

# Proteomics Analysis of Human Amniotic Fluid\*<sup>§</sup>

Chan-Kyung J. Cho<sup>‡§</sup>, Shannon J. Shan<sup>‡§</sup>, Elizabeth J. Winsor<sup>§</sup>,  
and Eleftherios P. Diamandis<sup>‡§¶</sup>

Amniotic fluid is a dynamic and complex mixture that reflects the physiological status of the developing fetus. In this study, the human amniotic fluid (AF) proteome of a 16–18-week normal pregnancy was profiled and analyzed to investigate the composition and functions of this fluid. Due to the complexity of AF, we utilized three different fractionation strategies to provide greater coverage. Two types of two-dimensional LC/MS/MS as well as an LC-SDS-PAGE-LC-MS/MS platform were used. A total of 16 AF samples between gestational ages of 16 and 18 weeks from women carrying chromosomally normal fetuses were analyzed by one of the three fractionation methods followed by a common reverse phase LC-MS/MS step. Mascot and The Global Proteome Machine engines were used to search the International Protein Index human database for peptide sequence identification. The list of proteins was generated by combining the results of both engines through the PeptideProphet of Scaffold software. All identified proteins were combined to generate the AF proteome comprising 1,026 unique gene matches or 842 non-redundant proteins. This list includes most of the currently used biomarkers for pregnancy-associated pathologic conditions such as preterm delivery, intra-amniotic infection, and chromosomal anomalies of the fetus. The subcellular localization, tissue expression, functions, and networks of the AF proteome were analyzed by various bioinformatic tools. These data will contribute to the better understanding of amniotic fluid function and to the discovery of novel biomarkers for prenatal diagnosis of fetal abnormalities. *Molecular & Cellular Proteomics* 6:1406–1415, 2007.

Amniotic fluid (AF)<sup>1</sup> is a complex and dynamic biological fluid that provides mechanical protection and nutrients as well

From the <sup>‡</sup>Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5G 1L5, Canada, <sup>§</sup>Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario M5T 3L9, Canada, and <sup>¶</sup>Department of Clinical Biochemistry, University Health Network and Toronto Medical Laboratories, Toronto, Ontario M5G 2C4, Canada

Received, March 1, 2007, and in revised form, May 1, 2007

Published, MCP Papers in Press, May 10, 2007, DOI 10.1074/mcp.M700090-MCP200

<sup>1</sup> The abbreviations used are: AF, amniotic fluid; AFP,  $\alpha$ -fetoprotein; FPLC, fast protein liquid chromatography; Ig, immunoglobulin; IPI, International Protein Index; SAX, strong anion exchange; SCX, strong cation exchange; 2DE, two-dimensional electrophoresis; bis-Tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; hCG, human chorionic gonadotropin.

as other molecules required for fetal growth and well being. Therefore, both the quantitative and qualitative integrity of AF are essential for normal development of the human fetus during pregnancy. During embryogenesis, AF is initially formed from maternal plasma that passed through fetal membranes. Because free diffusion occurs bidirectionally between the AF and the fetus across fetal skin, placenta, and umbilical cord from 10 to 20 weeks of gestation, AF composition becomes similar to that of fetal plasma during this period. Therefore, analysis of AF composition before skin keratinization that occurs between 19 and 20 weeks of gestation would reveal valuable information that may indicate physiological or pathological conditions of the fetus.

AF contains water, proteins, peptides, carbohydrates, lipids, hormones, and electrolytes. Among these components, studies have been done on the amino acids in AF, such as taurine, glutamine, and arginine, and some trophic mediators such as epidermal growth factor, transforming growth factor  $\alpha$  and  $\beta$ -1, and insulin-like growth factor I (1). Additionally expression levels of several proteins and cytokines were studied using immunoassays. For example, the levels of biomarkers for the most common chromosomal anomaly, Down syndrome, such as  $\alpha$ -fetoprotein (AFP), were thoroughly investigated both in maternal serum and in AF. Nevertheless little is known about the functions of these proteins or the majority of constituents of AF.

The collective profile of AF proteins has not as yet been assembled. Thus, little is known regarding the possible molecular interactions of proteins and their contributions in fetal development. Therefore, it is desirable to identify more proteins to explore their functions within the AF.

Recent technological advances in proteomics have been actively utilized to investigate AF proteins to better understand this complex biological fluid and to discover disease-specific biomarkers. Consequently many putative markers for such anomalies as premature rupture of amnion, intra-amniotic infection, and Down syndrome have been reanalyzed or newly discovered (2–4). However, none of these biomarkers have a high enough individual detection rate to be used on their own. Thus, the discovery of more efficient biomarkers that have a higher detection rate and specificity is highly desirable. Proteomics analysis of AF, therefore, could be an ideal first step in efforts to elucidate changes related to pathological conditions in the fetus.

Proteomic profiles of AF have been generated by several

groups using different methods since 1997 (5–9). More specifically, Liberatori *et al.* (5) identified 31 proteins by two-dimensional electrophoresis (2DE) followed by postseparation analysis techniques such as N-terminal sequencing. Nilsson *et al.* (6) were the first to use MS for profiling AF. They used LC-Fourier transform-ion cyclotron resonance MS to identify 58 proteins from AF. They were also the first to deplete albumin from AF to identify more proteins. In 2006, three groups analyzed normal AF supernatant. Park *et al.* (7) reported 37 proteins by using 2DE followed by matrix-assisted laser desorption/ionization time-of-flight MS; Michel *et al.* (8) identified 69 proteins from albumin-depleted AF by Off-Gel<sup>TM</sup> electrophoresis/LC-MS/MS; and Tsangaris *et al.* (9) reported 136 proteins by 2DE followed by matrix-assisted laser desorption/ionization-MS/MS. These different groups not only utilized different approaches as well as different protein databases but also used different levels of stringency for protein identification, making it difficult to assess the accuracy of each data set. For example, the group who identified the highest number of proteins searched their queries against the collective protein database and consequently identified many non-human proteins from human AF (9).

In this study, our objective was to generate an extensive profile of the normal human AF proteome. Similar to other biological fluids, AF contains a few major proteins that complicate proteome profiling. Depletion of these proteins certainly improves protein identification. However, the “sponge effect” of these major proteins (and especially of albumin) could lead to the loss of many low abundance proteins that may be valuable biomarkers. Therefore, we designed three complementary platforms to profile the AF proteome to maximize the number of proteins identified. As a result, we report here the most extensive list of normal human amniotic fluid proteins comprising 1,026 unique gene matches from at least 754 (and probably up to 842) different genes. The large number of the AF proteins allowed qualitative bioinformatics analysis and may provide the basis for further compositional and functional studies of AF.

