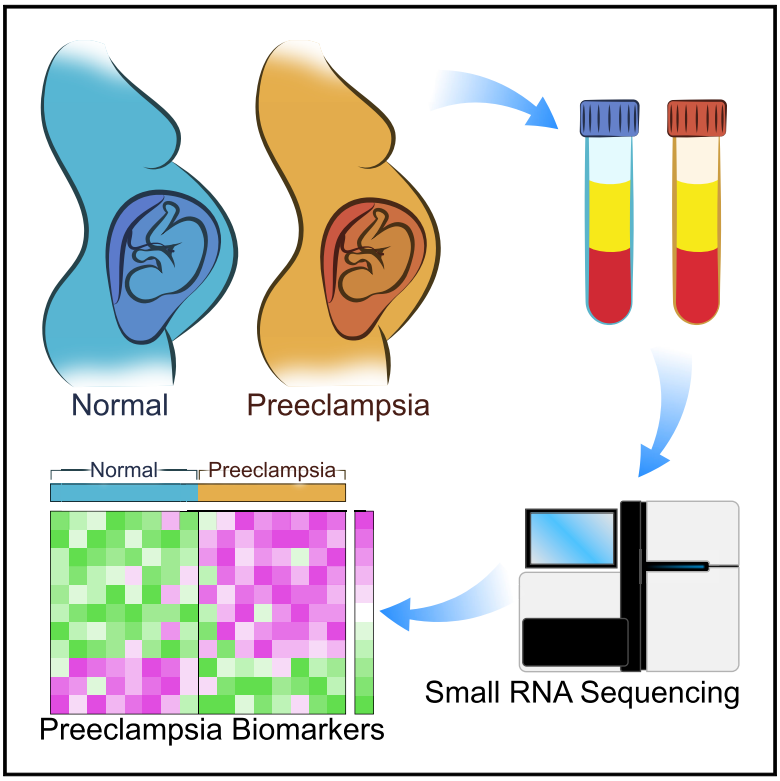
Cell Reports
Medicine

Article

Discovery and Verification of Extracellular miRNA Biomarkers for Non-invasive Prediction of Pre-eclampsia in Asymptomatic Women

Graphical Abstract



Authors

Srimeenakshi Srinivasan, Ryan Treacy, Tiffany Herrero, ..., Julja Burchard, J. Jay Boniface, Louise C. Laurent

Correspondence

llaurent@ucsd.edu

In Brief

Srinivasan et al. use maternal serum to identify and verify miRNA biomarkers in asymptomatic patients at elevated risk for later development of pre-eclampsia in independent cohorts. Biostatistical analyses show that the verification rate for the reversals using bivariate models is markedly higher than for univariate biomarkers.

Highlights

- Small RNA-seq of maternal serum in women who later developed pre-eclampsia
- Bioinformatic analyses identify univariate and bivariate miRNA biomarkers
- Many bivariate biomarkers contain miR-155-5p
- Deconvolution analyses verify several miRNAs to be placenta specific

Srinivasan et al., 2020, Cell Reports Medicine *1*, 100013
May 19, 2020
https://doi.org/10.1016/j.xcrm.2020.100013

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

Discovery and Verification of Extracellular miRNA Biomarkers for Non-invasive Prediction of Pre-eclampsia in Asymptomatic Women

Srimeenakshi Srinivasan,[1,11] Ryan Treacy,[2,11] Tiffany Herrero,[1,3,11] Richelle Olsen,[1,4] Trevor R. Leonardo,[1,5] Xuan Zhang,[6] Peter DeHoff,[1] Cuong To,[1] Lara G. Poling,[1] Aileen Fernando,[1] Sandra Leon-Garcia,[1] Katharine Knepper,[7] Vy Tran,[1] Morgan Meads,[7] Jennifer Tasarz,[1] Aishwarya Vuppala,[1] Soojin Park,[1] Clara D. Laurent,[1] Tony Bui,[1] Pike See Cheah,[6,8] Rachael Tabitha Overcash,[1] Gladys A. Ramos,[1] Hilary Roeder,[1,9] Ionita Ghiran,[10] Mana Parast,[7] The PAPR Study Consortium, Xandra O. Breakefield,[6] Amir J. Lueth,[2] Sharon R. Rust,[2] Max T. Dufford,[2] Angela C. Fox,[2] Durlin E. Hickok,[2] Julja Burchard,[2] J. Jay Boniface,[2] and Louise C. Laurent[1,12,*]

[1]Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Diego, La Jolla, CA 92037, USA
[2]Sera Prognostics, 2749 East Parleys Way, Salt Lake City, UT 84109, USA
[3]Department of Obstetrics and Gynecology, Stanford University, Palo Alto, CA 94305, USA
[4]Franciscan Maternal-Fetal Medicine Associates at St. Joseph, Tacoma, WA 98405, USA
[5]Department of Microbiology and Immunology, University of Illinois at Chicago, Chicago, IL 60607, USA
[6]Neurology and Radiology Services and Program in Neuroscience, Harvard Medical School, Massachusetts General Hospital, Boston, MA 02115, USA
[7]Department of Pathology, University of California, San Diego, La Jolla, CA 92037, USA
[8]Department of Human Anatomy, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor Darul Ehsan, Malaysia
[9]Kaiser Permanente San Diego, San Diego, CA 92120, USA
[10]Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA
[11]These authors contributed equally
[12]Lead Contact
*Correspondence: llaurent@ucsd.edu
https://doi.org/10.1016/j.xcrm.2020.100013

SUMMARY

Development of effective prevention and treatment strategies for pre-eclampsia is limited by the lack of accurate methods for identification of at-risk pregnancies. We performed small RNA sequencing (RNA-seq) of maternal serum extracellular RNAs (exRNAs) to discover and verify microRNAs (miRNAs) differentially expressed in patients who later developed pre-eclampsia. Sera collected from 73 pre-eclampsia cases and 139 controls between 17 and 28 weeks gestational age (GA), divided into separate discovery and verification cohorts, are analyzed by small RNA-seq. Discovery and verification of univariate and bivariate miRNA biomarkers reveal that bivariate biomarkers verify at a markedly higher rate than univariate biomarkers. The majority of verified biomarkers contain miR-155-5p, which has been reported to mediate the pre-eclampsia-associated repression of endothelial nitric oxide synthase (eNOS) by tumor necrosis factor alpha (TNF-α). Deconvolution analysis reveals that several verified miRNA biomarkers come from the placenta and are likely carried by placenta-specific extracellular vesicles.

INTRODUCTION

Placental dysfunction, for which the most common clinical manifestations are pre-eclampsia (PE) and intrauterine growth restriction (IUGR), is an important cause of fetal and maternal morbidity and mortality. Affecting approximately 5% of pregnancies,[1] PE is the second leading cause of maternal mortality[2,3] and the leading cause of medically indicated preterm birth (miPTB) in the United States, accounting for 15% of all PTBs.[4] PE is typically diagnosed by a combination of new-onset hypertension and proteinuria, but severe cases can be associated with maternal end organ damage, including cerebral edema, pulmonary edema, liver or kidney failure, hemolysis, or thrombocytopenia, placental abruption, seizures (eclampsia), or maternal and fetal death. The clinical manifestations of PE become apparent in the second half of pregnancy, but they arise from dysregulation of feto-placental development and/or maternal adaptation to pregnancy in early pregnancy. Low-dose aspirin therapy started between 12 and 28 weeks of gestation has been shown to decrease the risk of PE and IUGR in pregnancies with pre-existing hypertension, pre-existing diabetes, multifetal gestation, renal disease, autoimmune disease, and PE with an adverse pregnancy outcome in a prior pregnancy. For this reason, the US Preventative Task Force (USPTF) has

recommended prophylactic low-dose aspirin in pregnancies with these clinical risk factors for PE.[5] However, the majority of patients who develop PE or IUGR do not have known risk factors, and thus, it is an immediate priority to discover other methods for identification of high-risk pregnancies and to determine whether they would benefit from aspirin prophylaxis.

Early identification of pregnancies that have an elevated risk for developing PE would thus enable customization of prenatal care to incorporate the appropriate intensity of surveillance. It would also allow for selective enrollment of high-risk pregnancies for clinical trials on agents for prevention or treatment of PE. However, current strategies for early prediction of PE are limited by either suboptimal performance and/or clinical feasibility. Current modalities for first- and second-trimester risk assessment involve the assessment of maternal characteristics, measurement of specific analytes in the maternal blood, and sonographic measurement of the uterine artery pulsatility index.[6,7] The highest performing first-trimester risk assessment algorithm was based on a multivariate model incorporating a variety of maternal characteristics (maternal age, weight, height, race, smoking, assisted reproductive technologies, prior pregnancy with PE or small for gestational age [SGA] [<10th percentile], chronic hypertension, diabetes mellitus, lupus, antiphospholipid syndrome, and family history of PE),[1] serum analyte values (pregnancy-associated plasma protein A [PAPP-A] and placental growth factor [PLGF]),[2] mean arterial pressure, uterine artery pulsatility index, and reported detection rates of 95.3% for early (<34 week) PE, 45.6% for late PE, 55.5% for preterm SGA, and 44.3% for term SGA with a false positive rate (FPR) of 10%.[6] Vascular endothelial growth factor (VEGF), soluble FMS-like tyrosine kinase 1 (sFlt-1), and PLGF levels have shown promise as predictive biomarkers in the third trimester, primarily due to their high negative predictive value.[8–10]

Over the past decade, extracellular RNAs (exRNAs) in a variety of biofluids have been shown to have potential value as diagnostic and prognostic biomarkers for a variety of conditions, including cancer, heart disease, neurodegenerative disease, and liver injury (reviewed in Das et al.[11]). In this project, we have built on observations that there is exRNA of feto-placental origin in the maternal circulation,[12–16] suggesting that exRNAs may serve as biomarkers enabling non-invasive interrogation of placental function. Starting in 2011, a number of papers have reported on extracellular microRNA (miRNA) biomarkers associated with PE (Table 1).[17–35] Importantly, only two previous studies verified their initial findings in an independent cohort. One of these studies collected all samples after diagnosis, with the small discovery cohort (8 cases and 4 controls) being analyzed by small RNA sequencing (RNA-seq) and the verification cohort (38 cases and 32 controls) being analyzed by qRT-PCR.[22] In the other study, the discovery cohort samples (28 cases and 26 controls) were collected after diagnosis and analyzed by small RNA-seq and the verification cohort samples (only 6 cases and 10 controls) were collected pre-symptomatically and analyzed by qRT-PCR.[29]

This study was aimed at discovery and verification of extracellular miRNA predictors for PE. Cases and controls were selected from two studies in which maternal serum was collected from asymptomatic women between 17 and 28 weeks gestation and clinical outcomes were assessed after delivery. Cases and controls were divided into adequately sized discovery (49 cases and 92 controls) and verification (24 cases and 47 controls) cohorts. Small RNA-seq was used for biomarker discovery and verification, and univariate (single miRNA) and bivariate (ratios of pairs of miRNAs, also termed reversals) biomarkers were investigated. Key aspects of the study design were that discovery and verification were performed on independent sets of subjects and that the investigators who developed univariate and bivariate models were blinded to the clinical outcomes of subjects in the verification set. Candidate models were locked before verification analysis.

We were able to discover and verify both univariate extracellular miRNAs biomarkers and bivariate reversals that identified asymptomatic patients at elevated risk for later development of PE. The verification rate for reversals was markedly higher than for univariate biomarkers, indicating that the use of reversals may confer a degree of internal normalization that increases robustness.

RESULTS

Maternal serum samples were collected as part of two studies: the Placenta Study at UCSD and the PAPR Study from Sera Prognostics. The Placenta Study was a single-site, high-risk study that enrolled pregnant women with at least one risk factor for placental dysfunction, and the PAPR Study was a multi-site study that enrolled pregnant women without regard to risk factors for placental dysfunction (Figure S1A). For both studies, subjects were enrolled, maternal serum was collected between 17 and 28 weeks, and outcomes were obtained after delivery. Nineteen cases and 29 controls were selected from the Placenta Study, and 54 cases and 110 controls were selected from the PAPR study (selection criteria are listed in Figure S1B). The overall structure of the study is shown in Figure 1A. As described in detail in the STAR Methods, all of the Placenta Study samples were unblinded and included in the discovery cohort, and the PAPR Study samples were divided between the discovery and verification cohorts in such a way that the discovery cohort contained 141 subjects and the verification cohort contained 71 subjects, with similar distributions of cases and controls (Figure 1B) and gestational age at blood draw (GABD) (Figure 1C) in the two cohorts. There were no significant demographic or clinical differences between the discovery and verification cohorts (Table 2). As expected, in both cohorts, there was an earlier median gestational age of delivery and lower mean birthweight in the cases compared to controls. Of the other demographic or clinical variables, only BMI showed a significant difference between cases and controls (Table 2), with the cases having a higher BMI.

Given the potential for gestational age (GA)-dependent effects on expression of miRNAs, biomarker discovery and verification were each performed on the entire GA range, as well as for three GA windows: 17 weeks 0 days–21 weeks 5 days (early); 19 weeks 5 days–24 weeks 4 days (middle); and 22 weeks 2 days–28 weeks 0 days (late).

