

"This is the peer reviewed version of the following article: [Optimized plasma sample preparation and LC-MS analysis to support large-scale proteomic analysis of clinical trial specimens: Application to the Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) trial (2023)], which has been published in final form at <https://doi.org/10.1002/prca.202200106>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited."

Optimized plasma sample preparation and LC-MS analysis to support large-scale proteomic analysis of clinical trial specimens: Application to the Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) trial

Matthew B. O'Rourke*^{1,2,3}, Andrzej S. Januszewski*⁴, David R. Sullivan⁴, Imre Lengyel⁵, Alan J. Stewart⁶, Swati Arya⁶, Ronald Ma⁷, Sanjeev Galande⁸, Anandwardhan A. Hardikar⁴, Mugdha V. Joglekar⁴, Anthony C. Keech**⁴, Alicia J. Jenkins**⁴, Mark P. Molloy**^{1#}

¹ Bowel Cancer & Biomarker Lab, School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney

² Centre for Inflammation, Centenary Institute, Camperdown, NSW, 2050, Australia

³ School of Life Sciences, Faculty of Science, University of Technology Sydney, Ultimo, NSW, 2007, Australia

⁴ NHMRC Clinical Trials Centre, Faculty of Medicine and Health, The University of Sydney

⁵ Wellcome-Wolfson Institute for Experimental Medicine, School of Medicine, Dentistry and Biomedical Science, Queen's University Belfast, Belfast, UK

⁶ School of Medicine, University of St Andrews, Fife, UK

⁷ Chinese University of Hong Kong, Hong Kong, China

⁸ Indian Institute of Science Education and Research, Pune, India

* - equal first authors

** - equal senior authors

#Correspondance:

Prof. Mark P. Molloy

Email: m.molloy@sydney.edu.au

Keywords: Plasma, mass spectrometry, proteomics, biomarker, diabetes, fenofibrate

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/prca.202200106](https://doi.org/10.1002/prca.202200106).

This article is protected by copyright. All rights reserved.

Accepted Article

Statement of clinical relevance

Plasma biomarkers hold great promise to improve diagnosis and influence aspects of clinical practice, but their discovery is hampered by analytical challenges in study design and specimen analysis. Clinical studies require large-scale analysis of hundreds of specimens, so fit-for-purpose proteomic workflows that balance throughput and cost while ensuring quantitative rigor are required. Our aim was to evaluate human plasma sample suitability and preparation procedures for LC-MS which addresses these challenges. We determined that using neat, citrated plasma with online delipidation over 45 minutes of LC-MS sample acquisition time was a cost-effective way to quantify 172 abundant plasma proteins with high quantitative reproducibility. We report the intra-batch variability across 65 batches to be less than 2%. This demonstrates a viable approach to support large-scale clinical biomarker studies using human plasma.

Abstract

Purpose: Robust, affordable plasma proteomic biomarker workflows are needed for large-scale clinical studies. We evaluated aspects of sample preparation to allow LC-MS analysis of more than 1500 samples from the Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) trial of adults with type 2 diabetes.

Methods: Using LC-MS with data-independent acquisition we evaluated four variables: plasma protein depletion, EDTA or citrated anti-coagulant blood collection tubes, plasma lipid depletion strategies and plasma freeze-thaw cycles. Optimised methods were applied in a pilot study of FIELD participants.

Results: LC-MS of undepleted plasma conducted over a 45 min gradient yielded 172 proteins after excluding immunoglobulin isoforms. Cibachrome-blue-based depletion yielded additional proteins but with cost and time expenses, while immunodepleting albumin and IgG provided few additional identifications. Only minor variations were associated with blood collection tube type, delipidation methods and freeze-thaw cycles. From 65 batches involving over 1500 injections, the median intra-batch quantitative differences in the top 100 proteins of the plasma external standard was less than 2%. Fenofibrate altered seven plasma proteins.

Conclusions and Clinical Relevance: A robust plasma handling and LC-MS proteomics workflow for abundant plasma proteins has been developed for large-scale biomarker studies that balances proteomic depth with time and resource costs.

Introduction

Proteomics is increasingly used in clinical research to identify novel biomarkers for the prediction of health outcomes, including responses to interventions, suggest mechanisms of disease prediction and prevention, and facilitate the development of novel therapeutics^[1,2]. Such clinical studies are often extensive, and from heterogeneous subjects with potential impacts on proteomics analyses due to sample collection and handling^[3-5]. Here, we set out to develop a robust, cost-effective Liquid Chromatography - Mass Spectrometry (LC-MS) proteomic workflow to analyse large numbers of clinical plasma samples. Consideration was given to minimising the number of sample processing steps to control pre-analytical variance while achieving detailed proteomic coverage of abundant plasma proteins. Given the large scale of many clinical studies, including the one which we will undertake, there is a need to establish a robust method for data quality for both sample preparation

and subsequent mass spectrometric analysis, which would generally be acceptable for reporting undertaken clinical trials data.

Mass spectrometry-based proteomic analysis is often performed in discrete, relatively small studies where cohorts of samples are prepared in a single instance and then measured in the mass spectrometer in a single batch. This ensures that variation introduced through sample handling techniques and instrument performance are kept at a minimum^[6]. However, in the context of analysing thousands of plasma samples from clinical observational studies and trials, this is not feasible, requiring us to develop quality control (QC) methods and standardised protocols to ensure that each batch of samples could be obtained within a pre-defined quantitative threshold. Some reports have provided blueprints for proteomic biomarker discovery workflows compatible with large-scale studies. For example, Bruderer et al^[6]. established a capillary LC-MS/MS method with 300 µm I.D. columns to quantitate more than 1500 plasma samples, while Geyer et al^[7]. quantified 319 plasma proteomes in quadruplicate. These papers provide excellent examples of specific LC-MS workflows for plasma proteomics, but they remain demonstration studies for what can be achieved at scale in specialist labs and have not been widely implemented elsewhere. Moreover, important plasma sample characteristics and processing variables relevant to clinical studies were not explored in these reports.

