Computational Biology and Chemistry 75 (2018) 222–230

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/cbac

Research Article

Statistical and artificial neural network-based analysis to understand complexity and heterogeneity in preeclampsia

T. Murlidharan Nair[a,b,*]

[a] Department of Biological Sciences, Indiana University South Bend, South Bend, IN, United States
[b] Department of Computer Science/Informatics, Indiana University South Bend, South Bend, IN, United States

ARTICLE INFO

Keywords:
Preeclampsia
Biomarkers
Artificial neural networks
Calliper randomization

ABSTRACT

Preeclampsia is a pregnancy associated disease. It is characterized by high blood pressure and symptoms that are indicative of damage to other organ systems, most often involving the liver and kidneys. If left untreated, the condition could be fatal to mother and baby. This makes it important to delineate the complexities associated with the disease at a molecular level that would help develop methods for early diagnosis. In microarray-based studies, Textoris et al. and Mirzakhani et al. have analyzed the transcriptome with a view to identify biomarkers for preeclampsia. The current study has extensively analyzed these microarray data sets to understand the complexity and heterogeneity associated with preeclampsia. A statistical multiple comparisons-based approach has been used to identify features capable of distinguishing preeclampsia from normotensive cases. These features were then used to build an artificial neural network-based machine learning model that successfully classified the samples. Further, the machine learning model was used to delineate features critical for its internal representation by extending the calliper randomization approach to the analysis of microarray data. Functional analysis of the features identified by the calliper randomization approach revealed pathways that could be crucially involved in the mechanism of the underlying disease. Biological processes associated with the features identified have revealed among others, genes involved in reproductive processes to be differentially expressed.

1. Introduction

Placental disease of preeclampsia (PE) is defined by the onset of hypertension and proteinuria or dysfunction of major organs due to the onset of hypertension after 20 weeks of gestation (Chaiworapongsa et al., 2014a,b; Ngoc et al., 2006; Young et al., 2010). It is estimated that between 3 and 5% of all pregnancies are affected globally by PE. Further, it is alarming that 40–60% of maternal mortality has also been attributed to PE (Ngoc et al., 2006; Young et al., 2010). In the United States, PE accounts for 20% of maternal mortality (MacKay et al., 2001). It is unclear what factors contribute to PE. Studies show that ~7.5% of healthy nulliparous women may be affected, and multiparous women pregnant with a new partner could also have an elevated risk of PE comparable to nulliparous women (Tubbergen et al., 1999). Genetic predisposition to PE has also been found to be significant (Barton and Sibai, 2008). Preexisting morbidities that are associated with increased PE risk include chronic hypertension, diabetes mellitus, and blood clotting disorders (Barton and Sibai, 2008; Duckitt and Harrington, 2005).

The etiopathogenesis of PE has been explained using a hypothesized two-stage theory (Roberts and Hubel, 2009). Briefly, according to this theory, poorly profused placenta causes hypoxia (stage 1) resulting in the release of soluble factors such as reactive oxygen species, pro-inflammatory cytokines and antiangiogenic factors into the maternal circulation. This then leads to the clinical manifestation of PE (stage 2). In a metaphoric view of PE, pregnancy was considered akin to a car with accelerators and brakes (Ahmed and Ramma, 2015). The accelerators in this case were inflammation, oxidative stress and an anti-angiogenic state, while the braking system were the protective pathways, haem-oxygenase-1 (Hmox1) and cystathionine-γ-lyase (CSE). Failure of the protective pathways (brakes) that control the accelerators results in the manifestation of PE. The imbalance between vasodilation and vasoconstriction due to the soluble factors released into the maternal circulation possibly creates a condition of endothelial dysfunction and may persist after delivery. Another aspect, immunological dysfunction, has also been shown to play an important role in PE (Redman and Sargent, 2010; Cheng and Sharma, 2016; Rademacher et al., 2007). It is now an established fact that there is a complex interplay between the maternal immune system and placenta (Cheng and Sharma, 2016), but it is still unclear when these immunological

* Corresponding author at: Department of Biological Sciences, Indiana University South Bend, South Bend, IN, United States.
E-mail address: mnair@iusb.edu.

https://doi.org/10.1016/j.compbiolchem.2018.05.011
Received 20 November 2017; Received in revised form 11 April 2018; Accepted 9 May 2018
Available online 16 May 2018
1476-9271/

alternations begin. It has been hypothesized that dysregulated systemic and placental immunity may play a role in the onset of PE (Cheng and Sharma, 2016).

It is becoming increasingly clear that PE is a multifactorial condition that is genetically and immunologically governed. PE is also heterogeneous in its presentation, making it challenging to diagnose as well as to identify biomarkers (Cuffe et al., 2017). Given the risk associated with PE, biomarkers that can predict PE would allow for improved and timely intervention and prevent adverse outcomes. Earlier attempts to identify biomarkers have had limited success. Some of the potential biomarkers identified, viz. soluble endoglin, soluble vascular endothelial growth factor receptor and placental growth factor, were not able to successfully predict early PE (McElrath et al., 2012; Widmer et al., 2015). Since maternal symptoms of pregnancy are caused by factors released by placenta, they could potentially serve as diagnostic or predictive markers. However, none of the biomarkers identified thus far have been found to be clinically useful (Cnossen et al., 2005). This may be because of the complexity of association among the factors, and the methods used in their identification are unable to capture these complex cross associations. This is the problem that this paper attempts to explain using data sets from studies by Textoris et al. (2013) and Mirzakhani et al. (2016).

In the current study, the microarray data sets that were used to identify biomarkers for PE have been exhaustively analyzed both statistically, and by using a machine learning-based method. A simple approach capable of identifying possible features contained in data harboring complex correlations is presented. Functional analysis of the theoretically identified features is also presented.

2. Materials and methods

2.1. Microarray data

Microarray data generated by Textoris et al. (2013) was one of the data sets used in this study. The data is publicly available from NCBI gene expression omnibus (GSE48424). Briefly, the data consisted of blood RNA levels from 38 women that were part of their study: 19 patients with PE and 19 normotensive (NT) women selected based on age, weight, smoking status, race, gestational age and blood pH. Among the 19 patients with PE, 6 women were categorized as non-severe PE (nsPE) and 13 women as severe PE (sPE). The study had used a 4X44K Whole Human genome microarray G4112F chip manufactured by Agilent Technologies and followed the gene expression analysis method described in Thuny et al. (2012). For further details regarding the data, please refer to Textoris et al. (2013). In addition, the transcriptomic data generated as part of the Vitamin D antenatal asthma reduction trial (VDAART) was also analyzed (Mirzakhani et al., 2016). The gene expression data (GSE85307) in the VDAART study was profiled using Affymetrix GeneChip Arrays “Affymetrix Human Gene 1.0 ST Array”. The data constituted expression profile of 47 patients with PE and 110 normotensive cases.

