Quantitative Age-specific Variability of Plasma Proteins in Healthy Neonates, Children and Adults*

Stefan Bjelosevic‡§, Dana Pascovic¶, Hui Ping‡§, Vasiliki Karlaftis‡, Thiri Zaw¶, Xiaomin Song¶, Mark P. Molloy¶, Paul Monagle‡§, and Vera Ignjatovic‡§

Human blood plasma is a complex biological fluid containing soluble proteins, sugars, hormones, electrolytes, and dissolved gasses. As plasma interacts with a wide array of bodily systems, changes in protein expression, or the presence or absence of specific proteins are regularly used in the clinic as a molecular biomarker tool. A large body of literature exists detailing proteomic changes in pathologic contexts, however little research has been conducted on the quantitation of the plasma proteome in age-specific, healthy subjects, especially in pediatrics. In this study, we utilized SWATH-MS to identify and quantify proteins in the blood plasma of healthy neonates, infants under 1 year of age, children between 1–5 years, and adults. We identified more than 100 proteins that showed significant differential expression levels across these age groups, and we analyzed variation in protein expression across the age spectrum. The plasma proteomic profiles of neonates were strikingly dissimilar to the older children and adults. By extracting the SWATH data against a large human spectral library we increased protein identification more than 6-fold (940 proteins) and confirmed the concentrations of several of these using ELISA. The results of this study map the variation in expression of proteins and pathways often implicated in disease, and so have significant clinical implication. Molecular & Cellular Proteomics 16: 10.1074/mcp.M116.066720, 924–935, 2017.

Human blood plasma is a complex but ideal biological system containing proteins from a variety of cellular localizations. As such plasma is one of the best reporter systems in the clinical setting, as it communicates with most parts of the body and reflects changes in the state of the organism as a whole. Unlike serum, which is depleted of coagulation proteins that play a role not only in coagulation but also in processes such as inflammation, human plasma contains all the soluble proteins found in blood. Hence, human plasma is a commonly used sample for biomarker discovery.

Proteomics studies utilizing human plasma have, to date, focused largely on specific disease settings, on detecting the highest number of proteins or on the effect of certain drugs on changes in plasma protein expression. In addition, most such studies have focused on adults, with limited number of studies and hence limited knowledge of the plasma proteome in infants and children. The plasma proteomic studies in children have focused on examining changes in protein expression that occur during Cardiopulmonary Bypass Surgery (1), as well as in several disease states including: blood disorders (2), sickle cell disease (3), leukemia (4), and cystic fibrosis (5). There has, to date, only been one study that has focused on the physiological development of the plasma proteome by examining the changes in plasma protein expression that occur with age, from neonates through to adults (6). That study utilized two-dimensional difference in gel electrophoresis (2D-DIGE) methodology and described the concept of aging of the human plasma proteome by identifying significant age-specific changes in expression of up to 100 proteins spots.

The Human Proteome Organization (HUPO) Plasma Proteome Project (PPP) recognizes the importance of analyzing and understanding age related differences in the plasma proteome by identifying this as one of their scientific aims and

From the ‡Hematology Research Laboratory, Murdoch Childrens Research Institute, Melbourne, VIC 3052, Australia; §Department of Paediatrics, The University of Melbourne, Melbourne, VIC 3010, Australia; ¶Australian Proteome Analysis Facility, Macquarie University, Sydney, NSW 2113, Australia

Received January 23, 2017, and in revised form, February 26, 2017
Published, MCP Papers in Press, March 23, 2017, DOI 10.1074/mcp.M116.066720

Author contributions: M.P.M., P.M., and V.I. designed research; S.B., D.P., H.P., V.K., T.Z., and X.S. performed research; S.B., D.P., V.K., X.S., M.P.M., P.M., and V.I. analyzed data; S.B., D.P., M.P.M., P.M., and V.I. wrote the paper.

1 The abbreviations used are: ELISA, enzyme linked immunosorbent assay; 2D-DIGE, two-dimensional difference in-gel electrophoresis; HUPO, Human proteome Organization; PediOme, Pediatric Proteomics; DIA, data independent acquisition; TOFMS, time of flight mass spectrometry; ANOVA, analysis of variance; GPLD1, Gpl-specific phospholipase D; POSTN, periostin; A2M, alpha-2-macroglubulin; HRG, histidine rich glycoprotein; IDA, information dependent acquisition; PCA, principal component analysis; HBG1, hemoglobin subunit gamma-1; HBG2, hemoglobin subunit gamma-2; CO1A1, collagen alpha-1(I) chain; IGHA1, Ig alpha-1 chain C region; IGHA2, Ig alpha-2 chain C region; HBF, fetal hemoglobin; HgA, adult hemoglobin; TCERG1, transcription elongation regulator 1; ECHA, tri-functional enzyme subunit alpha; LC-MS, liquid chromatography mass spectrometry.
research priorities (7). In addition, the main aim of the recently established Pediatric Proteomics (PediOmE) initiative, a part of HUPO, is to advance the use of proteomics techniques to solve major issues in pediatric medicine through characterization of the pediatric proteome across a wide-variety of tissues and biological samples.

