Genomics Proteomics Bioinformatics 11 (2013) 335–344

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com

ORIGINAL RESEARCH

Screening Preeclamptic Cord Plasma for Proteins Associated with Decreased Breast Cancer Susceptibility

Hoi Pang Low [1], Ashutosh Tiwari [2], Jagadeesh Janjanam [2], Li Qiu [1], Chien-I Chang [1], William C. Strohsnitter [3], Errol R. Norwitz [3], Sun W. Tam [4,†], James E. Evans [4,‡], Karin M. Green [4], Joao A. Paulo [5], Mats Lambe [6], Chung-Cheng Hsieh [1,*]

[1] *Department of Cancer Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA*
[2] *Department of Chemistry, Michigan Technological University, Houghton, MI 49931, USA*
[3] *Department of Obstetrics and Gynecology, Tufts Medical Center, Boston, MA 02111, USA*
[4] *Proteomics and Mass Spectrometry Facility, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01545, USA*
[5] *Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA*
[6] *Department of Medical Epidemiology and Biostatistics, Karolinska Institute, SE171 77 Stockholm, Sweden*

Received 10 June 2013; revised 21 August 2013; accepted 3 September 2013
Available online 5 December 2013

KEYWORDS

Biomarkers;
Intrauterine environment;
Mass spectrometry;
Pregnancy;
Prenatal;
Preeclampsia

Abstract Preeclampsia, a complication of pregnancy characterized by hypertension and proteinuria, has been found to reduce the subsequent risk for breast cancer in female offspring. As this protective effect could be due to exposure to preeclampsia-specific proteins during intrauterine life, the proteomic profiles of umbilical cord blood plasma between preeclamptic and normotensive pregnancies were compared. Umbilical cord plasma samples, depleted of 14 abundant proteins, were subjected to proteomic analysis using the quantitative method of nanoACQUITY ultra performance liquid chromatography–mass spectrometry with elevated energy mode of acquisition^E (NanoUPLC-MS^E). Sixty-nine differentially expressed proteins were identified, of which 15 and 6 proteins were only detected in preeclamptic and normotensive pregnancies, respectively.

* Corresponding author.
E-mail: chung-cheng.hsieh@umassmed.edu (Hsieh CC).
† Current address: Department of Chemistry and Biochemistry, Clark University, Worcester, MA 01610, USA.
‡ Current address: Roskamp Institute Inc., Sarasota, FL 34243, USA.
Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

ELSEVIER Production and hosting by Elsevier

1672-0229/$ - see front matter
http://dx.doi.org/10.1016/j.gpb.2013.09.009

Additionally, expression of 8 proteins (gelsolin, complement C5, keratin type I cytoskeletal 10, pigment epithelium-derived factor, complement factor B, complement component C7, hemoglobin subunit gamma-2 and alpha-fetoprotein) were up-regulated in preeclampsia with a fold change of ⩾2.0 when compared to normotensive pregnancies. The identification of alpha-fetoprotein in preeclamptic umbilical cord blood plasma supported the validity of this screen as alpha-fetoprotein has anti-estrogenic properties and has previously been linked to preeclampsia as well as a reduced breast cancer risk. The findings of this pilot study may provide new insights into the mechanistic link between preeclampsia and potentially reduced breast cancer susceptibility in adult life.

Introduction

Breast cancer is the most common cancer that affects women in the United States. However, there are few effective interventions to lower breast cancer risk. Epidemiological and experimental studies suggest that the intrauterine environment influences breast cancer risk in the offspring [1]. Among the maternal, gestational and newborn characteristics, strong inverse associations with breast cancer risk have been found for prenatal exposure to preeclampsia [1–3]. In a large population-based study that examined prenatal factors and adult risk of breast cancer in a cohort of Swedish women, women born from a preeclamptic (PE) pregnancy had a 59% reduction in breast cancer risk (relative risk 0.41, 95% confidence interval 0.22–0.79) [2]. Additionally, a meta-analysis has shown that among all women, PE-born offspring had a 52% lowered risk of breast cancer (relative risk 0.48, 95% confidence interval 0.30–0.78) [3]. This potential protective effect is as strong as that associated with tamoxifen/raloxifene intervention [4] or physical exercise [5].

Preeclampsia is a complication of 5–8% of all pregnancies in the United States that brings significant morbidity and mortality for both mother and baby [6,7]. The hallmark characteristics of preeclampsia are a new-onset of hypertension and proteinuria after 20 weeks of gestation in a previously normotensive (N) woman. Although the underlying causes of the syndrome remain obscure, abnormal placenta development has been implicated [8]. There is also no compelling explanation for the inverse association between preeclampsia and subsequent breast cancer risk in the offspring, but a possible influence of the intrauterine environment has been implicated. Although preeclampsia has been associated with higher levels of umbilical cord alpha-fetoprotein (AFP) [9], insulin-like growth factor binding protein-1 (IGFBP-1) [10], leptin [11], triglycerides [12] and homocysteine [13], none of these efforts have included a proteomic screen on PE umbilical cord plasma to identify intrauterine biomarkers that might act synergetically or individually to be candidate cancer risk reduction molecules, as has been reported for AFP, a glycoprotein with anti-estrogenic properties [9,14].

As we hypothesize that the fetus of a PE pregnancy is exposed to preeclampsia-specific proteins during gestation that confer protection against breast cancer in the adult life, we performed a pilot study to compare the protein profile of umbilical cord plasma from PE pregnancies with that from N pregnancies by the quantitative proteomic method of nanoACQUITY ultra performance liquid chromatography–mass spectrometry with elevated energy mode of acquisition^{E} (NanoUPLC-MS^{E}) [15]. Absolute quantification of proteins by LCMS^{E} is a unique technology implemented on Q-TOF mass spectrometers, where accurate mass LCMS data were collected in an alternating low energy (MS) and elevated energy (MS^{E}) mode of acquisition [15]. Here we report the identification and quantification of 69 proteins in PE and N umbilical cord blood plasma, of which 15 and 6 proteins were only detected in PE and N pregnancies, respectively. Additionally, compared to N pregnancy, expression of 8 proteins was up-regulated more than twofold and expression of 7 proteins was down-regulated more than 1.5 fold in umbilical cord blood plasma from PE pregnancy.