Discovery and Verification of Univariate and Bivariate Predictors

In this study, we aimed to discover and verify individual univariate and bivariate (also termed reversals) predictors with

Table 1. Previously Published Studies Reporting Extracellular miRNA Biomarkers for Prediction or Diagnosis of Pre-eclampsia

Study and Population Origin	Biofluid	exRNA Isolation Method	exRNA Measurement Method	N; GA Sample Collection	Time Frame	Sensitivity, Specificity, AUC	miRNA PE > Control	miRNA PE < Control
Yang et al.[33] China	serum	mirVana (Ambion)	small RNA-seq (SOLiD, Applied Biosystems)	5 (4 pree [2 mild, 2 severe], 1 control)	third trimester, after diagnosis	–,–,–	hsa-miR-521, 520h, 517c-3p, 519d, 520 g, 517b-3p, 542-3p, 136, 518e, 125b, 125a-5p, 519a, 29a, let-7f-3p, 7a-3p	hsa-miR-223, 1260, 320c, 185, 1272, let-7f-5p, 7d-5p
Wu et al.[31] China (Han)	plasma	mirVana PARIS (Ambion)	miRNA microarray, 821 miRNAs (Agilent)	9 (5 severe PE, 4 normal)	37 and 40 weeks, after diagnosis	–,–,–, $p < 0.05$, fold-change ≥ 2	miR-574-5p, 26a,151-3p, 130a, 181a, 130b, 30d, 145, 103, 425, 221, 342-3p, and 24	miR-144 and 16
Hromadnikova et al.[20] Czech Republic	plasma	mirVana (Ambion)	qRT-PCR, 7 C19MC miRNAs	168 (24 mild PE, 39 severe PE, 27 FGR, 23 ghtn, 55 normal)	after diagnosis	–,–,–	hsa-miR-516-5p, 517-5p, 520-5p, 525-5p, 526a	
Li et al.[22] China	plasma	mirVanaTM miRNA Isolation Kit (Ambion)	discovery: small RNA-seq (SOLiD, Applied Biosystems) validation: qRT-PCR	discovery: 4 mild PE, 4 severe PE, 4 normal. validation: 16 mild PE, 22 severe PE, 32 normal.	34–37 weeks, after diagnosis	–,–,–	miR-141 and miR-29a (mild PE versus control)	miR-144 (mild PE and severe PE versus control)
Luque et al.[23] Barcelona	serum	miRNeasy Mini kit (QIAGEN)	TaqMan OpenArray Human MicroRNA Panel, 754 miRNAs, (Applied Biosystems)	31 PE, 44 normal	11–136 weeks, asymptomatic	–,–,–	miRNAs: miR-192, 143, and -125b	miR-127, -942, -126*, and -221, did not validate
Xu et al.[32] China	placenta and plasma	not given	placenta: mammalian miRNA chip array, 435 human miRNAs (V2.0, Capitalbio). Plasma: TaqMan qPCR.	53 (20 severe PE, 33 normal)	15–18 and 35–38 weeks, asymptomatic	–,–,–	miR-210	miR-18a, miR-19b1, and miR-92a1
Stubert et al.[28] Germany	serum	QIAamp Circulating Nucleic Acid Kit (QIAGEN)	TaqMan Low Density Array Human MicroRNA Card Set v3.0, 754 miRNAs (Applied Biosystems)	26 (13 HELLP, 13 controls)	28–41 weeks, at diagnosis	miR-122e: AUC = 0.82	miR-122, miR-758, and miR-133a (>2-fold)	miR-573, miR-766, and miR-409-3p
Ura et al.[30] Italy	sera	mirVana PARIS with 2× acid-phenol/ chloroform extraction	TaqMan Low Density Array Human MicroRNA Card Set v3.0, 754 miRNAs (Applied Biosystems)	48 (24 severe PE, 24 control)	12–14 weeks, asymptomatic	validated miR-1233, miR-520a, miR-210, miR-14	miR-1233; miR-650; miR-520a; miR-215; miR-210; miR-25; miR-518b; miR-193a-3p; miR-32; miR-204; miR-296-5p; miR-15	miR-126; miR-335; miR-144; miR-204; miR-668; miR-376a; miR-15b

(*Continued on next page*)

CellPress
OPEN ACCESS

Table 1. *Continued*

Study and Population Origin	Biofluid	exRNA Isolation Method	exRNA Measurement Method	N; GA Sample Collection	Time Frame	Sensitivity, Specificity, AUC	miRNA PE > Control	miRNA PE < Control
Yang et al.[34] China	plasma and placenta	TRIzol	small RNA-seq (SOLiD, Applied Biosystems)	5 (2 mild PE, 2 severe PE, 1 control)	244–283 days, at diagnosis (plasma) and delivery (placenta)	–,–,–	miR-126, miR-126*, miR-130a, miR-135b, miR-142-3p, miR-149, miR-188-5p, miR-18a, miR-18b, miR-203, miR-205, miR-224, miR-27a, miR-29a, miR-301a, miR-517c, miR-518-3p, miR-518e, miR-519d and miR-93	
Miura et al.[25] Japan	plasma	mirVana (Ambion)	qRT-PCR of 10 C19MC miRNAs	40 (20 cases [severe]: 20 controls)	27–34 weeks, after diagnosis	–,–,–	hsa-miR-518b, 1323, 516b, 516a-5p, 525-5p, 515-5p, 520h, 520a-5p, 519d, 526b	
Jairajpuri et al.[21] Bahrain	plasma	miRNeasy (QIAGEN)	custom miScript miRNA PCR array, 84 miRNAs (QIAGEN)	22 (15 cases [7 mild, 8 severe]: 7 controls)	third trimester, at delivery	–,–,–	hsa-miR- 215, 155, 650, 210, 21, 518b, 29a	hsa-miR-15b, 144, 18a, 19b1
Yoffe et al.[35] UK	plasma	miRNeasy (QIAGEN)	small RNA-seq (TruSeq, Illumina)	75 (35 cases [early onset]: 40 controls)	pre-symptomatic, 11 weeks 0 days to 13 weeks 6 days	0.72, 0.8, 0.86	hsa-miR-4433b, 221, hsa-let-7g	hsa-miR-182, 10b, 25, 99b, 143, 151a, 191, 146b, 486
Salomon et al.[27] Chile	plasma	miRNeasy Mini Kit (QIAGEN) from exosomes isolated by density gradient ultracentrifugation	small RNA-seq (TruSeq, Illumina)	47 (15 PE, 32 control)	11–14 weeks, 22–24 weeks, 32–36 weeks	–,–,–, $p < 0.05$	hsa-miR-486-5p, 423-5p, 451a, 107, 15a-5p, 335-5p, 92a-3p, 103a-3p	hsa-miR-126-3p
Gunel et al.[18] Turkey	plasma, placenta	mirVana miRNA Isolation kit (Ambion) plasma	G4870A SurePrint G3 Human v16 miRNA 8 × 60 K Microarray, 1368 miRNAs (Agilent)	36 (18 severe PE at 32 and 34 weeks, 18 control at 37 and 40 weeks)	32–40 weeks, at delivery and diagnosis	–,–,–	hsa-miR-877*, hsa-miR-118	hsa-miR-1539, hsa-let-7b*, hsa-miR-191*, hsa-miR-23c, hsa-miR-33b*, hsa-miR-425*, hsa-miR-4313, hsa-miR-550a, hsa-miR-933, hsa -miR-7f-1*
Gan et al.[17] China	plasma, urine cells	Trizol LS reagent (Invitrogen)	qRT-PCR, miR-210, miR-155, miR-125b-5p, miR-125a-5p	40 (20 PE, 20 control)	20–34 weeks, at diagnosis	miR-210: AUC = 0.750; miR-155: AUC = 0.703	miR-210 and miR-155 in plasma	

(*Continued on next page*)

Table 1. *Continued*

Study and Population Origin	Biofluid	exRNA Isolation Method	exRNA Measurement Method	N; GA Sample Collection	Time Frame	Sensitivity, Specificity, AUC	miRNA PE > Control	miRNA PE < Control
Timofeeva et al.[29] Russia	plasma	miRNeasy (QIAGEN)	small RNA-seq (NEBNext, NEB) on 12 cases from cohort 1; qRT-PCR on cohort 2	cohort 1– 54 (28 cases [16 early onset, 12 late onset]: 26 controls [10 < 34 weeks, 16 > 37 weeks]); at delivery. cohort 2–16 (6 cases: 10 controls); 11–13, 24–26, 30–32 weeks.		hsa-miR-423-5p: Sensitivity = 0.875; Specificity = 0.80; AUC = 0.844	hsa-miR-423-5p, 519a-3p, 629-5p, hsa-let-7c-5p cohort 1; hsa-miR-423-5p cohort 2 at 11–13 weeks	
Hromadnikova et al.[19] Czech Republic	plasma	mirVana (Ambion)	qRT-PCR of 6 C19MC miRNAs	97 (39 cases [21 pre-eclampsia, 18 IUGR]: 58 controls)	10–13 weeks	hsa-miR-517-5p: Sensitivity = 0.429; Specificity = 0.862; $p < 0.05$	hsa-miR-517-5p, 518b, 520h	
Martinez-Fierro et al.[24] Mexico	serum	miRNeasy Mini Kit (QIAGEN	TaqMan Low Density Array Human MicroRNA Card Set v2.0, focused on 51 C19MC miRNAs (Applied Biosystems)	34 (16 PE, 18 control)	12, 16, and/or 20 weeks, asymptomatic, longitudinal	–,–,–, $p < 0.05$	hsa-miR-520c-3p (16 weeks); hsa-miR-512-3p, 518f-3p, 520d-3p (20 weeks)	
Motawi et al.[26] Egypt	plasma	miRNeasy Mini Kit (QIAGEN) from exosomes isolated by ultracentrifugation	TaqMan miRNA qPCR of hsa-miR-136, 494, and 495	200 (100 PE [23 < 20 weeks, 77 > 20 weeks], 100 control [20 < 20 weeks], 80 > 20 weeks)	at diagnosis	<20 weeks: hsa-miR-136 0.95, 1.00, 1.000; hsa-miR-494 0.86, 0.95, 0.868; hsa-miR-495 0.90, 0.83, 0.940, $p < 0.05$	<20 weeks and >20 weeks: hsa-miR-136, 494 and 495	

List of nineteen previously published studies with key characteristics and findings. Column "Sensitivity, Specificity, AUC" lists information where available; "–,–,–" is listed for studies that did not report sensitivity, specificity, or AUC. See also Tables S1 and S2.

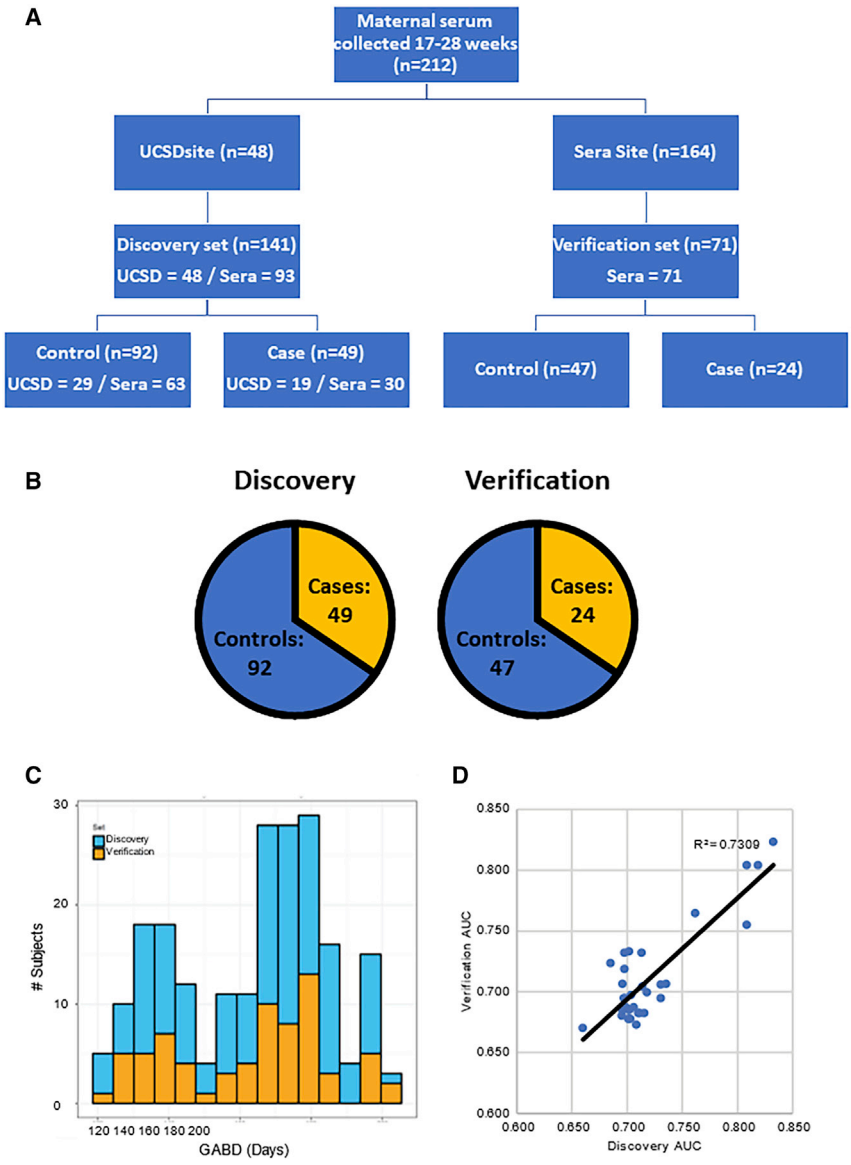


Figure 1. Study Design and Characteristics of Discovery and Verification Sets

(A) Overall study design. Maternal serum samples were collected at both UCSD site (n = 48) and Sera Prognostics site (n = 164). The samples were split into a discovery set (n = 141) and verification set (n = 71) with matching gestational ages and disease severity. Within each set, there were roughly 2× the number of controls as cases.

(B) Distribution of cases (49 in discovery and 24 in verification) and controls (92 in discovery and 24 in verification).

(C) Distribution of GABD.

(D) Correlation between discovery and verification AUCs. Scatterplot of the discovery and verification AUCs for candidate predictors for pre-eclampsia is shown. Best-fit linear trendline and r^2 value are shown.

See also Figure S1.

significant areas under the curve (AUCs) (lower 95% confidence intervals [CIs] did not include 0.5). In a future verification/validation study, we plan to test different combinations of multiple verified predictors and clinical parameters to identify the best-performing test for clinical use.

Data filtering and processing are described in detail in the STAR Methods section. Briefly, the small RNA-seq data were mapped using the exceRpt pipeline,[36] and the resulting miRNA data were filtered to remove miRNAs with >70% missing values; the raw counts for the remaining miRNAs are provided in Table S2, sheet 1. Sample-to-sample normalization was carried out through stabilization of variance and reduction in bias across distributions of read counts. Batch normalization was carried out using the PEER package.[37] Normalized and batch-corrected data are provided in Table S2, sheet 2.

In the discovery phase of the univariate analysis, individual miRNAs with single test significance ($p < 0.05$) were selected for each GABD window, resulting in identification of 14 individual miRNAs for the entire GABD window, 14 for the early GABD window, 11 for the middle GABD window, and 21 for the late GABD window (Tables S3 and S4, sheet 1). In the verification phase of the univariate analysis, miRNAs for which the lower 95% CI of the AUC was >0.5 were considered to have passed verification. This analysis identified only two candidate individual miRNAs that passed verification: one for the early GA window (hsa-miR-516b-5p) and one for the middle GA window (hsa-miR-941; Tables 3, S3, and S4, sheet 1). This verification rate is not unexpected, given the size of the verification set.

