High-Throughput Discovery and Characterization of Fetal Protein Trafficking in the Blood of Pregnant Women

Jill L. Maron, M.D, M.P.H. 1,* Gil Alterovitz, Ph.D. 2,* Marco Ramoni, Ph.D. 2 Kirby L. Johnson, Ph.D 3, and Diana W. Bianchi, M.D. 1,3

1Floating Hospital for Children at Tufts Medical Center, Department of Pediatrics, Division of Newborn Medicine 800 Washington Street, Box 44 Boston, MA 02111
2Children’s Hospital Informatics Program, Harvard-MIT Division of Health Sciences and Technology, Harvard Medical School Boston, MA 02115
3Floating Hospital for Children at Tufts Medical Center, Department of Pediatrics, Division of Genetics, 800 Washington Street, Boston, MA 02111

Abstract

Although the measurement of fetal proteins in maternal serum is part of standard prenatal screening for aneuploidy and neural tube defects, attempts to better understand the extent of feto-maternal protein trafficking and its clinical and biological significance have been hindered by the presence of abundant maternal proteins. The objective of this study was to circumvent maternal protein interference by using a computational predictive approach for the development of a noninvasive, comprehensive, protein network analysis of the developing fetus in maternal whole blood. From a set of 157 previously identified fetal gene transcripts, 46 were classified into known protein networks, and 22 downstream proteins were predicted. Statistically significantly over-represented pathways were diverse and included T-cell biology, neurodevelopment and cancer biology. Western blot analyses validated the computational predictive model and confirmed the presence of specific downstream fetal proteins in the whole blood of pregnant women and their newborns, with absence or reduced detection of the protein in the maternal postpartum samples. This work demonstrates that extensive feto-maternal protein trafficking occurs during pregnancy, and can be predicted and verified to develop novel noninvasive biomarkers. This study raises important questions regarding the biological effects of fetal proteins on the pregnant woman.

Keywords

Fetal Proteins; Feto-Maternal Trafficking Network Analysis; Prenatal Diagnosis

1. Introduction

Measurement of fetal proteins in maternal serum is part of standard prenatal screening for fetal aneuploidy and neural tube defects [1,2]. These markers, however, provide limited insight into the extent of feto-maternal protein trafficking and its biological and clinical significance. Attempts to conduct fetal proteomic analyses on maternal serum samples are hindered by
abundant maternal proteins that interfere with the detection of rare fetal proteins [3]. Additionally, access to proteins present in fetal whole blood in order to perform a systematic, comparative analysis between a pregnant woman and her fetus, is only possible in rare clinical scenarios [4]. In contrast, discarded amniotic fluid samples are more readily available and are purely fetal in origin. The study of fetal proteomics, therefore, has focused on 2-D gel electrophoresis and mass spectrometric analyses of mid-trimester normal amniotic fluid samples [5,6], amniocytes [7], and amniotic fluid obtained in the settings of preterm birth [8], preeclampsia [9], premature rupture of membranes (PROM) [10], intrauterine infection [11, 12] and aneuploidy [13,14,15]. In one comparative study, amniotic fluid and maternal plasma samples were obtained from the same woman at term [16].

We hypothesized that an accurate, comprehensive proteomic profile could be predicted from maternal whole blood using a protein interaction network. To do this, we used a list of 157 previously identified fetal gene transcripts [17]. After predicting the protein networks, and identifying the cellular locations and tissue expression profiles, the biological functions of each of the proteins were analyzed to better understand their origin, biology, and potential clinical application as a biomarker. To validate the predictive model, Western blot analyses were performed. The results show that extensive feto-maternal protein trafficking occurs during pregnancy and can be readily explored using a computational approach. The diverse nature of the fetal proteins identified raises important questions regarding the biological effects of these proteins on the pregnant woman.

2. Materials and Methods

Initial Gene Transcript List

This study was approved by the Tufts Medical Center Institutional Review Board. Briefly, previously identified fetal gene transcripts [17] were utilized to generate a predictive proteomic network. In the prior study, total RNA was extracted from the whole blood of nine women prior to and after delivery, and their newborns’ umbilical cord blood (n=10). Comparative microarray analyses were performed on all samples to identify gene transcripts that were present in the pregnant woman before she delivered and her own infant’s cord blood, but absent or significantly decreased in her postpartum sample. One hundred and fifty-seven gene transcripts were identified following strict statistical testing and adjustment for false discovery rates. Gene transcripts were confirmed by real time RT-PCR amplification and fetal specificity was confirmed by SNP analyses, as previously described [17].