This paper aims to share our experience in developing an LC-MS proteomic workflow to enable robust quantitation of plasma proteins from extensive clinical studies. In designing the workflow, we assessed four critical variables related to plasma collection and processing for proteomics LC-MS analyses: (1) depletion of abundant plasma proteins compared with neat plasma analysis, (2) approaches for plasma delipidation, (3) EDTA plasma compared with citrate plasma, and (4) impact of sample freeze-thaw cycles. We report a proof-of-concept pilot study that includes using the optimised workflow with plasma samples from 15 Fenofibrate Intervention and Event-Lowering in Diabetes (FIELD)^[8] trial participants pre-and post- 16 weeks of fenofibrate treatment, as well as our

subsequent quantitative results using an external plasma standard measured over an 18-month period by DIA LC-MS across 65 batches of 24 plasma samples per batch.

Materials and Methods

Materials

A citrated plasma sample was used as an external standard for batch QC assessment (Sigma Aldrich, North Rocks, Sydney, Australia). Methyl-tert-butyl-ether (MTBE), Methanol, Acetonitrile (ACN), Acetone, ethyl acetate, ammonium hydroxide, formic acid (FA), Tris-HCl pH8.0, chloroacetamide (CLA), tris(2-carboxyethyl)-phosphine (TCEP), monopotassium phosphate, triethylammonium bicarbonate (TEAB), chloroacetamide, sodium deoxycholate (SDC) and trifluoroacetic acid (TFA) were all of ultra-pure or LCMS grade and purchased from Sigma Aldrich North Rocks, Sydney, Australia. ProteoPrep® Immunoaffinity Albumin and IgG depletion Kits were from Sigma Aldrich North Rocks, Sydney, Australia). Aurum Affi-Gel Blue Mini Columns were from BioRad (Hercules, CA, USA). Sequencing grade trypsin was from Promega (Madison, WI, USA), and iRT peptides were from Biognosys (Schlieren, Switzerland). SDB-RPS disks were from Affinisep (Petit-Couronne, France).

Subjects

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Human Research Ethics Committee of The University of Sydney (08-2007/10216). All participants provided written informed consent.

For the experiments comparing abundant protein depletion, lipid depletion methods, blood collection tubes and assessing effects of sample freeze-thaw, we obtained plasma from five healthy volunteers aged between 27 and 59 yrs, three men, two women, mean age 40 years (± 13 years standard deviation).

In a pilot study, citrate plasma samples from 15 FIELD trial participants pre-and post-16-week run-in during the last 6-weeks of which once daily fenofibrate tablets (200 mg comiconised) were taken. Subsequently, we report the inter-assay CVs of the external plasma standard used for batch QC in 72 runs of batches of 24 FIELD samples.

Albumin and Immunoglobulin Depletion

The Australian Proteome Analysis Facility (Macquarie University, Sydney, Australia) performed the depletion of abundant proteins as per manufacturer's instructions. For immunoaffinity depletion of albumin and IgG, ProteoPrep immunoaffinity depletion columns were equilibrated in loading buffer (20 mM sodium phosphate, 150 mM NaCl, pH 7.4) and 50 μ l of neat plasma added. Samples were incubated in the columns for 15 min at room temperature, followed by washing twice, with 400 μ l loading buffer. The washes were combined and stored at -80°C as these contained the unbound plasma proteins. The bound fraction was discarded. For Cibacron Blue based depletion, an Aurum Affi-gel Cibacron Blue F3GA agarose gel microcolumn was used following the manufacturer's instructions. Briefly, columns were drained, and washed with loading buffer (20 mM sodium phosphate, pH 7.4). 125 μ l of neat plasma diluted in 375 μ l of loading buffer was then added to the column and spun using a benchtop centrifuge at 10,000 \times g for 20sec. This procedure was repeated with another 400 μ l of loading buffer, and the eluates were combined and stored at -80°C.

Processing of un-delipidated plasma

Plasma tryptic digests were prepared by transferring 10 μ l of plasma to a 1.5 ml Eppendorf tube and adding 800 μ l of 1% w/v SDC in 100 mM TEAB and vortexed. 100 μ l of diluted plasma was then removed to a fresh tube, and 1 μ l of 1 M TCEP was added together with 4 μ l of 1 M CLA. Samples were boiled at 95°C for 10 mins. Next, μ g of trypsin was added, and samples were digested overnight at 37°C.

Following digestion, 1 μ l of FA was added to precipitate the SDC, and the tubes were centrifuged at 18,000 \times g in an Eppendorf benchtop microfuge for 5 mins. The resulting supernatant was then desalted using the STAGE TIP desalting method below.

Offline De-lipidation

Three different methods of offline de-lipidation were tested.

Chloroform-methanol extraction

Plasma (10 μ l) was aliquoted into a clean 1.5ml Eppendorf tube, and then 50 μ l of cold MilliQ H₂O was added. Next, 75 μ l of 1 M monopotassium phosphate, 125 μ l of methanol, 375 μ l of chloroform and 150 μ l of cold water were added to the plasma, followed by 2 min of high-speed vortexing.