In 2012, a new proteomics approach known as SWATH-MS became available (8). This high-throughput, data independent acquisition (DIA) method allows confident identification and quantification of query peptides across large sample sets with a dynamic range of 4 orders of magnitude, and is as such well suited to studies of human plasma. A particular advantage of SWATH-MS is the ability to re-examine previously acquired data sets. As an emerging technology, to date only a handful of studies have utilized SWATH-MS for investigations of human plasma (9, 10). The first study investigated the performance and suitability of SWATH-MS for quantification of N-linked glycoproteins in human plasma and suggested that this combined approach is suitable for biomarker discovery and verification (9). The second study utilized SWATH-MS to quantify the variability in the plasma proteome in monozygotic and dizygotic twins, demonstrating large variability in protein abundance between twins (10).

The use of SWATH-MS technology in neonatal and pediatric plasma samples to determine variability and subsequent comparison to adult plasma remains an exciting research opportunity, and one that has, to date, not been explored. Therefore, this study aimed to utilize SWATH-MS to quantify age-specific variability of plasma proteins in healthy neonates, children, and adults.

**EXPERIMENTAL PROCEDURES**

**Participant Recruitment, Sample Collection and Storage—Blood samples used in this study were collected and processed according to established protocols (6, 11–15). Neonatal samples (within 72 h postbirth) were collected from healthy term neonates from the Family Birthing Unit or post-natal wards at the Royal Women’s Hospital, Melbourne. Eligibility criteria included: gestation >37 weeks, vaginal delivery, birth-weight >2500gm, APGAR at 5 min >7 and absence of systemic abnormalities. Pediatric samples were obtained from healthy children attending the hospital for minor elective day-surgery (e.g. tongue-tie release and circumcision) who were not receiving any medications and had no significant family history of major diseases, particularly diseases known to be associated with the process of aging (e.g. diabetes and cardiovascular disease). The pediatric cohort excluded adolescents taking oral contraceptive and smokers. Adult samples were obtained from healthy volunteers aged between 20 and 45 years of age who were not taking any medications (including oral contraceptive), were not smokers, and similarly, with no past medical history. Family history was assessed via a brief interview with the parents of the children and the adult volunteers themselves. Blood samples were collected in S-Monovette® tubes (Sarstedt, Australia), containing 1 volume of citrate per 9 volumes of blood and were centrifuged at 3000 rpm, 10 min, 10°C, with plasma stored at −80 °C, until testing.

This study was approved by the Royal Children’s Hospital Ethics in Human Research Committee, reference number 20031; and the Royal Women’s Hospital Human Research Ethics Committee (#02/08). Written informed consent was obtained from parents of children and from adult participants themselves.

**Plasma Sample Preparation—** 25 μl of human plasma sample was diluted in 475 μl of 50 mM ammonium bicarbonate (Sigma, St. Louis, MO) solution. Samples were reduced with 5 μM dithiothreitol (Bio-Rad, Hercules, CA) at 65 °C for 30 min, followed by alkylation with 10 mM iodoacetamide (Sigma) at room temperature for 30 min in the dark. One fifth of the reduced and alkylated sample (121 μl) was digested with sequencing-grade porcine trypsin (20 μg, Promega) overnight at 37 °C. The digested sample was diluted (>10.5) in 0.1% formic acid prior to all mass spectrometry analyses.

**Information Dependent Acquisition Mass Spectrometry—** A 5600 TripleTOF mass spectrometer (Sciex, Framingham, MA) coupled to an Eksigent Ultra-nanoLC-1D system (Eksigent Technologies, Dublin, CA) was employed for both IDA and SWATH-MS analysis. For IDA data, 10 μl of digested sample from each group was pooled, making four group pools (Neonate, <1 year, 1–5 year, adult) and 5 μl from each group was injected.

Peptides were loaded onto a reverse phase peptide C18 Captrap (Bruker, Billerica, MA) for pre-concentration and desalted for 5 min with the loading buffer (0.1% (v/v) formic acid, 2% (v/v) acetonitrile) at a flow rate of 10 μl per minute. After desalting, the peptide trap was switched in-line with an in-house packed analytical column (150 μm x 10 cm, solid core Halo C18, 160 Å, 2.7 μm (Bruker)). Peptides were eluted and separated from the column using the buffer B (99.9% (v/v) acetonitrile, 0.1% (v/v) formic acid) gradient starting from 2% and increasing to 10% in 10 min then to 35% over 78 min at a flow rate of 500 nL per minute. After peptide elution, the column was flushed with 95% buffer B for 15 min and re-equilibrated with 98% buffer A (0.1% (v/v) formic acid) for 15 min before next injection. Blanks were run between samples. In IDA mode, a TOFMS survey scan was acquired at m/z 350–1500 with 0.25 s accumulation time, with the ten most intense precursor ions (22–55; counts > 150) in the survey scan consecutively isolated for subsequent product ion scans. Dynamic exclusion was used with a window of 20 s. Product ion spectra were accumulated for 50 milliseconds in the mass range m/z 100–1500 with rolling collision energy.

**IDA Data Analysis—** Protein identification from IDA data was performed with ProteinPilot (v4.2, Sciex) using the Paragon algorithm. The search parameters were as follows: sample type: identification; cysteine alkylation: iodoacetamide; digestion: trypsin; instrument: TripleTOF 5600; special factors: none; ID focus: biological modifications; miss-cleavages: two. The Homo sapiens search database was obtained from UniProt (20,265 entries, Jan 2015). A reversed-decoy extended peptide assay library was constructed using the SWATHX-IDA search result of the 4 sample groups was imported into PeakView (v2.1) (Sciex) and used as a local peptide assay library. This library contained 151 proteins identified from un-depleted plasma samples. An extended peptide assay library was constructed using the SWATHX-tend script (16), which merged the local peptide assay library with a 10,000 protein human (17) spectral library downloaded from SWATHAtlas.