Computational Analyses

To generate the proteomic network, we converted the initial fetal transcripts into the corresponding translated proteins. Next, we automatically integrated information from a number of sources: Database of Interacting Proteins (DIP) [18], IntAct [19], Molecular INTeraction (MINT) [20], Biomolecular Interaction Network Database (BIND) [21], cPath [22], the Sanger Institute Interaction Map [23], Kyoto Encyclopedia of Genes and Genomes (KEGG) [24], and the Human Protein Reference Database (HPRD) [25]. The different protein identification numbers were converted to Uniprot accessions and NCBI Entrez Protein GI numbers. This was done by sequentially querying SeqHound [26] via remote Java Application Protocol Interface and AliasServer [27] through Simple Object Access Protocol (SOAP). Also, the International Protein Index (IPI) cross-reference indexes, Ensembl cross reference indexes, and Entrez Protein database were queried to match the disparate ID numbers with appropriate NCBI GI numbers. Next, SeqHound was used to find redundant GI numbers. The best annotated version of the protein (from a group of database entries referring to the same protein sequence) was then used. With a common identifier, the databases were then merged and duplicates removed. We extracted the relationships encoded in the integrated protein network.
between these proteins and neighbors with the network and then mapped this information to generate the proteomic network subset used here.

Use of Publicly Available Sources to Identify Protein Location and Function

Fetal proteins within the networks were entered into Uniprot and Ingenuity® to describe their sub-cellular locations, tissue expression profiles, and biological functions. To remain objective, each sub-cellular location described in Ingenuity® for any given protein was entered for analysis. Only known tissue expression profiles from Ingenuity® were utilized in this analysis, and each tissue known to express any given protein was counted. A “ubiquitously expressed” protein was described as such within Ingenuity®. This limited analysis was done in order to generate a concise tissue expression profile of the fetal proteins found in maternal blood. The use of other databases such as Entrez Gene or Unigene would have resulted in conflicting results and would have ultimately introduced bias. Statistically over-represented biological functions were identified through Ingenuity®.

Western Blot Analyses

To confirm the predictive computational model, we prospectively enrolled three pregnant women at full-term presenting to the Tufts Medical Center Labor and Delivery Unit for a scheduled cesarean section. After obtaining informed consent, 10 cc of whole blood were obtained from the pregnant women prior to delivery, from their infants’ umbilical cord at time of delivery and from the post-partum women again >48 hours following delivery. All blood was placed directly into EDTA Vacutainer® blood collection tubes (Becton, Dickinson) and remained at room temperature until further processing. Whole blood was diluted 10-fold with Dulbecco’s phosphate buffer saline. This dilution was chosen based on the appearance of clear protein band separation on SDS-PAGE gel electrophoresis following serial blood dilutions from two to 50 fold. Blood was then dissolved in SDS-PAGE loading buffer, which included a 0.05% bromophenol blue stain (Sigma) and DTT was added. Samples were boiled for 5 minutes to denature the proteins and then frozen at −20°C pending electrophoresis. All samples underwent only one freeze-thaw cycle. No immunoprecipitation or targeted protein reduction of the samples occurred.

Electrophoresis was performed in a 7.5% Tris-HCl Precast Ready Gel (BioRad) at 150V for one sample set at a time. Each sample set was comprised of the maternal, umbilical cord, and postpartum samples from a mother-infant pair; 15 µL of prepared blood was loaded into each well. Gels were transferred onto nitrocellulose paper (Whatman) for antibody staining. Three down-stream predicted proteins, Plexin A1 (goat anti-human polyclonal IgG, Santa Cruz Biotechnology), adenosine deaminase (goat anti-human polyclonal IgG, Santa Cruz Biotechnology), and VAV1 (rabbit anti-human polyclonal IgG, Santa Cruz Biotechnology), were probed. Alpha-fetoprotein (mouse anti-human monoclonal IgG, R & D Systems) served as a positive control of a known fetal protein in maternal blood [28].