Results

Identification of differentially expressed proteins between PE and N umbilical cord plasma

Screening of umbilical cord blood plasma samples by NanoUPLC-MS^{E} (as outlined in **Figure 1**) identified a total of 69 proteins (Table S1), of which 15 (**Table 1**) and 6 (**Table 2**) proteins were exclusive to PE and N pregnancies, respectively (**Figure 2**). An additional 48 proteins (Table S2) were detected in umbilical cord blood plasma from both PE and N pregnancies (Figure 2).

Identification of proteins with up-regulated expression in PE plasma

As daughters of PE births have a reduced risk for breast cancer [2,3], we hypothesize that proteins exclusive to umbilical cord plasma from PE pregnancies or up-regulated in PE plasma may play a role in reducing breast cancer risk. Of the 15 proteins that were exclusive to PE plasma (*i.e.*, absent in N plasma), hemoglobin subunit alpha (P69905) was present in the largest quantity of approximately 190 fmol and with a mean spectral count of 4.7 (Table 1). Additionally, serum amyloid

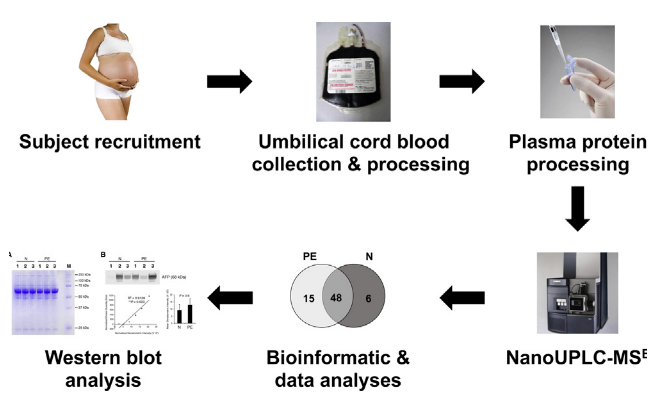

Figure 1 Experimental workflow

Table 1 Proteins exclusively expressed in preeclamptic umbilical cord plasma

Accession No.	Protein	Mean protein quantity (fmol)	Mean spectral counts
P69905	Hemoglobin subunit alpha	189.95	4.70
P02743	Serum amyloid P-component	27.12	1.44
P02751	Fibronectin	26.60	0.72
P02538	Keratin, type II cytoskeletal 6A or keratin 6A	23.48	0.55
P02671	Fibrinogen alpha chain	22.77	1.63
Q06033	Inter-alpha-trypsin inhibitor heavy chain H3	22.22	1.61
P02452	Collagen alpha-1(I) chain	16.65	0.55
P05543	Thyroxine-binding globulin or serpin peptidase inhibitor clade A member 7	10.21	0.67
P04004	Vitronectin	7.97	0.33
P02533	Keratin, type I cytoskeletal 14 or keratin 14	5.63	0.92
Q6UXB8	Peptidase inhibitor 16	5.16	0.25
P05155	Plasma protease C1 inhibitor or serpin peptidase inhibitor clade G member 1	5.11	0.56
P00736	Complement C1r subcomponent	3.15	0.33
P03952	Plasma kallikrein B	2.95	0.25
Q96PD5	N-acetylmuramoyl-L-alanine amidase	2.72	0.33

Note: Proteins are ranked in descending order based on mean protein quantity.

Table 2 Proteins exclusively expressed in normotensive umbilical cord plasma

Accession No.	Protein	Mean protein quantity (fmol)	Mean spectral counts
P01620	Ig kappa chain V-III region SIE	22.68	1.07
P01766	Ig heavy chain V-III region BRO	20.13	1.07
P00734	Prothrombin	11.20	1.28
P02647	Apolipoprotein A-I	9.46	0.45
P13671	Complement component C6	4.49	0.96
Q6S8J3	POTE ankyrin domain family member E	2.38	0.64

Note: Proteins are ranked in descending order based on mean protein quantity.

P-component (SAP; P02743), fibronectin (P02751), keratin type II cytoskeletal 6A or keratin 6A (P02538), fibrinogen alpha chain (P02671) and inter-alpha-trypsin inhibitor heavy chain H3 (Q06033) were present in amounts >20 fmol (Table 1).

To identify proteins that were up-regulated in PE relative to N plasma, we compared the mean quantities of each of the 48 proteins that were detected in both groups and identified 13 proteins that were up-regulated in PE plasma with a fold change $\geqslant 2.0$ when compared to that of N pregnancies (Table S1). An additional filtering criterion using a Bayes factor value >1 was applied and further reduced the number of proteins from 13 to 8 (**Table 3**). These proteins include complement C5 (C5; P01031), keratin type I cytoskeletal 10 or keratin 10 (P13645) and pigment epithelium-derived factor or serpin peptidase inhibitor clade F member 1 (PEDF or SERPINF1; P36955) that have a fold change >3 and a Bayes factor >5. Of note, this proteomic screen also detected the previously reported AFP (P02771) [9] (Table 3).

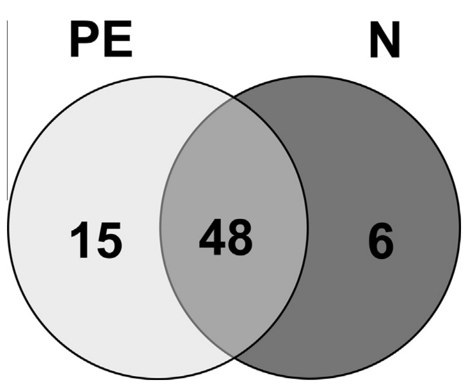

Figure 2 Venn diagram illustrates differentially expressed proteins
A Venn diagram shows that a total of 69 differentially expressed proteins were identified by comparing umbilical cord blood plasma from preeclamptic (PE) and normotensive (N) pregnancies, of which 15 and 6 proteins were detected only in PE and N plasma, respectively, whereas the remaining 48 proteins were present in both plasma.