2.2. Feature selection using statistical method

A statistical multiple comparison-based method was used to compare the expression values (Westfall, 1997). Comparing the differences in two population means may be achieved using an interval estimation or hypothesis testing approach, but comparison of several means can be done by ANOVA F-test. However, ANOVA F-test only provides information on whether the difference between the two means is significant. In order to obtain information on which means were different, one must resort to multiple comparison (see Fig. 1). The multcomp R package was used in the multiple comparison of the data (www.r-project.org). Exhaustive comparisons within and between the classes were carried out. Following the exhaustive comparison of expression levels from each probe, a unit score was given to each significant difference. Sums of scores were used to rank the probes. In case of the Textoris et al. data set (GSE48424), the samples were classified either as two classes (binary) or four classes (multiple). In the binary comparison analysis of the data, the categories of nsPE and sPE were clubbed into one group as PE and compared with NT. The features that were found to be significant were used for further analysis. In order to facilitate exhaustive comparison within and between each class of samples, each of the classes were further randomly divided as PE1, PE2, nsPE1, nsPE2, NT1 and NT2. This permitted within-class comparison and helped to understand the heterogeneity associated with the data.

In case of the VDAART transcriptome data set (GSE85307), the samples were classified as PE or NT. To delineate heterogeneity, within-class comparisons were done similarly. PE and NT samples were randomly divided into two sets each, PE1 and PE2, and NT1 and NT2. The samples that belonged to the same class (PE or NT) were then compared to each other.

2.3. Extracting complex features using artificial neural network-based calliper randomization approach

Artificial neural networks (ANNs) were conceptualized aspng mathematical models for the biological synapse (Rumelhart et al., 1986). ANNs are organized as a layered connected structure (see Fig. 2). Each layer consists of a number of simple computational units called neurons. The number of neurons in the input and output layer is dictated by the problem, but the number of neurons in the hidden layer that is responsible for capturing the nonlinearity is variable, and can be optimized for the problem at hand. Neurons in each layer communicate with the neurons in the following layer through variable weight connections where the learned information is stored. ANNs were trained using a supervised learning algorithm called “learning by back-propagation of errors” (Rumelhart et al., 1986).

Artificial neural network models were built to classify microarray data as PE or NT using the features obtained from the statistical analysis as discussed in section B. Models were built using the R package (RSNN). The learned information is stored as weight connections, but they do not reveal anything about the system being modeled. However, the powerful internal representation capability of neural networks was exploited to identify key features that were important for class separation. This was achieved by using a calliper randomization approach (Nair, 1997). Briefly, a calliper window of randomized input is presented to a trained model and the loss in prediction capability is used to score the features. The greater the loss in prediction, the more important are the features contained within the calliper window. Calliper windows of different sizes were used to delineate the features considered most important for recognition.

2.4. Functional analysis

Functional analysis of the features identified by the ANN-based calliper randomization method was done with the help of the Bioconductor packages clusterProfiler and ReactomePA, that uses Gene Ontology (GO) and the Reactome pathway database respectively (Yu et al., 2012; Yu and He, 2016).

3. Results and discussion

3.1. Multiple comparison of PE and NT

Exhaustive comparison of all probes in the data set GSE48424 were carried out between PE and NT in a between-class comparison. The samples from nsPE and sPE were combined into one class, PE, and compared with NT. The results of the comparison revealed that 2516 (6.13%) features had significantly different expression levels between PE and NT. This reveals a larger number of possible candidate biomarkers than originally reported (Textoris et al., 2013). Fig. 1 shows

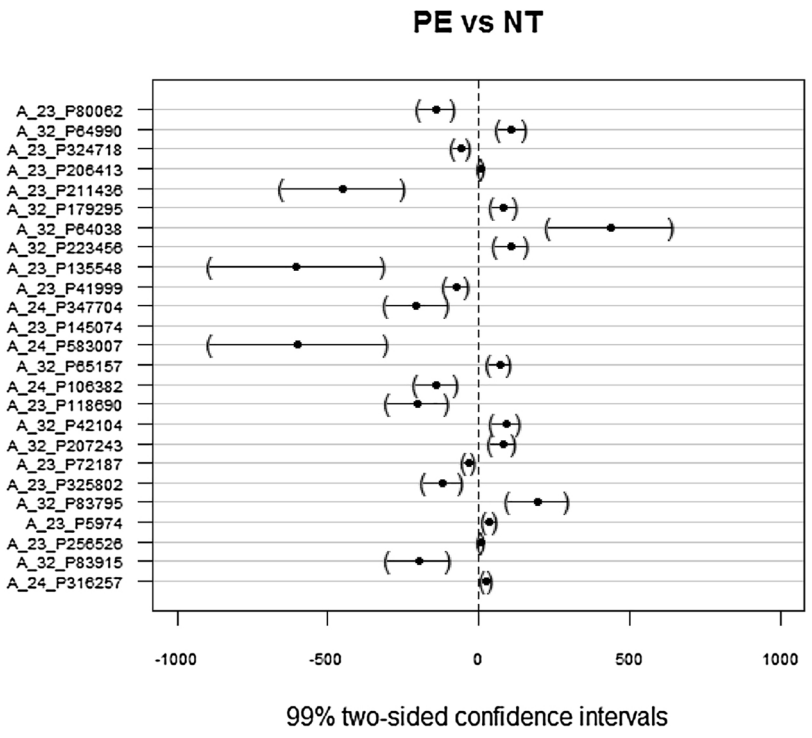

Fig. 1. Multiple comparison of expression levels for a subset of probes that showed significant difference between PE and NT samples.