**SWATH-MS—** For SWATH-MS experiments 5 μl of each digested sample was injected in sample order, with two technical replicates for each injection; one sample was run in triplicates once at the beginning and twice at the end of the series to check for potential instrument drift as detailed in Supplemental Information. Identical LC conditions were used as described above, with m/z window sizes determined based on precursor m/z frequencies in previous IDA data. SWATH variable window acquisition with a set of 60 overlapping windows (1 amu for the window overlap) was constructed covering the mass range of m/z 399.5 - 1249.5. In SWATH mode, first a TOFMS survey scan was acquired (m/z 350–1500, 0.05 s) then the 60 predefined m/z...
Supplemental Table S2

Student's peptide-level ratios for each protein. The reporting threshold required a log-transformed protein areas, and secondly by combining individual conditions were compared first by a two-sample Student's t-test of the log-transformed normalized protein peak areas, and pairwise comparisons of pairs of samples. The extracted transition ion peak areas, peptide peak areas and protein peak areas were exported in Excel for further statistical analysis.

Experimental Design and Statistical Rationale—Quantitative MS data was obtained from ten independent biological samples for each of the four age cohort groups. Two technical replicate were obtained for each biological sample in SWATH acquisition. The peptide ion peak areas were averaged for the replicate technical injections, then further scaled so that the sum of the ion intensities of each sample equal the maximum total ion intensity. The normalized ion intensities were summed for each peptide and protein. The entire protein level data was clustered using hierarchical clustering (Euclidean distance and complete linkage), and visualized using a principal component analysis to examine global trends. The data distribution was checked visually using boxplots and density plots (Supplemental information: data quality).

Two approaches were undertaken for determining differential expression: an analysis of variance on the log-transformed normalized protein peak areas, and pairwise comparisons of pairs of samples. For the analysis of variance, proteins were deemed to be differentially expressed if the ANOVA p value was less than 0.05, and the maximum protein fold change exceeded 1.5. We evaluated adjusting the ANOVA p values for multiple testing using the fdr criterion of Benjamini and Hochberg, however the combination of unadjusted approach compared with the 121 proteins identified using a Benjamini and Hochberg adjusted p value less than 0.05 (data not shown). The minimum fold change, of 1.51 was observed for Apolipoprotein-M, ranging up to greater than 300-fold change for Hemoglobin subunit gamma-2 which is significantly elevated in neonates. Specifically, there were 86 (80%) proteins with a maximum fold change ranging between 1.5 and 5; 11 (10%) proteins with a maximum fold change between 5- and 10-fold; 6 (6%) proteins with a maximum fold change between 10- and 50-fold and 4 (4%) proteins with a maximum fold change >50-fold.

RESULTS

Sample Characteristics—SWATH-MS analysis was performed on 10 individual samples from each of the following age groups: Neonates (5F/5M, within 72 h post-birth), <1 year olds (3F/7M, median age 7.7 months, range 6 months to 11 months), 1–5 year olds (5F/5M, median age 3.12 years, range 1.59 to 4.39 years) and adults (5F/5M, median age 31.40 years, range 22.02 to 44.97 years).

SWATH-MS—We used nondepleted plasma samples from the sample cohort and IDA to establish a local peptide assay library containing 146 proteins. SWATH acquisition of the 40 individual samples were obtained, and then matched to the local assay library for quantitation of peptide areas (supplemental Table S1).

Overall Differences Across the Age-spectrum From Neonates to Adults—Hierarchical clustering of the 40 plasma samples using SWATH quantitation of 146 proteins showed well separated clusters corresponding to the experimental age groups. Interestingly, this analysis clearly showed that plasma protein levels from the neonates were quite distinct from the three remaining age groups, with adults and one to five year olds clustering next to each other (Fig. 1). These trends were confirmed by principal component analysis, showing clear segregation based on age, with neonates clustering separately compared with all other age groups.

Given the strong separation of the samples with unsupervised analysis (clustering and PCA), not surprisingly a large percentage of the proteins were found to be differentially expressed based on the analysis of variance (ANOVA). A total of 107 proteins were found to have ANOVA p value < 0.05 and maximum fold change (ratio of highest average to the lowest average) at least 1.5 (Table I). This was a stricter approach compared with the 121 proteins identified using a Benjamini and Hochberg adjusted p value less than 0.05 (data not shown). The minimum fold change, of 1.51 was observed for Apolipoprotein-M, ranging up to greater than 300-fold change for Hemoglobin subunit gamma-2 which is significantly elevated in neonates. Specifically, there were 86 (80%) proteins with a maximum fold change ranging between 1.5 and 5; 11 (10%) proteins with a maximum fold change between 5- and 10-fold; 6 (6%) proteins with a maximum fold change between 10- and 50-fold and 4 (4%) proteins with a maximum fold change >50-fold.