#### MATERIALS AND METHODS

**Sample Collection and Preparation**—Human AF samples (8–10 ml) were obtained by amniocentesis from women at 16–18 weeks of gestation undergoing prenatal diagnosis mostly due to advanced maternal age ranging from 30 to 45 years after written informed consent. AF samples were centrifuged to collect amniocytes for cytogenetic analysis, and the cell-free supernatants were stored at –80 °C until use. 16 samples from chromosomally normal pregnancies were chosen randomly. The samples were fractionated by one of three methods as follows. Three samples were fractionated by strong anion exchange (SAX) liquid chromatography, another three were fractionated by strong cation exchange (SCX) liquid chromatography, and 10 samples were pooled together and fractionated by LC-SDS-PAGE (see Fig. 1). Samples were thawed and filtered through a 0.22- $\mu\text{m}$   $\times$  25-mm syringe-driven filter unit (Millipore). The pooled sample only was also filtered through an Amicon 50-kDa-cutoff centrifugal ultrafiltration device (Millipore), and the filtrate was collected.

All samples were depleted for IgG with a Protein A/G column (Bio-Rad). Samples were then dialyzed at 4 °C using a membrane with 3.5-kDa molecular mass cutoff (Spectra/Por) as follows: for 12 h in 5 liters of 1 mM ammonium bicarbonate buffer with one buffer exchange in the same buffer for the SCX approach and in 5 liters of 20 mM Tris buffer, pH 9.6, for the SAX approach.

**SAX Liquid Chromatography**—For the SAX and the LC-SDS-PAGE approaches, the dialyzed AF samples were directly loaded onto an HR10/10 column packed with SOURCE15Q media for strong anionic exchange (GE Healthcare). Fractionation was performed using a fast performance liquid chromatography (FPLC) system using 50 mM Tris-HCl, pH 9.6, as the running buffer and 1 M NaCl as the elution buffer for 1 h using a linear gradient at a flow rate of 3 ml/min. A total of 12 fractions were collected, dialyzed in 5 liters of 1 mM ammonium bicarbonate buffer for 12 h, and lyophilized to dryness for trypsin digestion.

**Trypsin Digestion**—Each lyophilized sample was denatured using 8 M urea, reduced with 200 mM dithiothreitol at 50 °C, and alkylated in 500 mM iodoacetamide at room temperature in a dark room. The sample was desalted using a NAP5 column (GE Healthcare). The sample was then lyophilized and resuspended in trypsin buffer (1:50, trypsin:protein concentration; 120  $\mu\text{l}$  of 50 mM ammonium bicarbonate, 100  $\mu\text{l}$  of methanol, 150  $\mu\text{l}$  of H<sub>2</sub>O) overnight at 37 °C (Promega, sequencing grade modified porcine trypsin). The sample was lyophilized to dryness. For the SCX approach only, the dried peptides were resuspended in 120  $\mu\text{l}$  of mobile phase A (0.26 M formic acid in 10% ACN) to be injected into the HPLC system.

**SDS-PAGE and In-gel Digestion**—Briefly each of the 12 fractions was dissolved in SDS-PAGE loading buffer, DTT was added, and samples were heated at 80 °C for 5 min and loaded in a single lane on a 1-mm-thick 4–12% bis-Tris gel (Invitrogen). After separation, the gel was stained with SimplyBlue<sup>TM</sup> SafeStain (Invitrogen). The intensely stained bands below 50 kDa were excised separately, and all of the in-between areas were cut into slices, resulting in a total of 10–12 slices per fraction. In-gel digestion was performed as follows. Bands were washed in 200  $\mu\text{l}$  of 30% methanol for 5 min, and then 200  $\mu\text{l}$  of 100% ACN was applied for 10 min. The bands were incubated in 10 mM DTT in 100 mM NH<sub>4</sub>HCO<sub>3</sub> for 30 min at 60 °C. The DTT solution was removed, and the bands were incubated with 200  $\mu\text{l}$  of 50 mM iodoacetamide in 100 mM NH<sub>4</sub>HCO<sub>3</sub> for 30 min in the dark. The iodoacetamide was then removed, and washes were performed with 500  $\mu\text{l}$  of distilled water followed by addition of 100  $\mu\text{l}$  of ACN. Then ACN was removed, and 50  $\mu\text{l}$  of the 0.01  $\mu\text{g}/\mu\text{l}$  trypsin solution was added. Proteins were digested overnight at 37 °C.

**SCX Liquid Chromatography**—For this approach, the samples were first digested with trypsin to generate peptides. The digested peptides, resuspended in 120  $\mu\text{l}$  of mobile phase A, were directly loaded onto a PolySULFOETHYL A<sup>TM</sup> column (The Nest Group, Inc.) containing hydrophilic anionic polymer (poly(2-sulfoethyl aspartamide)) with a pore size of 200 Å and a diameter of 5  $\mu\text{m}$ . Fractionation was performed using an HPLC system (Agilent 1100) for 1 h at a flow rate of 200  $\mu\text{l}/\text{min}$ . 1 M ammonium formate and 0.26 M formic acid in 10% ACN (mobile phase B) were added in a linear gradient. A protein cation exchange standard (Bio-Rad) was applied before each run to evaluate column performance. The eluent was monitored by UV absorbance at 280 nm. A total of 12 or 24 fractions were collected and lyophilized to dryness.

**On-line Reverse Phase Liquid Chromatography-Tandem Mass Spectrometry**—Each fraction from all fractionation schemes was resuspended in 80  $\mu\text{l}$  of 95% water, 0.1% formic acid, 5% ACN, 0.02% trifluoroacetic acid (Buffer A) and desalted using a ZipTip C<sub>18</sub> pipette tip (Millipore). The peptides were eluted in 4  $\mu\text{l}$  of 90% ACN, 0.1% formic acid, 10% water, 0.02% trifluoroacetic acid (Buffer B), and 80  $\mu\text{l}$  of Buffer A was added on top. Half of this volume (40  $\mu\text{l}$ ) was