Identification of proteins with down-regulated expression in PE plasma

Conversely, proteins with expression that was only detected or up-regulated in umbilical cord blood plasma from N pregnancies (Table 2) when compared to PE pregnancies (**Table 4**) are candidate proteins for increased breast cancer risk. These include proteins that were absent in PE plasma such as immunoglobulin (Ig) kappa chain V-III region SIE (P01620), Ig heavy chain V-III region SIE (P01766), prothrombin (P00734), apolipoprotein A-I (Apo A-I; P02647), complement component C6 (C6; P13671) and POTE ankyrin domain family member E (Q6S8J3) (Table 2). Of the 48 proteins that were detected in both PE and N plasma, expression of 13 proteins was up-regulated in N plasma with a fold change $\geqslant 1.5$ when compared to PE plasma (Table S1). Applying an additional filter criterion of identifying proteins with a Bayes factor >1

Table 3 Proteins with up-regulated expression in preeclamptic umbilical cord plasma

Accession No.	Protein	Mean normalized protein quantity (fmol)		Fold change (fmol)	Mean normalized spectral counts		Bayes factor	FDR
		N	PE		N	PE		
P06396	Gelsolin	7.54	58.69	7.79	0.80	3.85	1.16	0.709
P01031	Complement C5	14.44	95.14	6.59	2.02	8.81	6.12	0.000
P13645	Keratin, type I cytoskeletal 10 or keratin 10	134.35	509.73	3.79	10.22	24.41	5.10	0.000
P36955	Pigment epithelium-derived factor or serpin peptidase inhibitor clade F member 1	13.32	43.73	3.28	1.09	6.24	9.82	0.000
P00751	Complement factor B	50.01	144.62	2.89	5.54	14.88	1.43	0.515
P10643	Complement component C7	8.99	22.62	2.52	0.64	2.06	1.11	0.739
P69892	Hemoglobin subunit gamma-2	292.23	694.56	2.38	5.62	11.84	1.19	0.694
P02771	Alpha-fetoprotein	253.51	504.11	2.00	12.79	15.32	1.07	0.755

Note: Only proteins with fold change (fmol) $\geqslant$2.00 and Bayes factor >1 are listed, which are ranked in descending order of fold change (fmol). Fold change values are obtained by dividing mean fmol values of preeclamptic (PE) with mean fmol values of normotensive (N) pregnancies. Values for Bayes factor and FDR are obtained from QSPEC analysis of normalized spectral counts. FDR stands for false discovery rate.

Table 4 Proteins with up-regulated expression in normotensive umbilical cord plasma

Accession No.	Protein	Mean normalized protein quantity (fmol)		Fold change (fmol)	Mean normalized spectral counts		Bayes factor	FDR
		Normal	PE		Normal	PE		
P01009	Alpha-1-antitrypsin or serpin peptidase inhibitor clade A member 1	269.32	40.78	6.60	12.99	2.28	13.51	0.000
P08603	Complement factor H	238.23	40.71	5.85	30.86	8.28	1.33	0.635
P01834	Ig kappa chain C region	292.89	105.74	2.77	9.73	4.75	1.43	0.562
P01857	Ig gamma-1 chain C region	595.89	273.00	2.18	25.06	7.96	12.24	0.000
P51884	Lumican	41.86	24.36	1.72	3.14	1.93	1.28	0.667
P01023	Alpha-2-macroglobulin	1421.84	915.95	1.55	91.00	69.11	3.36	0.002
P02765	Alpha-2-HS glycoprotein	597.24	399.98	1.50	27.66	11.79	1.08	0.771

Note: Only proteins with fold change (fmol) $\geqslant$1.5 and Bayes factor >1 are listed, which are ranked in descending order of fold change (fmol). Fold change values are obtained by dividing mean fmol values of normotensive (N) with mean fmol values of preeclamptic (PE) pregnancies. Values for Bayes factor and FDR are obtained from QSPEC analysis of normalized spectral counts. FDR stands for false discovery rate.

decreased the number of proteins to 7 (Table 4). These proteins include alpha-1-antitrypsin or SERPINA1 (P01009) with the highest fold change of 6.6, followed by complement factor H (Factor H; P08603) with a fold change of approximately 5.9, Ig kappa chain C region (P01834), Ig gamma-1 chain C region (P01857), lumican (P51884), alpha-2 macroglobulin (P01023) and alpha-2-HS glycoprotein (P02765) (Table 4).

Protein functions and pathways

Proteins with expression exclusive (Table 1) or up-regulated in (Table 3) PE plasma (n = 23), and similarly proteins with expression absent (Table 2) or down-regulated (Table 4) in PE plasma (n = 13), were pooled and interrogated using the DAVID interface for their functions, cellular components, and participation in any pathways. Two proteins, both Ig chain regions (P01620 and P01766), were not recognized by DAVID. First, functional classification of proteins with exclusive or up-regulated expression, but not those with absent or down-regulated expression, in PE plasma resulted in a cluster consisting of 5 proteins, mainly components of the complement system, *i.e.*, complement C1r subcomponent (C1r), C5, C7, complement factor B (Factor B) and plasma protease C1 inhibitor or serpin peptidase inhibitor clade G member 1 (SERPING1). This observation was supported by KEGG pathway analysis, although 2 proteins with absent or down-regulated expression in PE plasma, namely C6 and Factor H, also participate in the complement cascade (Figure S1). Second, molecular function analysis identified a set of proteins (n = 5) whose expression were exclusive to or up-regulated in PE plasma, which display structural functions. These proteins include collagen alpha-1(I) chain, fibronectin, keratin 14, keratin 6A and keratin 10), which, except for fibronectin, played a role in ectoderm/epidermis development (Tables S3 and S4). Additionally, molecular functions that are unique to proteins with exclusive or up-regulated expression in PE plasma include serine-type endopeptidase activity (3 proteins: C1r, plasma kallifrein B and Factor B) and oxygen transporter activity (2 proteins: hemoglobin subunit alpha and hemoglobin subunit gamma 2) (Table S3).

