EuPA Open Proteomics 14 (2017) 1–10

Contents lists available at ScienceDirect

EuPA Open Proteomics

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

Comprehensive mass spectrometry based biomarker discovery and validation platform as applied to diabetic kidney disease

Scott D. Bringans[a,*], Jun Ito[a], Thomas Stoll[a], Kaye Winfield[a], Michael Phillips[b], Kirsten Peters[a,c], Wendy A. Davis[c], Timothy M.E. Davis[c], Richard J. Lipscombe[a]

[a] Proteomics International, PO Box 3008, Broadway, Nedlands, Perth, Western Australia, WA 6009, Australia
[b] Harry Perkins Institute of Medial Research, QQ Block, QEII Medical Centre 6 Verdun Street, Nedlands, Perth, Western Australia, WA 6009, Australia
[c] University of Western Australia, 35 Stirling Highway, Crawley, Perth, Western Australia, WA 6009, Australia

ARTICLE INFO

Article history:
Received 12 May 2016
Received in revised form 21 October 2016
Accepted 30 December 2016
Available online 5 January 2017

Keywords:
Biomarker
Diabetic kidney disease
MRM
iTRAQ
Diabetes

ABSTRACT

A protein biomarker discovery workflow was applied to plasma samples from patients at different stages of diabetic kidney disease. The proteomics platform produced a panel of significant plasma biomarkers that were statistically scrutinised against the current gold standard tests on an analysis of 572 patients. Five proteins were significantly associated with diabetic kidney disease defined by albuminuria, renal impairment (eGFR) and chronic kidney disease staging (CKD Stage $\geq$1, ROC curve of 0.77). The results prove the suitability and efficacy of the process used, and introduce a biomarker panel with the potential to improve diagnosis of diabetic kidney disease.
© 2017 The Authors. Published by Elsevier B.V. on behalf of European Proteomics Association (EuPA). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Protein biomarker discovery and validation of any disease is important for accurate and timely diagnosis as well as providing novel drug discovery targets and new options for therapeutic development. The most utilised biological source for biomarkers is plasma (or serum) which contains a snapshot of the physiological state of all tissues within the patient [1]. However, the discovery and validation of biomarkers for disease using mass spectrometry can be a lengthy and challenging process [2]. The design and methodology employed have huge impacts on the quality of the final results and their significance [2]. We describe a comprehensive workflow design for discovering and analytically validating a set of biomarker proteins specific for diabetic kidney disease (DKD) using mass spectrometry. An earlier pilot version of this study appeared as a technical note [3].

Diabetes is the largest cause of kidney disease (nephropathy) with 1 in 3 adults with diabetes having chronic kidney disease [4]. Worldwide, over 2 million people currently receive treatment for end stage renal disease (ESRD), although this number is likely to represent only 10% of people who actually need treatment to stay alive [5]. If DKD is detected early then appropriate intervention can help reduce further deterioration in kidney function before costly hospital-based care is required. The current gold standard tests for detecting early stage kidney disease are urinary albumin:creatinine ratio (ACR) and estimated glomerular filtration rate (eGFR), but the reliability of these results has been the subject of debate amongst clinicians [6,7]. Therefore, there is a current need to develop a more robust and specific alternative to ACR and eGFR for early detection of kidney disease.

The pipeline for proteomics-based biomarker development usually proceeds through several phases —discovery, verification and analytical validation. The discovery phase provides an initial list of proteins that may play a role in the course of disease progression. For mass spectrometry discovery, quantitative shotgun methods (such as iTRAQ) for assessing the relative concentrations of proteins are well established [8] and can provide that initial list of candidates biomarkers. This list must then be validated.

Analytical validation requires testing of the biomarker panel across a large cohort and is the primary barrier for biomarker development as the time, cost and reproducibility of such studies is burdensome. To overcome this, effective methods are required to validate potential biomarkers in large clinical cohorts. There are a variety of multiplexed assays for protein biomarker development including microarrays [9], fluorescence imaging [10] and immunoassays, in particular, enzyme-linked immunosorbent assay (ELISA) [11,12]. An emerging alternative platform to ELISA for multiplexed

* Corresponding author.
E-mail address: scott@proteomics.com.au (S.D. Bringans).

http://dx.doi.org/10.1016/j.euprot.2016.12.001
2212-9685/© 2017 The Authors. Published by Elsevier B.V. on behalf of European Proteomics Association (EuPA). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

biomarker analysis is targeted mass spectrometry or MRM (Multiple Reaction Monitoring)/SRM (Selective Reaction Monitoring), with the capacity for substantial multiplexing within a single liquid chromatography mass spectrometry (LCMS) run, an advantage when a large panel of biomarkers are required to be measured [12]. The recent development of data independent acquisition (DIA) workflows (MS/MS all, SWATH) presents a promising and novel MS quantitation technology [13,14]. However, DIA requires the latest generation of high-end MS instrumentation and it is yet to be established as a proven and robust MS-based quantitation technology across various biological matrices.

Prior to MRM validation the candidates must be individually verified where each protein is developed into a unique peptide signature, that clearly and specifically identifies and measures each peptide in turn. Targeted mass spectrometry is performed with a triple-quadrupole mass spectrometer, where precursor peptide ions are chosen as candidates to represent their respective protein. The precursor peptide ions are filtered by the first quadrupole and fragmented into product ions in the second quadrupole. Each product ion is then guided through the third quadrupole to the ion detector. The combination of a precursor and product ion pair is described as a "transition" and the amount of signal recorded by the detector forms the basis of protein quantitation by MRM.

The aim of this study was to find statistically meaningful plasma protein biomarkers specific to DKD through the development of a biomarker discovery and validation pipeline. These biomarkers would be incorporated into an early detection test that is more specific and robust than the current gold standard tests. This process was designed to start with a quantitative experiment using pooled, well-defined samples followed by preliminary validation on a relatively small pilot cohort, and then analytical validation in a much larger independent cohort. The combination of instrumentation common to any proteomics facility, with readily available mass spectrometry techniques (iTRAQ, MRM) and simply derived comprehensive labelling controls provides an ideal platform for plasma biomarker discovery and validation, as applied to DKD.

2. Materials and methods

All chemicals were sourced from Sigma-Aldrich (St Louis, MO, USA) unless otherwise stated.

2.1. Clinical samples

The clinical and demographic characteristics of the cohorts used in this study are shown in Supplementary Table S1. All clinical plasma samples were provided by the Fremantle Diabetes Study (FDS) which has been described in detail previously [15]. EDTA plasma was collected from all patients after an overnight fast and stored at −80 °C until required. The FDS collection protocol was approved by the Fremantle Hospital Human Rights Committee (07/397) with all patients providing informed written consent. In all patients kidney disease was measured by both albumin creatinine ratio (ACR) and estimated glomerular filtration rate (eGFR). The patients were classified by their ACR as normoalbuminuria ACR < 3 mg/mmol, microalbuminuria 3 ≤ ACR < 30 mg/mmol or macroalbuminuria ACR ≥ 30 mg/mmol. eGFR was estimated using the CKDEPI equation [16]. Chronic kidney disease (CKD) stage was determined using both ACR and eGFR according to current KDIGO (Kidney Disease Improving Global Outcomes, 2012) guidelines [17]. In the discovery phase, 20 samples from each of the three albuminuria groups were pooled and analysed by iTRAQ. A further 10 new individuals from each albuminuria group were used for preliminary validation of the putative biomarkers from the discovery phase using MRM. This provided the final MRM assay where the analytical validation phase measured the biomarkers from a further 572 independent patient samples with type 2 diabetes.