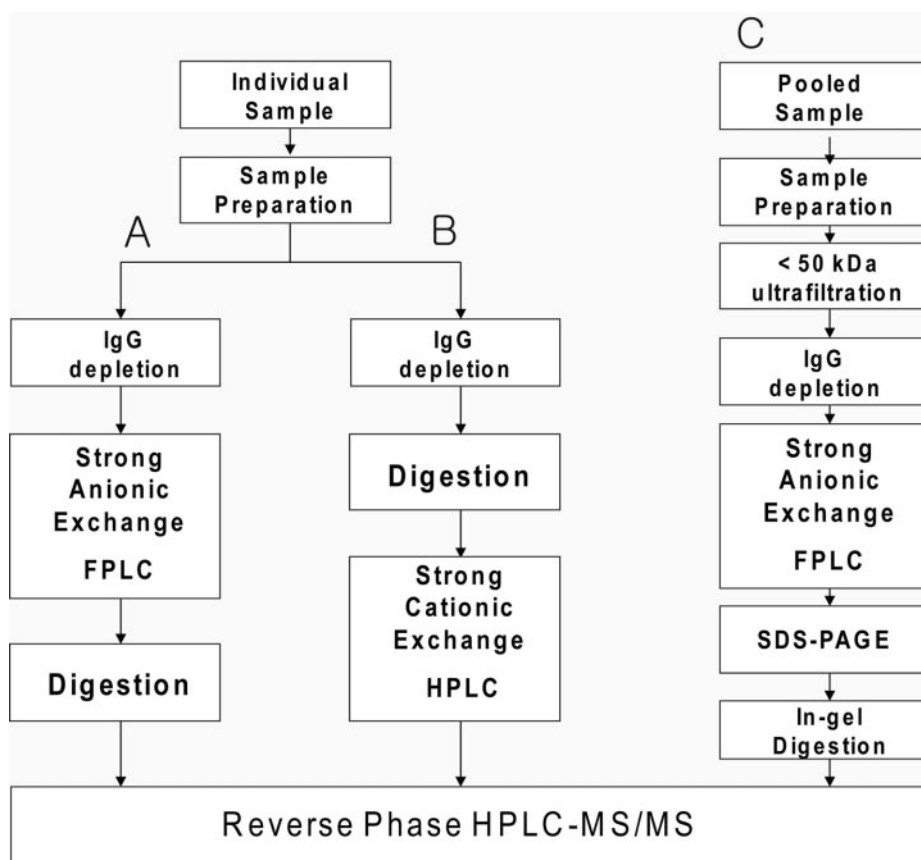


FIG. 1. The three sample preparation protocols used in this study. A, SAX-FPLC-LC-MS/MS. B, SCX-HPLC-LC-MS/MS. C, ultrafiltration with 50-kDa cutoff coupled to SAX-FPLC and followed by SDS-PAGE/LC-MS/MS.

loaded on an Agilent 1100 HPLC system by the autosampler onto a 2-cm  $C_{18}$  trap column (inner diameter, 200  $\mu\text{m}$ ), and the peptides were eluted onto a resolving 5-cm analytical  $C_{18}$  column (inner diameter, 75  $\mu\text{m}$ ). The 120-min gradient began at 0% Buffer B at 15  $\mu\text{l}/\text{min}$  for 5 min and then changed to 40% Buffer B for 103 min with a linear gradient, then to 65% Buffer B for 4 min, and finally to 100% Buffer B for 13 min. The peptides were subjected to nanoelectrospray ionization followed by MS/MS in an LTQ two-dimensional linear ion trap (Thermo Scientific) coupled on line to the HPLC system.

**Database Searching and Bioinformatics**—The resulting spectra from each fraction were searched separately against the IPI human database Version 3.16 by two database search engines: Mascot, Version 2.1.03 (Matrix Science) and The Global Proteome Machine Version 2.0.0.4 (Beavis Informatics Ltd.). The following parameters were used: (i) enzyme, trypsin; (ii) one missed cleavage allowed; (iii) fixed modification, carbamidomethylation of cysteines; (iv) variable modification, oxidation of methionine; (v) peptide tolerance, 3.0 Da; and (vi) MS/MS tolerance, 0.4 Da. The files were run on Scaffold Version 01\_05\_19 (Proteome Software). All DAT files, from Mascot, that were searched for all fractions from a single chromatography run (HPLC or FPLC) were loaded together as one “biological sample” within Scaffold, and all XML files, from The Global Proteome Machine, were loaded together as another biological sample. 95% peptide identification probability and 80% protein identification probability were used as the cutoffs for Scaffold, excluding proteins identified with lesser probability.

The sample reports were exported to Microsoft Excel from Scaffold, and relevant information and annotations for each protein were searched from databases including Swiss-Prot, Human Protein Reference Database, Entrez Gene, and the Plasma Proteome Database.

The false-positive error rate was calculated by analyzing all files with the same method except against a “sequence-reversed” IPI human database. The false-positive rate (FPR) was calculated as:  $\text{FPR} = \text{number of false peptides} / (\text{number of true peptides} + \text{number of false peptides})$ .

## RESULTS

### *Establishment of Complementary Methods for Proteomics Profiling*

Complex biological fluids such as serum or AF contain thousands of proteins ranging from high to low abundance. The major challenge for analyzing such proteomes is posed by high abundance proteins, especially albumin. In the case of serum, nearly 50% of the protein content is comprised by albumin. We verified by total protein and albumin analysis that albumin comprises nearly 70% of the protein content of AF with immunoglobulins (Igs) being the second most abundant fraction (data not shown). When Igs were removed by Protein A/G beads, the number of identified proteins increased by 38% (data not shown). Because removal of the bulk of Igs is straightforward, this step was included in all three fractionation protocols (Fig. 1). Albumin, unlike Igs, cannot be readily removed, and it also binds other proteins and peptides. Therefore, three different fractionation protocols were used, and different variations were tried within the context of each method (Fig. 1) to extend the depth of proteomics analysis.

The data are summarized below.

#### Number of Distinct Proteins Identified in AF

Three multidimensional separation methods coupled with MS/MS were used (Fig. 1). Raw files were searched by both Mascot and X!Tandem for more confident identification. To generate a statistically valid list of proteins, Scaffold was used to accommodate differences in algorithm and score calculation of the two search engines (10). The resulting lists of proteins from each run on Scaffold were combined into one collective list of proteins. Also all output files were searched with the reverse IPI human database, yielding the false-positive rate of ~3%.

To select only distinct proteins, we applied various elimination steps to generate the final non-redundant list. 1) We sorted the entries by IPI number and removed redundancies. 2) We removed all keratin entries (contaminants). 3) All Ig chains were removed because they are of lesser significance for proteome analysis, and they significantly increase redundancy and hinder statistical analysis of the proteome. Despite our effort to remove Igs by Protein A/G affinity chromatography, ~11% of all initial entries were attributed to Igs. This is not surprising because a similar trend has been observed with the plasma proteome (11). 4) Because one protein may have multiple IPI numbers, we sorted the proteins by their names and molecular weights and removed identical ones. After cleaning, the list included 1,026 proteins with distinct IPI number, name, and molecular weight (supplemental data: unique gene matches).

Different matches of one gene may reflect biologically significant different gene products such as splice variants, sequence variants, and cleavage products. In our method such a distinction is often difficult because protein prediction is based on peptide sequence searches. Therefore, a more stringent method was used to include only non-redundant proteins. We searched for gene names of each entry and removed all but one entry with the highest number of unique sequences. We acknowledge that this approach may remove some legitimate protein variants. Our final list included 842 proteins from 754 distinct genes and 88 proteins from uncharacterized genes (supplemental data: unique genes). Among the 842 unique proteins, 445 were identified with two or more unique peptides. Among 397 proteins that were identified by one unique peptide, at least 67 proteins were identified multiple times via more than one of the three methods. In total, at least 512 proteins were identified with high confidence, and the rest (330 proteins) were identified with at least 80% probability. Of the 512 high confidence proteins, 424 were identified for the first time in AF (supplemental data: unique genes).

