Red blood cell proteomics update: is there more to discover?

Angelo D'Alessandro, Monika Dzieciatkowska, Travis Nemkov, Kirk C. Hansen

Department of Biochemistry and Molecular Genetics, University of Colorado Denver - Anschutz Medical Campus, Aurora, CO, United States of America

Introduction

A great deal of interest has been attracted by back-of-the-envelope calculations suggesting that bacterial cells (recently revised to be ~3.9×10^{13}) outnumber the total number of host cells in the human body (3.0×10^{13}, estimated for a 70 kg "reference man")[1,2]. Although early estimates of ~2.8×10^{13} red blood cells (RBC) in adult humans[3] seem excessive in comparison to these numbers, it is universally accepted that RBC constitute the most abundant cell type in the body (~2.5×10^{13}) in foetal and post-natal life[3]. With an average lifespan of 120 days in circulation, every day ~2×10^{11} RBC are continuously cleared/released in the blood stream[3]. However, mature RBC are devoid of nuclei and organelles and each mature RBC contains ~270×10^{6} molecules of haemoglobin, accounting for approximately 90% of its dry weight[4]. It is not, therefore, surprising that RBC have been almost exclusively viewed as simple biological envelopes for circulating haemoglobin, in the light of their lack of nuclei and organelles and their clear and unique role in oxygen transport and gas homeostasis functions.

Understanding the complexity of RBC biology beyond gas transport is instrumental to the closely related fields of haematology and Transfusion Medicine. Approximately 108 million units of donated blood are collected globally every year (WHO, 2015). Transfusion of packed RBC is currently the only life-saving intervention for massively or chronically transfused recipients, such as trauma and sickle cell disease/blood cancer patients, respectively. Of note, trauma is currently the leading cause of death under the age of 59 (Center for Disease Control, 2010-2014), further highlighting the clinical relevance of transfusion of packed RBC.

In the last 14 years, proteomics technologies have been used to shed light on the 10% low abundance RBC proteome, revealing an unanticipated level of complexity proportional to the advancements in the analytical technologies and pre-fractionation strategies used to perform the analyses (Figure 1A)[5-13]. Strides in the field of mass spectrometry (MS)-based proteomics technologies have enabled the detection of up to 1,578 unique protein species in a single experiment, following extensive sample pre-fractionation with combinatorial hexapeptide ligand libraries, two-dimensional gel electrophoresis and nano-high performance liquid chromatography (HPLC)-MS/MS analysis[11]. This extensive dataset was integrated with previous -omics results to generate an early *in silico* interaction map of human RBC[9,14], and subsequent expansions to annotate up to 2,289 distinct gene products in mature erythrocytes[13].

Recently, ambitious proteomics projects have expanded our understanding of the human proteome[15,16]. However, none of these massive studies has focused on the RBC proteome, either under the assumption that the majority of the RBC proteome had been substantially unveiled by extant studies or to avoid the technical issues associated with the overwhelming abundance of RBC haemoglobin. The assumption that the RBC proteome had been fully elucidated has been recently challenged by in-depth proteomics studies, which combined advances in the field of proteomics technologies, MS-instrumentation and database searches to enable the identification of up to 2,838 unique proteins in adult and cord erythroid cells prior to or following reticulocyte maturation[12]. Here we highlight how these advances have made it possible to monitor ~2,000 RBC proteins by exploiting a routine laboratory proteomics workflow, using the GeLC-MS approach[17]. Since the presence of specific enzymes in the proteomics dataset does not necessarily indicate residual activity of the given enzyme, we integrated proteomics results with tracing metabolomics experiments by incubating human packed RBC with $^{13}C_{1,2,3}$-glucose-supplemented additives. The findings confirmed the presence and activity of cytosolic versions of Krebs cycle enzymes in mature packed RBC. The metabolomics tracing experiments also suggested that the present proteomics map is incomplete, by highlighting the presence and storage-dependent activation of enzymes that can catalyse pyruvate transamination reactions, an observation that paves the way for future, more in-depth analyses of the RBC proteome and opens new potential scenarios in the formulation of alternative storage additives for transfusion purposes.

Materials and methods

Blood collection and processing

Blood was collected from four healthy donor volunteers, in agreement with the Declaration of Helsinki.

