRGs. Spearman correlation analysis was performed to calculate the correlation coefficients between the DIRGs. A two-sided  $p$ -value of less than 0.05 was considered statistically significant. All statistical analyses were performed using R 4.0.0.

## RESULTS

### Landscape of DIRGs in PE Pregnancies

Twenty-seven DEGs were identified between normotensive and PE pregnancies according to the screening criteria of  $\log |FC| > 0.1$  and  $p < 0.05$  using the “limma” package based on the training dataset. We obtained 13 DIRGs by crossing the 27 DEGs with 2,499 IRGs, which were used to generate the Venn diagram (Figure 1A). In addition, 36 immune genes in our analysis overlapped with the original article according to the screening criteria of  $P < 0.05$ . The overlapping immune genes were listed in Supplementary Table 1. The expression of the 13 DIRGs in the normotensive and PE pregnancies are shown in Figures 1B,C. Both the heat map and the histogram showed that PI3, CCL18, CCL2, LTB, and CD48 were expressed at low levels in PE pregnancies compared to normotensive pregnancies. In contrast, the expression levels of LEP, CGB1, CDF15, LHB, CGB8, CGA, CGB5, and CRH were higher in the PE pregnancies than in the normotensive pregnancies. The chromosomal positions of the 13 DIRGs are shown in Figure 1D. We compared the immune infiltration between normotensive and PE pregnancies. The results revealed that the immune infiltration between normotensive and PE pregnancies have no statistical differences (Figure 2). It is generally known that PE pregnancies can cause premature birth at a gestational age of fewer than 37 weeks or an infant weight of less than 2.5 kg. This phenomenon is presented in Supplementary Figures 1A,B. Thus, we also investigated the relationship between the 13 DIRGs and clinical information (e.g., gestational age and infant weight). The results indicated that the 13 DIRGs were associated with premature birth (Supplementary Figures 1C,D).

### Construction and RF Model

The RF, GLM, and SVM models were each established using the training set. We explained the three models using the “DALEX” package and plotted the residual distribution to select the best model based on the test set. The RF model had the least sample residual (Supplementary Figure 2). Thus, the RF model was constructed to distinguish normotensive and PE pregnancies in our research. We analyzed the importance of the variables based on the RF model at the gestational age from 25 to 41 week (Figure 3A). Variable importance is ranked according to %IncMSE. “%IncMSE” represents an increase in

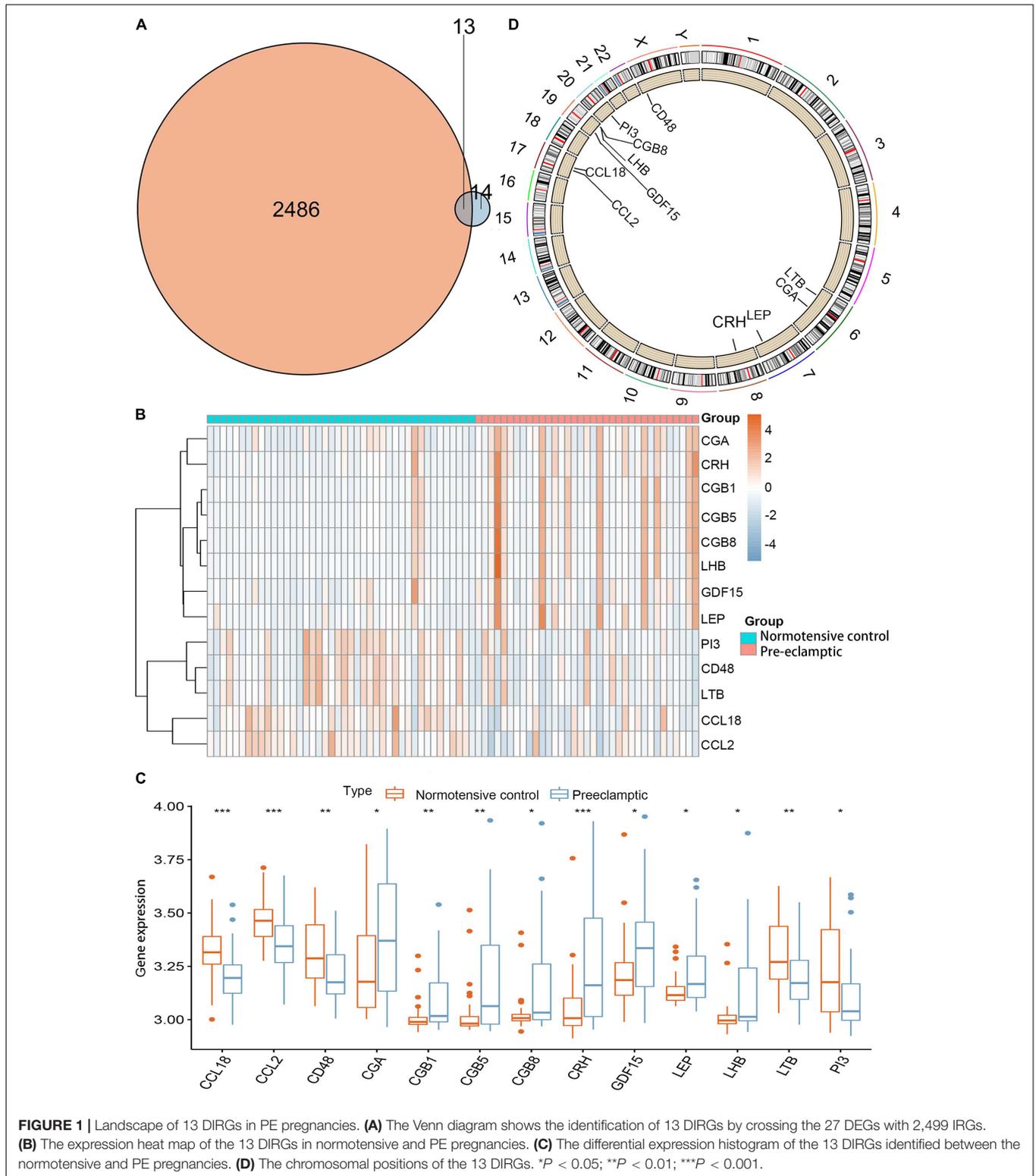
mean squared error. The more important a predictor gene is, the greater the prediction error will be, when its value is randomly replaced. Then, we performed a 10-fold cross validation to select the appropriate important variables and four most important explanatory variables (CRH, PI3, CCL18, and CCL2) from the RF model was selected for further evaluation. We calculated the risk score of each patients based on the regression coefficients using multivariate Cox proportional hazards regression (PHR) analysis. The risk score of each patients in training dataset was present in Supplementary Table 2. The time-dependent ROC curves indicated that the accuracy of the RF model is well based on the four most important explanatory variables at 30-, 36-, and 41-week (Figure 3B). The AUC value of the ROC curve increased with the elevated gestational weeks. The four most important explanatory variables have the most predictive power at 41-week during pregnancy. Correlation analysis of the 13 DIRGs revealed that the correlation coefficients between the four selected DIRGs were lower (Figure 4), suggesting the RF model exclude genes with similar functions to simplify the model and reduce unnecessary costs.

