(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 December 2013 (19.12.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/188686 A2

(51) International Patent Classification:
G01N 33/53 (2006.01)

(21) International Application Number:
PCT/US20 13/045709

(22) International Filing Date:
13 June 2013 (13.06.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
P1200368 15 June 2012 (15.06.2012) HU
61/699,193 10 September 2012 (10.09.2012) US

(71) Applicant: **WAYNE STATE UNIVERSITY** [US/US]; 5057 Woodward, Detroit, MI 48202 (US).

(72) Inventors: **TARCA, Adi,** L.; 49225 Rockerfeller Drive, Canton, MI 48188 (US). **THAN, Nandor;** 1160 Fairholme Road, Grosse Pointe Woods, MI 48236 (US). **JUHASZ, Gabor;** 1/c Pazmany Peter Setany, H-l 117 Budapest (HU). **KEKESI, Adrienna, Katalin;** 1/c Pazmany Peter Setany, H-l 117 Budapest (HU). **MEIRI, Hamutal;** 12 Barasani St., Suite 48, 69121 Tel Aviv (IL). **PAPP, Zoltan;** 4 Kutvolgyi Street, H-l 125 Budapest (HU). **ROMERO, Roberto;** 3990 John R., 4 Brush, Office 4734, Detroit, MI 48201 (US).

(74) Agents: **WINGER, C , Rachal** et al; K&L Gates, LLP, 925 Fourth Avenue, Suite 2900, Seattle, WA 98104-1 158 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available):* AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available):* ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

[Continued on next page]

(54) Title: BIOMARKER TEST FOR PREDICTION OR EARLY DETECTION OF PREECLAMPSIA AND/OR HELLP SYNDROME

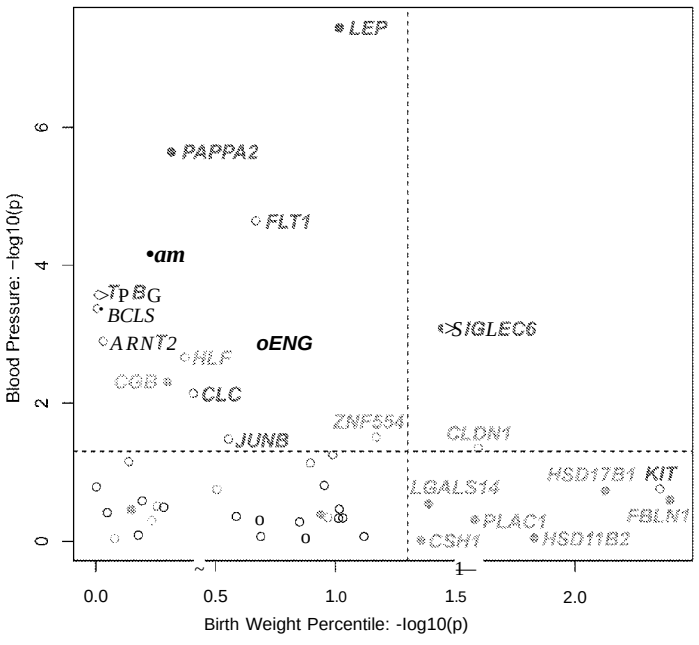

Figure 2C

(57) Abstract: Disclosed are specific biomarkers that allow for early testing of preeclampsia/HELLP syndrome. Thus, a method is provided predicting preeclampsia in a pregnant woman. Also disclosed is a kit comprising means for assaying a sample from a pregnant woman for the concentrations of the specific biomarkers.

WO 2013/188686 A2

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

— *with sequence listing part of description (Rule 5.2(a))*

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

BIOMARKER TEST FOR PREDICTION OR EARLY DETECTION OF PREECLAMPSIA AND/OR HELLP SYNDROME

FIELD OF THE DISCLOSURE

[0001] The present disclosure provides a test which can be used to predict and/or detect preeclampsia and/or Haemolysis, Elevated Liver enzymes and Low Platelets (HELLP) syndrome in pregnant women. More specifically, the disclosure provides a panel of biomarkers that can be used for early prediction and/or detection of preeclampsia and/or HELLP syndrome and may also allow the prediction and detection of closely related complications of pregnancy in early gestation such as including but not limited to implantation failure, and threatened and spontaneous miscarriage.

BACKGROUND OF THE DISCLOSURE

[0002] Preeclampsia is a syndrome defined by pregnancy-induced hypertension and proteinuria, which can lead to eclampsia (convulsions), and other serious maternal and/or fetal complications. Preeclampsia is originated in early gestation from the failure of implantation mechanisms and/or placental development, and is thus closely related to complications of pregnancy in early gestation such as including but not limited to implantation failure, and threatened and spontaneous miscarriage, Preeclampsia affects approximately 5-7% of pregnant women (approximately 8,370,000 pregnant women worldwide per year) and is a major cause of maternal and perinatal mortality. Furthermore, women with preeclampsia have an 8-fold higher risk of cardiovascular death later in their life, and offspring born from pregnancies affected by preeclampsia have an increased risk of metabolic and cardiovascular disease and mortality later in life.

[0003] The present diagnostic criteria for preeclampsia set by the United States National High Blood Pressure Education Program Working Group on High Blood Pressure in Pregnancy include new-onset hypertension coupled with proteinuria that develops after 20 weeks of gestation in women with previously normal blood pressures. These criteria further define preeclampsia as systolic or diastolic blood pressures of >140 and/or >90 mmHg, respectively, measured at two or more different time points, at least 4 hours (h) but not more than 1 week apart, as well as proteinuria of >300 mg

protein in a 24 h urine sample, or two random urine specimens obtained at least 4 h but not more than 1 week apart containing ≥1+ protein on a dipstick.

[0004] Based on the timing of the clinical manifestation, preeclampsia has been historically classified into different sub-forms, such as "term" (>37 weeks) and "preterm" (<37 weeks) or by using an alternative terminology "late-onset" and "early-onset" preeclampsia. The latter classification has not been uniformly used, but different studies have employed a range of gestational age cutoffs varying between 28 and 35 weeks for the distinction between early-onset and late-onset preeclampsia. Recently, it has been suggested to define 34 weeks as the gestational age cutoff between these two forms. It is important to note that preeclampsia may occur intrapartum or postpartum; thus, monitoring and evaluating the symptoms of preeclampsia should be continued during the postpartum period.

[0005] In 1954, it was first reported that preeclampsia may be associated with haemolysis, abnormal liver function and thrombocytopenia. Initially accepted to be a severe variant of preeclampsia, this group of symptoms later was suggested to constitute a separate clinical entity termed Haemolysis, Elevated Liver enzymes and Low Platelets (HELLP) syndrome. Supporting the idea that HELLP syndrome is a distinct condition, up to 20% of HELLP syndrome patients do not develop hypertension, 5-15% have minimal or no proteinuria and 15% show neither hypertension nor proteinuria. Moreover, laboratory findings in HELLP syndrome rarely correlate with the severity of hypertension or proteinuria.

[0006] In addition to the medical complications suffered by mothers and risks to the offspring, preeclampsia and HELLP syndrome cause approximately $7 billion in healthcare costs in the United States annually. Accordingly, there have been many attempts to provide a reliable predictive test for preeclampsia/HELLP syndrome. Previous attempts have involved assays for the concentrations of circulating biochemical markers in maternal blood but to date, the scientific literature on these approaches have been contradictory and inconclusive. There is a need in the art for new and improved methods of predicting and diagnosing these conditions.

SUMMARY OF THE DISCLOSURE

[0007] The present disclosure provides biomarker combinations that allow for the prediction and/or early detection of preeclampsia and/or HELLP syndrome, and may also allow the prediction and detection of closely related complications of pregnancy in early gestation such as but not limited to implantation failure, and threatened and spontaneous miscarriage.

[0008] One embodiment includes a method for assessing the presence or risk of preeclampsia in a female to determine the need for a treatment regimen comprising: determining levels of one or more of complement factor B; gelsolin isoform a precursor; hornerin, fetuin B; hemopexin precursor; apolipoprotein H precursor; fms-related tyrosine kinase 1; hydroxysteroid (17-β) dehydrogenase 1; leptin; lectin galactoside-binding soluble 14; pappalysin 2 or placenta-specific 1 in a biological sample obtained from the female; generating a dataset based on the determined levels; assessing the presence or risk of developing preeclampsia in the female based on the dataset; and determining a treatment regimen based on the assessed presence or risk.

[0009] Another embodiment includes a method for assessing the presence or risk of preeclampsia in a female to determine the need for a treatment regimen comprising: determining levels of one or more of complement factor B, hornerin, hemopexin precursor, hydroxysteroid (17-β) dehydrogenase 1, lectin galactoside-binding soluble 14 or pappalysin 2 in a biological sample obtained from the female; generating a dataset based on the determined levels; assessing the presence or risk of developing preeclampsia in the female based on the dataset; and determining a treatment regimen based on the assessed presence or risk.

[0010] In another embodiment, the assaying is performed for the levels of at least 3 biomarkers.

[001 1] In another embodiment, the sample is a blood sample.

[0012] In another embodiment the sample is other body fluid, secretion or excretion (such as but not limited to cervicovaginal fluid, saliva, or urine) sample.

[0013] In another embodiment the sample is an amniotic fluid sample.

[0014] In another embodiment the sample is fetal cells obtained invasively or non-invasive^.

[0015] In another embodiment, the sample is a placental sample.

[0016] In another embodiment, the biological sample is obtained before the 20th week of pregnancy, before the 19th week of pregnancy, before the 18th week of pregnancy, before the 17th week of pregnancy, before the 16th week of pregnancy, before the 15th week of pregnancy, before the 14th week of pregnancy, before the 13th week of pregnancy, before the 12th week of pregnancy, before the 11th week of pregnancy, before the 10th week of pregnancy, before the 9th week of pregnancy, before the 8th week of pregnancy, before the 7th week of pregnancy, before the 6th week of pregnancy, or after delivery.

[0017] In another embodiment, the treatment regimen is a therapeutic intervention.

[0018] In another embodiment, the therapeutic intervention prevents or reduces symptoms of preeclampsia before the symptoms manifest in the female and/or fetus.

[0019] Another embodiment includes a kit for assessing the presence or risk of preeclampsia in a female to determine the need for a treatment regimen comprising: detection mechanisms for determining levels of one or more of complement factor B; gelsolin isoform a precursor; hornerin, fetuin B; hemopexin precursor; apolipoprotein H precursor; fms-related tyrosine kinase 1; hydroxysteroid (17-β) dehydrogenase 1; leptin; lectin galactoside-binding soluble 14; pappalysin 2 or placenta-specific 1 in a biological sample obtained from the female; instructions how to (i) generate a dataset based on the determined levels; (ii) assess the presence or risk of developing preeclampsia in the female based on the dataset; and (iii) determine a treatment regimen based on the assessed presence or risk.

[0020] Another embodiment includes a kit for assessing the presence or risk of preeclampsia in a female to determine the need for a treatment regimen comprising: detection mechanisms for determining levels of one or more of complement factor B, hornerin, hemopexin precursor, hydroxysteroid (17-β) dehydrogenase 1, lectin galactoside-binding soluble 14 or pappalysin 2 in a biological sample obtained from the female; instructions how to (i) generate a dataset based on the determined levels;

(ii) assess the presence or risk of developing preeclampsia in the female based on the dataset; and (iii) determine a treatment regimen based on the assessed presence or risk.

[0021] In another embodiment, the kit includes detection mechanisms for at least three markers.

[0022] In another embodiment, the kit includes detection mechanisms for all markers described above.

[0023] In another embodiment, the methods and kits measure levels of at least one marker described in the figures and examples described herein.

BRIEF DESCRIPTION OF THE FIGURES

[0024] Figure 1. Figure 1 shows a genomic map of differentially expressed genes in preeclampsia. Circos visualization shows Chromosomes with solid lines in the inner circle. Curved lines connect the genomic coordinates of genes and transcription regulatory genes that are significantly correlated. Significance was determined by fitting a linear model between the expression level of gene and transcription regulatory gene pairs in all samples while controlling for FDR at 5%. Curves represent positive and negative correlations. The second circle shows the genomic location of genes with predominant placental expression (PPE) (black lines: non differentially expressed; grey lines: up- or down-regulated). The third and fourth circles show the locations of differentially expressed transcription regulatory genes and non-regulatory genes, respectively with inward-oriented bars (down-regulated) and outward-oriented bars (up-regulated). The height of the bars in the third and fourth circles represents the magnitude of gene expression changes.

[0025] Figure 2. Figure 2A-2C shows that two gene modules in preeclampsia are enriched in PPE genes and are associated with mean arterial blood pressure and birth weight percentile. 2A) Gene modules identified from WGCNA analysis of microarray data. Dysregulated placental gene expression could be characterized by five gene modules within the 1,409 differentially expressed genes in preeclampsia, marked with different shades of grey. The height plotted on the y-axis represents the distance metric (1-TOM) used by WGCNA. Out of 38 PPE genes (black vertical lines), 33 belonged to the grey-scale modules (lighter grey; n=22 and darker grey; n=1 1). These

two modules were also enriched in up-regulated and down-regulated genes marked under the modules with grey or black lines, respectively. 2B) Hierarchical clustering of qRT-PCR data obtained with 100 samples and selected 47 genes. Genes from light and dark grey modules clustered together in the validation sample-set. Importantly, 34 out of 60 samples from women with preeclampsia clustered tightly together. Pearson correlation was used for distance, and average for linkage. Samples (column leafs) were shaded according to patient groups and maturity status. 2C) Association of gene expression with mean arterial blood pressure and birth weight percentile. For each gene, a linear model was fitted (expression-blood pressure + birth weight percentile + gender + maturity status). The significance p-values (-logio of) for the two coefficients (blood pressure and birth weight percentile) were plotted for all 47 genes. Genes were shaded according to module membership (except black color for those not differentially expressed on the microarray). Filled circles represent PPE genes, dashed lines the significance threshold at p=0.05. Note that 7 out of 9 genes related to birth weight percentile are from the light grey module, while 10 out of 15 genes related to blood pressure are from the dark grey module.

[0026] Figure 3. Figure 3A-3C shows that the expression of dark grey module genes changes in the same direction in preeclampsia subgroups. 4A-B) In each barplot, the left and right panels show significant differences ("*") in preterm and term preeclampsia samples, respectively. Gene expression 4A) and protein immunostainings 4B) show similar patterns in sub-groups of preeclampsia. When the change with preeclampsia in the preterm samples was significantly different than the change with preeclampsia in the term samples, a "+" sign indicates this interaction. Semiquantitative immunoscorings for four proteins 4B) validated gene expression data. 4C) Representative images from the four immunostainings. The same placenta from a preterm control (left, 29 weeks) and from a patient with preterm preeclampsia with SGA (right, 31 weeks) is shown for the four immunostainings (40x magnifications).

[0027] Figure 4. Figure 4A-4UU show gene comparisons for *ARNT2; BCL3; BCL6; BTG2; CDKN1A; CGB3; CLC; CLDN1; CRH; CSH1; CYP19A1; DUSP1; ENG; ERVFRDE1; ERVWE1; ESRRG; FBLN1; FLT1; GATA2; GCM1; GH2; HLF; HSD11B2; HSD17B1; IKBKB; INSL4; JUNB; KIT; LEP; LGALS13; LGALS14; LGALS16;*

LGALS17A; MAPK13; NANOG; PAPPA; PAPPA2; PGF; PLAC1; POU5F1; SIGLEC6; TEAD3; TFAM; TFAP2A; TPBG; VDR; and *ZNF554* respectively.

DETAILED DESCRIPTION

[0028] As described above, there have been several attempts carried out over the past few years to develop and validate biomarkers for the early prediction of preeclampsia and/or HELLP syndrome; however, their results were not satisfactory. A possible reason for this is that the early diagnosis of syndromes with a heterogeneous molecular background cannot be solved with the utilization of only one or two biomarker molecules.

[0029] The present disclosure describes the use of a multidisciplinary systems biological approach which led to the identification of a unique gene-, protein- and hormone biomarker panel, which are repeatedly detectable in the 7-9th weeks of gestation. This aim of the present disclosure was achieved by the inclusion of several high-dimensional techniques: 1) whole genome transcriptomics of the placenta; 2) high-throughput qRT-PCR expressional profiling of the placenta; 3) high-throughput tissue microarray protein expression profiling of the placenta; 4) neural network analysis to select best combinations of candidate biomarkers to predict blood pressure and birth weight; 5) linear discriminant analysis model to provide sensitivity and specificity measures for preeclampsia prediction and 6) 2D-DIGE proteomics of maternal sera in early-pregnancy.

[0030] Whole-genome transcriptomics study of 17 placentas identified placenta- and pregnancy-specific genes differentially expressed in the placenta in preeclampsia and/or HELLP syndrome. The products of this set of genes can be identified in the maternal serum in large amounts in pregnancy, and thus, their expression and differential regulation is pregnancy-specific. However, their changes may not only be specific for preeclampsia, but also for other obstetrical syndromes.

[0031] High-throughput qRT-PCR expressional profiling of 100 placentas validated selected putative preeclampsia biomarkers at the RNA level.

[0032] High-throughput tissue microarray protein expression profiling of 100 placentas validated selected putative preeclampsia biomarkers at the protein level.

[0033] Neural network analysis supported the selection of the best combinations of

putative preeclampsia biomarker genes, which expression can predict blood pressure and birth weight.

[0034] The Linear Discriminant Analysis showed that the average sensitivity and specificity of transcriptomic biomarkers for the detection of preeclampsia was 91.5% and 75%, respectively.

[0035] 2D-DIGE proteomics of maternal sera in early-pregnancy revealed that the proteome of first trimester maternal blood in women with early-onset or late-onset preeclampsia differs from that of normal pregnant women, and these differences are partially different in the two subtypes of preeclampsia. Although these inflammatory and/or metabolic markers are not specific for pregnancy, they can differentiate between the two subtypes of preeclampsia. The combination of these transcriptomic and proteomic biomarker candidates resulted in a panel of molecules, which can detect preeclampsia-specific changes in maternal blood, and can also differentiate between the different subtypes of preeclampsia.

A number of methods for obtaining expression data can be used singly or in combination for determining expression patterns and profiles in the context of the present disclosure. For example, DNA and RNA expression patterns can be evaluated by northern analysis, PCR, RT-PCR, quantitative real-time RT-PCR analysis with TaqMan assays, FRET detection, monitoring one or more molecular beacon, hybridization to an oligonucleotide array, hybridization to a cDNA array, hybridization to a polynucleotide array, hybridization to a liquid microarray, hybridization to a microelectric array, molecular beacons, cDNA sequencing, clone hybridization, cDNA fragment fingerprinting, serial analysis of gene expression (SAGE), subtractive hybridization, differential display and/or differential screening.