Samples were then centrifuged at 18,000 RCF to remove the aqueous and organic phases (top and bottom liquid fractions). The tube was then dried in a vacuum desiccator for 5 min to remove any residual solvent or water. The protein pellet was then re-suspended in 200 μ l of 1% SDC 100 mM TEAB, and a 10 μ l aliquot was removed for BCA protein assay. The sample was processed with trypsin and peptides desalted as described above.

Acetone Precipitation

Acetone precipitation of proteins was performed by taking 10 μ l of plasma and adding 1 ml of ice-cold acetone. Samples were then left at -20°C for 1h before centrifuged at 18,000 \times g for 5 minutes to pellet the precipitated protein. 200 μ l of 1% SDC 100 mM TEAB was added to solubilise the protein pellet before 10 μ l was removed for BCA protein assay. The sample was processed with trypsin and peptides desalted as described above.

Methyl-Tert-Butyl Ether extraction

Neat plasma (84 μ l) was transferred into a 1.5 ml tube, and then 100 μ l of methanol and 340 μ l of MTBE were added. The sample was vortexed at high speed for 2 min, then centrifuged (18,000 \times g, 5

min) to separate phases before discarding both phases from the tube and drying the protein pellet. The dried pellet was re-solubilised in 200 μ l of 1% SDC 100 mM TEAB and 2 μ l of TCEP, 8 μ l of CLA added, and the sample boiled (10 minutes at 95°C). Trypsin was added in a 1:100 ratio and digested overnight before STAGE Tip desalting.

Stage tip processing

As described previously, a combined Stage tip digestion/desalting/delipidation workflow was conducted^[9]. Following tryptic digestion, 250 μ l of 99% ethyl acetate, 1% TFA was added to each sample and vortexed for 10 seconds. Samples were then processed in a multiplex batch utilising a 3D printed apparatus described in Harney et al.^[10].

Freeze-thaw cycles

EDTA plasma samples (1.2 ml in triplicates) from three volunteers were aliquoted (200 μ l) into six separate tubes and frozen at -80°C. Five tubes per subject were then removed from the freezer and allowed to thaw on a roller mixer at room temperature for 30 mins, then refrozen at -80°C. Samples remained at -80°C for one day before the next cycle. This procedure was continued stepwise until the final tubes were frozen and thawed six times. Each sample was prepared using the Stage tip, method and LCMS data collected in DDA mode as described below.

Blood collection tubes

Matched K2 EDTA and sodium citrate plasma from five volunteers were processed as per the Stage tip methodology described above. LCMS data was collected in data-independent acquisition mode as described below.

Liquid chromatography Mass spectrometry (LCMS) analysis

LC-MS was conducted with 2 μ g plasma tryptic digest, separated using a self-packed 150 μ m I.D. x 15 cm column containing 1.9 μ m C18 beads (ReproSil-Pur C18-AQ) on a 45 min gradient to a maximum

of 30% B (80% ACN 0.1% FA), (63 min method runtime) using an Ultimate 3000 RSLCnano system (Thermo Scientific) operating at capillary flow rate of 1.6 $\mu\text{L}/\text{min}$. Data was acquired using a Q Exactive HF mass spectrometer (Thermo Scientific).

For experiments investigating impacts of abundant protein depletion, de-lipidation, and freeze-thaw cycles, a Top15 data-dependent acquisition (DDA) mode was used on Q Exactive HF mass spectrometer, employing using the same chromatography, columns and run-length as DIA experiments. The following instrument settings were applied for DDA experiments: MS1 - AGC 3e6, Resolution 60K, Scan range 300-1650 m/z . MS2 -AGC 1e5, Resolution 15K, loop count 15.

For samples prepared using the Stage Tip workflow, data-independent acquisition (DIA) mode using variable m/z windows over the 300-1600 m/z range was used as previously described^[9].

Proteomic data analysis

For LC-MS data acquired using DDA, raw files were searched with MaxQuant V1.6.5 against the human proteome FASTA database (UP000005640, 20,350 entries) using the following settings: 1% FDR, reverse decoy mode, min peptides 1, FTMS mass tolerance 20 ppm, missed cleavages set to max 2. Modifications: oxidation of methionine, variable and carbamidomethyl (C) fixed, LFQ enabled with default settings.

LC-MS data acquired using DIA were analysed using Spectronaut Pulsar X v12.0.2 (Biognosys, Switzerland) following our methods using an enhanced spectral library derived from publicly accessible human plasma datasets as we previously described^[9] Raw files were searched in Spectronaut against the spectral library^[9] with the following settings: XIC extraction dynamic, precursor Q-value cutoff value 0.01, protein Q-value cutoff value 0.01. Peptides were quantified by a minimum of three transitions, protein quantification by top $n=5$ peptides. Data were then exported as a “Q-value Sparse” pivot report matrix with normalized protein areas. Missing values were represented with “filtered”. Perseus V1.6.1.2^[11] was used to log₂ transform data, and missing values

were imputed using the following strategy: random values within 30% of the median of the normal distribution were selected, followed by a reduction by three standard deviations. To compare the two groups, Student T-tests were used with $p < 0.05$, considered significant. To compare multiple groups, single factor ANOVA was used with $p < 0.05$ considered significant. For intra-batch analyses the Top 100 most abundant proteins from the external plasma standard were assessed for their pre-batch and post-batch area differences. The median area differences for each batch were recorded.