### Construction and Evaluation of the Nomogram Model

To facilitate the ability to predict PE pregnancies clinically, we constructed a nomogram model using the “rms” package based on the four selected DIRGs (CRH, PI3, CCL18, and CCL2) from the training dataset (Figure 5A). Calibration curves revealed that the predictiveness of the nomogram model was accurate (Figure 5B). The DCA curve was plotted to evaluate the clinical value of this nomogram model. The x-axis indicates the predicted probability, and the y-axis represents the net benefit. The oblique red line revealed that the nomogram model could benefit patients at high risk threshold from 0.1 to 0.9 (Figure 5C). We further evaluated the clinical impact curve based on the DCA curve to more intuitively assess the clinical impact of this model. We found that the predictive power of the nomogram model was remarkable. The predicted number of high-risk patients was greater than that of high-risk patients with an event (Figure 5D). We then test the nomogram model in the validation dataset based on the four selected DIRGs. Calibration curves revealed that the accuracy of the nomogram model was good (Supplementary Figure 3A). DCA curve and clinical impact curve indicated that patients can benefit from the nomogram at high risk threshold almost from 0.1 to 0.9 (Supplementary Figures 3B,C).

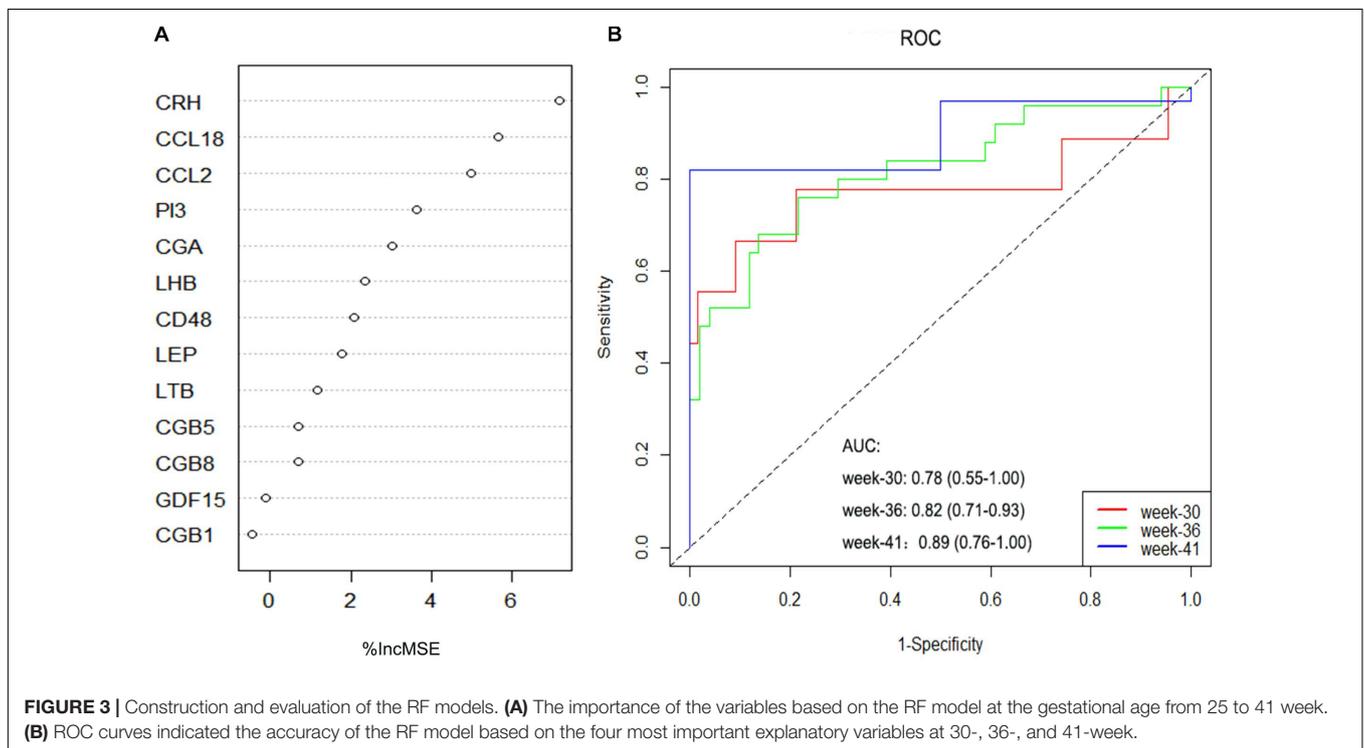
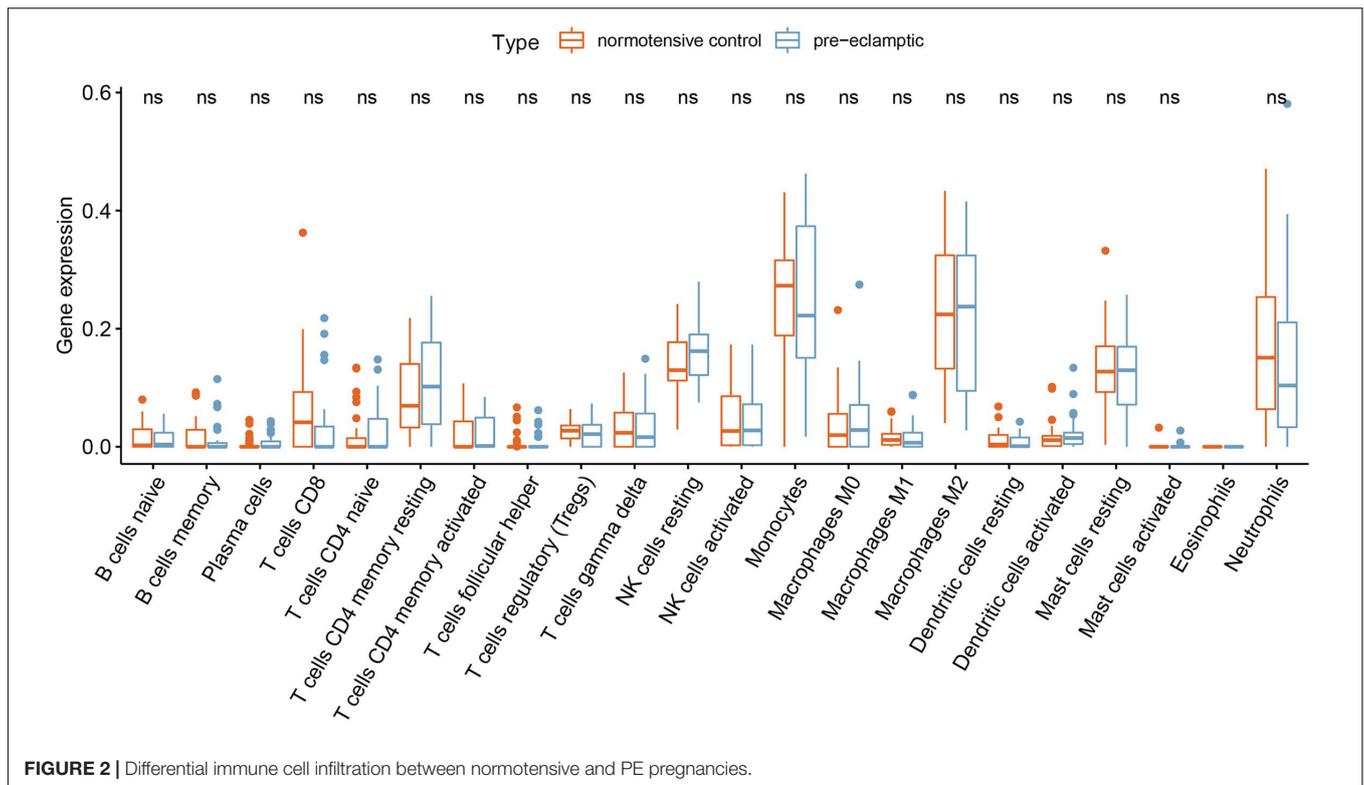
### Three Subgroups Obtained by Consensus Clustering