[0036] Gene expression changes can be related to epigenetic variations (e.g. DNA methylation). Epigenetic regulation mechanisms do not involve a change to the DNA sequence. Instead, epigenetic variations include covalent modification of DNA, RNA, and the proteins associated with DNA. These in turn can result in changes to the conformation of DNA and accessibility of regulators to the DNA. Such changes cannot be identified simply by gene sequencing. Janssen, B.G. et al., Particle and Fibre Toxicology, 10:22 (2013) studied methylation in placental tissue using methods published by Tabish, A.M. et al., *PLoS ONE* 2012, 7:e34674 and by Godderis, L. et al.,

Epigenomics 4:269-277 (201 2). MS-MLPA (Methylation-specific Multiplex ligation-dependent probe amplification) can be used to study methylation status of specific genes, for example in Proctor, M. et al., Clin. Chem. 52:1 276-1 283 (2006). Materials and methods for MS-MLPA as used in published studies can be obtained from MRC-Holland, Amsterdam, The Netherlands. Additional methods are reviewed and compared in Shen, L. et al., Curr. Opin. Clin. Nutr. Metab. Care. 10:576-81 (2007); Gu H et al., Nature Methods 7:1 33-1 38 (201 0); Bock C et al., Nature Biotech. 28:1 106-1114 (201 0); Harris RA et al., Nature Biotech. 28:1 097-1 105 (2010).

[0037] Protein expression patterns can be evaluated using any method that provides a quantitative measure and is suitable for evaluation of multiple markers extracted from samples. Exemplary methods include: ELISA sandwich assays, mass spectrometric detection, calorimetric assays, binding to a protein array (e.g., antibody array), or fluorescent activated cell sorting (FACS). Approaches can use labeled affinity reagents (e.g., antibodies, small molecules, etc.) that recognize epitopes of one or more protein products in an ELISA, antibody array, or FACS screen.

[0038] Typically, the term high-throughput refers to a format that performs at least about 100 assays, or at least about 500 assays, or at least about 1000 assays, or at least about 5000 assays, or at least about 10,000 assays, or more per day. When enumerating assays, either the number of samples or the number of protein markers assayed can be considered. Generally high-throughput expression analysis methods involve a logical or physical array of either the subject samples, or the protein markers, or both. Appropriate array formats include both liquid and solid phase arrays. For example, assays employing liquid phase arrays, e.g., for hybridization of nucleic acids, binding of antibodies or other receptors to ligand, etc., can be performed in multiwell or microtiter plates. Microtiter plates with 96, 384, or 1536 wells are widely available, and even higher numbers of wells, e.g., 3456 and 9600 can be used. In general, the choice of microtiter plates is determined by the methods and equipment, e.g., robotic handling and loading systems, used for sample preparation and analysis.

[0039] Alternatively, a variety of solid phase arrays can also be employed to determine expression patterns. Exemplary formats include membrane or filter arrays (e.g., nitrocellulose, nylon), pin arrays, and bead arrays (e.g., in a liquid "slurry"). Essentially any solid support capable of withstanding the reagents and conditions

necessary for performing the particular expression assay can be utilized. For example, functionalized glass, silicon, silicon dioxide, modified silicon, any of a variety of polymers, such as (poly)tetrafluoroethylene, (poly)vinylidenedifluoride, polystyrene, polycarbonate, or combinations thereof can all serve as the substrate for a solid phase array.

[0040] In one embodiment, arrays can include "chips" composed, e.g., of one of the above-specified materials. Polynucleotide probes, e.g., RNA or DNA, such as cDNA, synthetic oligonucleotides, and the like, or binding proteins such as antibodies or antigen-binding fragments or derivatives thereof, that specifically interact with expression products of individual components of the candidate library are affixed to the chip in a logically ordered manner, i.e., in an array. In addition, any molecule with a specific affinity for either the sense or anti-sense sequence of the marker nucleotide sequence (depending on the design of the sample labeling), can be fixed to the array surface without loss of specific affinity for the marker and can be obtained and produced for array production, for example, proteins that specifically recognize the specific nucleic acid sequence of the marker, ribozymes, peptide nucleic acids (PNA), or other chemicals or molecules with specific affinity.

[0041] Detailed discussion of methods for linking nucleic acids and proteins to a chip substrate, are found in, e.g., U.S. Pat. Nos. 5,143,854; 6,087,1 12; 5,21 5,882; 5,707,807; 5,807,522; 5,958,342; 5,994,076; 6,004,755; 6,048,695; 6,060,240; 6,090,556; and 6,040,138.

[0042] Microarray expression may be detected by scanning the microarray with a variety of laser or CCD-based scanners, and extracting features with software packages, for example, Imagene (Biodiscovery, Hawthorne, CA), Feature Extraction Software (Agilent), Scanalyze (Eisen, M. 1999. SCANALYZE User Manual; Stanford Univ., Stanford, Calif. Ver 2.32.), or GenePix (Axon Instruments).

[0043] In one embodiment, quantitative data obtained about the markers of interest and other dataset components can be subjected to an analytic process with chosen parameters. The parameters of the analytic process may be those disclosed herein or those derived using the guidelines described herein. The analytic process used to generate a result may be any type of process capable of providing a result useful for classifying a sample, for example, comparison of the obtained dataset with a reference

dataset, a linear algorithm, a quadratic algorithm, a decision tree algorithm, or a voting algorithm. The analytic process may set a threshold for determining the probability that a sample belongs to a given class. The probability preferably is at least 50%, or at least 60%, or at least 70%, or at least 80%, or at least 90% or higher.

[0044] The following examples further illustrate the present disclosure but should not be construed as limiting its scope in any way.

EXAMPLES

Example 1. Microarray study.

[0045] The research described in this Example was approved by the Health Science Board of Hungary and the Human Investigation Committee of Wayne State University. After obtaining informed consent, placental tissue samples were collected from Caucasian women at the First Department of Obstetrics and Gynecology, Semmelweis University. Specimens and data were stored anonymously. Pregnancies were dated to be between 8-12 weeks of gestation according to ultrasound scans. Patients with multiple pregnancies (twins, triplets, etc.) or fetuses having congenital or chromosomal abnormalities were excluded. Women were enrolled in the following homogenous groups: (1) preterm severe preeclampsia, with or without HELLP syndrome (n=12) and (2) preterm controls (n=5) (Table 1). Preeclampsia was defined according to the criteria set by the American College of Obstetricians and Gynecologists (Blood pressure: 140 mm Hg or higher systolic or 90 mm Hg or higher diastolic after 20 weeks of gestation in a woman with previously normal blood pressure; proteinuria: 0.3 g or more of protein in a 24-hour urine collection (usually corresponds with 1+ or greater on a urine dipstick test). Severe preeclampsia was defined according to Sibai et al., [Sibai, B et al. Pre-eclampsia. Lancet 2005;365:785-99]. Preterm controls had no medical complications, clinical or histological signs of chorioamnionitis, and delivered neonates with a birth weight appropriate-for-gestational age (AGA). C-section was performed in all preeclampsia cases due to severe symptoms, as well as in all controls due to previous C-section or malpresentation before 37 weeks of gestation.

Table 1

Groups	Preterm control (n=5)	Preterm preeclampsia with / without HELLP syndrome (n=12)
Maternal age (y)[b]	31.6 (31.5-34.3)	30.3 (26.1-35)
Primiparity[a]	40	66.7
Gestational age (week)[b]	31.0 (30.9-34.0)	31.2 (29.3-33.2)
Race[a]		
Caucasian	100	100
African American	0	0
Other	0	0
Systolic BP (mmHg)[b]	120 (120-120)	163 (160-170)[c]
Diastolic BP (mmHg)[b]	80 (70-80)	100 (100-101)[c]
Proteinuria[a]	0	100
Birth weight (g)[b]	1990 (1640-2210)	1065 (990-1420)
Cesarean delivery[a]	100	100

[a] Percentage [b] Median (IQR) [c] $p<0.01$

RNA isolation and microarray experiments

[0046] Placentas (n=17) were obtained immediately after delivery. Tissue specimens were excised from central cotyledons close to the umbilical cord in order to reduce the possible bias due to regional differences in gene expression, dissected from the choriodecidua on dry ice and stored at -80°C. Tissues were homogenized using a ThermoSavant FastPrep FP120 Homogenizer (Thermo Scientific, Wilmington, DE, USA) with Lysing MatrixD (MP Biomedicals, Illkirch, France). Total RNA was isolated using RNeasy Fibrous Tissue Mini Kit (Qiagen GmbH, Hilden, Germany), quantified with NanoDropl ooo (Thermo Scientific) and assessed by Agilent 2100 Bioanalyzer (Matriks AS, Oslo, Norway). Total RNAs (controls, n=5; preeclampsia, n=12) were labeled, and Cy3-RNAs were fragmented and hybridized to the Whole Human Genome Oligo Microarray G4112A on an Agilent scanner, (Agilent Technologies, Santa Clara, CA, USA), and processed with Agilent Feature Extraction software v9.5 according to the manufacturer's guidelines.

Data analysis

[0047] Demographics data were compared by the Fisher's exact test and Mann-Whitney test using SPSS version 12.0 (SPSS Inc., Chicago, IL, USA). Microarray data analysis was performed using the R statistical language and environment (website r-project.org). Microarray expression intensities were background-corrected using the "minimum" method in the "backgroundCorrect" function of the "limma" package. After $\log_2$ transformation, data were quantile-normalized. From the 41,093 probesets on the

array, 93 were removed before differential expression analysis because of lacking annotation in the array definition file (Agilent Technologies). Subsequently, an expression filter was applied to retain probesets with intensity greater than $\log_2$ in at least two samples, yielding a final matrix of 30,027 probesets (15,939 unique genes). Differential gene expression was assessed using a moderated t-test. P-values were adjusted using the false discovery rate (FDR) method. Target gene Entrez IDs for the probesets were determined using the R package "hgu41 12a.db". For probesets without annotation in the package, Entrez IDs were taken from the array definition file (Agilent Technologies). Probesets remaining un-annotated (without Entrez ID and/or gene symbol) were removed from further analysis. Probesets were defined as differentially expressed (n=1409) in this example if they had a FDR of <0.2 and a fold-change of ≥1.5. As used herein, "differential expression", "significantly differentially expressed", and similar terms generally mean that expression of a gene is significantly different based on a statistical power analysis, the results of which can be validated by qPCR at a 95% confidence interval.

[0048] The human U133A/GNF1 H microarray data on 79 human tissues, cells and cell lines from Symatlas microarray database [Su, Al et al. A gene atlas of the mouse and human protein-encoding transcriptomes. PNAS 2004;101:6062-67] was downloaded to search for human genes with predominant placental expression. A probeset was defined as having predominant placental expression, if its placental expression was 1) ≥1,000 fluorescence units; 2) six times higher than the 75th quantile of values in 78 other tissue and cell sources; and 3) two times higher than its expression in the tissue with the second highest expression. The resulting 215 probesets corresponded to 153 unique genes. An additional eleven genes not present on the microarray platform (Affymetrix, Santa Clara, CA, USA) used by Symatlas were added to this list based on their potential relevance. Out of 164 predominantly placental expressed genes, 157 were present on our Agilent array. These genes were tested for enrichment in differentially expressed genes compared to all genes on the array (1,409 out of 15,939) using Fisher's exact tests.

[0049] Chromosomal locations for all genes tested on the Agilent array were obtained from the R package "org.Hs.eg.db". Out of the 15,939 unique and 1,409 differentially expressed genes on the array, 15,935 and 1,408 could be assigned to

chromosomes, respectively. Mapping the microarray probe sets on the Affymetrix human U133A/GNF1H chips to ENTREZ identifiers was performed using the Bioconductor hgu133a.db and hgfocus.db packages. Chromosomal locations of the resulting list of genes were obtained from the package org.Hs.eg.db and from NCBI for the eleven additional genes. Enrichment analyses for chromosomes among PPE genes, differentially expressed genes, and differentially expressed genes encoding for transcriptional regulators were tested by Fisher's exact test. Chromosomal locations of PPE genes and differentially expressed genes (transcription regulators and non-transcription regulators) were visualized by Circos (Figure 1).

[0050] Weighted gene co-expression network analysis (WGCNA) was applied on the 1,409 differentially expressed genes across 17 samples to identify distinct regulation modules and prioritize candidate genes for qPCR verification. Gene pair-wise similarity (absolute Pearson correlation) matrix was first computed, then soft-thresholded by raising to the power of 10 (chosen based on the scalefree topology criterion) to obtain an adjacency matrix. The topology overlap matrix (TOM) was then derived from the adjacency matrix. The topology overlap measures the node interconnectedness within a network and was generalized to a weighted co-expression network. This measure defines similarity between two genes based on both correlations within themselves and outside with other genes. Gene distance matrix was defined as 1-TOM, and used for average linkage hierarchical clustering. A hybrid dynamic tree-cutting method was applied to obtain modules (tree clusters). Gene modules identified with this approach were further tested for enrichment in PPE genes using a Fisher's exact test. Transcription regulatory genes that were expressed at high levels (average log2 intensity >9) and co-expressed (absolute Pearson coefficient >0.8) with the most genes among PPE genes were treated as candidates for hub-genes in the module.

Example 2. Validation study.

Study groups, clinical definitions and sample collection

[0051] The research described in this Example was approved by the Institutional Review Boards of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH), Department of Health and Human Services (DHHS), and Wayne State University. After informed consent was obtained, placentas (n=100) were retrieved from the bank of biological

specimen of the Perinatology Research Branch (NICHD, NIH, DHHS). Pregnancies were dated to be between 8-12 weeks according to ultrasound scans. Patients with multiple pregnancies (twins, triplets, etc.) or fetuses having congenital or chromosomal abnormalities were excluded. Specimens and data were stored anonymously.

[0052] For qRT-PCR, tissue microarray, mRNA in situ hybridization, and laser capture microdissection, placentas were used from women selected from a large cohort into the following, homogenous patient groups: (1) preterm severe preeclampsia (PE; <36 weeks; n=20); (2) preterm severe preeclampsia associated with small-for-gestational age (SGA) (PE-SGA; <36 weeks; n=20); (3) preterm controls (PTC; <36 weeks; n=20); (4) term severe preeclampsia (TPE; >37 weeks; n=10); (5) term severe preeclampsia associated with SGA (TPESGA; >37 weeks; n=10); and (6) term controls (TC; >37 weeks; n=20). Women in these groups were predominantly of African American origin (Table 2). Term controls, consisting of normal pregnant women with (n=10) or without (n=10) labor, and preterm controls with preterm labor and delivery (n=20) had no medical complications or clinical or histological signs of chorioamnionitis, and delivered AGA neonates. Labor was defined by the presence of regular uterine contractions at a frequency of at least two contractions every 10 minutes with cervical changes resulting in delivery. Preeclampsia was defined according to the criteria set by the American College of Obstetricians and Gynecologists. Severe preeclampsia was defined according to Sibai et al., see above. SGA was defined as neonatal birth-weight below the 10th percentile for gestational age. C-section was performed in all preeclampsia cases due to severe symptoms and in controls due to previous C-section or malpresentation.

Table 2

Groups	Preterm control (n=20)	Preterm preeclampsia (n=20)	Preterm preeclampsia with SGA (n=20)	Term control (n=20)	Term preeclampsia (n=10)	Term preeclampsia with SGA (n=10)
Maternal age (y)[b]	22 (20-28.5)	23.5 (21-27)	22.5 (19.5-30)	22 (21-32)	19 (19-35)	26.5 (19-31)
Primiparity[a]	20	40	25	15	40	10
Gestational age (week)[b]	32.3 (28.2-34.9)	31.4 (29.6-33.6)	31.8 (29.7-34.4)	38.6 (38-39.1)	39.1 (38.6-39.6)	38.4 (37.3-38.9)
Race[a]						
Caucasian	5	10	10	15	0	0

African American	95	90	90	80	100	100
Other	0	0	0	5	0	
Systolic BP (mmHg)[b]	116 (110-125)	177 (166-187)[c]	171 (164-189)[c]	121 (111-134)	173 (165-178)[c]	169 (164-190)[c]
Diastolic BP (mmHg)[b]	65 (59-71)	105 (103-111)[c]	108 (94-118)[c]	70 (64-73)	106 (102-110)[c]	102 (97-104)[c]
Proteinuria[b]	0	3 (2-3)[c]	3 (3-3)[c]	0	3 (1-3)[c]	3 (1-3)[c]
Birth weight (g)[b]	1635 (1075-2715)	1488 (1050-1908)	1173 (908-1650)	3215 (3110-3335)	3123 (2990-3200)	2405 (2205-2555)[c]
Birth weight percentile[b]	40.5 (31.9-53.4)	22.7 (18.3-32.9)[d]	6.7 (1-8.6)[c]	46 (37.2-63)	37.1 (28.5-48.8)	1.1 (1-3.5)[c]
Cesarean delivery[a]	45	80[d]	75	55	40	20

[a] Percentage; [b] Median (IQR); [c] $p<0.001$; [d] $p<0.05$

Total RNA isolation, cDNA generation and quantitative real-time RT-PCR

[0053] Total RNA was isolated from snap-frozen placental villous tissues (n=100) with TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and Qiagen RNeasy kit (Qiagen, Valencia, CA, USA) according to the manufacturers' recommendations. The 28S/18S ratios and the RNA integrity numbers were assessed using an Agilent Bioanalyzer 2100 (Agilent Technologies), RNA concentrations were measured with NanoDrop1000 (Thermo Scientific). Five hundred ng of total RNA was reverse transcribed with High Capacity cDNA Reverse Transcription Kit using random hexamers (Applied Biosystems). TaqMan Assays (Applied Biosystems) were used for high-throughput gene expression profiling on the Biomark™ qRT-PCR system (Fluidigm, San Francisco, CA, USA) according to the manufacturers' instructions.

Tissue microarray (TMA) construction, immunostaining and immunoscoring

[0054] TMAs were constructed from FFPE villous tissue blocks (n=100). Briefly, three 20x35mm recipient blocks were made of Paraplast X-Tra tissue embedding media (Fisher Scientific, Pittsburgh, PA, USA). One mm diameter cores from tissue blocks were transferred in triplicate into recipient paraffin blocks using an automated tissue arrayer (Beecher Instruments, Inc., Silver Spring, MD, USA). Five μιτι sections cut from TMAs were placed on silanized slides and stained with antibodies and reagents manually, or either on a Ventana Discovery autostainer (Ventana Medical

Systems, Inc, Tucson, AZ, USA) or a Leica BOND-MAX™ autostainer (Leica Microsystems, Wetzlar, Germany). Images were captured with an Olympus BX41 microscope (Olympus America Inc., Center Valley, PA, USA). Immunostainings were semiquantitatively scored by two examiners blinded to the clinical information with a modified immunoreactive score. Immunostaining intensity was graded as follows: 0=negative, 1=weak, 2=intermediate, and 3=strong. All villi in a random field of each of three cores were evaluated by both examiners, and scores within each core were averaged to represent target protein quantity of that core. Thus, each placenta had three scores corresponding to three cores examined, and group comparisons using these scores were conducted in a same way as for qRT-PCR data.

Histopathologic evaluation of the placenta

[0055] Placental tissue samples (n=1 00) were taken by systematic random sampling, fixed in 10% neutral-buffered formalin, and embedded in paraffin. Five μιτι sections were cut from the villous tissue blocks, stained with hematoxylin and eosin, and examined using bright-field light microscopy by two anatomic pathologists blinded to the clinical information. Histopathologic changes were defined according to published criteria. "Maternal underperfusion" and "fetal vascular thrombo-occlusive disease" scores were calculated by summing the number of different pathologic lesions consistent with these lesion categories present in a given placenta.