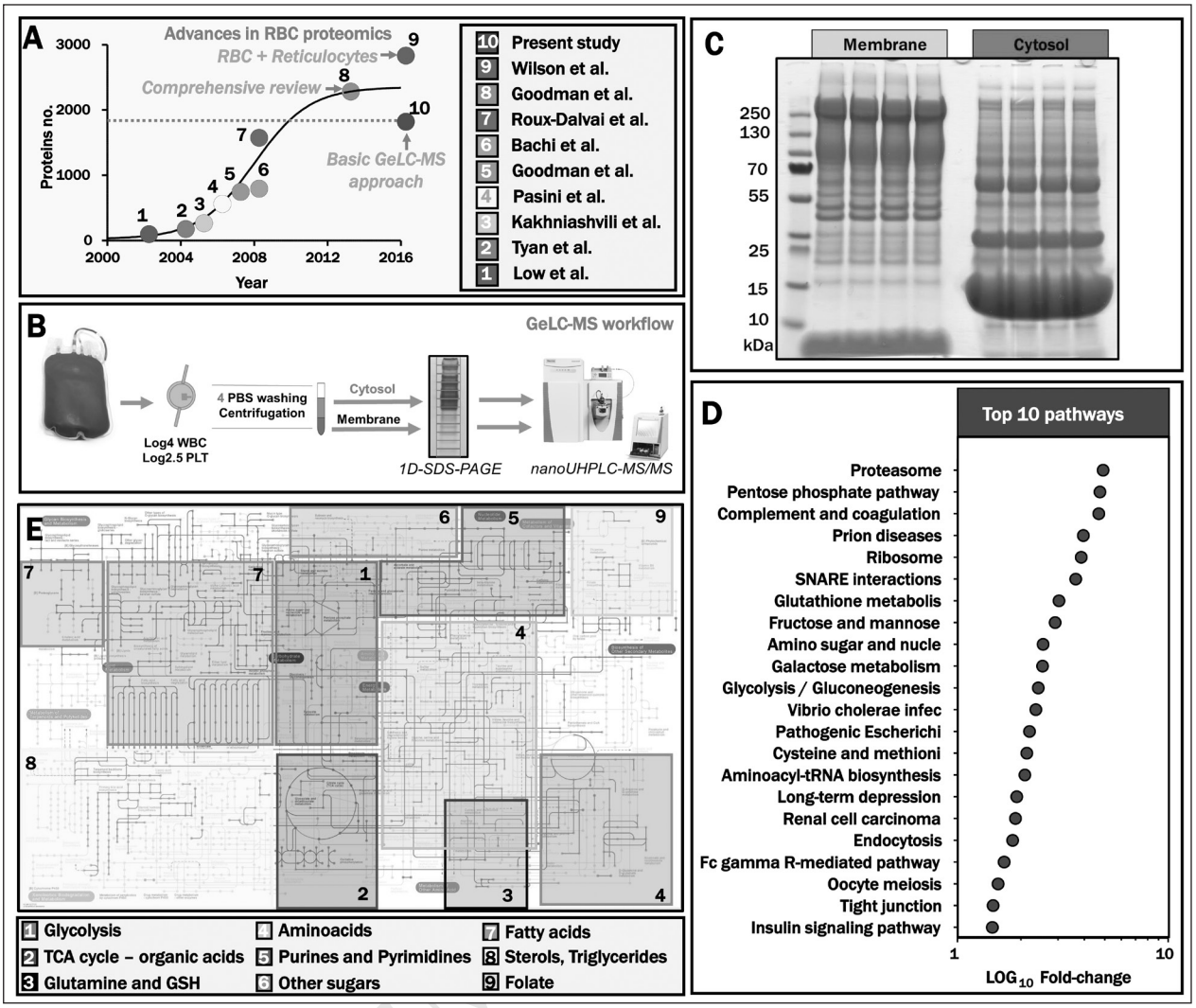

Figure 1 - The RBC proteome: an update.
(A) Proteomics technologies are continuously expanding our understanding of RBC complexity. (B) Here, proteomics analyses were performed through the GeLC-MS proteomics workflow[17]. (C) A representative 1D-SDS-PAGE preparative gel for RBC membrane and cytosol proteomics analyses. (D) Top pathways resulting from gene ontology term enrichment for biological function through DAVID (accessible at https://david.ncifcrf.gov/). (E) Metabolic enzymes identified in the RBC proteome were mapped against the KEGG pathway database (matching nodes are highlighted in dark grey; general pathways are indicated in the legend at the bottom of panel). A high-resolution version of this panel is provided in Online Supplementary Figure S1.