Consensus clustering was used to divide the PE cases into several subgroups to explore the function of the 13 DEGs in PE based on their expression profiles. We plotted a CDF curve that allows us to determine at the  $k$  number. The CDF can have an approximate maximum when we choose the most suitable  $k = 3$  (Figure 6A). Figure 6B shows the delta area score of the CDF curve from  $k = 2-9$ . We found that the relative increase in the delta area score tends to be stable after  $k = 3$ . Thus, the PE cases could be divided into three subgroups. Figure 6C provides a view of the



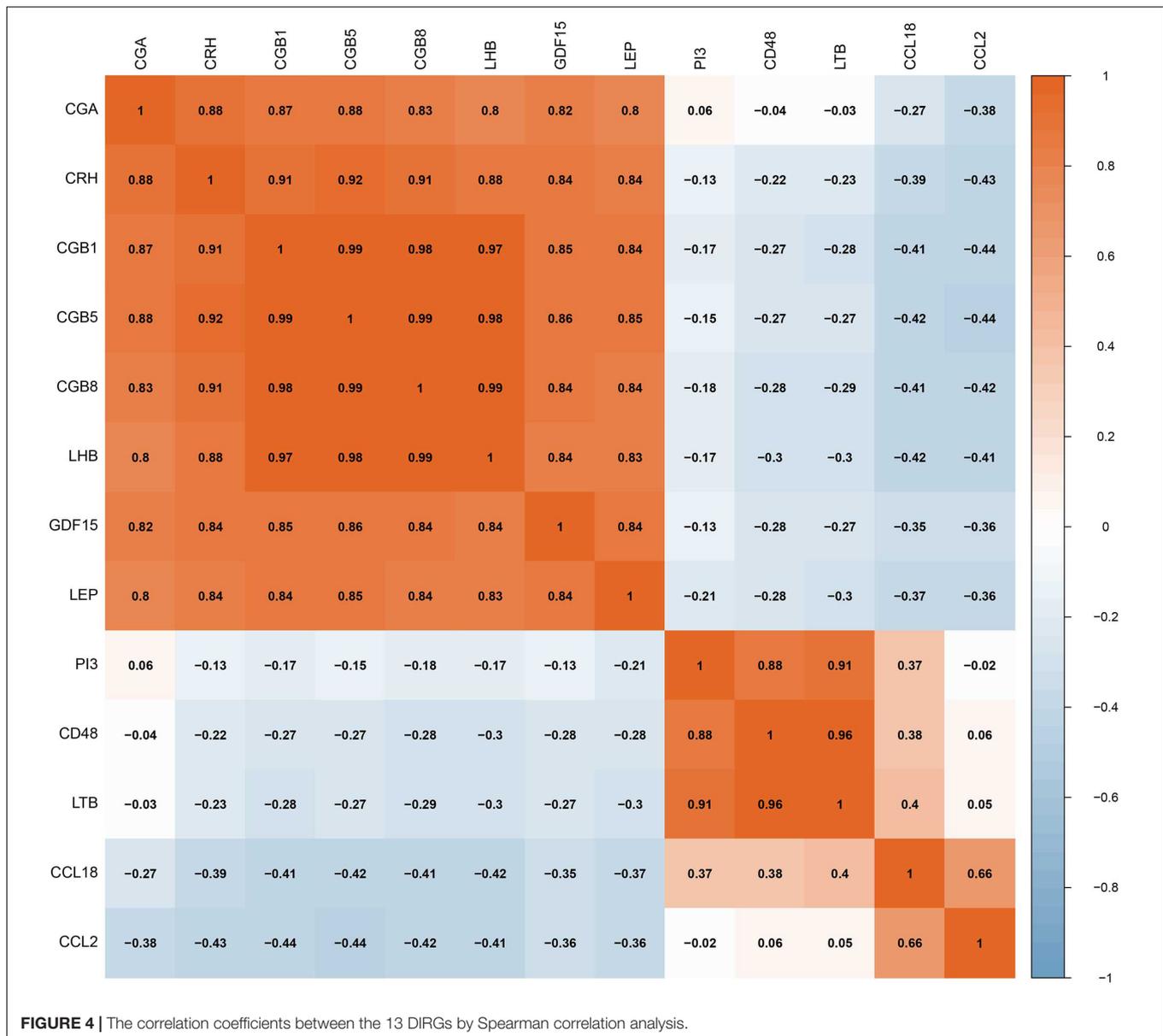
item cluster membership across the different  $k$  numbers to track the cluster history. The matrix heat maps were clearly separated when  $k = 3$  (Figure 6D). We also explored the expression of the 13 DIRGs in the three subgroups (clusters A, B, and C).