The resulting hierarchical clustering of these 107 proteins confirmed a similar pattern to that observed for the unsupervised clustering of the overall data (Fig. 2A). These proteins could be grouped into 6 specific clusters of expression (Fig. 2B). The list of proteins included in each cluster is presented in supplemental Table S2. The most interesting of these were clusters 2 and 4, with 25 and 30 proteins, respectively, showing increased expression with increasing age; and clusters 1 and 5, with 14 and 9 proteins, respectively, showing decreased expression with increasing age. Proteins included in these clusters represent the KEGG pathways of complement and coagulation cascade, with the most significant gene ontology functions being: acute inflammatory response, protein activation cascade, complement activation and regulation of humoral immune response.
Ingenuity Pathway Analysis—The top enriched pathways for the 107 differentially expressed proteins (based on the ANOVA) are shown in Fig. 3. Although 14 pathways were shown to be represented by the differentially expressed proteins, the top 5 pathways in order of significance were: Acute phase response signaling (18.1% overlap); LXR/RXR Activation (19.5% overlap); FXR/RXR Activation (17.4% overlap); Complement system (34.2% overlap) and Coagulation system (34.3% overlap). Fig. 4 shows the top 3 identified networks as Hematological system development and function, Organismal functions, Developmental disorder (Fig. 4A); Neurological disease, Lipid metabolism, Molecular transport (Fig. 4B) and Humoral immune response, Inflammatory disease, Immunological disease (Fig. 4C). Diseases and disorders, as well as physiological system development and functions identified to be of significance are listed in Table II.

Extending the PediOme With a Large Repository Based Spectral Library—As it is notoriously difficult to generate an extensive protein profile of nondepleted plasma by MS because of the presence of many highly abundant proteins we evaluated whether more proteins could be detected by searching the SWATH data against a large human repository based spectral library containing spectra for more than 10,000 human proteins (17). We used SWATHXtend software to merge this library with our local library, ensuring reliable retention time alignments for extraction of SWATH acquired data. With this approach using nondepleted, nonfractionated plasma we extracted quantitative information from 940 proteins (extraction FDR 1%, two or more peptides required) (supplemental Table S3, Fig. 1C), a greater than 6-fold increase compared with the locally generated IDA based spectral library. These proteins are reported to be secreted but are commonly under-reported in proteomic analyses of plasma, so we used ELISA to confirm their presence and concentrations. The expression patterns for five proteins with the largest age-specific changes in expression are shown in Fig. 5. These are: Hemoglobin subunit gamma-1 (HBG1); Hemoglobin subunit gamma-2 (HBG2); collagen alpha-1 I chain (CO1A1); Ig alpha-1 chain C region (IGHA1), and Ig alpha-2 chain C region (IGHA2).

ELISA Validation—ELISA confirmed the expression of periostin and GPLD1 in plasma samples from each age group. GPLD1 expression (Fig. 6A) was greatest in adults with declining abundance with age (40 μg/ml in adults, 18 μg/ml neonates). Periostin showed the inverse trend, but with much lower expression (0.6 μg/ml in neonates, 0.01 μg/ml in adults). It was significantly downregulated (Fig. 6B) in all age groups in comparison to neonates, and significantly downregulated in Adults in comparison to 1–10 years and 1–5-year age groups. This agrees with the SWATH data and demonstrates that an extended human library is a useful strategy for deeper plasma profiling.

We also validated expression of two other proteins, with our findings in agreement with the SWATH data Alpha-2-macroglobulin (A2M) was found to be significantly downregulated (Fig. 6C) in all age groups (<1 year, 1–5 years and Adult) in comparison to neonates. The 1–5-year group was found to be significantly up-regulated in comparison to adults. Histidine rich glycoprotein (HRG) (Fig. 6D) was shown to be up-regulated in the 1–5-year group when compared with neonates, but downregulated in <1 year and Adults.

Pairwise Comparisons of Protein Expression Using an Extended Plasma Library—Pairwise comparisons between specific age groups were also performed. Proteins that were
| Accession | Protein Name          | Uniprot Accession | Mean (1 y.o.) | Mean (5 y.o.) | Mean (adult) | Fold Change | p Value |
|-----------|-----------------------|-------------------|--------------|--------------|--------------|-------------|----------|
| O95445    | APOM_HUMAN            | P01000            | 2.20E+06     | 1.88E+06     | 1.91E+06     | 1.51        | 6.32E-03 |
| P08603    | FIBB_HUMAN            | P00516            | 1.25E+07     | 1.40E+07     | 1.40E+07     | 1.00        | 0.52E-01 |
| P02675    | FIBC_HUMAN            | P00517            | 1.51E+08     | 1.06E+08     | 1.10E+08     | 1.05        | 3.05E-01 |
| P02679    | FIBG_HUMAN            | P00518            | 1.45E+08     | 1.30E+08     | 1.35E+08     | 1.03        | 3.90E-01 |
| P00734    | THRB_HUMAN            | P00519            | 1.33E+07     | 1.35E+07     | 1.40E+07     | 1.06        | 1.50E-01 |
| P06727    | APOA4_HUMAN           | P00520            | 1.05E+07     | 1.05E+07     | 1.10E+07     | 1.05        | 3.50E-01 |
| Q14624    | ITIH4_HUMAN           | P00521            | 1.26E+07     | 1.52E+07     | 1.70E+07     | 1.13        | 1.99E-07 |
| P02751    | FINC_HUMAN            | P00522            | 1.63E+07     | 2.27E+07     | 2.84E+07     | 1.74        | 1.32E-05 |
| P02752    | FINC_HUMAN            | P00523            | 1.63E+07     | 2.27E+07     | 2.84E+07     | 1.74        | 1.32E-05 |
| P04217    | A1BG_HUMAN            | P00524            | 8.36E+06     | 1.52E+06     | 1.77E+06     | 1.30        | 2.79E-03 |
| P03952    | KLKB1_HUMAN           | P00525            | 8.74E+06     | 1.68E+06     | 1.90E+06     | 1.11        | 3.99E-08 |
| P01880    | IGHD_HUMAN            | P00526            | 2.08E+07     | 2.02E+07     | 2.58E+07     | 1.29        | 1.87E-03 |