Packed RBC were stored in CP2D-AS-3 (Haemonetics, Braintree, MA, USA), as previously reported[18]. Samples were leucofiltered (RC2D, Haemonetics) to achieve a $\log_4$ reduction in white blood cells (WBC), to $<1\times10^6$ WBC/L or <0.5 WBC/1×10^6 RBC, improving upon previous erythrocyte enrichment protocols[11]. To remove any residual trace of plasma (90% of which is removed by filtration), platelets ($\log_{2.5}$-filtered) and WBC, the RBC were further washed three times with PBS + PMSF (154 mM NaCl, 10 mM phosphate buffer, pH 7.4, containing 0.1 mM phenylmethylsulfonyl fluoride)[11]. At each step supernatant and top RBC layers were removed to eliminate residual contaminating cells/plasma, prior to the final separation of RBC cytosol and membrane fractions[11].

GeLC-MS proteomics analysis

Proteomics analyses were performed using a GeLC-MS approach[17]. Briefly, 30 µg of protein extracts were run on a 4-12% gradient one-dimensional SDS-PAGE gel. Eighteen bands were cut from each lane, trypsin digested upon reduction and alkylation of unmodified cysteine residues, and analysed by nano-UHPLC-MS/MS (Easy-nLC1000, QExactive HF - positive ion mode, Thermo Fisher, Waltham, MA, USA). A volume of 8 µL of sample was injected through a trapping column

for desalting (ZORBAX 300SB-C18, 5×0.3 mm, 5 µm particle size, Agilent Technologies, Santa Clara, CA, USA) and then switched online at 600 nL/min over a home-made 100 µm × 150 mm fused silica capillary packed with Synergi Hydro C18 resin (4 µm, 80 Å from Phenomenex, Torrance, CA, USA). Samples were run on a 350 nL/min 90 min linear gradient of 5-32% ACN with 0.1% formic acid to separate the peptides. The mass spectrometer was operated in the positive ion mode, scan range m/z 400-2,000, followed by collision-induced dissociation MS/MS of the 20 most intense precursor ions at a normalised collision energy of 35 eV. Raw data files were converted into .mgf files which were then merged using an in-house script. Error tolerant searches were performed using Mascot (v. 2.4) against the human UniprotKB database (release date 2015.1.8), including decoy sequences (cysteine carbamidomethylation and methionine oxidation set as fixed and variable modifications, respectively). Mass tolerances for membrane and vesicle data were set to ±15 ppm for precursor ions and ±0.6 Da for fragment ions. For all Mascot search results, peptide spectral matches (PSM) were filtered at a 95% confidence threshold (excluding PSM with an expectation value >0.05).