The heat map and the histogram showed that CCL18, CD48, LTB, and PI3 were highly expressed in cluster B, whereas CCL2 was highly expressed in cluster A, and CGA, CGB1, CGB5, CGB8, CRH, GDF15, LEP, and LHB were highly expressed in



cluster C compared to the other subgroups (**Figures 7A,B**). The principal component analysis (PCA) indicated that the expression of the 13 DIRGs could completely distinguish the three subgroups (**Figure 7C**).

We then test the consensus clustering in validation dataset based on the 12 DEGs (GDF15 gene could not be found in the validation dataset) and found similar grouping results (**Supplementary Figure 4**). The heat map and the histogram



revealed that except for CCL2 gene, the difference distribution of the other genes in the three groups was similar (**Supplementary Figures 5A,B**). The principal component analysis (PCA) indicated that the expression of the 12 DIRGs could also distinguish the three subgroups (**Supplementary Figure 5C**).

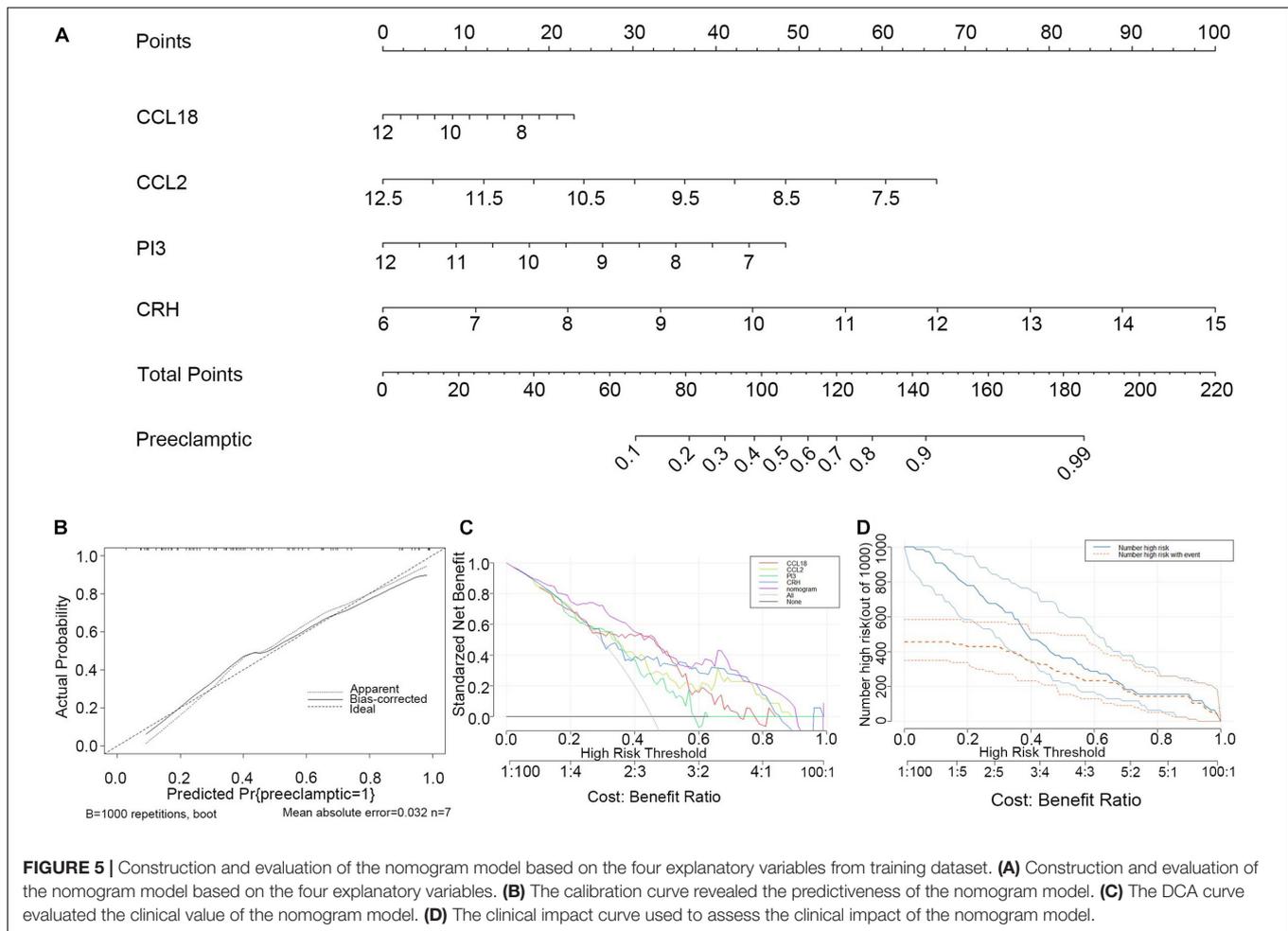
### Clinical Traits and Immune Cell Infiltration Characteristics of the Three Subgroups

To explore the significance of the three subgroups, we compared the gestational age and infant weight between the three subgroups by Kruskal-Wallis tests based on the training dataset. We found that the cluster C subgroup had a lower gestational age and infant weight than the other two subgroups (**Figures 8A,B**). Differential analysis of the immune cell infiltration levels in the

three subgroups revealed that infiltration by the memory B cells, naïve CD4<sup>+</sup> T cells,  $\gamma\delta$  T cells, monocytes, and neutrophils in the cluster B subgroup was higher than in the other subgroups while resting dendritic cells and M2 macrophages were found in higher numbers in the cluster C subgroup (**Figure 8C**).