2.2. Immunodepletion

A Human 14 Multiple Affinity Removal (MARS14) column (Agilent Technologies, Australia) was used to chromatographically remove the 14 most abundant proteins from human plasma samples according to the manufacturer's protocol. In brief, 20 μL of each plasma sample was loaded onto a 4.6 × 50 mm MARS 14 HPLC column in an Agilent 1100 HPLC system (Agilent Technologies, Australia) with UV detection and fraction collection. Immunodepleted proteins were monitored at 280 nm and collected in a 6 min window at a flow rate of 0.125 mL/min over a 25 min run. The columns were used for 250 injections each as per the manufacturer's recommendation.

2.3. Discovery phase – iTRAQ

2.3.1. Sample preparation

Proteins in the immunodepleted samples were trypsin digested and labelled with iTRAQ reagent according to the manufacturer's protocol (Sciex, Framingham, MA, USA). The labelling of samples was as follows: reference control plasma – label 114, normoalbuminuria group – label 115, microalbuminuria group – label 116, macroalbuminuria – label 117. After labelling the samples were mixed 1:1:1:1 based on their protein content.

2.3.2. First dimension ion exchange

Labelled peptides were desalted on a Strata-X 33 μM polymeric reversed phase column (Phenomenex, Torrance, CA, USA) by firstly rinsing the column with 100% methanol (Thermo Fisher Scientific, San Jose, CA, USA), 100% acetonitrile (Fisher Scientific, Australia), 100% milliQ H_2O followed by loading of the labelled peptides, washing with milliQ H_2O and elution with 80% acetonitrile plus 0.1% formic acid. The dried sample was dissolved in 100 μL of 10 mM KH_2PO4, 10% acetonitrile, pH 3.0 and separated by strong cation exchange liquid chromatography (SCX) on an Agilent 1100 HPLC system (Agilent Technologies) using a PolySulfoethyl column (4.6 × 100 mm, 5 μm, 300 Å, Nest Group, Southborough, MA, USA). Peptides were eluted with a linear gradient of 0–400 mM KCl in 10 mM KH_2PO4, 10% acetonitrile, pH 3.0 at a flow rate of 0.5 mL/min. A total of 40 × 1 min fractions were combined into 8 fractions based on an even amount of peptide absorbance at 280 nm for each. The 8 fractions were desalted on a Strata-X 33 μM polymeric reversed phase column as detailed previously. The 8 fractions were then dried under vacuum.

2.3.3. Second dimension reverse phase nanoLC onto MALDI plates

Dried peptide fractions from the SCX were dissolved in 0.1% trifluoroacetic acid and loaded onto a C18 pre-column and then separated on a C18 PepMap100, 3 μm column (Dionex, Sunnyvale, CA, USA) using the Ultimate 3000 nano HPLC system (Dionex). A gradient of 10–40% acetonitrile in 0.1% trifluoroacetic acid at a flow rate of 300 nL/min was used with the eluent mixed 1:3 with matrix solution (including Calibration Mixture) and spotted onto a 384 well Opti-TOF plate (Sciex) using a Probot Micro Fraction Collector (Dionex).

2.3.4. MALDI mass spectrometry

The spotted plates were analysed using a 4800 TOF/TOF system (Sciex). The Nd:YAG laser was set at 355 nm and a frequency of 200 Hz, with 400 shots per spot for MS data acquisition and MS data acquired for singly charged peptides in the mass range of 800–

4000 *m/z*. A job-wide interpretation method selected the 20 most intense precursor ions above a signal/noise ratio of 20 from each spectrum for MS/MS acquisition but only in the spot where their intensity was at its peak. MS/MS spectra were acquired with 4000 laser shots per selected ion with a mass range of 60 to the precursor ion.

Protein identification and iTRAQ quantification were performed using ProteinPilot™ 2.0.1 Software (Sciex). MS/MS spectra were searched against a human protein database. Search parameters were: Sample type, iTRAQ 4plex (peptide labelled); Cys alkylation, MMTS; Digestion, trypsin; Instrument, 4800; Special factors, none; Species, none; Quantitate tab, checked; ID focus, Biological modifications, Search effort, thorough; Detected protein threshold (unused ProtScore), 1.3–which corresponds to proteins identified with >95% confidence. The peak area ratios of the iTRAQ reporter ions for each peptide reflected the relative abundances of the peptides. For quantification analysis protein expression ratios were computed by the Paragon Algorithm (Protein Pilot) based on the peak area ratios of the peptides accounting for the same protein.

2.3.5. iTRAQ initial biomarker selection

Each of the significant differentially expressed proteins from the iTRAQ analysis of the FDS cohort was catalogued. To ensure as inclusive a selection of potential biomarkers as possible, the primary significance level was broadened from selection of those proteins with a default p value of <0.05 to include a secondary list of proteins with differential expression ratio of >2 and <0.5 and also those proteins where the p value was <0.1. Only a single replicate iTRAQ experiment was performed on pooled samples, with the study design then moving to individual sample analysis with greater statistical power.

2.4. Verification and validation phase – MRM

2.4.1. Reference plasma control

A reference plasma control sample was obtained from pooling plasma from three healthy volunteers. This was aliquoted and stored at −80 °C, and used throughout the study. The reference plasma control sample was labelled with ^{18}O water and termed "Std18". The same amount of Std18 was spiked into each patient sample (1:1) prior to LC-MRM/MS analysis to correct for spray efficiency and ionisation differences between analytical runs. The Std18 was prepared as per references [18–20].

2.4.2. Preparation of plasma for MRM

Each immunodepleted plasma sample was concentrated in a 10 kDa Vivaspin 6 concentrator (Sartorius AG, Goettingen, Germany) and reconcentrated in 1 M triethylammonium bicarbonate buffer (TEAB) by centrifugation. The protein concentration in 2 μL of depleted plasma was measured with an infrared-based method (Direct Detect, Merck Millipore, Darmstadt, Germany) according to the protocol provided by the manufacturer. Depleted plasma protein samples were reduced with 30 mM Tris(2-carboxyethyl) phosphine, alkylated with 30 mM iodoacetamide and digested with 0.2 μg/μL trypsin (trypsin:plasma protein ratio of 1:20). To stop the digestion reaction, samples were boiled, then desalted on Strata-X 33 μm polymeric reversed phase columns (Phenomenex) and dried down in a speedvac.

2.4.3. Verification of MRMs

MRM assays were developed using a 4000 QTRAP system (Sciex) equipped with a NanoSpray source with data analysis and refinement using Skyline Software [21] (University of Washington, Seattle, WA, USA). Multiple MRM transitions were developed per peptide for each putative protein biomarker for both the light version of the peptide and the ^{18}O-labelled heavy version of the peptide. Cross contamination testing was carried out with injections of unlabelled plasma while monitoring transitions for both unlabelled and ^{18}O labelled versions of each peptide. No significant peaks with a S/N > 3 were detected for any of the 13 final peptides in the heavy label peptide. To determine the linear range of detection of each unlabelled peptide a dilution series study was performed in duplicate with the light version and ^{18}O-labelled heavy version of a reference plasma sample. The linear quantification ranges and R^2 values are shown in Supplementary Table S2.

