



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Tandem mass tag proteomic and untargeted metabolomic profiling reveals altered serum and CSF biochemical datasets in iron deficient monkeys



Brian J Sandri^{a,b}, Jonathan Kim^c, Gabriele R. Lubach^d, Eric F Lock^c, Candace Guerrero^e, LeeAnn Higgins^e, Todd W Markowski^e, Pamela J Kling^f, Michael K Georgieff^{a,b}, Christopher L. Coe^c, Raghavendra B Rao^{a,b,*}

^a Division of Neonatology, Department of Pediatrics, University of Minnesota, Minneapolis, MN, USA

^b Masonic Institute for the Developing Brain, University of Minnesota, Minneapolis, MN, USA

^c Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, USA

^d Harlow Center for Biological Psychology, University of Wisconsin, Madison, WI, USA

^e Department of Biochemistry, Molecular Biology, & Biophysics, University of Minnesota, Minneapolis, Minnesota, USA

^f Division of Neonatology, Department of Pediatrics, University of Wisconsin, Madison, WI, USA

ARTICLE INFO

Article history:

Received 20 July 2022

Revised 26 August 2022

Accepted 6 September 2022

Available online 11 September 2022

Keywords:

Rhesus Macaque

Multiomics

Synaptogenesis

ABSTRACT

The effects of early-life iron deficiency anemia (IDA) extend past the blood and include both short- and long-term adverse effects on many tissues including the brain. Prior to IDA, iron deficiency (ID) can cause similar tissue effects, but a sensitive biomarker of iron-dependent brain health is lacking. To determine serum and CSF biomarkers of ID-induced metabolic dysfunction we performed proteomic and metabolomic analysis of serum and CSF at 4- and 6-months from a nonhuman primate model of infantile IDA. LC/MS/MS analyses identified a total of 227 metabolites and 205 proteins in serum. In CSF, we measured 210 metabolites and 1,560 proteins. Data were either processed from a Q-Exactive (Thermo Scientific, Waltham, MA) through Progenesis Q1 with accurate mass and retention time compar-

* Corresponding author's email address and Twitter handle

E-mail address: raghurao@umn.edu (R.B. Rao).

isons to a proprietary small molecule database and Metlin or with raw files imported directly from a Fusion Orbitrap (Thermo Scientific, Waltham, MA) through Sequest in Proteome Discoverer 2.4.0.305 (Thermo Scientific, Waltham, MA) with peptide matches through the latest Rhesus Macaque HMDB database. Metabolite and protein identifiers, p-values, and q-values were utilized for molecular pathway analysis with Ingenuity Pathways Analysis (IPA). We applied multi-way distance weighted discrimination (DWD) to identify a weighted sum of the features (proteins or metabolites) that distinguish ID from IS at 4-months (pre-anemic period) and 6-months of age (anemic).

© 2022 The Author(s). Published by Elsevier Inc.
 This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Specifications Table

Subject	Omics: General Omics: Metabolomics Omics: Proteomics Neuroscience: Developmental
Specific subject area	Multomics and the long-term physiological consequences of early-life iron deficiency.
Type of data	Table Image Chart Graph Figure Illustration
How the data were acquired	Metabolomic data were acquired via ultra high-performance liquid chromatography (UHPLC) inline with a Thermo Q Exactive Quadrupole Orbitrap high-resolution mass spectrometer. Metabolite identifications were achieved through comparison of MS2 fragment data using Metlin and Progenesis Q1 software. Quantitative proteomic data were acquired with a Shimadzu UHPLC system and the Thermo Fusion Orbitrap mass spectrometer. Identifications were processed through SEQUEST in the Proteome Discover environment.
Data format	Raw Analyzed Annotated
Description of data collection	Serum and CSF from 12 iron sufficient (IS) and 7 ID rhesus macaques were analyzed for hematological and metabolomic profiles. Due to experimental balance constraints a subset of 6 IS and 6 ID monkeys were proteomically profiled. Proteomic data were normalized to a common mastermix control using Scaffold Q+ software (Proteome Software, Portland, OR)
Data source location	Institution: University of Minnesota – Twin Cities Campus City: Minneapolis and St. Paul, MN Country: United States of America GPS coordinates: 44.97356, -93.20992
Data accessibility	All supplemental figures, tables, and data methods are located at https://doi.org/10.5281/zenodo.7023976 Proteomic data repository: PRIDE Data identification number: PXD028275 Direct URL to the data: http://central.proteomexchange.org/cgi/GetDataset?ID=PX028275 Metabolomic data repository: MetaboLights Data identification number: MTBL53388 Direct URL to the data: https://www.ebi.ac.uk/metabolights/MTBL53388