The set of proteins with exclusive or upregulated expression share similar functions with those whose expression is either absent or downregulated in PE plasma. These include enzyme peptidase inhibitor activities such as endopeptidase inhibitor and serine-type endopeptidase inhibitor activities. However, more proteins (n = 6) are found in the protein set with exclusive or up-regulated expression in PE plasma (Table S3) than in the set with absent or down-regulated expression in PE plasma (n = 3) (Table S5). The protein sets also have members that participate in immune inflammatory responses and

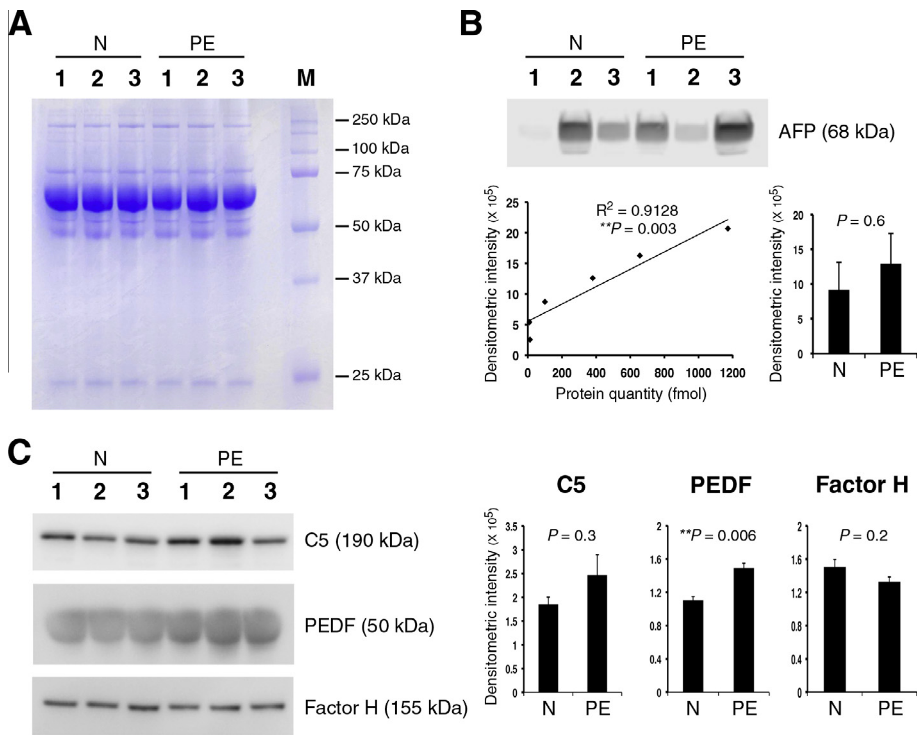

Figure 3 Validation of protein expression by Western blot analysis
Western blot analysis was performed to detect the presence of alpha-fetoprotein (AFP), complement C5 (C5), pigment epithelium-derived factor (PEDF), and complement factor H (factor H) in normotensive (N1, N2 and N3) and preeclamptic (PE1, PE2 and PE3) umbilical cord blood samples. **A**. A Coomassie blue-stained SDS–PAGE gel showed equal loading of 10 μg protein per lane of all samples tested. **B**. Western blot analysis for AFP showed a significant association between the densitometric intensities of the bands with the corresponding normalized protein quantity (R^2 = 0.91, P = 0.003 significant at $^{**}P < 0.01$, linear regression analysis) but a nonsignificant (P = 0.6, t-test) up-regulation of AFP in PE when compared to N samples. **C**. Western blot analysis and quantification densitometric analysis showed that expression of PEDF was significantly up-regulated in PE when compared to N samples (P = 0.006 significant at $^{**}P < 0.01$, t-test), whereas up-regulation for C5 expression and down-regulation for Factor H expression in PE plasma were not significant (P = 0.3 and 0.2, respectively, t-test). Error bars represent standard error of the mean (SEM).

complement activation (Tables S4 and S6). Consistent with the presence of these proteins in plasma, the majority of these proteins were clustered to the extracellular region or extracellular space in the cellular component analysis (Tables S7 and S8).

Validation of protein expression by Western blot analysis

To validate our findings, several proteins were selected for western blot analysis (**Figure 3**). For representative proteins with up-regulated expression in PE plasma, we chose 2 proteins with high fold change and Bayes factor values, *i.e.*, C5 and PEDF. We also examined the expression of AFP, which has been previously reported in PE plasma [9] and therefore would serve as a positive control in our analysis. Factor H was selected as the representative protein with up-regulated expression in N plasma. The loading of equal amounts (10 μg) of total protein for each of the plasma samples analyzed was confirmed by Coomassie Blue-stained polyacrylamide gel visualization (Figure 3A) before the transfer to polyvinylidene difluoride (PVDF) membranes. First, the quantification feature of the NanoUPLC-MS^E method was validated as the densitometric intensities for the bands of each sample immuno-positive for the AFP antibody corresponded significantly with the normalized fmol quantities determined from the bioinformatic analysis (Figure 3B; R^2 = 0.91, P = 0.003, linear regression analysis). Second, although there were substantial variations within each group, which is not uncommon in human plasma samples, the mean densitometric intensities of the bands for AFP and C5 were higher, but not significantly (P = 0.6 and 0.3, respectively, t-test), for PE samples when compared to N samples (Figure 3B and C). However, the mean band intensity for PEDF was significantly (P = 0.006, t-test) higher in PE than N samples (Figure 3C). Conversely, the mean densitometric band intensity for Factor H was higher, but not significantly (P = 0.2, t-test), in N plasma when compared to that of PE plasma (Figure 3C). These initial observations of protein expression are in accordance with the findings of the proteomic screen although confirmation of more protein expression using more samples is needed.

Discussion

To explore the mechanisms by which a PE intrauterine environment might confer protection against the subsequent development of breast cancer in female offspring [2,3,16], we employed the quantitative proteomic technology of NanoUPLC-MS^E, a unique method to compare the relative abundance of protein between samples based on their absolute quantity from LC-MS data of tryptic peptides. In this method, the average MS signal response for the three most intense tryptic peptides per mole of protein has been found to be constant within a coefficient of variation of less than $\pm 10\%$ [15]. Hence, given an internal standard, the absolute quantity of each protein could be determined in a complex mixture without the cumbersome generation of a calibration-response curve for specific polypeptides using numerous external reference peptides or in the use of radiolabeled amino acids.