The FASTA file for each biomarker protein sequence was imported into the Skyline program; the precursor and fragment ions for each peptide derived from the protein were generated by performing *in silico* digestion. The peptide filter conditions were as follows: the precursor length range was set at 7 to 21 amino acids, and peptides with repeat arginine (Arg, R) or lysine (Lys, K) residues were not used. If proline (Pro) was next to an arginine (Arg, R) or lysine (Lys, K) residue, the peptide was not used. Useful proteotypic peptide information from literature and repositories (PeptideAtlas [22], MRMaid [23]) was incorporated and the selection of transitions was supported by spectral libraries (Institute for Systems Biology [24], National Institute of Standards and Technology [25], Global Proteome Machine organisation [26], Bibliospec [27]). Transitions for ^{18}O-labelled peptides were created by selecting the "$^{18}O(2)$" isotope modification in the Peptide Settings tab in Skyline. The final list of transitions with precursor/product ion, collision energy (CE) and declustering potential (DP) values are provided in Supplementary Table S3.

Initially, to confirm the iTRAQ data the same pooled samples were analysed with the optimised MRM method on the QTRAP and the data viewed in Skyline to verify the MRM assay and the presence of the ^{18}O-stable isotope labelled reference plasma version of each peptide. For preliminary validation these samples were followed by a pilot study on 3 new sets of 10 patients from each of the normo, micro and macroalbuminuria groups. The comparison of these individuals dictated the final MRM assay that was ultimately applied to the large clinical cohort of 572 patients for analytical validation.

2.4.4. MRM mass spectrometry

Relative peptide quantitation MRM analyses were performed with a 4000 QTRAP mass spectrometer coupled with a Dionex Ultimate 3000 nano-HPLC system. To reduce the void volume and obtain sharper intensity peaks no pre-column was used and a small sample loop (100 μm ID capillary tubing containing 1 μL sample) was inserted in the autosampler. A 1 μL volume of loading buffer (98% H_2O, 2% ACN, 0.05% TFA) containing 1:1 (v/v) ratio of tryptic unlabelled and ^{18}O-labelled reference plasma peptides was loaded onto a 15 cm Zorbax 300SB-C18 (Agilent Technologies) analytical column. Peptides were separated in a 90 min LC run with a linear gradient of 2 to 30% buffer B (100% ACN + 0.1% formic acid) at a flow rate of 400 nL/min. The conditions set in Analyst v1.4 MS software (Sciex) for scheduled MRM analysis of plasma biomarker peptides by the 4000 QTRAP interfaced with a nanospray source were as follows: ion spray (IS); 2900 V, interface heater temperature (IHT); 200 °C; collision gas (CAD); HIGH, ion source gas 1 (GS1); 30 and curtain gas (CUR); 10. MS parameters for declustering potential (DP) and collision energy (CE) were calculated by the Skyline program and MS resolutions for Q1 and Q3 were set at low.

In the MS settings for scheduled MRM, polarity was set to positive, MRM detection window was 360 s, target scan time was 4 s and pause time between mass ranges was 5 milliseconds. To verify plasma peptide sequences, an MRM triggered MS/MS experiment was performed with enhanced product ion (EPI) experiments targeting specific transition pairs. The following settings were used: IDA (Information Dependant Acquisition) first level criteria: 1–2 most intense peaks which exceed 5000 counts

per second, with rolling collision energy and a mass tolerance of 250 mDa. Following this, two EPI scans were carried out with the following conditions to obtain MS/MS spectra: scan range of 200–1200 *m/z*, scan rate of 1000 Da/s, positive polarity, number of scans to sum = 2, product of 30 Da and total scan time (including pauses) was 2.7 s. A Mascot (Matrix Science, London, UK) search and/or manual inspection of MS/MS fragmentation spectra were performed to confirm peptide identity. In addition, interference testing was carried out with Skyline by using the dot product (dotp) cutoff >0.8 between the acquired MSMS spectrum and the final MRM relative transition heights as well as between the MRM and the spectral library obtained MSMS spectra. Interfering transitions were removed to maintain the integrity of the data. In the analytical validation phase of the study, a single scheduled MRM experiment was used to quantify the selected unlabelled and ^{18}O-labelled target transitions. MRM assays were created using the Scheduled MRM algorithm on Analyst.

2.4.5. Skyline data collation

All transition peaks were visually checked in Skyline and wrong/missed peaks corrected manually. Un-labelled/^{18}O-labelled peptide ratios were exported from Skyline for further analysis.

MRM-quantified biomarker peptide peak areas were adjusted to 1 μg peptide 'on column' by multiplying them by a normalisation factor based on their infrared-calculated protein concentration. The normalised unlabelled peak areas were then divided by their respective ^{18}O-labelled peak area to obtain the final unlabelled/^{18}O-labelled peak area ratios.

The limit of detection (LOD) for MRM peak intensity ratio analysis was a signal to noise ratio of greater than 3:1 for transition peak area: background peak area. This LOD was based on a previous MRM quantitative analysis of plasma peptides with a 4000 QTRAP [28]. Additionally, all transitions for both unlabelled and ^{18}O-labelled transitions with minimum peak height intensities below 1000 counts were excluded from analysis. Total peak area and background area values were obtained from viewing the peptide results grid in Skyline software. The unlabelled and ^{18}O-labelled transitions were required to have the same retention time and same order of transition intensity. This is in addition to confirmation of the peptide's identity with a full scan MS/MS spectrum that had been previously performed.

2.4.6. Synthetic peptide

A ^{13}C^{15}N stable isotope-labelled synthetic peptide sequence from Complement factor H-related protein 2 (CFHR2) was obtained from Sigma-Aldrich (USA), with purity of greater than 95% as an AQUA™ peptide. The cysteine residue of the peptide was alkylated to the stable S-carboxyamidomethylcysteine (CAM) form. The peptide sequence from the CFHR2 protein was LVYPSCEE [K_^{13}C^{15}N] (RMM = 1132.5 Da) where the terminal K residue had both ^{13}C and ^{15}N isotopes as denoted.

2.5. Data analysis

Demonstrations of MRM assay linearity, technical reproducibility and sensitivity were performed with synthetic CFHR2 peptide LVYPSCEEK. A dilution series of five known concentrations (from 500 attomoles/μL to 200 femtomoles/μL loaded on column) of the synthetic CFHR2 peptide were tested. This standard was spiked into the post-tryptic digested immunodepleted reference plasma and measured by MRM six times. To create the linear regression standard curve, the peptide concentration was calculated from the average peak area quantified in Skyline MRM software analysis program. Standard deviation and CV calculations for peptide concentration were also performed from the average peak area value to determine technical reproducibility of the assay for the synthetic CFHR2 peptide.

The standard curve was also used to verify the LOD and limit of quantification (LOQ) of the synthetic CFHR2 peptide in the MRM assay. The LOD was set to an average peak area from six technical replicates of at least 3 times above background signal and the LOQ was set to an average peak area of at least 10 times above background signal [28].