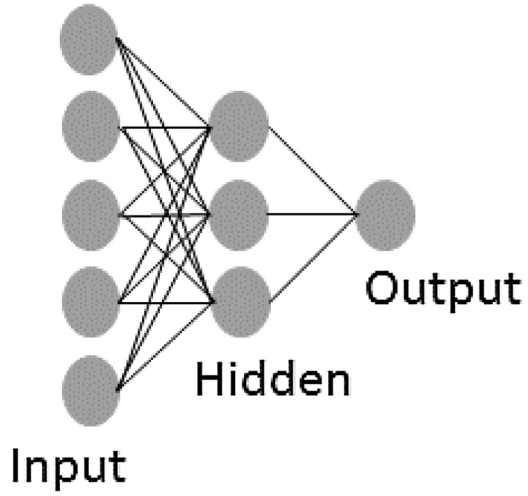

Fig. 2. A general three-layer fully connected artificial neural network. Expression values of the features identified using the multiple comparison method were presented as input and the output was either 1 (PE) or 0 (NT).

the multiple comparison between PE and NT. The plot shows the probes whose expression levels were significantly different between PE and NT samples. Significant difference is indicated by the fact that the intervals did not intersect the zero line. In the interest of brevity, a small subset that showed a significant difference between PE and NT is plotted. If the features obtained using multiple comparison are significantly different between the two classes, then they should be capable of class separation using standard clustering methods. Fig. 3 shows the hierarchical clustering and biplot-principal component analysis using the extracted features. Hierarchical clustering was done using Cluster 3.0 (Eisen et al., 1998). The features were able to cluster the two classes; however, there were five samples that were incorrectly classified (sPE10, nsPE5, sPE11, sPE12, and NT9). Biplot PCA analysis was used to understand the overall structure of the features. Positive and negative correlations in expression are taken into account in a PCA analysis. The length of the arrow is a representation of the variance in the data for that sample. The angle between the vectors represent the correlation between the samples. Positively correlated samples have zero degree angle between them, while those with negative correlation are separated by 180 degrees. Samples that are orthogonal have no correlation between them. The biplot analysis shows that the PE and NT samples, while clearly different from each other, have significant spread between them. It is noteworthy to point out that the PE samples that were incorrectly clustered in hierarchical clustering also tend to be more similar to NT samples, as inferred from the PCA analysis. Further, the biplot analysis also helps to understand the differences between and within PE and NT samples. It is clear that there is significant heterogeneity with PE and NT samples which was also noted by Textoris et al. (2013). This makes it a challenge in identifying unique biomarkers capable of correctly distinguishing PE and NT. Further, none of the protein serological biomarkers that were delineated using a multiomics approach (Liu et al., 2013) were part of the features identified. This further underscores the complexity and heterogeneity of the underlying disease and the challenge in identifying a specific biomarker as a common denominator.

With a view to further understand the heterogeneity and variation due to samples, within- and between-class comparisons were carried out after dividing the samples as discussed in material and methods section. Table 1 gives the percentage of features that were differentially expressed in a within- and between-class comparison. It is interesting to note that the NT samples were more homogeneous than the nsPE and sPE samples. In total, there were ~2200 features that were differentially expressed between nsPE and sPE. This clearly points to the fact that there is a significant heterogeneity within the PE samples, thus making the identification of unique biomarkers extremely challenging. It also adds to the complex interplay that exists within the transcriptome that makes the identification of biological signatures for the disease more difficult using simple statistical methods. A machine learning method like artificial neural networks that are capable of capturing complex correlations from the data may potentially be more useful in understanding the complexity associated with the transcriptome.

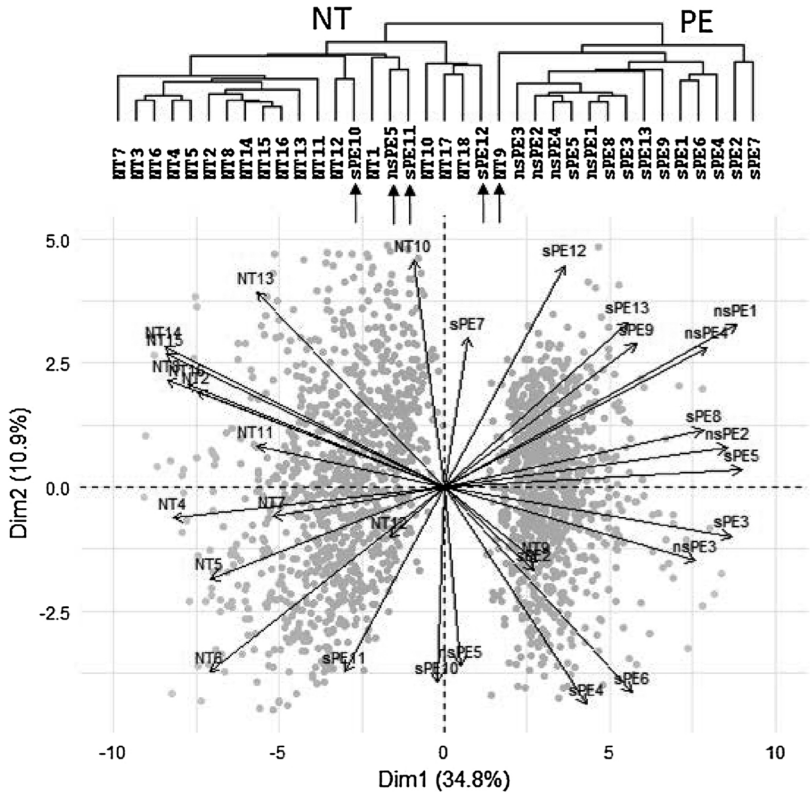

Fig. 3. Cluster analysis and biplot PCA using the extracted features. Features incorrectly clustered are pointed to by ↑.

Table 1
Within- and between-class comparison of differentially expressed features.

	NT1	NT2	sPE1	sPE2	nsPE1
NT2	65				
sPE1	277	1209			
sPE2	206	26	143		
nsPE1	1274	695	411	303	
nsPE2	1319	670	651	867	512

3.2. Artificial neural network-based analysis of the features identified from the GSE48424 data set

Using the features that were found to be significant in the multiple comparison analysis, a three-layer neural network model was built to distinguish PE from NT. All the 2516 features obtained from the statistical analysis were presented to the network as input with one hidden layer consisting of 15 neurons. The data set were partitioned into training (85%) and test data sets (15%). The test data sets were not presented to the neural net at the time of training and was used to evaluate the performance of the trained model. The network was capable of learning and classifying the features as belonging to PE or NT classes. Fig. 4 shows the progressive decrease in error as the network learned the patterns from the data. The neural network was clearly able to generalize the patterns as both the training and test error profiles were monotone decreasing. The ROC curves for training and testing data sets are shown in Fig. 5.

The network model was able to successfully classify all the test data to their respective classes indicating that the network was able to learn the patterns of gene expression that distinguished the two classes. However, a network model being a black-box model is incapable of providing information on which features are relevant in the classification process. It is noteworthy to mention that neural networks are capable of capturing second and higher order correlations from data.