### DISCUSSION

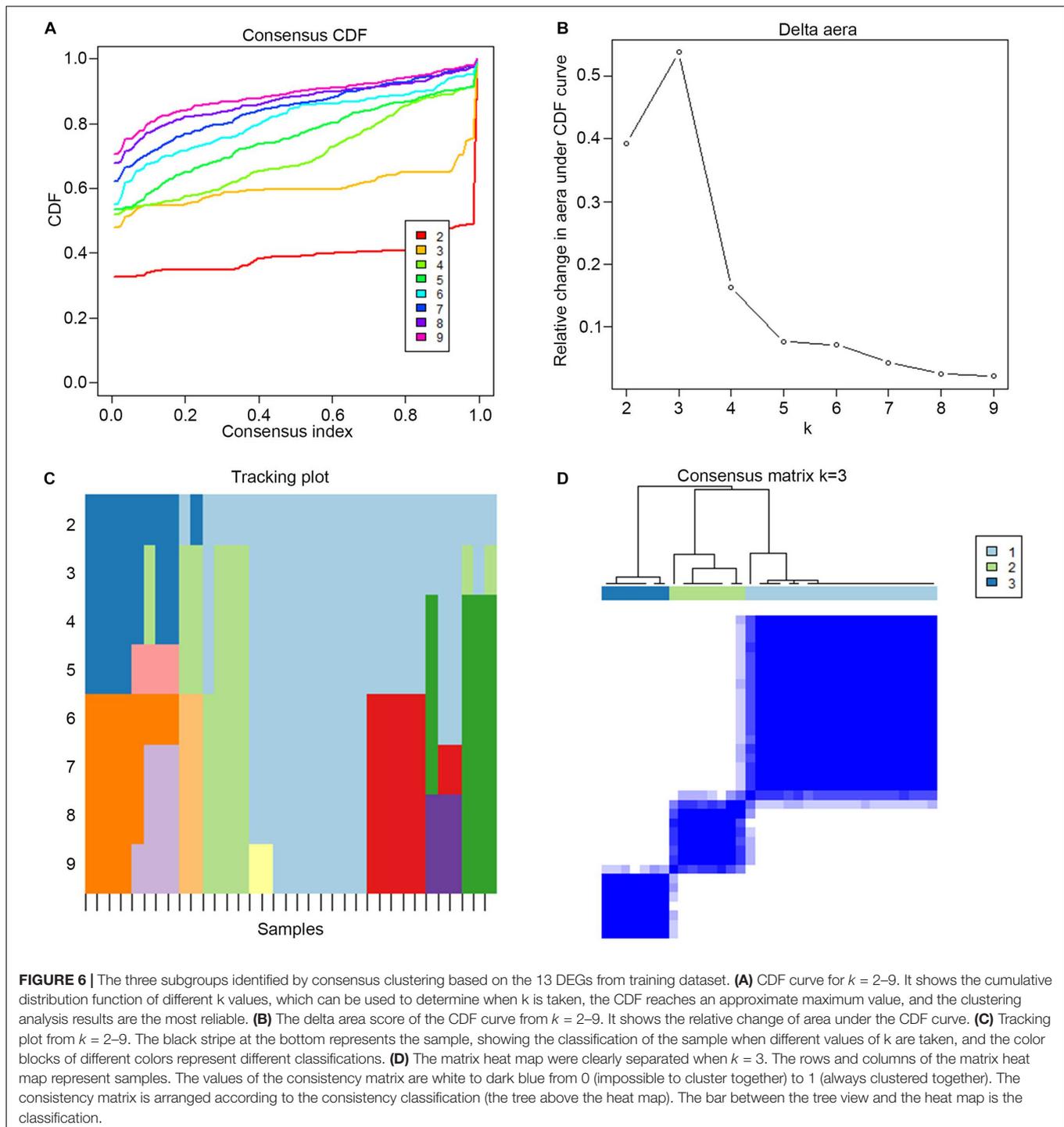
PE is the main cause of maternal mortality (Abalos et al., 2013). There is currently no gold standard for diagnosis or effective preventive and treatment methods. The immune system consists of cytokines, chemokines, inhibitory receptors, and ligands, and immune cells that play important roles in normal pregnancy (Kubes and Jenne, 2018). An immune imbalance may lead



to pregnancy-specific complications, including PE (Schumacher et al., 2018). In our research, we explored the significance of immune-related genes in PE and established a basis for screening diagnosis markers and individualized immunotherapy for PE.

We screened 13 immune-related genes (PI3, CCL18, CCL2, LTB, CD48, LEP, CGB1, CDF15, LHB, CGB8, CGA, CGB5, and CRH) based on significant differential expression analysis between normotensive and PE pregnancies. PI3, CCL18, CCL2, LTB, and CD48 were downregulated in PE. In contrast, LEP, CGB1, CDF15, LHB, CGB8, CGA, CGB5, and CRH were upregulated. Most of the significant immune-related genes belonged to the category of antimicrobials and cytokines that function in the reaction to bacteria, inflammatory responses, and immune processes. We also explored the relationship between the 13 DIRGs and premature birth. The results suggested that low expression of PI3, CCL18, CCL2, LTB, and CD48 and high expression of LEP, CGB1, CDF15, LHB, CGB8, CGA, CGB5, and CRH might be involved in the occurrence of premature delivery. These data are consistent with the adverse outcome of premature delivery caused by PE. RF is an integrated algorithm composed of a decision tree, which can be used as a classification tool. The decision tree produces a tree model, classifies the data, and forecasts by repeatedly distinguishing each variable. RF has