(continued on next page)

Related research article

B.J. Sandri, J. Kim, G.R. Lubach, E.F. Lock, C. Guerrero, L. Higgins, T.W. Markowski, P.J. Kling, M.K. Georgieff, C.L. Coe, R.B. Rao, Multiomic profiling of iron-deficient infant monkeys reveals alterations in neurologically important biochemicals in serum and cerebrospinal fluid before the onset of anemia, *Am J Physiol Regul Integr Comp Physiol.* 322 (2022) R486–R500.
<https://doi.org/10.1152/ajpregu.00235.2021>

Value of the Data

- Our data provides both proteomic and metabolomic information in the early iron-deficient and anemic period of rhesus macaques at 4- and 6- months. Because these data capture the period immediately prior to full-blown anemia the data can be examined for useful markers of impending organ damage due to iron-deficiency (ID) that may not be reflected in traditional hematological measures.
- These data can be beneficial to researchers seeking to examine the role of iron on early-life development with both immediately consequential biomarkers (metabolites) and potentially long-term changes (proteins). By examining multiple biological compartments and time points greater physiological and temporal insight can be realized.
- Although our focus was mainly on the effects of ID upon early-life development the data could be reused to understand biomarkers associated with normal development and to provide further insight into the timing and efficacy of future micronutrient studies.

1. Data Description

We generated quantitative proteomic, metabolomic, and hematological profiles of serum and CSF from *rhesus macaques* at 4- and 6- months of age. The data were processed as previously described [1–5] and canonical correlation analysis (CCA) was performed to examine significant trends (Fig. 1). Resultant data was examined using Ingenuity Pathways Analysis (IPA) to identify molecular pathways perturbed by early-life ID.

Metabolomic profiles for all samples can be found on the MetaboLights repository. Individual .RAW files are available for reverse phase (RP) positive, RP negative, and HILIC phases for serum and CSF at both time points. Metabolite IDs were matched through data files using Progenesis QI (Waters Inc., Milford, MA) program. Data processed through this pipeline interprets and aligns chromatographic peaks, then compares metabolite abundances across designated groups. Mass to charge ratio (M/Z) and column retention time comparisons to the Metlin small molecule MS/MS database (Scripps Research Institute, La Jolla, California) and a generated small molecule database were used to assign identifications to compounds presenting unique M/Z-retention time values. Filtered and curated lists can be found in the supplemental materials.

Proteomic RAW files from all three TMT-quantitative runs are in the PRIDE repository. Peptides were grouped into protein families based on alignment derived from the UniProt *Macaca mulatta* (Rhesus macaque) proteome (ID: UP000006718) using Scaffold Q+ software (Proteome Software, Portland, OR). For each protein we list the ratio of expression of ID/IS, p-value, and false-discovery rate.