With a view to capture the features that are considered important in the classification process by the neural network, the calliper randomization approach (Nair et al., 1995; Nair, 1997) first introduced by the author in the analysis of sequences was extended to this analysis. Noise was introduced into a calliper window of input features and then presented to the trained network model to assess the prediction capability of the model. The loss in prediction capability of the network model was used as a measure of the relative importance of the features contained in the window for classification. In effect, the approach helps in opening up the black box of the neural network model and provides an insight into the pattern recognition process of a machine learning model. Fig. 6 shows the calliper error associated when calliper windows of different sizes are randomized. The larger the error, the more relevant are the associated features. Features that emerged as relevant in the analysis were further analyzed for biological significance.

3.3. Functional analysis of features

In total, 1856 features were identified as relevant by the machine learning method. These were further filtered for annotation and probes querying the same gene. The resulting 1303 features were analyzed for biological relevance. The GO database was used to characterize these genes using the clusterProfiler package (Yu et al., 2012; Gene Ontology, 2015; Ashburner et al., 2000). The feature sets were analyzed for the molecular function, biological processes and cellular components. The molecular function analysis using GO revealed that the features belonged to four categories. These were ubiquitin-like protein ligase binding, ubiquitin protein ligase binding, protein transported activity, and phosphatidylinositol bisphosphate phosphatase activity. The biological process category analysis using GO revealed that the feature sets contained at least 23 different biological processes. The biological processes associated with the features are summarized in Fig. 7. Cellular component analysis of the identified features revealed that a disparate set of components were associated with the features. The cellular components associated with the features are summarized in Fig. 8. It is interesting that one of the categories of biological processes associated

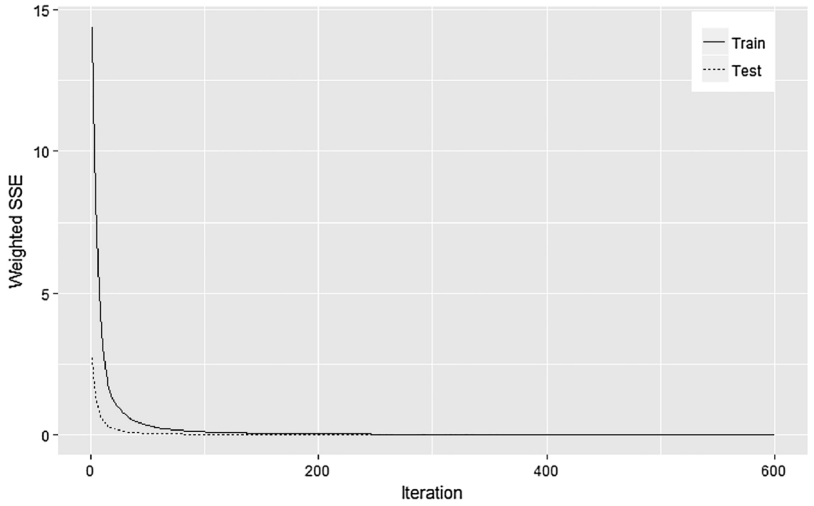

Fig. 4. Error profile of training and test data set.

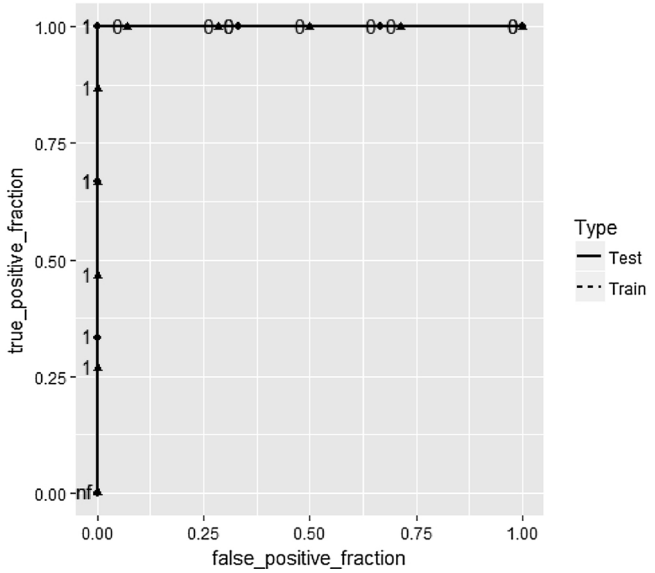

Fig. 5. ROC curve showing the performance of the neural network model.

with the features identified is vesicle organization. Vesicle organization may play an important role in the resolution of inflammation pathways (Perucci et al., 2017).

There are several possible mechanisms that may be responsible for the pathogenesis of preeclampsia (Young et al., 2010). While it is important to understand the association of the features with respect to the cellular components and biological processes they may be associated with, it would be crucial to delineate the pathways involved, and understand how they may be interrelated. The pathways associated with the features identified were delineated by querying the Reactome pathway knowledgebase using the ReactomePA Bioconductor package (Fabregat et al., 2016; Yu and He, 2016; Croft et al., 2014). The results of the analysis reveal that features identified are associated with several pathways that were already known to play an important role in preeclampsia. For instance, it has been found that the preimplantation factor (PIF) protein levels in placentas from pregnancies affected by preeclampsia or intra-uterine growth restriction had significantly lower levels of PIF. It is also known that PIFs' effect on placental apoptosis is mediated by TP53 (Moindjie et al., 2016). Several of the features identified have been associated with the TP53 pathway. Pathways are not isolated entities, and studies on complexities in biological systems have revealed that different pathways have synergistic effects on each other. UpSet is a convenient way to visualize the possible interaction/association between pathways (Lex et al., 2014).