Notably, the screen detected the previously reported AFP in the umbilical cord blood plasma of PE pregnancies [9]. In epidemiological studies, elevated levels of AFP, a glycoprotein produced by the fetal liver and yolk sac with known anti-estrogenic properties, have been independently associated with a reduced breast cancer risk [17] as well as indirectly associated with a reduction in breast cancer based on its association with low birth weight [18], ethnicity [19], multiple births [20] and hypertensive disorders of pregnancy [9,21]. In addition to AFP, the set of proteins identified to be up-regulated in PE samples can also be linked to anti-cancer effects. For example, PEDF or SERPINF1, a 50 kDa glycoprotein that was first discovered as a factor secreted by retinal pigment epithelial cells with neuronal differentiation and neurotrophic capabilities [22], has been reported to have additional roles in neuroprotection, anti-angiogenesis and anti-tumorigenesis [23]. Indeed, both AFP [14] and PEDF [24] have been explored for their roles as anti-cancer agents. To our knowledge, this is the first report of a significant up-regulation of PEDF in PE plasma.

Besides playing a role in the pathological events of preeclampsia [25,26], the identification of components of the complement system from PE umbilical cord plasma also points to a potential role of an anti-tumor immune response. It is conceivable that complement activation might be able to eliminate cancer cells and thus contribute to a lower cancer risk. For example, SAP, a member of the pentraxin family that is found exclusively in our PE plasma samples, may function to clear such defective cells via complement activation [27]. Additionally, cancer cells may evade the immune surveillance of complement-mediated lysis by expressing membrane-bound and soluble complement inhibitors, including Factor H [28]. This is consistent with our hypothesis that proteins that are down-regulated in PE plasma, such as Factor H, are potential molecules of increased cancer risk. Hence, cancer-specific modulation of an anti-tumor immune response by the complement system either via the blockage of complement inhibitors or inhibition of complement activation itself is promising [29,30].

The significance of identifying a set of keratins with exclusive or up-regulated expression in PE plasma is unclear. Keratin (or cytokeratin) expression is used to mark the epithelial cell lineages in mammary development [31] and for the evaluation of breast cancer subtypes [32]. While keratin can be a contaminant introduced by sample handling, expression of keratin 10 has been reported to be up-regulated in lymph node metastasis of liver [33]. Arguably, unlike secretory proteins such as AFP, PEDF and components of the complement system, the presence of cytoskeletal proteins in the plasma might indicate that they are no longer associated with the cells, thus conferring protection.

Many of the potential proteins of low- and high-risk breast cancer from our screen display endopeptidase inhibitor activities. Although the role of such activities in cancer is not fully understood, it has been proposed that endopeptidase activity is important in invasion where the action of proteinases is required for tumor cells to penetrate the extracellular matrix and the basement membrane [34]. Thus, it follows that the identification of many proteins with peptidase inhibitor activity from PE plasma is consistent with the model of decreased cancer risk. However, counter-intuitively, the identified serine-type endopeptidase inhibitor alpha-1-antitrypsin/SERPINA1 has been associated with the development potential and poor prognosis of gastric cancer [35], colorectal cancer [36], lung cancer [37] and insulinomas [38]. It has been suggested that alpha-1-antitrypsin may function to modulate host-immunodefence mechanisms in favor of tumor cells and promote blood circulation within tumor tissues for tumor development [36,37]. Hence, the down-regulation of alpha-1-antitrypsin expression in PE plasma is again consistent with a model of decreased cancer risk by preeclampsia.

Our screen identified 13 potential high-risk breast cancer proteins whose expression is decreased in PE or up-regulated in N pregnancies (Tables 2 and 4). Previously reported high-risk proteins include insulin-like growth factor-1 (IGF-1), 1,25-dihydroxyvitamin D, insulin-like growth factor binding protein-3 (IGFBP-3), ghrelin, high-density lipoprotein (HDL)-cholesterol and Apo A-1 [10–12]. Hence, other than Apo A-I, which was also picked up by our screen, we report here 12 potential novel proteins that are postulated to be associated with a higher prenatal risk of breast cancer due to fetal programming. As an example, the expression of lumican, an abundant small leucine-rich proteoglycan in breast stroma, was significantly correlated with mammographic density, an important risk factor for breast cancer [39]. Future research is needed to determine the relevance of these proteins in relation to breast cancer risk and novel therapies can be designed to silence such potential high-risk molecules.

Current cancer research has focused mainly on the analysis of pathological samples from patients whereby proteins identified would be limited to those produced after disease formation. Consistent with a fetal-origin-of-disease hypothesis, we look for prenatal factors in the intrauterine environment using umbilical cord blood plasma samples. Although our findings support research for the analysis of the prenatal environment using umbilical cord blood to identify factors that mediate a protective effect toward disease outcomes, such as that of cancer, it must be pointed out that not all biomarkers associated with preeclampsia play a role in reducing breast cancer risk and it is unclear how these molecules, if they do play such a role, mediate their effects from the *in utero* environment into adult life. Speculatively, such molecules could have brought about genetic and/or epigenetic changes to the stem cells of the developing embryo and these 'altered' cells become more resistant or susceptible to disease onset in later life [40].

In summary, this pilot study identified candidate proteins in umbilical cord blood plasma that might play a role in the prevention of breast cancer later in life. The validity of our screen for prenatal factors is supported by the identification of several

PE proteins that have anti-estrogenic, anti-angiogenic and anti-tumorigenic activities, properties that are consistent with a lowered risk of a hormone-responsive cancer such as breast cancer. However, given the small sample size, conclusions are preliminary. Upon validation with a larger sample size and testing in animal models, such preeclampsia-associated proteins can serve as biomarkers for identifying individuals with different susceptibilities to breast cancer, give us insight into the potential prenatal mechanisms by which preeclampsia mediates its protective effect against breast cancer in female offspring, and lead to potential applications in cancer surveillance and prevention.

Materials and methods

Experimental workflow

The experimental workflow is summarized in Figure 1. Briefly, subjects with PE and N pregnancies were recruited and, at the time of birth, their umbilical cord blood was collected, processed and the plasma stored at −80 °C prior to analysis. The plasma was immuno-depleted of 14 most abundant proteins with the Seppro IgY14 column and processed for analysis by the proteomic technology of NanoUPLC-MS^E. Selected differentially expressed proteins were validated by western blot analysis.