2.5.1. Statistical analysis

All statistical analyses were performed in SPSS for Windows (version 21; SPSS Inc., Chicago, IL, USA). The data presented uses the relative concentration of a protein between samples, consequently the validation MRM results were produced as a set of ratios of unlabelled/^{18}O-labelled peak areas from the set of transitions for each peptide. These ratios were normalised to the median value for each peptide. All biomarker peak area ratios were natural ($\log_e$) transformed to normalise their distribution. To confirm candidate biomarkers in the pilot study, two-way comparisons using a Mann-Whitney test for non-parametric data were performed between the albuminuric groups (normo- versus micro-, micro- versus macro-, and normo- versus macroalbuminuria). If a protein met the criteria of $p < 0.1$ for at least one peptide then that protein was considered for the validation phase. In the analytical validation phase, Spearman's rank order correlation (ρ) was used to investigate the relationship between each biomarker, ACR, eGFR and CKD. A two-tailed significance level of $p < 0.05$ was used for these analyses.

The diagnostic relationship between plasma biomarker concentrations and i) microalbuminuria, ii) eGFR<60 mL/min per 1.73 m^2 and iii) CKD stage were examined using multivariate logistic regression modelling (forward conditional variable selection with $p < 0.05$ for entry and >0.10 for removal). All protein biomarkers with bivariate $p \leq 0.20$ were considered for entry in a forward stepwise manner. To avoid overfitting the models, the cohort was randomly split into train and test sub-cohorts (80:20 split, respectively), with the ratio of people with the outcome of interest similar in both groups [29]. The train sub-cohort was used for model development for each diagnostic outcome, while the test sub-cohort was used for model validation. The clinical and demographic characteristics of train and test sub-cohorts are shown in Supplementary Table S1. The discriminative ability of each model was assessed by the area under the curve (AUC) produced by receiver operating characteristic (ROC) curves. The Youden Index was used to determine the optimal predicted probability cut-off to achieve maximum sensitivity and specificity for each model [30]. Other measures of diagnostic performance were based on the optimal cut-off; false positive and false negative rate, positive and negative predictive value, and diagnostic odds ratio. To compare the performance of each biomarker model to the current gold standard, firstly, the ability of patient gold standard ACR and eGFR data to correctly identify individuals with eGFR < 60 mL/min per 1.73 m^2 and ACR $\geq$ 3 mg/mmol, respectively was assessed. Secondly, the biomarker models for eGFR < 60 mL/min per 1.73 m^2 and ACR $\geq$ 3 mg/mmol were compared to their respective gold standard tests to evaluate performance. Finally, the biomarker model for the combined outcome of CKD (incorporating eGFR < 60 mL/min per 1.73 m^2 and ACR $\geq$ 3 mg/mmol) was compared to all other models.

3. Results and discussion

3.1. Study design

The most critical aspect of the initial phase of any biomarker project is to select high quality patient cohorts. This includes

careful patient selection that is representative of the clinical question to be addressed as well as consistent collection protocols to minimise any potential degradation of protein *via* freeze thaw cycles or extended periods of room temperature exposure. The plan for this study is shown in Fig. 1 and describes the DKD biomarker discovery, preliminary and analytical validation workflow from start to finish.

In the discovery phase of this project the patient cohort was chosen to provide samples of people with varying and distinct levels of kidney disease as measured by albuminuria as a complication of diabetes. Pools of samples were labelled with iTRAQ reagents and analysed on a MALDI TOF/TOF mass spectrometer. The potential biomarkers derived from this data were converted into a series of MRM transitions, and verified for data quality. Along with the 3 original pooled samples a further 3 groups of 10 individuals representing normo-, micro- and macroalbuminuria were tested in a randomised order by targeted MS. This pilot study of 30 individuals provided sufficient data to derive a final MRM assay list for validating on the larger independent clinical cohort of 572 patients.

3.2. Discovering biomarker candidates – iTRAQ

The iTRAQ results as analysed by Protein Pilot (Sciex) provided a list of proteins and their relative expression levels between the three albuminuria (normo-, micro-, and macroalbuminuria) groups. From this, a primary list and a secondary list of potential biomarkers were derived. The selection criteria to produce the overall list were deliberately kept broad so as not to exclude any protein that may be of interest. Proteins were excluded where the quantitative or identification evidence was poor. When comparing the three albuminuria groups against each other the iTRAQ data provided a combined potential biomarker list of 32 proteins. Due to the variability of replicate iTRAQ data (for example [31–33] show only 43–61% common identifications across 3 iTRAQ replicates) the current workflow was designed based on a single iTRAQ experiment followed by a complementary small scale targeted mass spectrometry pilot study. This orthogonal validation technique transformed the list of potential biomarkers into a final analytical assay for large scale validation.

3.3. Verification and validation by MRM

MRM transitions were developed for each of the 32 potential biomarkers using the following parameters. Preliminary MRM transition lists were generated by a series of steps which included; downloading protein candidate sequences, digesting proteins *in silico* in conjunction with a filter (e.g. 7–21 amino acids, 0 missed cleavage) and selecting a minimum of 4 transitions per peptide (precursor charge z2, product charge z1). Useful proteotypic peptide information from literature and repositories was incorporated with support from spectral libraries. A reference plasma control was then used to search for and verify each of the peptides selected for each of the potential biomarkers. If any MRM evidence was found after analysis in Skyline then a full MS/MS spectrum was acquired with an MRM triggered MS/MS run. The third piece of evidence after MRM signal and MS/MS data was confirmation of the suitability of the peptide for ^{18}O-labelling. If all evidence was satisfactory then the biomarker was added to the final biomarker assay. Of the 32 potential biomarkers, 25 met the above criteria.

The MRM assay of the 25 potential biomarkers (41 peptides, 254 transitions) was used to quantify the proteins in a preliminary validation of 3×10 individuals representing each 'label' from the iTRAQ experiments. Of the 25 potential biomarkers, 8 were found to be significantly different between the normo-, micro- and macroalbumuric groups in this pilot study.

These 8 proteins formed the final biomarker MRM assay (13 peptides, 64 transitions) and were applied in the last phase of the process to the analytical validation on 572 individual patient plasma samples.

3.4. Stability and reproducibility of the proposed DKD protein biomarker panel

In order to compare and measure protein concentrations over a long period of time there is a requirement for a standard control sample to provide a fixed reference point for all measurements. The use of an ^{18}O-labelling technique provides an elegant solution by labelling every peptide in a reference plasma sample to produce a "universal" standard [18–20]. With this method the two C-terminal oxygen atoms on each peptide are exchanged from ^{16}O to ^{18}O with the reaction catalysed by trypsin. This results in a 4 Da shift from the unlabelled peptide allowing easy discernment of each form of the peptide in the mass spectrometer. This method is both cheap and comprehensive, allowing every valid MRM for each peptide to have a reference point for comparisons between samples and across time. To complement this global internal standard, a biomarker peptide was synthesised as an alternative isotopically labelled standard, allowing an accurate measure of the reference plasma variability.