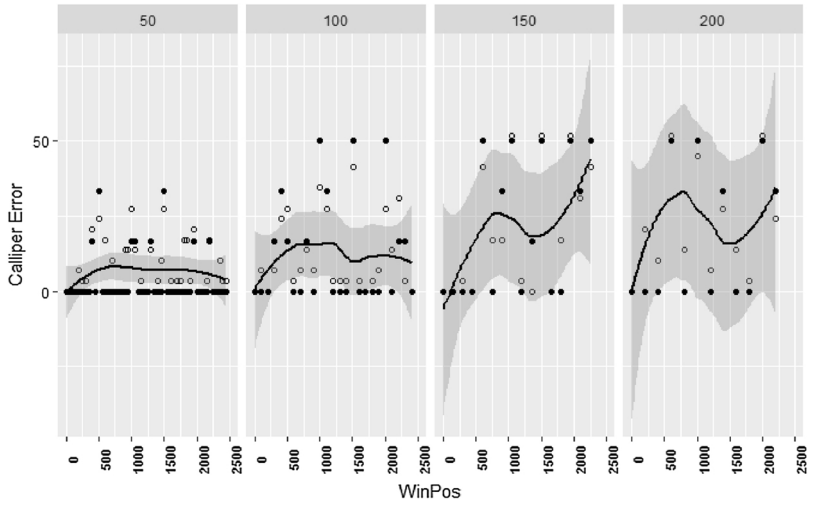

Fig. 6. Calliper errors for varying window sizes (50, 100, 150, and 200). Peaks correspond to regions important for the network in its internal representation.

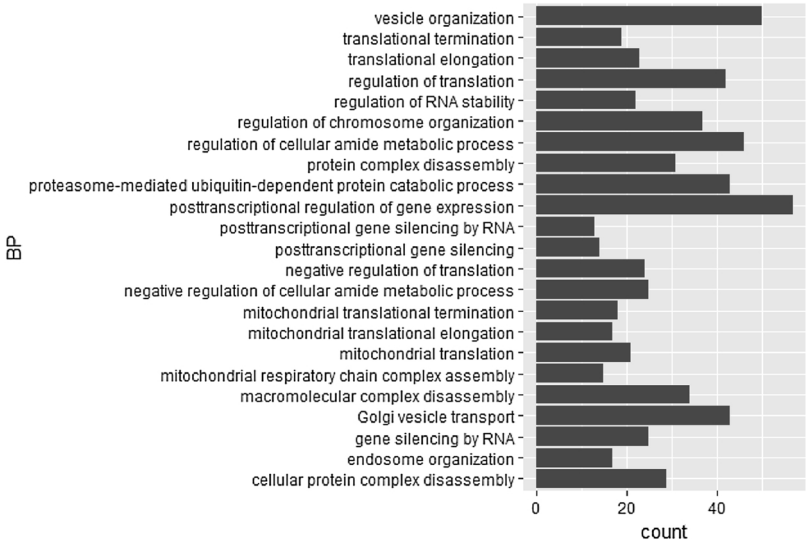

Fig. 7. Biological processes associated with features obtained by GO analysis.

Fig. 9 shows the pathways associated with the features and the different interactions. The bar graph depicts the number of genes involved. One of the pathways that stands out is the TRIF mediated TLR3/TLR4 signaling. Toll-like receptors (TLRs) are transmembrane proteins that constitute a family of pattern recognition receptors. These are shown to be involved in innate immunity that is responsible for distinguishing infectious non-self and non-infectious self (Medzhitov and Janeway, 2002). Innate immune response at maternal-fetal interface is crucial for the success of pregnancy (Koga et al., 2014).

4. Analysis of the VDAART transcriptome

With a view to gain an insight into the complexity of the molecular nature of preeclampsia, the VDAART transcriptome was analyzed specifically to understand heterogeneity and to determine if the features identified from the GSE48424 data set were capable of differentiating preeclampsia and normotensive cases in the VDAART GSE85307 data set (Mirzakhani et al., 2016).

The features identified from the analysis of the GSE48424 data set were compared to features in the VDAART data set (GSE85307) using BioMart (Durinck et al., 2009). All of the features did not have a one-to-one correspondence since the technologies used in generating the transcriptome were different. However, it was possible to map 2245 features, but since the features were not exactly the same, it was not possible to test the performance of the artificial neural network model built using features from the GSE48424 data set using mapped features from the VDAART data set. Thus, an entirely new neural network model was trained using the mapped features. It is noteworthy to point out that these were mapped features as opposed to features identified using feature selection methods. The neural network was only able to classify 60% of the samples. This further points to the heterogeneity associated with preeclampsia. Thus, it was necessary to identify and study the differential expression to delineate features from the VDAART data set.

4.1. Feature selection using multiple comparison and neural network modeling of VDAART transcriptome

Exhaustive analysis of the VDAART transcriptome (GSE85307) was done similarly using the statistical multiple comparison method previously described. The exhaustive analysis revealed that there were 262 features that were significantly different. Even from this data set, none of the features identified using a multiple comparison method were part of the serological protein markers identified by Liu et al. (2013). This again suggests that there is significant heterogeneity with the

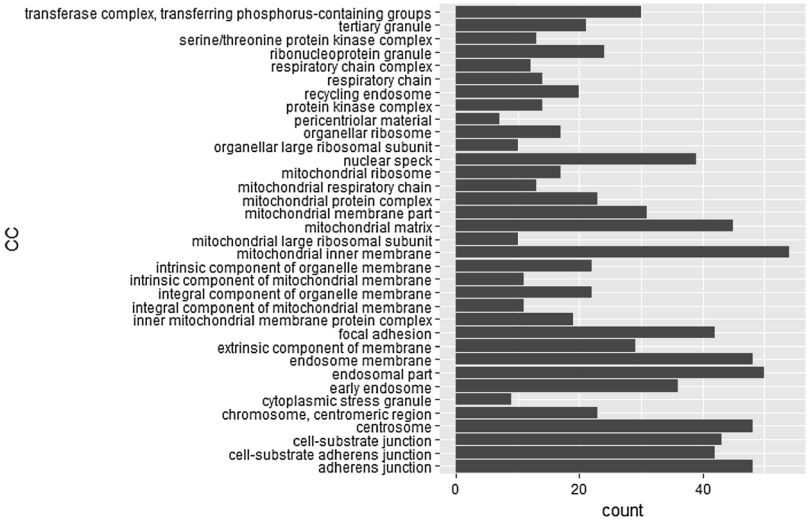

Fig. 8. Cellular components associated with the features obtained by GO analysis.