**Table 1:** ANOVA analysis showing the mean variation in protein expression across the age spectrum (neonates to adults). Results are expressed as mean protein expression (1 y.o. − 5 y.o. − adult) and fold change.
| P07358 | COB, HUMAN | 7.55E-05 | 2.24 | 8.63E-05 | 2.25 | 1.68E-10 | 2.25 | Lumican | Complement component C8 beta chain |
| P01874 | LF/H11001 | 0.6 | 8.94E-06 | 2.25 | 1.04E-06 | 2.25 | 2.42E-05 | 2.25 | Alpha-1-glycoprotein 1 |
| P02763 | LF/H11001 | 1.54E-06 | 2.25 | 1.44E-06 | 2.25 | 4.86E-07 | 2.25 | Ceacam-3 | Leucine-rich 3 |
| P00314 | LF/H11001 | 1.44E-06 | 2.25 | 6.87E-07 | 2.25 | 2.79E-07 | 2.25 | Leukemia, V17 |
| P00450 | LF/H11001 | 0.6 | 6.87E-07 | 2.25 | 2.79E-07 | 2.25 | 2.79E-07 | 2.25 | Leukemia, V17 |
| P02750 | A2GL, HUMAN | 9.15E-06 | 2.39 | 1.19E-05 | 2.39 | 1.15E-05 | 2.39 | Leucine-rich alpha-2-glycoprotein |
| P02763 | LF/H11001 | 1.54E-06 | 2.25 | 1.44E-06 | 2.25 | 4.86E-07 | 2.25 | Ceacam-3 | Leucine-rich 3 |

**Table continued...**
Table I—continued

| Uniprot accession | Protein name       | Mean-n Ye | Mean--<1 y.o. | Mean--1–5 y.o. | Mean--adult | Fold change | ANOVA adjusted p value | Protein mean–neonate | Protein mean–1–5 y.o. | Protein mean–adult | Fold | ANOVA-adjusted p value | Protein mean–neonate | Protein mean–1–5 y.o. | Protein mean–adult |
|-------------------|-------------------|-----------|---------------|----------------|--------------|-------------|------------------------|----------------------|----------------------|---------------------|-------|------------------------|----------------------|----------------------|---------------------|
| P01877            | P02452            | CO1A1     | 1.40E-04      | 1.06E-05       | 6.81E-05     | 8.42        | 5.90E-09              | 7.83                 | 4.90E-03             | 1.44E-05 | 1.44E-06 | 3.51E-11              | 2.09E-07             | 2.14E-02             | 4.72E-06             |
| P01878            | P04438            | HV319     | 1.17E-05      | 1.07E-05       | 1.02E-04     | 1.72        | 8.10                   | 3.35E-03             | 5.90E-03             | 3.35E-03 | 4.84E-05 | 8.01               | 1.04E-05             | 1.06E-05             | 6.49E-05             |
| P01879            | P01591            | IGHJ      | 1.34E-04      | 1.07E-05       | 1.02E-04     | 1.72        | 8.10                   | 3.35E-03             | 5.90E-03             | 3.35E-03 | 4.84E-05 | 8.01               | 1.04E-05             | 1.06E-05             | 6.49E-05             |
| P01861            | P01871            | CD5L      | 1.60E-02      | 1.04E-02       | 1.02E-04     | 1.72        | 8.10                   | 3.35E-03             | 5.90E-03             | 3.35E-03 | 4.84E-05 | 8.01               | 1.04E-05             | 1.06E-05             | 6.49E-05             |
| P08682            | P69891            | HBA       | 1.66E-05      | 1.07E-05       | 1.02E-04     | 1.72        | 8.10                   | 3.35E-03             | 5.90E-03             | 3.35E-03 | 4.84E-05 | 8.01               | 1.04E-05             | 1.06E-05             | 6.49E-05             |
| P08681            | P01876            | IGHA1     | 6.11E-06      | 1.07E-05       | 1.02E-04     | 1.72        | 8.10                   | 3.35E-03             | 5.90E-03             | 3.35E-03 | 4.84E-05 | 8.01               | 1.04E-05             | 1.06E-05             | 6.49E-05             |
| P08680            | P01875            | IGHA2     | 7.27E-05      | 1.07E-05       | 1.02E-04     | 1.72        | 8.10                   | 3.35E-03             | 5.90E-03             | 3.35E-03 | 4.84E-05 | 8.01               | 1.04E-05             | 1.06E-05             | 6.49E-05             |

TABLE 1: The table shows the proteome age-related changes in plasma proteins in human populations. The data is presented as mean expression values across multiple age groups, with statistical analysis using ANOVA. The table highlights proteins that are differentially expressed, withFold change values indicating up-regulation or down-regulation compared to a reference group. Proportions and percentages are not available in the table, but the fold change and p-values provide insights into the significance of these changes. The table includes key proteins such as Ig gamma-4 chain C region, Ig mu heavy chain V-III region JON, and Haptoglobin alpha chain, which have been shown to play critical roles in various biological processes.