Results

Abundant protein depletion

It is well-established that the high concentration of some proteins in plasma, such as immunoglobulins and albumin, limits the depth of proteome analysis that can be achieved in a single run LC-MS experiment^[12,13]. Affinity-based depletion methods have been used to remove these highly abundant proteins to address this high dynamic range issue^[14]. However, for large-scale studies as proposed here, added costs and time delays arising from using depletion would need to be justified. Therefore, we assessed the depth of proteome analysis that could be achieved with neat plasma compared with two affinity depletion approaches; specifically immunodepletion of albumin and IgG (DepS) and non-specific Cibacron blue dye-based depletion (Dep G). Using a 45 min LC gradient and single-shot Top 15 DDA method, we detected a core group of 146 proteins in all three conditions. The use of Dep G revealed an additional 80 unique proteins, Dep S had seven unique proteins, while neat plasma showed 19 unique proteins (Fig 1 and Supplementary Data 1). When this was further investigated, Dep G resulted in a high number of additional immunoglobulin protein identifications, with 27 being attributable to heavy and light chain variable regions. A smaller subset of proteins was found when immunoglobulins were discounted from depleted and non-depleted data (See Figure 1B). Immunodepletion with DepS offered no tangible advantage to increasing proteome access compared with use of neat plasma. Dep G provided access to 40 additional proteins over neat plasma, although this comes with additional costs and sample handling

steps. We detected Apolipoprotein AI (Apo-AI) in neat plasma, and it appeared to be depleted in the two other methods. This is a relevant finding as Apo-AI, a key component of High-Density Lipoprotein (HDL), is a biomarker of the lipid drug fenofibrate activity. We, therefore, considered that analysis using neat plasma is the most facile approach that affords time and cost advantages and would minimise quantitative variability due to sample handling. This came at the expense of 40 fewer proteins that could be detected using DepG.

Plasma De-lipidation methods

Given that plasma contains numerous lipid species that may accumulate on stationary phase^[15] and having selected neat plasma as the starting matrix, we tested various delipidation procedures with the objective to maximise LC column and MS system stability needed to analyse hundreds of clinical samples. We tested four delipidation methods: acetone, methyl-tert-butyl ether, methanol Chloroform (MEOHCL), and ethyl acetate integrated into the STAGE tip purification protocol^[16].

We used DDA to assess the recovery of proteins from triplicate preparations of each method (Figure 2). Unsurprisingly, we observed a high overlap amongst the methods, although interestingly, ethyl acetate integrated with STAGE tip desalting led to the recovery of 26 unique proteins.

Similarly, the three other extraction methods recovered 23 proteins absent from the ethyl acetate STAGE tip purification (Supplementary Data 2). This demonstrates that the selection of de-lipidation strategies has a minor effect on protein identifications. Of the methods tested herein, we chose to adopt the ethyl acetate integrated workflow for its simplicity in manual handling that is compatible with 96-well plate-based processing of samples.

Comparison of blood collection tubes

Most plasma proteomic studies utilise EDTA plasma, and there is little information regarding comparisons with sodium citrate plasma. This is relevant to the FIELD study as only citrate plasma

was available for the proteomic analyses. There were no significant differences in the protein quantitation between plasma sample types, except five proteins (Table 1, Supplementary data 3).

We observed higher abundance of haemoglobins in EDTA plasma, while citrate plasma showed higher abundances of fibrinogen and fibronectin^[17].

Effects of multiple freeze-thaw cycles on stored plasma

We conducted sample freeze-thaw events to assess the effects of six freeze-thaw cycles on the quantitation of proteins. Overall, there was no significant variation in protein abundance between the cycles (ANOVA on 270 proteins was non-significant) (see supplementary data 4). To further illustrate this, Figure 3 shows the relative abundance of apolipoproteins, displaying minimal variance across six freeze-thaw events.

A pilot study of FIELD samples

We applied the above determined optimal conditions, using neat, sodium citrate plasma with ethyl acetate STAGE tip delipidation and DIA LC-MS strategy to a cohort of 15 FIELD trial participant's plasma samples obtained before and after active run-in (6 weeks fenofibrate treatment). 222 proteins were detected across each of the 30 plasma samples (Supplementary Table 5). A differential analysis comparing pre-and post-treatment groups identified seven significantly differentially regulated proteins ($P < 0.05$) (see Table 2). The major HDL protein constituents, apolipoproteins AI and A (Apo-AI and Apo-AII), showed a non-significant increase post-6-weeks of fenofibrate in this small cohort. There was a significant decrease in the abundance of kallistatin ($\log_2FC -0.13$, $p=0.04$), a serine proteinase inhibitor that some studies suggest is a biomarker for microvascular complications in diabetes patients^[18].

Quantitative reproducibility in large scale analyses

The above methods were applied to a cohort of FIELD trial participants (1,560 samples data to be reported elsewhere). Here we report a summary of the intra-batch quantitative reproducibility associated with LC-MS analysis of the external QC plasma standard. The external plasma standard was acquired pre- and post- 65 batches of 24 samples and the difference in areas of the Top 100 most abundant proteins used to assess batch acceptability. The median intra-batch difference was 1.6%, with the medians for 95% of the data within the range -9% to 11%. (Figure 4 and Supplementary Data 6).

Discussion

Our objective was to evaluate plasma sample handling workflows to optimise a protocol for the large-scale analysis of human plasma for clinical studies which could be used for subsequent analyses of FIELD trial samples. Key factors evaluated included the reproducible quantitative measurement of protein data with appropriate quality control measures, high throughput capability and a simplified sample preparation protocol to minimise variability caused by operator processing. To achieve this, we explored the capabilities of several sample preparation techniques. We applied quality control measures to ensure stable instrument performance and low sample processing variability that could be achieved across a cohort of over 1560 human samples.

