Impact of Hemolysis on Multi-OMIC Pancreatic Biomarker Discovery: Derisking Precision Medicine Biomarker Development

Richard Searfoss
 BERG LLC

Punit Shah
 BERG LLC

Kennedy Ofori-Mensa
 BERG LLC

Valerie Bussberg
 BERG LLC

Vladimir Tolstikov
 BERG LLC

Bennett Greenwood
 BERG LLC

Hongyan Li
 BERG LLC

Corinne DeCicco
 Beth Israel Deaconess Medical Center

Elder Granger
 BERG LLC

Leonardo O. Rodrigues
 BERG LLC

Eric Grund
 BERG LLC

A. James Moser
 Beth Israel Deaconess Medical Center

Rangaprasad Sarangarajan
 BERG LLC

Michael A. Kiebish (✉ michael.kiebish@berghealth.com)
 BERG LLC   https://orcid.org/0000-0002-5678-0072

Niven R. Narain
 BERG LLC

Research

Abstract

Cancer biomarker discovery is critically dependent on the integrity of biofluid, and tissue samples acquired from study participants. Multi-omic profiling of candidate protein, lipid, and metabolite biomarkers is confounded by timing and fasting status of sample collection, participant demographics and treatment exposures of the study population. Contamination by hemoglobin, whether caused by hemolysis during sample preparation or underlying red cell fragility, contributes 0 – 10 g/L of extraneous protein to plasma, serum, and Buffy coat samples and may interfere with biomarker detection and validation. We analyzed 617 plasma, 701 serum, and 657 buffy coat samples from a 7 year longitudinal multi-omic biomarker discovery program evaluating 400+ participants with or at risk for pancreatic cancer, known as Project Survival™. Hemolysis was undetectable in 93.1% of plasma and 95.0% of serum samples, whereas only 37.1% of Buffy coat samples were free of contamination by hemoglobin. Regression analysis of multi-omic data demonstrated a statistically significant correlation between hemoglobin concentration and the resulting pattern of analyte detection and concentration. Although hemolysis had the greatest impact on identification and quantitation of the proteome, distinct differentials in metabolomics and lipidomics were also observed and correlated with severity. We conclude that quality control is vital to accurate detection of informative molecular differentials using OMIC technologies and that caution must be exercised to minimize the impact of hemolysis as a factor driving false discovery in large cancer biomarker studies.

Introduction

The current healthcare ecosystem is rapidly evolving toward deploying precision medicine strategies for increasing optimal stratification of patients to improve clinical outcomes. These actions will predominantly focus on the use of molecular, digital, and clinical biomarkers that will characterize patients on multiple dimensions of phenotypic presentation. Standardization of quality parameters governing sample collection are important to ensure accuracy and reproducibility of potential discoveries ultimately easing translation back into the clinic. Molecular markers, whether genetic, proteomic, lipidomic or metabolomic, hold tremendous promise to deconvolute the biological presentation of patients. The composition of adaptive biological molecules (proteins, lipids, and metabolites) can be significantly influenced by patient demographics, pharmacological agents, and sample handling processes which can hinder potential biomarker discovery and development.

Hemolysis represents a common sample processing outcome and can be due to handling, but also disease etiology rendering red blood cells (RBC) more labile for lysis. Hemolysis can occur for a variety of reasons and leads to the release of free hemoglobin into blood collection samples [1]. Due to some medical conditions, or as the result of taking certain medications, this breakdown of RBC's can be increased. Hemolysis has the potential to drastically alter the observed proteome of buffy coat samples due to contamination of hemoglobin and other high-abundance proteins seen in RBC. RBC are mainly comprised of hemoglobin and carbonic anhydrase-1, contributing 97% and 1% of the entire RBC proteome, respectively[2]. The buffy coat fraction of whole blood has been observed to be less than 1% of the blood by volume[3]. As a result, even minor contamination of RBC into the other fractions, or increased hemolysis due to medical reasons, can increase the concentration of hemoglobin and carbonic anhydrase-1 and potentially impact the observed proteome.

Guidelines governing omics analysis of clinical samples have been developed over the past decade as the use of such platforms has been broadly adopted in R&D and clinical trial assessments [4, 5]. This includes

standardized sample preparation approaches and techniques, quality controls, and the recommended size of cohorts required to ensure statistical significance of potential findings. However, protocols for quality controls regarding sample collection are deficient. Several key challenges have already been demonstrated in using biofluids for biomarker discovery, such as chemical modifications of proteins or sample degradation during storage. Further, utilizing plasma and serum, which is often employed for convenience of collection, exhibits a wide dynamic range of protein concentrations, making the identification of low abundance potential biomarkers all the more challenging. One potentially impactful occurrence that should be included is the effect of hemolysis, which can directly contribute to both aforementioned challenges.

Herein, we performed mass spectrometry based lipidomics, metabolomics and proteomics analysis of plasma and serum from over 420 individuals in pancreatic biomarker clinical trial. Buffy coat samples were only subjected to proteomics analysis based on the sample amount obtained. A subset of samples obtained were impacted by hemolysis resulting in contamination of the matrix of interest. A comprehensive assessment of expressional patterns of proteins, lipids and metabolites was performed to identify hemolytic contamination in these samples. The proteome of buffy coat was most impacted, resulting in expressional changes of proteins originating from red blood cells. The use of markers impacted by hemolysis should considered with caution for exploration as biomarkers.

Methods

Study Design

There were 420 patients enrolled in this study: 224 males and 196 females. These fell into one of the five categories as follows; healthy volunteers: 33, patients with pancreatitis: 113, early pancreatic cancer: 67, local pancreatic cancer: 115 and metastatic pancreatic cancer: 92. All volunteers participating in this clinical study NCT02781012 gave their informed consent for inclusion before they participated in the study. Research use of the samples was conducted in accordance with the terms outlined within the informed consent form and the terms set forth therein and with the tenets of the Declaration of Helsinki and its later amendments or comparable ethical standards.

Sample Collection

Whole blood samples were collected via venipuncture into EDTA tubes. All samples were processed and frozen at -80ºC within 3 hours of the blood draw. The plasma fraction was separated using centrifugation at 1200 x g for 10 minutes at room temperature and was aliquoted into separate tubes and frozen. During centrifugation, the buffy coat layer also separated from the red blood cells. The buffy coat layer was collected and diluted with 8mL RPMI buffer, transferred into a 50mL Leucosep tube, and centrifuged at 1200 x g for 10 minutes at room temperature to separate the buffy coat layer further from the red blood cells. Buffy coat was washed three times with PBS and pelleted to remove solution. Finally, the buffy coat was resuspended in 200µL of PBS and split between two tubes before being frozen at -80ºC. A separate vial of blood was collected for serum sample collection in serum separator tubes and was left at room temperature for 30–45 minutes to allow for the clot to form. Serum separator tubes were then centrifuged at 1200 x g for 10 minutes at room temperature. Separated serum was aliquoted and frozen at -80ºC.

Detection Of Hemolysis

Upon receipt, all samples were accessioned and qualitatively assigned a colorimetric hemolysis score of 1−3 for plasma and serum and 0−4 for buffy coat following the color scale in Fig. 1[6]. A score of zero was reserved for buffy coat samples appearing clear to opaque white when buffy coat cells were most pure. Given the natural yellowish appearance of plasma and serum, a score of zero was never given, and a score of 1 was considered most pure.