Statistical analysis and evaluation of qRT-PCR data

[0056] Demographics data were compared by the Fisher's exact test and Mann-Whitney test using SPSS version 12.0 (SPSS). qPCR data were analyzed using the AACt method in the R statistical environment (website r-project.org). Data was first normalized to the reference gene *(RPLPO)* and batch effect was adjusted through calibrator samples. Log_2 mRNA relative concentrations were obtained for each sample as -**ACt**$_{(gene)}$ = *Ct(RPLPO) - Ctfgene)-* The surrogate gene expression values (-**ACt**$_{gene)}$ were used to perform a hierarchical clustering with 1-Pearson correlation distance and average linkage. Between group comparisons (in which groups were predefined based on the clinical characteristics of the patients) were performed by fitting a linear model on -ACt values, using as covariates the group variable indicator while allowing for an interaction between the group variable and the maturity status of the fetus (term vs. preterm).

[0057] Besides these group comparisons, the analysis was extended to include all 100 patients in the validation phase, to test for the association between gene expression and blood pressure as well as birth weight while adjusting for gestational age. All variables in the latter analysis were treated as continuous. P-values of <0.05 were considered significant.

[0058] A neural networks based approach was used to determine the best combination of 2-8 genes that would best predict blood pressure and birth weight at the same time based on qRT-PCR data. Samples (n=100) were randomly split into 10 equal and balanced (with respect to the presence of preeclampsia) cross-validation folds. At each fold, 90% of the samples were used to rank the genes in an univariate fashion for predicting blood pressure and birth weight separately, and the best 15 genes for each of the two outcomes were retained using simple linear model, adjusting for gender and maturity (term/preterm). Then all gene-combinations were used as inputs in a neural network model that was trained to predict both blood pressure and birth weight using the training data. The remaining 10% of the samples were used to determine the Average Absolute Relative prediction Error (AARE) of the neural network for each gene-combination. The cross-validation procedure was repeated 10 times, splitting therefore the data into 10 different 10-fold partitions, for a total of 100 training and 100 test sets of samples. The number of times that a given gene-combination was found in the top 5% of combinations (the smallest AARE) was recorded and used to rank the combinations of genes for their ability to predict both blood pressure and birth weight percentile. A linear discriminant analysis (LDA) model was used to provide a realistic measure of the sensitivity and specificity of a model predicting the disease status (preeclampsia vs controls). LDA was performed starting with six genes *(FLT1, HSD17B1, LEP, LGALS14, PAPPA2,* and *PLAC1)* chosen from the results of the neural network analysis as being top 2 best predictors and / or highly placenta specific genes, and then was also repeated with a restricted set of four genes *(HSD17B1, LGALS14, PLAC1,* and *PAPPA2).* The 100 samples were split repeatedly at random in two parts: a training part (80% of the samples) was used to fit a LDA model, and a test part (20% of the samples) was used to compute the sensitivity and specificity of the fitted model. The estimates for sensitivity and specificity were averaged over 100 such trials to give a robust estimate. At each trial, the genes

considered were ranked using a t-test and the optimal number of them to be included in the LDA model is determined using the performance of LDA model via an internal 3-fold cross-validation process. The procedure is described in more detail elsewhere (Tarca, A.L., Than, N.G., & Romero, R. Methodological Approach from the Best Overall Team in the IMPROVER Diagnostic Signature Challenge. *Systems Biomedicine* submitted, (201 3).

[0059] The model had two sets of parameters. The mean expressions values (-ACt values) of the four genes in the two groups were (m):

	HSD17B1	LGALS14	PLAC1	PAPPA2
$\mathbf{m}_{Ctl}$	0.091715	-1.77314	-1.81852	-0.65007
$\mathbf{m}_{PE}$	-1.05794	-2.8269	-2.69142	1.092378

and the variance-covariance matrix (∑) was:

	HSD17B1	LGALS14	PLAC1	PAPPA2
HSD17B1	1.677801	1.583045	1.354269	0.26431
LGALS14	1.583045	2.082332	1.533739	0.551178
PLAC1	1.354269	1.533739	1.654183	0.331429
PAPPA2	0.26431	0.551178	0.331429	1.904354

For any new individual we assumed that the expression profile x was available, e.g. x=

HSD17B1	LGALS14	PLAC1	PAPPA2
-1.29723	-2.68723	-2.9415	2.099069

[0060] The posterior probability for each patient class (preeclampsia vs. controls) was computed from the new profile values and model parameters using the multivariate normal formula:

$$P(\mathbf{x} \mid Ctl) = \frac{1}{(2\pi)^{N/2} \mid \Sigma \mid^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{Ctl})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_{Ctl}) \right) = 0.00149$$

$$p(\mathbf{x} \mid PE) = \frac{1}{(2\pi)^{N/2} \mid \Sigma \mid^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{PE})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_{PE}) \right) = 0.0196$$

[0061] When p(x|PE)> p(x|Ctl), the sample was classified as preeclampsia. When p(x|Ctl)>p(x|PE), the sample was classified as control.

[0062] The statistical R package was used to compute these probabilities using the following syntax, assuming that these parameters are loaded into R first, and the *mvtnorm* library is also loaded:

```
> x
        HSD17B1   LGALS14     PLAC1   PAPPA2
10796 -1.297228 -2.687228 -2.941502 2.099069
> m
          HSD17B1   LGALS14     PLAC1     PAPPA2
Control  0.091715 -1.773139 -1.818516 -0.6500658
PE      -1.057936 -2.826896 -2.691423  1.0923776
> sigma
          HSD17B1   LGALS14     PLAC1    PAPPA2
HSD17B1 1.6778008 1.5830448 1.3542685 0.2643097
LGALS14 1.5830448 2.0823321 1.5337394 0.5511775
PLAC1   1.3542685 1.5337394 1.6541829 0.3314286
PAPPA2  0.2643097 0.5511775 0.3314286 1.9043535
> pCtl=dmvnorm(nx, m[1,], sigma,log=FALSE)
> pPE=dmvnorm(nx, m[2,], sigma,log=FALSE)
> pCtl
      10796
0.001489049
> pPE
     10796
0.01960817
```

Example 3. Maternal serum proteomics

Study groups, clinical definitions and sample collection

[0063] All women were enrolled in a prospective, longitudinal, multicenter study in prenatal community clinics of the Maccabi Healthcare Services, Israel between August 2002 and March 2003. Pregnancies were dated according to the last menstrual period and verified by first trimester ultrasound. Patients with multiple pregnancies (twins, triplets, etc.) or fetuses having congenital or chromosomal abnormalities were excluded. The collection and investigation of human clinical samples were approved by the Maccabi Institutional Review Board, experimental procedures and data analyses were approved by the Health Science Board of Hungary and the Human Investigation Committee of Wayne State University. Informed consent was obtained from women prior to sample collection. Specimens and data were stored anonymously.

[0064] Preeclampsia was defined as hypertension that developed after 20 weeks (systolic or diastolic blood pressure >140 or >90 mmHg, respectively, measured at two different time points, 4h to 1 week apart) coupled with proteinuria (>300mg in a 24h urine collection or >2+ on a dipstick) according to the International Society for the Study of Hypertension in Pregnancy. Preeclampsia was defined severe, if 1) severe

hypertension (systolic or diastolic blood pressure ≥160 or ≥110 mmHg) was coupled with proteinuria; 2) if hypertension was coupled with severe proteinuria (>5g/24h or ≥3 on a dipstick), or 3) if maternal multi-organ involvement was present, such as pulmonary edema, oliguria, abnormal liver function, epigastric or right upper-quadrant pain, thrombocytopenia, or severe central nervous symptoms including seizures. Small-for gestational age was defined as neonatal birth weight below the 10th percentile for gestational age. Healthy controls had no medical or obstetric complications and delivered a neonate with a birth-weight appropriate for gestational age.

[0065] Peripheral blood samples were obtained by venipuncture in the first trimester from women who subsequently developed preterm severe preeclampsia (<36 weeks; n=5), term severe preeclampsia (>37 weeks; n=5), as well as healthy controls (>37 weeks; n=10) matched for gestational age at blood draw (Table 3). Samples were kept for 1–2h at room temperature (RT) and then centrifuged at 10,0000 for 10min. Sera were collected, stored at 2-8 ^{o}C for up to 48h until transferred to the Maccabi Central Laboratory, and then stored in aliquots at -20 ^{o}C until shipped on dry ice to Hungary.

Table 3.

Groups	Controls for PE with SGA	Preeclampsia with SGA	Controls for term PE	Term preeclampsia
Number of cases	5	5	5	5
Gestational age at blood draw (week)	10 (9-11)	8 (8-9)	8 (8-9)	9 (8-10)
Gestational age at delivery (week)	39.7 (38.6-40.0)	34.9 (29.3-35.3)	38.7 (38.6-41.0)	38.1 (38.0-38.1)
Systolic BP (mmHg)	105 (104-110)	160 (150165)	110 (110118)	150 (140160)
Diastolic BP (mmHg)	60 (60-70)	100 (100-100)	67 (63-68)	100 (90-100)
Proteinuria	0	4 (3-4)	-	3 (3-4)
Birth weight (gram)	2955 (2900-3100)	1720 (975-1800)	2955 (2900-3100)	3200 (3150-3210)

Median (IQR)

I. Discovery phase

Sample preparations, immunodepletion of high-abundance serum proteins

[0066] Sera were immunodepleted at Biosystems International Ltd. (Debrecen, Hungary) for 14 highly abundant serum proteins on an Agilent 1100 HPLC system

using Multiple Affinity Removal LC Column -Human 14 (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's protocol. To improve the resolution of 2D gels, immunodepleted serum samples were liophylized, and then delipidated and salt depleted at Proteome Services, Ltd. (Budapest, Hungary). Briefly, one volume of all samples was mixed with four volumes of methanol and was thoroughly vortexed. Subsequently, one volume of chloroform was added to these mixtures, which were vortexed again followed by the incorporation of three volumes of water (HPLC grade). After centrifugation at 14,000rpm for 5min at 4°C, the upper phases were discarded. Three volumes of methanol were then added and the resultant mixtures were vortexed and centrifuged again. The supernatants were discarded and the pellets containing the precipitated plasma proteins were air-dried for 10min. The delipidated and salt-depleted plasma protein samples were dissolved in lysis buffer (7M urea; 2M thiourea; 20mM Tris; 5mM magnesium acetate, 4% CHAPS) and their pH was adjusted to 8.0.

Fluorescent labeling and two-dimensional differential in-gel electrophoresis (2D-DIGE)

[0067] Protein concentrations of the immunodepleted, desalted and delipidated serum samples were between 2-4µg/µl as determined with PlusOne Quant Kit (GE Healthcare, Pittsburgh, PA, USA). Samples were equalized for protein content, and then 5µg of each protein sample was labeled with CyDye DIGE Fluor Labeling kit for Scarce Samples (saturation dye) (GE Healthcare) at a concentration of 4nmol/5pg protein according to the manufacturer's instructions. Individual samples from cases (n=10) and controls (n=10) were labeled with Cy5. An internal standard reference sample was pooled from equal amounts (2.5pg) of all individual samples in this experimental set and was labeled with Cy3. Then, 5µg of each Cy5-labeled individual sample was merged with 5µg of the Cy3-labeled reference sample, and these 20 mixtures were run in 2x10 gels simultaneously. Briefly, labeled proteins were dissolved in IEF buffer containing 0.5% ampholytes, 0.5% DTT, 8M urea, 30% glycerin, 2% CHAPS and were rehydrated passively onto 24cm IPG 20 strips (pH3-10, GE Healthcare) for at least 14h at RT. After rehydration, the IPG strips were subjected to first dimension IEF for 24h to attain a total of 80kVh. Focused proteins were reduced by equilibrating with a buffer containing 1% mercaptoethanol for 20min. After reduction, IPG strips were loaded onto 10% polyacrylamide gels (24x20cm) and SDS-

PAGE was conducted at 10W/gel in the second dimension. Then, gels were scanned in a Typhoon TRIO+ scanner (GE Healthcare) using appropriate lasers and filters with the PMT biased at 600V. Images in different channels were overlaid using selected colors and the differences were visualized using Image Quant software (GE Healthcare). Differential protein analysis was performed using the Differential In-gel Analysis (DIA) and Biological Variance (BVA) modules of the DeCyder 6.0 software package (GE Healthcare).

Identification of differentially expressed protein spots

[0068] The internal standard reference sample representative of every protein present in all experiments was loaded equally in all gels, and thus, provided an average image for the normalization of individual samples. The determination of the relative abundance of the fluorescent signal between internal standards across all gels provided standardization between the gels, removing experimental variations and reducing gel-to-gel variations. According to the standard proteomic protocol, the threshold for differential expression was set at 1.05-fold minimum fold-change. A p-value was determined for each protein spot using the Student's t-test by the BVA module of the DeCyder software (GE Healthcare). A p-value of <0.05 was considered statistically significant.

II. Preparative phase

Sample preparation, fluorescent labeling, 2D-DIGE

[0069] The density of spots in the case of Colloidal Coomassie Blue labeling depends only the concentration of protein in the sample, however the density of spots in the case of saturation dyes labeling depends on the number of cysteines of the labeled proteins too, because the saturation dyes labeling method labels all available cysteines on each protein. This results in the same pattern with different density among samples on the analytical and the preparative gels rendering identification more difficult. To eliminate this problem for the exact identification of proteins in spots of interest, the preparative 2D electrophoresis was performed using CyDye saturation fluorescent labeling and Colloidal Coomassie Blue labeling in the same gel. A total of 800µg of proteins per each of the two gels ran. Briefly, the 10-10 immunodepleted serum samples in the "preterm" and "term" comparisons were pooled together and the salt-depletion step was repeated three-times. Five µg protein from each of these two

pooled samples was labeled with Cy3, merged with 800µg of unlabeled proteins from the same sample and resolved in the dry-strip. After separation of the first dimension, focused proteins were first reduced by equilibrating with a buffer containing 1% mercaptoethanol for 20min, and then alkylated with a buffer containing 2.5% iodoacetamide for 20min. Following electrophoresis, gels were scanned in a Typhoon TRIO+ scanner as described above, the differentially expressed spots were matched among the "master" analytical and the fluorescent preparative gel image using Biological Variance (BVA) modules of the DeCyder 6.0 software package (GE Healthcare). The resolved protein spots were visualized by the Colloidal Coomassie Blue G-250 staining protocol. Differentially expressed individual spots were excised from the gels to compare the images.

III. Identification phase

In-gel digestion

[0070] The excised protein spots were analyzed at the Proteomics Research Group of the Biological Research Center of the Hungarian Academy of Sciences (Szeged, Hungary); the detailed protocol is entitled "In-Gel Digest Procedure" described in the website "msfacility.ucsf.edu/ingel.html" and reproduced below: Briefly, salts, SDS and Coomassie brilliant blue were washed out, disulfide bridges were reduced with dithiothreitol, and then free sulfhydryls were alkylated with iodoacetamide. Digestion with side-chain protected porcine trypsin (Promega) proceeded at 37°C for 4h, and the resulting peptides were extracted.

[0071] In-Gel Digest Procedure

1. Wearing gloves and sleeve protectors, wipe down ALL surfaces in the hood with methanol/water moistened lint-free cloth, including the outside of all your tubes (make sure to not wipe off the labeling!), the outside and inside of the Speed Vac and centrifuge, tube racks, bottles etc. Wipe razor blades with methanol-soaked lint-free cloth.
2. Prepare the following solutions:
   25 mM NH_4HCO_3 (100 mg/50 ml)
   25 mM NH_4HCO_3 in 50% ACN
   50% ACN/5% formic acid (may substitute TFA or acetic acid)
   12.5 ng/µL trypsin in 25mM NH_4HCO_3 (freshly diluted)

3. Dice each gel slice into small pieces (1 mm2) and place into 0.65 mL siliconized tubes (PGC Scientific).
4. Add ~100µL (or enough to cover) of 25mM NH_4HCO_3/50% ACN and vortex for 10 min.
5. Using gel loading pipet tip, extract the supernatant and discard.
6. Repeat steps 3 and 4 once or twice.
7. Speed Vac the gel pieces to complete dryness (~ 20 min).
   For low-level proteins (<1 pmol), especially those separated by 1-D SDS-PAGE, reduction and alkylation is recommended. These procedures are performed after step 6.
   a. Prepare fresh solutions:
      10 mM DTT in 25 mM NH_4HCO_3 (1.5 mg/mL)
      55 mM iodoacetamide in 25 mM NH_4HCO_3 (10 mg/mL)
   b. Add 25 µL (or enough to cover) 10 mM DTT in 25 mM NH_4HCO_3 to dried gels. Vortex and spin briefly. Allow reaction to proceed at 56°C for 1 hr.
   c. Remove supernatant, add 25 µL 55 mM iodoacetamide to the gel pieces. Vortex and spin briefly. Allow reaction to proceed in the dark for 45 min. at room temperature.
   d. Remove supernatant (discard). Wash gels with ~100 µL NH_4HCO_3, vortex 10 min, spin.
   e. Remove supernatant (discard). Dehydrate gels with ~100µL (or enough to cover) of 25 mM NH_4HCO_3 in 50% ACN, vortex 5 min, spin. Repeat one time.
   f. Speed Vac the gel pieces to complete dryness (-20 min). Proceed with trypsin digest.
8. Add trypsin solution to just barely cover the gel pieces. Estimate the gel volume and add about 3x volume of trypsin solution. This volume will vary from sample to sample, but on average -5-25 µL is sufficient.
9. Rehydrate the gel pieces on ice or at 4°C for 10 min. Spin. Add 25mM NH_4HCO_3 as needed to cover the gel pieces.
10. Spin briefly and incubate at 37°C for 4 hours - overnight.

Extraction of Peptides

1. Transfer the digest solution (aqueous extraction) into a clean 0.65 ml_ siliconized tube.
2. To the gel pieces, add 30 µl_ (enough to cover) of 50% ACN/5% formic acid, vortex 20-30min., spin, sonicate 5 min. Repeat.
3. Vortex the extracted digests, spin and Speed Vac to reduce volume to 10 µl_.
4. Either proceed with C18 ZipTip (Millipore) cleanup or analyze with LC-MS. Add 2-5 µl_ of 5% formic acid. When analyzing low levels of protein, concentrate the petides by eluting from ZipTips using 3µl_ of elution solution, into a clean 0.65 ml_ siliconized tube.
5. Use 1µl_ of the unseparated digests for analysis by MALDI.

Matrices for unseparated digests:

a-cyano-4-hydroxycinammic acid in 50% ACN/1% TFA (10 mg/mL).

2,5-dihydroxybenzoic acid (DHB), saturated solution in water.

References:

Rosenfeld, et al., Anal. Biochem. (1992) 203(1), 173-179.

Hellman, et al., Anal. Biochem. (1995) 224(1), 451-455.

LC-MS/MS

[0072] Samples were analyzed on a Waters Acquity nanoUPLC system online coupled to an ion trap tandem mass spectrometer (LCQ Fleet, ThermoScientific) in information-dependent acquisition mode, where MS acquisitions (1s survey scans) were followed by CID analyses (3s MS/MS scans) on computer-selected multiply charged ions. HPLC conditions included in-line trapping onto a nanoACQUITY UPLC trapping column (Symmetry, C18 5 µιτι, 180 µιτι x 20 mm) (15 µl/min with 3% solvent B) followed by a linear gradient of solvent B (10 to 50% in 40min, flow rate: 250nl/min; nanoACQUITY UPLC BEH C18 Column, 1.7µηη, 75µηη x 200mm). Solvent A: 0.1% formic acid in water, solvent B: 0.1% formic acid in acetonitrile. LC-MS/MS analysis was performed in "triple play" mode in the mass range of m/z: 450-1600.