#### Detection of Low Abundance Proteins

Of 842 non-redundant proteins, 167 (20%) were identified via all three fractionation platforms (Fig. 2). Most, but not all,

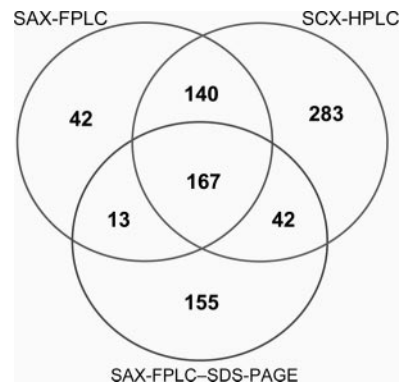


FIG. 2. Number of proteins identified using the three sample preparation methods. In total, 842 proteins were identified.

of the 167 proteins were of high abundance. Other proteins of high abundance were detected by two platforms. For example,  $\alpha_{1B}$ -glycoprotein (molecular mass, ~80 kDa) and  $\alpha_1$ -antichymotrypsin (molecular mass, ~65 kDa) were identified with over 10 unique peptides from each of the SAX-FPLC and SCX-HPLC platforms, but they were not detected by the SAX-FPLC-SDS-PAGE platform. This is likely due to the fact that the latter method was designed to remove albumin and other proteins of molecular mass of 50 kDa or higher. 195 (23%) of the proteins were identified by two platforms. The rest of the proteins (480 proteins, 57%) were identified by one platform (Fig. 2). This finding by no means reflects the reproducibility of our approach because the three different fractionation protocols were designed to complement rather than reproduce each other.

#### Characterization of the Amniotic Fluid Proteome

**Subcellular Localization**—Each identified protein was assigned a subcellular localization based on information from Swiss-Prot, Entrez Gene, and Gene Ontology databases. When one protein is known to be localized in more than one cellular compartment, all of the categories were counted non-exclusively. Fig. 3 shows the cellular distribution of 558 identified proteins with known localization. The majority are extracellular (42%) and membrane (26%) proteins. By searching the Plasma Proteome Database we found that 304 (36%) of the 842 proteins have also been found in plasma. This does not mean that the remaining 538 proteins are exclusive to AF because the plasma proteome list is still growing.

**Tissue Expression**—Tissue expression of each protein was searched from Swiss-Prot, Entrez Gene, and Gene Ontology databases. 10 functional categories were selected based on the number of hits per organ. When one protein is expressed in more than one tissue, only the major tissue of expression was counted (otherwise we would have identified too many tissues for the majority of proteins). Fig. 4 shows 301 proteins with known tissue expression information (supplemental data: unique genes). Some of the organs to which many proteins were attributed include kidney, placenta, lung, liver, and heart. 24

FIG. 3. Classification of 558 AF proteins by subcellular localization.

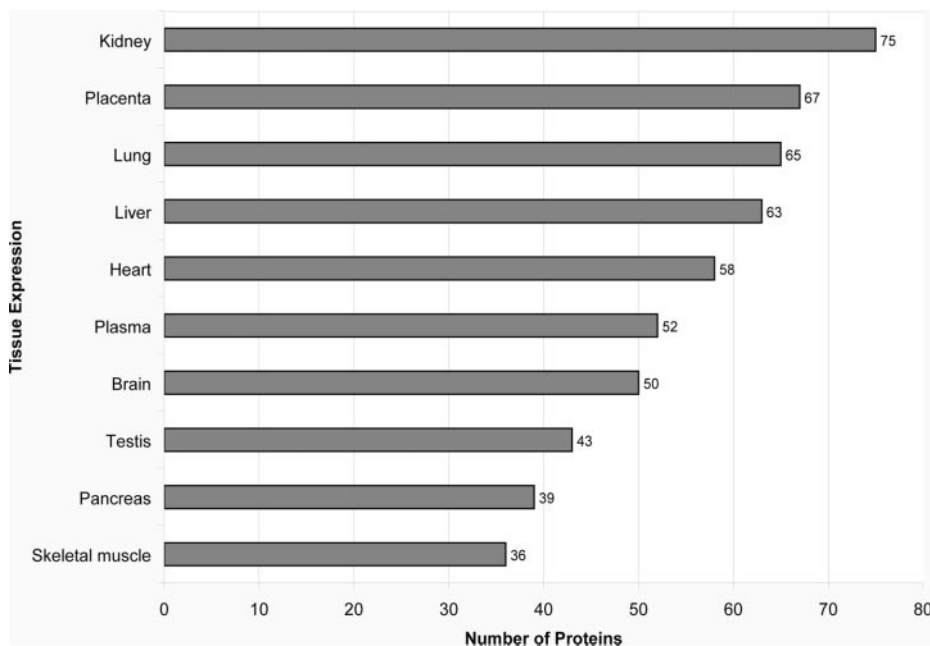
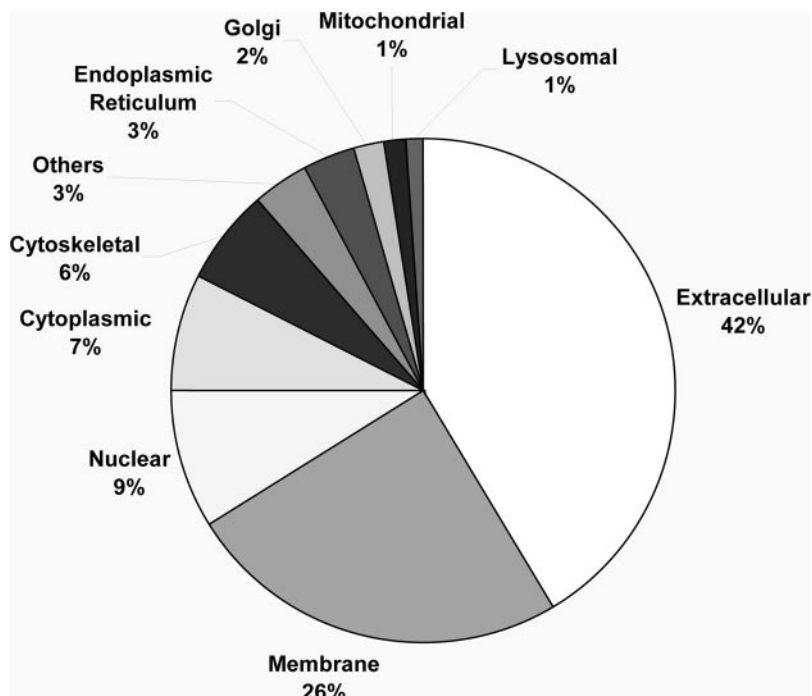


FIG. 4. Tissue expression profile of AF proteins. Tissue expression profiles of 301 proteins were collated to identify major tissues from which AF proteins are expressed. Shown are the top 10 tissues with the most number of proteins. Other major tissues include testis, intestine, pancreas, thymus, and spleen (30–35 proteins each).

proteins were specifically annotated as being expressed from embryonic organs/tissues (supplemental data: unique genes).

**Assignments of Molecular Function**—We utilized Ingenuity (Ingenuity Systems) to retrieve known functions of each protein. 221 (of 842) were matched with functions. Because one protein may have multiple functions we selected the functions with  $p$  value  $<0.015$ . Fig. 5 shows the top 15 of 72 different functions based on significance. Major categories included cellular movement, development and function of organs, cellular growth and proliferation, cancer, and cell-to-cell signal-

ing. More specific functions and the names of genes for each function are shown in the supplemental data. For example, the function “cancer” includes specific functions such as apoptosis, cell cycle regulation, and migration.