Proteomics

Protein Extraction

65µL of raw plasma/serum was filtered through a pre-wet 0.22µm cellulose acetate spin filter. 40µL of the filtered plasma/serum was pipetted onto another pre-wet 0.22µm cellulose acetate spin filter and combined with 20µL of 80mg/mL lipid removal agent (LRA). The mixture was placed on a shaker for 30 minutes and then centrifuged. The resulting filtrate was roughly 40µL in volume and was combined with 120µL of Agilent Buffer A. The sample was then loaded into vials and placed on the Agilent 1260 series HPLC, and the top 14 abundant proteins were depleted using the Multi-Affinity Removal Column 14 from Agilent. The depleted samples were collected into vials and protein concentration was determined using the Bradford Assay.

Buffy coat samples were lysed with a lysis buffer containing 5M Urea, 50mM Tris-HCl pH 8.3, 0.1% SDS, 1% Protease and Phosphatase Inhibitor Cocktail, and Optima LC/MS Water. 100µL of lysis buffer was added to each sample and mixed by pipetting up and down, and then the whole sample was immediately transferred out of the sample vial and into a 1.5mL Eppendorf tube. Each sample was sonicated with four 3-second pulses at 20% amplification to fully lyse the cells. Sonicated samples were centrifuged at 17,000 x g for 10 minutes, and the supernatant was then used in the Bradford Assay to determine the protein concentration.

Trypsin Digestion

Extracted proteins were trypsin digested as previously described[7]. In brief, proteins were reduced with 10mM Tris(2-carboxyethyl) Phosphine (TCEP) and alkylated with 18.75mM iodoacetamide before being precipitated in acetone overnight and digested with trypsin the next day.

TMT Labeling of Peptides

10-plex TMT (Tandem Mass Tag) reagents were used to label peptides from all samples, allowing relative quantitation (Thermo Fisher Scientific). For buffy coat samples, a reference sample was created by pooling an aliquot of peptides from each individual sample. For plasma and serum, a reference sample was purchased from BioIVT and prepped alongside clinical samples (Westbury, CT). Peptides (20 µg) from each of the samples were dissolved in 20 µL of 200 mM triethylammonium bicarbonate (TEAB), pH 8.5 solution, and mixed with 20 µL of TMT reagent that was freshly dissolved in 256 µL of anhydrous acetonitrile, LCMS grade. Channel 126 was used for labeling the pooled reference sample in all matrices analyzed in this study. After 1h incubation at RT, the reaction was quenched by adding 8 µL 5% hydroxylamine. Peptides labeled by different TMT reagents were then mixed, dried and desalted on C18 Spin columns. Desalted peptides were dried in a vacuum centrifuge and stored at -20°C until LC-MS/MS analysis.

Mass Spectrometry

LC-MS/MS analysis was performed using a Waters nanoAcquity 2D LC system coupled to a Thermo Q Exactive Plus MS. TMT-labeled MPs were resolved over 12 fractions, 90-minute gradient per fraction, and fractionated using two-dimensional reversed-phase chromatography prior to MS analysis. In the first dimension, Buffer A is 20mM Ammonium formate pH 9.5 and Buffer B is 100 % Acetonitrile. Peptides are separated on a Basic reverse phase compatible column BEH C18 trap column, 300µM x 50mm, particle size 5µM (waters) into 12 Fractions with %B 7.4, 10.8, 12.6, 14, 15.3, 16.7, 18.3, 20.4, 23.5, 50, 65 and wash. In the second dimension, peptides were separated in a 90min gradient from 5%B to 85%B.

The eluting peptides were sprayed into the mass spectrometer using electrospray ionization and a data dependent Top 15 acquisition method was used to fragment candidate ions. Full MS survey scans were collected at a resolution of 35,000, scan range of 400–1800 Thompsons (Th; Th = Da/z), followed by MS/MS scans at a resolution of 35,000 with a 1.2 Th isolation window. Only ions with a + 2 to + 4 charge were considered for isolation and fragmentation. Data was searched using Proteome Discoverer 1.4 using SEQUEST and Mascot algorithms and uniprot database.

Metabolomics

Metabolomics Analysis.

Plasma and serum samples for metabolomics analysis were prepared as previously described[8–12]. Metabolite extraction was achieved using a mixture of isopropanol, acetonitrile, and water at a ratio of 3:3:2 v/v. Extracts were divided in to three parts: 75 µL for gas chromatography combined with time-of-flight high-resolution mass spectrometry, 150 µL for reversed-phase liquid chromatography coupled with high-resolution mass spectrometry, and 150 µL for hydrophilic interaction chromatography with liquid chromatography and tandem mass-spectrometry, and analyzed as previously described [8–12]. We used the NEXERA XR UPLC system (Shimadzu, Columbia, MD, USA), coupled with the Triple Quad 5500 System (AB Sciex, Framingham, MA, USA) to perform hydrophilic interaction liquid chromatography analysis, NEXERA XR UPLC system (Shimadzu, Columbia, MD, USA), coupled with the Triple TOF 6500 System (AB Sciex, Framingham, MA, USA) to perform reversed-phase liquid chromatography analysis, and Agilent 7890B gas chromatograph (Agilent, Palo Alto, CA, USA) interfaced to a Time-of-Flight Pegasus HT Mass Spectrometer (Leco, St. Joseph, MI, USA). The GC system was fitted with a Gerstel temperature-programmed injector, cooled injection system (model CIS 4). An automated liner exchange (ALEX) (Gerstel, Muhlheim an der Ruhr, Germany) was used to eliminate cross-contamination from the sample matrix that was occurring between sample runs. Quality control was performed using metabolite standards mixture and pooled samples, applying the methodology previously described [13–16]. A quality control sample containing a standard mixture of amino and organic acids purchased from Sigma-Aldrich as certified reference material, was injected daily to perform an analytical system suitability test and to monitor recorded signals day to day reproducibility as previously described[8–12]. A pooled quality control sample was obtained by taking an aliquot of the same volume of all samples from the study and injected daily with a batch of analyzed samples to determine the optimal dilution of the batch samples and validate metabolite identification and peak integration. Collected raw data was manually inspected, merged, imputed and normalized by the sample median. Metabolite identification was performed using in house authentic standards analysis. Metabolite annotation was used utilizing recorded retention time and retention indexes,

recorded MS^n and $HRAMS^n$ data matching with METLIN, NIST MS, Wiley Registry of Mass Spectral Data, HMDB, MassBank of North America, MassBank Europe, Golm Metabolome Database, SCIEX Accurate Mass Metabolite Spectral Library, MzCloud, and IDEOM databases.

