Detection and Identification of Novel Metabolomic Biomarkers in Preeclampsia

Louise C. Kenny, MBChB (Hons), PhD, MRCOG, David Broadhurst, PhD, Marie Brown, PhD, Warwick B. Dunn, PhD, Christopher W. G. Redman, MBBChir, MA, FRCP, FRCOG, Douglas B. Kell, BA (Hons), MA, DPhil, and Philip N. Baker, DM, FRCOG

In a previous study, the ability of gas chromatography time-of-flight mass spectrometry to detect potential metabolic biomarkers in preeclampsia was demonstrated. In this study, the authors sought to validate their preliminary findings in an entirely different patient cohort using a complementary, novel, and powerful combination of analytical tools (ultra performance liquid chromatography and LTQ Orbitrap mass spectrometry system). Eight metabolites that appeared in the authors' previous patient cohort were identified as being statistically significant ($P < .01$) as discriminatory biomarkers. The chemical identities of these 8 metabolites were established using authentic chemical standards. They included uric acid, 2-oxoglutarate, glutamate, and alanine. This is the first study to identify, in an unbiased manner, a series of small-molecular-weight metabolites that effectively detect preeclampsia in plasma. The identity of these metabolites provides new insights into the pathology of this condition and raises the possibility of the development of a predictive test.

Key words: Preeclampsia, biomarkers, metabolomics.

Preeclampsia is a common pregnancy-associated disorder that complicates 2% to 5% of pregnancies. The disease remains a leading cause of maternal death and is responsible for significant prenatal morbidity and mortality, accounting for up to 20% of all neonatal intensive care admissions.[1] Furthermore, preeclampsia carries health care implications in adult life. Infants born of pregnancies complicated by preeclampsia are at an increased risk of hypertension, heart disease, and diabetes.[2]

The pathogenesis of preeclampsia is thought to involve inappropriate adaptation of the interface between the maternal vasculature and the developing placenta early in pregnancy, which subsequently leads to the development of a poorly perfused fetoplacental unit.[3,4] In this model, continuing poor perfusion of the placenta is proposed to result in the secretion of factor(s) into the maternal circulation. These cause "activation" of the vascular endothelium. The clinical syndrome of preeclampsia results from widespread changes in endothelial cell function in both small and large vessels.[5-7]

There is currently no accurate way of predicting preeclampsia, and consequently, clinicians are unable to offer either targeted surveillance or potential preventative therapies to those at greatest risk. Widespread plasma alterations precede the clinical onset of preeclampsia and, therefore, there is intense interest in the identification of predictive biomarkers. Candidate proteins have been investigated as risk determinants for preeclampsia, both in isolation and in combination with other markers, but have limited sensitivity and specificity.[3] Preeclampsia is

From the Maternal and Fetal Health Research Centre, University of Manchester, St Mary's Hospital, Manchester, United Kingdom (LCK, PNB); the Department of Obstetrics and Gynaecology, University College Cork, Cork University Maternity Hospital, Wilton, Cork, Ireland (LCK, DBK); School of Chemistry and the Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, the University of Manchester, Manchester, United Kingdom (DB, MB, WBD, D/K); Nuffield Department of Obstetrics and Gynaecology, John Radcliffe Hospital, University of Oxford, Oxford, United Kingdom (CWGR); and School of Medicine and the Manchester Interdisciplinary Biocentre, the University of Manchester, Manchester, United Kingdom (PNB).

DB, MB, and WBD contributed equally to this work.

Address correspondence to: Louise C. Kenny, MBChB (Hons), PhD, MRCOG, Department of Obstetrics and Gynaecology, University College Cork, Cork University Maternity Hospital, Wilton, Cork, Ireland. E-mail: l.kenny@ucc.ie.