Normalization of extracellular miRNA datasets has proven to be challenging. Standard normalization approaches, such as the use of spike-in synthetic oligonucleotides, ''housekeeping'' small RNAs, or bioinformatic methods commonly used for cellular long RNA-seq datasets, have not been successful. Even for studies of miRNAs in cells and tissues, it has been advocated that sample-set-specific normalizers be used,[38] and it is commonly accepted that normalization of exRNA datasets is even more challenging.[39] We reasoned that, for pairs of endogenous miRNAs, the expression of each miRNA might serve as an endogenous control for the other and therefore produce more reproducible features than the abundances measured for individual miRNAs. We implemented this paired normalization approach by forming ratios of individual miRNA abundances, termed ''reversals,'' by adapting the method described in Price et al.[40]. In pregnancy, this approach has been applied to the development of prognostic biomarkers of spontaneous preterm birth and PE based on serum protein abundances.[41,42] After log

Table 2. Study Subject Demographics and Clinical Characteristics of 141 Patients in the Discovery Cohort and 71 Patients in the Verification Cohort

	Discovery Case (n = 49)	Discovery Control (n = 92)	p Value Discovery Case versus Control	Verification Case (n = 24)	Verification Control (n = 47)	p Value Verification Case versus Control	p Value Discovery versus Verification
Mean maternal age (year)	30.1 ± 6.4	29.0 ± 6.1	0.29	26.8 ± 6.7	27.7 ± 6.1	0.59	0.88
Median gravidity	3 ± 1.7	2 ± 1.7	0.24	2 ± 2.5	2 ± 1.0	0.72	0.85
Median parity	1 ± 1.1	1 ± 1.5	0.93	1 ± 2.3	1 ± 1.2	0.75	0.96
Mean BMI	30.7 ± 10.1	26.1 ± 7.8	0.02	33.7 ± 9.6	27.6 ± 7.5	0.01	0.95
No. subjects with diabetes	10 (20.4%)	9 (9.8%)	0.12	5 (20.8%)	5 (10.6%)	0.29	1.0
Race/ethnicity			0.40			0.57	0.11
White – non-Hispanic	24 (49.0%)	50 (54.3%)		7 (29.2%)	20 (42.6%)		
Hispanic	12 (24.5%)	25 (27.2%)		13 (54.2%)	17 (36.2%)		
African-American	7 (14.3%)	9 (9.8%)		3 (12.5%)	8 (17.0%)		
Asian	4 (8.2%)	3 (3.3%)		0 (0.0%)	1(2.1%)		
Pacific Islander	1 (2.0%)	0 (0.0%)		0 (0.0%)	0 (0.0%)		
Other	1 (2.0%)	5 (5.4%)		1 (4.2%)	1 (2.1%)		
Median GABD (week)	24.3 ± 3.0	24.0 ± 3.1	0.31	24.1 ± 2.9	23.5 ± 3.3	0.39	0.71
Median GA delivery (week)	36.3 ± 6.4	39.2 ± 1.2	<0.01	36.2 ± 3.2	39.3 ± 2.5	<0.01	0.67
No. preterm deliveries	27 (55.1%)	3 (3.3%)	<0.01	15 (60%)	3 (6.5%)	<0.01	0.49
Mean birthweight (g)	2,589.5 ± 922.7	3,464.8 ± 465.4	<0.01	2,733.6 ± 893.5	3,213.0 ± 675.9	0.03	0.87
No. IUGR	11 (7.9%)	none[a]	NA	5 (7.0%)	none[a]	NA	1.0

[a]As IUGR was an inclusion criterion for the case group, there were no subjects with IUGR in the control group.

transformation, the miRNA abundance data approximated a normal distribution, enabling assessment of relative expression in an arithmetic, geometric, or power relationship. Geometric relationships (calculated as log ratios) are commonly used to generate models from biological data, and power relationships (calculated as ratios of normally distributed values) are a feature of risk analysis[43] and have been used successfully in analysis of cDNA microarrays.[44] We therefore decided to evaluate the performance of these two types of data transformation on our dataset. For the discovery cohort, we compared the results obtained when the reversals were constructed as the log values of the ratios of normalized counts (geometric) and the ratios of the log values (power). We found that the latter was preferred because it resulted in better separation of cases and controls, as visualized in the first two principal components of a principal-component analysis (PCA) (Figures S2A and S2B). The reversals constructed by the ratios of the log values also resulted in increased stability and magnitude of performance in least absolute shrinkage and selection operator (LASSO) analysis. Thus, we used this approach for generation of bivariate features (reversals) for both the discovery and verification portions of our study.

Reversals were selected by ranking performance in bootstrapped resampling with replacement, as detailed in the STAR Methods section. Briefly, we derived five ranks from the following statistics, each computed across 1,000 iterations of cross-validation: (1) the mean of the cross-validation AUCs; (2) the lower 25th percentile of the cross-validation AUCs; (3) the mean of the squared Pearson correlation coefficient between the reversal scores with diagnosis of PE case (1) or control (0); (4) the lower 25th percentile of the squared Pearson correlation coefficient between the reversal scores and the diagnosis of PE case (1) or control (0); and (5) the square of the differences in the case mean and control mean reversal scores (i.e., the squared mean shift). Each rank was then inverted, and all ranks were summed for each reversal to obtain the final ranking.

For the bivariate analysis, the top 50 reversals for each GABD window were selected for testing in the verification phase, in which those for which the lower 95% CI AUC > 0.5 were considered to have passed verification (Table S4, sheet 2). This analysis identified one reversal in the full GABD window, four in the early window, two in the middle window, and 23 in the late window that passed verification in the same window (Tables 3, S3, and S4, sheet 2). The verification rates are not unexpected, given the size of the verification set, with the exception of an unusually high rate of verification in the late window.

Overall, we observed that the verification rate was markedly higher for the late GABD window reversals (23 verified out of 50 reversals identified in discovery) compared to both the univariate predictors (2 verified out of 14 identified in discovery across all GABD windows) and the reversal of the other three GABD windows (1, 4, and 2 out of 50 for the full, early, and middle windows, respectively; Table S3). We attribute the superior performance of reversals compared to univariate predictors to the "internal normalization" gained by using a ratio of values for a pair of miRNAs measured in the same sample, which would be expected to minimize the technical variability that might be introduced during sample collection, processing, storage, and analysis. The late GABD window may have the best performance because of the larger number of samples (compared to the early and middle GABD windows) and lower GA-dependent biological

Table 3. Univariate Predictors and Reversals Selected in Discovery and Confirmed in Blinded Verification

Univariate/ Bivariate	Numerator	Denominator	GABD Window	Discovery Mean AUC—Full	Discovery Mean AUC—Early	Discovery Mean AUC—Middle	Discovery Mean AUC—Late	Verification Mean AUC—Full	Verification Mean AUC—Early
B	hsa-miR-127-3p	hsa-miR-485-5p	F	0.66[a]	0.57	0.61[a]	0.70[a]	0.67[b]	0.70[b]
B	hsa-miR-4732-3p	hsa-miR-941	E	0.55	0.81[a]	0.57	0.58	0.53	0.80[a]
B	hsa-miR-1273h-3p	hsa-miR-3173-5p	E	0.57	0.83[a]	0.57	0.53	0.62[a]	0.82[a]
B	hsa-miR-155-5p	hsa-miR-3173-5p	E	0.56	0.76[a]	0.55	0.50	0.58	0.76[a]
B	hsa-miR-150-3p	hsa-miR-193b-5p	M	0.63[a]	0.76[a]	0.73[a]	0.58	0.70[a]	0.58
B	hsa-miR-1285-3p	hsa-mir-378c	M	0.56	0.61[a]	0.70[a]	0.54	0.65[b]	0.63[b]
B	hsa-miR-4732-3p	hsa-miR-381-3p	E	0.53	0.82[a]	0.63[a]	0.57	0.50	0.80[a]
B	hsa-miR-320b	hsa-miR-155-5p	L	0.58	0.62[a]	0.60[a]	0.70[a]	0.54	0.70[a]
B	hsa-miR-181a-5p	hsa-miR-155-5p	L	0.59	0.58	0.59	0.70[a]	0.52	0.70[a]
B	hsa-miR-26b-5p	hsa-miR-155-5p	L	0.59	0.53	0.61[a]	0.69[a]	0.55	0.65[a]
B	hsa-let-7g-5p	hsa-miR-155-5p	L	0.59	0.58	0.60	0.70[a]	0.53	0.69[a]
B	hsa-miR-4443	hsa-miR-155-5p	L	0.57	0.64[a]	0.56	0.72[a]	0.53	0.73[a]
B	hsa-miR-425-5p	hsa-miR-155-5p	L	0.58	0.60	0.58	0.70[a]	0.54	0.69[a]
B	hsa-miR-146a-5p	hsa-miR-155-5p	L	0.57	0.63[a]	0.54	0.70[a]	0.57	0.61[a]
B	hsa-miR-25-3p	hsa-miR-155-5p	L	0.59	0.57	0.57	0.71[a]	0.52	0.75[a]
B	hsa-miR-151a-3p	hsa-miR-155-5p	L	0.58	0.63[a]	0.57	0.71[a]	0.54	0.70[a]
B	hsa-miR-320a	hsa-miR-155-5p	L	0.58	0.62[a]	0.61[a]	0.71[a]	0.54	0.69[a]
B	hsa-miR-30d-5p	hsa-miR-155-5p	L	0.58	0.64[a]	0.56	0.71[a]	0.54	0.71[a]
B	hsa-miR-126-3p	hsa-miR-155-5p	L	0.58	0.62[a]	0.57	0.70[a]	0.55	0.67[a]
B	hsa-miR-146b-5p	hsa-miR-155-5p	L	0.57	0.64[a]	0.55	0.70[a]	0.56	0.63[a]
B	hsa-let-7i-5p	hsa-miR-155-5p	L	0.60[a]	0.60[a]	0.61[a]	0.74[a]	0.55	0.69[a]
B	hsa-miR-26a-5p	hsa-miR-155-5p	L	0.56	0.65[a]	0.56	0.70[a]	0.54	0.69[a]
B	hsa-miR-625-3p	hsa-miR-155-5p	L	0.58	0.64[a]	0.59	0.73[a]	0.54	0.67[a]
B	hsa-miR-423-5p	hsa-miR-155-5p	L	0.59	0.60	0.59	0.70[a]	0.55	0.73[a]
B	hsa-miR-451a	hsa-miR-155-5p	L	0.61[a]	0.57	0.60[a]	0.72[a]	0.54	0.69[a]
B	hsa-miR-125a-5p	hsa-miR-155-5p	L	0.58	0.63[a]	0.61[a]	0.71[a]	0.51	0.75[a]
B	hsa-miR-99a-5p	hsa-miR-155-5p	L	0.58	0.60[a]	0.56	0.70[a]	0.56	0.68[a]
B	hsa-let-7b-5p	hsa-miR-155-5p	L	0.60	0.60	0.61[a]	0.71[a]	0.54	0.75[a]
B	hsa-miR-516b-5p	hsa-miR-155-5p	L	0.55	0.72[a]	0.47	0.70[a]	0.54	0.72[a]
B	hsa-miR-363-3p	hsa-miR-155-5p	L	0.59	0.58	0.61[a]	0.70[a]	0.53	0.73[a]
U	hsa-miR-516b-5p		E	0.49	0.81[a]	0.64[a]	0.59	0.54	0.75[a]
U	hsa-miR-941		M	0.65	0.69[a]	0.70[a]	0.64[a]	0.56	0.49

Information pertaining to the GABD window, performance (area under the curve [AUC]) in the discovery and verification cohorts, tissue sources (calculated), and associated carrier subclasses (calculated) is listed. Mean AUCs for discovery and verification sets for each GABD window. Chromosome: chromosome on which each miRNA is encoded. miRNA Cluster: if a given miRNA is located in a miRNA cluster, a "Y" is entered. Tissue Atlas: for each miRNA. The cell or tissue type with the highest percentage, and with percentages within 10 percentage points of the highest percentages, are listed in decreasing abundance in the "max/(max − 10%)" columns. Associated Carrier Subclasses: for miRNAs that show differential enrichment in one or more carrier subclasses by immunoprecipitation, the enriched subclass(es) are listed. See also Tables S3 and S4 and Figure S2. E, early; F, full; L, late; M, middle.

[a]Mean AUC values between 0.6 and 0.8.

[b]miRNAs for which the verification mean AUCs for each GABD window and averaged across all GABD windows columns are >0.6.

variability (compared to the full GABD window). Another indication that bivariate analysis may provide more robust results than univariate analysis was seen when we examined the relative expression of the individual miRNAs comprising the reversals in cases compared to controls (Table S4, sheet 3: discovery [num/denom]: direction [case versus ctrl]/Wilcoxon p value and verification [num/denom]: direction [case versus ctrl]/Wilcoxon p value). Here, we saw that, when examined individually, 4/9 of the denominator miRNAs were significantly (Wilcoxon $p < 0.05$, indicated in red font) differentially expressed between

Verification Mean AUC—Middle	Verification Mean AUC—Late	Chromosome Numerator	Chromosome Denominator	Numerator	Denominator	Tissue Atlas Max/(Max − 10%) Numerator	Tissue Atlas Max/(Max − 10%) Denominator	Carrier Subclasses Numerator	Carrier Subclasses Denominator
0.68[b]	0.68[b]	chr14	chr14	Y	Y	placenta/liver	platelets	CD63	
0.63[a]	0.68[a]	chr17	chr20	Y	Y	RBC	liver	input_AGO2	input_CD63
0.57	0.56	chr16	chr14			platelets	platelets		
0.59	0.51	chr21	chr14			lymphocytes	platelets		
0.71[a]	0.76[a]	chr19	chr16		Y	lymphocytes	liver		
0.70[b]	0.67[b]	chr02|chr07	chr10			RBC	liver	input_AGO2	input
0.46	0.56	chr17	chr14	Y	Y	RBC	placenta	input_AGO2	
0.59	0.68[a]	chr01	chr21			liver/RBC	lymphocytes	PLAP	
0.54	0.69[a]	chr01|chr09	chr21			placenta	lymphocytes		
0.51	0.72[a]	chr02	chr21			liver/placenta	lymphocytes	PLAP	
0.56	0.68[a]	chr03	chr21			RBC	lymphocytes	CD63	
0.56	0.68[a]	chr03	chr21			liver	lymphocytes	PLAP	
0.54	0.69[a]	chr03	chr21	Y		platelets/placenta/liver/RBC	lymphocytes	CD63	
0.47	0.68[a]	chr05	chr21			platelets	lymphocytes	input_CD63	
0.56	0.69[a]	chr07	chr21	Y		RBC	lymphocytes	input	
0.55	0.68[a]	chr08	chr21			platelets	lymphocytes	CD63	
0.61[a]	0.67[a]	chr08	chr21			placenta/liver	lymphocytes	PLAP	
0.57	0.68[a]	chr08	chr21	Y		placenta	lymphocytes	PLAP	
0.54	0.73[a]	chr09	chr21			platelets	lymphocytes	CD63	
0.58	0.69[a]	chr10	chr21			liver	lymphocytes	CD63	
0.55	0.71[a]	chr12	chr21			RBC	lymphocytes	CD63	
0.55	0.69[a]	chr12|chr03	chr21			platelets/liver/placenta	lymphocytes	CD63	
0.48	0.69[a]	chr14	chr21			platelets	lymphocytes	CD63	
0.46	0.72[a]	chr17	chr21	Y		platelets	lymphocytes	AGO2	
0.46	0.70[a]	chr17	chr21	Y		RBC	lymphocytes	AGO2	
0.57	0.70[a]	chr19	chr21	Y		placenta	lymphocytes	PLAP	
0.46	0.71[a]	chr21	chr21	Y		liver	lymphocytes		
0.46	0.73[a]	chr22	chr21	Y		RBC	lymphocytes		
0.64[a]	0.69[a]	chr19	chr21	Y		placenta	lymphocytes	PLAP	
0.54	0.69[a]	chrX	chr21	Y		RBC	lymphocytes		
0.70[a]	0.52	chr19		Y		placenta		PLAP	
0.73[a]	0.62	chr20		Y		liver		input_CD63	

cases and controls in the discovery set although 2/9 were also differentially expressed in the verification set. However, for the individual component miRNAs in the reversals, the direction of differential expression was not always preserved between the discovery and verification cohorts. Taken together, these findings suggest that the relative concentrations of pairs of extracellular miRNAs, rather than the absolute levels of individual miRNAs, are more robust predictive biomarkers of disease.