Lipidomics

Structural Lipidomic Analysis

A cocktail of deuterium-labeled and odd chain phospholipid standards from diverse lipid classes was added to 25 μL of thawed serum or plasma. Standards were chosen to represent each lipid class and were prepared at concentrations proportional to the endogenous amount per sample matrix to provide the most accurate quantitation and dynamic range for each lipid species. 4 mL chloroform:methanol (1:1, v/v) was added to each sample and the lipid extraction was performed as described. Lipid extraction was automated using a customized sequence on a Hamilton Robotics STARlet system (Hamilton, Reno, NV) to meet the high-throughput requirements. Lipid extracts were dried under nitrogen and reconstituted in 68 μL chloroform:methanol (1:1, v/v). Samples were flushed with nitrogen and stored at -20°C. Samples were diluted 50 fold in isopropanol:methanol:acetonitrile:water (3:3:3:1, by volume) with 2 mM ammonium acetate in order to optimize ionization efficiency in positive and negative modes. Electrospray ionization-MS was performed on a TripleTOF® 5600⁺ (SCIEX, Framingham, MA), coupled to a customized direct injection loop on an Ekspert microLC200 system (SCIEX) as described[17, 18].

Mediator Lipidomic Analysis

A mixture of deuterium-labeled internal standards was added to aliquots of 100 μL serum or plasma, followed by 3x volume of sample of cold methanol (MeOH). Samples were vortexed for 5 minutes and stored at – 20°C overnight. Cold samples were centrifuged at 14,000g at 4°C for 10 minutes, and the supernatant was then transferred to a new tube and 3 mL of acidified H_2O (pH 3.5) was added to each sample prior to C18 SPE columns (Thermo Pierce) and performed as described[19]. The methyl formate fractions were collected, dried under nitrogen, and reconstituted in 50 μL MeOH:H_2O (1:1, v/v). Samples were transferred to 0.5 mL tubes and centrifuged at 20,000 g at 4°C for 10 minutes. Thirty-five μL of supernatant were transferred to LC–MS vials for analysis using the BERG LC–MS/MS mediator lipidomics platform as described.

Data Analysis

Proteins that had missing values in more than 85% of samples were considered unreliable, and therefore removed from further analysis. Data was normalized according to a median centering and variance scaling approach applied across samples [20, 21]. Batches due to study cohort were corrected using an empirical Bayesian framework, ComBat [22, 23]. Briefly, this method performed location and scale adjustments based on estimated batch effect parameters per protein and returned a corrected dataset for further analysis. The data was then used for identifying differential expression between different hemolysis score in plasma, serum and buffy coat. Missingness was calculated as the proportions of missing proteins in each sample.

Results

Workflow, Design and Summary

To evaluate the impact of hemolysis on biomarker discovery utilizing a multi-omics platform, we compared proteins, lipids, and metabolites identified across plasma, serum, and buffy coat samples (proteomics only) acquired from 420 non-diseased and pancreatic cancer patients. Workflow of Proteomics, Lipidomic and Metabolomic analysis is shown in Fig. 1. A hemolysis score was recorded for each sample, ranging from 0−4 for buffy coat and 1−3 for plasma and serum. A summary of the distribution of hemolysis scores within each sample type can be found in Fig. 2. Buffy coat yielded the largest hemolyzed samples 37.1% - #0, 25.1% - #1, 24.8% -#2, 12.4% - #3, 0.4% - #4 hemolysis. Protocol of isolation of buffy coat from blood may be one of the major reasons for the large number of contaminated buffy coat samples.

In proteomics, 7302, 1971, 2146 proteins were identified and quantified in buffy coat, serum and plasma, respectively, using TMT labeling and 2D online LC-MS/MS. After filtering the data for proteins that have less than 85% missing values, a total of 3648, 453, and 492 proteins in buffy coat, serum and plasma, respectively, were obtained and used for further analysis. In lipidomics, 1318 structural lipids and 106 mediator lipids were identified and quantified after data filtration for analysis in plasma and serum samples. In metabolomics, a total of 514 and 508 metabolites were identified and quantified in plasma and serum samples, respectively, after data filtering and kept for further analysis.

Differentially Expressed Metabolites And Lipids

Lipidomics analysis revealed no significant changes in lipid expression for mediator lipidomics data when comparing samples with hemolysis scores of 2 + to 1 in both plasma and serum. However, for structural lipidomics analysis, 5 lipids were found to be down regulated, and 2 lipids up regulated in plasma, and 14 lipids were down regulated, and 11 lipids up regulated in serum (Table 1). More profound effects were seen in the metabolomics data. When comparing samples with hemolysis scores of 2 + to 1, a total of 51 metabolites were found to be down regulated and a total of 25 upregulated due to hemolysis in plasma (Table 1). For the same comparison in serum, 93 metabolites were down regulated and 21 were upregulated due to hemolysis (Table 1). A summary of these results can be found in supplemental table 1.

Table 1

	Proteomics			Signaling Lipidomics		Structural Lipidomics		Metabolomics	
Matrix	Buffy Coat	Plasma	Serum	Plasma	Serum	Plasma	Serum	Plasma	Serum
Downregulated Species	250	2	2	0	0	5	14	51	93
Upregulated Species	208	22	13	0	1	2	11	25	21
* Differentially expressed species due to hemolysis									

Missingness

A subset of samples with the lowest hemolysis score was created, in this case, a score of 0 for buffy coat samples and a score of 1 for plasma and serum samples. This subset was used to filter the proteins, and only the proteins that have less than 85% missing values were kept in the full proteomics data. The missing proportions of proteins for each sample were computed, and samples were then grouped by hemolysis score of 0: 244 samples, score 1: 165 samples, score 2: 163 samples, score 3+: 85 samples in buffy coat (Fig. 2). The boxplots clearly indicate that as the hemolysis score of a sample increases, the number of proteins that are identified across the set within the sample decreases, and the medians of proportions of missing proteins are 0.299, 0.353, 0.406, 0.410 for the groups with hemolysis score 0, 1, 2, 3+, respectively (Fig. 3). This can be explained by an increase in the signal derived from the more abundant hemoglobin proteins contributed from the lysed red blood cells, suppressing the signal of the less abundant proteins and changing the dynamic range of the protein content that would ideally be identified from samples with little to no hemolytic contamination.

Differential Expressed Proteins

To assess the effect of hemolysis on relative protein expression in buffy coat, comparisons between hemolysis groups were performed as shown by volcano plots (Fig. 4A). Overall, 657 samples were included in this analysis. A total of 3,647 proteins were identified when assessing the differentially expressed proteins between samples with a score of 0 vs. 1 (Fig. 4A), with 394 differently expressed proteins down-regulated and 310 proteins up-regulated at a 1.3 fold change threshold and a p-value of 0.05. Comparing samples with a score of 0 vs. 2 Fig. 4B, a total of 701 proteins were consistently identified across all samples, with 348 proteins differently expressed proteins down-regulated and 251 proteins upregulated at a 1.3 fold-change threshold and a p-value of 0.05 (Supplemental Table 1). Lastly, we compared samples with a score of 0 vs. 3 + Fig. 4C, and a total of 592 proteins were consistently identified across all samples, with 238 proteins differently expressed proteins down-regulated and 187 proteins up-regulated at a 1.3 fold-change threshold and a p-value of 0.05. Hemolysis not only impacted the proteins identified but also impacted the quantitation of the differentially expressed proteins.