$^{13}C_{1,2,3}$-glucose tracing metabolomics experiment

CP2D-AS-3 leucofiltered RBC concentrates were spiked with 1.1 mM $^{13}C_{1,2,3}$-glucose (SIGMA Aldrich, St. Louis, MO, USA), resulting in a ~20% addition to the total glucose (unlabeled ^{12}C-glucose monohydrate in AS-3 is ~5.55 mM[18]). Samples were sterilely harvested on storage days 2, 7 and on a weekly basis until the end of the shelf-life of the unit (storage day 42). RBC and supernatants were separated through gentle centrifugation, and measurements of $^{13}C_3$-alanine were performed as previously reported[18,19] using UHPLC-MS (Vanquish, Q Exactive - Thermo Fisher, San Jose, CA, USA). RBC were collected on a weekly basis and immediately extracted in ice-cold lysis/extraction buffer (methanol:acetonitrile:water 5:3:2) at 1:3 dilutions. Samples were then agitated at 4 °C for 30 min followed by centrifugation at 10,000 *g* for 15 min at 4 °C. Protein and lipid pellets were discarded, while supernatants were stored at –80 °C prior to metabolomic analyses. Twenty µL of RBC extracts were injected into an UHPLC system (Vanquish, Thermo Fisher) and run on a Kinetex C18 column (150 x 2.1 mm i.d., 1.7 µm particle size - Phenomenex, Torrance, CA, USA) at 250 µL/min (phase A: Optima H_2O, 0.1% formic acid; phase B: acetonitrile, 0.1% formic acid). The UHPLC system was coupled online with a Q Exactive mass spectrometer (Thermo Fisher), scanning in Full MS mode (2 µscans) at 70,000 resolution in the 60-900 m/z range, 4 kV spray voltage, 15 sheath gas and 5 auxiliary gas, operated in negative and then positive ion mode (separate runs). Metabolite assignments and isotopologue distributions were determined using Maven software (Princeton, NJ, USA), upon conversion of .raw files into .mzXML format through MassMatrix (Cleveland, OH, USA), correction for natural abundance of ^{13}C, and assignments confirmed against a heavy label standard for alanine (^{13}C ^{15}N-alanine and >650 standard compounds - IROATech, Sigma Aldrich, St. Louis, MO, USA). Reproducibility of extraction efficiency, chromatographic and MS technical stability were tested by determining coefficients of variation for the quantitation of spikes in heavy labelled standard mixes (0.25 µM final concentration - Cambridge Isotopes Laboratories, Inc., Tewksbury, MA, USA) and the xenometabolite 5-fluorouracil (2.5 µM final concentration - F6627-1G, Sigma Aldrich).

Absolute quantitation was performed by exporting the values for integrated peak areas of light metabolites and heavy labelled standards (1 µM - Cambridge Isotope Laboratories, Inc.) into GraphPad Prism 5.0.

Results and discussion

Proteomics analyses were performed using a GeLC-MS workflow (Figure 1B). Sample pre-fractionation strategies to remove haemoglobin and the most abundant RBC proteins (e.g. carbonic anhydrase)[11,20,21] had previously been used to make the low abundance RBC proteome amenable to detection through analytical proteomics workflow. Here, preliminary 1D-SDS-PAGE fractionation was adopted as part of a routine proteomics workflow. The resulting gel lanes for membranes and cytosol protein extracts where excised into 18 different bands, trypsin-digested prior to nanoUHPLC-MS peptidomics analysis. This workflow resulted in the identification of a total of 1,826 proteins (Online Supplementary Table SI). Bioinformatic analysis suggests that RBC can engage in many different pathways other than just gas transport (Figure 1D, Online Supplementary Table SII). In particular, mapping of the experimental results into the most comprehensive KEGG metabolic pathway (ko01100) revealed an unanticipated metabolic complexity of human RBC (highlighted in red - Figure 1E; Online Supplementary Figure S1). For example, here we identified cytosolic versions of enzymes involved in tri- and dicarboxylic acid metabolism, such as acetyl-coA citrate lyase, isocitrate dehydrogenase 1, pyruvate carboxylase, malate dehydrogenase 1 and fumarase (Figure 2A). Additionally, enzymes potentially promoting cytosolic reactions of one-carbon metabolism (serine hydroxymethyltransferase 1 and methyltetrahydrofolate reductase 1) have been identified (Figure 2B). Some of the enzymes mentioned above have the capacity to catalyse reactions that can generate