Database search and data interpretation

[0073] Raw data files were converted into searchable peak list Mascot generic files (*.mgf) with the Mascot Distiller software v2.1.1.0. (Matrix Science, Inc, London, UK). The resulting peak lists were searched against a human subdatabase of the non-redundant protein database of the National Center for Biotechnology Information

(NCBInr 2008.07.1 8., Bethesda, MD, USA; 6,833,826 sequences) in MS/MS ion search mode on an in-house Mascot server v2.2.04 using Mascot Daemon software v2.2.2. (Matrix Science Inc). Monoisotopic masses with peptide mass tolerance of ±50ppm and fragment mass tolerance of ±0.1 Da were submitted. Carbarn idomethylation of Cys was set as fixed modification, and acetylation of protein N-termini, methionine oxidation, and pyroglutamic acid formation from peptide N-terminal Gin residues were permitted as variable modifications. Acceptance criteria was set to at least two significant (peptide score>40, p<0.05) individual peptides per protein.

RESULTS

I. Differentially expressed genes in preeclampsia are enriched among predominantly placental expressed genes and on three chromosomes

[0074] Because the pathogenesis of preeclampsia originates from the placenta, new biomarker candidates predominantly expressed in the placenta as well as gene-regulatory networks involved in the placental pathogenesis of preeclampsia with a systems biological approach were sought. Analysis of a microarray dataset revealed 1,409 differentially expressed unique genes in preterm preeclampsia compared to preterm controls. From these differentially expressed genes, 137 were found to encode for proteins with functions in transcription regulation (transcription factors, co-activators, or co-repressors). Analysis of BioGPS microarray data and previous evidence revealed 164 unique genes predominantly expressed in the placenta, from which 157 were present on our microarray platform.

[0075] Differentially expressed genes in preeclampsia were highly enriched (OR=3.4, $p=6.9x10^{-9}$) in PPE genes (38 out of 157) when compared to all genes on the array. When investigating chromosomal locations of genes of interest, it was found that differentially expressed genes were enriched in genes located on Chromosomes 6 and 7 (OR=1 .54, pFDR=1 .6x1 0^{-3}, and OR=1 .42, pFDR=0.02, respectively). Interestingly, Chromosome 19 was over-represented in differentially expressed transcription regulatory genes (OR=2.6, pFDR=0.02), and genes with predominant placental expression (OR=2.5, $p=1x10^{-4}$). These enrichments are in accordance with the fact that Chromosome 19 harbors large primate and placenta-specific gene families (e.g. *CGBs, LGALSs, PSGs)* and zinc finger transcription factor gene families. Visualization

of gene expression and co-expression data supports a potential regulatory "hub" role for Chromosome 19 in placental gene expression in primates and its dysregulation in preeclampsia.

[0076] Figure 1. Genomic map of differentially expressed genes in preeclampsia. Circos visualization shows Chromosomes with solid lines in the inner circle. Curved lines connect the genomic coordinates of genes and transcription regulatory genes that are significantly correlated. Significance was determined by fitting a linear model between the expression level of gene and transcription regulatory gene pairs in all samples while controlling for FDR at 5%. Curves represent positive and negative correlations. The second circle shows the genomic location of PPE genes (black lines: non differentially expressed; grey lines: up- or down-regulated). The third and fourth circles show the locations of differentially expressed transcription regulatory genes and non-regulatory genes, respectively with inward-oriented bars (down-regulated) and outward oriented bars (up-regulated). The height of the bars in the third and fourth circles represents the magnitude of gene expression changes.

[0077] Figure 2. Two gene modules in preeclampsia are enriched in PPE genes and are associated with mean arterial blood pressure and birth weight percentile. 2A) Gene modules identified from WGCNA analysis of microarray data. Dysregulated placental gene expression could be characterized by five gene modules within the 1,409 differentially expressed genes in preeclampsia, marked with different shades of grey. The height plotted on the y-axis represents the distance metric (1-TOM) used by WGCNA. Out of 38 PPE genes (black vertical lines), 33 belonged to the grey-scale modules (lighter grey; n=22 and darker grey; n=1 1). These two modules were also enriched in up-regulated and down-regulated genes marked under the modules with grey or black lines, respectively). 2B) Hierarchical clustering of qRT-PCR data obtained with 100 samples and selected 47 genes. Genes from light and dark grey modules clustered together in the validation sample-set. Importantly, 34 out of 60 samples from women with preeclampsia clustered tightly together. Pearson correlation was used for distance, and average for linkage. Samples (column leafs) were shaded according to patient groups and maturity status. 2C) Association of gene expression with mean arterial blood pressure and birth weight percentile. For each gene, a linear model was fitted (expression-blood pressure + birth weight percentile + gender +

maturity status). The significance p-values (-logi $_0$ of) for the two coefficients (blood pressure and birth weight percentile) were plotted for all 47 genes. Genes were shaded according to module membership (except black color for those not differentially expressed on the microarray). Filled circles represent PPE genes, dashed lines the significance threshold at p=0.05. Note that 7 out of 9 genes related to birth weight percentile are from the light grey module, while 10 out of 15 genes related to blood pressure are from the dark grey module.

II. Differentially expressed genes in preeclampsia cluster into major regulatory modules

[0078] In order to identify regulatory modules of genes and transcription regulatory genes, which may drive dysregulated placental gene expression, a WGCNA analysis with the differentially expressed genes on the microarray was conducted. Out of 1,409, 1,403 genes were assigned to four modules containing 506, 442, 381 , and 74 genes. Of interest, 33 out of 38 genes with predominant placental expression belonged to the light grey (n=22) and dark grey (n=1 1) modules. The light grey module was enriched in down-regulated (OR=1 .88, p=2.59x10 $^{-8}$), while the dark grey module was enriched in up-regulated (OR=6.47, p<2.2x1 0^{-16}) genes, suggesting the presence of distinct dysregulated gene-networks in preterm preeclampsia.

[0079] Among up-regulated genes in the dark grey module was FLT1 , which has a pathogenic role in preeclampsia by producing increased amounts of soluble Flt-1 and driving blood pressure elevation. Up-regulated genes with predominant placental expression included *CRH, LEP, PAPPA2, SIGLEC6* and novel biomarker candidates. Among down-regulated, PPE genes in the light grey module were regulators of fetal growth *{CSH1, HSD11B2),* metabolism *(ESRRG),* estrogen synthesis *(HSD17B1),* stress hormone metabolism *(HSD11B2)* and immune regulation of placentation *(LGALS14).*

III. Differentially expressed genes in the dark grey and light grey modules are associated with blood pressure and birth-weight percentile, respectively

[0080] To validate the described results on a large patient population with different ethnic origin and with various subtypes of preeclampsia (preterm and term, with or without SGA), 47 genes for high-throughput expression profiling were selected, if they were: 1) differentially expressed on the microarray, predominantly placental

expressed, specifically by the syncytiotrophoblast, and potentially secreted; 2) transcription regulatory genes with high co-expression with PPE genes; and 3) other genes with relevant role in trophoblast differentiation (e.g. *GCM1),* trophoblast-specific gene expression (e.g. *TEAD3)* or pathogenesis of preeclampsia (e.g. *ENG, LGALS13).*

[0081] Hierarchical clustering of qRT-PCR data showed that 34 out of 60 placentas from women with preeclampsia clustered together. This was also true for the genes belonging to the light and dark grey modules (Figure 2B). Based on the possible involvement of these modules in distinct pathogenic pathways, revealing their biological relevance in a novel way was attempted. "Phenotype analysis" showed that 7 out of 9 genes related to birth weight percentile were from the light grey module, while 10 out of 15 genes related to blood pressure were from the dark grey module (Figure 2C).

[0082] Placental histopathologic data was also assessed. It was found that the expression of genes in the dark grey module was significantly associated with the presence of "fetal vascular thrombo-occlusive disease" *(SIGLEC6, ENG, TPBG)* and "maternal underperfusion" (top associations: *LEP, FLT1, TPBG, ENG),* conditions consistent with placental hypoxia and/or ischaemia. The majority of the light grey module genes (top associations: *CLDN1, HSD17B1, CSH1, PLAC1, LGALS14)* was significantly associated with the presence of "maternal underperfusion".

[0083] In addition, using a classical approach, group comparisons between controls and two groups of preeclampsia at term and preterm separately were performed. It was found that qRT-PCR data validated microarray results in 72% (34/47 genes). Tissue microarray immunostainings for four selected proteins validated the microarray data at the protein level for this module.

[0084] Figure 3. The expression of dark grey module genes changes in the same direction in preeclampsia subgroups. 3A-B) In each barplot, the left and right panels show significant differences ("*") in preterm and term preeclampsia samples, respectively. Gene expression 3A) and protein immunostainings 3B) show similar patterns in sub-groups of preeclampsia. When the change with preeclampsia in the preterm samples was significantly different than the change with preeclampsia in the term samples, a "+" sign indicates this interaction. Semiquantitative immunoscorings

for four proteins 3B) validated gene expression data. 3C) Representative images from the four immunostainings. The same placenta from a preterm control (left, 29 weeks) and from a patient with preterm preeclampsia with SGA (right, 31 weeks) is shown for the four immunostainings (40x magnifications).

[0085] In the light grey module, gene expression changes were also more severe in the preterm groups of preeclampsia than in term. Some genes had significant dysregulation both at term and preterm (e.g. *LGALS13, LGALS14),* while others only at preterm (e.g. *CSH1).* These data also reflect to the heterogeneous placental pathology behind the pathogenesis of preeclampsia, and the more severe pathologies in preterm.

[0086] Figure 4A-UU show gene comparisons for *ARNT2; BCL3; BCL6; BTG2; CDKN1A; CGB3; CLC; CLDN1; CRH; CSH1; CYP19A1; DUSP1; ENG; ERVFRDE1; ERVWE1; ESRRG; FBLN1; FLT1; GATA2; GCM1; GH2; HLF; HSD11B2; HSD17B1; IKBKB; INSL4; JUNB; KIT; LEP; LGALS13; LGALS14; LGALS16; LGALS17A; MAPK13; NANOG; PAPPA; PAPPA2; PGF; PLAC1; POU5F1; SIGLEC6; TEAD3; TFAM; TFAP2A; TPBG; VDR;* and *ZNF554* respectively.

IV. Transcriptomic biomarkers

[0087] The results of the neural network based analysis was a set of combinations of 2 to 8 genes, as assessed by the number of times they were retained as best predictors of blood pressure and birth-weight percentile when using different subsets of the training samples (Table 4). From these sets of combinations, six genes *{FLT1, HSD17B1, LEP, LGALS14, PAPPA2,* and *PLAC1)* were selected as being top 2 best predictors and / or highly placenta specific genes.

Table 4

Best 2 predictors	**X times out of 100 training-test sessions in top 5%**
HSD17B1 / PAPPA2	52
HSD17B1 / LEP	35
LEP / LGALS13	32
LGALS14 / PAPPA2	30
LEP / LGALS14	26
FLT1 / HSD17B1	25
ENG / LGALS13	23

CRH / LGALS14	*21*
CSH1 / PAPPA2	*21*
FLT1 / LGALS14	**19**

Best 3 predictors	
HSDl 7B1 / KIT/ PAPPA2	**38**
CRH / HSDl 7B1 / PAPPA2	**35**
CSH1 / HSDl 7B1 / PAPPA2	**35**
HSDl 7B1 / LGALS13 / PAPPA2	**35**
CLC/ HSDl 7B1 / PAPPA2	**33**
CLDN1 / HSDl 7B1 / PAPPA2	**33**
FBLN1 / HSDl 7B1 / PAPPA2	**33**
CGB3 / HSDl 7B1 / PAPPA2	**32**
HSDl 7B1 / PAPPA2 / PLAC1	**31**
CSH1 / LEP / SIGLEC6	**30**

Best 4 predictors	
FBLNl / HSD1 7B1 / LEP / SIGLEC6	**34**
CRH / HSD11B2 / HSDl 7B1 / LGALS14	**33**
CLC/ HSDl 7B1 / KIT/ PAPPA2	**31**
CLC/ HSDl 7B1 / LGALS13 / PAPPA2	**31**
CRH / HSDl 7B1 / LGALS14 / PAPPA2	**29**
HSD1 7B1 / LEP / LGALS13 / PAPPA2	**28**
CRH / LGALS14 / PAPPA2 / TPBG	**26**
CLC/ CRH / LGALS14 / PAPPA2	**25**
CRH / HSDl 7B1 / KIT / PAPPA2	**25**
CRH / HSD1 7B1 / LEP / LGALS14	**25**

Best 5 predictors	
HSDl 7B1 / LEP / LGALS13 / PAPPA2 / SIGLEC6	**32**
CRH / HSD11B2 / HSDl 7B1 / LGALS14 / PAPPA2	**31**
CLC/ CRH / KIT / LGALS14 / PAPPA2	**30**
CSH1 / FLT1 / HSD1 7B1 / LEP / SIGLEC6	**29**
HSD11B2 / HSDl 7B1 / LEP / LGALS13 / SIGLEC6	**29**
CRH / CSH1 / ENG / LGALS14 / PAPPA2	**28**
CRH / CSH1 / LGALS14 / PAPPA2 / TPBG	**28**
CRH / HSD11B2 / HSDl 7B1 / LGALS13 / TPBG	**28**
FBLNl / LEP / LGALS14 / PAPPA2 / SIGLEC6	**28**
FLTl / HSDl 1B2 / LEP / LGALS13 / SIGLEC6	**28**

Best 6 predictors	
CRH / HSD11B2 / HSDl 7B1 / LGALS13 / LGALS14 / TPBG	**37**

CRH / CSH1 / HSD11B2 / HSD17B1 / LGALS14 / PAPPA2	**33**
CLC/ CRH / CSH1 / KIT / LGALS14 / PAPPA2	**31**
HSD11B2 / KIT/ LEP / LGALS14 / SIGLEC6 / TPBG	**30**
CRH / CSH1 / HSD11B2 / LGALS14 / PAPPA2 / TPBG	**29**
CSH1 / FBLN1 / HSD17B1 / LEP/ PAPPA2 / SIGLEC6	**29**
CLC/ CRH / CSH1 / FBLN1 / LGALS14 / PAPPA2	**28**
CRH / CSH1 / HSD11B2 / HSD17B1 / LGALS14 / SIGLEC6	**28**
CRH / HSD11B2 / HSD17B1 / KIT/ LGALS14 / PAPPA2	**28**
ENG / FBLN1 / HSD17B1 / LEP / PAPPA2 / SIGLEC6	**28**

Best 7 predictors	
CRH / HSD11B2 / HSD17B1 / LGALS13 / LGALS14 / PAPPA2 / TPBG	**40**
CRH / CSH1 / HSD11B2 / HSD17B1 / LGALS13 / LGALS14 / PAPPA2	**36**
CRH / HSD11B2 / HSD17B1 / KIT/ LGALS13 / LGALS14 / TPBG	**34**
CGB3 / CRH / HSD11B2 / HSD17B1 / LGALS13 / LGALS14 / TPBG	**31**
CRH / CSH1 / FLT1 / HSD11B2 / HSD17B1 / LGALS14 / PAPPA2	**31**
CLC/ CRH / CSH1 / FBLN1 / LGALS14 / PAPPA2 / PLAC1	**30**
CRH / CSH1 / HSD11B2 / HSD17B1 / LGALS14 / PAPPA2 / PLAC1	**30**
CRH / CSH1 / HSD11B2 / LGALS13 / LGALS14 / PAPPA2 / TPBG	**30**
CRH / CSH1 / LGALS13 / LGALS14 / PAPPA2 / PLAC1 / TPBG	**30**
CRH / CSH1 / ENG / HSD11B2 / HSD17B1 / LGALS14 / PAPPA2	**29**

Best 8 predictors	
CRH / CSH1 / HSD11B2 / HSD17B1 / LGALS13 / LGALS14 / PAPPA2 / TPBG	**47**
CRH / CSH1 / FLT1 / HSD11B2 / HSD17B1 / LGALS13 / LGALS14 / TPBG	**37**
CGB3 / CRH / HSD11B2 / HSD17B1 / LGALS13 / LGALS14 / PAPPA2 / TPBG	**34**
CLC/ CRH / HSD11B2 / HSD17B1 / KIT / LGALS13 / LGALS14 / TPBG	**33**
CRH / FLT1 / HSD11B2 / HSD17B1 / LGALS13 / LGALS14 / PAPPA2 / TPBG	**33**
CLC/ CRH / CSH1 / FBLN1 / FLT1 / KIT / LGALS14 / PAPPA2	**32**
CRH / CSH1 / HSD11B2 / HSD17B1 / LGALS13 / LGALS14 / SIGLEC6 / TPBG	**32**
CRH / CSH1 / FBLN1 / HSD11B2 / HSD17B1 / LGALS13 / LGALS14 / TPBG	**31**
CRH / CSH1 / FLT1 / HSD11B2 / HSD17B1 / LGALS13 / LGALS14 / PAPPA2	**31**
CLC/ CLDN1 / CRH / FBLN1 / HSD11B2 / HSD17B1 / KIT/ LGALS14	**30**

[0088] The Linear Discriminant Analysis showed that the average sensitivity and specificity of the selected set of transcriptomic biomarkers for the detection of preeclampsia was 91.5% and 75%, respectively.

V. Differentially expressed proteins in maternal serum in preeclampsia

[0089] In the discovery phase, in the comparison of samples taken from healthy pregnant women with normal pregnancy outcome and those who subsequently developed preterm severe preeclampsia, 2080-2460 protein spots were identified on

the gels. There were 39 protein spots, which were differentially expressed (29 down-regulated and 10 up-regulated) in at least 3 out of the 5 disease samples. The biggest difference in disease samples was 3.1-fold up-regulation and 3.1-fold down-regulation.

[0090] In the comparison of samples taken from healthy pregnant women with normal pregnancy outcome and those who subsequently developed term severe preeclampsia, 2130-2380 protein spots were identified on the gels. There were 20 protein spots, which were differentially expressed (11 down-regulated and 9 up-regulated) in at least 3 out of the 5 disease samples. The biggest difference in disease samples was a 3.9-fold up-regulation and a 4.5-fold down-regulation.

[0091] In the preparative phase, in the comparison of samples taken from healthy pregnant women with normal pregnancy outcome and those who subsequently developed preterm severe preeclampsia, there were -2380 protein spots identified on the gels. From the 39 previously identified differentially expressed spots in preterm preeclampsia, 29 (25 down-regulated and 4 up-regulated) was identified and excised from the gels.

[0092] In the comparison of samples taken from healthy pregnant women with normal pregnancy outcome and those who subsequently developed term severe preeclampsia, there were -2350 protein spots identified on the gels. From the 20 previously identified differentially expressed spots, 18 (11 down-regulated and 7 up-regulated) was identified and excised from the gels.