DISCUSSION

Although previous studies have assessed the aging of the plasma proteome in pediatric and adult populations (6), this is the first study to have utilized SWATH-MS technology to quantify and analyze variability in differential protein expression in both adult and pediatric age-specific contexts. An added advantage of this methodology was the ability to utilize ingenuity pathway analysis to elucidate enriched biochemical pathways and pathways involved in disease processes. Indeed, ANOVA analysis of the variation in protein expression across multiple age groups shows that proteins involved in hematopoietic development and function, the immune system, and physical growth undergo significant change as the plasma proteome ages.

We identified 107 proteins that varied in abundance over the four age groups examined. The most significant as identified by ANOVA analysis play a role in hematopoietic development, and immune function. Hemoglobin subunit gamma-2 (HBG2) is a protein encoded by the HBG2 gene, and has been shown to play an important role in the development of the hematopoietic system (18). In the developing fetus, HBG1 and HBG2 code for two gamma chain proteins, which along with corresponding alpha chains, make up fetal hemoglobin (HbF). At birth, HbF undergoes replacement by mature adult hemoglobin (HbA), in a process that continues over the first 6 months of age (19). As shown in Table I, the mean expression levels of HBG2 decreased significantly with age as HbA replaced HbF. In a similar manner, collagen alpha-1I chain (CO1A1), a protein that forms part of the alpha-1I chain in Type I collagen also undergoes decreased expression from neonates to adults. As Type I collagen is found in abundance in bone, tendons and the dermis (20), a likely explanation for this is that CO1A1 expression spikes early in life as a result of growth, and is then downregulated in adulthood to a maintain-
nance function. In contrast, Ig alpha-1 chain C region (IGHA1), a protein that plays a role in the formation of IgA antibodies (21), was shown in our analysis to exhibit higher expression as the proteome aged, with largest expressional changes observed in the neonatal to 1–5 age groups. This suggests that rapid development of the immune system occurs in the neonatal to early childhood period, and continues to mature in adolescent years and adulthood.

An advantage to utilizing SWATH-MS was the ability to re-mine the data by using a significantly larger assay library for matching. Here we took advantage of an extensive human assay library containing spectra from over 10,000 proteins and re-extracted quality matching peptides from SWATH data of undepleted and nonfractionated plasma. This increased the detected peptides by greater than 6-fold, enabling more thorough peptidome quantification, representing the largest set to date of SWATH quantified proteins in nondepleted human plasma. ELISA validation of two putatively secreted proteins, periostin and GPLD-1 confirmed their SWATH measured levels in plasma. Liu et al. (10) used SWATH analysis of human plasma and reported the quantitation of 342 proteins in 232 plasma samples collected longitudinally from pairs of twins. They detected GPLD-1, but not periostin and the annotated subcellular localization of these 342 protein was varied. This included nuclear proteins (e.g. Transcription elongation regulator 1(TCERG1), Heterogeneous nuclear ribonucleoproteins), mitochondrial proteins (e.g. GRP75, Trifunctional enzyme subunit alpha (ECH), Adenylate kinase 2) and Golgi proteins (e.g. Golgi membrane protein GP73), among others with non-obvious functional roles in circulation. With these observations there is good reason to utilize the large-scale, proteome wide library approach as we have shown here as this readily increases the depth of plasma proteome coverage using LC-MS without sample immunodepletion. Interestingly, a related approach enabled quantitation of over 1500 proteins in mouse plasma, many of which were of cellular origin (22). Of course, secondary approaches should attempt to validate the detection of cellular proteins observed in plasma. Type 1 error cannot be excluded when utilizing large-scale libraries of this heterogeneous nature, and one strategy is to add stringency by requiring evidence for multiple peptides, in this case at least 2 per protein.

We used pathway knowledge to identify the most enriched pathways, as well as, functions and diseases. Acute phase response signaling, involved in the acute nonspecific immune response, constituted the most enriched canonical pathway. Given the previous evidence that IGHA1 expression increased as the subjects aged, this is further indication that immune development is a dynamic and continuously evolving system. Moreover, identification of IL-6 as

**Fig. 2.** A, Heatmap of the 107 differentially expressed proteins from the local assay library identified using ANOVA. The clustering patterns were obtained using a correlation-based distance and complete linkage. Blue - proteins with decreased expression; Red - proteins with increased expression. B, Clusters and protein abundance trends obtained by hierarchical clustering of the differentially expressed proteins identified using an ANOVA. x-axis represents the four age groups (1, Neonates; 2, <1 year olds; 3, 1–5 year olds; 4, Adults). y-axis represents log Normalized Area of the expression of each relevant protein (gray), with the overall mean presented as a colored line incorporating standard deviation (2SD) bars. The separation in three clusters was based on cutting the dendrogram at a distance of 0.1 and coloring the resulting sample clustering.
the top upstream regulator (a cytokine involved in the inflammatory response, and a mediator of the acute phase response (23)), the inflammatory response and the humoral immune response as the most enriched disease context and physiological function respectively, further supports this contention.