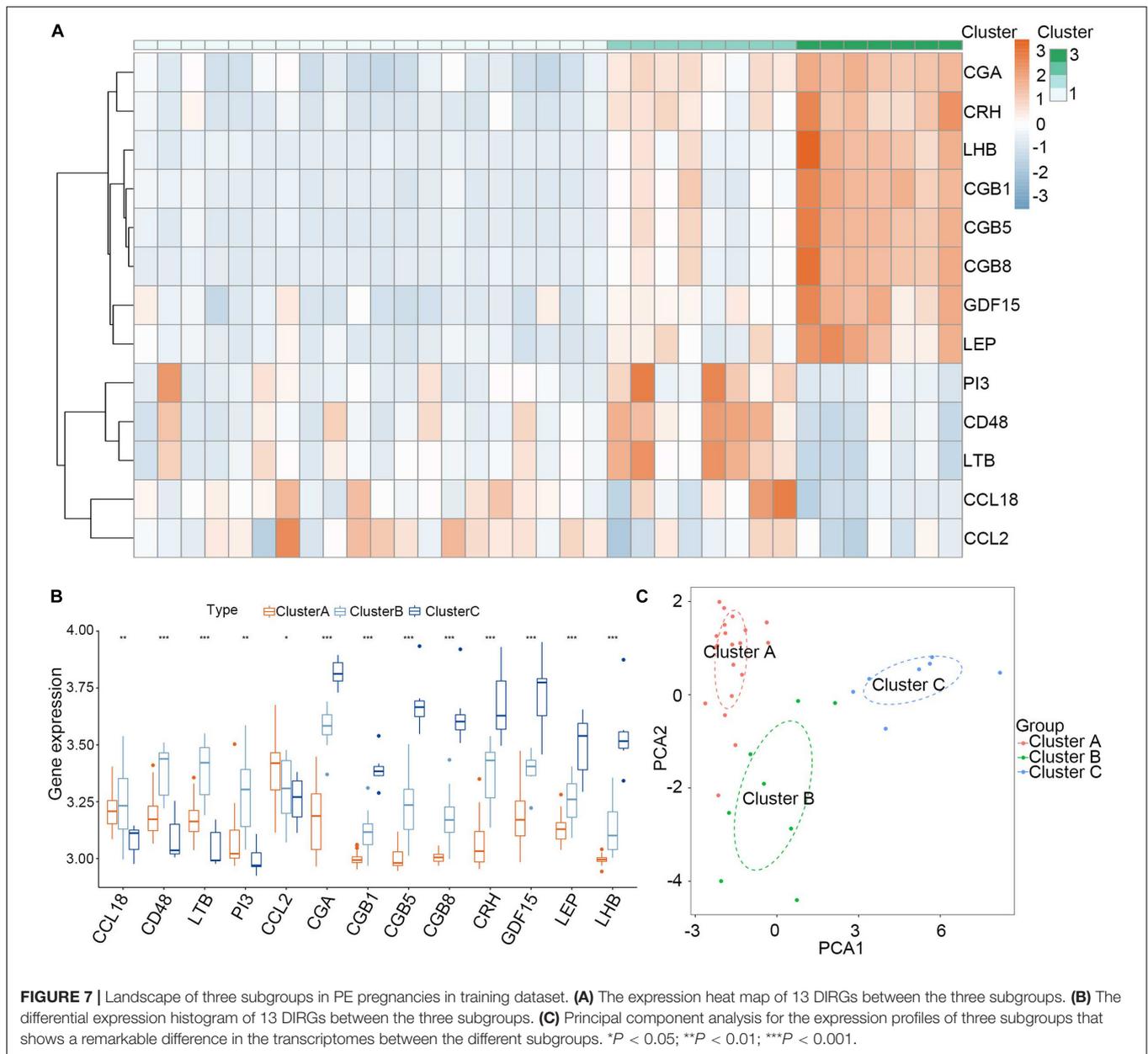
numerous merits, such as judging the importance of features, does not easily overfit, and its training speed is relatively fast (Jeong et al., 2020). GLM is a widely used statistical model that is mainly used to solve a binary classification problem, and represents the possibility of something happening (Qu and Luo, 2015). SVM was first proposed by Cortes and Vapnik in 1963 (Ben-Dor et al., 2000). It has many unique advantages in solving small sample, nonlinear, and high-dimensional pattern recognition (Huang et al., 2018). We constructed different models based on these methods and found that the RF model was the best model for identifying the most suitable biomarkers for diagnosing PE. Indeed, the DCA and clinical impact curves demonstrated that PE patients could benefit from the nomogram generated using the four explanatory variables (CRH, PI3, CCL18, and CCL2) identified using the RF model in both training dataset and testing dataset. We used consensus clustering analysis of the 13 DIRGs to classify PE pregnancies into three subgroups. CCL2 was higher expression in cluster A subgroup compared with the other groups, while CCL18, CD48, LTB, PI3 were higher expression in cluster B subgroup, and CGA, CGB1, CGB5, CGB8, CRH, CDF15, LEP, LHB were higher expression in cluster C subgroup. We also found similar grouping results in testing dataset based on the 12 DEGs (GDF15 gene could not be



found in the validation dataset). The heat map and the histogram revealed that except for CCL2 gene, the difference distribution of the other genes in the three groups was similar, which revealed the stability of the grouping by consensus clustering analysis.

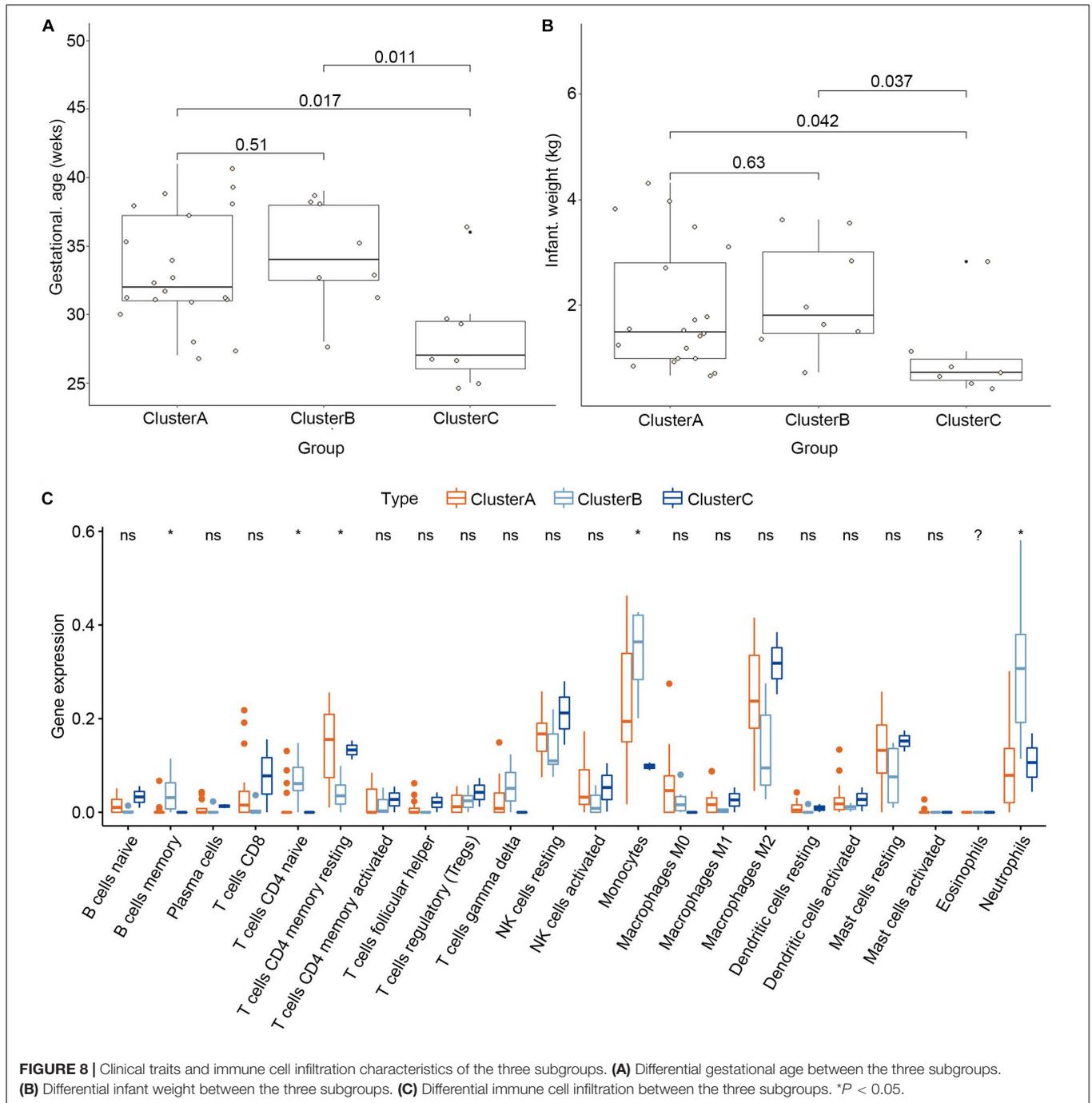
Proteins related to the genes higher expression in cluster C subgroup were mainly hormone-associated protein and the cluster C subgroup had a lower gestational age and infant