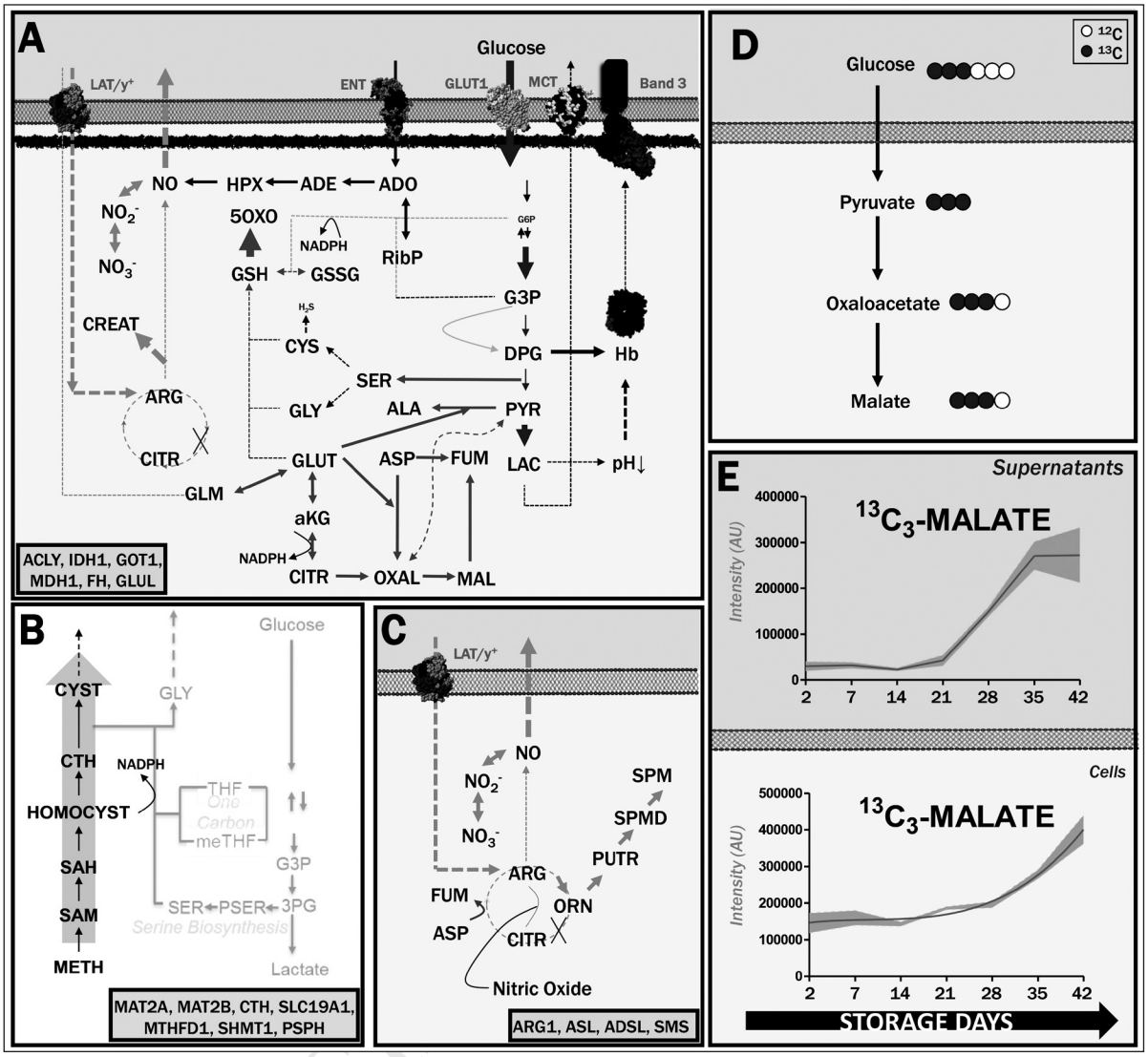

Figure 2 - Proteomics and tracing experiments reveal the presence of previously unappreciated metabolic pathways in mature RBC.
(A) An overview of the central metabolic pathways that are reasonably active in RBC in the light of proteomics evidence. Di- and tri-carboxylates can be metabolised by the enzymes indicated in the bottom left corner of the panel. (B) Enzymes identified in the cytosol of RBC in the present study which may be involved in sulphur/methionine synthesis and serine/folate metabolism are shown. Some reactions catalysed by IDH1 (A) or MTHFD1 (B) can generate NADPH without the activation of the oxidative phase of the pentose phosphate pathway (A). (C) An overview of metabolic reactions in nitrogen/arginine metabolism that are theoretically possible in RBC in the light of the enzymes identified in this study. (D) An overview of the $^{13}C_{1,2,3}$-glucose labelling experiment and expected isotopologue distribution in downstream glucose oxidation, pyruvate carboxylation and malate dehydrogenase 1-dependent generation of malate. (E) Experimental results of $^{13}C_3$-alanine detection in supernatants and cell extracts of stored RBC incubated with $^{13}C_{1,2,3}$-glucose. Median ± interquartile ranges are shown. Enzyme abbreviations are consistent with Uniprot names in Online Supplementary Table SI.

NADPH, a key antioxidant reducing equivalent until now thought to be produced exclusively by reactions from the oxidative phase of the pentose phosphate pathway in RBC. Additionally, we found evidence of the presence of enzymes involved in nitrogen metabolism (incomplete urea cycle) and polyamine synthesis in the RBC cytosol (Figure 2C).