Proteins were selected for Western blot analysis following a comprehensive literature search for a known association between the protein and fetal development, as well as the availability of an antibody for Western blot analysis. Additionally, proteins were chosen if they were predicted by the computational model and not directly derived from the original fetal gene list. For each antibody, primary and secondary concentrations were 1:1000; blots were incubated for one hour at room temperature. To control for the initial loaded protein concentration in each well, the blots were subsequently stripped (Thermo Scientific) for 30 minutes at 37°C and were re-exposed to ensure dissolution of bands. Glyceraldehyde phosphate dehydrogenase (GAPDH) was utilized as a control for the presence of protein and to normalize initial protein concentration on each blot (mouse anti-human monoclonal antibody, Millipore). Primary and secondary antibody staining concentrations for GAPDH were 1:1000.
3. Results

The predicted fetal proteomic network contained 103,657 protein relationships. Of the initial 157 initial fetal RNA transcripts identified, 46 were classified into known protein networks, and 222 downstream predicted proteins were identified (Figures 1a and 1b, Supplemental Data Table 1). Statistically significantly over-represented pathways included B- and T-cell biology, hematopoiesis, glucose regulation, cancer biology, and neurodevelopment.

Characterization of Proteome

Sub-cellular Location—The sub-cellular locations of each of the proteins are shown in Figure 2. A small majority (24%) could be found in the nucleus, with 22% located in the cell membrane and 20% dispersed in the cytoplasm. Only a small percent (4%) were in the extracellular space.

Tissue Expression—The tissue expression profile of the proteins identified in this analysis are depicted in Figure 3. Proteins expressed in the liver, kidney, peripheral blood cells, heart, lung and brain are highly represented. The vast majority of all other organ tissues are represented in this analysis.

Network Construction for Biological Processes—All identified proteins were entered into Ingenuity® to assess their biological function. There were 60 biological function categories that were statistically over-represented by the fetal protein networks. The top 20 are depicted in Figure 4; the full list can be found in Supplemental Data Table 2. Cellular Growth and Proliferation was the most significant (p<10^{-14}), however Cellular Development, Tissue Development, Embryonic Development and the development of the Nervous System, Connective Tissue, Cardiovascular System, Skeletal and Muscular Systems, Lymphoid Tissue, and Hepatic System, were all statistically over represented within this data set. Interestingly, Cancer and Immune Cell Trafficking were also over expressed.

Western Blot Analyses

Western blot analyses detected three downstream predicted proteins as proof of principle. These proteins were Plexin A1 (Q9U1W2), adenosine deaminase (P00813), and VAV1 (P15498). Plexin A1 and VAV were both present in antepartum and umbilical samples, but not in the postpartum sample, suggesting clearance (Figures 5a,b). Uniform detection of GAPDH on the same blots for each sample served as both a control and normalization value for initial protein concentration. Adenosine deaminase was also present in the antepartum and umbilical cord blood samples, but did not completely disappear from the postpartum sample (Figure 5c). Our known control, alpha-fetoprotein, was detected in all blood samples (Figure 5d).

4. Discussion

Using a novel computational predictive model, rather than a direct assay such as 2-D gel electrophoresis or mass spectrometry, we have shown that extensive and diverse feto-maternal protein trafficking occurs during pregnancy and can be readily detected noninvasively in maternal whole blood. With an innovative and validated computational approach, we have circumvented a major limitation to fetal proteomic analysis of maternal blood, namely the abundance of maternal proteins interfering with the detection of fetal proteins. This work thus opens a rich area for further study with broad implications for understanding both the fundamental biology involved as well as medical applications.
By entering fetal gene transcripts previously identified in maternal whole blood into a computational predictive model, we were able to develop a comprehensive proteomic network of the term neonate. While our results are inherently linked to the genes identified, previously unrecognized downstream proteins have emerged, and their presence was validated by Western blot analyses. In our experimental model, we validated three proteins, plexin A1, adenosine deaminase, and VAV1, involved in different networks within our analyses. Plexin A1 is a member of a family of molecules first found to play an important role in axon guidance within the central nervous system [29]. As receptors to the semaphorin class of proteins, plexin A1 also plays a role in cardiac morphogenesis [30], and in the immune system [31]. Adenosine deaminase catalyzes the hydrolysis of adenosine to inosine and deficiency of this enzyme causes a form of severe combined immunodeficiency disease (SCID) [32]. The proto-oncogene, VAV1, is a contributor to hematopoiesis and is actively involved in T-cell and B-cell development and activation [33]. Thus, the results of our protein analysis shows upregulation of pathways involved in the development of the nervous and cardiovascular systems, as well as immune system pathways.