Further, comparisons between samples with no visual hemolysis (scores of 0 for buffy coat, scores of 1 for plasma and serum) were made to samples with visual hemolysis (scores of 1 – 3 + for buffy coat, scores of 2 + for plasma and serum). Differential expression of proteins was observed using volcano plots shown in Fig. 4, using a threshold of 1.3 fold change with a corresponding p-value of 0.05 to be considered differentially expressed. Overall, 250 proteins were found to be downregulated and 208 upregulated in buffy coat. Similarly, in plasma and serum, 2 proteins were found to be down regulated in samples scored 2 + compared to samples scored 1. A total of 22 proteins in plasma and 13 proteins in serum were found to be up regulated in the same comparison. A summary of these results can be found in supplemental table 1.

Impact Of Hemolysis On Hemoglobin

To study hemolysis via protein identification and relative quantitation, we assessed the expression of Hemoglobin Subunit Alpha (HBA1), Hemoglobin Subunit Beta (HBB), and Hemoglobin Subunit Delta (HBD) across all sample types and grouped by hemolysis score within each sample type. Hemolysis is generally classified as the lysis of RBC in circulation or during sample preparation, and as hemoglobin is one of the most

abundant proteins in red blood cells, the hemoglobin expression increases due to hemolysis (Fig. 5A) and increased stepwise with increasing hemolysis score. A similar pattern was seen in both plasma and serum, with lower levels observed in samples with a hemolysis score of 1, and significantly higher levels observed in samples scored 2+ (Fig. 5b and Fig. 5C).

We also assessed the expression of carbonic anhydrase (CA1), histone H2B type 1-L (HIST1H2BL), and ubinuclein-2 (UBN2) (Fig. 6). CA1 is another major protein found in RBC's and is responsible for processing carbon dioxide in the body. The expression of CA1 is low in samples classified with a hemolysis score of 0, and increases similar to the hemoglobin protein expression with increasing hemolysis score (Fig. 6). HIST1H2BL and UB2 are both nuclear proteins whose identification is expected in buffy coat samples and not from red blood cells. HIST1H2BL and UBN2 expression follow the expected result, with higher expression in samples with hemolysis score of 0 and lower expression with increasing hemolysis score (Fig. 6), indicating signal suppression of these proteins as a result of hemolysis.

Discussion

Translation of biomarkers into clinical practice requires the comprehensive understanding of the impact of sample handling to avoid false discovery of processing markers rather than disease associated biomarkers. Adaptive omic technologies such as proteomics, lipidomics, and metabolomics demonstrate tremendous promise associating the patient phenotypic with causal biology but are also significantly impacted by red blood cell contamination in plasma, serum, or buffy coat. In the current study, we uncovered that buffy coat was the most significantly affected by hemolysis in a prospective biomarker study investigating pancreatic cancer and at-risk populations. The incidence of hemolysis was independent of disease conditions (data not shown) but did influence detection and quantification of analytes.

Contamination by proteins found in RBC from hemolysis has also been demonstrated in red blood cell storage in an occurrence known as storage lesions. Storage lesions are progressive changes in the morphology, biochemistry, and function of RBC during storage that result in changes in the viability of the RBC and accumulation of contaminating proteins and cells. These changes in RBC ultimately lead to hemolysis, and consequently, a release of the cytosolic contents into solution[1, 24]. A study observing changes in the protein distribution of RBC supernatant over a storage period identified appreciable increases in proteins, including carbonic anhydrase 1 and 2 (CA1 and CA2), peroxiredoxin-1 and – 2 (PRDX1 and PRDX2), and catalase, as well as others, due to hemolysis of RBC over time in these storage lesions[25]. Similarly, our findings also conclude these proteins to be contaminants in plasma, serum, and buffy coat due to hemolysis that may occur in vivo or during sample processing.

The identification of proteins in a sample depends on the dynamic range of the proteins. Identifying less abundant proteins in a sample via LC-MS/MS analysis is challenging at low concentrations as current mass spectrometry capabilities allow for identification over a range of 3–4 orders of magnitude[26]. Hemolysis increases the hemoglobin content in the sample of interest. Given that hemoglobin accounts for 97% of the composition of RBC's, with carbonic anhydrase accounting for another 1%, this can create significant suppression of signal of low abundant proteins in the biofluid of choice for a proteomic study[27]. In proteomics, sample quantitation is performed using equal volume of fluid or equal concentration of protein content. In this study the equal concentration of proteins was used for semi-quantitation, supplemented by

Tandem Mass Tags for protein quantitation. The general hypothesis is that the samples are identical with minor changes. Quantitation of proteins is impacted due to hemolysis which leads to an increase in concentration of red blood cell proteins. As contamination increases, the proportion of proteins of interest in the sample decreases and can lead to inaccurate quantitation and false discovery of the biomarkers. Hemolyzed samples should be avoided in omics studies to minimize data analysis variability and data interpretation errors. The use of differentially expressed species (supplemental table 1) as biomarkers of disease in any study should be viewed with caution due to hemolysis. For instance, carbonic anhydrase-1 has been demonstrated as a biomarker in serum for prostate cancer[28]. Further, peroxiredoxin-2 was identified as a biomarker in a panel of proteins from plasma for Anderson-Fabry disease[29]. While this may in fact be the case, careful consideration should be taken into sample quality while testing to avoid false positives, and analysis should be performed to conclude these proteins had little to no contribution to their signal from sample handling issues or hemolysis.

In clinical settings, omics analysis on serum, plasma or buffy coat samples requires caution while handling samples to avoid hemolysis. Following a set protocol is required when collecting and handling samples and any deviation in sample handling needs to be recorded. In some cases, even after all sample handling precautions have been taken, hemolysis may still occur due to underlying biological factors. In these scenarios, various methods can be used for data analysis to minimize the impact of contamination of proteins like hemoglobin. One approach is to ignore any contributing red blood cell proteins as a biomarker, if considered as contamination. A second approach is to use proteins such as hemoglobin or carbonic anhydrase to normalize the data and specifically normalize only non-red blood cell containing proteins. This can minimize the impact due to hemolysis in quantitation. However, any attempt that might minimize this effect may not completely negate the impact due to hemolysis. A third approach is to move towards equal volume quantitation compared to equal concentration quantitation, however, this might require technical advancements in instrumentation and technology. Identification of contaminating proteins cannot be avoided, and expression of those protein rise with the increase in the hemolysis score. Sophisticated LC-MS/MS technology, biochemical procedures for sample preparation and advance bioinformatics tools need to be used for omics analysis in precision medicine. Using stringent purification procedures are of key importance in using blood samples for identification and application of biomarkers.

This study comprehensively assessed omics variables significantly impacted by the increase in hemolysis score in buffy coat and plasma/serum. Differences were identified that were associated with increasing hemolysis score, including missingness of proteins identified. Integration of lipidomics, metabolomics and proteomics data provided an expanded, comprehensive insight of the impact of hemolysis. Overall, our results will serve as a comprehensive resource to the biomarker community in the field of blood analysis. Diagnostic applications will be able to leverage these proteins, lipids and metabolites identified as hemolytic contamination for future biomarker studies.

Declarations

Ethics approval and consent to participate

This study was IRB approved and all patients consented to participate.