Reproductive Sciences Vol. 15 No. 6 July 2008 591-597
DOI. 10.1177/1933719108316908

Table 1. Demographic Data for Patients From Whom Plasma Samples Were Taken[a]

	Normal Outcome (n = 20)	Preeclampsia (n = 20)
Parity	0 (0-4)	0 (0-4)
Gestational age at sampling (days)	243 (186-282)	245 (186-272)
Systolic BP at booking (mm Hg)	100 (90-130)	120 (90-150)
Diastolic BP at booking (mm Hg)	60 (50-80)	74 (60-90)
Maximum systolic (mm Hg)	100 (90-130)	170 (140-220)[b]
Maximum diastolic (mm Hg)	60 (50-82)	115 (100-130)[b]
Delivery gestation (days)	283 (272-292)	253 (191-274)[b]
Birth weight (g)	3369 (2857-4168)	2153.5 (630-3675)[b]

Abbreviation: BP, blood pressure.
[a]Data provided as median (range).
[b]Preeclampsia versus normal outcome, $P < .05$.

undoubtedly a multisystem disorder, and the manifestations of the disease seem unlikely to be related to a single protein.

We have therefore adopted a new approach to identifying biomarkers by focusing on metabolic changes at the time of diagnosis. Metabolomics is the data-driven study of the different patterns of metabolites within living organisms, tissues, and cells. Although metabolomics is complementary to transcriptomics and proteomics, it also has several distinct advantages, in particular the fact that metabolomics changes are amplified (both experimentally and on theoretical grounds) over those of the proteome and the transcriptome.[8,9]

We previously reported preliminary results of an anonymous metabolomic screen of plasma from women with preeclampsia and described (but did not then identify chemically) 3 highly discriminatory metabolites.[10] However, we did not at that time have access to a separate patient cohort for the necessary validation assays.[11] In the present study, we have used a separate technology and have studied plasma derived from a different population of women with preeclampsia to validate our preliminary findings and to identify the metabolites involved.

METHODS

Participants

Plasma samples (cohort 1) were obtained from 20 primiparous Caucasian women with preeclampsia diagnosed according to the International Society for the Study of Hypertension in Pregnancy (ISSHP) guidelines from a single maternity unit in England.[12] Controls were obtained from the same antenatal population and were matched for maternal age, parity, and body mass index and for gestational age at sampling. Plasma samples were only retained from controls for this study if they subsequently experienced an uncomplicated pregnancy. Further demographic details for this cohort are provided in Table 1.

In our previous study,[10] plasma samples (cohort 2) were obtained from the GOPEC archive. GOPEC was a British Heart Foundation–funded multicenter collaborative study involving 10 UK university departments of obstetrics and gynaecology. Within this study, 1000 low-risk Caucasian primiparous women who nevertheless developed preeclampsia, defined according to ISSHP guidelines, were recruited and sampled between 1999 and 2003. The study protocol is described in detail elsewhere.[13] Eighty-seven women within the archive who had donated blood to the study antenatally (after diagnosis and within a week prior to delivery) were identified and were matched with 87 normal pregnant controls for maternal age, parity, and body mass index and for gestational age at sampling. Controls were obtained from antenatal clinics in Manchester and Dundee.

The 2 studies combined make up a substantial cohort of 216 patients. Sourcing these cohorts from different locations, at different times, and analyzing their plasma samples on different mass spectral technology means that any metabolites discovered across both cohorts are likely to be robust and not due to any location/time/technology confounding factors.

Sample Collection

Blood samples were taken at the time of recruitment. Samples were collected into precooled glass tubes containing ethylenediaminetetraacetic acid using the Vacutainer system (Becton Dickinson, Franklin Lakes, New Jersey) and immediately centrifuged at 1500*g* for 15 minutes at

4°C. Plasma was then removed and stored in aliquots at −80°C until required. The collection and storage conditions were identical for samples taken from both patients and controls in both archives.

Technology

The LTQ Orbitrap (Thermo Fisher Scientific, Waltham, Massachusetts) is a high-mass electrospray ionization mass spectrometer that allows very high accuracy registration of atomic mass/charge.[14,15] The accurate determination of molecular and fragment ion masses can assist greatly in assigning the elemental composition of an unknown compound and, consequently, help with chemical structure determination. This is because a mass measurement accuracy of approximately 1 to 3 ppm, especially in conjunction with isotope analyses,[16] allows the direct estimation of molecular formulae. In combination with small particle/ultrahigh pressure liquid chromatographic separations, in this example using an Acquity UPLC system (Waters Corp, Milford, Massachusetts), a combination of high chromatographic and mass resolution and high mass accuracy allows a previously unachievable information-rich biological picture to be obtained.