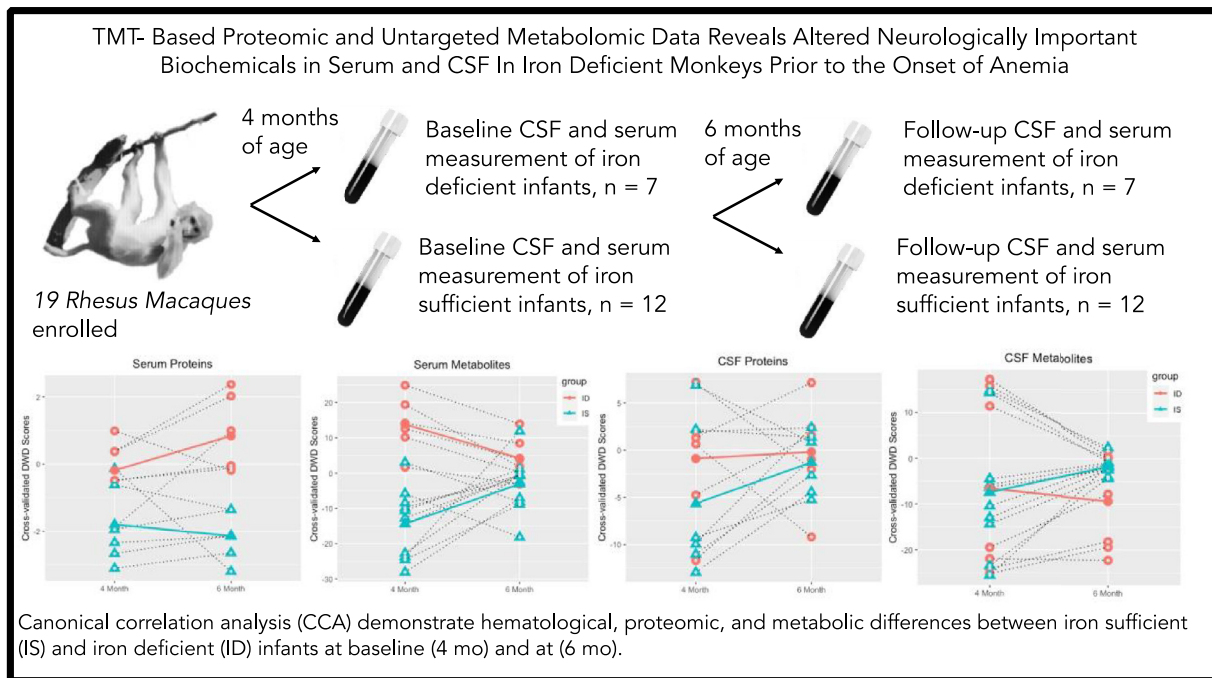


Fig. 1. Experimental workflow for monkey samples. Rhesus Macaques were generated at the University of Wisconsin – Madison with pregnant dams fed a diet that can produce ID infants. In this study, 12 IS and 7 ID infants were enrolled with serum and CSF samples obtained at 4- and 6- months for the purpose of obtaining hematological, and quantitative proteomic and metabolomic data. Multiway-distance-weighted discriminatory analysis of several compartments, time points, and conditions demonstrate differential hematological, proteomic, and metabolic profiles based on iron status.

2. Experimental Design, Materials and Methods

2.1. Monkeys

Participants were selected from a colony of rhesus monkeys (*Macaca mulatta*) from 19 mother-reared infants starting at 4-months of age. Blood and CSF samples were acquired to obtain hematological status, proteomic and metabolomic profiles at 4- and 6-months of age. This experiment and all procedures approved by the Animal Care and Use Committees at the University of Minnesota and the University of Wisconsin.

2.2. Hematology

The traditional measures of IDA include MCV < 60.0 fL, Hgb < 10.0 g/dL, RDW > 15.0%, and ZnPP/H > 150 μ mol/mol. All these assays were determined at a certified clinical lab (UnityPoint Health Meriter Labs, Madison, WI). Zinc protoporphyrin/heme ratio (ZnPP/H) was determined on site using a hemato-fluorometer (Aviv Bio-medical Sciences, Lakewood, NJ).

Abbreviations of hematological acronyms: red blood cell count (RBC), mean corpuscular volume (MCV), hematocrit (HCT), mean corpuscular Hgb (MCH), and RBC distribution width (RDW).

2.3. Separation of proteomic and metabolomic fractions

Serum and CSF samples were separated into primarily proteomic or metabolomic fractions through application of Amicon Ultra, 3kDa spin filtration (Millipore, Burlington, MA). Retained fraction contained proteins that were used for proteomic analysis whilst the flow-thru contained primarily metabolites and peptides < 3kDa. Metabolites were isolated through cold-methanol extraction while proteins underwent urea-based denaturation followed by tryptic digestion.

2.4. Data collection - metabolomics

Both CSF and blood samples were separated through ultra-high-performance liquid chromatography (UHPLC) using the Scientific Ultimate 3000 RS UHPLC platform (Thermo Scientific, Waltham, MA). Reversed phase (RP) chemistry methods were conducted using an Acuity BEH particle C-18 column measuring 2.1 \times 100 mm and heated to 40°C with a flow rate of 0.4 mL/min. RP positive mode consisted of an aqueous phase containing 0.1% formic acid with buffer B containing 99.9% acetonitrile and 0.1% formic acid. The gradient started at 2% B and ramped to 25% B in 60 seconds, then 80% B in 7 minutes. It reached 100% B in 6 min and was held at 100% B for 120 seconds. RP negative mode consisted of 10 mM ammonium acetate pH 9 for buffer A and buffer B was 100% acetonitrile. The gradient ramped identically to RP positive mode.