Consent for publication

All authors consent to publication.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

RS: Drafted the work or substantively revised it, provided substantial contributions to the conception and design of the work and acquisition, analysis

PS: Drafted the work or substantively revised it, provided substantial contributions to the conception and design of the work and acquisition, analysis

KO: Design of the work and acquisition, analysis

VB: Design of the work and acquisition, analysis and interpretation of data

VT: Design of the work and acquisition, analysis and interpretation of data

BG: Design of the work and acquisition, analysis and interpretation of data

HL: Analysis and interpretation of data

CD: Analysis and interpretation of data

EG: Substantively revised it, analysis and interpretation of data

LOR: Analysis and interpretation of data

EG: Drafted the work or substantively revised it, analysis and interpretation of data

AJM: Drafted the work or substantively revised it, provided substantial contributions to the conception and design of the work and acquisition, analysis

RS: Drafted the work or substantively revised it, provided substantial contributions to the conception and design of the work and acquisition, analysis

NRN: Drafted the work or substantively revised it, provided substantial contributions to the conception and design of the work and acquisition, analysis

References

1. Chen D, Schubert P, Devine DV. Proteomic analysis of red blood cells from donors exhibiting high hemolysis demonstrates a reduction in membrane-associated proteins involved in the oxidative response. Transfusion. 2017;57(9):2248–56.

2. Barasa B, Slijper M. Challenges for red blood cell biomarker discovery through proteomics. Biochimica et Biophysica Acta (BBA)-Proteins Proteomics. 2014;1844(5):1003–10.

3. Mescher AL: **Junqueira's basic histology: text and atlas**, vol. 12: McGraw-Hill Medical 13th ed. New York; 2013.

4. Bittremieux W, Tabb DL, Impens F, Staes A, Timmerman E, Martens L, Laukens K. Quality control in mass spectrometry-based proteomics. Mass Spectrom Rev. 2018;37(5):697–711.

5. Geyer PE, Voytik E, Treit PV, Doll S, Kleinhempel A, Niu L, Müller JB, Buchholtz ML, Bader JM, Teupser D. Plasma proteome profiling to detect and avoid sample-related biases in biomarker studies. EMBO molecular medicine. 2019;11(11):e10427.

6. Smith M, Chan Y, Dolci A, Kellogg M, McCudden C, McLean M: **Hemolysis, icterus, and lipemia/turbidity indices as indicators of interference in clinical laboratory analysis, approved guideline**. *Wayne, PA, USA: Clinical and Laboratory Standards Institute* 2012.

7. Sturtz LA, Wang G, Shah P, Searfoss R, Raj-Kumar P-K, Hooke JA, Fantacone-Campbell JL, Deyarmin B, Cutler ML, Sarangarajan R. Comparative analysis of differentially abundant proteins quantified by LC–MS/MS between flash frozen and laser microdissected OCT-embedded breast tumor samples. Clin Proteomics. 2020;17(1):1–12.

8. Baskin AS, Linderman JD, Brychta RJ, McGehee S, Anflick-Chames E, Cero C, Johnson JW, O'Mara AE, Fletcher LA, Leitner BP. Regulation of human adipose tissue activation, gallbladder size, and bile acid metabolism by a β3-adrenergic receptor agonist. Diabetes. 2018;67(10):2113–25.

9. Drolet J, Tolstikov V, Williams BA, Greenwood BP, Hill C, Vishnudas VK, Sarangarajan R, Narain NR, Kiebish MA. Integrated metabolomics assessment of human dried blood spots and urine strips. Metabolites. 2017;7(3):35.

10. Gacias M, Gaspari S, Santos P-MG, Tamburini S, Andrade M, Zhang F, Shen N, Tolstikov V, Kiebish MA, Dupree JL. Microbiota-driven transcriptional changes in prefrontal cortex override genetic differences in social behavior. elife. 2016;5:e13442.

11. Kiebish MA, Cullen J, Mishra P, Ali A, Milliman E, Rodrigues LO, Chen EY, Tolstikov V, Zhang L, Panagopoulos K. Multi-omic serum biomarkers for prognosis of disease progression in prostate cancer. Journal of translational medicine. 2020;18(1):1–10.

12. Tolstikov V, Nikolayev A, Dong S, Zhao G, Kuo M-S. Metabolomics analysis of metabolic effects of nicotinamide phosphoribosyltransferase (NAMPT) inhibition on human cancer cells. PloS one. 2014;9(12):e114019.

13. Bajad S, Shulaev V. LC-MS-based metabolomics. In: Metabolic Profiling. edn.: Springer; 2011. pp. 213–28.

14. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, Brown M, Knowles JD, Halsall A, Haselden JN. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nature protocols. 2011;6(7):1060–83.

15. Want EJ, Masson P, Michopoulos F, Wilson ID, Theodoridis G, Plumb RS, Shockcor J, Loftus N, Holmes E, Nicholson JK. Global metabolic profiling of animal and human tissues via UPLC-MS. Nature protocols. 2013;8(1):17–32.

16. Yuan M, Breitkopf SB, Yang X, Asara JM. A positive/negative ion–switching, targeted mass spectrometry–based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. Nature protocols. 2012;7(5):872.

17. Gao F, McDaniel J, Chen EY, Rockwell H, Lynes MD, Tseng Y-H, Sarangarajan R, Narain NR, Kiebish MA. Monoacylglycerol analysis using MS/MSALL quadruple time of flight mass spectrometry. Metabolites. 2016;6(3):25.

18. Lynes MD, Leiria LO, Lundh M, Bartelt A, Shamsi F, Huang TL, Takahashi H, Hirshman MF, Schlein C, Lee A, et al. The cold-induced lipokine 12,13-diHOME promotes fatty acid transport into brown adipose tissue. Nat Med. 2017;23(5):631–7.

19. Powell WS. Extraction of eicosanoids from biological fluids, cells, and tissues. Methods in molecular biology. 1999;120:11–24.

20. Durbin BP, Hardin JS, Hawkins DM, Rocke DM. A variance-stabilizing transformation for gene-expression microarray data. Bioinformatics. 2002;18(suppl_1):105–10.

21. Huber W, Von Heydebreck A, Sültmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics. 2002;18(suppl_1):96–104.

22. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118–27.

23. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882–3.

24. Yoshida T, Prudent M, D'Alessandro A. Red blood cell storage lesion: causes and potential clinical consequences. Blood Transfusion. 2019;17(1):27.

25. D'Alessandro A, Dzieciatkowska M, Hill RC, Hansen KC. Supernatant protein biomarkers of red blood cell storage hemolysis as determined through an absolute quantification proteomics technology. Transfusion. 2016;56(6):1329–39.

26. Zubarev RA. The challenge of the proteome dynamic range and its implications for in-depth proteomics. Proteomics. 2013;13(5):723–6.

27. Weisenhorn EM, Riley NM, Hess JR, Raife TJ, Coon JJ. Multi-omics evidence for inheritance of energy pathways in red blood cells. Molecular Cellular Proteomics. 2016;15(12):3614–23.

28. Takakura M, Yokomizo A, Tanaka Y, Kobayashi M, Jung G, Banno M, Sakuma T, Imada K, Oda Y, Kamita M: **Carbonic anhydrase I as a new plasma biomarker for prostate cancer**. *International Scholarly Research Notices* 2012, **2012**.