Since identification of an enzyme in a proteomics dataset does not necessarily imply that this protein preserves its catalytic function in the mature erythrocyte, tracing experiments were designed by incubating leucocyte-filtered packed RBC with $^{13}C_{1,2,3}$-glucose prior to refrigerated storage for up to 42 days. Potential time-dependent incorporation of heavy

labelling in oxaloacetate and malate was predicted by the identification of pyruvate carboxylase and malate dehydrogenase 1 (cytosolic) in the proteomics dataset (Online Supplementary Figure S1). An overview of the predicted labelling pattern from $^{13}C_{1,2,3}$-glucose is provided in Figure 2D. MS-based metabolomics tracing experiments confirmed the previously reported storage-dependent accumulation of malate in packed RBC[18], suggesting that a minor percentage (<1%) of the total malate actually derives from heavy glucose catabolism through glycolysis, carboxylation of glucose-derived pyruvate and conversion of the latter substrate into malate (isotopologue M+3 of oxaloacetate and malate, respectively; Figure 2E). This observation also suggests that the majority of malate accumulating in stored RBC derives from sources other than glucose, aspartate or citrate-derived oxaloacetate representing the most likely candidate. While future studies will shed further light on this issue, in the present study the integration of proteomics results with evidence from metabolomics experiments in the presence of stable isotope tracers reveals for the first time that cytosolic versions of Krebs cycle enzymes are present and enzymatically active in mature RBC, an observation with clear potential implications for the future design and development of novel storage additives for improved *ex vivo* preservation of mature erythrocytes.

Extensive, tracing metabolomics experiments suggest that our protein coverage is not comprehensive. Metabolic labelling of packed RBC with $^{13}C_{1,2,3}$-glucose shows a storage duration dependency in pyruvate transamination to alanine (resulting in the accumulation of the $^{13}C_3$-alanine product in stored RBC and supernatants; Figure 3A,B). Even though our labelling data and previous targeted studies[22]but are unable to transport this amino acid across their cell membrane. Consequently, erythrocytes rely on de novo glutamate biosynthesis from α-ketoglutarate and glutamine to maintain intracellular levels of glutamate. Erythrocytic glutamate biosynthesis is catalyzed by three enzymes, alanine aminotransferase (ALT confirmed the presence and activity of glutamate pyruvate transaminase in human RBC, we did not identify this enzyme in our dataset. Similarly, we did not identify other low abundance RBC enzymes that may play adaptive regulatory roles in healthy or pathological RBC, such as sphingosine kinase 1[23].