For both preparations, the column was attached to a heated electro-spray ionizing source mounted to a Q-Exactive Quadrupole-Orbitrap mass spectrometer (Thermo Scientific, Waltham, MA). The voltage spray was set to 3.4 kilo volts, gas flow at the sheath was set to 50 with auxiliary gas set to 8. The Slens RF was set to 55 volts with the probe heater at 400° C and the capillary heated to 320°C. We acquired MS1 spectra within the 70–1050 m/z window with a resolution of 70,000 at 200 m/z. We used a max injection time of 200 milliseconds and automatic gain control (AGC) of 1.0E6. After acquiring the MS1 only data in full scan mode we identified metabolites of interest through MS2 fragmentation studies. The MS2 scans were conducted using a dynamic inclusion list of targets from MS1. High energy fragmentation was achieved with gradually increased collisional energy from 30 – 60 with a detector setting 17,500 resolution, AGC of 1e5 ions and 100 milliseconds of injection time.

2.5. Data analysis - Metabolomic

Data files from the Q Exactive were imported into Progenesis Q1 (Waters Inc., Burlington, MA) software. This program can interrogate chromatographic information and generate comparative quantitation across multiple comparisons. We ran pooled and blank controls to establish baseline and false signals which can then be used to filter experimental data. Metabolites with high variation, low signal, or insufficient peak width were discarded. Features that met these stringent criteria were then identified through comparison with an internally generated library of small molecules and comparison to the Metlin small molecule MS/MS database (Scripps Research Institute, La Jolla, California). Identifications with a dot product scoring of 700 or greater were retained as valid pending fragmentation analysis.

2.6. Trypsin digestion - Proteins

Prior to trypsin digestion concentration of protein was assayed using Bradford reagent (Bio-Rad, Hercules, CA). After determination of protein concentration, a 20-microgram amount was added to a microfuge tube with lysis buffer (7 M urea, 2 M thiourea, 0.4 M ammonium bicarbonate, pH 8.0). All samples were incubated at 37°C for 45 minutes. Four-fold dilution was achieved through addition of mass spec-grade water (Fisher Scientific, Pittsburgh, PA) and Trypsin Gold reagent was added at a ratio of 1 microgram Trypsin Gold to 25 micrograms of protein. Digestion proceeded in a dry heat incubator set to 37°C for 16 hours. Upon completion samples were frozen at -80°C for 30 minutes and speed vacuumed to dryness. 1mL C18 column (Restek, Bellefonte, PA) cleanup was performed by vacuum manifold induction. Proteomic analysis was limited to 6 ID and 6 control IS samples to maintain balance and match available channels present in the TMT reagent kit.

2.7. Data collection - Proteomics

Samples labeled with the Tandem Mass Tag (Thermo Scientific, Waltham, MA) were suspended in ammonium formate buffer and fractionated using high pH with C18 chromatography. A hot-sleeve Column Heater (Analytical Sales & Products, Pompton Plains, NJ) was used to maintain proper temperature on a Prominence (Shimadzu, Columbia, MD) HPLC system with a Gemini NX C18 cartridge (Phenomenex, Torrance, CA) and a C18 XBridge column measuring 0.15 meters with an internal diameter of 2.1 millimeters and 5-micron particle size (Waters Corporation, Milford, MA). Buffer A consisted of 20 mM ammonium formate, pH 10 in 98 parts water and 2 parts acetonitrile with Buffer B containing 20 mM ammonium formate, pH 10 in 10 parts water and 90 parts acetonitrile. The gradient flow was 200 microliters a minute ramping from 2% to 7% buffer B over 30 seconds then 7% to 15% buffer B over 7 and a half minutes. Buffer B ramped from 15% to 35% over $\frac{3}{4}$ of an hour and from 35% to 60% in 15 minutes. Fractions were obtained in 2-minute intervals with UV monitoring at 215 and 280 nanometers. Peptides were concatenated according to early or late status as previously described [2].