Sample Preparation for Cohort 1

All reagents used were of HPLC grade purity (Sigma-Aldrich CHROMASOLV, Dorset, United Kingdom).

Serum samples were prepared by spiking 200-μL aliquots with an internal standard solution (50 μL; 0.17 mg/mL succinic d_4 acid, malonic d_2 acid, glycine d_5 in water), vortex mixing for 15 seconds, and deproteinization by addition of 600 μL methanol followed by vortex mixing for 15 seconds and centrifugation (13 487*g*, 15 minutes). Supernatants were transferred to Eppendorf tubes and lyophilized (HETO VR MAXI vacuum centrifuge attached to a HETO CT/DW 60E cooling trap; Jouan, Gydevang, Denmark). Samples were reconstituted in 200 μL water prior to analysis.

Analysis by Acquity UPLC Coupled to LTQ Orbitrap

Samples were analyzed, in a random order, by an Acquity UPLC coupled to a LTQ-Orbitrap mass spectrometry system (Thermo Fisher Scientific, Bremen, Germany) operating in electrospray ionization mode. Chromatographic separations were performed at a flow rate of 0.4 mL/min employing an Acquity UPLC BEH 1.7-μm C_{18} column (2.1 × 100 mm; Waters Corp) using an ultra performance liquid chromatography system (Waters Corp). The column was eluted with 0.1% formic acid in water (A) and 0.1% formic acid in methanol (B). The column was held at 100% A for 1 minute and subsequently ramped to 100% B (curve 5) over 15 minutes, followed by a 4-minute hold at 100% B before a rapid return to 100% A and a hold for 2 minutes. The column and samples were maintained at temperatures of 50°C and 10°C, respectively. A 10-μL sample volume was introduced onto the column. Fifty percent of the column effluent was transferred to the mass spectrometer.

Samples were analyzed twice, once in positive and once in negative ion mode. Centroid mass spectrometry scans were acquired in the mass range of 50 to 1000Th using the Orbitrap mass analyzer operating with a target mass resolution of 30 000 (full width at half maximum as defined at m/z 400) and a scan time of 0.4 seconds. Mass accuracies predominantly were observed to be less than 3 ppm over a wide dynamic range with external mass calibration.

Identification of Metabolites by Exact Mass Matching

Identification of possible chemical formulae for the exact mass was based on the list of formulae at http://fiehnlab.ucdavis.edu/projects/Seven_Golden_Rules/. Matching was done in the negative electrospray ionization data to the $[M - H]-$ adduct and in positive to the $[M + H]$, $[M + Na]$,[17] and $[M + NH_4]+$ adducts.

Primary matching of the possible formulae to compounds in the KEGG (http://www.genome.jp/kegg/) and the Human Metabolome (http://www.hmdb.ca/) databases was undertaken. Secondary matching was performed on data from internal reference metabolite standards, the Lipid Maps Web site (http://www.lipidmaps.org/) and the Japanese metabolites consortium (http://www.metabolome.jp/).

The identities of metabolites of particular interest were confirmed by comparing the raw spectra with those of authentic chemical standards, further matching peaks by retention time and their accurate mass.