3.4.1. ^{18}O-labelled reference plasma

The ^{18}O-labelled reference plasma (Std18) was created to provide a common point of reference for all relative unlabelled/^{18}O-labelled peptide peak area measurements. The stability of Std18 ^{18}O-labelling over time was tested with a sub-batch of freshly reconstituted Std18. After aliquots were tested the remainder was stored at −20 °C for two weeks. Ratio comparisons of eight biomarker peak areas showed high levels of labelling efficiency with the median labelling performance of 98.2% at t = 0

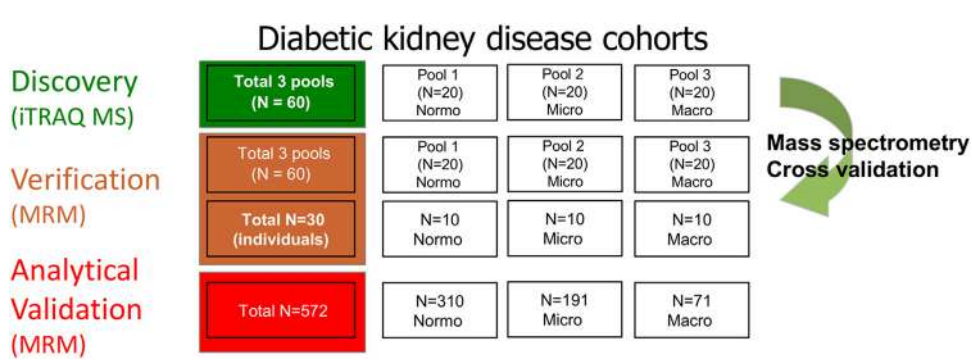

Fig 1. Workflow for discovery and validation of diabetic kidney disease biomarkers. Total numbers of patient samples analysed, either in pools (as denoted) or individually. The uncoloured boxes denote the breakdown of samples into the normo-, micro- and macroalbuminuria categories as labelled.

and 96.3% at t = 2 weeks. The worst performing peptide had a loss of 3.3% labelling over the 2 weeks with the median loss across all peptides at only 1.0% demonstrating the robustness of ^{18}O-labelling when stored at −20 °C for up to two weeks (Fig. 2). The small percentage of unlabelled peptide present in the Std18 is a fixed amount for all analyses and when comparing the changing protein ratios from different patients the influence of this small fixed value is negated.

3.4.2. Intra-and inter-day CV analysis of Std18 and synthetic peptide control

Modern mass spectrometers are sensitive to subtle changes in their operating environment. This can cause MRM-quantified absolute peak areas of the same peptide from the same reference standard to show significant variation between runs. This can be overcome by using a fixed amount of labelled reference plasma with each sample. Due to the use of peak area ratios, any instrument variation is minimised as the unlabelled peptide co-elutes with the fixed labelled peptide and experiences identical chromatographic and ionisation conditions. To confirm this hypothesis a fixed amount (100 fmoles) of a synthetic ^{13}C^{15}N isotope-labelled peptide was spiked into unlabelled/^{18}O-labelled samples just prior to MRM analysis. Analysis of these samples was spread over a 3 month time period, with the reference plasma control sample included between every batch of 20 samples. In total, 5 intra-day and 9 inter-day control plasma samples were run in duplicate.

The peak areas of the synthetic standard in the 14 reference plasma controls were compared in Skyline to the corresponding ^{18}O-labelled reference peptide. Despite considerable peak intensity variation across the 14 controls over time, the overall peak area profile for the synthetic peptide was highly similar to that of the ^{18}O-labelled counterpart (Fig. 3a and b). Furthermore, when the ratio of these two standards is calculated (Fig. 3c), this provides an intra-day CV of 5.9% and an inter-day CV of 8.1% from the 14 controls.

Overall, this analysis demonstrated that the mass spectrometer does vary in its response to the same amount of material on column over time but the important aspect is that the response ratio between the spiked and labelled peptides is the same. This demonstrates that even when the mass spectrometer has variation in the ionisation of peptides the labelled standard captures this variation and therefore corrects for it in the calculation of ratios of unlabelled/^{18}O-labelled peaks. This ensures that samples acquired on different days, weeks or months can be compared against each other because the ^{18}O-labelled reference plasma is capturing the instrument variation on that day at that time. The inclusion of a known fixed amount of synthetic peptide also provides the opportunity to derive absolute quantification of proteins if this is desired, although it was not essential in this study. In this example the absolute quantification range for the CFHR2 protein in plasma was 0.94 ng–2.1 μg/mL. The LOD was 0.18 ng/mL with the LOQ at 0.94 ng/mL.

3.5. Analytical validation of diagnostic biomarkers

The final biomarker MRM assay of 8 proteins derived from the discovery and pilot study was deployed to test the much larger independent cohort of 572 individuals with type 2 diabetes. This involved examination of ACR values as well as eGFR and the CKD classification (KDIGO 2012). Of the 572 patients, 53.1% were male, with a mean±SD age of 66.6 ± 10.6 years and median [inter quartile range] diabetes duration of 10.0 [3.0-16.0] years. In the cohort, 54.4% had normoalbuminuria, 33.4% microalbuminuria, and 12.2% macroalbuminuria. Mean eGFR was 79.7 ± 21.2 mL/min/1.73m^2, 16.1% had eGFR<60 mL/min/1.73m^2 and 69.6% were taking an ACE-I and/or ARB. 50% of the diagnostic cohort had CKD defined by eGFR < 60 mL/min/1.73m^2 or ACR ≥ 3.0 mg/mmol (CKD stage ≥ 1, Fig. 4).

Table 1 details the biomarker correlations to ACR and eGFR. Of the 8 candidate biomarkers, 5 were significantly correlated with ACR (APOA4, CFHR2, HBB, IBP3 and AMBP, all $p < 0.05$), and 5 with eGFR (APOA4, APOC3, CFHR2, IBP3 and AMBP, all $p < 0.05$) (Table 1). Four proteins in particular, APOA4, CFHR2, IBP3 and AMBP, showed significant correlations with both ACR and eGFR for at least one peptide.

To show the diagnostic ability of the biomarkers, multivariate logistic regression models were developed in the train sub-cohort for ACR ≥ 3 mg/mmol, eGFR < 60 mL/min/1.73 m^2 and CKD, and validated in the test sub-cohort (Table 2). There were no significant

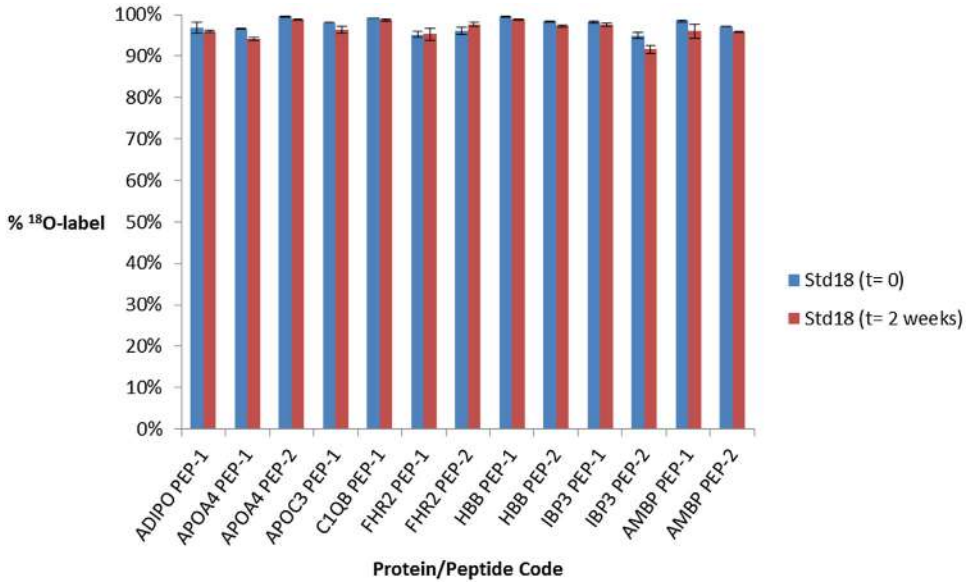

Fig 2. Stability of Std18 ^{18}O-labelling over time. Three replicates (n = 3) of Std18 biomarker peptides at t = 0 and t = 2 weeks (stored at −20°C) were analysed by MRM. The peak area ratios of ^{18}O-labelled peptides were divided by the combined unlabelled and labelled peak areas to determine the% of labelled peptide with their averages shown in percentage terms. Error bars are 1 standard deviation from the average peak area ratios.