These combined fractions were then dried and brought up in solvent 98 parts water, 2 parts acetonitrile, and 0.01 parts formic acid prior to injection into the Fusion Orbitrap LC/MS system (Thermo Scientific, Waltham, MA). Data were acquired by loading fractions on an Easy-nLC 1000 HPLC (Thermo Scientific, Waltham, MA) in tandem with an Orbitrap Fusion (Thermo Scientific, Waltham, MA). Peptides were introduced via a fused silica PicoTip Emitter (New Objective, Woburn, MA) packed in-house with ReproSil-Pur C18-AQ (1.9 μm particle, 120 \AA pore; Dr. Maish GmbH Ammerbuch, Germany) with the column kept at 55°C at a constant flow of 300 microliters per minute. Gradient parameters were 5% to 22% buffer B (99.9% acetonitrile, 0.1% formic acid) over 45 minutes followed by ramping from 22% to 35% for 25 minutes and 35% to 95% over the final ten minutes. This column directly fed the nanospray in line with the Orbitrap

Fusion mass spectrometer (Thermo Scientific, Waltham, MA). Voltage was kept at 2.1 kilovolts for positive mode and the capillary heated to 275°C. The mass spectrometer surveyed masses in the range of 380 – 1580 m/z at a resolution of 60,000 at 100 m/z with an AGC of 1.0E6 over a 250-millisecond injection. The twelve ions with the most intensity from the MS1 stage were fragmented by collisional dissociation with stabilized collisional energy of 35%, and detector settings of 60k resolution, AGC 5E4 ions, 250 milliseconds maximum injection time and Fourier transform first mass mode was fixed at 110 m/z. To prevent sampling of highly abundant proteins we set dynamic exclusion to a 40 second window with ± 10 ppm mass tolerance.

The drastic difference in proteins quantified in CSF vs serum was determined to be a shielding effect from the many species of highly-abundant protein in serum that can cause further complications with removal schemes [6]. The MS1 sampling space can become physically overwhelmed with ion species derived from these highly-abundant proteins [7]. This shielding issue is much less pronounced in CSF since only albumin is present whereas serum contains a host of immunoglobulins, apolipoproteins, along with albumin.

2.8. Database search parameters - Proteomics

The proteomic raw mass spectra data were analyzed with SEQUEST (XCorr only) in Proteome Discoverer 2.4.0.305 (Thermo Fisher Scientific, Waltham, MA). The UniProt (Uniprot.org) *Macaca mulatta* (tax ID 9544) Universal Proteome protein sequence database from Aug 18, 2020 (UP000013456) merged with the common lab contaminant sequences from <https://www.thegpm.org/crap/>, with a total of 17,804 entries, was utilized for the database search. Sequest searched the data with a fragment ion mass tolerance of 0.080 Da and a parent ion tolerance of 15 ppm. Carbamidomethyl cysteine, pyro-glutamic acid, deamidation of asparagine and glutamine, oxidation of methionine, acetyl of the protein N-terminus, and TMT 10-plex of lysine and peptide N-terminus were specified in Sequest as variable modifications.

2.9. Data analysis - Quantitation

We deployed Scaffold Q+ (version 4.10, Proteome Software, Portland, OR) to quantitate Tandem Mass Tag (TMT) based quantitation at the peptide level. Peptide-based identifications required an 8% probability in order to pass a false discovery rate (FDR) of 1% or less using the Percolator posterior error probability calculation [8]. Only protein identifications that had 99% or greater probability, contained 2 or more peptides, and an FDR of 1% or less were accepted. The Protein Prophet algorithm [9] was used to assign probabilities to the identified proteins. Protein families were utilized in cases where the algorithm could not definitively assign isoforms based on MS/MS data and to satisfy parsimony principles. Quantification channels were normalized using the algorithm described in i-Tracker [10]. Intra- and inter-run normalization was achieved through integrative intensity-based ANOVA based on median intensity [11]. For serum experiments 17125_17126_17127, 14,522 spectra in the experiment at the given thresholds, 12,064 (83%) were included in quantitation. For CSF experiments 17561_17562_17563, 56,229 spectra in the experiment at the given thresholds, 46,067 (82%) were included in quantitation.

2.10. Statistical analysis - Multiomics

We deployed a novel distance weighted discrimination (DWD) [12] with multiway components in the R environment (V4.0) through the MultiwayClassification package (V1.0) in order to categorized the association of proteomic, hematological, and metabolomic profiles with ID condition, age, and fluid compartment. Through the use of multiway DWD [13] we are able to consider multiple weighted sums capable of discriminated the sample classes across time.