While we expected to detect the predicted proteins in the pregnant mother and her infant, we were surprised to find a complete clearance of plexin A1 and VAV1 in the post-partum samples. We speculated that proteins may not be as rapidly cleared from the maternal circulation, given that maternal immunoglobulin G (IgG) protein is detectable in the fetus and exhibits its effects for days to weeks [34]. Additionally in our prior work, a fetal gene in the maternal circulation was considered to be significant if it was present in both the antepartum and umbilical cord blood samples, but absent or significantly decreased in the postpartum sample, and made rigorous statistically significant criteria. Therefore, the reduction, but not complete clearance of adenosine deaminase in the postpartum samples in this study was expected. It is probable that additional studies would reveal varying clearance profiles of fetal proteins from the maternal circulation.

An in depth analysis of the proteins identified revealed that the majority of the predicted proteins could be found within the cell. Only 4% of all proteins within the network were localized extracellularly. In contrast, 42% of the proteins described in the proteomic analysis of amniotic fluid supernatant could be found in the extracellular space [4]. This is important, as it suggests that the cellular nature of whole blood is undoubtedly contributing to our detection of predominantly cellular proteins. It is unclear at this time whether the intracellular proteins detected are present in the maternal circulation as a result of ongoing fetal cell apoptosis or intact fetal cells that are lysed during the preparation protocol. Additional experiments are needed to clarify this issue.

The expression profiles show that the fetal proteins detected in the pregnant woman’s whole blood originate from a diverse group of tissues and organs from the developing fetus. Development proteomic networks dominate the functional characterization of the predicted proteins, illustrating the potential clinical application of this technology as a means to monitor normal and abnormal fetal development. Interestingly, other statistically over-represented biological functions identified included ‘immune cell trafficking’ and ‘cancer’. It is probable that up-regulation of immune cell trafficking within the fetus is necessary for immune tolerance between mother and fetus during gestation. Additionally, while it is understood that many proteins first identified in cancer biology are also seen in the developing embryo and fetus, our work reveals that there is direct trafficking of these proteins into the maternal circulation during pregnancy. These statistically significant biological functions highlight important physiological processes that occur in the pregnant woman and the fetus, and illustrate how the novel proteomic analytical approach described here may also contribute to the fields of immunology, oncology and prenatal diagnosis.
Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROM</td>
<td>Premature rupture of membranes</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde Phosphate Dehydrogenase</td>
</tr>
</tbody>
</table>

Acknowledgments

We would like to thank Dr. Guang Wang of the Molecular Cardiology Research Institute at Tufts Medical Center and Helene Stroh for their contribution to this work, as well as the physicians, nurses, and staff of the Tufts Medical Center Department of Obstetrics and Gynecology. This work was funded by NIH grants HD42053-06 to Dr. Diana Bianchi and K99 LM009826 to Dr. Gil Alterovitz.

References


Figure 1a.
Figure 1b.

**Figure 1.**

* a. A two-dimensional zoomed-in partial depiction of the proteomic network. Red circles represent the fetal proteins derived from the initial biomarkers previously described. Yellow circles represent downstream predicted proteins.

* b. A three-dimensional zoomed-in partial depiction of the proteomic network. Red circles represent the fetal proteins derived from the initial biomarkers previously described. Yellow circles represent downstream predicted proteins.
Figure 2.
Subcellular location of predicted proteins as assigned by Ingenuity®.
Figure 3.
Tissue expression profiles of the predicted proteins within the network.
Figure 4.
Statistically over-represented biological functions as assigned by Ingenuity®. The complete list of biological functions can be found in Supplemental Data 2.
Figure 5.
Western blot images showing the clearance of inferred fetal protein from maternal blood (i.e. presence of the protein prior to delivery and its absence or partial clearance following birth, coupled with the presence of the protein in umbilical cord blood). Upper panels: a) VAV1 protein, with a molecular weight of 95 kiloDaltons (kDa), b) Plexin A1 protein, with a molecular weight of 220 kDa, c) Adenosine Deaminase, with a molecular weight of 41 kDa, and d) Alpha-Fetoprotein, with a molecular weight of 69 kDa. GAPDH protein detection (lower panels) is included as a positive control. A: antepartum maternal blood sample; U: umbilical cord blood sample; P: postpartum maternal blood sample.