Removal of highly abundant plasma proteins

It has been previously established that there is an advantage to using top N abundant protein depletion for increasing the detectable proteome in plasma when conducting conventional nanoLC-MS experiments ^[19]. Our investigations using 45min LC gradient with capillary flow and with DDA mode acquisition demonstrated a moderate increase in the number of detected proteins can be achieved following non-selective partitioning using Cibacron blue dye (Fig 1). At the same time, specific depletion of albumin and IgG afforded no significant benefit in our LC-MS/MS workflow. Several factors may contribute to this finding including incomplete depletion given the plasma

concentration of albumin and IgG is extremely high and there is considerable inter-patient heterogeneity in disease states, non-specific binding causing inter-sample quantitative variability, while other detractive aspects include additional sample handling steps and high reagent costs. Taken together, we failed to find sufficient evidence that plasma depletion combined with single-shot LC-MS acquisition offered benefit compared to the use of neat plasma in our workflow designed to robustly quantitate the most highly abundant plasma proteins in thousands of samples. The consumable and data acquisition cost of running a single plasma sample in our laboratory is currently ~ \$20AUD. This includes a budget for instrument time, reagents and related consumables. When applied to the FIELD cohort of 1690 samples (1,560 plasma samples + 130 QC samples), the projected cost of the study was exceeding \$30,000AUD. The DepG depletion at \$15AUD per sample would add an additional \$23,400AUD in experimental costs for this study without a considerable gain in proteomic depth. Non-specific depletion methods such as DepG may further compromise an investigation by depleting potentially important biomarkers, such as Apo-AI in our case, a known biomarker for fenofibrate response and potential modulate of diabetes complications. In this study, we demonstrated that undepleted, neat plasma could be effectively analysed by LC-MS on a large-scale while maintaining high-quality instrument performance, and this strategy avoids complications associated with partial or incomplete depletions. The limitation to this approach is proteome depth, which can only be addressed using more extensive sample fractionation methods and come with added time and cost impacts.

Evaluation of de-lipidation methods

Human plasma contains as much as 471 mg/dL weight of lipids primarily sterol lipids, glycerophospholipids, glycerolipids, sphingolipids amongst other classes and species^[20]. Moreover, large, hydrophobic lipids may be retained on reversed-phase LC columns and contribute to poor reproducibility^[15] and column blocking^[10] so we rationalised we should deplete these to meet our objective for a robust and stable LC-MS system to quantitate thousands of clinical specimens. Three

different offline lipid removal methods were trialed, including one-step protein precipitations using acetone, and, more complex workflows, which separate lipids and proteins by inducing phases in the aqueous and organic layers. A drawback of the two-phase methods is the need to carefully pipette around the protein pellet, which sits at the interface of both phases, to remove the aqueous and organic phases. The use of offline methods of de-lipidation also creates a need to perform offline desalting. The manual handling required to perform these two separate steps of the workflow adds to the potential for introducing quantitative bias. We also trialed an online method of de-lipidation and de-salting in a single multiplexed workflow. The Mann group published the Stop And Go Extraction (STAGE tip) method in 2003^[16] is now a well-established method for desalting peptide samples. Hanley *et al.*,^[10] further adapted it for high-throughput analysis by using a custom-designed 3D printed tip holder in microplate format, enabling parallel analysis of 96 samples.

We found that the online de-lipidation method was the most convenient for handling multiple samples. With the aid of a combination style repeating pipette (e.g. Eppendorf Combipette M4), the ability to prime, wash and elute peptides was rapid and reproducible. One downside of the online method is that the solvents used for washing and elution are highly volatile and hazardous, meaning that venting had to be employed to remove vapours following centrifugation. The performance of our workflow without a delipidation step was not accessed.

Blood anti-coagulants

Most plasma proteomic studies deploying LC-MS utilise samples collected into EDTA anti-coagulant tubes^[21,22]. There is a paucity of proteomic studies examining plasma proteins collected from other types of anti-coagulant tubes. This is relevant to the FIELD study where citrated plasma samples are available for specimens collected over 12 years ago^[8]. We performed a pairwise comparison of matched citrate and EDTA preserved plasma from a group of donors and observed minimal quantitative differences except for haemoglobin, fibrinogen and fibronectin levels. The suspected antagonistic nature of fenofibrate on cardiovascular complications has been linked to fibrosis related

cardiovascular remodelling^[23]; therefore the elevated detection of these clotting related proteins when using citrate was deemed useful. Contrastingly the use of EDTA has been previously reported to increase the osmotic fragility of erythrocytes leading to increased haemolysis^[24]. When EDTA was used, we found that elevated levels of haemoglobin were encountered, consistent with minor red blood cell haemolysis, which was absent in the citrated samples. This data confirmed the utility of sodium citrate plasma for LC-MS proteomic analyses.

Effects of multiple freeze thaw cycles on stored plasma

One of the considerations of this study was the effect of multiple freeze-thaw cycles. Parallel mRNA analyses were also planned for the specific cohort of chosen samples, so it was important to determine which analysis would need to take priority. RNA is typically less stable and prone to degradation when frozen and thawed multiple times; therefore the effects of various freeze-thaw cycles were evaluated to see the impact on the detectable proteome. A subset of (non-FIELD study) donor plasma was collected, and aliquoted, before being frozen and thawed six times. After observing no significant global variation, we chose a subset of apolipoproteins and complement proteins for this analysis. Our results showed that there was little effect on the abundance or detection of these markers of fenofibrate action. We attribute this to the denaturing conditions used for protein digestion, meaning that if proteins were degraded by freeze-thaw, this might not have been detectable in the subsequent LCMS peptide analysis. We note that it has previously been reported that the concentration of plasma proteins remains stable when specimens are stored at -80°C^[25].