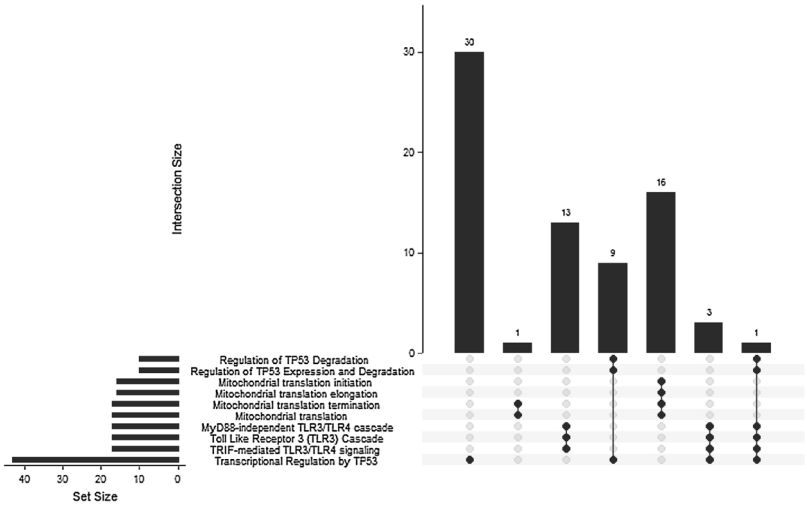

Fig. 9. Association between pathways obtained from reactome pathway knowledgebase.

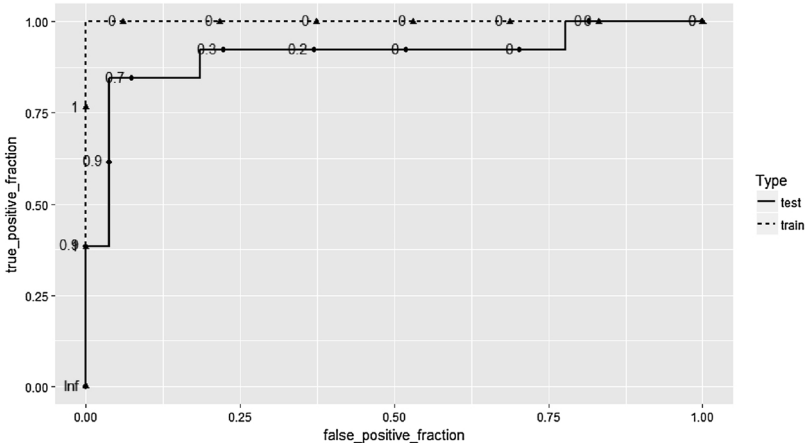

Fig. 10. ROC plot of the performance of the neural network model using the features from the VDAART data set.

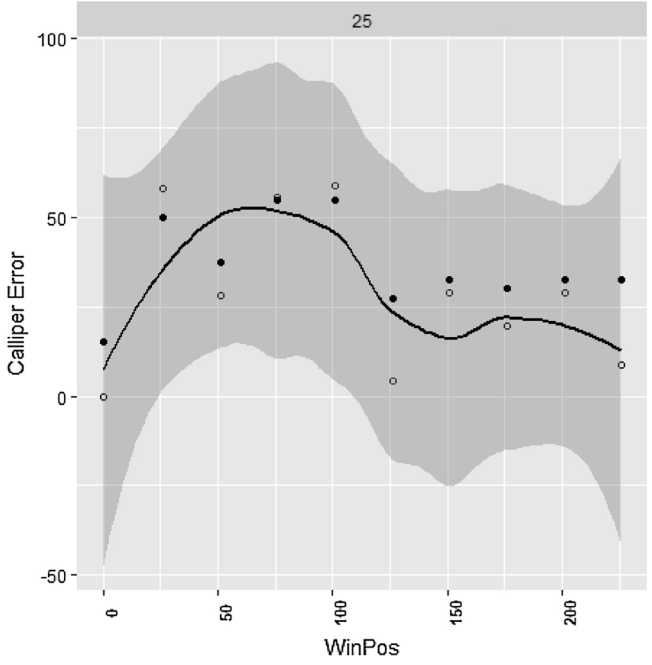

Fig. 11. Calliper errors for window of size 25. Peaks correspond to regions important for the network in its internal representation.

underlying molecular basis of the disease and the feature sets need to be analyzed to delineate complex correlations. Towards that, these 262 features were used to build a neural network model to distinguish preeclampsia from normotensive cases. The ROC curves for training and test data sets are shown in Fig. 10. While the network was able to accurately classify all the training data sets, not all of the test data sets were correctly classified (AUC = 0.908). The trained model was then used to delineate features that exhibited higher-order correlations by using the calliper-randomization approach. Fig. 11 shows the regions that were considered important by the model for pattern recognition.

To understand the inherent variability, the samples were further analyzed to delineate the within-class variability associated with the VDAART transcriptome data. The preeclampsia and normotensive data sets were analyzed separately by randomly dividing each of the two classes further into two groups. The two randomly divided groups that belonged to the same class (PE or NT) were compared with each other. Thus, by comparing two random groups of PE with each other, it was possible to determine the extent of variability within the PE class. The same was true for the NT class. The results showed that there were 347 features that were significantly different within PE and there were 397 features that were significantly different within NT.

Functional profile of the 261 features obtained using multiple comparison was done using gene ontology. Fig. 12 shows the different biological processes associated with the features. These results not only reveal commanalities with the processes identified by Mirzakhani et al. (2016) but also reveal that genes involved in reproductive processes

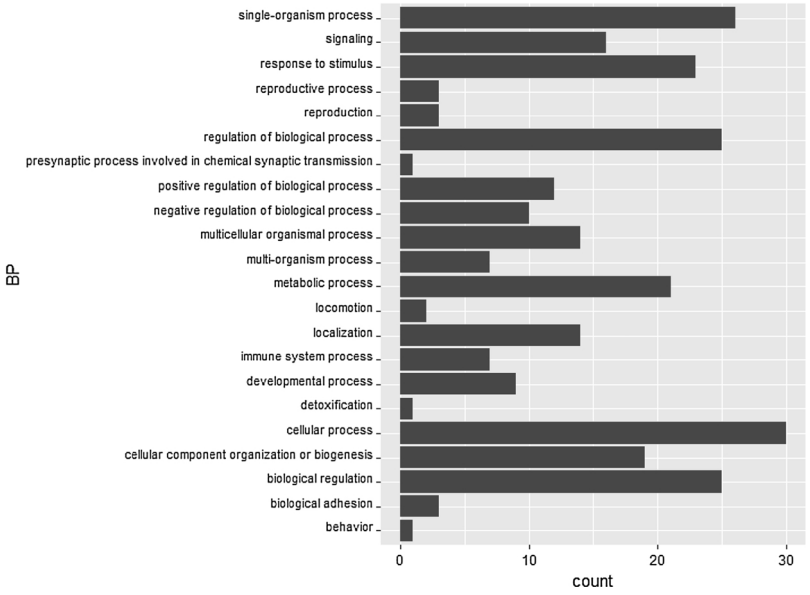

Fig. 12. Functional analysis of the features associated with the VDAART data sets.

were also part of the 261 features. The genes included PARK7, TSSK3, and DIAPH3. It is noteworthy to point out that PARK7 also known as DJ-1 is known to have a putative role in oxidative stress and hypoxic change and has been shown to have elevated expression in patients with severe preeclampsia (Kwon et al., 2013). CHFCR5 gene grouped as part of the immune system process has also been known to demonstrate differential methylation of its promoter and has been noted to be a gene of interest in preeclampsia (Ching et al., 2014).