29. Hollander Z, Dai DL, Putko BN, Yogasundaram H, Wilson-McManus JE, Thompson RB, Khan A, West ML, McManus BM, Oudit GY. Gender-specific plasma proteomic biomarkers in patients with Anderson−Fabry disease. Eur J Heart Fail. 2015;17(3):291−300.
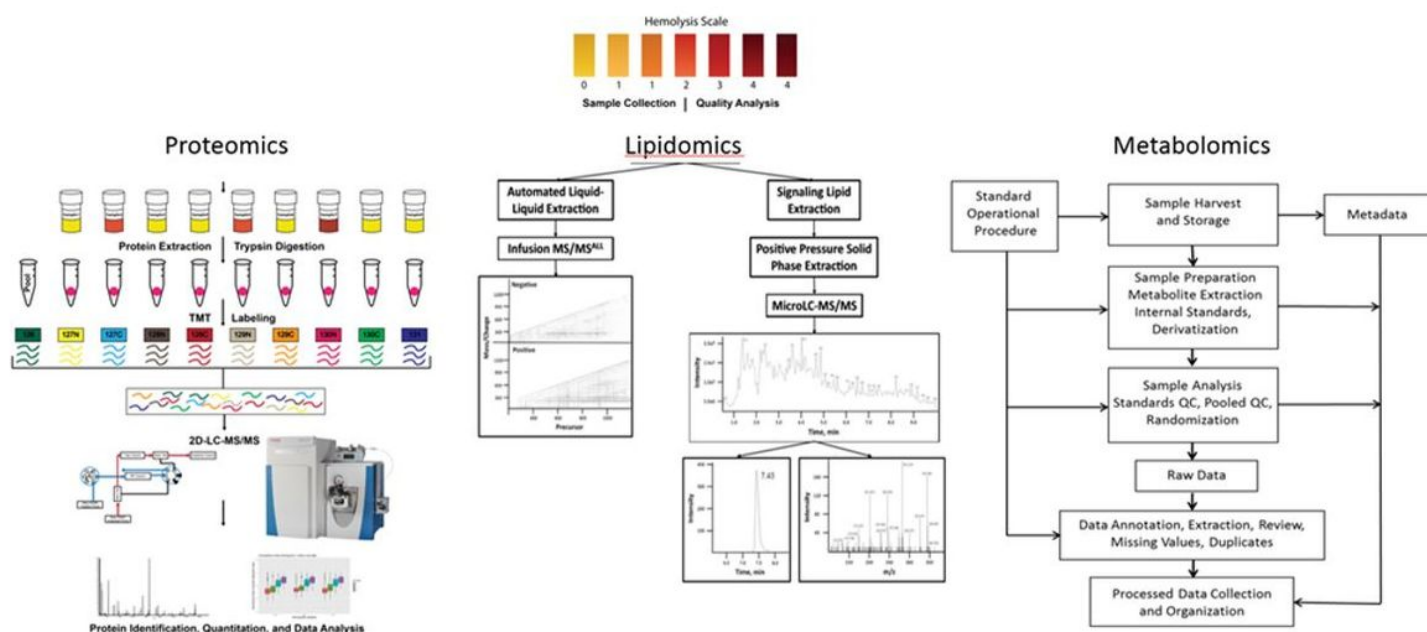
Figures

Figure 1

Workflow of the methods used to study the impact of Hemolysis. Initially, clinical samples were assigned a hemolysis score of 0-4 following the hemolysis scale color legend. In proteomics, plasma and serum were filtered and depleted of the top 14 most abundant proteins, and buffy coat cells were lysed. Proteins were extracted and digested with trypsin before being labeled with TMT 10-Plex. TMT-labeled peptides were analyzed using 2D LC-MS/MS platform and quantified using Proteome Discoverer v1.4. In lipidomics, structural lipids were extracted via liquid/liquid extraction method on an automated Hamilton Robotics STARlet system. Extracted lipids were analyzed via direct injection electrospray ionization TOF-MS. Further, mediator lipids were acidified and extracted using SPE. Eluted lipids were dried and resuspended for LC-MS analysis. In metabolomics, metabolites were extracted in organic conditions and analyzed using gas chromatography− mass spectrometry (GC/MS), reversed-phase liquid chromatography−mass spectrometry (RP-LC/MS), and hydrophilic interaction chromatography−liquid chromatography−tandem mass spectrometry (HILIC-LC/MS/MS). Post-processing of data included inspection, merging, and imputation.
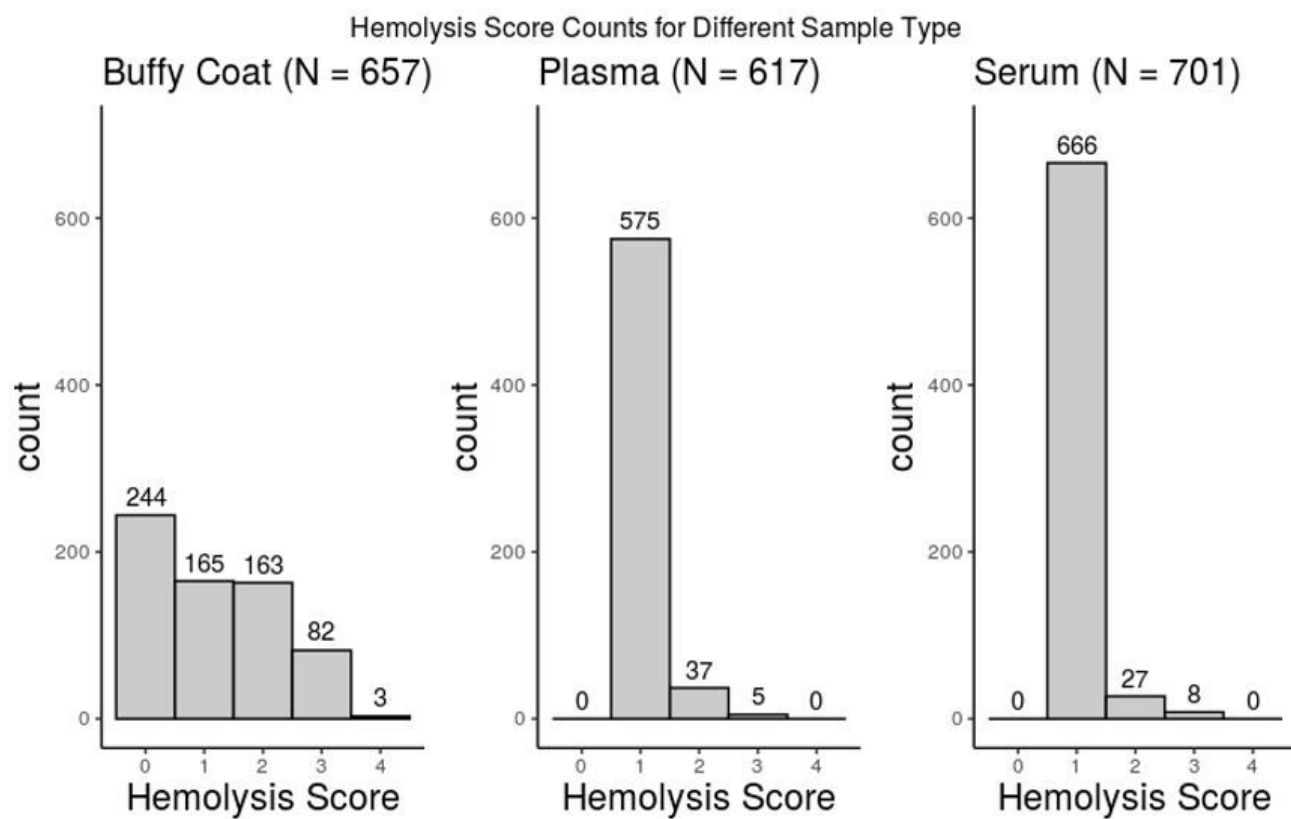
Hemolysis Score Counts for Different Sample Type