[0093] In the identification phase, the following differentially expressed proteins could be identified:

A) Preterm preeclampsia

No	Direction	Gene symbol	ID	Protein Name
1	DOWNinPE	A1BG	gi|21071030	alpha 1B-glycoprotein precursor
2	DOWNinPE	AGT	gi|532198	Angiotensinogen
3	DOWNinPE	APOA4	gi|178757	apolipoprotein A-IV precursor
4	DOWNinPE	APOL1	gi|12408013	apolipoprotein L-I
5	DOWNinPE	CP	gi|4557485	ceruloplasmin precursor
6	DOWNinPE	C1QB	gi|399140	complement C1q subcomponent subunit B precursor
7	DOWNinPE	C7	gi|45580688	complement component 7 precursor
8	DOWNinPE	C4	gi|2347136	complement component C4
9	DOWNinPE	CFB	gi|291922	complement factor B
10	DOWNinPE	CFH	gi|148745112	complement factor H
11	DOWNinPE	GSN	gi|4504165	gelsolin isoform a precursor

12	DOWNinPE	HPX	gi|386789	hemopexin precursor
13	DOWNinPE	HRG	gi|4504489	histidine-rich glycoprotein precursor
14	DOWNinPE	IGFALS	gi|4826772	insulin-like growth factor binding protein, acid labile subunit
15	DOWNinPE	KNG1	gi|3774864 1	kininogen 1
16	DOWNinPE	PLG	gi|387026	Plasminogen
17	DOWNinPE	PAEP	gi| 182093	pregnancy-associated endometrial alpha2-globulin
18	DOWNinPE	GC	gi| 139641	vitamin D-binding protein precursor
19	UPinPE	APOH	gi| 153266841	apolipoprotein H precursor
20	UPinPE	C4	gi|23471 36	complement component C4

B) Term preeclampsia

No	Direction	Gene symbol	ID	Protein Name
1	DOWNinPE	SERPINA3	gi|177809	alpha-1-antichymotrypsin
2	DOWNinPE	CP	gi|4557485	ceruloplasmin precursor
3	DOWNinPE	C7	gi|45580688	complement component 7 precursor
4	DOWNinPE	CFB	gi|291922	complement factor B
5	DOWNinPE	GSN	gi|4504165	gelsolin isoform a precursor
6	DOWNinPE	HRNR	gi|28557150	hornerin
7	DOWNinPE	ITIH2	gi|55958063	inter-alpha (globulin) inhibitor H2
8	UPinPE	AGT	gi|532198	angiotensinogen
9	UPinPE	CFB	gi|291922	complement factor B
10	UPinPE	FETUB	gi|49902016	fetuin B (alpha-2 Heremans-Schmid glycoprotein)
11	UPinPE	GSN	gi|4504165	gelsolin isoform a precursor
12	UPinPE	ITIH4	gi|31542984	inter-alpha (globulin) inhibitor H4
13	UPinPE	CD14	gi|3983127	monocyte antigen CD14 precursor
14	UPinPE	PEDF	gi|189778	pigment epithelial-differentiating factor
15	UPinPE	PLG	gi|387026	plasminogen
16	UPinPE	GC	gi|139641	vitamin D-binding protein precursor

[0094] In each two comparisons, those candidates were selected which were differentially expressed in all disease specimens, had the highest fold-change, and the strongest p-value: complement factor B, gelsolin isoform a precursor, hornerin, fetuin B, hemopexin precursor, and apolipoprotein H precursor.

[0095] Unless otherwise indicated, all numbers expressing quantities of ingredients, properties such as molecular weight, reaction conditions, and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about." Accordingly, unless indicated to the contrary, the numerical parameters set forth in the specification and attached claims are approximations that may vary depending upon the desired properties sought to be obtained by the present disclosure. At the very least, and not as an attempt to limit the application of the doctrine of equivalents to the scope of the claims, each numerical parameter should at

least be construed in light of the number of reported significant digits and by applying ordinary rounding techniques. Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the invention are approximations, the numerical values set forth in the specific examples are reported as precisely as possible. Any numerical value, however, inherently contains certain errors necessarily resulting from the standard deviation found in their respective testing measurements.

[0096] The terms "a," "an," "the", and similar referents used in the context of describing the invention (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. Recitation of ranges of values herein is merely intended to serve as a shorthand method of referring individually to each separate value falling within the range. Unless otherwise indicated herein, each individual value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., "such as") provided herein is intended merely to better illuminate the invention and does not pose a limitation on the scope of the invention otherwise claimed. No language in the specification should be construed as indicating any non-claimed element essential to the practice of the invention.

[0097] Groupings of alternative elements or embodiments of the invention disclosed herein are not to be construed as limitations. Each group member may be referred to and claimed individually, or in any combination with other members of the group or other elements found herein. It is anticipated that one or more members of a group may be included in, or deleted from, a group for reasons of convenience and/or patentability. When any such inclusion or deletion occurs, the specification is deemed to contain the group as modified thus fulfilling the written description of all Markush groups used in the appended claims.

[0098] Certain embodiments of this invention are described herein, including the best mode known to the inventors for carrying out the invention. Of course, variations on these described embodiments will become apparent to those of ordinary skill in the art upon reading the foregoing description. The inventor expects skilled artisans to employ such variations as appropriate, and the inventors intend for the invention to be

practiced otherwise than specifically described herein. Accordingly, this invention includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above described elements in all possible variations thereof is encompassed by the invention unless otherwise indicated herein or otherwise clearly contradicted by context.

[0099] Specific embodiments disclosed herein may be further limited in the claims using "consisting of or "consisting essentially of language. When used in the claims, whether as filed or added per amendment, the transition term "consisting of" excludes any element, step, or ingredient not specified in the claims. The transition term "consisting essentially of limits the scope of a claim to the specified materials or steps and those that do not materially affect the basic and novel characteristic(s). Embodiments of the invention so claimed are inherently or expressly described and enabled herein.

[0100] Furthermore, numerous references have been made to patents and printed publications throughout this specification. Each of the above-cited references and printed publications are individually incorporated herein by reference in their entirety.

In closing, it is to be understood that the embodiments of the invention disclosed herein are illustrative of the principles of the present invention. Other modifications that may be employed are within the scope of the invention. Thus, by way of example, but not of limitation, alternative configurations of the present invention may be utilized in accordance with the teachings herein. Accordingly, the present invention is not limited to that precisely as shown and described.

CLAIMS

1. A method for assessing the presence or risk of preeclampsia/HELLP syndrome, in a female to determine the need for a treatment regimen comprising:

determining levels of one or more of complement factor B; gelsolin isoform a precursor; hornerin, fetuin B; hemopexin precursor; apolipoprotein H precursor; fms-related tyrosine kinase 1; hydroxysteroid (17-β) dehydrogenase 1; leptin; lectin galactoside-binding soluble 14; pappalysin 2; or placenta-specific 1 in a biological sample obtained from the female;

generating a dataset based on the determined levels;

assessing the presence or risk of developing preeclampsia in the female based on the dataset; and

determining a treatment regimen based on the assessed presence or risk.

2. A method for assessing the presence or risk of preeclampsia in a female to determine the need for a treatment regimen comprising:

determining levels of one or more of complement factor B, hornerin, hemopexin precursor, hydroxysteroid (17-β) dehydrogenase 1, lectin galactoside-binding soluble 14, or pappalysin 2 in a biological sample obtained from the female;

generating a dataset based on the determined levels;

assessing the presence or risk of developing preeclampsia in the female based on the dataset; and

determining a treatment regimen based on the assessed presence or risk.

3. A method for assessing the risk of complications closely related to preeclampsia/HELLP syndrome in pregnancy in early gestation wherein said complications are selected from, but not limited to, implantation failure, and threatened and spontaneous miscarriage, in a female to determine the need for a treatment regimen comprising:

determining levels of one or more of complement factor B; gelsolin isoform a precursor; hornerin, fetuin B; hemopexin precursor; apolipoprotein H precursor; fms-related tyrosine kinase 1; hydroxysteroid (17-β) dehydrogenase 1; leptin; lectin galactoside-binding soluble 14; pappalysin 2; or placenta-specific 1 in a biological sample obtained from the female;

generating a dataset based on the determined levels;

assessing the presence or risk of developing preeclampsia in the female based on the dataset; and

determining a treatment regimen based on the assessed presence or risk.

4. A method according to claim 1, 2, or 3, wherein the assaying is performed for the levels of at least 3 biomarkers.

5. A method according to claim 1, 2, 3, or 4, wherein the sample is a blood sample.

6. A method according to claim 1, 2, 3, or 4 wherein the sample is other body fluid, secretion, or excretion (such as, but not limited to, cervicovaginal fluid, saliva, or urine) sample.

7. A method according to claim 1, 2, 3, or 4 wherein the sample is amniotic fluid sample.

8. A method according to claim 1, 2, 3, or 4, wherein the sample is fetal cells obtained invasively or non-invasively.

9. A method according to claim 1, 2, 3, or 4, wherein the sample is a placental sample.

10. A method according to claim 1, 2, 3, 4, 5, or 6 wherein the biological sample is obtained before the 20th week of pregnancy, before the 19th week of pregnancy, before the 18th week of pregnancy, before the 17th week of pregnancy, before the 16th week of pregnancy, before the 15th week of pregnancy, before the 14th week of pregnancy, before the 13th week of pregnancy, before the 12th week of pregnancy, before the 11th week of pregnancy, before the 10th week of pregnancy, before the 9th week of pregnancy, before the 8th week of pregnancy, before the 7th week of pregnancy, before the 6th week of pregnancy, or after delivery.

11. A method according to claim 1, 2, 3, 4, 5, 6, or 7 wherein the treatment regimen is a therapeutic intervention.

12. A method according to claim 11 wherein the therapeutic intervention prevents or reduces symptoms of preeclampsia before the symptoms manifest in the female and/or fetus.

13. A kit for assessing the presence or risk of preeclampsia in a female to determine

the need for a treatment regimen comprising: detection mechanisms for determining levels of one or more of complement factor B; gelsolin isoform a precursor; hornerin, fetuin B; hemopexin precursor; apolipoprotein H precursor; fms-related tyrosine kinase 1; hydroxysteroid (17-β) dehydrogenase 1; leptin; lectin galactoside-binding soluble 14; pappalysin 2; or placenta-specific 1 in a biological sample obtained from the female; instructions how to (i) generate a dataset based on the determined levels; (ii) assess the presence or risk of developing preeclampsia in the female based on the dataset; and (iii) determine a treatment regimen based on the assessed presence or risk.

14. A kit for assessing the presence or risk of preeclampsia in a female to determine the need for a treatment regimen comprising: detection mechanisms for determining levels of one or more of complement factor B, hornerin, hemopexin precursor, hydroxysteroid (17-β) dehydrogenase 1, lectin galactoside-binding soluble 14, or pappalysin 2 in a biological sample obtained from the female; instructions how to (i) generate a dataset based on the determined levels; (ii) assess the presence or risk of developing preeclampsia in the female based on the dataset; and (iii) determine a treatment regimen based on the assessed presence or risk.

15. A kit according to claim 13 or 14 wherein the kit includes detection mechanisms for at least three markers.

16. A kit according to claim 13 or 14 wherein the kit includes detection mechanisms for all markers of claim 13 or 14.

17. A kit according to claim 13 or 14 wherein the kit includes detection mechanisms for at least one marker of claim 13 or 14.