Subject recruitment

The study protocol was approved by the institutional review boards of the University of Massachusetts Medical School and Tufts Medical Center, and informed consent was obtained from all participating subjects. Study subjects were recruited among pregnant women who delivered at Tufts Medical Center, Boston, MA. All subjects were 18 years or older with a singleton pregnancy, HIV- and hepatitis B-negative, and the fetus was free of anomalies by ultrasound examination. Preeclampsia and its severity were diagnosed using standard clinical criteria [7]. Briefly, preeclampsia was diagnosed by the presence of a persistently elevated systolic blood pressure (BP) $\geqslant$140 mmHg and/or diastolic BP $\geqslant$90 mmHg after 20 weeks of gestation along with a 24-h urinary protein output of $\geqslant$300 mg or $\geqslant$1+ proteinuria on dipstick.

Umbilical cord blood collection and processing

Infants were delivered according to standard obstetrical practices. Umbilical cord blood was collected from the umbilical vein using a blood collection bag containing 35 ml of citrate–phosphate–dextrose anticoagulant (Fenwal, Lake Zurich, IL). Samples were centrifuged at 20 °C for 30 min at 400*g* within 24 h of collection. The top plasma layer was harvested into 2-ml cryovials and stored at −80 °C prior to use. In this pilot study, umbilical cord plasma samples were collected from three PE pregnancies (PE1, PE2 and PE3) and equal number of N controls (N1, N2 and N3) for proteomic analysis.

Plasma protein processing

Samples were depleted of abundant proteins using the Seppro IgY14 Spin Column Kit (Sigma–Aldrich, St. Louis, MO) according to manufacturer's instructions. After depletion, the samples were concentrated by filtration on a 10-kDa cutoff Amicon Ultra Centrifugal Filter Unit (Millipore, Billerica, MA) until less than 200 µl of concentrated solution remained. Protein concentration was determined by the Bradford method. The concentrated samples were prepared in 50 mM ammonium bicarbnonate with 0.1% RapiGest. The proteins were then reduced in 5 mM dithiothreitol at 60 °C for 30 min and alkylated in 15 mM iodoacetamide by incubation in the dark at room temperature for 30 min. A tryptic digestion was performed at 30 °C overnight with 1.5 µg of trypsin (Promega, Madison, WI). The RapiGest was acid cleaved by the addition of trifluoroacetic acid (TFA) to a final concentration of 0.5% at 37 °C for 45 min. The particulates were centrifuged down and the peptides in the supernatant were passed through a 0.22 µm Amicon Ultrafree-MC spin-filter (Millipore) before analysis by NanoUPLC-MS^E.

NanoUPLC-MS^E analysis

Samples were individually analyzed by the non-data dependent absolute protein quantification method described by Silva and colleagues [15]. Before each injection, the protein concentration of the sample was adjusted to 500 ng/µl with 0.1% TFA. A tryptic digest of yeast alcohol dehydrogenase (ADH; 200 fmol/µl; Waters, Milford, MA) was added to the digests as an internal standard for quantification. Nanoflow separations of the tryptic peptides were performed on a NanoUPLC system (Waters). The column temperature was maintained at 35 °C. Partial loop sample injections (2 µl) were performed in a random order with three analyses of each sample. The UPLC system was equilibrated with 5% acetonitrile (ACN)/0.1% formic acid (FA) (mobile phase A). Separation of injected tryptic peptides was achieved by application of a linear 60-min gradient from 3% to 90% ACN in 0.1% FA which was passed through both the trapping (SymmetryC18, 180 µm × 20 mm; Waters) and analytical columns. A lock mass solution (200 fmol/µl of Glu-fibrinopeptide; Sigma) in 30% ACN/0.1% FA in water was delivered via the auxiliary solvent pump at 300 nl/min into the reference sprayer of the NanoLockSpray source. Alternating high and low collision energy mass spectra of NanoUPLC eluates were acquired using a Q-TOF Premier mass spectrometer (Waters). Nanospray ionization was performed using uncoated, pulled fused silica emitters (New Objective, Woburn, MA) at a potential of 3.5 kV. The time-of-flight (TOF) analyzer was operated in the V mode with a typical mass resolution of 10,000. Alternating low and high collision energy scans were used to obtain low collision energy (4 eV) MS spectra and programmed high collision energy (ramped from 15 to 40 eV) MS^E mixed product ion spectra, both acquired at 0.6 s/scan. The instrument was calibrated using 13 fragment ions from a high energy scan of the $[M+2H]^{2+}$ ion from Glu-fibrinopeptide. To correct any shifts in masses that occurred during analyses, a single-point lock mass correction was performed at 30 s intervals using the $[M+2H]^{2+}$ ion from Glu-fibrinopeptide (m/z 785.8426). The raw files acquired were processed and searched against the database of human proteins generated from the UniProtKB/Swiss-Prot protein sequence database downloaded on September 9, 2009 (www.expasy.org), using the ProteinLynx Global Server (PLGS) Identity^E version 2.4 software (Waters).

Bioinformatic and data analysis

Scaffold

Scaffold (version 3.00.03, Proteome Software, Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at >95% probability as specified by the Peptide Prophet algorithm [41]. Protein identifications were accepted if they could be established at >99% probability as assigned by the Protein Prophet algorithm [42] and contained at least 3 identified peptides. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

Fold change determination

Relative abundance of each protein was assessed by first converting intensity values to fmol amounts by normalization to 200 fmol ADH internal control added in each run. As the average MS signal of the three most intense peptides from each protein is proportional to the abundance of that protein [15], the total amount in fmol of all the proteins identified for each run was determined and the fmol amounts of each protein was then normalized so that the sum of each run was the same for all the runs. Mean values for all the runs of each sample and the mean value for each of the two groups of samples were calculated. Fold changes were determined by comparing the means of the two groups. A value >2.0 or >1.5 (based on the upper bound of the 95% confidence interval) was used as our significance threshold for proteins whose expression was up-regulated in PE and N pregnancies, respectively.

Spectral counting

Since spectral counts correlate with protein abundance [43], relative abundance of each protein was also assessed by normalization of spectral counts based on the total spectral counts of each run, as reported previously [44]. Normalized spectral counts were subjected to QSPEC analysis [45] to obtain Bayes factor and false discovery rate (FDR) values as additional filtering criteria.