Statistical Analyses

The statistical analysis takes the form of a hypothesis-driven experimental validation. The chemically identified

Table 2. Summary of the 8 Discriminatory Metabolites Validated in the Present Study[a]

	Cohort 1				Cohort 2			
Name	KW *P* Value	ANOVA *P* Value	ROC Area Hull	Fold Increase	KW *P* Value	ANOVA *P* Value	ROC Area Hull	Fold Increase
Alanine	2.E-03	1.E-02	0.83	1.22	2.E-04	2.E-05	0.70	1.82
2-Hydroxy-3-methyl-butanoic acid	1.E-02	2.E-02	0.79	1.42	5.E-03	9.E-03	0.68	1.52
2-Ethyl-3-hydroxypropionic acid	4.E-04	1.E-04	0.87	1.35	3.E-06	2.E-06	0.75	1.46
2-Oxoglutaric acid	1.E-02	1.E-02	0.78	1.27	1.E-08	4.E-07	0.73	1.73
Glutamic acid	9.E-03	8.E-03	0.80	1.39	2.E-08	1.E-08	0.78	2.29
Xylitol or ribitol	2.E-03	2.E-03	0.84	1.27	1.E-03	2.E-03	0.71	1.24
Uric acid	5.E-04	1.E-04	0.88	1.38	7.E-05	9.E-03	0.69	2.60
Uric acid*	6.E-06	2.E-07	0.92	1.59				
Creatinine	2.E-03	7.E-04	0.84	1.32	4.E-03	7.E-05	0.67	1.41

Abbreviations: KW, Kruskal–Wallis; ANOVA, analysis of variance; ROC, receiver–operator characteristic.

[a]The table gives the values of the significance as determined from the Kruskal–Wallis test and by standard ANOVA, as well as the area under the convex hull of the ROC curve for the different populations. Creatinine and alanine were detected solely by positive electrospray ionization, whereas uric acid was detected using both polarities (*indicates positive ionization). The other metabolites were detected using negative electrospray ionization. It was not possible to discriminate xylitol from ribitol on the basis of their mass spectra or their retention times.

metabolites found to be significant in our earlier study (cohort 2) were compared with the metabolite peaks found in this study (cohort 1). The exact mass values were calculated for 45 significant metabolites from cohort 2. The raw spectral data for cohort 1 were then searched by matching these exact masses to binned mass windows of width 0.002 *m/z*. If a match was found, then the significance of that peak (peak intensity is defined as the area under a particular retention time window at a particular exact mass window) between "case" and "control" populations was tested (the null hypothesis is that for a given metabolite peak all samples are drawn from the same population, or from 2 populations with the same mean). As this strategy represents an experimental validation, concerns about type I errors are reduced. Thus, the critical *P* value was set to .01. Any match that passed this significance test was then verified using an authentic chemical standard. Before choosing which significance test to apply, each metabolite peak was checked for within-class kurtosis and for within-class goodness of fit to a normal distribution using the Lilliefors test. For a given metabolite peak, if either control or case samples had kurtosis >3, or failed the Lilliefors test (where the null hypothesis, that the sample set has a normal distribution, is rejected if the *P* value <.05), then the nonparametric rank-based Wilcoxon rank sum (Mann–Whitney) test was used[18]; otherwise the classical 1-way analysis of variance test was used. In addition, receiver–operator characteristic (ROC) curves were created (not shown). The area under the ROC curve is considered a good nonparametric indicator of discriminatory ability (0.5 = no discrimination; 1 = perfect discrimination).

All statistical analyses were carried out using the Matlab scripting language (http://www.mathworks.com/). All algorithms used are implemented such that any missing values are ignored. Scripts are available on request.

RESULTS

Because of the uncertainty of reasonable hypotheses about the detailed origins of preeclampsia, we have taken a data-driven metabolomics approach[19] to establish biomarkers that might be diagnostic—and perhaps eventually prognostic—of disease development. As described in the Methods section, our strategy was to take all the significant metabolites from the previous study (cohort 2) and, on the basis of their known molecular masses, search for these in the cohort 1 raw spectra. Of the 45 metabolites that were compared in this way, 8 were matched with a significance value $P < .01$, of which the chemical identities of 7 were validated by running authentic standards and comparing the results both mass spectrally and by their retention index (2-ethyl-3-hydroxypropionate is yet to be verified). Table 2 shows these metabolites with their respective *P* value for each of the cohort tests, whereas Figure 1 shows Box-whisker plots of the distribution (case/control) of these 8 chemically identified markers.