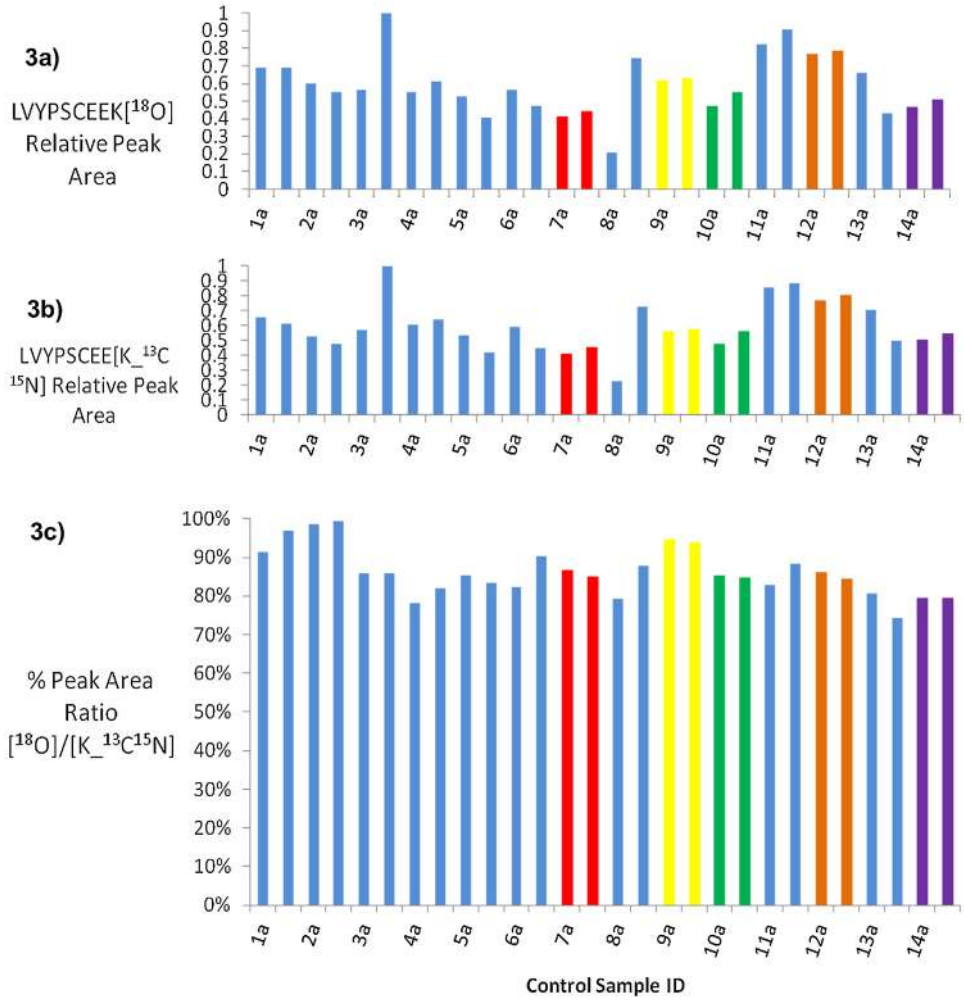

Fig 3. Intra- and inter-day peak area profiles of ^{18}O- and $^{13}C^{15}N$-labelled CFHR2 peptides. Peak areas of ^{18}O-labelled LVYPSCEEK peptide in 14 reference plasma controls (Fig 3a) and spiked 100 fmoles of synthetic $^{13}C^{15}N$-labelled peptide LVYPSCEEK (Fig 3b), quantified by MRM. The controls were numbered 1–14, with the same colour (except blue) used for intra-day duplicate samples. Inter-day samples and their duplicates (indicated as 'a' and 'b' samples) are coloured in blue. The% Peak Area Ratio for the 18O-labelled / synthetic $^{13}C^{15}N$-labelled peptides is shown in Fig 3c.The intra-day CV was 5.9% and the inter-day CV was 8.1%.

	PROTEINUREA – assessed by ACR (mg/mmol)			
eGFR (mL/min /1.73m²)	NORMAL ACR 0 --- <3	MILD ACR >3 --- <30	HEAVY ACR >30	
≥ 90	135	61	25	Not CKD CKD 0
60-89	151	88	20	CKD 1
45-59	18	14	11	
30-44	7	22	6	CKD 2
15-29	0	6	5	CKD 3
<15	0	0	3	

Fig 4. Stratification of patient cohort by ACR, eGFR and CKD risk. The cohort of 572 patients is shown as the distribution according to ACR and eGFR categories and the associated CKD risk (KDIGO).

Table 1
Biomarker correlation of MRM data for Pilot study (ACR) and Validation study (ACR and eGFR) for cohort of 572 patients.

Protein Name	UniprotKB Accession	Peptide	N = 30 Pilot study	N = 572 Validation Study	
			Mann Whitney test	Spearman's rho p-value	
			ACR p-value	ACR p-value	eGFR p-value
Adiponectin	ADIPO	(Pep1) GDIGETGVPGAEGPR	0.008	0.251	0.089
Apolipoprotein A-IV	APOA4	(Pep1) LEPYADQLR	>0.1	**0.002**	<**0.001**
		(Pep2) ISASAEELR	0.083	<**0.001**	<**0.001**
Apolipoprotein C-III	APOC3	(Pep1) DALSSVQESQVAQQAR	0.056	0.701	**0.004**
Complement C1q subcomponent subunit B	C1QB	(Pep1) IAFSATR	0.002	0.063	0.382
Complement factor H-related protein 2	CFHR2	(Pep1) TGDIVEFVCK	>0.1	0.090	<**0.001**
		(Pep2) LVYPSCEEK	0.030	**0.010**	<**0.001**
Hemoglobin subunit beta	HBB	(Pep1) SAVTALWGK	0.052	<**0.001**	0.355
		(Pep2) VNVDEVGGEALGR	0.052	<**0.001**	0.346
Insulin-like growth factor-binding protein 3	IBP3	(Pep1) ALAQCAPPPAVCAELVR	0.083	<**0.001**	0.060
		(Pep2) FLNVLSPR	0.069	<**0.001**	**0.019**
Protein AMBP	AMBP	(Pep1) TVAACNLPIVR	>0.1	**0.017**	**0.049**
		(Pep2) EYCGVPGDGDEELLR	0.037	0.210	<**0.001**

The significant biomarker proteins are shown with MRM data correlations to ACR for the Pilot study and both ACR and eGFR for the Validation study. The Pilot study shows the best p-value for comparison between macro/micro/normoalbuminuria groups against ACR using a Mann Whitney test for non-parametric data. The Validation study data for 572 patient plasma samples shows the MRM peptide data with the corresponding Spearman's rho p-value for correlation to ACR (mg/mmol) and eGFR (mL/min/1.73 m^2) values. For the Validation study **bold** values indicate $p < 0.05$.