FIELD sample pilot study

To evaluate the effectiveness of the optimised workflow, a pilot study was conducted using stored FIELD study citrate plasma samples from patients treated with fenofibrate for six weeks, with fasted venous blood being taken pre-and post-treatment. Minor increases in the major HDL lipoproteins

ApoA-I and Apo-AII were observed after 6-weeks of fenofibrate, although this trend did not reach statistical significance in this small cohort. Despite the small cohort size, several proteome changes were observed, including a change in the amount of circulating kallistatin reported previously as being associated with cardiovascular complications in people with diabetes^[18]. While the small sample size of this pilot cohort means that these preliminary conclusions need to be further validated, it clearly demonstrated a fit-for-purpose analytical workflow for studying FIELD samples.

Evaluation of quality control measures over time

We utilised book-ended batching with an external plasma standard to monitor batch quality. We used the median area difference for the Top 100 proteins, aiming to achieve <15% intra-batch difference. On the occasions when this metric was exceeded, it prompted us to examine causes which were primarily due to declining MS detection sensitivity, or occasionally because of human errors in sample preparation pipetting. The sample preparation workflow itself is robust as we never experienced a failure using that procedure. For the LC-MS using our setup with capillary flow it was typical to achieve 150 injections of 2 µg raw plasma digest before the system sensitivity deteriorated and failed batch QC. This was rectified by replacing the self-packed 150 µm LC column, changing the transfer capillary and cleaning the MS front-end ion optics to restore sensitivity, a process which would consume 1-3 days. This was a necessary process to enable quantitation of less abundant proteins, which would otherwise result in missing data.

In summary, we have reported a facile and robust human plasma sample preparation and LC-MS workflow suitable for large-scale proteomic analysis using neat plasma. The use of neat plasma has the advantage of avoiding complications due to additional sample processing steps, as well as time and reagent cost savings. We envisage this workflow which balances proteomic depth against cost and time can be applied to most clinical proteomic studies using human plasma where quantitative analysis of the major plasma proteins is the objective.

Associated Data

Mass spectrometry data is available via ProteomeXchange with the identifier PDX029732.

Competing interests

The authors declare no competing interests

Funding

This work was performed by funding from The University of Sydney (CIA Jenkins) and funds provided by the National Health and Medical Research Council (Australia) APP1147897

Authors' contributions

MBOR: manuscript draft, carried out sample preparation, performed mass spectrometry and data analysis, **ASJ:** critically reviewed data and manuscript draft, **DRS:** conceptualisation, funding acquisition, access to specimens and study oversight, **IL:** study design, validation, **AJS:** methodology, validation, **SA:** investigation, methodology, **RCM:** funding acquisition, **SG:** funding acquisition, **AAH:** funding acquisition, supervision, **MVJ:** supervision, **ACK:** conceptualisation, funding, access to specimens and study oversight, **AJJ:** conceptualisation, funding, sample provision, manuscript revision, **MPM:** conceptualisation, methodology, funding acquisition, critical review of data, manuscript revision.

References

- [1] Geyer, P. E., Holdt, L. M., Teupser, D., & Mann, M. (2017). Revisiting biomarker discovery by plasma proteomics. *Mol Syst Biol*, *13*(9), 942. doi: 10.15252/msb.20156297
- [2] Ehardt, H. A., Root, A., Sander, C., & Aebersold, R. (2015). Applications of targeted proteomics in systems biology and translational medicine. *Proteomics*, *15*(18), 3193-3208. doi: 10.1002/pmic.201500004
- [3] Pavlou, M. P., Diamandis, E. P., & Blasutig, I. M. (2013). The long journey of cancer biomarkers from the bench to the clinic. *Clin Chem*, *59*(1), 147-157. doi: 10.1373/clinchem.2012.184614
- [4] Ignjatovic, V., Geyer, P. E., Palaniappan, K. K., Chaaban, J. E., Omenn, G. S., Baker, M. S., . . . Schwenk, J. M. (2019). Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection to Achieving Translational Data. *Journal of proteome research*, *18*(12), 4085-4097. doi: 10.1021/acs.jproteome.9b00503
- [5] Liu, Y., Buil, A., Collins, B. C., Gillet, L. C., Blum, L. C., Cheng, L. Y., . . . Aebersold, R. (2015). Quantitative variability of 342 plasma proteins in a human twin population. *Mol Syst Biol*, *11*(1), 786. doi: 10.15252/msb.20145728
- [6] Bruderer, R., Muntel, J., Muller, S., Bernhardt, O. M., Gandhi, T., Cominetti, O., . . . Reiter, L. (2019). Analysis of 1508 Plasma Samples by Capillary-Flow Data-Independent Acquisition Profiles Proteomics of Weight Loss and Maintenance. *Mol Cell Proteomics*, *18*(6), 1242-1254. doi: 10.1074/mcp.RA118.001288
- [7] Geyer, P. E., Holdt, L. M., Teupser, D., & Mann, M. (2017). Revisiting biomarker discovery by plasma proteomics. *Molecular systems biology*, *13*(9), 942-942. doi: 10.15252/msb.20156297
- [8] Effects of long-term fenofibrate therapy on cardiovascular events in 9795 people with type 2 diabetes mellitus (the FIELD study): randomised controlled trial. (2005). *The Lancet*, *366*(9500), 1849-1861. doi: 10.1016/S0140-6736(05)67667-2
- [9] O'Rourke, M. B., Sahni, S., Samra, J., Mittal, A., & Molloy, M. P. (2021). Data independent acquisition of plasma biomarkers of response to neoadjuvant chemotherapy in pancreatic ductal adenocarcinoma. *J Proteomics*, *231*, 103998. doi: 10.1016/j.jprot.2020.103998
- [10] Harney, D. J., Hutchison, A. T., Hatchwell, L., Humphrey, S. J., James, D. E., Hocking, S., . . . Larance, M. (2019). Proteomic Analysis of Human Plasma during Intermittent Fasting. *Journal of proteome research*, *18*(5), 2228-2240. doi: 10.1021/acs.jproteome.9b00090
- [11] Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., . . . Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, *13*(9), 731-740. doi: 10.1038/nmeth.3901
- [12] Lee, P. Y., Osman, J., Low, T. Y., & Jamal, R. (2019). Plasma/serum proteomics: depletion strategies for reducing high-abundance proteins for biomarker discovery. *Bioanalysis*, *11*(19), 1799-1812. doi: 10.4155/bio-2019-0145
- [13] Geyer, Philipp E., Kulak, Nils A., Pichler, G., Holdt, Lesca M., Teupser, D., & Mann, M. (2016). Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Systems*, *2*(3), 185-195. doi: <https://doi.org/10.1016/j.cels.2016.02.015>
- [14] Zhao, Y., Chang, C., Qin, P., Cao, Q., Tian, F., Jiang, J., . . . Qian, X. (2016). Mining the human plasma proteome with three-dimensional strategies by high-resolution Quadrupole Orbitrap Mass Spectrometry. *Analytica Chimica Acta*, *904*, 65-75. doi: <https://doi.org/10.1016/j.aca.2015.11.001>
- [15] Han, X., & Gross, R. W. (2005). Shotgun lipidomics: electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples. *Mass Spectrom Rev*, *24*(3), 367-412. doi: 10.1002/mas.20023