We observed a strong correlation between the discovery and verification AUCs for all verified univariate predictors and

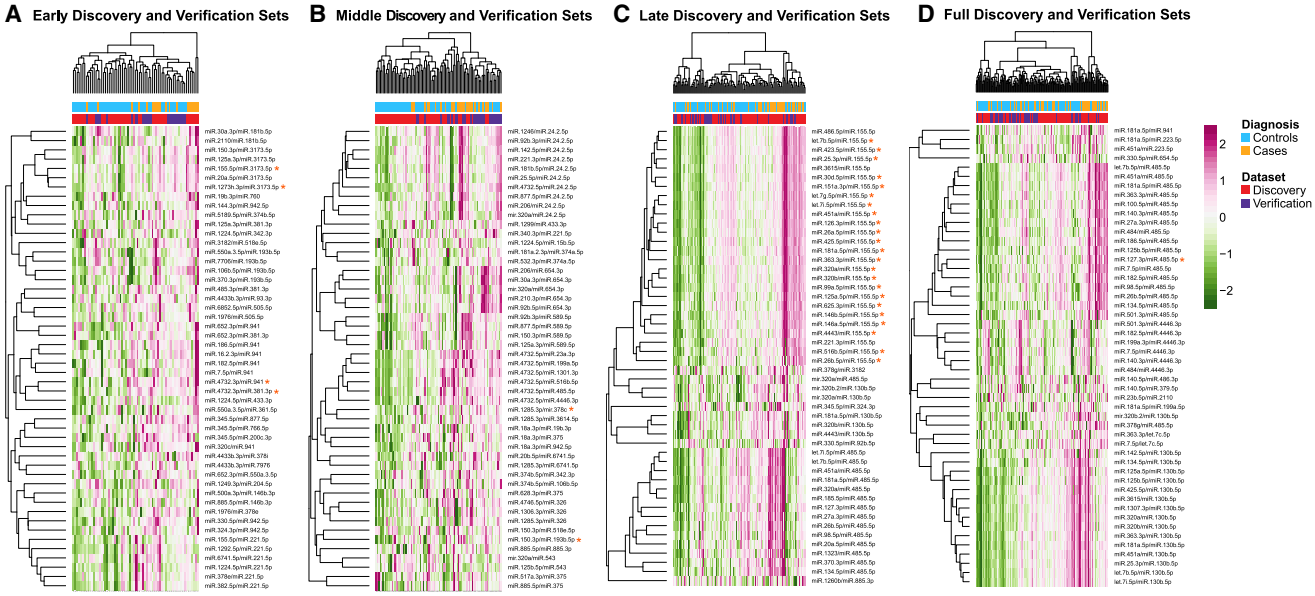


Figure 2. Heatmaps of Reversals Selected in Discovery for Each GABD Window

All heatmaps show normalized reversal scores for both discovery and verification subjects, for the 50 reversals selected in discovery. Trackbars above the heatmaps indicate pre-eclampsia cases (orange) and controls with normal pregnancy outcomes (turquoise) and membership in the discovery (red) or verification (blue) set. The color bar is in units of standard deviations of reversal score distributions. Asterisks indicate reversals that passed verification.

(A) Data for 50 reversals identified in discovery for the early GABD window, for the early GABD subjects (n = 67; discovery and verification sets).

(B) Middle GABD window for the middle GABD subjects (n = 93).

(C) Late GABD window for the late GABD subjects (n = 132).

(D) Full GABD window for all subjects (n = 212).

See also Tables S3 and S4.

reversals (Figure 1D), indicating similar performance in two independent sets of subjects.

For the early, middle, and late GABD windows, we used the normalized reversal scores from all cases and controls for the verified reversals to generate PCA plots (Figures S2C–S2E). For the full GABD windows, there was only one verified reversal, which was not sufficient to generate a PCA plot. The PCA plots for the early and late GABD windows showed good separation between cases and controls. For the middle GABD window, outlier cases and controls were most clearly separated—this may be due to the small number (two) of verified reversals in this GABD window. To examine the relationships between reversal scores and diagnosis (cases and controls) on a more granular level, we generated heatmaps for the top 50 reversals for each GABD window (Figures 2A–2D). We observed the expected clustering segregation of cases from controls, and intermixing of samples from the discovery and verification sets. We also saw that, for each of the verified reversals (indicated by the red asterisks), there were several other reversals (which often shared either the numerator or denominator miRNA) that displayed similar patterns but did not pass verification. This suggests that there may be certain "high-value" numerators and denominators (especially hsa-miR-485-5p for the late and full GABD windows) that warrant further investigation as components of potential multivariate predictors. As well, a larger dataset may allow verification of additional reversals.

Predictors Discovered in One GABD Window May Verify in Other GABD Windows

It is of clinical interest to identify predictors that perform well across a broad range of gestational ages. Thus, in addition to determining whether each univariate predictor and reversal identified in discovery was verified in the same GABD window, we determined its performance across all four GABD windows. We found that several predictors did pass our verification threshold (5^{th} percentile AUC > 0.5) in other GABD windows (italicized in Table S4, sheet 3, verification: GABD window). For each predictor that verified in at least one GABD window, we provide the mean AUC for each GABD window for both the discovery and verification cohorts (Table S4, sheet 3, discovery mean AUC: full, early, middle, late, and verification mean AUC: full, early, middle, late). We saw from these results that some of the predictors, particularly hsa-miR-127-3p/hsa-miR-485-5p, hsa-miR-1285-3p/hsa-mir-378c, and hsa-miR-331-3p (mean verification AUCs bolded in column C [numerator] and D [denominator] in Table S4, sheet 3), performed well in the verification cohort across all GABD windows.

Some of the Same miRNAs Are Shared among Multiple Reversals or between Univariate Predictors and Reversals

Four miRNAs (hsa-miR-485-5p, hsa-miR-941, hsa-miR-3173-5p, and hsa-miR-155-5p) were found in the denominators or more than one reversal, with hsa-miR-155-5p being in the

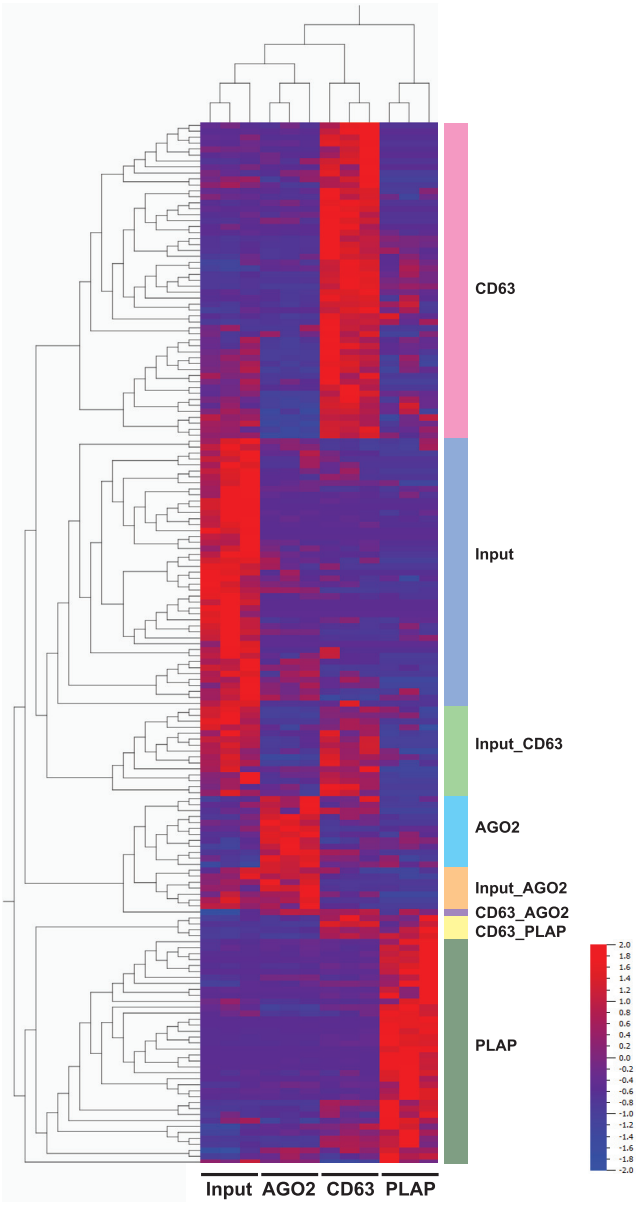


Figure 3. Cell and Tissue miRNA Expression and Deconvolution Analysis

(A) PCA plot showing unsupervised clustering of cell and tissue types by miRNA profiling data.

(B) Hierarchical clustering of cell and tissue types with heatmap of differentially expressed miRNAs ($q < 10^{-12}$).

(C) Box-and-whisker plot of deconvolution results, indicating the percent contribution of each cell/tissue type to the extracellular miRNA profiles of each of the maternal serum samples in this study (both discovery and verification cohorts; n = 212).

(D) Bar graphs showing, for each cell/tissue type, the number of numerators/univariate predictors (top) and denominators (bottom) for which that cell/tissue type was a major contributor.

See also Table S5.

denominator of all 23 of the reversals discovered and verified in the late GABD window (late/late), as well as in the numerator of one of the early/early reversals. hsa-miR-26b-5p was the numerator in one late/late and one full, late/middle reversal. Both of the univariate predictors that were discovered and verified in the same GABD window (hsa-miR-516b-5p and hsa-miR-941) were also members of verified reversals (underlined in Table 3).

Verified Predictors Include Placenta-Associated miRNAs

Verified predictors include members of two placenta-associated miRNA clusters, one located on chromosome 14 (Tables 3 and S4, sheet 3, highlighted in blue in numerator chromosome, denominator chromosome, and in miRNA cluster)[45] and the other on chromosome 19 (Tables 3 and S4, sheet 3, highlighted in red in numerator chromosome, denominator chromosome, and in miRNA cluster).[46]

To assess the likely cell/tissue source of the miRNAs comprising each predictor, we performed small RNA-seq on peripheral blood mononuclear cells (PBMCs), red blood cells (RBCs), and platelets collected by centrifugation from human plasma; granulocytes, lymphocytes, and monocytes isolated from human plasma by fluorescence-activated cell sorting; and adult human brain, heart, intestine, kidney, liver, lung, pancreas, and human placenta collected from 17 to 28 weeks gestation (sample level data: Table S5, sheet 1; data averaged for each cell/tissue type: Table S5, sheet 2). miRNAs that are highly significantly ($q < 10^{-12}$) differentially expressed among cell/tissue types were identified (Figures 3A and 3B; Table S5, sheets 3 and 4) and combined with the raw exRNA data from the discovery and verification cohorts (Tables S2, sheet 1 and S5, sheet 5) for deconvolution analysis to estimate the fractional contribution of each cell/tissue type to the overall extracellular miRNA profile of maternal serum (Figure 3; Table S5, sheet 6). Finally, the results of the deconvolution analysis (Table S5, sheet 6) were combined with the averaged cell/tissue miRNA expression data (Table S5, sheet 2) to estimate the fractional contribution of each cell/tissue type to the amount of each extracellular miRNA in maternal serum (Table S5, sheet 7). This information was then extracted for the miRNAs comprising each univariate predictor and reversal (Table S4, sheet 3, tissue atlas columns), and the cell/tissue types contributing the highest percentage, or a percentage within ten percentage points of the highest percentage, were listed (Table S4, sheet 3: tissue atlas, max/[max − 10%]). We observed that liver, RBC, placenta, and platelets contributed most strongly to the overall maternal extracellular miRNA population (Figure 3C). These cell/tissue types were also the predominant sources of many of the extracellular miRNAs predictors (Table S4, sheet 3: tissue atlas, max/[max − 10%]), but liver was underrepresented for both the numerators and denominators and lymphocytes were overrepresented for the denominators (Figure 3D). Overall, the placenta was identified as a major contributor for 12 of the 30 verified reversals and 1 of the 2 verified univariate predictors. Interestingly, in the reversals, the same cell/tissue type was a major contributor to both the numerator and denominator for only three reversals.

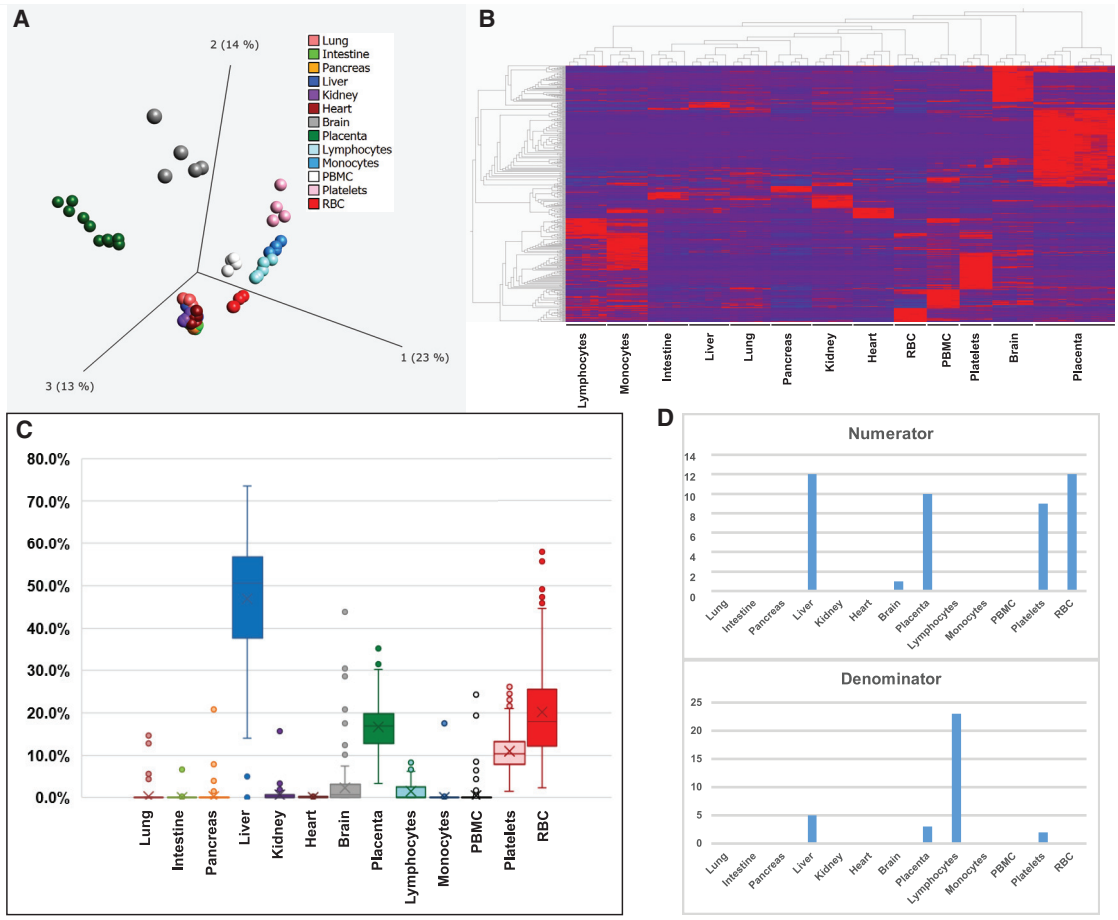


Figure 4. miRNAs Associated with Different Carrier Subclasses

Heatmap showing eight sets of co-expressed miRNAs in pooled third-trimester control sample (n = 1; technical replicates = 3) identified by hierarchical clustering. Each set of miRNAs is labeled with the associated carrier subclasses. See also Table S6.