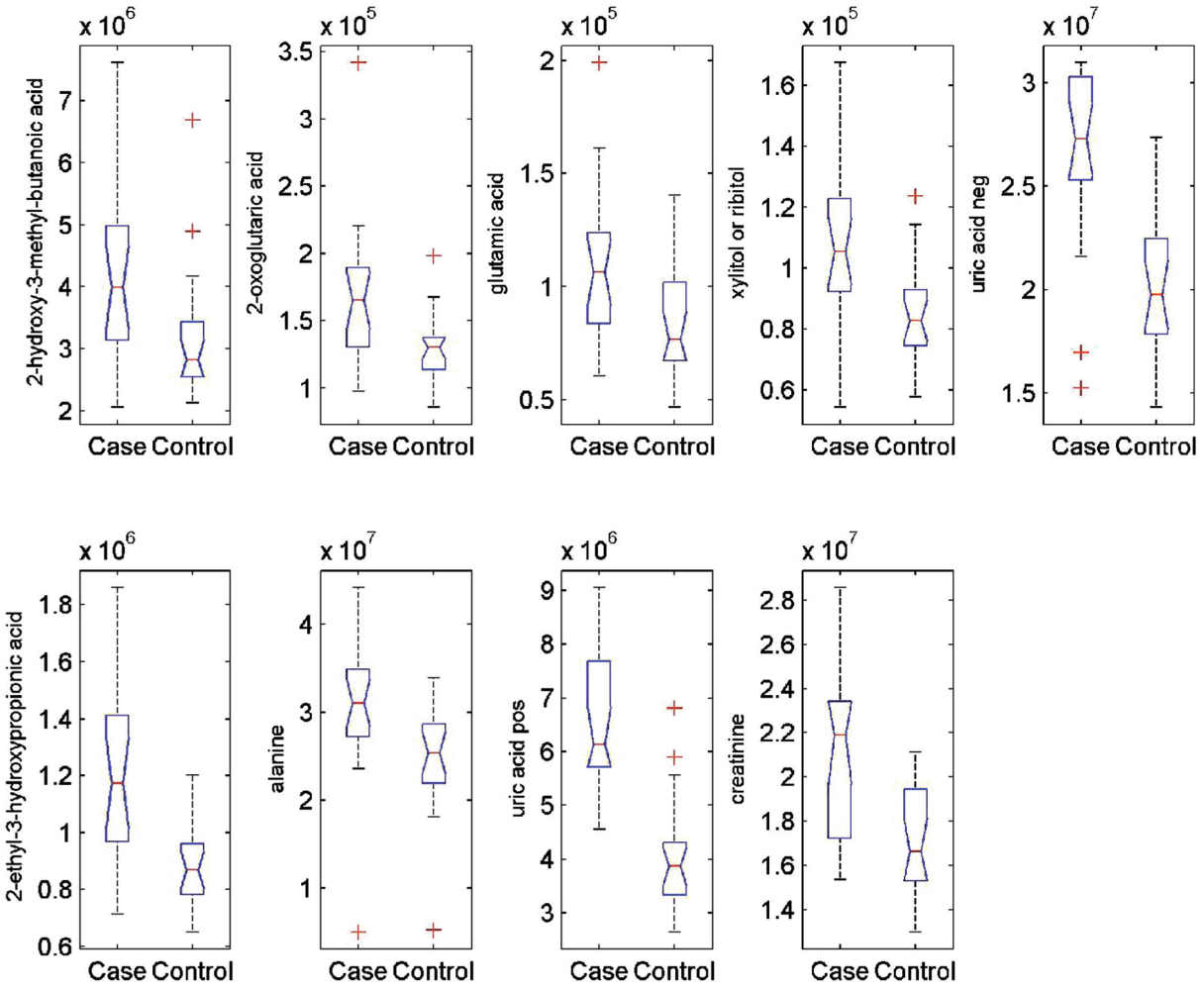

Figure 1. Box-whisker plots illustrating the discrimination between cases and controls in cohort 1 for the 8 discriminatory metabolites identified in the present work. Notches display the variability of the median between samples. The width of a notch is computed so that box plots whose notches do not overlap (as above) have different medians at the 5% significance level. The significance level is based on a normal distribution assumption, but comparisons of medians are reasonably robust for other distributions. Comparing box plot medians is like a visual hypothesis test, analogous to the *t* test used for means. For statistical significance values, see Table 1.