Ontology and pathway analysis

Gene ontology analysis [46] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis [47] were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Database (http://david.abcc.ncifcrf.gov/) interface [48] for the proteins that were unique or significantly up-regulated in the two groups. For gene ontology analysis, only molecular functions, biological processes and cellular components with $P < 0.05$ (as specified by DAVID) were considered.

Western blot analysis

Plasma samples (10 μg total protein/well) were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE). A non-reducing SDS–PAGE was run for C5 detection [49]. Proteins were transferred to PVDF membranes and blocked with 5% skim milk powder (for C5, Factor H and PEDF) or 5% bovine serum albumin (for AFP) in tris-buffered saline (TBS) with 0.05% Tween 20 (TBS-T) for 2 h at room temperature. After washing with TBS-T, the appropriate blot was incubated overnight at 4 °C with one of the following antibodies: anti-human Factor H (goat, 1:2000; AF4779, R&D Systems, Minneapolis, MN), anti-human AFP antibodies (chicken, 1:2000; AF1369, R&D Systems), mouse anti-human C5 (mouse, 1:500; sc-70476, Santa Cruz Biotechnology, Santa Cruz, CA) or anti-human PEDF (mouse, 1:500; sc-53921, Santa Cruz Biotechnology). After washing at room temperature, the blots were incubated for 2 h at room temperature with 1:5000 to 1:10,000 of an appropriate secondary antibody conjugated to horseradish peroxidase. The protein of interest was detected using the ECL Western Blotting Detection System (GE Healthcare Bio-Sciences/Amersham Biosciences, Piscataway, NJ) in a LAS-4000 Luminescent Image Analyzer (Fujifilm Life Sciences, Valhalla, NY). Densitometry was performed using the image analysis software Multi Gauge V3.0 (Fujifilm Life Sciences). Statistical significance was determined by *t*-test and linear regression analyses at a two-sided $P < 0.05$.

Authors' contributions

HPL, ML and CCH participated in conception and design of the study, data analysis and interpretation, and manuscript writing. AT and JJ participated in the western blotting analysis and manuscript writing. LQ and CIC participated in sample processing and collection of data. WCS and ERN participated in subject recruitment, data analysis and manuscript writing. SWT, JEE, KMG and JAP participated in proteomic analysis, assembly of data, and manuscript writing. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Acknowledgements

The authors acknowledge the late Dr. Todd Savarese for his contribution during the early stages of the study, and Gerald M. Seixas, Jr. and the staff at Proteome Software Inc. for assistance with bioinformatics analyses. This work was supported by a grant from the US National Institutes of Health (Grant No. R01CA090902).

Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.gpb.2013.09.009.

References

[1] Potischman N, Troisi R. In-utero and early life exposures in relation to risk of breast cancer. Cancer Causes Control 1999;10:561–73.

[2] Ekbom A, Hsieh CC, Lipworth L, Adami HO, Trichopoulos D. Intrauterine environment and breast cancer risk in women: a population-based study. J Natl Cancer Inst 1997;89:71–6.