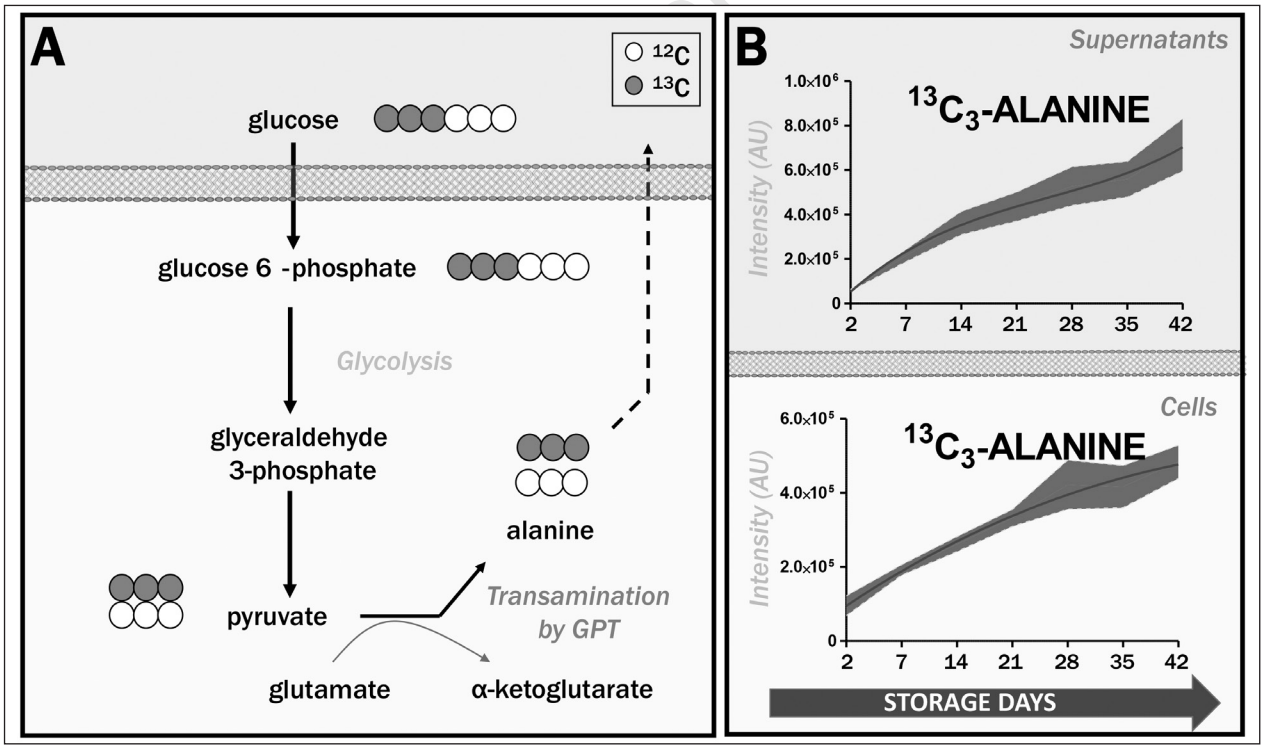

Figure 3 - Tracing experiments suggest that some enzymes that have not been identified in the updated proteomics dataset may actually be expressed and active in mature RBC.
(A) An overview of the $^{13}C_{1,2,3}$-glucose labelling experiment and expected isotopologue distribution in downstream glucose oxidation and pyruvate transamination reactions. (B) The experimental results of 13C3-alanine detection in supernatants and cell extracts of stored RBC incubated with $^{13}C_{1,2,3}$-glucose. Median ± interquartile ranges are shown.

Conclusions

The present dataset, although not definitive, supports the idea that there is still ample opportunity for discovery in one of the most historically studied human cells. Recent comprehensive updates of the RBC proteome complexity -only some of which are referenced here- have been detailed using advanced proteomics technologies and investigating both mature and immature mitochondria-endowed (reticulocyte) RBC fractions[12]. Further extensions of RBC protein complexity and biochemistry are anticipated, stemming from the application of advancing bioanalytical technologies. Meanwhile, this study could serve as a reference for the feasibility of large-scale RBC proteomics studies using routine laboratory analytical strategies while delivering results as comprehensive as those obtainable only with state-of-the-art technologies less than a decade ago. Combined advances in discovery-mode and functional -omics technologies are thus currently available to perform large-scale validation studies and enable the identification of novel modes of therapeutic intervention for the most abundant host cell in the human body.

Authorship contributions

MD, ADA performed the proteomics analyses and organised the resulting data. ADA, TN performed metabolomics analyses. KCH supervised the experiments and provided critical proteomics expertise and key instrumentation. ADA wrote the manuscript and all the Authors critically contributed to its finalisation. ADA and MD share first authorship.

Keywords: mass spectrometry, proteomics, error tolerant, erythrocyte, haematology.

Disclaimer of conflicts of interest

ADA received funds from the National Blood Foundation, Early Career Grant 2016 cycle. Though unrelated to the contents of the manuscript, the Authors disclose that ADA, TN and KCH are part of Endura LLC. ADA is a consultant for New Health Sciences.