Placenta-Associated miRNA Predictors Are Associated with CD63+ and PLAP+ Carrier Subclasses

Recent work supports the existence of distinct carrier subclasses, each of which is associated with a specific repertoire of molecular cargo, including miRNAs.[47,48] To explore whether miRNA predictors were carried by specific carrier subclasses, we enriched for canonical extracellular vesicles (EVs), placenta-associated EVs, and ribonucleoprotein complexes (RNPs) from pooled serum from third-trimester pregnant women using magnetic beads conjugated to antibodies raised against CD63 (a commonly used EV surface marker), PLAP (a placental EV-associated surface marker),[49] and AGO2 (a component of the RNA-induced silencing complex and associated with a large fraction of the extracellular miRNAs that are not associated with EVs),[50] respectively. We then performed small RNA-seq on these immunoaffinity enriched samples, as well as the input pooled serum (Table S6, sheet 1) and then identified miRNAs that were significantly ($q \leq 0.05$) differentially expressed among these groups (Table S6, sheet 2). Hierarchical clustering allowed us to identify eight co-expressed sets of miRNAs, each of which had a characteristic pattern of enrichment in one or more carrier subclass (Table S6, sheet 2; Figure 4). The three expected sets of miRNAs that showed non-overlapping associations with CD63, AGO2, or PLAP indicate that certain miRNAs are loaded into distinct carrier subclasses that display only one of these three markers. The two sets of miRNAs that were enriched for two markers (CD63_AGO2 and CD63_PLAP) suggest that some miRNAs are associated with either two carrier subclasses or with a single carrier subclass displaying both markers. The two sets of miRNAs that were strongly detected in both the input and associated with one of the markers (Input_CD63 and Input_AGO2) are consistent with certain miRNAs being associated with two carrier subclasses: one displaying either CD63 or AGO2 and one that does not display any of the three tested markers. Finally, the set of miRNAs detected in unfractionated pregnant serum, but not associated with any of the three tested markers (input), indicates that there remain one or more other carrier subclasses that do not display any of the three tested markers.

The carrier subclasses with which each miRNA predictor was associated were then extracted and listed in Table 3: associated carrier subclasses and Table S4, sheet 3. Of the 13 miRNAs for

which placenta was a major contributor, 6 were associated with the PLAP subclass, 4 with the CD63 subclass, and 3 with none of the categories in Figure 4. It is noteworthy that the denominators for most of the reversals were not assigned to any of the categories in Figure 4. As noted above, if a given miRNA were associated with an as-yet-unidentified carrier subclass, we would expect it to be assigned to the "input" category. Therefore, the unassigned miRNAs are those that are present at similar levels in all tested subclasses, as well as the input; even representation across carrier subclasses would be a good feature for a broadly useful normalizer and may be why unassigned miRNAs were preferentially selected as denominators for reversals.

DISCUSSION

In this study, we were able to identify and verify extracellular miRNA biomarkers for prediction of PE. An important result we noted was that not only were there many fewer univariate predictors compared to reversals identified in the discovery cohort, but a markedly lower percentage of the univariate candidates passed the AUC cutoff in the verification cohort. We attribute this finding to variability in small RNA-seq data obtained from exRNA samples arising from both experimental variability during the exRNA isolation process and biological variability from heterogeneity in the representation of the various carrier subclasses among serum samples collected from different individuals (as discussed in Murillo et al.[47]). These sources of variability cannot be accounted for using standard normalization methods, thus making measurements of individual miRNAs difficult to compare between samples. In our bivariate analyses, ratios of pairs of miRNAs (rather than measurements of single miRNAs, as in the univariate analyses) were tested, allowing the two miRNAs in each pair (which we also refer to as a reversal) to normalize each other. We decided to include log transformation into our analysis because gene expression levels have been shown to be lognormally distributed.[51] For lognormally distributed data, linear scale analyses are dominated by high outliers and make detection of downregulation difficult, making log transformation a logical step to incorporate. However, it was initially unclear whether we should use the log values of the ratios, representing a geometric relationship between miRNA abundances, or the ratios of the log values, representing a power relationship. We therefore calculated both sets of values for the discovery cohort and found that there was better separation of cases and controls using the ratios of the log values, even prior to selection of best miRNA reversals. Ratios of normally distributed values, such as our log-transformed miRNA abundances, are used frequently in risk analysis.[43] Ratios of log values have been shown to be particularly useful for examining a change in the rate of incidence of a clinical event. Relative log survival is an unbiased estimate of the relative hazard.[52] Thus, for miRNAs whose abundance is related to PE-free pregnancy "survival" versus incidence of PE, the ratios of logs provide a useful metric. Assessed in the combination of training and verification data, the first principal components of the verified reversals in the early, middle, and late blood draw windows show a strong separation between cases and controls while capturing the majority of variance. These observations suggest that the reversals are likely to distinguish cases from controls in similar populations. Intriguingly, many of our verified reversals contained one miRNA highly expressed in placenta and one likely non-placental miRNA, and no reversals were composed of two miRNAs that were both highly expressed in placenta. This suggests that miRNAs expressed by non-target tissues may serve as internal normalizers that enable more robust measurement of target-tissue-associated exRNA biomarkers or that normalization of placental to maternal contributions improves predictions for fetal/maternal dyad disease states like PE.

We observed that several of the predictors that were identified in a given GABD window from the discovery cohort performed well across multiple GABD windows in the verification cohort. This suggests that it may be possible to develop a clinical predictive assay with good performance across a relatively broad GA range. The performance and robustness of such a test may be enhanced by constructing a multianalyte assay, which in addition to extracellular miRNA predictors may incorporate clinical parameters and other molecular biomarkers.

Two clusters of miRNAs have been found to be of particular significance in placental biology. A study by Bentwich et al.[46] identified a placenta-specific cluster of miRNAs on the long arm of chromosome 19 (chr19q13), which is commonly referred to as the C19MC cluster. A subsequent publication from our laboratory found that this cluster was also highly expressed in pluripotent human embryonic stem cells but was rapidly downregulated during differentiation.[53] miRNAs on the long arm of chromosome 14 (chr14q32) have also been shown to be highly expressed in the placenta and embryonic stem cells and to regulate gene expression during development.[45,53] Our verified predictors contained several miRNAs that were encoded in the C19MC and chr14q32 miRNA clusters.

We used miRNA expression data from a variety of cell and tissue types to determine the likely sources of our extracellular miRNA biomarkers. For the reversals, liver, RBC, placenta, and platelets were the most frequent major contributors of the numerator miRNAs and lymphocytes were the major contributor for the large majority of denominator miRNAs. We also used miRNA expression data from samples enriched from pooled pregnant serum samples by immunoaffinity separation using magnetic beads conjugated to antibodies raised against CD63, AGO2, and PLAP to determine the carrier subclass association of our extracellular miRNA biomarkers. Nearly half of the miRNA biomarkers for which placenta was a major contributor were associated with PLAP, and one-third were associated with CD63, suggesting that placental EVs and canonical EVs are important carriers of placentally derived extracellular miRNAs. It is important to note that our approach for estimating the contribution of each cell/tissue type to the level of specific miRNAs in the serum assumes that the intracellular level of each miRNA is reflected in the population of miRNAs released by that cell/tissue into the serum. However, there is evidence that there is selective RNA cargo loading into EVs, RNPs, and other carriers.[54] Moreover, we recognize that our dataset does not include all cell and tissue types. To refine these calculations, it will be necessary to obtain profiles of the exRNAs released by each cell and tissue type.

Among the collected demographic and clinical variables, we observed that, as expected, the cases had a significantly earlier

gestational age of delivery and lower birthweight icompared to controls. We also noted a significantly higher BMI in cases compared to controls, which is consistent with previous literature reporting an elevated risk of PE in obese gravidas.[55]

As mentioned in the Introduction, we found nineteen previous studies reporting on extracellular miRNAs associated with PE, with limited overlap in identified miRNAs among studies (Table 1).[17–35] These studies can be divided into twelve discovery studies, which used large qRT-PCR panels, microarrays, or small RNA-seq,[18,22,23,27–35] and seven targeted studies, which performed qRT-PCR on small numbers of selected miRNAs,[17,19–21,24–26] most commonly members of the placenta-specific miRNA cluster located at chr19q13 (C19MC).[19–21,24,25] Eleven C19MC miRNAs were identified in four of the discovery studies (the other eight discovery studies did not identify any C19MC miRNAs),[29,30,33,34] but only three specific miRNAs were shared among at least two studies: hsa-miR-517c-3p and hsa-miR-518e-3p were shared between Yang et al.[33] and Yang et al.,[34] and hsa-miR-519a-3p was common to Yang et al.[33] and Timofeeva et al.[29] (Table S1). Of the nineteen C19MC miRNAs identified as differentially expressed between PE and control in at least one targeted study, five were found by two studies. Jairajpuri et al.[21] used a candidate approach, targeting 84 miRNAs identified in previous placental RNA and exRNA studies to be associated with PE, but found that only 43 of these were detectable in their exRNA samples. Of these, nine overlapped with extracellular miRNA biomarkers found in other exRNA studies (Table S1). All five of the overlapping miRNAs that were higher in PE than control in the Jairajpuri et al.[21] data were also consistently higher in the other studies: hsa-miR-650[30]; hsa-miR-29a[22,33,34]; hsa-miR-210[17,30,32]; hsa-miR-518b[19,25,30]; and hsa-miR-155-5p.[17] Three of the four overlapping miRNAs that were lower in PE than control in the Jairajpuri et al.[21] data were also consistently lower in the other studies: hsa-miR-144-3p[22,30,31]; hsa-miR-19b1[32]; and hsa-miR-15b-5p[30]. The fourth overlapping miRNA that was lower in PE than control for the Jairajpuri et al.[21] study was also lower in PE in Xu et al.[32] but higher in PE in Yang et al.[34] It is notable that the C19MC miRNAs were largely seen in the studies that compared PE cases after diagnosis with gestational age-matched non-PE controls. Of the five studies that used a discovery approach on pre-symptomatic subjects,[23,27,30,32,35] the sample sizes were quite small (15–35 cases/24–40 controls), and only one C19MC miRNA was identified as a biomarker in one study.[30] We compared our results to these previous studies and found that 11 of our miRNA predictors overlapped with biomarkers present at higher levels in the serum or plasma of patients with PE (or who later developed PE) compared to controls and 5 overlapped with biomarkers previously reported to be lower in PE compared to controls (Figures S2F and S2G). All overlaps between the miRNAs identified in this study with prior studies are shown in Table S4, sheet 3: overlap with literature. Of these overlapping miRNAs, miR-155-5p, which was present in 24 of our verified reversals, is of particular interest. It has been reported to be more highly expressed in placentas from pre-eclamptic compared to normal pregnancies and to suppress cell invasion in HTR-8/SVneo trophoblast cell line through repression of eNOS expression.[56] It was later reported that, in human umbilical vein endothelial cells (HUVECs), aspirin could prevent tumor necrosis factor alpha (TNF-α)-induced endothelial dysfunction by repressing downstream hsa-miR-155-5p expression and thereby derepressing eNOS.[57] It is therefore possible that hsa-miR-155-5p may not only be a biomarker for prediction and diagnosis of PE but also may be a functional mediator of PE pathogenesis.

We speculate that extracellular miRNA biomarkers for PE may be indicators of placental or maternal tissue stress and/or serve as signaling molecules between the placenta and maternal tissues or between maternal tissues. Three extracellular miRNA biomarkers identified in this study have been previously associated with hypertension. hsa-miR-26b-5p and hsa-miR-7-5p were found to be upregulated in the plasma of non-pregnant patients with hypertension and left ventricular hypertrophy (LVH) compared to normotensive patients or patients with hypertension but no LVH.[58] hsa-miR-181a-5p mimic has been shown to decrease blood pressure in hypertensive mice.[59] In our analysis, hsa-miR-26b-5p appears to be predominantly derived from the liver and placenta, hsa-miR-7-5p from the brain, and hsa-miR-181a-5p from the placenta.

Our study has applied an unbiased approach to discovery and verification of extracellular miRNA biomarkers for prediction of PE. The rigor of our study design, including adequate numbers of cases and controls for both the discovery and blinded verification phases of analysis, has enabled us to develop an approach to extracellular miRNA biomarker discovery/verification, which will be generalizable to other diseases. The candidate predictors from this study will now need to be validated on a large independent cohort, as individual biomarkers or as components of multianalyte assays, which may include not only combinations of extracellular miRNA predictors but also clinical parameters, such as history of severe PE, kidney disease, chronic hypertension, or abnormal analytes during first or second trimester screening. Validated clinical assays for predicting the risk of clinically relevant PE will allow targeting of clinical resources to high-risk cases, while sparing low-risk patients unnecessary anxiety. They will also enable identification of high-risk cases for clinical studies aimed at personalized administration of aspirin, as well as development of preventative and therapeutic modalities.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human subjects
  - Study subject enrollment
  - Clinical data collection, adjudication of pregnancy outcome, and selection of cases and controls for analysis
- METHOD DETAILS

- Maternal serum
- Placenta tissue
- Blood cells
- Immunoprecipitation of exRNA carriers
- RNA extraction and small RNA sequencing library construction
- Small RNA sequencing

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data analysis

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.xcrm.2020.100013.