5. Conclusion

Statistical and artificial neural network analysis of preeclampsia transcriptome (GSE48424 and GSE85307) revealed that there are features that are distinct to preeclampsia in each case. The within- and between-class comparison of the features from the data sets revealed that there are features within the same class that may show differential expression. This could account for the challenge in identifying unique features that are capable of distinguishing preeclampsia from normotensive cases. Further, using the ANN-based calliper randomization approach, it was possible to identify a subset of features that were critically important for imparting knowledge to the network model. While these were not experimentally verified, functional analysis using GO and the reactome pathway knowledgebase revealed that the features identified were associated with biological processes and pathways that were known to be important in disease development. However, the features that were identified from each data set were distinct and thus any conclusion about using a particular feature as a biomarker must be done with great caution and by taking into consideration the heterogeneity that exists within populations, as well as the plurality of molecular disruptions that could result in an underlying disease.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.compbiolchem.2018.05.011.

References

Ahmed, A., Ramma, W., 2015. Unravelling the theories of pre-eclampsia: are the protective pathways the new paradigm? Br. J. Pharmacol. 172 (6), 1574–1586. URL https://www.ncbi.nlm.nih.gov/pubmed/25303561.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat. Genet. 25 (1), 25–29. URL https://www.ncbi.nlm.nih.gov/pubmed/10802651.

Barton, J.R., Sibai, B.M., 2008. Prediction and prevention of recurrent preeclampsia. Obstet. Gynecol. 112 (2 Pt 1), 359–372. URL https://www.ncbi.nlm.nih.gov/pubmed/18669736.

Chaiworapongsa, T., Chaemsaithong, P., Korzeniewski, S.J., Yeo, L., Romero, R., 2014a. Pre-eclampsia: Part 2. Prediction, prevention and management. Nat. Rev. Nephrol. 10 (9), 531–540. URL https://www.ncbi.nlm.nih.gov/pubmed/25003612.

Chaiworapongsa, T., Chaemsaithong, P., Yeo, L., Romero, R., 2014b. Pre-eclampsia: Part 1. Current understanding of its pathophysiology. Nat. Rev. Nephrol. 10 (8), 466–480. URL https://www.ncbi.nlm.nih.gov/pubmed/25003615.

Cheng, S.B., Sharma, S., 2016. Preeclampsia and health risks later in life: an immunological link. Semin. Immunopathol. 38 (6), 699–708. URL https://www.ncbi.nlm.nih.gov/pubmed/27339196.

Ching, T., Song, M.A., Tiirikainen, M., Molnar, J., Berry, M., Towner, D., Garmire, L.X., 2014. Genome-wide hypermethylation coupled with promoter hypomethylation in the chorioamniotic membranes of early onset pre-eclampsia. Mol. Hum. Reprod. 20 (9), 885–904.

Cnossen, J.S., Mol, B.W., van Der Post, J.A., Bossuyt, P.M., Ter Riet, G., 2005. World Health Organization systematic review of screening tests for preeclampsia. Obstet. Gynecol. 105 (5 Pt 1), 1151–1152. author reply 1152. URL https://www.ncbi.nlm.nih.gov/pubmed/15863566.

Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R., Jassal, B., Jupe, S., Matthews, L., May, B., Palatnik, S., Rothfels, K., Shamovsky, V., Song, H., Williams, M., Birney, E., Hermjakob, H., Stein, L., D'Eustachio, P., 2014. The reactome pathway knowledgebase. Nucleic Acids Res. 42 (Database issue), D472–7. URL https://www.ncbi.nlm.nih.gov/pubmed/24243840.

Cuffe, J.S.M., Holland, O., Salomon, C., Rice, G.E., Perkins, A.V., 2017. Review: placental derived biomarkers of pregnancy disorders. Placenta 54, 104–110. URL https://www.ncbi.nlm.nih.gov/pubmed/28117143.

Duckitt, K., Harrington, D., 2005. Risk factors for pre-eclampsia at antenatal booking: systematic review of controlled studies. BMJ 330 (7491), 565. URL https://www.ncbi.nlm.nih.gov/pubmed/15743856.

Durinck, S., Spellman, P.T., Birney, E., Huber, W., 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRT. Nat. Protocols 4, 1184–1191.

Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D., 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U. S. A. 95 (25), 14863–14868. URL http://www.pnas.org/content/95/25/14863.abstract.

Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., Matthews, L., May, B., Milacic, M., Rothfels, K., Shamovsky, V., Webber, M., Weiser, J., Williams, M., Wu, G., Stein, L., Hermjakob, H., D'Eustachio, P., 2016. The reactome pathway knowledgebase. Nucleic Acids Res. 44 (D1), D481–7. URL https://www.ncbi.nlm.nih.gov/pubmed/26656494.

Gene Ontology, C., 2015. Gene ontology consortium: going forward. Nucleic Acids Res. 43 (Database issue), D1049–56. URL https://www.ncbi.nlm.nih.gov/pubmed/25428369.

Koga, K., Izumi, G., Mor, G., Fujii, T., Osuga, Y., 2014. Toll-like receptors at the maternal–fetal interface in normal pregnancy and pregnancy complications. Am. J. Reprod. Immunol. 72 (2), 192–205.

Kwon, H.-S., Hwang, H.-S., Sohn, I.-S., Park, S.-H., 2013. Expression of DJ-1 proteins in placentas from women with severe preeclampsia. Eur. J. Obstet. Gynecol. Reprod. Biol. 168 (1), 40–44.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., Pfister, H., 2014. Upset: visualization of intersecting sets. IEEE Trans. Vis. Comput. Graph. 20 (12), 1983–1992 URL < Go to ISI > ://WOS:000344991700043.

Liu, L.Y., Yang, T., Ji, J., Wen, Q., Morgan, A.A., Jin, B., Chen, G., Lyell, D.J., Stevenson, D.K., Ling, Xuefeng, B., Butte, A.J., 2013 Nov. Integrating multiple 'omics' analyses identifies serological protein biomarkers for preeclampsia. BMC Med. 11 (1), 236.