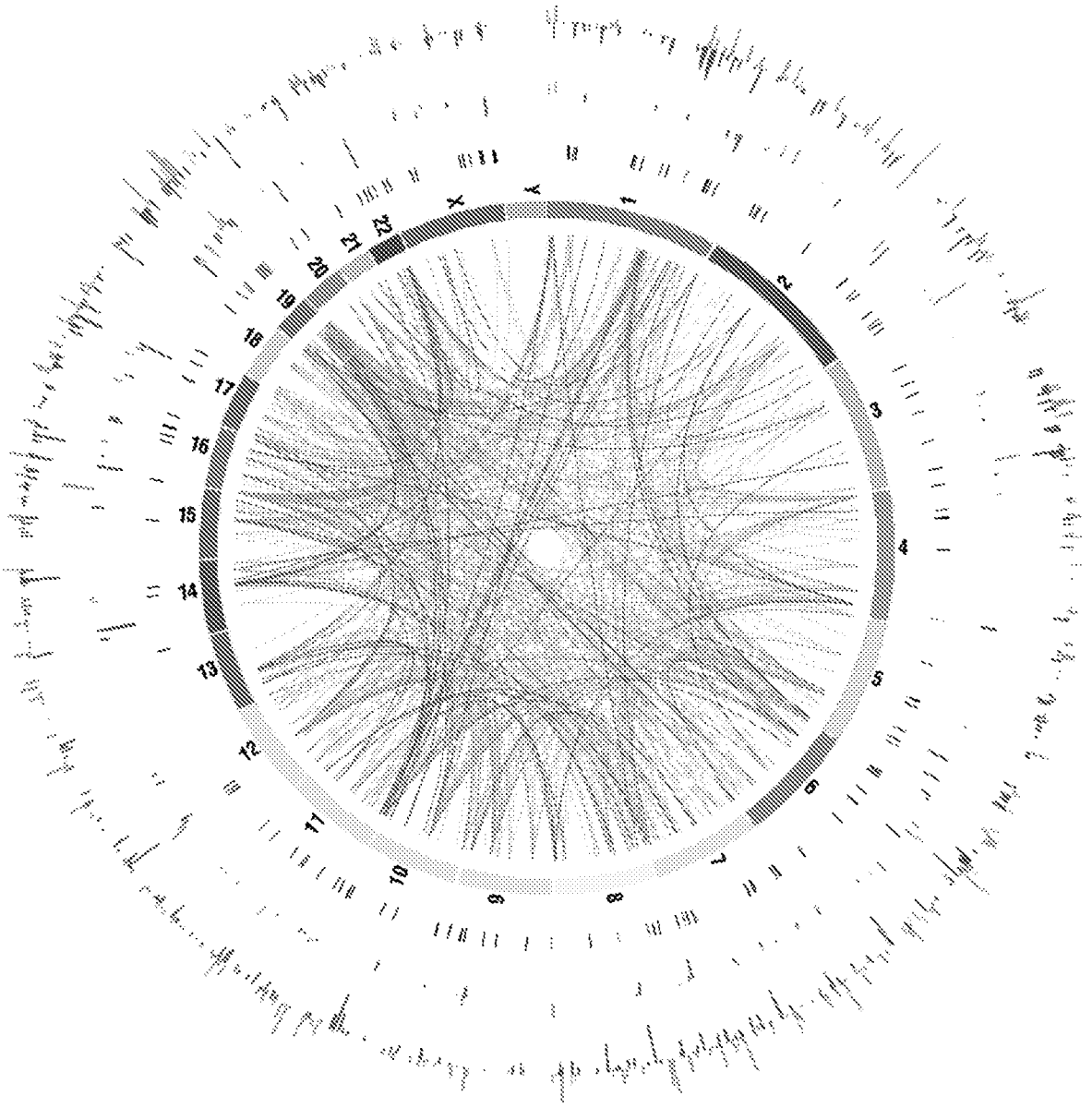

Figure 1

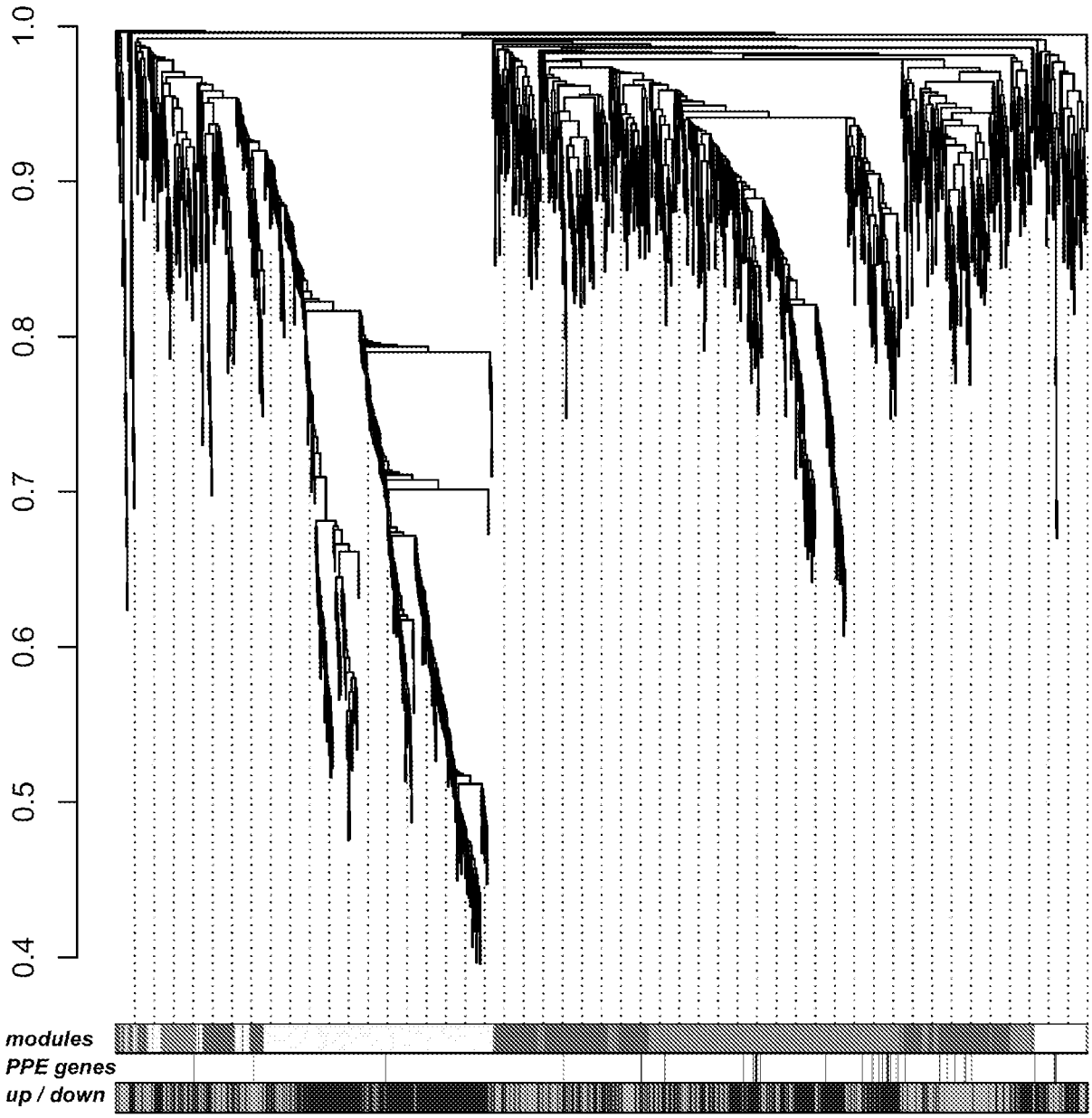

Figure 2A

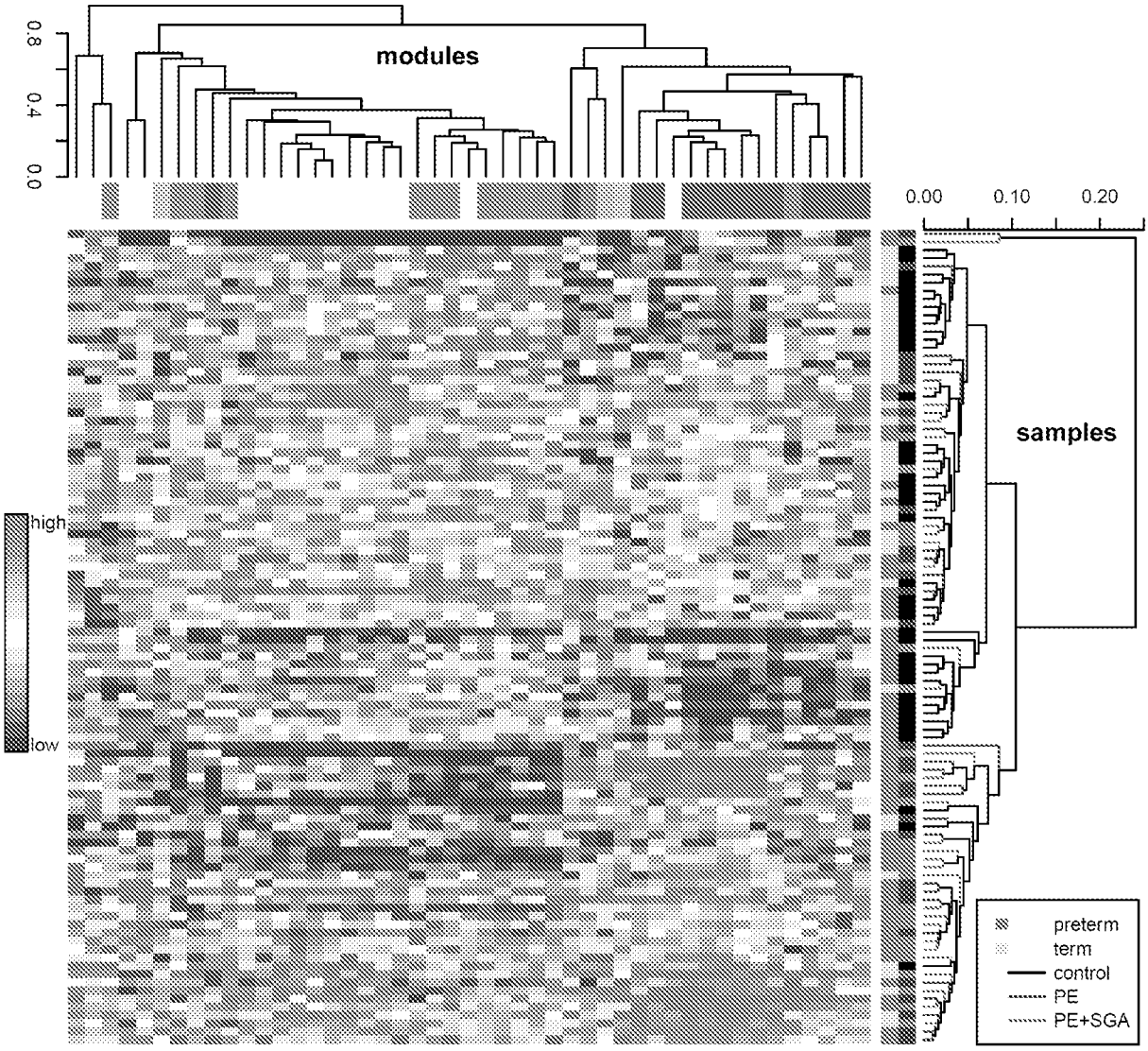
modules
0.8
0.4
0.0
0.00
0.10
0.20
samples
high
low
preterm
term
control
PE
PE+SGA

Figure 2B

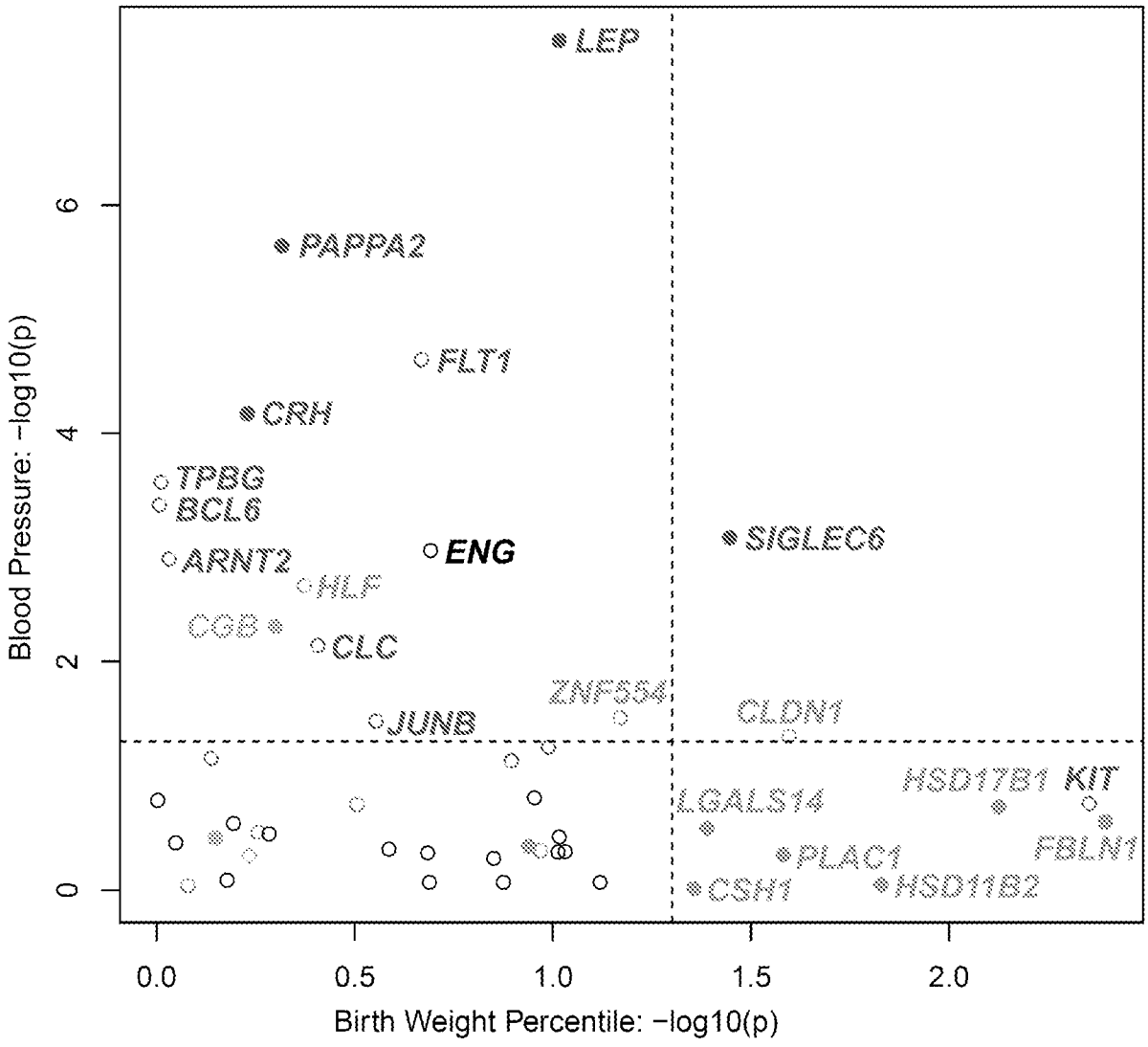
LEP
PAPPA2
FLT1
CRH
TPBG
BCL6
ENG
SIGLEC6
ARNT2
HLF
CGB
CLC
ZNF554
JUNB
CLDN1
HSD17B1
KIT
LGALS14
PLAC1
FBLN1
CSH1
HSD11B2
Blood Pressure: −log10(p)
0
2
4
6
0.0
0.5
1.0
1.5
2.0
Birth Weight Percentile: −log10(p)

Figure 2C

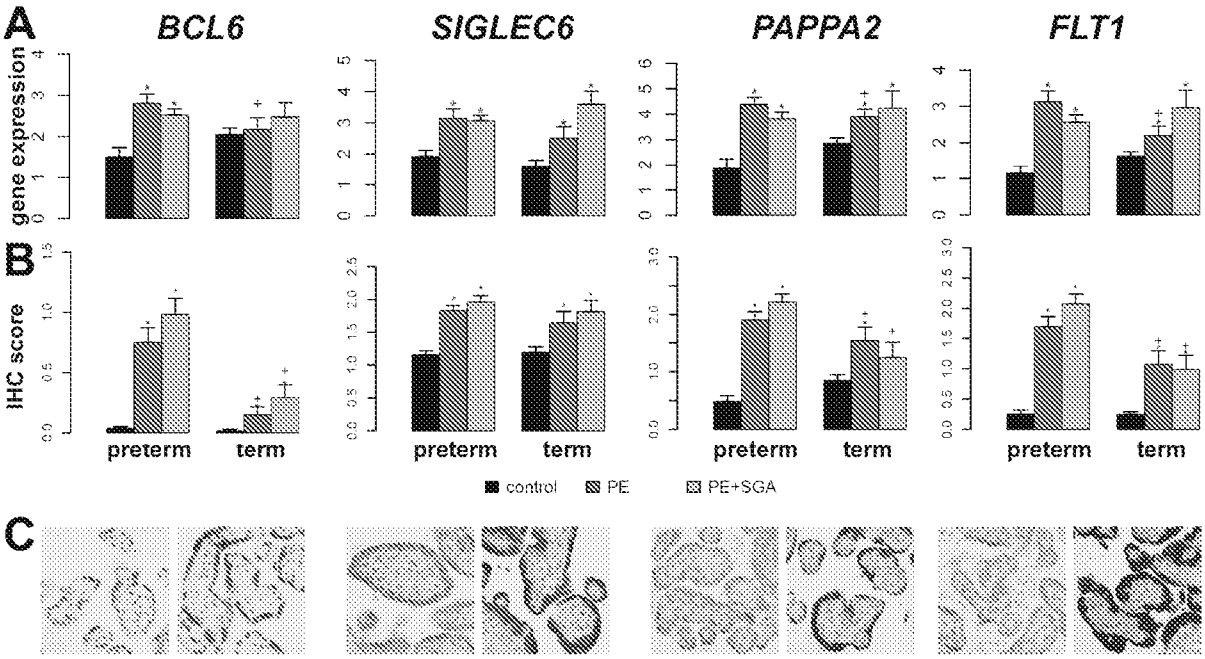
A
BCL6
SIGLEC6
PAPPA2
FLT1
gene expression
B
IHC score
preterm
term
control
PE
PE+SGA
C