Indeed, quantitation of age-specific variability of plasma proteins has clear and relevant application to the clinic. Although changes in proteomic expression in disease-specific contexts have been well characterized in the literature and mentioned previously in this paper, identification of expression changes in age-specific contexts is a potentially powerful tool for clinical use. For example, blood borne pathologies such as beta-thalassemias and other related conditions often feature prolonged gamma chain HBG2 production, long into early childhood (24). Thus, proteomic

![Fig. 3. Top enriched pathways for the 107 differentially expressed proteins based on the ANOVA analysis, ranked in increasing order of their Fisher exact test p value (Ingenuity pathway analysis, default background comparison). Green - proteins with decreased expression in neonates; Red - proteins with increased expression in neonates; White - no overlap with the data set; Orange - log (p value), Fisher exact test, right tailed.](image)

![Fig. 4. Most significant networks identified using Ingenuity Pathway Analysis. A, Hematological System Development and Function, Organismal Functions, Developmental Disorder; B, Neurological Disease, Lipid Metabolism, Molecular Transport; C, Humoral Immune Response, Inflammatory Response, Immunological Disease. Red denotes up-regulation in neonates compared with adults, green denotes down-regulation in neonates. White denotes a molecule not in the data set. The legend for the shapes is as follows: Double circle, Complex/Group; Trapezium, transporter; Diamond, enzyme; Horizontal diamond, Peptidase; Circle, other.](image)
analysis of patient plasma, in reference to age-specific protein reference ranges, may assist clinicians in the diagnostic phase, or help to elucidate the severity of the relevant pathology.

In this study we have shown that the plasma proteome undergoes significant change as the system ages. Apart from identifying proteins with various functional roles in systems such as hematopoietic development, the immune response and growth, we also identified the most highly enriched pathways, upstream regulators, disease contexts and functions in comparison to the default background. Future studies should be directed at further elucidating age-specific variability in protein expression and, importantly, its links to disease processes.

**TABLE II**

*Ingenuity Pathway Analysis of diseases and disorders and development and function based on the comparison of the 108 differentially expressed proteins to the default background*

| Diseases and Disorders                           | p value          | # Molecules |
|-------------------------------------------------|------------------|------------|
| Inflammatory response                           | 1.01E-04-5.02E-23| 58         |
| Metabolic disease                               | 1.20E-04-4.73E-22| 46         |
| Immunological disease                           | 9.85E-05-6.15E-21| 39         |
| Neurological disease                            | 1.23E-04-3.86E-19| 46         |
| Psychological disorders                         | 2.73E-08-3.86E-19| 3          |
| Cell-to-cell signaling and interaction          | 1.49E-04-1.38E-20| 52         |
| Cellular movement                               | 1.39E-04-3.29E-18| 41         |
| Cell death and survival                         | 1.15E-04-5.86E-12| 49         |
| Cellular development                            | 6.94E-05-2.94E-11| 27         |
| Lipid metabolism                                | 1.49E-04-1.55E-10| 36         |

**Physiological system development and function**

|                                                               | p value          | # Molecules |
|---------------------------------------------------------------|------------------|------------|
| Humoral immune response                                       | 8.95E-05-5.02E-23| 20         |
| Hematological system development and function                | 1.04E-04-8.00E-20| 51         |
| Immune cell trafficking                                       | 1.04E-04-3.28E-19| 43         |
| Organismal functions                                          | 7.21E-05-4.35E-17| 22         |
| Tissue morphology                                             | 9.53E-05-6.84E-13| 4          |

**FIG. 5.** *Age-specific expression patterns of five proteins of interest selected for discussion.* Expression patterns for five proteins with the largest age-specific changes in expression. A, Hemoglobin subunit gamma-1 (HBG1); B, Hemoglobin subunit gamma-2 (HBG2); C, collagen alpha-1I chain (CO1A1); D, Ig alpha-1 chain C region (IGHA1) and E, Ig alpha-2 chain C region (IGHA2).
DATA AVAILABILITY

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner repository with the data set identifier PXD005661.

Reviewer account details: Username: reviewer18033@ebi.ac.uk, Password: auk6Amra.

Acknowledgments—We thank members of the anesthetic and surgical departments of the Royal Children’s Hospital for facilitation of sample collection.

* This study was supported by the Victorian Government’s Operational Infrastructure Support Program. This research has been supported in part by Bioplatforms Australia Ltd and facilitated by access to Australian Proteome Analysis Facility supported by the Australian Government’s National Collaborative Research Infrastructure Scheme.

□ This article contains supplemental material.

To whom correspondence should be addressed: Hematology Research Laboratory, Murdoch Childrens Research Institute, Department of Pediatrics, The University of Melbourne, Tel.: +61 3 9936 6520; Fax: +61 3 9348 1391; E-mail: vera.ignjatovic@mcri.edu.au.

REFERENCES

1. Lull, M. E., Freeman, W. M., Myers, J. L., Midgley, F., Kimatian, S. J., Undar, A., and Vrana, K. E. (2006) Plasma proteomics: a noninvasive window on pathology and pediatric cardiac surgery. ASAIO J. 52, 562–566
2. Michiels, J. J., Brenneman, Z., Gaddis, A., van der Planken, M., Schroyens, W., van de Velde, A., and van Vliet, H. (2006) Classification and characterization of hereditary types 2A, 2B, 2C, 2D, 2E, 2M, 2N and 2U (unclassifiable) von Willebrand disease. Clin. Appl. Thromb. Hemost. 12, 397–420
3. Kakhniashvili, D. G., Griko, N. B., Bulla, L. A., Jr., and Goodman, S. R. (2005) The proteomics of sickle cell disease: profiling of erythrocyte membrane proteins by 2D-DIGE and tandem mass spectrometry. Exp. Biol. Med. 230, 787–792
4. Yocum, A. K., Busch, C. M., and Felix, C. A. (2006) Proteomics-based strategy to identify biomarkers and pharmacological targets in leukemias with t(4;11) translocations. J. Proteome Res. 5, 2743–2753
5. Roxo-Rosa, M., da Costa, G., and Luider, T. M. (2006) Proteomic analysis of nasal cells from cystic fibrosis patients and noncystic fibrosis control individuals: search for novel biomarkers of cystic fibrosis lung disease. Proteomics 6, 2314–2325
6. Ignjatovic, V., Lai, C., Summerhayes, R., Mathesius, U., Tawfikis, S., Perrughi, M. A., and Monagle, P. (2011) Age-related differences in plasma proteins: how plasma proteins change from neonates to adults. PLoS ONE 6, e17213
7. Omenn, G. S., ed HUPO. (2006) Exploring the Human Plasma Proteome, pp. 1–33, Wiley-VCH Verlag GmbH&Co., Weinheim.