This methodology could identify a weighted sum for both metabolomic and proteomic data thereby discriminating ID from IS profiles at both time points (4- and 6- months) for all four major datasets: CSF metabolome, serum metabolome, CSF proteome, and serum proteome. Cross-validation was achieved through leave-one-out analysis with the data for one subject removed and the remaining subjects used to generate feature weights that were then tested against the removed subject profile to generate the distance score. This analysis was repeated with each subject scored according to this process. T-tests were performed at 4- and 6-month time points to validate the DWD scores between ID and IS conditions. Repeated measures ANOVA was deployed protein and metabolite with categorical variables for ID status (ID/IS), time (4 months or 6 months), an interaction effect, and random effects for each dataset. False discovery rate adjustments [14] were calculated throughout the tests as previously described thereby obtaining Q values through pairwise comparison methods. To further understand the physiological significance of the observed data, proteins and metabolites were considered individual analytes curated through the Ingenuity Pathway Analyses (IPA) (Qiagen, Hilden, Germany) molecular pathways modeling software.

Ethics Statements

The 19 monkeys in this study were derived from a colony at the University of Wisconsin, the Institutional Animal Care and Use (IACUC) approved all procedures, protocol number: L005931. All infants were born from singleton full-term pregnancies born to different mothers and raised with a standardized environment. Rhesus monkeys raised in this model have low iron stores at birth and continue to experience a continual loss of iron due to rapid growth and solely feeding on their mother's milk, which is a poor source of dietary iron. CSF and blood samples were collected in a sterile manner using approved sedation and clinical techniques. All animal treatment and care complied with the ARRIVE guidelines and were in accordance with the National Institutes of Health guide for the care and use of laboratory animals (NIH Publications No. 8023, revised 1978).

CRediT Author Statement

Brian Sandri: Conceptualization, Methodology, Investigation, Resources, Data Analysis, Writing – Original Draft, Writing – revised draft, Visualization, Data Curation, and Project Administration; **Jonathan Kim:** Data Analysis, Software, Visualization; **Gabriele Lubach:** Conceptualization, Methodology, Validation, Investigation, Resources; **Eric Lock:** Writing – original draft, Data Analysis, Software, Visualization; **Candace Guerrero:** Methodology, Investigation; **LeeAnn Higgins:** Analysis; **Todd Markowski:** Methodology, Investigation; **Pamela Kling:** Writing – original draft, Conceptualization **Michael Georgieff:** Writing – original draft, Conceptualization, Funding Acquisition; **Christopher Coe:** Writing – Original Draft, Conceptualization, Funding Acquisition; **Raghavendra Rao:** Writing – original draft, Conceptualization, Funding Acquisition, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

[Proteomics \(Original data\)](#) (PRIDE).

Serum Proteomics (Original data) (PRIDE).
CSF Metabolomics (Original data) (MetaboLights).
Serum Metabolomics (Reference data) (MetaboLights).

Acknowledgments

Funding: This work was supported by grants NIHRO1HD089989, NIHRO1HD080201, NIHRO1HD057064, NIGMSR01GM130622, and NIHRO1HD39386. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.dib.2022.108591](https://doi.org/10.1016/j.dib.2022.108591).