Figure 3

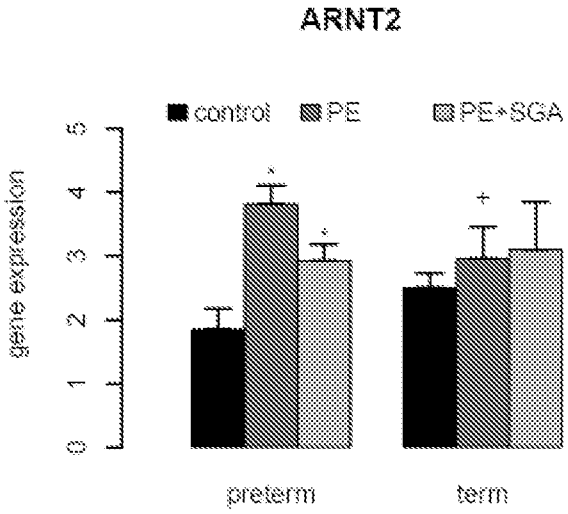

Figure 4A

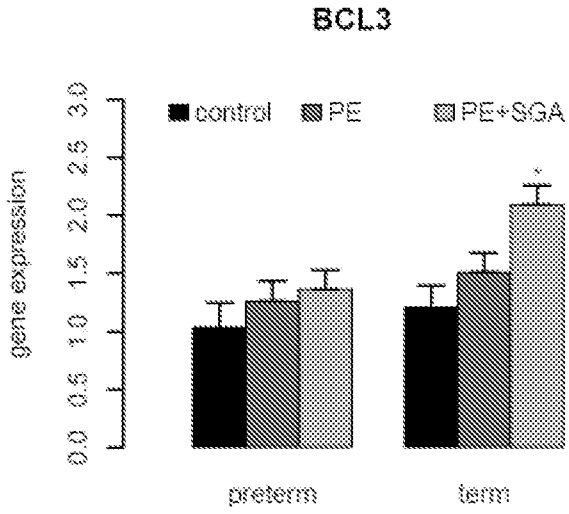

Figure 4B

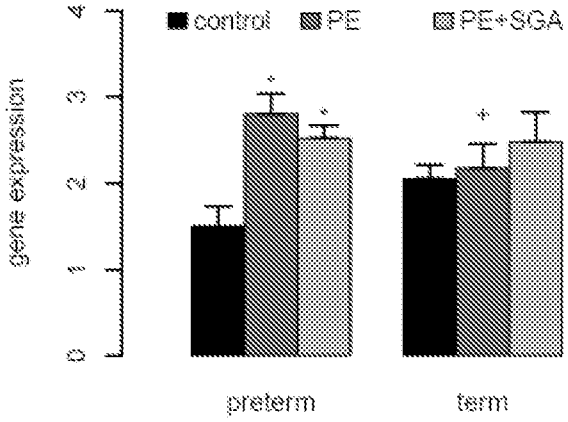

Figure 4C

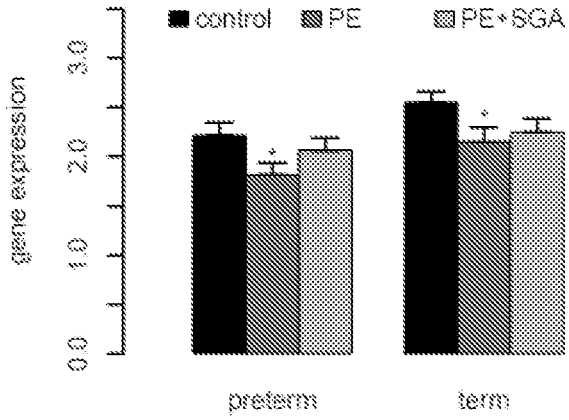

Figure 4D

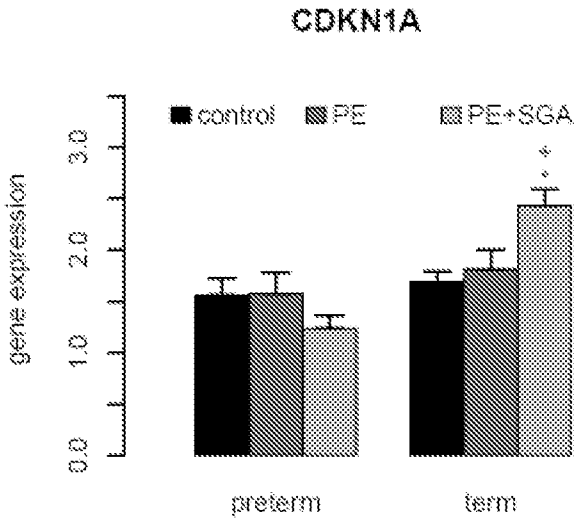

Figure 4E

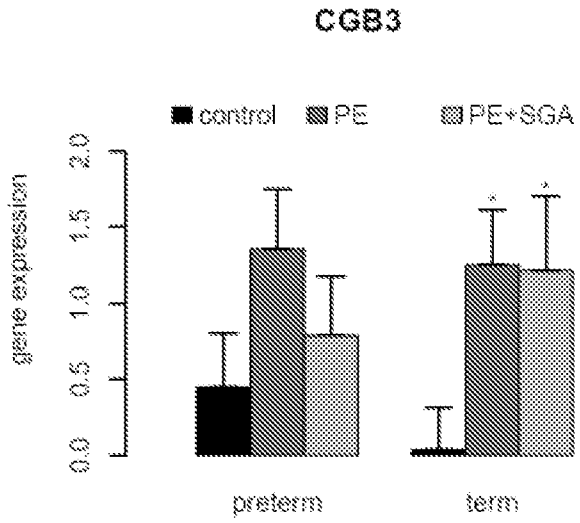

Figure 4F

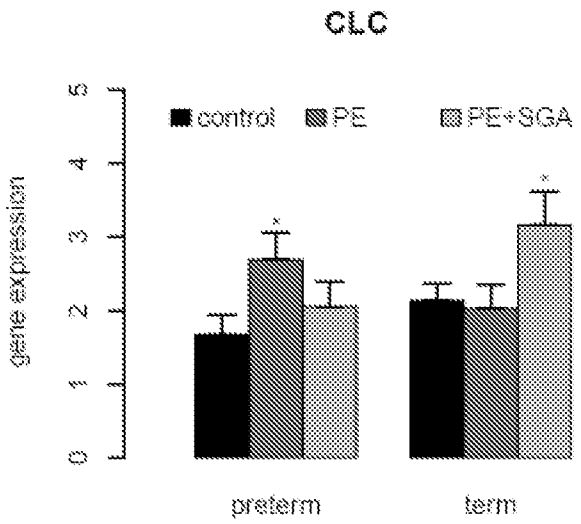

Figure 4G

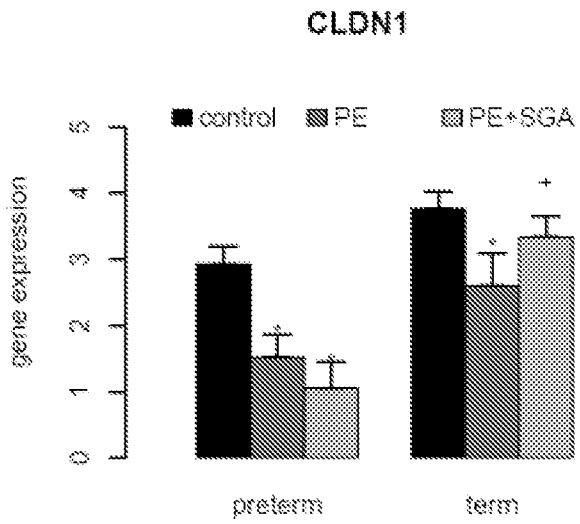

Figure 4H

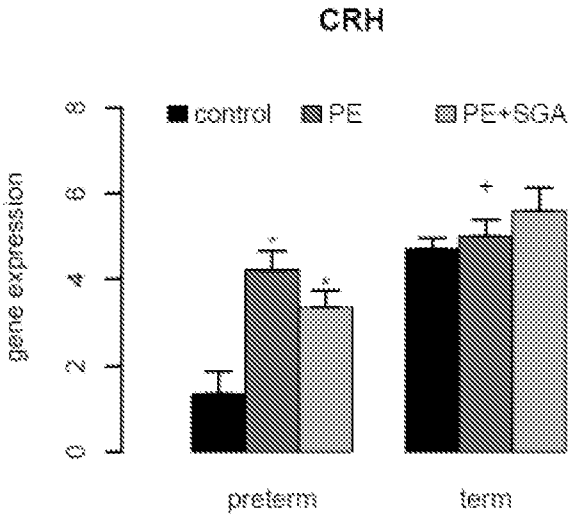

Figure 4I

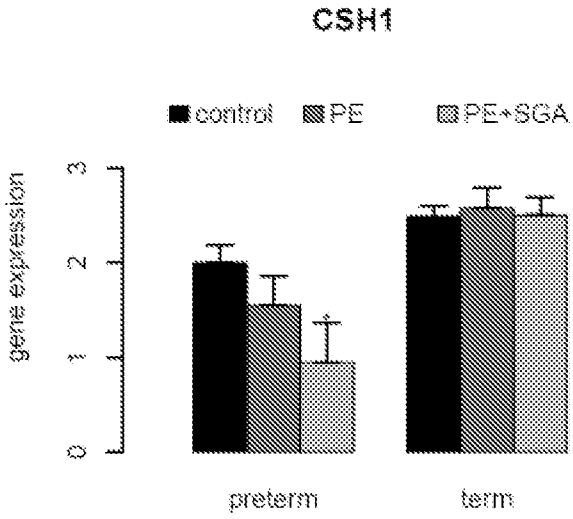

Figure 4J

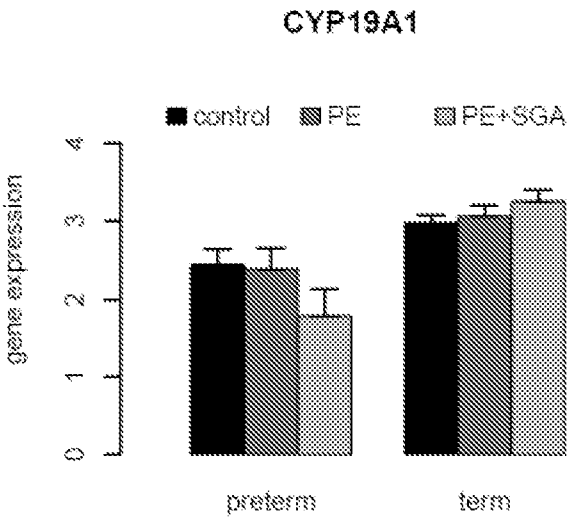

Figure 4K

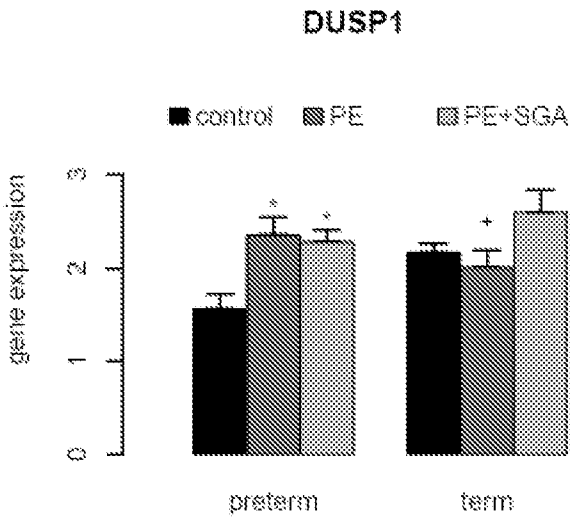

Figure 4L

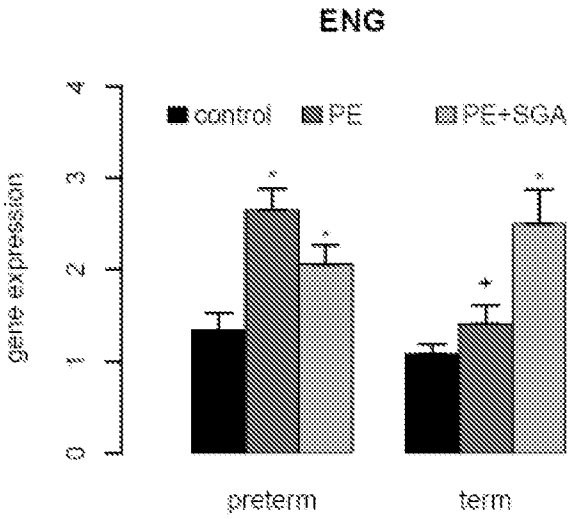

Figure 4M

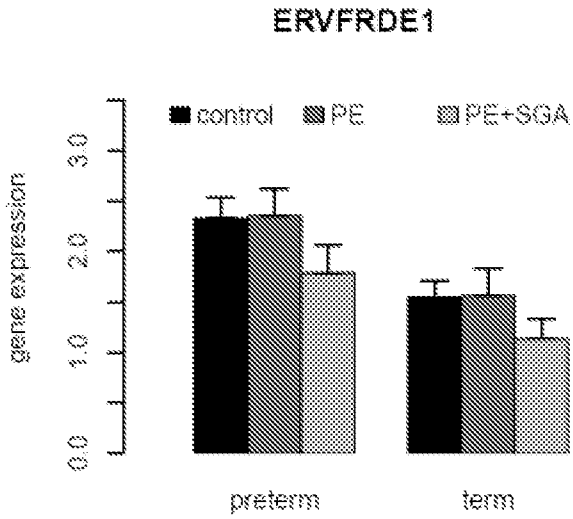

Figure 4N

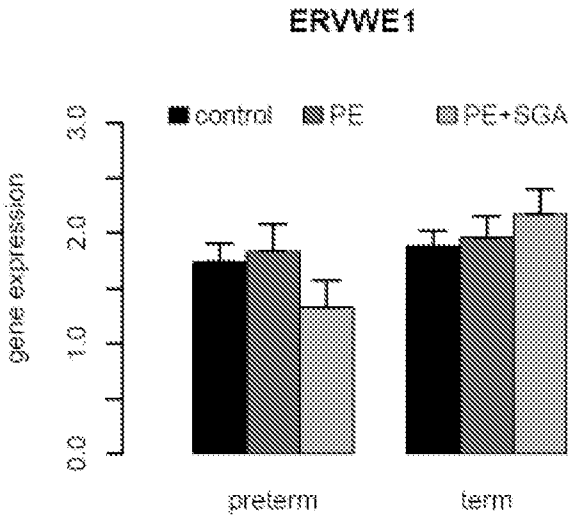

Figure 4O

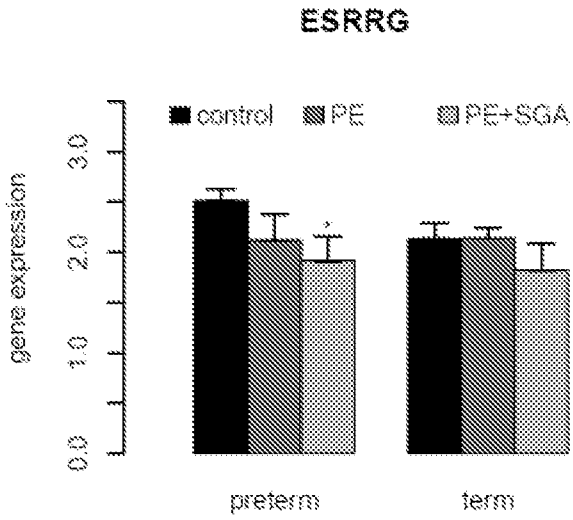

Figure 4P

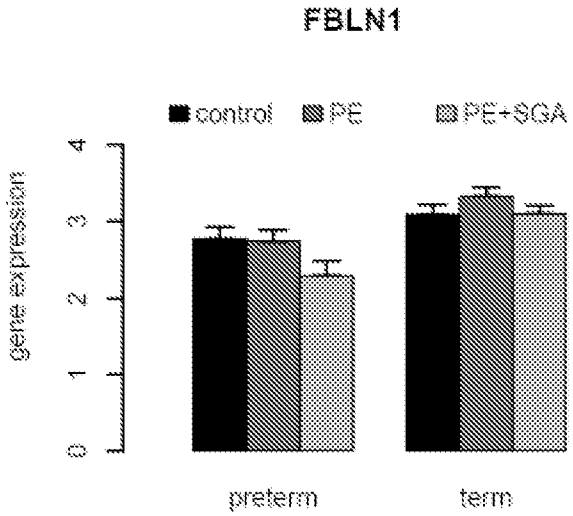

Figure 4Q

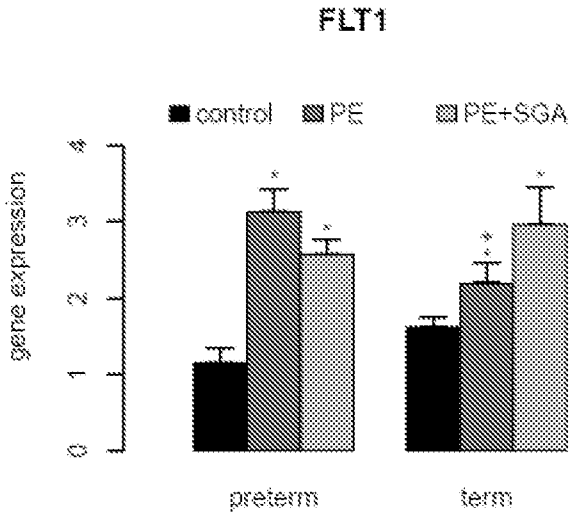

Figure 4R

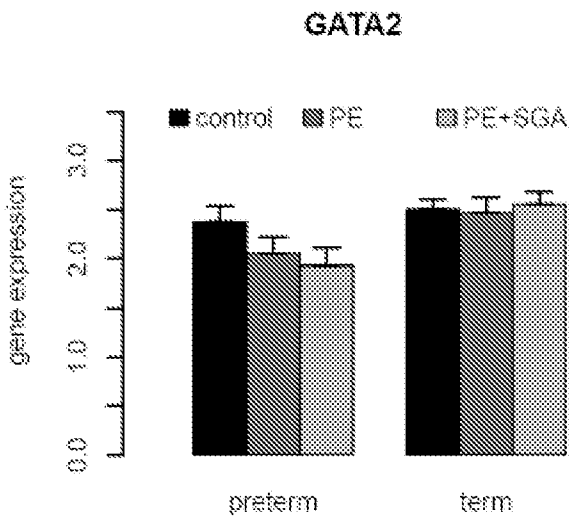

Figure 4S

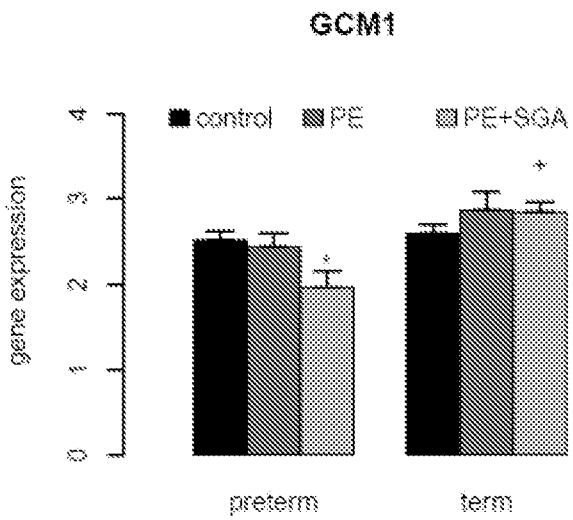

Figure 4T

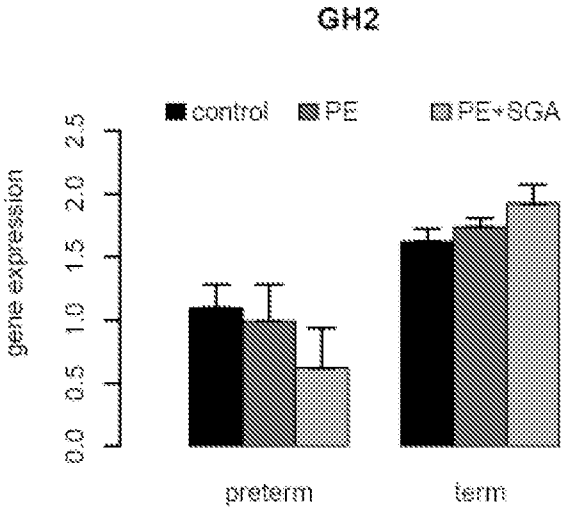

Figure 4U

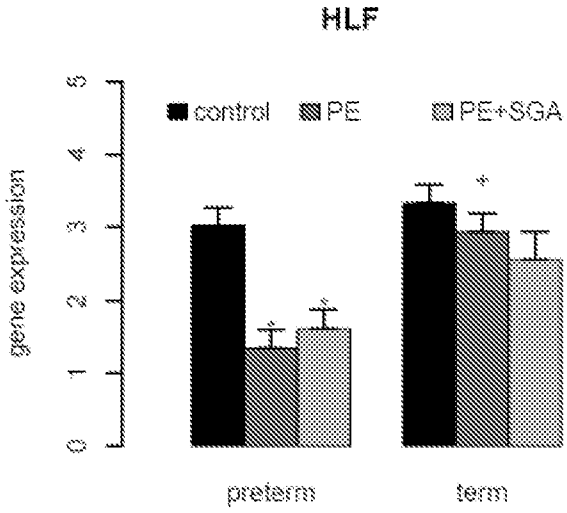

Figure 4V

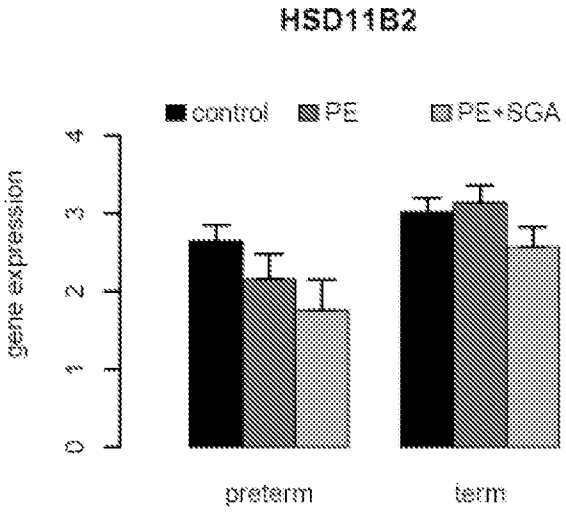

Figure 4W

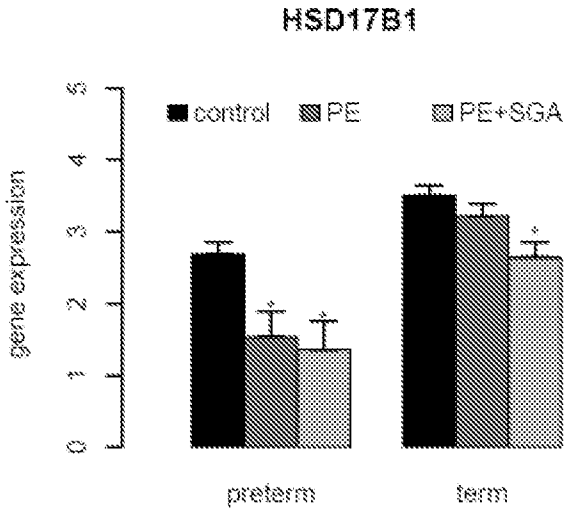

Figure 4X

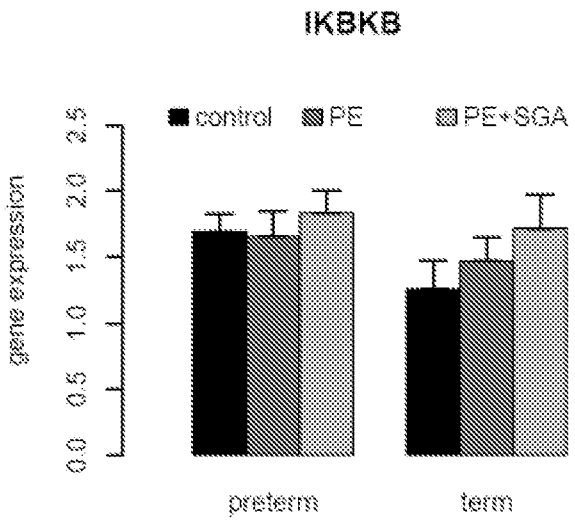

Figure 4Y

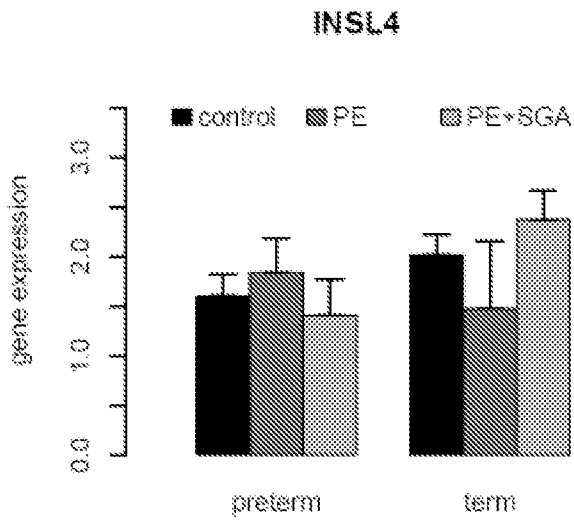

Figure 4Z

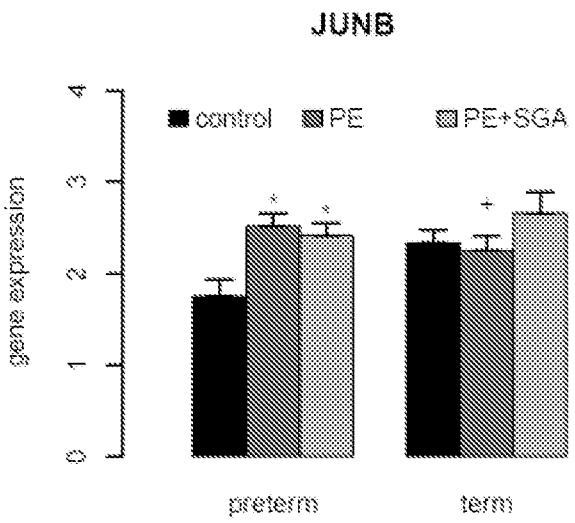

Figure 4AA

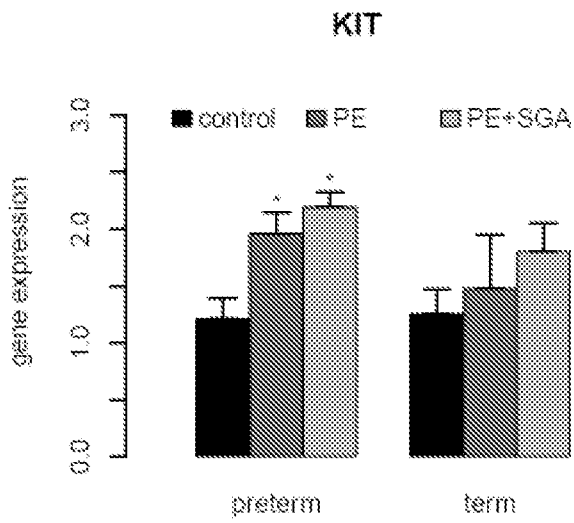

Figure 4BB

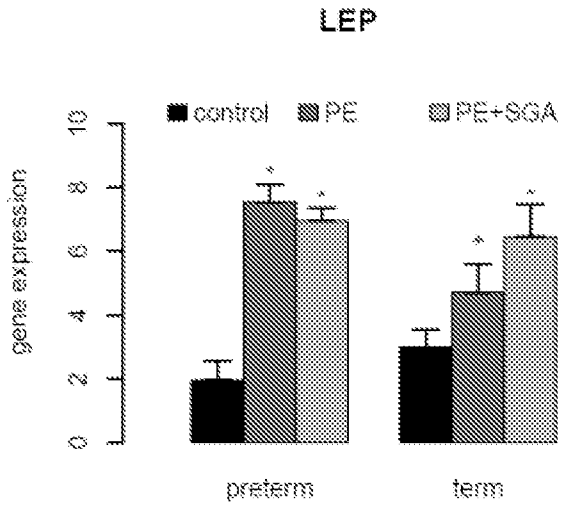

Figure 4CC

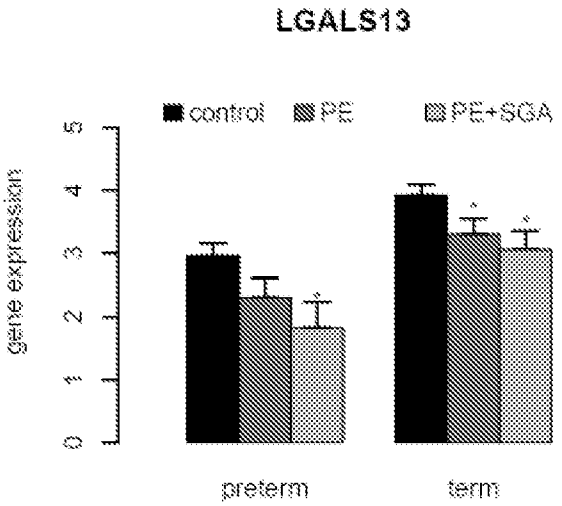

Figure 4DD

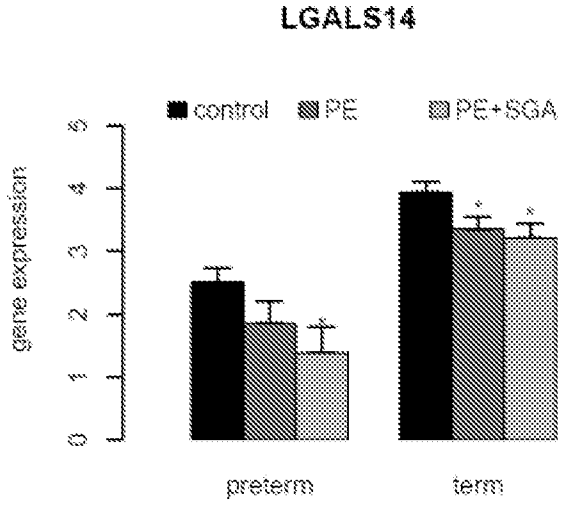

Figure 4EE

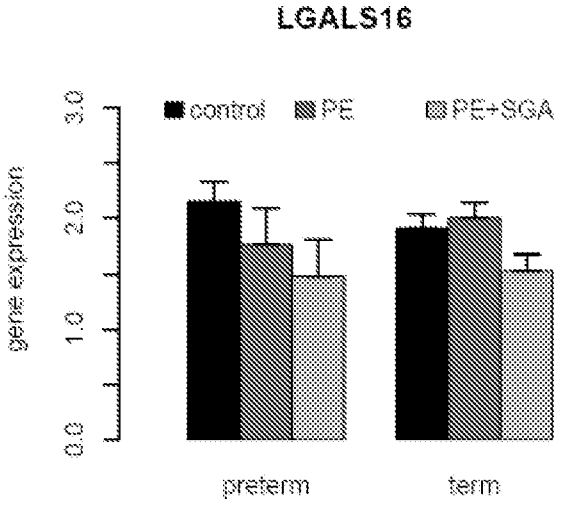