**Network Construction for Biological Processes**—Assignment of biological processes and subsequent construction of networks was done using Ingenuity software. We found that at least 227 different proteins are components of existing molecular networks. A total of 27 networks were constructed. One with the highest score is shown in Fig. 6. This particular

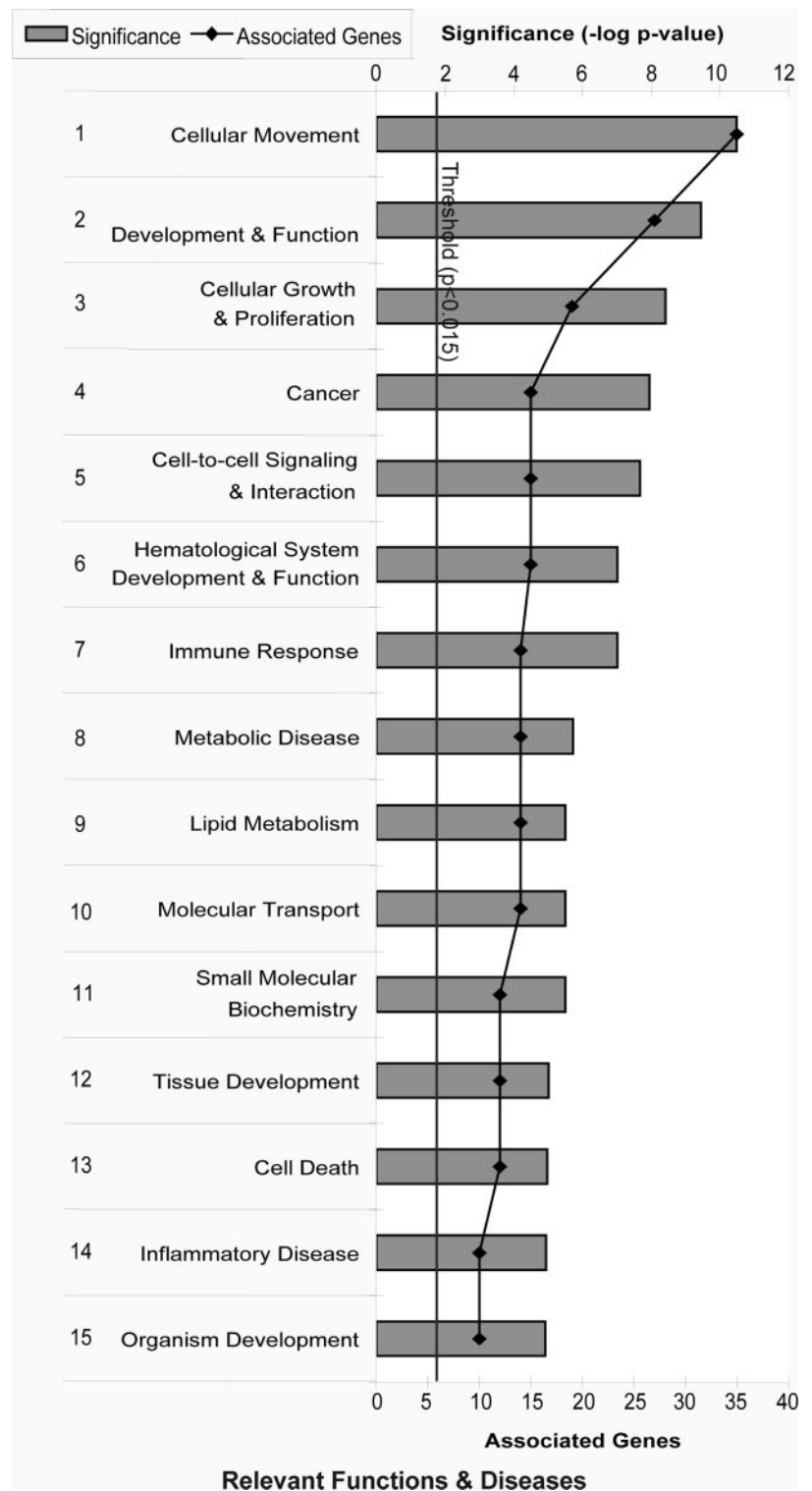


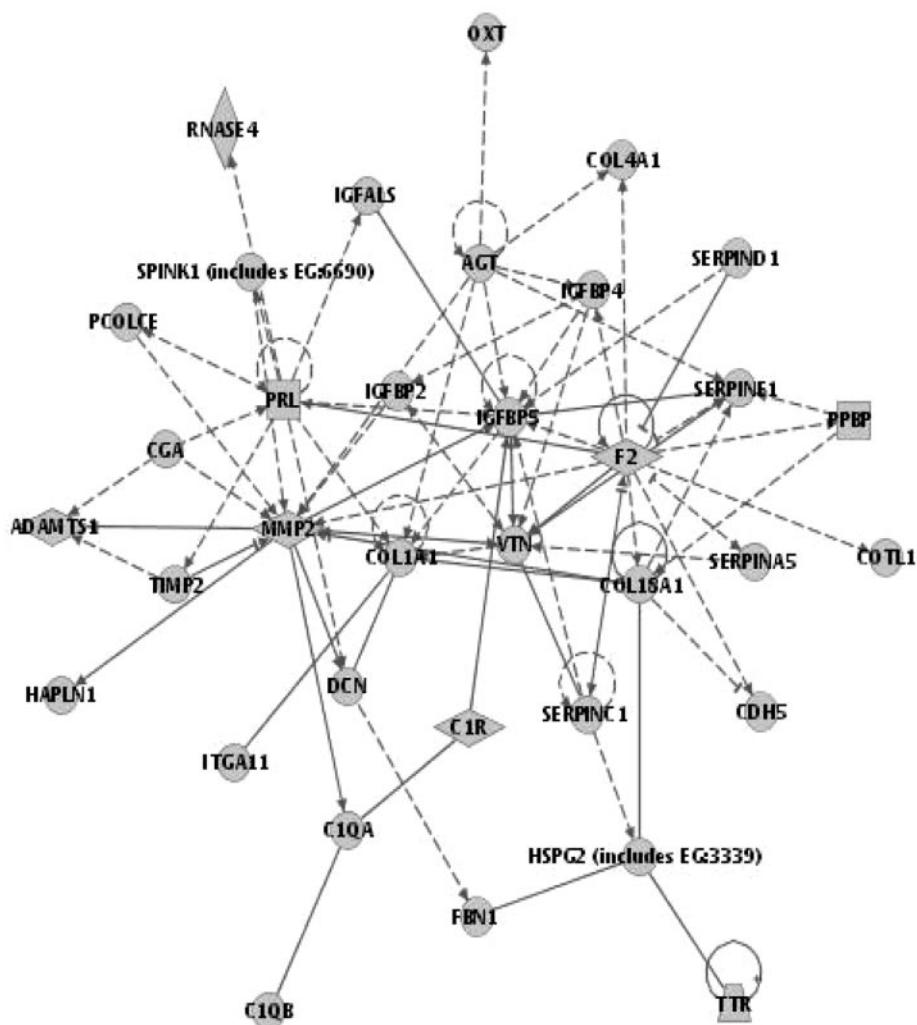
FIG. 5. **Top 15 functions of the proteins identified from AF.** 221 genes were matched to 72 functions with significance above 1.83 ( $p$  value  $<0.015$ ). The top 15 functions involving the 211 genes are shown here with the *number* of associated genes that were identified for each function.

network shows 35 genes that work together for cardiovascular system development, and proteins of all of these 35 genes were found in our AF proteome. A table with functions, involved genes, and significance scores of all 27 networks is presented in the supplemental data.