[3] Xue F, Michels KB. Intrauterine factors and risk of breast cancer: a systematic review and meta-analysis of current evidence. Lancet Oncol 2007;8:1088–100.
[4] Vogel VG, Costantino JP, Wickerham DL, Cronin WM, Cecchini RS, Atkins JN, et al. Effects of tamoxifen vs raloxifene on the risk of developing invasive breast cancer and other disease outcomes: the NSABP Study of Tamoxifen and Raloxifene (STAR) P-2 trial. JAMA 2006;295:2727–41.
[5] Lee I, Oguma Y. Physical activity. In: Schottenfeld D, Fraumeni JF, editors. Cancer epidemiology and prevention. New York: Oxford University Press; 2006. p. 449–67.
[6] Walker JJ. Pre-eclampsia. Lancet 2000;356:1260–5.
[7] ACOG Committee on Obstetric Practice. ACOG practice bulletin. Diagnosis and management of preeclampsia and eclampsia. Number 33, January 2002. Int J Gynecol Obstet 2002;77:67–75 and Obstet Gynecol 2002;99:159–67.
[8] Redman CW, Sargent IL. Pre-eclampsia, the placenta and the maternal systemic inflammatory response – a review. Placenta 2003;24:S21–7.
[9] Vatten LJ, Romundstad PR, Odegård RA, Nilsen ST, Trichopoulos D, Austgulen R. Alpha-foetoprotein in umbilical cord in relation to severe pre-eclampsia, birth weight and future breast cancer risk. Br J Cancer 2002;86:728–31.
[10] Díaz E, Halhali A, Luna C, Díaz L, Avila E, Larrea F. Newborn birth weight correlates with placental zinc, umbilical insulin-like growth factor I, and leptin levels in preeclampsia. Arch Med Res 2002;33:40–7.
[11] Ayidin S, Guzel SP, Kumru S, Aydin S, Akin O, Kavak E, et al. Serum leptin and ghrelin concentrations of maternal serum, arterial and venous cord blood in healthy and preeclampsia pregnant women. J Physiol Biochem 2008;64:51–9.
[12] Catarino C, Rebelo I, Belo L, Rocha-Pereira P, Rocha S, Castro EB, et al. Fetal lipoprotein changes in pre-eclampsia. Acta Obstet Gynecol Scand 2008;87:628–34.
[13] Napolitano PG, Wakefield CL, Elliot DE, Doherty DA, Magann EF. Umbilical cord plasma homocysteine concentrations at delivery in pregnancies complicated by pre-eclampsia. Aust N Z J Obstet Gynaecol 2008;48:261–5.
[14] Mizejewski GJ. The alpha-fetoprotein-derived growth inhibitory peptide 8-mer fragment: review of a novel anticancer agent. Cancer Biother Radiopharm 2007;22:73–98.
[15] Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ. Absolute quantification of proteins by $LCMS^E$: a virtue of parallel MS acquisition. Mol Cell Proteomics 2006;5:144–56.
[16] Trichopoulos D. Hypothesis: does breast cancer originate in utero? Lancet 1990;335:939–40.
[17] Richardson BE, Peck JD, Wormuth JK. Mean arterial pressure, pregnancy-induced hypertension, and preeclampsia: evaluation as independent risk factors and as surrogates for high maternal serum alpha-fetoprotein in estimating breast cancer risk. Cancer Epidemiol Biomarkers Prev 2000;9:1349–55.
[18] Nagata C, Iwasa S, Shiraki M, Shimizu H. Estrogen and alpha-fetoprotein levels in maternal and umbilical cord blood samples in relation to birth weight. Cancer Epidemiol Biomarkers Prev 2006;15:1469–72.
[19] Lambe M, Trichopoulos D, Hsieh CC, Wuu J, Adami HO, Wide L. Ethnic differences in breast cancer risk: a possible role for pregnancy levels of alpha-fetoprotein? Epidemiology 2003;14:85–9.
[20] Lambe M, Hsieh C, Tsaih S, Ekbom A, Adami HO, Trichopoulos D. Maternal risk of breast cancer following multiple births: a nationwide study in Sweden. Cancer Causes Control 1996;7:533–8.
[21] Thompson WD, Jacobson HI, Negrini B, Janerich DT. Hypertension, pregnancy, and risk of breast cancer. J Natl Cancer Inst 1989;81:1571–4.
[22] Tombran-Tink J, Chader GG, Johnson LV. PEDF: a pigment epithelium-derived factor with potent neuronal differentiative activity. Exp Eye Res 1991;53:411–4.
[23] Ek ET, Dass CR, Choong PF. PEDF: a potential molecular therapeutic target with multiple anti-cancer activities. Trends Mol Med 2006;12:497–502.
[24] Dass CR, Ek ET, Choong PF. PEDF as an emerging therapeutic candidate for osteosarcoma. Curr Cancer Drug Targets 2008;8:683–90.
[25] Feinberg BB. Preeclampsia: the death of Goliath. Am J Reprod Immunol 2006;55:84–98.
[26] Lynch AM, Salmon JE. Dysregulated complement activation as a common pathway of injury in preeclampsia and other pregnancy complications. Placenta 2010;31:561–7.
[27] Du Clos TW, Mold C. Pentraxins (CRP, SAP) in the process of complement activation and clearance of apoptotic bodies through Fc receptors. Curr Opin Organ Transplant 2011;16:15–20.
[28] Ajona D, Castaño Z, Garayoa M, Zudaire E, Pajares MJ, Martinez A, et al. Expression of complement factor H by lung cancer cells: effects on the activation of the alternative pathway of complement. Cancer Res 2004;64:6310–8.
[29] Sivasankar B, Longhi MP, Gallagher KME, Betts GJ, Morgan BP, Godkin AJ, et al. CD59 blockage enhances antigen-specific $CD4^+$ T cell responses in humans: a new target for cancer immunotherapy? J Immunol 2009;182:5203–7.
[30] Markiewski MM, DeAngelis RA, Benencia F, Ricklin-Lichtsteiner SK, Koutoulaki A, Gerard C, et al. Modulation of the antitumor immune response by complement. Nat Immunol 2008;9:1225–35.
[31] Gusterson BA, Stein T. Human breast development. Semin Cell Dev Biol 2012;23:567–73.
[32] Shao MM, Chan SK, Yu AMC, Lam CCF, Tsang JYS, Lui PCW, et al. Keratin expression in breast cancers. Virchows Arch 2012;461:313–22.
[33] Zong J, Guo C, Liu S, Sun MZ, Tang J. Proteomic research progress in lymphatic metastases of cancers. Clin Transl Oncol 2012;14:21–30.
[34] Siewinski M, Gutowicz J, Zarzycki A, Mikulewicz W. Role of cysteine endopeptidases in cancerogenesis. Cancer Biother Radiopharm 1996;11:169–76.
[35] Tahara E, Ito H, Taniyama K, Yokozaki H, Hata J. $Alpha_1$-antitrypsin, $alpha_1$-antichymotrypsin, and $alpha_2$-macroglobulin in human gastric carcinomas: a retrospective immunohistochemical study. Hum Pathol 1984;15:957–64.
[36] Karashima S, Kataoka H, Itoh H, Maruyama R, Koono M. Prognostic significance of alpha-1-antitrypsin in early stage of colorectal carcinomas. Int J Cancer 1990;45:244–50.
[37] Higashiyama M, Doi O, Kodama K, Yokouchi H, Tateishi R. An evaluation of the prognostic significance of alpha-1-antitrypsin expression in adenocarcinomas of the lung: an immunohistochemical analysis. Br J Cancer 1992;65:300–2.
[38] de Sá SV, Corrêa-Giannella ML, Machado MC, Krogh K, de Almeida MQ, Pereira MAA, et al. Serpin peptidase inhibitor clade A member 1 as a potential marker for malignancy in insulinomas. Clin Cancer Res 2007;13:5322–30.
[39] Alowami S, Troup S, Al-Haddad S, Kirkpatrick I, Watson PH. Mammographic density is related to stroma and stromal proteoglycan expression. Breast Cancer Res 2003;5:R129–35.
[40] Savarese TM, Low HP, Baik I, Strohsnitter WC, Hsieh CC. Normal breast stem cells, malignant breast stem cells, and the perinatal origin of breast cancer. Stem Cell Rev 2006;2:103–10.
[41] Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 2002;74:5383–92.
[42] Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 2003;75:4646–58.
[43] Liu H, Sadygov RG, Yates 3rd JR. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem 2004;76:4193–201.

[44] Paulo JA, Kadiyala V, Lee LS, Banks PA, Conwell DL, Steen H. Proteomic analysis (GeLC–MS/MS) of ePFT-collected pancreatic fluid in chronic pancreatitis. J Proteome Res 2012;11:1897–912.

[45] Choi H, Fermin D, Nesvizhskii AI. Significance analysis of spectral count data in label-free shotgun proteomics. Mol Cell Proteomics 2008;7:2373–85.

[46] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9.

[47] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acid Res 2000;28:27–30.

[48] Dennis Jr G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 2003;4:P3.

[49] Tiwari A, Liba A, Sohn SH, Seetharaman SV, Bilsel O, Matthew CR, et al. Metal deficiency increases aberrant hydrophobicity of mutant superoxide dismutases that cause amyotrophic lateral sclerosis. J Biol Chem 2009;284:27746–58.