Figure 2

Hemolysis score distribution of sample count buffy coat (N=657), Plasma (N=617), and Serum (N=701) samples included in this study.
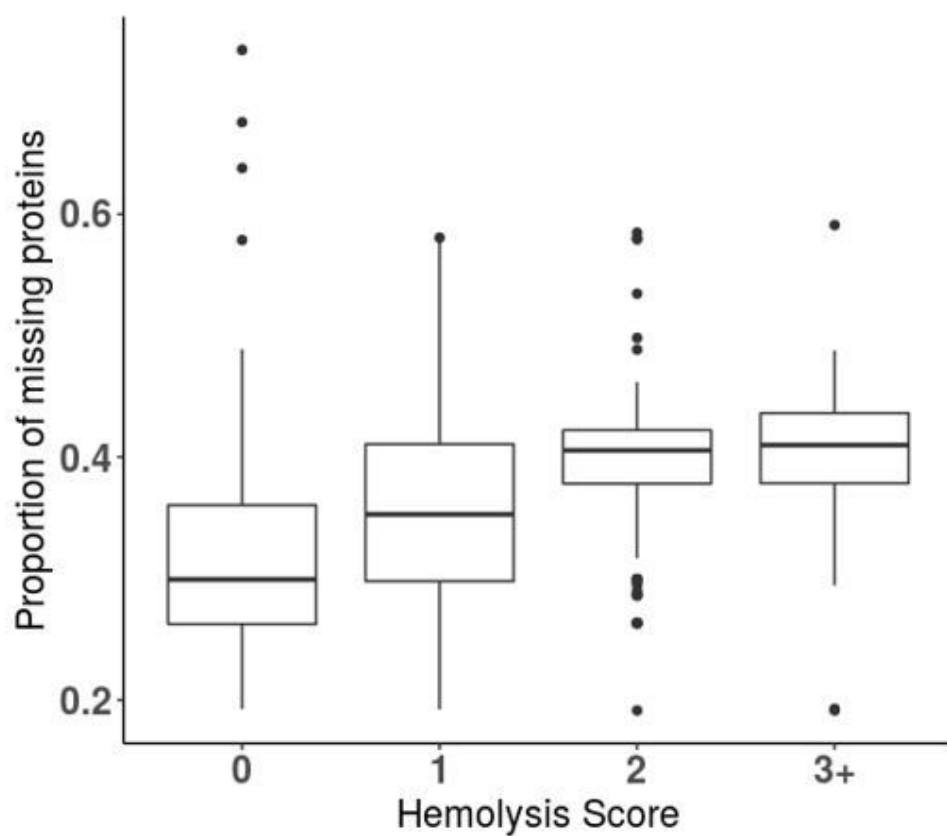
Figure 3

Boxplots visualizing relative proportion of missing proteins. Missing proteins were determined by comparing total number of proteins identified across all samples of each type, and number of proteins not identified in each sample relative to total identified proteins 3648, groups by hemolysis score of 0, 1, 2 and 3.
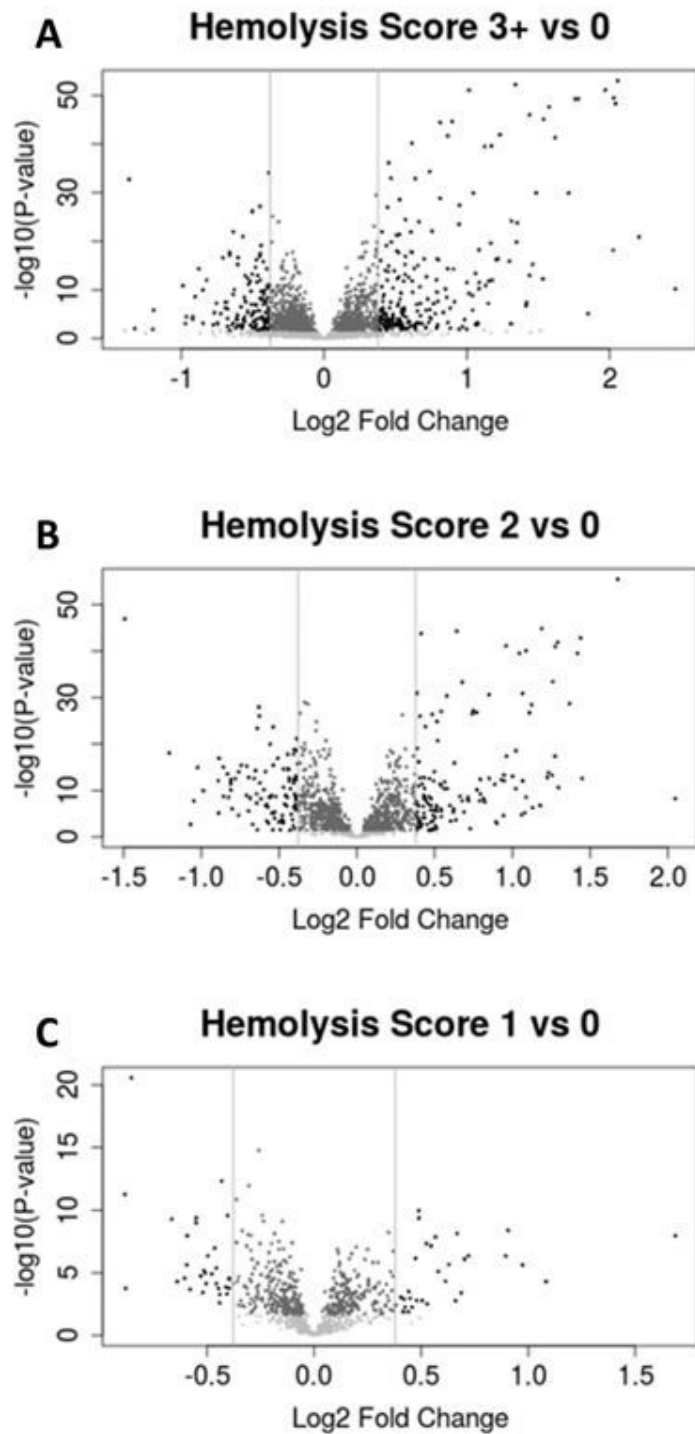
Figure 4