weight than the other groups. Therefore, boldly speculate that a large number of hormone levels are positively correlated with the risk of PE and the conjecture was highly consistent with previous studies (Makrigiannakis et al., 2018; Al-Kaabi et al., 2020; Daskalakis et al., 2020). Antonis Makrigiannakis suggested that corticotropin releasing hormone (CRH) may contribute to preeclampsia (Makrigiannakis et al., 2018). The



$\beta$ -hCG encoded by CGB5 and CGB8 genes was demonstrated positive correlation with preeclampsia (Al-Kaabi et al., 2020). Georgios Daskalakis reported that preeclampsia is associated with increased leptin (LEP) (Daskalakis et al., 2020). In addition, we found that memory B cells, naïve CD4+T cells,  $\gamma\delta$  T cells, monocytes, and neutrophils had higher infiltration in the cluster B subgroup. Similarly, resting dendritic cells and M2 macrophages showed higher infiltration in the cluster C group. We speculate that immunosuppression may cause a disorder in the internal environment, leading to premature birth. Additional research is needed to explore the relationship between immune cell infiltration and the PE subgroups, gestational age, and infant weight of the subgroups of PE patients.

This study had some limitations. Firstly, the sample tissues came from decidual tissue rather than blood, so we could not determine whether the selected diagnostic markers would be suitable for measuring in blood samples. Secondly, our research results are only predictions. These results need to be verified by basic science and clinical studies. Thirdly, since the decidual basalis samples are taken at the occurrence of PE, it will be interesting to see whether the biomarkers cytokines can be detected in serum. However, we searched the data sets in the database and found no transcriptome data that met the requirements. Fourthly, we did not verify our research results using other datasets due to the limitations of the data in the database. Finally, the key immune-related genes used for PE diagnosis were selected based on the placental tissue samples at



the gestational age from 25 to 41 week. Whether these genes have the potential to be biomarkers for early diagnosis need further experimental studies.

### CONCLUSION

In conclusion, our study selected four explanatory variables and established a nomogram model to predict PE. We also identified three subgroups with different clinical outcomes and immune cell

infiltration. Future studies on these three subgroups may provide direction for individualized treatment of PE patients.

### DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This work was supported by the Natural Science Foundation of Liaoning Province (No. 20180530076) and the Key Research and Development Joint Program Liaoning Province (No. 2020JH2/10300146).

## ACKNOWLEDGMENTS

We thank the authors who provided the GEO public datasets.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.579709/full#supplementary-material>