#### *Comparison of the Present Proteome with Previous Publications*

We combined human AF proteome entries from previous publications to generate one list (2, 4–9, 12). All keratins and Ig chains were removed as explained above. We also re-

FIG. 6. An example of a network constructed based on gene matches found in the AF proteome. 227 genes from the AF proteome were found to be components of 27 significant networks. The network shown scored the highest of the 27 networks as proteins from all of the 35 genes shown were identified in AF. The functions of this network include cardiovascular system development and function, cellular movement, and skeletal and muscular development and function.



moved redundancies as described above and eliminated non-human proteins. A combined list of 184 proteins was generated (supplemental data: publications). Fig. 7 shows a Venn diagram that compares the 842 proteins identified from our study with the 184 proteins from previous publications. All proteins that were previously reported in at least three publications (Fig. 7, Groups III, IV, and V) were identified in this study. For proteins that were identified in two publications, 15 of 19 were also found here. Finally 54 of 140 proteins that were identified in only one publication were also identified in this study. Assuming that all of the proteins identified in previous publications and in this study are correct, a total of 936 proteins have been identified from human AF so far.

We sorted the 842 proteins by the number of unique peptides identified (supplemental data: unique genes). The number of unique peptides is generally accepted as a semiquantitative measure of protein abundance. We used data from SAX-FPLC and SCX-HPLC only as these protocols do not exclude any proteins other than Igs. The top 15 proteins with the largest number of unique peptides are presented in Table

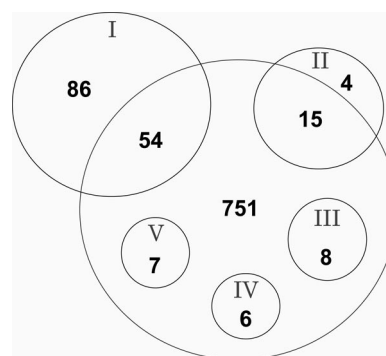


FIG. 7. Number of overlapping proteins from previous publications. The Venn diagram shows overlap between 184 proteins from previous studies and 842 proteins from this study. The largest circle indicates the proteins found by our group. Group I contains proteins listed only in one publication; Group II indicates proteins reported from two studies. Groups III, IV, and V are proteins reported in three or more publications.

TABLE I  
Top 15 high abundance proteins from AF

Amniotic fluid proteome <sup>a</sup>		Rank	Plasma proteome (13)	
Gene	Protein		Gene	Protein
<i>ALB</i>	Albumin	1	<i>ALB</i>	Albumin
	Immunoglobulins	2		Immunoglobulins
<i>FN1</i>	Fibronectin	3	<i>TF</i>	Serotransferrin
<i>TF</i>	Serotransferrin	4	<i>FG</i>	Fibrinogen
<i>C3</i>	Complement C3	5	<i>AMBP</i>	$\alpha_1$ -Microglobulin
<i>SERPINA1</i>	$\alpha_1$ -Antitrypsin	6	<i>SERPINA1</i>	$\alpha_1$ -Antitrypsin
<i>CP</i>	Ceruloplasmin	7	<i>C3</i>	Complement C3
<b><i>AFP</i></b> <sup>b</sup>	$\alpha$ -Fetoprotein	8	<b><i>HP</i></b>	Haptoglobin
<i>GC</i>	Vitamin D-binding protein	9	<i>APOA1</i>	Apolipoprotein A-I
<b><i>POSTN</i></b>	Periostin	10	<b><i>APOB</i></b>	Apolipoprotein B
<i>APOA1</i>	Apolipoprotein A-I	11	<i>ORM1</i>	$\alpha_1$ -Acid glycoprotein
<i>SERPINC1</i>	Antithrombin III	11/12		Lipoprotein
<b><i>TGFBI</i></b>	Transforming growth factor $\beta$ -induced protein ig-h3 precursor	13	<i>CFH</i>	Factor H
<i>AMBP</i>	$\alpha_1$ -Microglobulin	14	<i>CP</i>	Ceruloplasmin
<i>PLG</i>	Plasminogen	15	<i>C4</i>	Complement C4

<sup>a</sup> Proteins were ranked by the number of unique peptides identified from SAX and SCX methods. The top 15 proteins from the plasma proteome are provided for comparison.

<sup>b</sup> Proteins in bold are found in high abundance in either AF or plasma; other proteins are found in high abundance in both fluids.

I along with the top 15 proteins from the human plasma proteome (13). Among these 15 abundant proteins from AF,  $\alpha$ -fetoprotein, transforming growth factor  $\beta$ -induced protein ig-h3 precursor, and periostin are found at relatively low concentrations in plasma and are not included in the top 15 plasma proteins. Conversely haptoglobin, which is one of the most abundant proteins in plasma, is found at low concentration in AF, and apolipoprotein B was identified in AF by only one group (8), indicating its low abundance in AF.

#### DISCUSSION

The complex nature of biological fluids requires well established methodologies for depletion of high abundance proteins and efficient sample fractionation before proteomics analysis. Among various methods, LC-MS/MS allows identification of proteins in a high throughput fashion unlike the rather slow and laborious 2DE-MS/MS methods. To avoid the masking effect of high abundance proteins, methods based on dye affinity resins (Cibacron blue), Protein A/G, and IgY combinations have been tried (14, 15). These approaches have two major disadvantages. First, albumin or other high abundance proteins bind to small proteins or peptides, a phenomenon called the "sponge" effect. Thus, elimination of proteins of interest along with high abundance proteins is likely. Second, the most specific and effective immunoaffinity depletion columns are costly and have limited life span and capacity.

Removal of immunoglobulins may still eliminate some small proteins or peptides that bind to them. However, compared with other proteins, Ig peptides tend to swamp the database searches; thus the benefits of removal outweigh the potential loss. Therefore, we used Protein A/G beads to remove Igs from each sample to minimize their interference. Although there were still residual Igs left behind (11% of 1,026 identified gene matches were Igs), the total number of identified proteins markedly increased when Igs were depleted. Two dif-

ferent types of ion exchange chromatography were used based on the premise that proteins masked by one type of ion exchange chromatography may be detected with the other type. SAX-FPLC followed by SDS-PAGE and in-gel trypsin digestion was also performed after depleting proteins with molecular mass greater than 50 kDa by ultrafiltration. Although this method has a number of disadvantages, including low recovery of total protein, 155 proteins were uniquely identified. Fig. 2 shows that the majority of proteins of our proteome (480 proteins, 57%) were identified by only one platform, and 101 of them (21%) were identified with two or more peptides. These 101 proteins comprise 24% of the "new" proteins with two peptides or more. The SDS-PAGE/SCX-HPLC platform identified the highest number of proteins that fall into this category. Our results indicate that combining these approaches allowed for greater coverage of the AF proteome.