References

- [1] B.J. Sandri, A. Kaplan, S.W. Hodgson, M. Peterson, S. Avdulov, L. Higgins, T. Markowski, P. Yang, A.H. Limper, T.J. Griffin, P. Bitterman, E.F. Lock, C.H. Wendt, Multi-omic molecular profiling of lung cancer in COPD, *Eur. Respir. J.* (2018) 52, doi:[10.1183/13993003.02665-2017](https://doi.org/10.1183/13993003.02665-2017).
- [2] B.J. Sandri, L. Masvidal, C. Murie, M. Bartish, S. Avdulov, L. Higgins, T. Markowski, M. Peterson, J. Bergh, P. Yang, C. Rolny, A.H. Limper, T.J. Griffin, P.B. Bitterman, C.H. Wendt, O. Larsson, Distinct cancer-promoting stromal gene expression depending on lung function, *Am. J. Respir. Crit. Care Med.* 200 (2019) 348–358, doi:[10.1164/rccm.201801-0080OC](https://doi.org/10.1164/rccm.201801-0080OC).
- [3] B.J. Sandri, G.R. Lubach, E.F. Lock, M.K. Georgieff, P.J. Kling, C.L. Coe, R.B. Rao, Early-life iron deficiency and its natural resolution are associated with altered serum metabolomic profiles in infant rhesus monkeys, *J. Nutr.* (2019), doi:[10.1093/jn/nxz274](https://doi.org/10.1093/jn/nxz274).
- [4] B.J. Sandri, G.R. Lubach, E.F. Lock, P.J. Kling, M.K. Georgieff, C.L. Coe, R.B. Rao, Correcting iron deficiency anemia with iron dextran alters the serum metabolomic profile of the infant Rhesus Monkey, *Am J Clin Nutr* 113 (2021) 915–923, doi:[10.1093/ajcn/nqaa393](https://doi.org/10.1093/ajcn/nqaa393).
- [5] B.J. Sandri, J. Kim, G.R. Lubach, E.F. Lock, C. Guerrero, L. Higgins, T.W. Markowski, P.J. Kling, M.K. Georgieff, C.L. Coe, R.B. Rao, Multiomic profiling of iron-deficient infant monkeys reveals alterations in neurologically important biochemicals in serum and cerebrospinal fluid before the onset of anemia, *Am J Physiol Regul Integr Comp Physiol* 322 (2022) R486–R500, doi:[10.1152/ajpregu.00235.2021](https://doi.org/10.1152/ajpregu.00235.2021).
- [6] M. Pietrowska, A. Wlosowicz, M. Gawin, P. Widlak, MS-Based proteomic analysis of serum and plasma: problem of high abundant components and lights and shadows of albumin removal, *Adv Exp Med Biol* 1073 (2019) 57–76, doi:[10.1007/978-3-030-12298-0_3](https://doi.org/10.1007/978-3-030-12298-0_3).
- [7] R. Millioni, S. Tolin, L. Puricelli, S. Sbrignadello, G.P. Fadini, P. Tessari, G. Arrigoni, High abundance proteins depletion vs low abundance proteins enrichment: comparison of methods to reduce the plasma proteome complexity, *PLOS ONE* 6 (2011) e19603, doi:[10.1371/journal.pone.0019603](https://doi.org/10.1371/journal.pone.0019603).
- [8] L. Käll, J.D. Storey, W.S. Noble, Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry, *Bioinformatics* 24 (2008) i42–48, doi:[10.1093/bioinformatics/btn294](https://doi.org/10.1093/bioinformatics/btn294).
- [9] A.I. Nesvizhskii, A. Keller, E. Kolker, R. Aebersold, A statistical model for identifying proteins by tandem mass spectrometry, *Anal Chem* 75 (2003) 4646–4658, doi:[10.1021/ac0341261](https://doi.org/10.1021/ac0341261).
- [10] I.P. Shadforth, T.P.J. Dunkley, K.S. Lilley, C. Bessant, i-Tracker: for quantitative proteomics using iTRAQ, *BMC Genomics* 6 (2005) 145, doi:[10.1186/1471-2164-6-145](https://doi.org/10.1186/1471-2164-6-145).
- [11] A.L. Oberg, D.W. Mahoney, J.E. Eckel-Passow, C.J. Malone, R.D. Wolfinger, E.G. Hill, L.T. Cooper, O.K. Onuma, C. Spiro, T.M. Therneau, H.R. Bergen, Statistical analysis of relative labeled mass spectrometry data from complex samples using ANOVA, *J. Proteome Res* 7 (2008) 225–233, doi:[10.1021/pr700734f](https://doi.org/10.1021/pr700734f).
- [12] X. Qiao, H.H. Zhang, Y. Liu, M.J. Todd, J.S. Marron, Weighted distance weighted discrimination and its asymptotic properties, *J. Am. Stat. Assoc.* 105 (2010) 401–414, doi:[10.1198/jasa.2010.tm08487](https://doi.org/10.1198/jasa.2010.tm08487).
- [13] T. Lyu, E.F. Lock, L.E. Eberly, Discriminating sample groups with multi-way data, *Biostatistics* 18 (2017) 434–450, doi:[10.1093/biostatistics/kxw057](https://doi.org/10.1093/biostatistics/kxw057).
- [14] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple TESTING, *J. R. Stat. Soc. Ser. B (Methodological)*, 57 (1995) 289–300.