## REFERENCES

- Abalos, E., Cuesta, C., Grosso, A. L., Chou, D., and Say, L. (2013). Global and regional estimates of preeclampsia and eclampsia: a systematic review. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 170, 1–7. doi: 10.1016/j.ejogrb.2013.05.005
- Al-Kaabi, M. A., Hamdan, F. B., and Al-Matubi, H. (2020). Maternal plasma kisspeptin-10 level in preeclamptic pregnant women and its relation in changing their reproductive hormones. *J. Obstet. Gynaecol. Res.* 46, 575–586. doi: 10.1111/jog.14208
- Ben-Dor, A., Bruhn, L., Friedman, N., Nachman, I., Schummer, M., and Yakhini, Z. (2000). Tissue classification with gene expression profiles. *J. Comp. Biol.* 7, 559–583. doi: 10.1089/106652700750050943
- Bhattacharya, S., Andorf, S., Gomes, L., Dunn, P., Schaefer, H., Pontius, J., et al. (2014). ImmPort: disseminating data to the public for the future of immunology. *Immunol. Res.* 58, 234–239. doi: 10.1007/s12026-014-8516-1
- Daskalakis, G., Bellos, I., Nikolakea, M., Pergialiotis, V., Papapanagiotou, A., and Loutradis, D. (2020). The role of serum adipokine levels in preeclampsia: a systematic review. *Metab. Clin. Exp.* 106:154172. doi: 10.1016/j.metabol.2020.154172
- Grandi, S. M., Filion, K. B., Yoon, S., Ayele, H. T., Doyle, C. M., Hutcheon, J. A., et al. (2019). Cardiovascular disease-related morbidity and mortality in women with a history of pregnancy complications. *Circulation* 139, 1069–1079. doi: 10.1161/circulationaha.118.036748
- Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., and Xu, W. (2018). Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics* 15, 41–51.
- Jeong, B., Cho, H., Kim, J., Kwon, S. K., Hong, S., Lee, C., et al. (2020). Comparison between statistical models and machine learning methods on classification for highly imbalanced multiclass kidney data. *Diagnostics (Basel)* 10:415. doi: 10.3390/diagnostics10060415
- Kubes, P., and Jenne, C. (2018). Immune responses in the liver. *Ann. Rev. Immunol.* 36, 247–277.
- Lai, H., Nie, L., Zeng, X., Xin, S., Wu, M., Yang, B., et al. (2020). Enhancement of heat shock protein 70 attenuates inducible nitric oxide synthase in preeclampsia complicated with fetal growth restriction. *J. Matern. Fetal Neonatal Med.* 13, 1–9. doi: 10.1080/14767058.2020.1789965
- LaMarca, B., Cornelius, D., and Wallace, K. (2013). Elucidating immune mechanisms causing hypertension during pregnancy. *Physiology (Bethesda Md)* 28, 225–233. doi: 10.1152/physiol.00006.2013
- LaMarca, B., Cornelius, D. C., Harmon, A. C., Amaral, L. M., Cunningham, M. W., Faulkner, J. L., et al. (2016). Identifying immune mechanisms mediating the hypertension during preeclampsia. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 311, R1–R9.
- Ma, Y., Ye, Y., Zhang, J., Ruan, C. C., and Gao, P. J. (2019). Immune imbalance is associated with the development of preeclampsia. *Medicine* 98:e15080. doi: 10.1097/md.00000000000015080
- Makrigiannakis, A., Vrekoussis, T., Zoumakis, E., Navrozoglou, I., and Kalantaridou, S. N. (2018). CRH receptors in human reproduction. *Curr. Mol. Pharm.* 11, 81–87.
- Perez-Sepulveda, A., Torres, M. J., Khoury, M., and Illanes, S. E. (2014). Innate immune system and preeclampsia. *Front. Immunol.* 5:244. doi: 10.3389/fimmu.2014.00244
- Phippis, E. A., Thadhani, R., Benzing, T., and Karumanchi, S. A. (2019). Preeclampsia: pathogenesis, novel diagnostics and therapies. *Nat. Rev. Nephrol.* 15, 275–289. doi: 10.1038/s41581-019-0119-6
- Qu, Y., and Luo, J. (2015). Estimation of group means when adjusting for covariates in generalized linear models. *Pharm. Stat.* 14, 56–62. doi: 10.1002/pst.1658
- Rahimzadeh, M., Norouzian, M., Arabpour, F., and Naderi, N. (2016). Regulatory T-cells and preeclampsia: an overview of literature. *Expert Rev. Clin. Immunol.* 12, 209–227. doi: 10.1586/1744666x.2016.1105740
- Redman, C. W., and Sargent, I. L. (2010). Immunology of pre-eclampsia. *Am. J. Reprod. Immunol.* 63, 534–543. doi: 10.1111/j.1600-0897.2010.00831.x
- Robillard, P. Y., Dekker, G., and Chaouat, G. (2014). Fourteen years of debate and workshops on the immunology of preeclampsia. Where are we now after the 2012 workshop? *J. Reprod. Immunol.* 10, 62–69. doi: 10.1016/j.jri.2013.06.003
- Sahay, A. S., Patil, V. V., Sundrani, D. P., Joshi, A. A., Wagh, G. N., Gupte, S. A., et al. (2014). A longitudinal study of circulating angiogenic and antiangiogenic factors and AT1-AA levels in preeclampsia. *Hypertens. Res.* 37, 753–758. doi: 10.1038/hr.2014.71
- Say, L., Chou, D., Gemmill, A., Tunçalp, Ö, Moller, A. B., and Daniels, J. (2014). Global causes of maternal death: a WHO systematic analysis. *Lancet Glob. Health* 2, e323–e333.

**Supplementary Figure 1** | Correlations between the clinicopathological features and the 13 DIRGs. (A) Differential gestational age between normotensive and PE pregnancies. (B) Differential infant weight between normotensive and PE pregnancies. (C) The differential expression of the 13 DIRGs among distinct gestational ages. (D) The differential expression of the 13 DIRGs among distinct infant weights.

**Supplementary Figure 2** | Residual distributions of the RF, GLM, and SVM models.

**Supplementary Figure 3** | Construction and evaluation of the nomogram model based on the four explanatory variables from validation dataset. (A) Construction and evaluation of the nomogram model based on the four explanatory variables. (B) The calibration curve revealed the predictiveness of the nomogram model. (C) The DCA curve evaluated the clinical value of the nomogram model. (D) The clinical impact curve used to assess the clinical impact of the nomogram model.

**Supplementary Figure 4** | The three subgroups identified by consensus clustering based on the 12 DEGs from validation dataset.

**Supplementary Figure 5** | Landscape of three subgroups in PE pregnancies in validation dataset. (A) The expression heat map of 12 DIRGs between the three subgroups. (B) The differential expression histogram of 12 DIRGs between the three subgroups. (C) Principal component analysis for the expression profiles of three subgroups that shows a remarkable difference in the transcriptomes between the different subgroups.

**Supplementary Table 1** | Thirty-six immune genes in our analysis overlapped with the original article according to the screening criteria of  $P < 0.05$ .

**Supplementary Table 2** | The risk score of each patients in training dataset.

- Schumacher, A., Sharkey, D. J., Robertson, S. A., and Zenclussen, A. C. (2018). Immune cells at the fetomaternal interface: how the microenvironment modulates immune cells to foster fetal development. *J. Immunol.* 201, 325–334. doi: 10.4049/jimmunol.1800058
- Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26, 1572–1573. doi: 10.1093/bioinformatics/btq170
- Yong, H. E., Melton, P. E., Johnson, M. P., Freed, K. A., Kalionis, B., Murthi, P., et al. (2015). Genome-wide transcriptome directed pathway analysis of maternal pre-eclampsia susceptibility genes. *PLoS One* 10:e0128230. doi: 10.1371/journal.pone.0128230
- Zhang, B., Wu, Q., Li, B., Wang, D., Wang, L., and Zhou, Y. L. (2020a). m(6)A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. *Mol. Cancer* 19:53.
- Zhang, S., Zhang, E., Long, J., Hu, Z., Peng, J., Liu, L., et al. (2019). Immune infiltration in renal cell carcinoma. *Cancer Sci.* 110, 1564–1572. doi: 10.1111/cas.13996
- Zhang, Z., Wang, P., Zhang, L., Huang, C., Gao, J., Li, Y., et al. (2020b). Identification of key genes and long noncoding RNA-associated competing endogenous RNA (ceRNA) networks in early-onset preeclampsia. *BioMed. Res. Int.* 2020:1673486.
- Zhou, Z., Cheng, Y., Jiang, Y., Liu, S., Zhang, M., Liu, J., et al. (2018). Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis. *Int. J. Biol. Sci.* 14, 124–136. doi: 10.7150/ijbs.22619

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Wang, Li, Song and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.