It is comforting that among these was uric acid, which has previously been recognized as diagnostic of clinical disease manifestations.[20] This was also true in a recent study of biomarkers for heart failure,[21] and uric acid may also be observed under ischemic conditions.[22] It is, of course, well known that ischemic injury is a major contributor to the pathology of preeclampsia,[23–28] and another of our discriminatory metabolites was 2-oxoglutarate, a TCA cycle intermediate that represents a partial oxidation product that may be expected to accumulate under conditions of a limited oxidative capacity,[21] and it is closely linked metabolically to glutamate. Sugar alcohols such as xylitol/ribitol also represent a more reduced kind of molecule than their parent sugars, again consistent with the view that they might be sentinels of anoxia or ischemia. We do not have specific hypotheses for the other metabolites, such as 2-hydroxy-3-methylbutanoate and 2-ethyl-3-hydroxypropionate (these do not appear in any of the databases mentioned above save that the former is in http://www.hmdb.ca). However, it is anticipated that the novel identification of such metabolites will contribute to enhanced understanding of the pathogenesis of the condition. It is also possible that these metabolites may be increased prior to the development

of clinical signs and symptoms of the condition and thus this may also lead to the development of predictive/diagnostic tests. Clinicians are currently unable to offer targeted surveillance with the aim of optimizing the time of delivery and reducing the morbidity and mortality associated with preeclampsia. Furthermore, there are several emerging interventional and therapeutic strategies but effective prediction is required before preventative therapy is seen as justifiable and acceptable.

In due time, given the emerging knowledge of the human metabolic network,[29] it will be appropriate to develop a systems biology model that combines the metabolomics with quantitative network information.[8,30]

CONCLUSIONS

Our previous work[10] showed that small molecules could indeed discriminate samples taken from preeclamptics from those of matched controls, but we did not then identify the metabolites. In addition, such studies require validation using a separate cohort.[11] We have now carried out this validation using a novel technology, the Orbitrap mass spectrometer, that allows extremely high mass resolution measurements to be performed. Knowledge of the exact mass of candidate biomarkers and comparison with those of known metabolites led to the identification of no fewer than 8 metabolites that were discriminatory in both the original cohort and the present, new validation set. Armed with knowledge of these metabolites we can now test specific hypotheses about their potential roles in the etiology of this most important cardiovascular disease.

ACKNOWLEDGMENTS

DBK thanks the BBSRC, EPSRC, RSC, and BHF and LCK and PNB thank Tommy's the Baby Charity, NHS NEAT funding, and the BHF for financial support. This is a contribution from the Manchester Centre for Integrative Systems Biology (http://www.mcisb.org).