- [16] Rappsilber, J., Ishihama, Y., & Mann, M. (2003). Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Analytical Chemistry*, 75(3), 663-670. doi: 10.1021/ac026117i
- [17] Skeppholm, M., Wallén, N. H., Blombäck, M., & Kallner, A. (2008). Can both EDTA and citrate plasma samples be used in measurements of fibrinogen and C-reactive protein concentrations? *Clin Chem Lab Med*, 46(8), 1175-1179. doi: 10.1515/cclm.2008.219
- [18] McBride, J. D., Jenkins, A. J., Liu, X., Zhang, B., Lee, K., Berry, W. L., . . . Ma, J.-x. (2014). Elevated Circulation Levels of an Antiangiogenic SERPIN in Patients with Diabetic Microvascular Complications Impair Wound Healing through Suppression of Wnt Signaling. *Journal of Investigative Dermatology*, 134(6), 1725-1734. doi: <https://doi.org/10.1038/jid.2014.40>
- [19] Polaskova, V., Kapur, A., Khan, A., Molloy, M. P., & Baker, M. S. (2010). High-abundance protein depletion: comparison of methods for human plasma biomarker discovery. *Electrophoresis*, 31(3), 471-482. doi: 10.1002/elps.200900286
- [20] Quehenberger, O., Armando, A. M., Brown, A. H., Milne, S. B., Myers, D. S., Merrill, A. H., . . . Dennis, E. A. (2010). Lipidomics reveals a remarkable diversity of lipids in human plasma. *J Lipid Res*, 51(11), 3299-3305. doi: 10.1194/jlr.M009449
- [21] Sadagopan, N. P., Li, W., Cook, J. A., Galvan, B., Weller, D. L., Fountain, S. T., & Cohen, L. H. (2003). Investigation of EDTA anticoagulant in plasma to improve the throughput of liquid chromatography/tandem mass spectrometric assays. *Rapid Commun Mass Spectrom*, 17(10), 1065-1070. doi: 10.1002/rcm.1023
- [22] Randall, S. A., McKay, M. J., & Molloy, M. P. (2010). Evaluation of blood collection tubes using selected reaction monitoring MS: Implications for proteomic biomarker studies. *Proteomics*, 10(10), 2050-2056. doi: <https://doi.org/10.1002/pmic.200900517>
- [23] Lebrasseur, N. K., Duhaney, T. A., De Silva, D. S., Cui, L., Ip, P. C., Joseph, L., & Sam, F. (2007). Effects of fenofibrate on cardiac remodeling in aldosterone-induced hypertension. *Hypertension*, 50(3), 489-496. doi: 10.1161/hypertensionaha.107.092403
- [24] Kafka, M., & Yermiahu, T. (1998). The effect of EDTA as an anticoagulant on the osmotic fragility of erythrocytes. *Clin Lab Haematol*, 20(4), 213-216. doi: 10.1046/j.1365-2257.1998.00014.x
- [25] Zander, J., Bruegel, M., Kleinhempel, A., Becker, S., Petros, S., Kortz, L., . . . Teupser, D. (2014). Effect of biobanking conditions on short-term stability of biomarkers in human serum and plasma. *Clin Chem Lab Med*, 52(5), 629-639. doi: 10.1515/cclm-2013-0705

Table 1. Differentially abundant proteins in EDTA and sodium citrate plasma samples.