ACKNOWLEDGMENTS

This publication is part of the NIH Extracellular RNA Communication Consortium paper package and was supported by the NIH Common Fund's exRNA Communication Program. We would like to thank the Harvard Brain Bank and Lifesharing for adult tissue samples and Luc Laurent for illustrating the graphical abstract. This work was funded by NIH UH3TR000906 (S.S., T.H., R.T., R.O., T.R.L., X.Z., C.T., A.F., S.L.-G., K.K., V.T., M.M., J.T., A.V., S.P., R.T.O., G.A.R., H.R., D.E.H., J.B., J.J.B., and L.C.L.) and NIH U01HL126494 (S.S., X.Z., P.D., L.G.P., C.D.L., T.B., X.O.B., and L.C.L.).

AUTHOR CONTRIBUTIONS

Conceptualization, L.C.L., J.J.B., and J.B.; Software and Formal Analysis, J.B., R.T., A.J.L., and C.T.; Investigation, S.S., T.H., R.T.O, T.R.L., X.Z., P.D., L.G.P., A.F., S.L.-G., K.K., V.T., M.M., M.T.D., C.D.L., T.B., and P.S.C.; Resources, I.G. and The PAPR Study Consortium; Data Curation, S.P., R.T.O., G.A.R., H.R., J.T., and A.V.; Writing – Original Draft, S.S., R.T., T.H., L.G.P., J.B., and L.C.L.; Writing – Review & Editing, S.S., R.T., J.B., D.E.H., J.J.B., and L.C.L.; Visualization, R.T., J.B., and L.G.P.; Supervision, L.C.L., J.J.B., J.B., D.E.H., and X.O.B.; Project Administration, D.E.H., A.C.F., S.R.R., and M.P.; Funding Acquisition, L.C.L.

DECLARATION OF INTERESTS

L.C.L. is a site PI for the Sera Prognostics TREETOP study. J.B.B., J.B., D.E.H., A.C.F., M.T.D., S.R.R., A.J.L., and R.T. are employees and shareholders in Sera Prognostics. Sera Prognostics has filed a provisional patent application (United States Provisional Application no. 62/777,576) on the findings presented in this manuscript.

Received: June 7, 2019
Revised: November 1, 2019
Accepted: April 21, 2020
Published: May 19, 2020

REFERENCES

1. Bartsch, E., Medcalf, K.E., Park, A.L., and Ray, J.G.; High Risk of Preeclampsia Identification Group (2016). Clinical risk factors for preeclampsia determined in early pregnancy: systematic review and meta-analysis of large cohort studies. BMJ *353*, i1753.
2. Ness, R.B., and Roberts, J.M. (1996). Heterogeneous causes constituting the single syndrome of preeclampsia: a hypothesis and its implications. Am. J. Obstet. Gynecol. *175*, 1365–1370.
3. Sibai, B.M. (2003). Diagnosis and management of gestational hypertension and preeclampsia. Obstet. Gynecol. *102*, 181–192.
4. Sibai, B.M. (2006). Preeclampsia as a cause of preterm and late preterm (near-term) births. Semin. Perinatol. *30*, 16–19.
5. Henderson, J.T., Whitlock, E.P., O'Conner, E., Senger, C.A., Thompson, J.H., and Rowland, M.G. (2014). Low-dose aspirin for prevention of morbidity and mortality from preeclampsia: a systematic evidence review for the U.S. Preventive Services Task Force. Ann. Intern. Med. *160*, 695–703.
6. Poon, L.C., Syngelaki, A., Akolekar, R., Lai, J., and Nicolaides, K.H. (2013). Combined screening for preeclampsia and small for gestational age at 11-13 weeks. Fetal Diagn. Ther. *33*, 16–27.
7. Yliniemi, A., Makikallio, K., Korpimaki, T., Kouru, H., Marttala, J., and Ryynanen, M. (2015). Combination of PAPPA, fhCGβ, AFP, PlGF, sTNFR1, and maternal characteristics in prediction of early-onset preeclampsia. Clin. Med. Insights Reprod. Health *9*, 13–20.
8. Levine, R.J., Maynard, S.E., Qian, C., Lim, K.H., England, L.J., Yu, K.F., Schisterman, E.F., Thadhani, R., Sachs, B.P., Epstein, F.H., Sibai, B.M., Sukhatme, V.P., and Karumanchi, S.A. (2004). Circulating angiogenic factors and the risk of preeclampsia. N. Engl. J. Med. *350*, 672–683.
9. Tjwa, M., Luttun, A., Autiero, M., and Carmeliet, P. (2003). VEGF and PlGF: two pleiotropic growth factors with distinct roles in development and homeostasis. Cell Tissue Res. *314*, 5–14.
10. Caillon, H., Tardif, C., Dumontet, E., Winer, N., and Masson, D. (2018). Evaluation of sFlt-1/PlGF ratio for predicting and improving clinical management of pre-eclampsia: experience in a specialized perinatal care center. Ann. Lab. Med. *38*, 95–101.
11. Das, S., Extracellular RNA Communication Consortium; Ansel, K.M., Bitzer, M., Breakefield, X.O., Charest, A., Galas, D.J., Gerstein, M.B., Gupta, M., Milosavljevic, A., McManus, M.T., et al. (2019). The Extracellular RNA Communication Consortium: establishing foundational knowledge and technologies for extracellular RNA research. Cell *177*, 231–242.
12. Ng, E.K.O., Tsui, N.B.Y., Lau, T.K., Leung, T.N., Chiu, R.W.K., Panesar, N.S., Lit, L.C.W., Chan, K.-W., and Lo, Y.M.D. (2003). mRNA of placental origin is readily detectable in maternal plasma. Proc. Natl. Acad. Sci. USA *100*, 4748–4753.
13. Ge, Q., Liu, Q., Bai, Y., Wen, T., and Lu, Z. (2005). A semi-quantitative microarray method to detect fetal RNAs in maternal plasma. Prenat. Diagn. *25*, 912–918.
14. Go, A.T.J.I., Visser, A., Mulders, M.A.M., Blankenstein, M.A., van Vugt, J.M.G., and Oudejans, C.B.M. (2004). Detection of placental transcription factor mRNA in maternal plasma. Clin. Chem. *50*, 1413–1414.
15. Tsui, N.B., Chim, S.S., Chiu, R.W., Lau, T.K., Ng, E.K., Leung, T.N., Tong, Y.K., Chan, K.C., and Lo, Y.M. (2004). Systematic micro-array based identification of placental mRNA in maternal plasma: towards non-invasive prenatal gene expression profiling. J. Med. Genet. *41*, 461–467.
16. Poon, L.L.M., Leung, T.N., Lau, T.K., and Lo, Y.M.D. (2000). Presence of fetal RNA in maternal plasma. Clin. Chem. *46*, 1832–1834.
17. Gan, L., Liu, Z., Wei, M., Chen, Y., Yang, X., Chen, L., and Xiao, X. (2017). MiR-210 and miR-155 as potential diagnostic markers for pre-eclampsia pregnancies. Medicine (Baltimore) *96*, e7515.
18. Gunel, T., Hosseini, M.K., Gumusoglu, E., Kisakesen, H.I., Benian, A., and Aydinli, K. (2017). Expression profiling of maternal plasma and placenta microRNAs in preeclamptic pregnancies by microarray technology. Placenta *52*, 77–85.
19. Hromadnikova, I., Kotlabova, K., Ivankova, K., and Krofta, L. (2017). First trimester screening of circulating C19MC microRNAs and the evaluation of their potential to predict the onset of preeclampsia and IUGR. PLoS ONE *12*, e0171756.
20. Hromadnikova, I., Kotlabova, K., Ondrackova, M., Kestlerova, A., Novotna, V., Hympanova, L., Doucha, J., and Krofta, L. (2013). Circulating C19MC microRNAs in preeclampsia, gestational hypertension, and fetal growth restriction. Mediators Inflamm. *2013*, 186041.
21. Jairajpuri, D.S., Malalla, Z.H., Mahmood, N., and Almawi, W.Y. (2017). Circulating microRNA expression as predictor of preeclampsia and its severity. Gene *627*, 543–548.

22. Li, H., Ge, Q., Guo, L., and Lu, Z. (2013). Maternal plasma miRNAs expression in preeclamptic pregnancies. Biomed. Res. Int. *2013*, 970265.

23. Luque, A., Farwati, A., Crovetto, F., Crispi, F., Figueras, F., Gratacos, E., and Aran, J.M. (2014). Usefulness of circulating microRNAs for the prediction of early preeclampsia at first-trimester of pregnancy. Sci. Rep. *4*, 4882.

24. Martinez-Fierro, M.L., Garza-Veloz, I., Gutierrez-Arteaga, C., Delgado-Enciso, I., Barbosa-Cisneros, O.Y., Flores-Morales, V., Hernandez-Delgadillo, G.P., Rocha-Pizaña, M.R., Rodriguez-Sanchez, I.P., Badillo-Almaraz, J.I., et al. (2018). Circulating levels of specific members of chromosome 19 microRNA cluster are associated with preeclampsia development. Arch. Gynecol. Obstet. *297*, 365–371.

25. Miura, K., Higashijima, A., Murakami, Y., Tsukamoto, O., Hasegawa, Y., Abe, S., Fuchi, N., Miura, S., Kaneuchi, M., and Masuzaki, H. (2015). Circulating chromosome 19 miRNA cluster microRNAs in pregnant women with severe pre-eclampsia. J. Obstet. Gynaecol. Res. *41*, 1526–1532.

26. Motawi, T.M.K., Sabry, D., Maurice, N.W., and Rizk, S.M. (2018). Role of mesenchymal stem cells exosomes derived microRNAs; miR-136, miR-494 and miR-495 in pre-eclampsia diagnosis and evaluation. Arch. Biochem. Biophys. *659*, 13–21.

27. Salomon, C., Guanzon, D., Scholz-Romero, K., Longo, S., Correa, P., Illanes, S.E., and Rice, G.E. (2017). Placental exosomes as early biomarker of preeclampsia: potential role of exosomal microRNAs across gestation. J. Clin. Endocrinol. Metab. *102*, 3182–3194.

28. Stubert, J., Koczan, D., Richter, D.U., Dieterich, M., Ziems, B., Thiesen, H.J., Gerber, B., and Reimer, T. (2014). miRNA expression profiles determined in maternal sera of patients with HELLP syndrome. Hypertens. Pregnancy *33*, 215–235.

29. Timofeeva, A.V., Gusar, V.A., Kan, N.E., Prozorovskaya, K.N., Karapetyan, A.O., Bayev, O.R., Chagovets, V.V., Kliver, S.F., Iakovishina, D.Y., Frankevich, V.E., et al. (2018). Corrigendum to "Identification of potential early biomarkers of preeclampsia" [Placenta (2018) 61-71]. Placenta *63*, 61.

30. Ura, B., Feriotto, G., Monasta, L., Bilel, S., Zweyer, M., and Celeghini, C. (2014). Potential role of circulating microRNAs as early markers of preeclampsia. Taiwan J. Obstet. Gynecol. *53*, 232–234.

31. Wu, L., Zhou, H., Lin, H., Qi, J., Zhu, C., Gao, Z., and Wang, H. (2012). Circulating microRNAs are elevated in plasma from severe preeclamptic pregnancies. Reproduction *143*, 389–397.

32. Xu, P., Zhao, Y., Liu, M., Wang, Y., Wang, H., Li, Y., Zhu, X., Yao, Y., Wang, H., Qiao, J., et al. (2014). Variations of microRNAs in human placentas and plasma from preeclamptic pregnancy. Hypertension *63*, 1276–1284.

33. Yang, Q., Lu, J., Wang, S., Li, H., Ge, Q., and Lu, Z. (2011). Application of next-generation sequencing technology to profile the circulating microRNAs in the serum of preeclampsia versus normal pregnant women. Clin. Chim. Acta *412*, 2167–2173.

34. Yang, S., Li, H., Ge, Q., Guo, L., and Chen, F. (2015). Deregulated microRNA species in the plasma and placenta of patients with preeclampsia. Mol. Med. Rep. *12*, 527–534.

35. Yoffe, L., Gilam, A., Yaron, O., Polsky, A., Farberov, L., Syngelaki, A., Nicolaides, K., Hod, M., and Shomron, N. (2018). Early detection of preeclampsia using circulating small non-coding RNA. Sci. Rep. *8*, 3401.

36. Rozowsky, J., Kitchen, R.R., Park, J.J., Galeev, T.R., Diao, J., Warrell, J., Thistlethwaite, W., Subramanian, S.L., Milosavljevic, A., and Gerstein, M. (2019). exceRpt: a comprehensive analytic platform for extracellular RNA profiling. Cell Syst. *8*, 352–357.e3.

37. Åstrand, M. (2003). Contrast normalization of oligonucleotide arrays. J. Comput. Biol. *10*, 95–102.

38. Schwarzenbach, H., da Silva, A.M., Calin, G., and Pantel, K. (2015). Data normalization strategies for microRNA quantification. Clin. Chem. *61*, 1333–1342.

39. Endzelinš, E., Berger, A., Melne, V., Bajo-Santos, C., Sobolevska, K., Ābols, A., Rodriguez, M., Šantare, D., Rudnickiha, A., Lietuvietis, V., et al. (2017). Detection of circulating miRNAs: comparative analysis of extracellular vesicle-incorporated miRNAs and cell-free miRNAs in whole plasma of prostate cancer patients. BMC Cancer *17*, 730.

40. Price, N.D., Trent, J., El-Naggar, A.K., Cogdell, D., Taylor, E., Hunt, K.K., Pollock, R.E., Hood, L., Shmulevich, I., and Zhang, W. (2007). Highly accurate two-gene classifier for differentiating gastrointestinal stromal tumors and leiomyosarcomas. Proc. Natl. Acad. Sci. USA *104*, 3414–3419.

41. Saade, G.R., Boggess, K.A., Sullivan, S.A., Markenson, G.R., Iams, J.D., Coonrod, D.V., Pereira, L.M., Esplin, M.S., Cousins, L.M., Lam, G.K., et al. (2016). Development and validation of a spontaneous preterm delivery predictor in asymptomatic women. Am. J. Obstet. Gynecol. *214*, 633.e1–633.e24.

42. Verlohren, S., Herraiz, I., Lapaire, O., Schlembach, D., Moertl, M., Zeisler, H., Calda, P., Holzgreve, W., Galindo, A., Engels, T., et al. (2012). The sFlt-1/PlGF ratio in different types of hypertensive pregnancy disorders and its prognostic potential in preeclamptic patients. Am. J. Obstet. Gynecol. *206*, 58.e1–58.e8.

43. Hayya, J., Armstrong, D., and Gressis, N. (1975). A note on the ratio of two normally distributed variables. Manage. Sci. *21*, 1215–1354.