Table 2
Diagnostic performance of biomarker models of microalbuminuria (ACR ≥3 mg/mmol), eGFR < 60 mL/min/1.73m^2 and CKD compared to gold standard ACR and eGFR tests in type 2 diabetes.

	Diagnostic Test	Diagnosis	AUC	True Positive Rate (Sensitivity)	False Positive Rate (1-Specificity)	DOR
	Gold Standard ACR	eGFR < 60 mL/min/1.73m^2	N/A	73%	40%	4.0
	Gold Standard eGFR	ACR ≥ 3 mg/mmol	N/A	26%	8%	4.0
TRAIN (n = 459)	BM (eGFR model)[a]	eGFR < 60 mL/min/1.73m^2	0.80	73%	25%	8.3
	BM (ACR model)[a]	ACR ≥ 3 mg/mmol	0.68	72%	42%	3.6
	BM (CKD model)[a]	CKD ≥ 1	0.68	56%	25%	3.9
TEST (n = 113)	BM (eGFR model)[a]	eGFR < 60 mL/min/1.73m^2	0.81	88%	32%	14.9
	BM (ACR model)[a]	ACR ≥ 3 mg/mmol	0.71	52%	15%	6.0
	BM (CKD model)[a]	CKD ≥ 1	0.77	56%	15%	7.6

The performance of the models in the train and test sub-cohorts is shown.

- BM (eGFR model) (APOA4_Pep 2, APOC3_Pep 1, CFHR2_Pep 1, IBP3_Pep 2).
- BM (ACR model) (APOA4_Pep 1, C1QB_Pep 1, CFHR2_Pep 2, IBP3_Pep 2).
- BM (CKD model) (APOA4_Pep 1, CFHR2_Pep 2, IBP3_Pep 2).

[a] BM, Biomarker model; AUC, area under curve; DOR, diagnostic odds ratio. BM models were developed in train sub-cohort and validated in test sub-cohort.

differences in clinical or demographic characteristics between the two sub-cohorts (Supplementary Table S1).Independent associates consistent across the three outcomes were APOA4, CFHR2, and IBP3 (Table 2). Table 2 presents the diagnostic performance of each of the biomarker models compared to the gold standard ACR and eGFR diagnostic tests. The ACR and eGFR data for each individual were used as the gold standard for diagnosing the opposite outcome. When the ACR data were used to diagnose eGFR < 60 mL/min/1.73 m^2 it performed well, with a True Positive and False Positive rate of 73% and 40%, respectively. In the opposite analysis, the patient's eGFR data had very poor True Positive rate (26%_ but excellent False Positive rate (8%) when used to diagnose ACR ≥ 3 mg/mmol (Table 2). Comparing the biomarker ACR and eGFR models to the respective gold standard diagnostic tests, it can be seen that there was an improvement in diagnostic performance for both of the biomarker models in the test sub-cohort. The biomarker eGFR model had an improved True Positive rate (88% vs 73%) and a reduced False Positive rate (32% vs 40%) over the gold standard ACR for diagnosing eGFR < 60 mL/min/1.73 m^2, while the biomarker ACR model had improved True Positive rate (52% vs 26%), but poorer False Positive rate (15% vs 8%). The diagnostic odds ratio (DOR) for the eGFR and ACR biomarker models were significantly better than those of the gold standards (Table 2, eGFR 14.9 vs 4.0, ACR 6.0 vs 4.0). The biomarker CKD model, combining both ACR and eGFR, has an AUC of 0.77 with a True Positive rate of 56% and a False Positive rate of 15%. The benefit of this test is that it is capable of diagnosing people that are normally missed by individual gold standard ACR or eGFR tests, for example, people with normoalbuminuria that have eGFR < 60 mL/min/1.73 m^2 (Fig. 4, n = 25), or those with eGFR ≥ 60 mL/min/1.73 m^2 and micro- or macroalbuminuria (Fig. 4, n = 194). These data suggest that the biomarker CKD model is an effective alternative to current gold standard diagnostic tests and could replace the need for collection of urine and blood for analysis in two separate tests.

As expected for a biomarker discovery to validation pipeline there was a level of attrition in moving from 32 iTRAQ derived potentials to 25 verifiable potential biomarkers to an MRM assay for 8 candidate biomarkers and finally to a panel of 5 significant biomarkers with correlation to ACR (Fig. 5). This attrition demonstrates the importance of initial wide selection criteria to

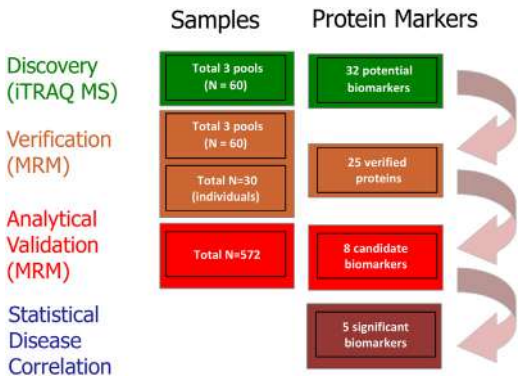

Fig 5. Biomarker progression from discovery to ACR correlation. The progression of potential biomarkers identified from the discovery iTRAQ analysis through to those that were statistically correlated to ACR.

allow for any marker of potential to be included in the initial discovery phase. Concomitant to this finding is that the majority of candidate biomarkers identified in the pilot study remained significant in the much larger cohort. The significance of the final results vindicates the choice of workflow in developing biomarkers for DKD. The initial pooled iTRAQ experiment followed by the pilot study by MRM of a small group of individual patients allowed the final analytical validation to take place on an independent large cohort.

4. Conclusions

Currently over 400 million people have diabetes [34] and 1 in 3 adults with diabetes have chronic kidney disease. The biomarker discovery pipeline detailed in this paper illustrates that a comprehensive study starting with a small number of patient plasma samples can ultimately produce a diagnostic test that has advantages over the current gold standards (ACR, eGFR). Use of highly stratified patient samples and the broad mass spectrometry platform has produced a panel of biomarkers for DKD that have ultimately been analytically validated in a large independent cohort of 572 patients with significant correlations with the current measures of disease. The multivariate analysis provided a panel of markers that performed well when either ACR or eGFR was the gold standard. Importantly, however, the utility of the panel may be best expressed when diagnosing CKD with a single test rather than requiring both urine and plasma collection and analysis. Improved testing would allow earlier intervention and therefore result in better patient outcomes. The ^{18}O-labelling of the reference plasma was a key tool to provide global relative measurements over an extended analysis timeframe. The use of a separate isotope labelled synthetic peptide provided confirmation that the ^{18}O-labelling was accounting for any instrument variation. The approach described above is therefore a straightforward yet effective strategy for protein biomarker discovery through to analytical validation, and has the capacity to provide an improved diagnostic test for DKD.

Acknowledgements

Partial funding for the validation study was provided by the Federal Government of Australia under a Commercialisation Australia Grant (CAU05922). The initial discovery and pilot validation study was conducted with the support of the (Western Australian) Centre for Food and Genomic Medicine. The MS analyses were performed in facilities provided by the Lotterywest State Biomedical Facility-Proteomics node and Bioplatforms Australia at the Harry Perkins Institute for Medical Research.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.euprot.2016.12.001.