Figure 4FF

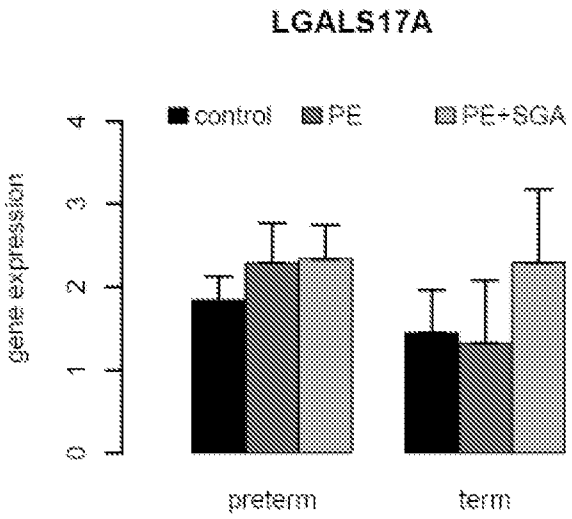

Figure 4GG

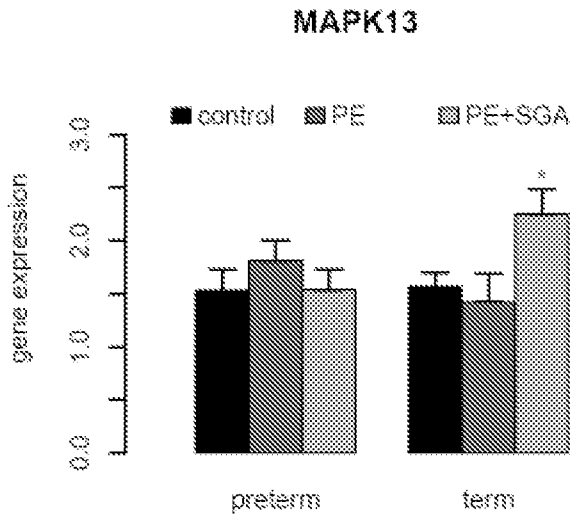

Figure 4HH

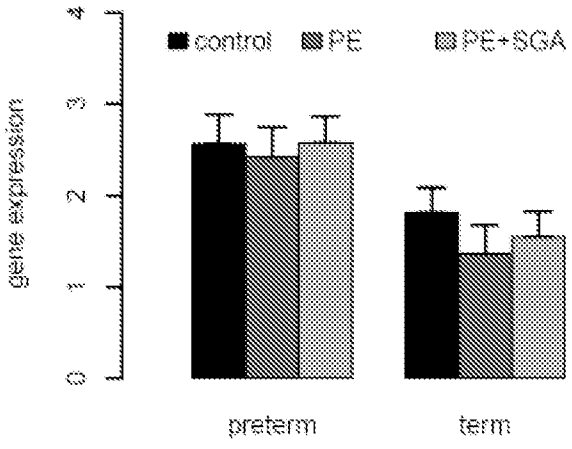

Figure 4II

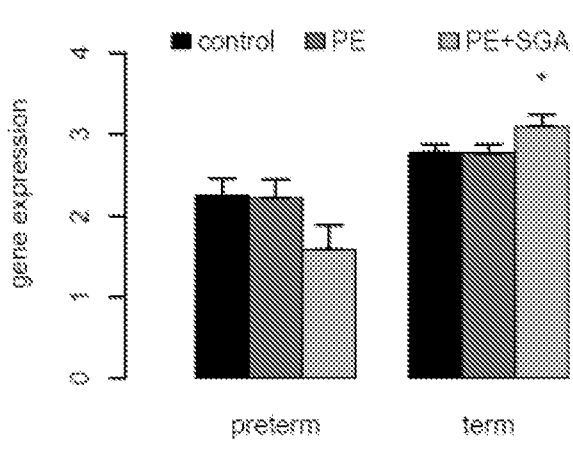

Figure 4JJ

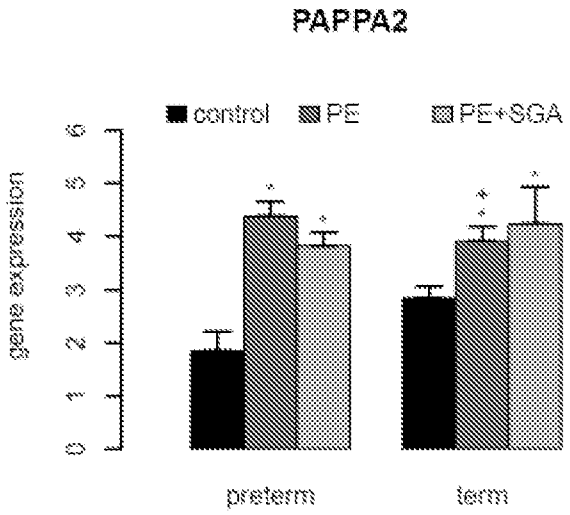

Figure 4KK

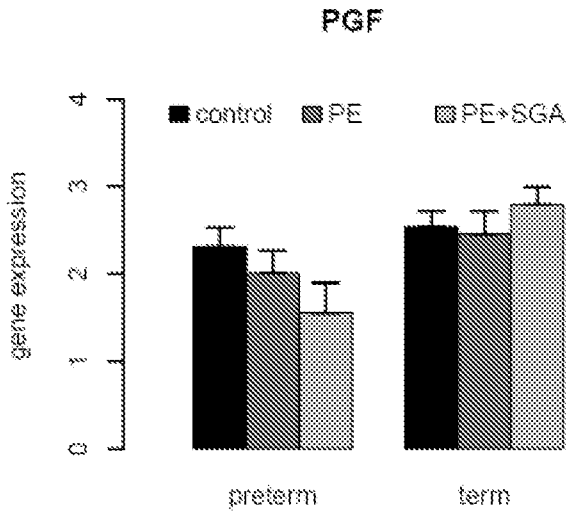

Figure 4LL

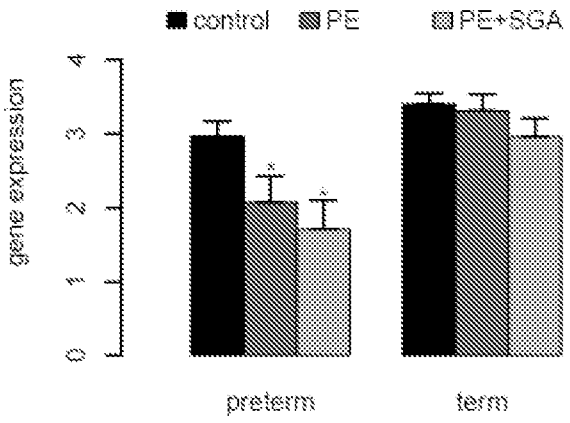

Figure 4MM

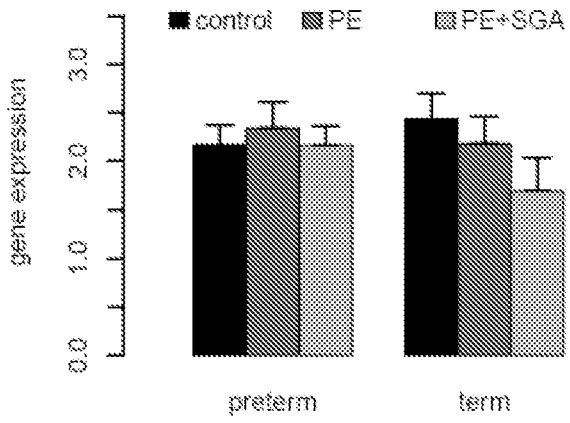

Figure 4NN

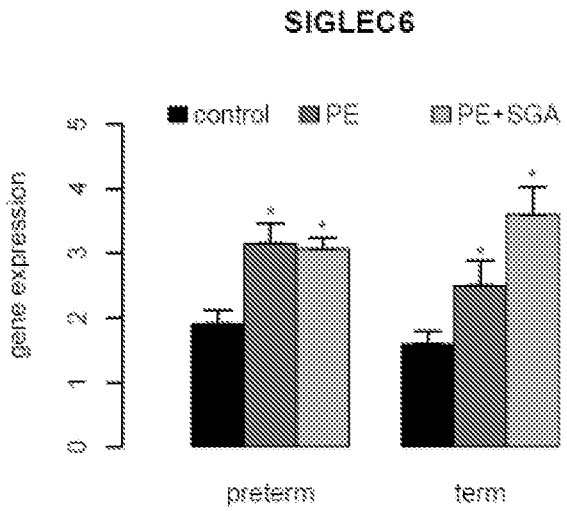

Figure 4OO

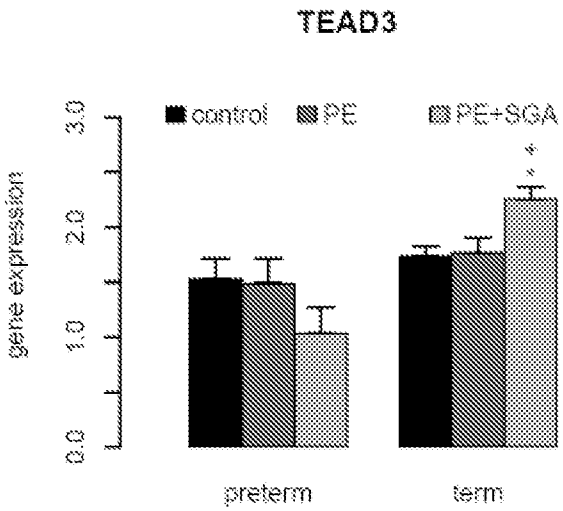

Figure 4PP

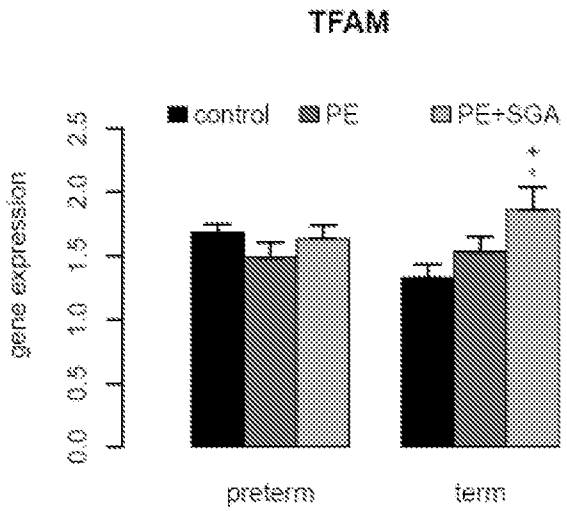

Figure 4QQ

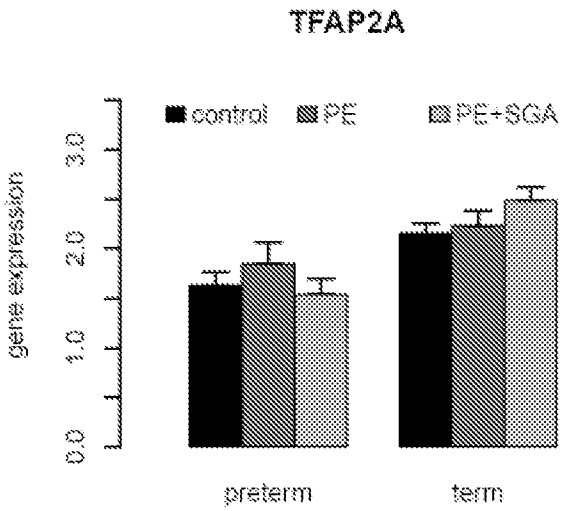

Figure 4RR

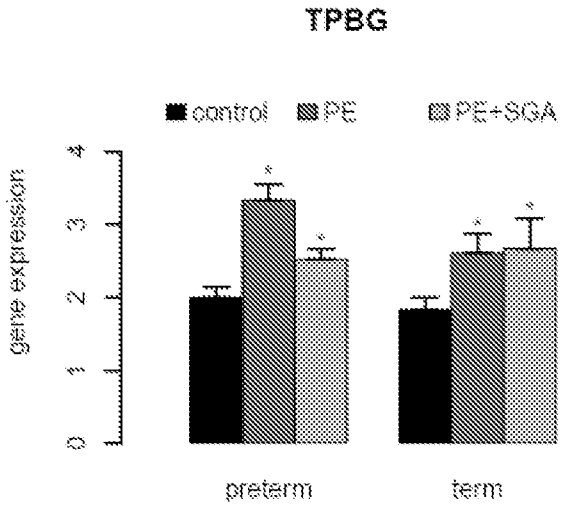

Figure 4SS

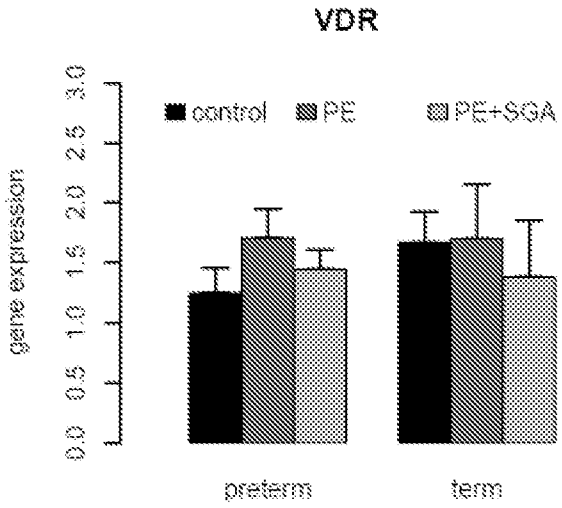

Figure 4TT

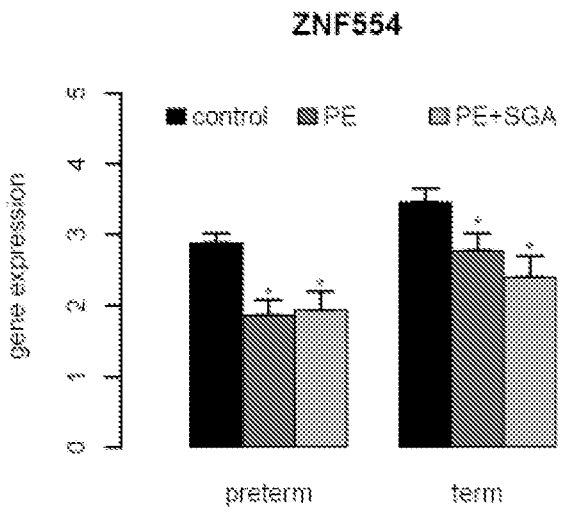

Figure 4UU

专利名称(译)	用于预测或早期检测先兆子痫和/或hellp综合征的生物标志物测试		
公开(公告)号	EP2861989A2	公开(公告)日	2015-04-22
申请号	EP2013803743	申请日	2013-06-13
[标]申请(专利权)人(译)	韦恩州立大学 GENESIS THERANOSTIX KORLATOLT FELELOSSEGU TARSASAG		
申请(专利权)人(译)	韦恩州立大学 塞麦尔维斯大学 GENESIS THERANOSTIX KORLATOLT FELELOSSEGU TARSASAG		
当前申请(专利权)人(译)	韦恩州立大学 塞麦尔维斯大学 GENESIS THERANOSTIX KORLATOLT FELELOSSEGU TARSASAG		
[标]发明人	TARCA ADI L THAN NANDOR JUHASZ GABOR KEKESI ADRIENNA KATALIN MEIRI HAMUTAL PAPP ZOLTAN ROMERO ROBERTO		
发明人	TARCA, ADI, L. THAN, NANDOR JUHASZ, GABOR KEKESI, ADRIENNA, KATALIN MEIRI, HAMUTAL PAPP, ZOLTAN ROMERO, ROBERTO		
IPC分类号	G01N33/53 C12Q1/00 G01N33/68		
CPC分类号	G01N33/689 G01N2800/368 C12Q1/6883 G01N2800/50		
代理机构(译)	Lengyel的，索尔特		
优先权	2012000368 2012-06-15 HU 61/699193 2012-09-10 US		
其他公开文献	EP2861989A4		
外部链接	Espacenet		

摘要(译)

公开了允许早期检测先兆子痫/ HELLP综合征的特定生物标志物。因此，提供了预测孕妇中的先兆子痫的方法。还公开了一种试剂盒，其包含用于测定来自孕妇的样品浓度的特定生物标志物的手段。