MacKay, A.P., Berg, C.J., Atrash, H.K., 2001. Pregnancy-related mortality from pre-eclampsia and eclampsia. Obstet. Gynecol. 97 (4), 533–538. URL https://www.ncbi.nlm.nih.gov/pubmed/11275024.

McElrath, T.F., Lim, K.H., Pare, E., Rich-Edwards, J., Pucci, D., Troisi, R., Parry, S., 2012. Longitudinal evaluation of predictive value for preeclampsia of circulating angiogenic factors through pregnancy. Am. J. Obstet. Gynecol. 207 (5), 407. e1-7. URL https://www.ncbi.nlm.nih.gov/pubmed/22981320.

Medzhitov, R., Janeway, C.A.J., 2002. Decoding the patterns of self and nonself by the innate immune system. Science 296 (5566), 298–300.

Mirzakhani, H., Litonjua, A.A., McElrath, T.F., O'Connor, G., Lee-Parritz, A., Iverson, R., Macones, G., Strunk, R.C., Bacharier, L.B., Zeiger, R., Hollis, B.W., Handy, D.E., Sharma, A., Laranjo, N., Carey, V., Qiu, W., Santolini, M., Liu, S., Chhabra, D., Enquobahrie, D.A., Williams, M.A., Loscalzo, J., Weiss, S.T., 2016. Early pregnancy vitamin D status and risk of preeclampsia. J. Clin. Invest 126 (12), 4702–4715.

Moindjie, H., Santos, E.D., Gouesse, R.J., Swierkowski-Blanchard, N., Serazin, V., Barnea, E.R., Vialard, F., Dieudonne, M.N., 2016. Preimplantation factor is an anti-apoptotic effector in human trophoblasts involving p53 signaling pathway. Cell Death Dis. 7 (12), e2504. URL https://www.ncbi.nlm.nih.gov/pubmed/27906186.

Nair, T.M., 1997. Calliper randomization: an artificial neural network based analysis of *E. coli* ribosome binding sites. J. Biomol. Struct. Dyn. 15 (3), 611–617. URL https://www.ncbi.nlm.nih.gov/pubmed/9440007.

Nair, T.M., Tambe, S.S., Kulkarni, B.D., 1995. Analysis of transcription control signals using artificial neural networks. Comput. Appl. Biosci. 11 (3), 293–300. URL https://www.ncbi.nlm.nih.gov/pubmed/7583698.

Ngoc, N.T., Merialdi, M., Abdel-Aleem, H., Carroli, G., Purwar, M., Zavaleta, N., Campodonico, L., Ali, M.M., Hofmeyr, G.J., Mathai, M., Lincetto, O., Villar, J., 2006. Causes of stillbirths and early neonatal deaths: data from 7993 pregnancies in six developing countries. Bull. World Health Org. 84 (9), 699–705. URL https://www.ncbi.nlm.nih.gov/pubmed/17128339.

Perucci, L.O., Correa, M.D., Dusse, L.M., Gomes, K.B., Sousa, L.P., 2017. Resolution of inflammation pathways in preeclampsia – a narrative review. Immunol. Res.

Rademacher, T.W., Gumaa, K., Scioscia, M., 2007. Preeclampsia, insulin signalling and immunological dysfunction: a fetal, maternal or placental disorder? J. Reprod. Immunol. 76 (1–2), 78–84. URL https://www.ncbi.nlm.nih.gov/pubmed/17537518.

Redman, C.W., Sargent, I.L., 2010. Immunology of pre-eclampsia. Am. J. Reprod. Immunol. 63 (6), 534–543. URL https://www.ncbi.nlm.nih.gov/pubmed/20331588.

Roberts, J.M., Hubel, C.A., 2009. The two stage model of preeclampsia: variations on the theme. Placenta 30 (Suppl. A), S32–7. URL https://www.ncbi.nlm.nih.gov/pubmed/19070896.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323 (6088), 533–536. http://dx.doi.org/10.1038/323533a0.

Textoris, J., Ivorra, D., Ben Amara, A., Sabatier, F., Menard, J.P., Heckenroth, H., Bretelle, F., Mege, J.L., 2013. Evaluation of current and new biomarkers in severe pre-eclampsia: a microarray approach reveals the VSIG4 gene as a potential blood biomarker. PLoS ONE 8 (12), e82638. URL https://www.ncbi.nlm.nih.gov/pubmed/24349325.

Thuny, F., Textoris, J., Amara, A.B., Filali, A.E., Capo, C., Habib, G., Raoult, D., Mege, J.L., 2012. The gene expression analysis of blood reveals S100A11 AND AQP9 as potential biomarkers of infective endocarditis. PLoS ONE 7 (2), e31490. URL https://www.ncbi.nlm.nih.gov/pubmed/22319637.

Tubbergen, P., Lachmeijer, A.M., Althuisius, S.M., Vlak, M.E., van Geijn, H.P., Dekker, G.A., 1999. Change in paternity: a risk factor for preeclampsia in multiparous women? J. Reprod. Immunol. 45 (1), 81–88. URL https://www.ncbi.nlm.nih.gov/pubmed/10660264.

Westfall, P.H., 1997. Multiple testing of general contrasts using logical constraints and correlations. J. Am. Stat. Assoc. 92 (437), 299–306.

Widmer, M., Cuesta, C., Khan, K.S., Conde-Agudelo, A., Carroli, G., Fusey, S., Karumanchi, S.A., Lapaire, O., Lumbiganon, P., Sequeira, E., Zavaleta, N., Frusca, T., Gulmezoglu, A.M., Lindheimer, M.D., 2015. Accuracy of angiogenic biomarkers at 20 weeks' gestation in predicting the risk of pre-eclampsia: a WHO multicentre study. Pregnancy Hypertens. 5 (4), 330–338. URL https://www.ncbi.nlm.nih.gov/pubmed/26597750.

Young, B.C., Levine, R.J., Karumanchi, S.A., 2010. Pathogenesis of preeclampsia. Annu. Rev. Pathol. 5, 173–192. URL https://www.ncbi.nlm.nih.gov/pubmed/20078220.

Yu, G., He, Q.Y., 2016. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol. Biosyst. 12 (2), 477–479. URL https://www.ncbi.nlm.nih.gov/pubmed/26661513.

Yu, G., Wang, L.G., Han, Y., He, Q.Y., 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics 16 (5), 284–287. URL https://www.ncbi.nlm.nih.gov/pubmed/22455463.