REFERENCES

1. Confidential Enquiry into Stillbirths and Deaths in Infancy (CESDI). *Fifth Annual Report*. Maternal and Child Health Research Consortium; 1998.
2. Barker DJ, Gluckman PD, Godfrey KM, et al. Fetal nutrition and cardiovascular disease in adult life. *Lancet*. 1993;341:938-941.
3. Hayman R, Brockelsby J, Kenny L, Baker P. Preeclampsia: the endothelium, circulating factor(s) and vascular endothelial growth factor. *J Soc Gynecol Investig*. 1999;6:3-10.
4. Pijnenborg R, Anthony J, Davey DA, et al. Placental bed spiral arteries in the hypertensive disorders of pregnancy. *Br J Obstet Gynaecol*. 1991;98:648-655.
5. Kenny LC, Baker PN, Kendall DA, Randall MD, Dunn WR. Differential mechanisms of endothelium-dependent vasodilator responses in human myometrial small arteries in normal pregnancy and pre-eclampsia. *Clin Sci (Lond)*. 2002;103:67-73.
6. Roberts JM, Taylor RN, Musci TJ, Rodgers GM, Hubel CA, McLaughlin MK. Preeclampsia: an endothelial cell disorder. *Am J Obstet Gynecol*. 1989;161:1200-1204.
7. Rodgers GM, Taylor RN, Roberts JM. Preeclampsia is associated with a serum factor cytotoxic to human endothelial cells. *Am J Obstet Gynecol*. 1988;159:908-914.
8. Kell DB. Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Discov Today*. 2006;11:1085-1092.
9. Raamsdonk LM, Teusink B, Broadhurst D, et al. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol*. 2001;19:45-50.
10. Kenny LC, Dunn WB, Ellis DI, et al. Novel biomarkers for pre-eclampsia detected using metabolomics and machine learning. *Metabolomics*. 2005;1:227-234.
11. Broadhurst D, Kell DB. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics*. 2006;2:171-196.
12. Davey DA, MacGillivray I. The classification and definition of the hypertensive disorders of pregnancy. In: *Hypertension in Pregnancy: the Proceedings of the 16th Study Group of the Royal College of Obstetricians and Gynaecologists*. London, UK: RCOG; 1986:401-407.
13. GOPEC Consortium. Disentangling fetal and maternal susceptibility for pre-eclampsia: a British multicenter candidate-gene study. *Am J Hum Genet*. 2005;77:127-131.
14. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R. The Orbitrap: a new mass spectrometer. *J Mass Spectrom*. 2005;40:430-443.
15. Makarov A, Denisov E, Kholomeev A, et al. Performance evaluation of a hybrid linear ion trap/Orbitrap mass spectrometer. *Anal Chem*. 2006;78:2113-2120.
16. Kind T, Fiehn O. Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics*. 2007;8:105.
17. Friedrich T, Roth M, Helm-Kruse S, Jeltsch A. Functional mapping of the EcoRV DNA methyltransferase by random mutagenesis and screening for catalytically inactive mutants. *Biol Chem*. 1998;379:475-480.
18. Hollander M, Wolfe DA. *Nonparametric Statistical Methods*. New York, NY: Wiley; 1973.
19. Kell DB, Oliver SG. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and

hypothesis-driven science in the post-genomic era. *Bioessays*. 2004;26:99-105.
20. Redman CW, Bonnar J. Plasma urate changes in pre-eclampsia. *BMJ*. 1978;1:1484-1485.
21. Dunn WB, Broadhurst DI, Sasalu D, et al. Serum metabolomics reveals many novel metabolic markers of heart failure, including pseudouridine and 2-oxoglutarate. *Metabolomics*. In press.
22. Sabatine MS, Liu E, Morrow DA, et al. Metabolomic identification of novel biomarkers of myocardial ischemia. *Circulation*. 2005;112:3868-3875.
23. Borzychowski AM, Sargent IL, Redman CW. Inflammation and pre-eclampsia. *Semin Fetal Neonatal Med*. 2006;11:309-316.
24. Hubel CA. Oxidative stress in the pathogenesis of preeclampsia. *Proc Soc Exp Biol Med*. 1999;222:222-235.
25. Redman CW, Sargent IL. Placental debris, oxidative stress and pre-eclampsia. *Placenta*. 2000;21:597-602.
26. Redman CW, Sargent IL. Pre-eclampsia, the placenta and the maternal systemic inflammatory response—a review. *Placenta*. 2003;24(suppl A):S21-S27.
27. Sargent IL, Borzychowski AM, Redman CW. NK cells and human pregnancy—an inflammatory view. *Trends Immunol*. 2006;27:399-404.
28. Sharma JB, Sharma A, Bahadur A,Vimala N, Satyam A, Mittal S. Oxidative stress markers and antioxidant levels in normal pregnancy and pre-eclampsia. *Int J Gynecol Obstet*. 2006;94: 23-27.
29. Duarte NC, Becker SA, Jamshidi N, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci USA*. 2007;104:1777-1782.
30. Kell DB. Metabolomics, machine learning and modelling: towards an understanding of the language of cells. *Biochem Soc Trans*. 2005;33:520-524.