Volcano plots showing a comparison of protein expression between 3 vs 0, 2 vs 0 and 1 vs 0. The expression of protein ratio to the QCP was exhibited as Log2 fold and compared to -Log10 of p-value. Significant proteins required minimum 1.5 fold-change difference and maximum p-value of 0.05.
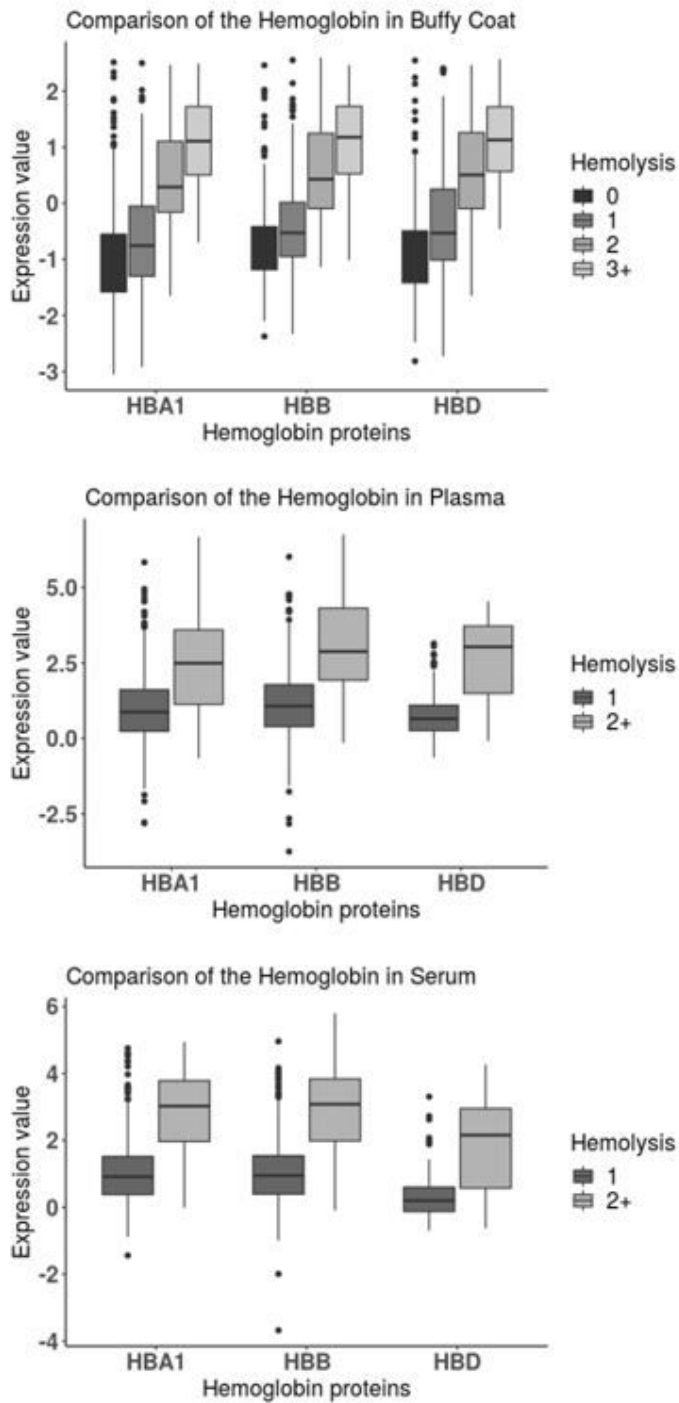
Figure 5

Expression of hemoglobin proteins in buffy coat, plasma, and serum (HBA1 = Hemoglobin Subunit Alpha, HBB = Hemoglobin Subunit Beta, HBD = Hemoglobin Subunit Delta) as a relative measure of hemolysis. Samples were grouped into hemolysis score 1 and 2+ for plasma and serum due to low number of score 3 and 4 samples in these matrices.
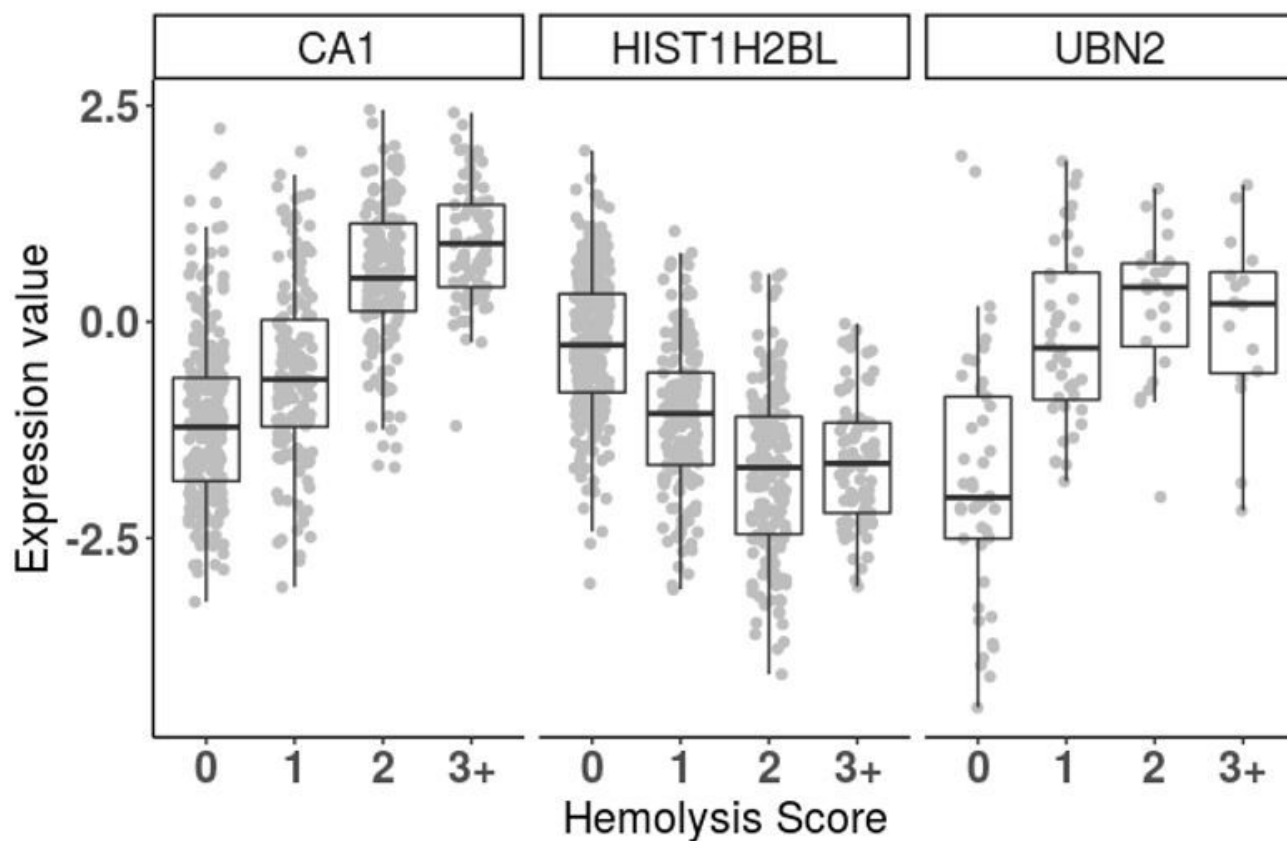
Figure 6

is a boxplot of buffy coat proteins expressions of (CA1 = Carbonic anhydrase, HIST1H2BL = Histone H2B type 1-L, UBN2 = Ubinuclein-2) that were identified as significantly differentially expressed proteins in comparison to their Hemolysis score of 0,1,2 and 3+. Expression values are log2 ratio to the pool.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalTable.xlsx