Despite interest regarding composition and functions of amniotic fluid, there have been limited attempts to generate an in-depth analysis of its proteome. Collectively eight previous studies have identified 184 proteins, excluding keratins and Igs. Moreover previous studies on AF proteins lacked the stringency required for high confidence identification of proteins. In this study, we used a conservative identification strategy with a low false-positive rate (~3%) and successfully identified 751 new proteins. When compared with the proteins from previous studies, most of the proteins that were not identified here were identified by only one study (86 of 184 proteins; Fig. 7), whereas all proteins found three times or more were also identified in this study.

Fig. 5 hints at molecular functions of AF that are as yet unknown. An analysis of networks for biological processes, as shown in Fig. 6, will not only provide clues on how different proteins in AF interact to aid in the normal development of



TABLE II

Markers for pathological conditions that were identified in amniotic fluid

AMBP,  $\alpha$ -1-microglobulin precursor.

Pathology	Markers <sup>a</sup>
Down syndrome	<b>hCG <math>\beta</math> chain</b> , <u>AFP</u> , <b>inhibin A (<math>\beta</math>-A chain)</b> , glyceraldehyde-3-phosphate dehydrogenase (16), AMBP, collagen $\alpha$ 1 (I) (4), collagen $\alpha$ 1 (III) (4), collagen $\alpha$ 1 (V) (4), basement membrane-specific heparin sulfate proteoglycan core protein, IBP-1 (4), <b>activin A (<math>\beta</math>-A chain) (17)</b>
Trisomy 13	<u>AFP</u> , <u>hCG <math>\beta</math> chain</u>
Trisomy 18	<u>AFP</u> , <u>hCG <math>\beta</math> chain</u>
Preterm delivery	<b>CD163 (18)</b>
Fetoplacental hypoxemia	activin A ( $\beta$ -A chain) (19)
Intra-amniotic infection	calgranulin A (20), calgranulin B (12), calgranulin C (21), vitamin D-binding protein (12), IGFBP-1 (12), <b>neutrophil defensin-1 (21)</b>
Pre-eclampsia	fibronectin (22), <b>intercellular adhesion molecule-1 (23)</b> , <b>plasminogen activator inhibitor (24)</b>
Rupture of the membrane	prolactin (8), AFP, IGFBP-1 (25), <b>hCG <math>\beta</math> chain</b> , fibronectin (8), agrin (2), plasma retinol-binding protein precursor (2), apolipoprotein A-I (2), B-factor (2)
<i>Ureaplasma</i> infection	<b>intercellular adhesion molecule-1 (26)</b>

<sup>a</sup> Currently used markers as well as putative markers previously identified in AF are listed with the corresponding pathologies. Underlined proteins represent clinically used markers. Bold proteins were identified by MS for the first time in this study.

fetus but also allows exploration for novel biomarkers based on known functional pathways.

Many proteins present in AF are of interest because their expression reflects physiological and pathological conditions of the fetus and/or pregnancy. For example, AFP is actively investigated because a change in its concentration in maternal serum correlates with several abnormalities of the fetus, including spina bifida, Down syndrome, trisomy 13, and trisomy 18. The levels of proteins like AFP in maternal serum are highly related to pregnancy. These proteins are produced and secreted by either the fetus or the placenta as pregnancy progresses. Therefore, the expression of such markers in AF has been actively studied for pathologies such as Down syndrome, trisomies 13 and 18, intra-amniotic infection, preterm delivery, pre-eclampsia, rupture of membrane, and fetoplacental hypoxia.

Our study identified many of the proteins that are putative or currently used as markers of fetal well-being. Of the proteins in Table II, several maternal serum markers were identified here for the first time via MS, including hCG and the  $\beta$ -A chains of activin and inhibin A for Down syndrome. Also, several putative markers were found including CD163 for preterm delivery, neutrophil defensin-1 for intra-amniotic infection, intercellular adhesion molecule-1 and plasminogen activator inhibitor for pre-eclampsia, hCG for rupture of the membrane, and intercellular adhesion molecule-1 for *Ureaplasma* infection. These data further support the view that quantitative shotgun proteomics analysis of AF may be a feasible and effective method to screen multiple pathologies in the future. It should be noted, however, that AF is a very dynamic mixture, the composition of which is known to differ depending on gestational age. Therefore, the data of Figs. 3–6 may be specific for gestational ages of 16–18 weeks. This period was chosen because most amniocenteses are done during this gestational age.

Although AF is initially created from infiltrated maternal serum, its composition is known to change as pregnancy proceeds. The top 15 proteins from the plasma proteome and the AF proteome from this study were compared, and some significant differences were noted (Table I). Moreover only 36% of the AF proteins were also found in the plasma proteome. Even if it is taken into account that the plasma proteome is not complete, these results suggest that AF has a significantly different composition than plasma and is associated with different functions. Some of the top functions of AF proteins, as shown in Fig. 5, include cellular movement, development and function of organs, cellular growth and proliferation, apoptosis and cell cycle (shown as cancer), cell-to-cell signaling and interaction, hematological system development and function, and immune response, all involving at least 15 proteins each from our lists. More detailed functions and the names of proteins responsible for each function are listed separately (supplemental data: functions). It is highly likely that these functions are closely associated with the molecular and physiological pathways operating during this period of pregnancy.

The major challenge and the bottleneck in proteomics lie between protein discovery and target validation. Proteins with potential as biomarkers should be selected for further comparative analysis of expression and structural modifications with samples from chromosomally normal and abnormal pregnancies. Currently used markers have insufficient individual detection rate and specificity; therefore the use of multiple markers is necessary. For example, current prenatal screening for Down syndrome includes three to five different maternal serum markers. It is now perceived that an integrated test or fingerprinting may reflect a patient's status better than one marker alone. Therefore, an MS-derived systematic expression profiling of proteins from a network (similar to the one shown in Fig. 6) for an affected biological process may reflect

an effective screening tool.

In summary, 1,026 unique gene matches from 842 genes were identified from chromosomally normal human amniotic fluid acquired at 16–18 weeks of gestation. Various aspects of the AF proteome including functions and tissue expression were explored utilizing bioinformatics tools. We believe that this study provides the starting point for further functional studies of AF as well as for discovery of novel biomarkers for pregnancy-associated abnormalities.

**Acknowledgments**—We thank Vathany Kulasingam, Girish Sardana, Chris Smith, and Dr. Barry Hoffman for helpful discussions and advice.

\* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ The on-line version of this article (available at <http://www.mcponline.org>) contains supplemental material.

|| To whom correspondence should be addressed: Dept. of Pathology and Laboratory Medicine, Mount Sinai Hospital, 60 Murray St., 6th Fl., Rm. 6-201, Toronto, Ontario M5T 3L9, Canada. Tel.: 416-586-8443; Fax: 416-619-5521; E-mail: ediamandis@mtnsina.on.ca.