References

[1] L.A. Liotta, M. Ferrari, E. Petricoin, Clinical proteomics: written in blood, Nature 425 (2003) 905.
[2] S. Ray, P.J. Reddy, R. Jain, K. Gollapalli, A. Moiyadi, S. Srivastava, Proteomic technologies for the identification of disease biomarkers in serum: advances and challenges ahead, Proteomics 11 (11) (2011) 2139–2161.
[3] ABSciex technote avaialble at http://www.absciex.jp/Documents/Downloads/Literature/Biomarker-Pipeline-Diabetes-4250211-01.pdf.
[4] Centers for Disease Control and Prevention, National Chronic Kidney Disease Fact Sheet, U.S. Department of Health and Human Services, Atlanta, GA, 2014. http://www.cdc.gov/diabetes/pubs/pdf/kidney_factsheet.pdf.
[5] W.G. Couser, G. Remuzzi, S. Mendis, M. Tonelli, The contribution of chronic kidney disease to the global burden of major noncommunicable diseases, Kidney Int. 80 (December (12)) (2011) 1258–1270.
[6] C.A. Houlihan, C. Tsalamandris, A. Akdeniz, G. Jerums, Albumin to creatinine ratio: a screening test with limitations, Am. J. Kidney Dis.: Off. J. Natl. Kidney Found. 39 (6) (2002) 1183–1189.
[7] A.K. Mottl, K.S. Kwon, M. Mauer, E.J. Mayer-Davis, S.L. Hogan, A.V. Kshirsagar, Normoalbuminuric diabetic kidney disease in the U: S. population, J. Diabetes Complications 27 (2) (2013) 123–127.
[8] J.A. Westbrook, J. Noirel, J.E. Brown, P.C. Wright, C.A. Evans, Quantitation with chemical tagging reagents in biomarker studies, Proteomics Clin. Appl. 9 (2015) 295–300.
[9] H.J. Hye Jin Lee, A.W. Wark, R.M. Corn, Microarray methods for protein biomarker detection, Analyst 133 (8) (2008) 975–983.
[10] M.Q. Smith, C.A. Staley, D.A. Kooby, T. Styblo, W.C. Wood, L. Yang, Multiplexed fluorescence imaging of tumor biomarkers in gene expression and protein levels for personalized and predictive medicine, Curr. Mol. Med. 9 (8) (2009) 1017–1023.
[11] K. Kohler, H. Seitz, Validation processes of protein biomarkers in Serum—a cross platform comparison, Sensors 12 (2012) 12710–12728.
[12] C.E. Parkera, C.H. Borchers, Mass spectrometry based biomarker discovery, verification, and validation –Quality assurance and control of protein biomarker assays, Mol. Oncol. 8 (4) (2014) 840–858.
[13] L.C. Gillet, P. Navarro, S. Tate, H. Rost, N. Selevsek, et al., Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis, Mol. Cell. Proteomics 11 (6) (2012).
[14] N. Selevsek, C.Y. Chang, L.C. Gillet, P. Navarro, O.M. Bernhardt, et al., Reproducible and consistent quantification of the Saccharomyces cerevisiae proteome by SWATH-mss spectrometry, Mol. Cell. Proteomics 14 (3) (2015) 739–749.
[15] T.M. Davis, D.G. Bruce, W.A. Davis, Cohort profile: the fremantle diabetes study, Int. J. Epidemiol. 42 (2) (2013) 412–421.
[16] A.S. Levey, L.A. Stevens, C.H. Schmid, Y.L. Zhang, A.F. Castro, 3rd, Feldman HI et al: a new equation to estimate glomerular filtration rate, Ann. Intern. Med. 150 (9) (2009) 604–612.
[17] A.S. Levey, P.E. de Jong, J. Coresh, M. El Nahas, B.C. Astor, K. Matsushita, et al., The definition, classification, and prognosis of chronic kidney disease: a KDIGO Controversies Conference report, Kidney Int. 80 (1) (2011) 17–28.
[18] W. Qian, T. Liu, V.A. Petyuk, M.A. Gritsenko, B.O. Petritus, et al., Large-scale quantitative discovery proteomics enabled by the use of an O-labeled universal reference sample, J. Proteome Res. 8 (2009) 290–299.
[19] B.O. Petritus, W. Qian, D.G. Camp II, R. Smith, A simple procedure for effective quenching of trypsin activity and prevention of ^{18}O-labeling back-exchange, J. Proteome Res. 8 (2009) 2157–2163.
[20] R. Niles, H.E. Witkowska, S. Allen, S.C. Hall, S.J. Fisher, M. Hardt, Acid-catalyzed oxygen-18 labeling of peptides, Anal. Chem. 81 (2009) 2804–2809.
[21] B. MacLean, D.M. Tomazela, N. Shulman, M. Chambers, G.L. Finney, B. Frewn, et al., Skyline: an open source document editor for creating and analyzing targeted proteomics experiments, Bioinformatics 26 (7) (2010) 966–968.
[22] Peptide Atlas, available at http://www.peptideatlas.org.
[23] MRMaid, available at http://138.250.31.29/mrmaid/.
[24 Institute for Systems Biology, spectral database available at http://www.peptideatlas.org/speclib/#ISB.
[25] National Institute of Standards and Technology, spectral database available at http://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:cdownload.
[26] Global Proteome Machine, databases available at http://gpmdb.thegpm.org/.
[27] Bibliospec, spectral library available at https://skyline.gs.washington.edu/labkey/project/home/software/BiblioSpec/begin.view.
[28] J. Stahl-Zeng, V. Lange, R. Ossola, K. Eckhardt, W. Krek, R. Aebersold, et al., Sensitivity detection of plasma proteins by multiple reaction monitoring of N-Glycosites, Mol. Cell. Proteomics 6 (2007) 1809–1817.

[29] F. Bagherzadeh-Khiabani, et al., A tutorial on variable selection for clinical prediction models: feature selection methods in data mining could improve the results, J. Clin. Epidemiol. 71 (2016) 76–85.
[30] W.J. Youden, Index for rating diagnostic tests, Cancer 3 (1) (1950) 32–35.
[31] F. Song, A. Poljak, N.A. Kocchan, M. Raftery, H. Brodaty, G.A. Smythe, P.S. Sachdev, Plasma protein profiling of mild cognitive impairment and Alzheimer's disease using iTRAQ quantitative proteomics, Proteome Sci. 12 (5) (2014).
[32] Y. Ji, J. Ji, F. Sun, Y. Zeng, X. He, J. Zhao, Y. Yu, W. Wu, Quantitative proteomics analysis of chondrogenic differentiation of C3H10T1/2 mesenchymal stem cells by iTRAQ labelling coupled with on-line two-dimensional LC/MS/MS, Mol. Cell. Proteomics 9 (3) (2010) 550–564.
[33] H. Wang, S. Alvarez, L.M. Hicks, Comprehensive comparison of iTRAQ and label-free LC-based quantitative proteomics approaches using two Chlamydomonas reinhardtii strains of interest for biofuels engineering, J. Proteome Res. 11 (2012) 487–501.
[34] WHO, Global Report on Diabetes, (2016) . (Available at) www.who.int/diabetes/global-report/en/.