44. Chen, Y., Dougherty, E.R., and Bittner, M.L. (1997). Ratio-based decisions and the quantitative analysis of cDNA microarray images. J. Biomed. Opt. *2*, 364–374.

45. Seitz, H., Royo, H., Bortolin, M.-L., Lin, S.-P., Ferguson-Smith, A.C., and Cavaillé, J. (2004). A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. Genome Res. *14*, 1741–1748.

46. Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., et al. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. Nat. Genet. *37*, 766–770.

47. Murillo, O.D., Thistlethwaite, W., Rozowsky, J., Subramanian, S.L., Lucero, R., Shah, N., Jackson, A.R., Srinivasan, S., Chung, A., Laurent, C.D., et al. (2019). exRNA Atlas analysis reveals distinct extracellular RNA cargo types and their carriers present across human biofluids. Cell *177*, 463–477.e15.

48. Srinivasan, S., Yeri, A., Cheah, P.S., Chung, A., Danielson, K., De Hoff, P., Filant, J., Laurent, C.D., Laurent, L.D., Magee, R., et al. (2019). Small RNA sequencing across diverse biofluids identifies optimal methods for exRNA isolation. Cell *177*, 446–462.e16.

49. Dragovic, R.A., Southcombe, J.H., Tannetta, D.S., Redman, C.W.G., and Sargent, I.L. (2013). Multicolor flow cytometry and nanoparticle tracking analysis of extracellular vesicles in the plasma of normal pregnant and pre-eclamptic women. Biol. Reprod. *89*, 151.

50. Turchinovich, A., Weiz, L., Langheinz, A., and Burwinkel, B. (2011). Characterization of extracellular circulating microRNA. Nucleic Acids Res. *39*, 7223–7233.

51. Bengtsson, M., Ståhlberg, A., Rorsman, P., and Kubista, M. (2005). Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. Genome Res. *15*, 1388–1392.

52. Perneger, T.V. (2008). Estimating the relative hazard by the ratio of logarithms of event-free proportions. Contemp. Clin. Trials *29*, 762–766.

53. Laurent, L.C., Chen, J., Ulitsky, I., Mueller, F.-J., Lu, C., Shamir, R., Fan, J.-B., and Loring, J.F. (2008). Comprehensive microRNA profiling reveals a unique human embryonic stem cell signature dominated by a single seed sequence. Stem Cells *26*, 1506–1516.

54. Wei, Z., Batagov, A.O., Schinelli, S., Wang, J., Wang, Y., El Fatimy, R., Rabinovsky, R., Balaj, L., Chen, C.C., Hochberg, F., et al. (2017). Coding and noncoding landscape of extracellular RNA released by human glioma stem cells. Nat. Commun. *8*, 1145.

55. Roberts, J.M., Bodnar, L.M., Patrick, T.E., and Powers, R.W. (2011). The role of obesity in preeclampsia. Pregnancy Hypertens. *1*, 6–16.

56. Li, X., Li, C., Dong, X., and Gou, W. (2014). MicroRNA-155 inhibits migration of trophoblast cells and contributes to the pathogenesis of severe

preeclampsia by regulating endothelial nitric oxide synthase. Mol. Med. Rep. *10*, 550–554.

57. Kim, J., Lee, K.S., Kim, J.H., Lee, D.K., Park, M., Choi, S., Park, W., Kim, S., Choi, Y.K., Hwang, J.Y., et al. (2017). Aspirin prevents TNF-α-induced endothelial cell dysfunction by regulating the NF-κB-dependent miR-155/eNOS pathway: role of a miR-155/eNOS axis in preeclampsia. Free Radic. Biol. Med. *104*, 185–198.
58. Kaneto, C.M., Nascimento, J.S., Moreira, M.C.R., Ludovico, N.D., Santana, A.P., Silva, R.A.A., Silva-Jardim, I., Santos, J.L., Sousa, S.M.B., and Lima, P.S.P. (2017). MicroRNA profiling identifies miR-7-5p and miR-26b-5p as differentially expressed in hypertensive patients with left ventricular hypertrophy. Braz. J. Med. Biol. Res. *50*, e6211.
59. Marques, F.Z., and Charchar, F.J. (2015). microRNAs in essential hypertension and blood pressure regulation. In microRNA: Medical Evidence, G. Santulli, ed. (Springer), pp. 215–235.
60. Coarfa, C., Pichot, C., Jackson, A., Tandon, A., Amin, V., Raghuraman, S., Paithankar, S., Lee, A.V., McGuire, S.E., and Milosavljevic, A. (2014). Analysis of interactions between the epigenome and structural mutability of the genome using Genboree Workbench tools. BMC Bioinformatics *15* (*Suppl 7*), S2.
61. Riehle, K., Coarfa, C., Jackson, A., Ma, J., Tandon, A., Paithankar, S., Raghuraman, S., Mistretta, T.A., Saulnier, D., Raza, S., et al. (2012). The Genboree Microbiome Toolset and the analysis of 16S rRNA microbial sequences. BMC Bioinformatics *13* (*Suppl 13*), S11.
62. Subramanian, S.L., Kitchen, R.R., Alexander, R., Carter, B.S., Cheung, K.H., Laurent, L.C., Pico, A., Roberts, L.R., Roth, M.E., Rozowsky, J.S., et al. (2015). Integration of extracellular RNA profiling data using metadata, biomedical ontologies and Linked Data technologies. J. Extracell. Vesicles *4*, 27497.
63. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics *12*, 77.
64. Stegle, O., Parts, L., Durbin, R., and Winn, J. (2010). A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput. Biol. *6*, e1000770.
65. Parts, L., Stegle, O., Winn, J., and Durbin, R. (2011). Joint genetic analysis of gene expression data with inferred cellular phenotypes. PLoS Genet. *7*, e1001276.
66. Ripley, B.D. (2009). Stochastic Simulation (Wiley).
67. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-CD63 Antibody	BD PharMingen	Cat# 556019; RRID:AB_396297
Anti-AGO2 Antibody	Abcam	Cat# ab57113; RRID:AB_2230916
Anti-PLAP Antibody	Abcam	Cat# ab118856; RRID:AB_10900125
Biological Samples		
Maternal Serum	UCSD Placental Dysfunction Clinic	
Maternal Serum	Sera Prognostics Repository	
Critical Commercial Assays		
miRNeasy micro kit	QIAGEN	217084
mirVana miRNA Isolation Kit, without phenol	ThermoFisher Scientific	AM1561
RNA Clean & Concentrator-5	Zymo Research	R1013
DNA Clean & Concentrator-5	Zymo Research	D4013
Quant-iT RiboGreen RNA Assay Kit	ThermoFisher Scientific	R11490
Quant-iT PicoGreen dsDNA Assay Kit	ThermoFisher Scientific	P11496
Agilent RNA 6000 Pico Kit	Agilent Technologies	5067-1513
Agilent RNA 6000 Nano Kit	Agilent Technologies	5067-1511
Bioanalyzer High Sensitivity DNA Analysis	Agilent Technologies	5067-4626
NEBNext Small RNA Library Prep Set for Illumina (Multiplex Compatible)	New England Biolabs	E7330L
Deposited Data		
Small RNA-seq data, miRNA	This paper	See https://www.ncbi.nlm.nih.gov/gap/, phs002016.v1.p1
Software and Algorithms		
exceRpt small RNA-seq pipeline for exRNA profiling	Genboree Bioinformatics	http://www.genboree.org/site/

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Louise C. Laurent (llaurent@ucsd.edu), Professor and Vice Chair for Translational Research, Director of Perinatal Research, Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California San Diego, La Jolla, CA, USA.

Materials Availability

Serum or RNA generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement (some samples are depleted).

Data and Code Availability

The small RNA-seq datasets generated during this study are available at dbGaP https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002016.v1.p1,

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human subjects

Research on human samples were conducted following written informed consent under IRB protocols approved by the Human Research Protections Program at UCSD. Biofluid and RNA samples were labeled with study identifiers; no personally identifiable information was shared among participating laboratories. Raw data will be deposited in a controlled-access database (dbGAP).

Study subject enrollment

Maternal serum was collected between 17-28 weeks gestation. Samples were obtained from the high-risk Placental Study at the University of California, San Diego and from the average-risk Proteomic Assessment of Preterm Risk (PAPR) Study at Sera Prognostics. Eligibility criteria for the two studies are listed in Figure S1. Eligibility criteria for the UCSD Placenta Study included: abnormal first or second trimester analytes defined by PAPP-A < 0.3 MoM, AFP > 2.5 MoM, Inhibin > 2.0 MoM, and Estradiol < 0.30 MoM *and/or* prior adverse pregnancy outcome attributable to preeclampsia *and/or* maternal co-morbidities associated with increased risk for preeclampsia.

Clinical data collection, adjudication of pregnancy outcome, and selection of cases and controls for analysis

After delivery, relevant clinical data were abstracted from the Clarity Clinical Data Warehouse, which houses clinical data exported from UCSD's EPIC Electronic Medical Record for quality improvement and research uses. These data were then used by adjudicators to determine the clinical outcome for each case, with each case adjudicated by two OB/GYN physicians, at least one of which was Board-Certified in Maternal Fetal Medicine. For the samples from the Sera Prognostics PAPR Study, clinical diagnoses were abstracted by clinical research staff at each participating site from the subjects' medical records. No source document verification or adjudication of diagnoses was performed. From the UCSD Placenta Study, 19 cases and 29 controls were selected. From the Sera PAPR Study, 54 cases and 110 controls were selected.

METHOD DETAILS

Maternal serum

Maternal blood was collected by peripheral venipuncture into BD Vacutainer serum blood collection tubes (Becton Dickinson), held at room temperature for at least 10 minutes and centrifuged at 2000 x g for 10 minutes. The serum was divided into 1 mL aliquots and stored at −80°C until RNA extraction was performed.

Placenta tissue

Placenta tissue samples (≤0.5 cm x 0.5 cm x 0.5 cm) were collected after elective termination procedures (5-22 weeks gestational age) or delivery (22-42 weeks), and immediately placed in RNAlater (ThermoFisher). After storage in RNAlater for 24 hours-7 days, the tissue samples were transferred into clean microfuge tubes and stored at −80°C until RNA extraction.

Adult tissue samples (≤0.5 cm x 0.5 cm x 0.5 cm) were collected at the time of organ harvest for organ donation and immediately placed in RNAlater (ThermoFisher). After storage in RNAlater for 24 hours-7 days, the tissue samples were transferred into clean microfuge tubes and stored at −80°C until RNA extraction.

Blood cells

PBMC, Platelets, and RBCs

Human blood samples were collected with written consent from donors ≥ 18 years of age under an IRB protocol approved by the Human Research Protections Programs at UCSD. Biofluid samples were labeled with study identifiers.

Whole blood was collected from two male and two female healthy non-pregnant adult donors, 22-50 years of age. For each donor, blood was collected in the following order: ~8mL into a serum BD Vacutainer collection tube (Becton Dickinson, PN 368045) followed by 3x4.5mL into CTAD (0.11 M buffered trisodium citrate, 15 M theophylline, 3.7 M adenosine, 0.198 M dipyridamole) collection tubes (Becton Dickinson, PN 367947). The serum tubes were held at room temperature for 20 minutes prior to centrifugation at 2000 x g for 5 minutes with no brake. 500μL aliquots were transferred from the clear upper serum layer into screw cap 2mL centrifuge tubes and frozen at −80C until they were processed.

Peripheral blood mononuclear cells (PBMC), platelets, and washed red blood cells (RBC) were purified from the CTAD tubes. Wide bore pipette tips were used at all relevant steps to reduce cell shearing and lysis.

For platelets, the CTAD tubes were centrifuged at 100 x g for 20 minutes with no brake and all but ~100μL of the supernatant was added to a fresh 15mL conical centrifuge tube. Freshly prepared Prostaglandin I2 (PGI2) (Abcam, ab120912-1mg) was added to ~2uM final concentration. The Platelet Rich Plasma (PRP) was then centrifuged at 100 x g for 20 minutes with no brake, and all but ~100μL of the supernatant was added to a fresh 15mL conical centrifuge tube. To pellet the platelets, this tube was centrifuged at 800 x g for 20 minutes with no brake. The platelet pellet was washed without pellet resuspension in 10mL of Platelet Wash Buffer (PWB) (1X wash buffer:10mM Tris pH 7.5 138mM NaCl 1.8mM CaCl2 0.49mM MgCl2 1uM PGI2) The material was centrifuged at 800 x g for 10 minutes with no brake and the supernatant material was removed to near completion. The platelet pellet was gently resuspended in 2mL of PWB and transferred to a 2mL centrifuge tube. The mixture was centrifuged at 800 x g for 10 minutes with no brake and the supernatant material was removed to near completion. The platelet pellet was stored at −80C until processed.

PBMCs and RBCs were purified from the material remaining after the first CTAD tube centrifugation step. For the PBMCs, the remaining PRP, buffy coat, and a small portion of the RBCs were combined by patient and transferred from the CTAD tubes into a fresh 15mL tube. Sufficient PGI2 was added such that the concentration would be 2uM when PWB was added to a 10mL total volume. The material was gently inverted several times to mix and centrifuged at 100x g for 20 minutes with no brake. The supernatant material was removed to near completion and the pellet was mixed in 10mL of RBC lysis buffer (150mM NH4Cl, 10mM NaHCO3, 1.27mM

EDTA) placed at room temperature for 20 minutes. The material was centrifuged at 500 x g for 5 minutes and the supernatant was discarded. The pellet was washed twice with 10mL of Dulbecco's phosphate-buffered saline (DPBS) each time and centrifuged as before. The pellet was gently resuspended in 2mL of DPBS and the material was transferred to a 2mL centrifuge tube and centrifuged as before. The supernatant was carefully removed and the pellet material in the tube was placed at −80C until processed.

The remaining RBCs within the CTAD tubes were combined by patient into a 50mL conical tube and DPBS was added to 50mL. The cells were centrifuged at 500 x g for 5 minutes with no brake and the supernatant was decanted. This washing process was repeated two more times. 200μL aliquots of the remaining washed RBC pellet were transferred into 2mL screw cap tubes and stored at −80C until processed.

Lymphocytes and Monocytes

Human peripheral blood was obtained from health adult volunteers in accordance with the guidelines of the Institutional Review Board of Beth Israel Deaconess Medical Center after informed consent was obtained in accordance with the Declaration of Helsinki. Ten milliliters of blood from healthy donors were collected via cubital venipuncture into a syringe prefilled with 2.3 mL of 6% Dextran 500 (Sigma-Aldrich, St. Louis, MO) and 1 mL of 3.2% Sodium Citrate (Sigma-Aldrich). After gentle mixing the blood was sedimented for 45 minutes with the syringe's nozzle up. The RBC-free fraction was washed once by centrifugation at 2000xg for 10 minutes. The resulting pellet was resuspended in 0.5mL of $HBSS^{2+}$.