References

1) Bianconi E, Piovesan A, Facchin F, et al. An estimation of the number of cells in the human body. Ann Hum Biol 2013; **40**: 463-71.
2) Sender R, Fuchs S, Milo R. Revised estimates for the number of human and bacteria cells in the body. PLoS Biol 2016; **14**: e1002533.
3) Ballas SK. Erythrocyte concentration and volume are inversely related. Clin Chim Acta 1987; **164**: 243-4.
4) D'Alessandro A, Kriebardis AG, Rinalducci S, et al. An update on red blood cell storage lesions, as gleaned through biochemistry and omics technologies. Transfusion 2015; **55**: 205-19.
5) Low TY, Seow TK, Chung MC. Separation of human erythrocyte membrane associated proteins with one-dimensional and two-dimensional gel electrophoresis followed by identification with matrix-assisted laser desorption/ionization-time of flight mass spectrometry. Proteomics 2002; **2**: 1229-39.
6) Tyan Y-C, Jong S-B, Liao J-D, et al. Proteomic profiling of erythrocyte proteins by proteolytic digestion chip and identification using two-dimensional electrospray ionization tandem mass spectrometry. J. Proteome Res 2005; **4**: 748-57.
7) Kakhniashvili DG, Bulla LA, Goodman SR. The human erythrocyte proteome analysis by ion trap mass spectrometry. Mol Cell Proteomics 2004; **3**: 501-9.
8) Pasini EM, Kirkegaard M, Mortensen P, et al. In-depth analysis of the membrane and cytosolic proteome of red blood cells. Blood 2006; **108**: 791-801.
9) Goodman SR, Kurdia A, Ammann L, et al. The human red blood cell proteome and interactome. Exp Biol Med 2007; **232**: 1391-408.
10) Bachi A, Simó C, Restuccia U, et al. Performance of combinatorial peptide libraries in capturing the low-abundance proteome of red blood cells. 2. Behavior of resins containing individual amino acids. Anal Chem 2008; **80**: 3557-65.
11) Roux-Dalvai F, Gonzalez de Peredo A, Simó C, et al. Extensive analysis of the cytoplasmic proteome of human erythrocytes using the peptide ligand library technology and advanced mass spectrometry. Mol Cell Proteomics 2008; **7**: 2254-69.
12) Wilson MC, Trakarnsanga K, Heesom KJ, et al. Comparison of the proteome of adult and cord erythroid cells, and changes in the proteome following reticulocyte maturation. Mol Cell Proteomics 2016; **15**: 1938-46.
13) Goodman SR, Daescu O, Kakhniashvili DG, Zivanic M. The proteomics and interactomics of human erythrocytes. Exp Biol Med 2013; **238**: 509-18.
14) D'Alessandro A, Righetti PG, Zolla L. The red blood cell proteome and interactome: an update. J Proteome Res 2010; **9**: 144-63.
15) Kim M-S, Pinto SM, Getnet D, et al. A draft map of the human proteome. Nature 2014; **509**: 575-81.
16) Wilhelm M, Schlegl J, Hahne H, et al. Mass-spectrometry-based draft of the human proteome. Nature 2014; **509**: 582-7.
17) Dzieciatkowska M, Hill R, Hansen KC. GeLC-MS/MS analysis of complex protein mixtures. Methods Mol Biol 2014; **1156**: 53-66.
18) D'Alessandro A, Nemkov T, Kelher M, et al. Routine storage of red blood cell (RBC) units in additive solution-3: a comprehensive investigation of the RBC metabolome. Transfusion 2015; **55**: 1155-68.
19) Reisz JA, Wither MJ, Dzieciatkowska M, et al. Oxidative modifications of glyceraldehyde 3-phosphate dehydrogenase regulate metabolic reprogramming of stored red blood cells. Blood 2016; **128**: e32-42.
20) Ringrose JH, van Solinge WW, Mohammed S, et al. Highly efficient depletion strategy for the two most abundant erythrocyte soluble proteins improves proteome coverage dramatically. J Proteome Res 2008; **7**: 3060-3.
21) D'Amici GM, Rinalducci S, Zolla L. Depletion of hemoglobin and carbonic anhydrase from erythrocyte cytosolic samples by preparative clear native electrophoresis. Nat Protoc 2012; **7**: 36-44.
22) Ellinger JJ, Lewis IA, Markley JL. Role of aminotransferases in glutamate metabolism of human erythrocytes. J Biomol NMR 2011; **49**: 221-9.
23) Zhang Y, Berka V, Song A, et al. Elevated sphingosine-1-phosphate promotes sickling and sickle cell disease progression. J Clin Invest 2014; **124**: 2750-61.

Arrived: 18 November 2016 - Accepted: 20 November 2016
Correspondence: Angelo D'Alessandro
Department of Biochemistry and Molecular Genetics
University of Colorado Denver Anschutz Medical Campus
12801 E. 17th Ave, RC1 South L18-9118
Aurora, CO 80045, USA
e-mail: angelo.dalessandro@ucdenver.edu