8. Gillet, L. C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol. Cell. Proteomics 11, 1535–9484

9. Liu, Y., Huttenhain, R., Surinova, S., Gillet, L. C., Mouritsen, J., Brunner, R., Navarro, P., and Aebersold, R. (2013) Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. Proteomics 13, 1247–1256

10. Liu, Y., Buil, A., Collins, B. C., Gillet, L. C., Blum, L. C., Cheng, L. Y., Vitek, O., Mouritsen, J., Lachance, G., Spector, T. D., Dermitzakis, E. T., and Aebersold, R. (2015) Quantitative variability of 342 plasma proteins in a human twin population. Mol. Syst. Biol. 11, 786

11. Monagle, P., Barnes, C., Ignjatovic, V., Furmedge, J., Newall, F., Chan, A., De Rosa, L., Hamilton, S., Ragg, P., Robinson, S., Auldast, A., Crock, C., Roy, N., and Rowlands, S. (2006) Developmental haemostasis: Impact for clinical haemostasis laboratories. Thromb. Haemost. 95, 362–372

12. Ignjatovic, V., Straka, E., Summerhayes, R., and Monagle, P. (2010) Age-specific differences in binding of heparin to plasma proteins. J. Thromb. Haemost. 6, 1290–1294

13. Ignjatovic, V., Summerhayes, R., Newall, F., and Monagle, P. (2009) The in vitro response to low-molecular-weight heparin is not age-dependent in children. Thromb. Haemost. 103, 855–856

14. Attard, C., van der Straaten, T., Kariatlis, V., Monagle, P., and Ignjatovic, V. (2013) Developmental haemostasis: age-specific differences in the quantity of haemostatic proteins. J. Thromb. Haemost. 11, 1850–1854

15. Attard, C., Kubitz, D., Monagle, P., and Ignjatovic, V. (2014) The in vitro anticoagulant effect of Rivaroxaban in Neonates. Blood Coagulation Fibrinolysis 25, 237–240

16. Wu, J. X., Song, X., Pascoveci, D., Zaw, T., Care, N., Krisp, C., and Molloy, M. P. (2016) SWATH Mass Spectrometry Performance Using Extended Peptide MS/MS Assay Libraries. Mol. Cell. Proteomics 15, 2501–2514

17. Rosenberger, G., Koh, C. C., Guo, T., Rost, H. L., Kouvonen, P., Collins, B. C., Heusel, M., Liu, Y., Caron, E., Vichalkovski, A., Faini, M., Schubert, O. T., Faridi, P., Ebhardt, H. A., Matondo, M., Lam, H., Bader, S. L., Campbell, D. S., Deutsch, E. W., Moritz, R. L., Tate, S., and Aebersold, R. (2014) A repository of assays to quantify 10,000 human proteins by SWATH-MS. Sci. Data 1, 140031

18. Higgs, D. R., Vickers, M. A., Wilkie, A. O., Pretorius, I. M., Jarman, A. P., and Weatherall, D. J. (1989) A review of the molecular genetics of the human alpha-globin gene cluster. Blood 73, 1081–104

19. Pereira, C., Relvas, L., Bento, C., Abade, A., Ribeiro, M. L., and Manco, L. (2015) Polymorphic variations influencing fetal hemoglobin levels: association study in beta-thalassemia carriers and in normal individuals of Portuguese origin. Blood Cells Mol. Dis. 54, 315–320

20. Mann, V., and Raiston, S. H. (2003) Meta-analysis of COL1A1 Sp1 polymorphism in relation to bone mineral density and osteoporotic fracture. Bone 32, 711–717

21. Takahashi, K., Raska, M., Stuchlova Horynova, M., Hall, S. D., Poulsen, K., Kilian, M., Hiki, Y., Yuzawa, Y., Moldoveanu, Z., Julian, B. A., Renfrow, M. B., and Novak, J. (2014) Enzymatic sialylation of IgA1 O-glycans: implications for studies of IgA nephropathy. PloS One 9, e99026

22. Malmström, E., Kilsgård, O., Hauri, S., Smeds, E., Herwald, H., Malmström, L., Malmström, J. (2016) Large-scale inference of protein tissue origin in gram-positive sepsis plasma using quantitative targeted proteomics. Nat. Commun. 6, 10261

23. Heinrich, P. C., Castell, J. V., and Andus, T. (1990) Interleukin-6 and the acute phase response. Biochem. J. 265, 621–636

24. Neishabury M, Zamani S, Azarkeivan A, Abedini SS, Darvish H, Zamani F, and Najmabadi H. (2012) The modifying effect of Xmn1-HBG2 on thalassemic phenotype is associated with its linked elements in the beta globin locus control region, including the palindromic site at 5HS4. Blood Cells Mol Dis. 48, 1–5.