#### REFERENCES

- Underwood, M. A., Gilbert, W. M., and Sherman, M. P. (2005) Amniotic fluid: not just fetal urine anymore. *J. Perinatol.* **25**, 341–348
- Vuadens, F., Benay, C., Crettaz, D., Gallot, D., Sapin, V., Schneider, P., Bienvenut, W. V., Lemery, D., Quadroni, M., Dastugue, B., and Tissot, J. D. (2003) Identification of biologic markers of the premature rupture of fetal membranes: proteomic approach. *Proteomics* **3**, 1521–1525
- Thadikkaran, L., Crettaz, D., Siegenthaler, M. A., Gallot, D., Sapin, V., Iozzo, R. V., Queloz, P. A., Schneider, P., and Tissot, J. D. (2005) The role of proteomics in the assessment of premature rupture of fetal membranes. *Clin. Chim. Acta* **360**, 27–36
- Tsangaris, G. T., Karamessinis, P., Kolialexi, A., Garbis, S. D., Antsaklis, A., Mavrou, A., and Fountoulakis, M. (2006) Proteomic analysis of amniotic fluid in pregnancies with Down syndrome. *Proteomics* **6**, 4410–4419
- Liberatori, S., Bini, L., De Felice, C., Magi, B., Marzocchi, B., Raggiaschi, R., Frutiger, S., Sanchez, J. C., Wilkins, M. R., Hughes, G., Hochstrasser, D. F., Bracci, R., and Pallini, V. (1997) A two-dimensional protein map of human amniotic fluid at 17 weeks' gestation. *Electrophoresis* **18**, 2816–2822
- Nilsson, S., Ramstrom, M., Palmblad, M., Axelsson, O., and Bergquist, J. (2004) Explorative study of the protein composition of amniotic fluid by liquid chromatography electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *J. Proteome Res.* **3**, 884–889
- Park, S. J., Yoon, W. G., Song, J. S., Jung, H. S., Kim, C. J., Oh, S. Y., Yoon, B. H., Jung, G., Kim, H. J., and Nirasawa, T. (2006) Proteome analysis of human amnion and amniotic fluid by two-dimensional electrophoresis and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proteomics* **6**, 349–363
- Michel, P. E., Crettaz, D., Morier, P., Heller, M., Gallot, D., Tissot, J. D., Reymond, F., and Rossier, J. S. (2006) Proteome analysis of human plasma and amniotic fluid by off-gel isoelectric focusing followed by nano-LC-MS/MS. *Electrophoresis* **27**, 1169–1181
- Tsangaris, G. T., Kolialexi, A., Karamessinis, P., Anagnostopoulos, A. K., Antsaklis, A., Fountoulakis, M., and Mavrou, A. (2006) The normal human amniotic fluid supernatant proteome. *In vivo* **20**, 479–490
- Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383–5392
- Anderson, N. L., Polanski, M., Pieper, R., Gatlin, T., Tirumalai, R. S., Conrads, T. P., Veenstra, T. D., Adkins, J. N., Pounds, J. G., Fagan, R., and Lobley, A. (2004) The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* **3**, 311–326
- Gravett, M. G., Novy, M. J., Rosenfeld, R. G., Reddy, A. P., Jacob, T., Turner, M., McCormack, A., Lapidus, J. A., Hitti, J., Eschenbach, D. A., Roberts, C. T., Jr., and Nagalla, S. R. (2004) Diagnosis of intra-amniotic infection by proteomic profiling and identification of novel biomarkers. *J. Am. Med. Assoc.* **292**, 462–469
- Tirumalai, R. S., Chan, K. C., Prieto, D. A., Issaq, H. J., Conrads, T. P., and Veenstra, T. D. (2003) Characterization of the low molecular weight human serum proteome. *Mol. Cell. Proteomics* **2**, 1096–1103
- Ogata, Y., Charlesworth, M. C., and Muddiman, D. C. (2005) Evaluation of protein depletion methods for the analysis of total-, phospho- and glycoproteins in lumbar cerebrospinal fluid. *J. Proteome Res.* **4**, 837–845
- Zolotarjova, N., Martosella, J., Nicol, G., Bailey, J., Boyes, B. E., and Barrett, W. C. (2005) Differences among techniques for high-abundant protein depletion. *Proteomics* **5**, 3304–3313
- Lubec, G., Labudova, O., Cairns, N., and Fountoulakis, M. (1999) Increased glyceraldehyde 3-phosphate dehydrogenase levels in the brain of patients with Down's syndrome. *Neurosci. Lett.* **260**, 141–145
- Wallace, E. M., D'Antona, D., Shearing, C., Evans, L. W., Thirunavukarasu, P., Ashby, J. P., Shade, M., and Groome, N. P. (1999) Amniotic fluid levels of dimeric inhibins, pro-alpha C inhibin, activin A and follistatin in Down's syndrome. *Clin. Endocrinol. (Oxf.)* **50**, 669–673
- Vogel, I., Grove, J., Thorsen, P., Moestrup, S. K., Ulbjerg, N., and Moller, H. J. (2005) Preterm delivery predicted by soluble CD163 and CRP in women with symptoms of preterm delivery. *Br. J. Obstet. Gynaecol.* **112**, 737–742
- Jenkin, G., Ward, J., Hooper, S., O'Connor, A., de Kretser, D., and Wallace, E. (2001) Feto-placental hypoxemia regulates the release of fetal activin A and prostaglandin E(2). *Endocrinology* **142**, 963–966
- Ruetschi, U., Rosen, A., Karlsson, G., Zetterberg, H., Rymo, L., Hagberg, H., and Jacobsson, B. (2005) Proteomic analysis using protein chips to detect biomarkers in cervical and amniotic fluid in women with intra-amniotic inflammation. *J. Proteome Res.* **4**, 2236–2242
- Buhimschi, I. A., Christner, R., and Buhimschi, C. S. (2005) Proteomic biomarker analysis of amniotic fluid for identification of intra-amniotic inflammation. *Br. J. Obstet. Gynaecol.* **112**, 173–181
- Stubbs, T. M., Lazarchick, J., and Horger, E. O., III. (1984) Plasma fibronectin levels in preeclampsia: a possible biochemical marker for vascular endothelial damage. *Am. J. Obstet. Gynecol.* **150**, 885–887
- Krauss, T., Kuhn, W., Lakoma, C., and Augustin, H. G. (1997) Circulating endothelial cell adhesion molecules as diagnostic markers for the early identification of pregnant women at risk for development of preeclampsia. *Am. J. Obstet. Gynecol.* **177**, 443–449
- Gao, M., Nakabayashi, M., Sakura, M., and Takeda, Y. (1996) The imbalance of plasminogen activators and inhibitor in preeclampsia. *J. Obstet. Gynaecol. Res.* **22**, 9–16
- Rutanen, E. M., Karkkainen, T. H., Lehtovirta, J., Uotila, J. T., Hinkula, M. K., and Hartikainen, A. L. (1996) Evaluation of a rapid strip test for insulin-like growth factor binding protein-1 in the diagnosis of ruptured fetal membranes. *Clin. Chim. Acta* **253**, 91–101
- Hadar, A., Ytzchaki, O., Shani-Shrem, N., Horowitz, S., Mazor, M., and Horowitz, S. (2006) Amniotic fluid intercellular adhesion molecule-1 is a marker for Ureaplasma infection in women with preterm labor. *Am. J. Obstet. Gynecol.* **195**, 1601–1606