Protein Accession number	Protein Names	Protein Descriptions	AVG Log2 Ratio EDTA v Citrate	Q-value
P02675	FIBB_HUMAN	Fibrinogen beta chain	-0.59	1.05E-42
P02671	FIBA_HUMAN	Fibrinogen alpha chain	-0.59	3.92E-40
P02751	FINC_HUMAN	Fibronectin	-1.69	3.02E-39
P68871	HBB_HUMAN	Haemoglobin subunit beta	2.21	5.52E-37
P69905	HBA_HUMAN	Haemoglobin subunit alpha	2.18	3.50E-21

Accepted Article

Table 2. Showing significantly regulated proteins in 15 patients treated with fenofibrate (complete protein information can be seen in supplementary data 5)

Protein accession number	Protein name	Log₂ FC Post-Pre	P-value
P02787	Serotransferrin	0.18	0.007
B9A064	Immunoglobulin lambda-like polypeptide 5	-0.17	0.017
A0A0C4DH24	Immunoglobulin kappa variable 6-21	-1.97	0.022
P04003	C4b-binding protein alpha chain	0.20	0.024
P04350	Tubulin beta chain	2.66	0.033
P20742	Pregnancy zone protein	-0.21	0.037
P29622	Kallistatin	-0.13	0.040

Accepted Article

Figure Legends

Figure 1. A. Venn diagram showing a comparison of the number of proteins identified using different depletion methods. **Dep G** represents samples depleted with Cibacron blue dye, **Dep S** represents samples depleted serum albumin and IgG. **B.** List of unique proteins detected using each method (immunoglobulin chain proteins are not shown)

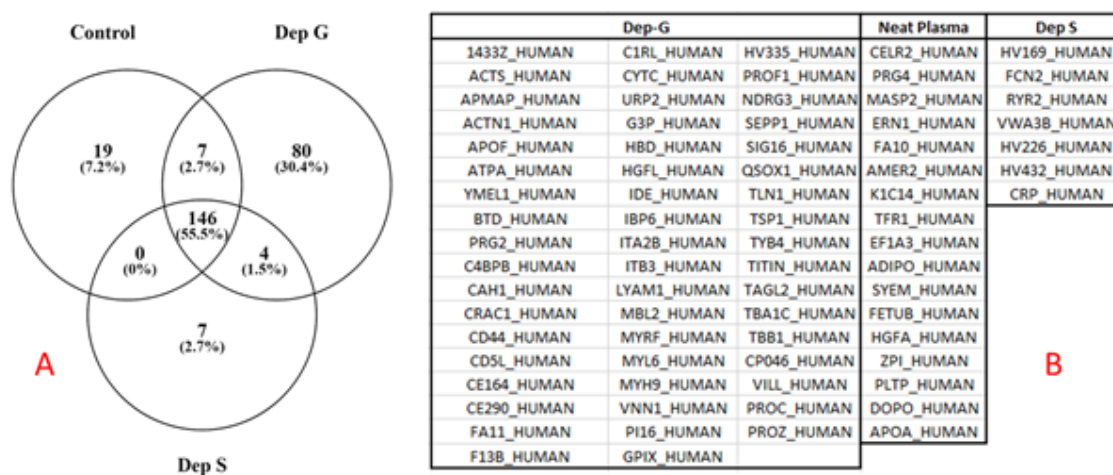


Figure 2. Venn diagram comparing number of proteins recovered following various delipidation methods. Methyl-tert-butyl-ether (MTBE), methanol chloroform (MEOHCL), acetone and ethyl acetate STAGE tip protocol.

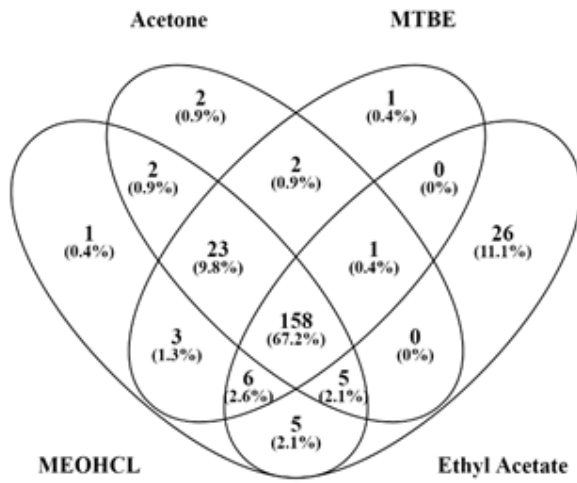


Figure 3. Relative abundances of apolipoproteins are stable across six freeze-thaw events

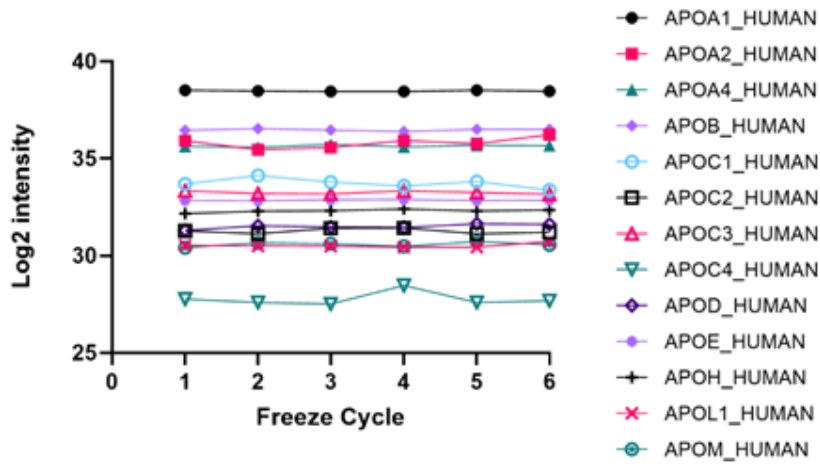


Figure 4. Intra-batch protein abundance differences of external standard (post-batch versus pre-batch) for Top 100 proteins in 65 batches.. Major dash – median, minor dashes 25 percentile and 75 percentile. Major dots – 5 and 95 percentiles.