The cells were sorted using a Becton Dickinson FACSAria IIu cell sorter equipped with five lasers (350nm, 405nm, 488nm, 561nm, and 640nm). The cell populations were sorted through a 70 μm nozzle tip at a sheath pressure of 70 psi and a drop drive frequency of 90–95 kHz. A highly pure sorting modality (4-way purity sorting for FACS Aria, Masks at 0-32-0) was chosen for cell sorting. The flow rate was maintained at an approximate speed of 10,000 events/second. Lymphocytes and monocytes were gated based on FSC/SSC properties. The FSC values are proportional to the diameter of the interrogated cells, whereas the SSC values provides information about the internal complexity of the interrogated cell or its granularity. Sorted cells were collected in 5 mL polypropylene tubes containing 1 mL collection medium (RPMI supplemented with 50% FBS, 100 μg/ml gentamicin, 4 mM L-glutamine, 20 mM HEPES) and stored at −80C until processed.

Immunoprecipitation of exRNA carriers

Antibody biotinylation: Antibodies raised agaist CD63, AGO2, and PLAP were used. Sodium azide was removed from antibody stocks using the Zeba spin desalting column (7K MWCO, 0.5 ml, Thermo Fisher Scientific, Cat#89882). Antibodies were then biotinylated using the EZ-Link Sulfo-NHS-LC-Biotin reagent (ThermoFisher, Cat#21327), following manufacturer's protocol. Briefly, 10 mM biotin solution was prepared by dissolving 1 mg of no-weight Sulfo-NHS-LC-Biotin in 180μL ultrapure water (purified by Milli-Q Biocel System). Appropriate volume of biotin was added to antibody in order to gain about 20-fold excess biotin-to-antibody molar ratios. The mixture was incubated at room temperature for 2 hr. The biotinylated antibody was then filtered using another desalting column and the final concentration of the biotinylated antibody was measured using a NanoDrop UV spectrophotometer (ThermoFisher) based on absorption at 280 nm.

Magnetic bead preparation: Dynabeads MyOne Streptavidin T1 (Invitrogen, Cat#65601) suspension was transferred to 2.0 mL microcentrifuge tube and placed on the DynaMag-2 magnetic rack followed by aspiration of supernatant. The tube was removed from the magnetic rack and washed with 0.01% Tween-20. Washing step was repeated twice. For blocking purpose, the beads were washed 3 times in PBS containing 0.1% BSA prior to use.

Immunoprecipitation: The immunoprecipitation procedure was performed by incubating the serum with antibody conjugated beads. Briefly, serum from pregnant females was thawed and diluted 1:1 with double filtered 1X PBS (PierceTM 20X PBS, ThermoFisher, Cat#28348). Every 1,000μL of serum was invert-mixed with 6 μg biotinylated antibody for 20 min at RT on a HulaMixer® Sample Mixer (ThermoFisher) at 10 rpm. Then, 390μL of washed Dynabeads was added to the mixture and invert-mixed for 25 min at RT on a Hula mixer at 10 rpm. The mixture was then washed three times with 0.1% BSA and subjected to RNA extraction.

RNA extraction from Dynabeads: RNA was extracted using the miRNeasy mini kit (QIAGEN, Cat#217004) following manufacturer's protocol. In brief, the Dynabeads were subjected to phenol/chloroform extraction step for RNA extraction using Qiazol Lysis Reagent (QIAGEN, Cat#79306) followed by chloroform. The aqueous phase was used as input into the miRNeasy procedure and the RNA was eluted in 14 μL of nuclease-free water. To avoid contamination with genomic DNA, the RNA samples were also treated with deoxyribonuclease I (DNase I, Invitrogen). The quality of RNA was assessed by using the RNA 6000 Nano Pico Kit (Agilent Technologies, Cat#5067-1513) and the Bioanalyzer 2100 (Agilent Technologies). The eluted RNA was dried down using a speedvac, and used as input into the small RNaseq library preparation process. Small RNaseq libraries were generated and size selected as described below.

RNA extraction and small RNA sequencing library construction

Serum

RNA was extracted from 500 μL of maternal serum using the miRNeasy Micro kit (QIAGEN) according to the manufacturer's protocol with a few modifications. Briefly, 2.5 mL of the QIAzol Lysis Reagent was added to the serum and incubated for 5 min. To this 500 μL of chloroform was added, incubated for 3 min and centrifuged for 15 min at 12,000 x g at 4°C. The RNA in the aqueous phase was precipitated by adding 1.5 x volumes of 100% ethanol and then loaded on to MinElute spin column and centrifuged at 1,000 x g for 15 s. The columns were then washed with 700 μL Buffer RWT, 500 μL Buffer RPE and 500 μL 80% ethanol consecutively by centrifuging

for 15 s at ≥ 8000 x g. After a final drying spin at full speed for 5 min, RNA was eluted in 35 μL RNase-free water directly to the center of the spin column membrane and centrifuging for 1 min at 100 x g followed by 1 min at full speed. The RNA was then concentrated using the Zymo RNA Clean and Concentrator-5. 60 μL of the RNA binding buffer and 90 μL of 100% ethanol was added to 30 μL of RNA, transferred to the Zymo-Spin IC columns and centrifuged at 2000xg for 30 s. The column was washed with 700 μL and 400 μL of RNA wash buffer and centrifuged at full speed for 30 s and 2 min respectively. The RNA was then eluted into a final volume of 9 μL RNase free water. The size distribution and quality of the extracted RNA was verified on Agilent RNA 6000 Pico chips using the Agilent 2100 Bioanalyzer instrument.

Tissues

RNA was extracted from the placental and adult tissue samples using the miRVANA miRNA Isolation Kit (Ambion) using the manufacturer's Total RNA protocol. 400μL of frozen (approximately −70°C) 1mm silica/zirconia beads (BioSpec Products PN 10079110Z) were added to each frozen tissue piece, along with 800μL RNA lysis solution, and placed into a MINIBEADBEATER (BioSpec Products) for one minute. The resultant material was immediately centrifuged at 17,000 *xg* for 5 minutes. To the supernatant (600 μL), 60 μL miRNA Homogenate Additive was added, vortexed and incubated on ice for 10 min. 600 μL of Acid Phenol was added, vortexed for 30 s and centrifuged for 5 min at max speed. The aqueous phase was transferred to a fresh tube and 1.25 volumes of 100% ethanol was added. The solution was transferred to filter tube, spun at 10,000 xg for 30 s and then washed once with 700μL wash solution 1 and 2x with 500μL wash solution 2/3 at 10,000 xg for 30 s. After a drying spin max speed for 2 min, RNA was eluted with 100 μL of 95°C RNase free water at max speed for 30 s. The extracted RNA was quantified using the RiboGreen reagent (ThermoFisher). The size distribution and quality of the extracted RNA was verified on Agilent RNA 6000 Nano chips using the Agilent 2100 Bioanalyzer instrument.

Blood cells:

Small RNaseq libraries were prepared from 1.2 μL input RNA using the NEBNext Small RNA Sequencing Library Preparation kit (New England BioLabs), using a mosquito HTS automated nanoliter liquid handler (TTP Labtech). For the automation, the reaction volume was reduced to 1/5th of the manufacturer's recommended volume and the adaptors were diluted to 1/6th of the manufacturer's recommended concentration.

Libraries were then cleaned and concentrated using the Zymo DNA Clean and Concentrator-5 kit (Zymo Research) with a 25 μL elution volume and quantified using the Quant-iT Picogreen DNA Assay High Sensitivity kit (ThermoFisher). The size distributions of the library products were determined using the Agilent High Sensitivity DNA chip on the Agilent 2100 Bioanalyzer instrument. The libraries were then pooled (up to 48 samples/pool) based on their concentrations and their size distribution, to obtain similar numbers of miRNA reads among libraries. The pooled libraries were then subjected to size selection on the PippinPrep instrument to remove unincorporated adapters and primers and adaptor-dimers. For exRNA from biofluid samples, including immunoprecipitation experiments, a 115-180 bp size selection window was used and for placental tissues samples, a 120-135 bp window was used.

Small RNA sequencing

Libraries were sequenced on a HiSeq 4000 with single-end 75 bp reads. Some samples were sequenced more than once (technical replicates) to obtain higher number of reads.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis

Clinical data

Clinical data were analyzed using Student t- test and Mann-Whitney U Test where appropriate (SPSS version 24).

Biofluid exRNA small RNaseq data:

Small RNaseq data from the exRNA samples were processed, including adaptor trimming and mapping to miRBase (miRbase v.21) to yield Raw Count data, using the ExceRpt small RNA sequencing data analysis pipeline version 4.6.2 with minimum insert length set at 10 nt and no mismatches permitted on the Genboree workbench (http://genboree.org/theCommons/projects/exrna-tools-may2014/wiki/Small%20RNA-seq%20Pipeline).[60–62]

The Placental Dysfunction Clinic samples were unblinded and included in the Discovery set. The Sera samples were initially blinded and were divided between Discovery and Verification sets, in a manner that resulted in a similar distribution of gestational age at blood draw (GABD), and of the proportions of preeclampsia cases to non-preeclamptic controls across all GABD and in 1- and 3-week windows of GABD between the Discovery and Verification sets.

Filtering was performed to remove individual miRNAs with > 70% missing values. The raw counts for the remaining miRNAs are provided in Table S2, Sheet 1. Read counts were log2 transformed. Sample-to-sample normalization was carried out through stabilization of variance and reduction in bias across distributions of read counts. Variance stabilizing transformation and bias reduction are useful for making high- and low- read-count samples and miRs more tractable, as stabilizing variance reduces heteroskedasticity and reducing bias removes sample-wide mean shifts. The PEER package (Sanger Institute) was run to reduce batch effects while retaining biological variation.[37] Replicate data was then collapsed to single values. AUCs were generated with the pROC package, using the Delong and bootstrap methods to establish the confidence intervals (CIs).[63–65] Analysis was performed using R 3.4.3. Normalized data are provided in Table S2, Sheet 2.

Four windows of GABD were considered: full window (119-196 days), early window (119-152), middle window (138 – 172 days), and late window (156 – 196 days). Univariate miRNA models were fit to the entire dataset range of gestational age at blood draw (GABD) and to early, middle and late GABD windows.

For the Discovery phase, univariate models with significant chi-square p values (p value < 0.05) between residual and null deviance were selected for each GABD window (Table S4, Sheet 1). Univariate models for which the lower confidence interval (CI) area under the curve (AUC) was > 0.5 were considered to have passed Verification (Tables 3 and S4, Sheet 1 and Sheet 3).

In Discovery, bivariate reversals (ratios of log of miRNA counts) were ranked by an inverse rank sum using 1000 bootstraps.[66] For each iteration, the AUC, the squared correlation between the ratio and a 1/0 conversion of the diagnosis column (PE = 1, control = 0), and the mean difference between cases and controls were calculated. Five ranks were derived from the resulting statistics across the 1000 iterations: 1) the mean of the AUCs of the reversal; 2) the lower 25% quantile of the AUCs of the reversal; 3) the mean of the squared correlation; 4) the lower 25% of the squared correlation; and 5) the square of the differences between the case/control mean shift. Each rank was then inverted and summed for each reversal to obtain the final ranking (Table S4, Sheet 2). Reversals were considered to pass Verification if their lower CI did not cross 0.5 in the Verification dataset using the DeLong method for CI calculation (Tables 3 and S4, Sheet 2 and Sheet 3).

Power analysis examined the power of the Verification set to detect non-random classifier performance, based on a one-sided test (power.roc.test, pROC package) for confidence intervals not containing an AUC of 0.5. Results estimated that the blood draw windows containing one-third of the Verification set would have 80% power to detect: AUCs of 0.65 whose 60% confidence intervals did not include 0.5; AUCs of 0.7 whose 80% confidence intervals did not include 0.5; AUCs of 0.75 whose 90% confidence intervals did not include 0.5; and AUCs of 0.8 whose 95% confidence intervals did not include 0.5.

Placenta and Adult Tissue small RNaseq data

Small RNaseq data from the tissue samples were trimmed and mapped using the exceRpt pipeline[36] version 4.6.2 with minimum insert length set at 15 nt and no mismatches permitted. Scaled data (expressed as reads per million total miRNA reads) are provided in Table S5, Sheet 1, and scaled expression values averaged for each miRNA and each cell/tissue type are provided in Table S5, Sheet 2). Differential expression analysis using the sample level data (Table S5, Sheet 1) was performed using the Multigroup Comparison function in Qlucore (qlucore.com); data for highly significantly (q-value $< 10^{-12}$) differentially expressed miRNAs are shown in Table S5, Sheet 3.

Estimation of fractional contribution of each cell/tissue type to each miRNA

First, deconvolution analysis to calculate the fractional contribution of each cell/tissue type to the overall miRNA content of maternal serum was performed using the CIBERSORT package,[67] which employs a linear support vector regression model to estimate proportions. To construct the input dataset, we took the intersection between the miRNAs that passed our detection filter for the maternal serum extracellular miRNA data (Table S2, Sheet 1) and were differentially expressed among tissue types (Table S5, Sheet 1). For the "gene expression signature" input file (Table S5, Sheet 4), we extracted the miRNA expression data averaged for each cell/tissue type for the miRNAs in this intersect set from Table S5, Sheet 2. For the "gene expression profile" input file (Table S5, Sheet 5) we extracted the raw extracellular miRNA expression data for each exRNA sample in the Discovery and Verification cohorts for the miRNAs in this intersect set from Table S2, Sheet 1. The CIBERSORT output is provided in Table S5, Sheet 6, and shows for each exRNA sample the percent of the overall miRNA profile accounted for by each cell/tissue type.

The results from the deconvolution analysis were then combined with the miRNA expression values averaged for each cell/tissue type (Table S5, Sheet 1) to calculate the contribution of each cell/tissue type to the total expression level for each miRNA. Specifically, for each miRNA and cell/tissue type, the expression of that miRNA in that cell/tissue type was multiplied by the fractional contribution of that cell/tissue type to the overall miRNA profile of maternal serum (averaged across all of the exRNA samples). The resulting values were scaled across cell/tissue types to compute the percent of that miRNA present in maternal serum that was contributed by each cell/tissue type (Table S5, Sheet 7).

Small RNaseq data from immunoprecipitation experiments

Small RNaseq data from the tissue samples were trimmed and mapped using the exceRpt pipeline[36] version 4.6.2 with minimum insert length set at 15 nt and no mismatches permitted. Scaled data (expressed as reads per million total miRNA reads) are shown in (Table S6, Sheet 1). Multigroup differential expression analysis was performed using Qlucore (Qlucore.com) and miRNAs that were significantly differentially expressed ($q < 0.05$) between at least 2 groups (input, CD63, AGO2, PLAP) were identified (Table S6, Sheet 2).