Mass Spectrometry-based Serum Proteomics for Biomarker discovery and validation

Santosh D. Bhosale[1], Robert Moulder[1], Petri Kouvonen[1], Riitta Lahesmaa[1] and David R. Goodlett[1,2]

[1]Turku Centre for Biotechnology, University of Turku, Finland,

[2]Department of Pharmaceutical Science, University of Maryland, Baltimore, Maryland, USA,

Abstract/Summary:

Blood protein measurements are used frequently in the clinic in the assessment of patient health. Nevertheless, there remains the need for new biomarkers with better diagnostic specificities. With the advent of improved technology for bioanalysis and the growth of biobanks including collections from specific disease-risk cohorts, the plasma proteome has remained a target of proteomics research towards the characterization of disease related biomarkers. The following protocol presents a workflow for serum/plasma proteomics including details of sample preparation both with and without immunoaffinity depletion of the most abundant plasma proteins, and methodology for selected reaction monitoring mass spectrometry validation.

Key words: Serum/Plasma, Label-free quantification, Selected reaction monitoring, Proteomics, Mass spectrometry

1. Introduction:

Mass spectrometry based proteomics has remained in focus as a method of biomarker discovery from clinical samples, in particular serum or plasma (1). Ideally, quantitative data of specific

marker panels could enable the clinician to predict the subclinical status and decide upon the therapeutic management of a disease process (2, 3). Currently, with the maturation of omics platforms and systems biology approaches, it has become possible to improve sample throughput to a level suitable for biomarker discovery and validation on moderate size (4).

Although proteomics literature includes many examples where putative biomarkers have been identified by mass spectrometry, the list of approved and implemented markers still remains marginal. Following candidate identification it is crucial that the plausibility of the identified and but putative marker is established in order to propagate their movement from the discovery to the translational pipeline (5). ELISA assays are sensitive and easy to implement and are accepted as the standard approach for validation. However, in recent years there has been a growing use of selected reaction monitoring (SRM) mass spectrometry for validation of putative markers discovered by proteomics (6). Specifically, SRM is used for the quantification of selected peptides in a targeted manner from a complex sample matrix. The SRM technique, protocol and its application have previously been reviewed in detail, for which the reader is referred to reviews by the Aebersold group (7-8). In comparison to ELISA, SRM provides the advantages of facilitating the monitoring of multiple targets in a single analysis, and the targets can be directly interpolated from the discovery data. As SRM is a peptide centric approach it is less susceptible to the influence of sample ageing, whereas with ELISA conformational changes in the proteins may influence the interaction with the epitope.

To enable unambiguous SRM validation of the differentially abundant proteins detected in a discovery experiment, it is essential to target peptides that uniquely identify the target protein (i.e. proteotypic peptides) that are preferably stable (i.e. not prone to oxidation or other variable modification), neither too long or short, and frequently observed in the discovery data with a reproducible retention time. Peptide tandem mass spectra are used to create SRM transitions lists, which are derived from the parent and fragment ion *m/z* values, that are unique to a given

peptide sequence and retention time. To realize the full capabilities of SRM it is beneficial to establish retention indices for the targeted peptides. Spiking with retention time standards, even in the discovery phase, can be thus used and enable the scheduling of a series of targeted measurements (23). In order to authenticate the peptide targets isotopically labelled synthetic peptides should be obtained (e.g. incorporating heavy lysine and arginine containing the following heavy isotopes, $^{13}C_6{}^{15}N_2$ and $^{13}C_6{}^{15}N_4$, respectively).

Blood serum is an easily accessible biofluid, which in essence carries a biochemical record of an individual's health status. Its proteomic analysis, however, is challenged by the wide range of protein abundances that are characteristic of its composition. Albumin alone, for instance, constitutes around half of serum protein composition by weight, and the next 12 proteins account for another 45%. The dominance of albumin therefore limits sample loading and the depth of profiling (9). A frequently necessary, but debatable, consideration is whether to deplete the abundant proteins or not. Targeted removal of the more abundant proteins can be used to alter the range of protein abundance, although the depletion step can be influenced by non-targeted and thus non-specific interactions (10). Analysis of the non-depleted sample, on the other hand, provides a better record of the key serum proteins, although is less likely to provide useful quantitative data for the lower abundance proteins (11). Several manufacturers provide antibody affinity media that may be used to target the removal of the most abundant proteins. These can be used in a chromatographic or spin-cartridge format. Resins are available for albumin alone, albumin and the IgGs, as well as the 6, 7, 12, 14 and 20 most abundant proteins. Frequently researchers have removed 12-14 proteins using such resins, which are commercially available from companies such as Agilent and Sigma (12-13). The chromatographic approach has been a popular choice, although the recent availability of single use cartridges has provided new opportunities for throughput and scalability (14).

Here we present a pipeline for serum proteomics biomarker discovery and validation, indicating workflows that use immunoaffinity depletion as well as one that does not shown in **Figure 1** and **Figure 2**.

2. Materials:

Equipment/Instrument:

1. For LC-based immunoaffinity depletion with chromatographic columns of 4 to 6.6 mm i.d., use a HPLC system capable of delivering a flow rate between 0.1 ml to 1.0 ml together with a UV detector (280 nm) and fraction collector. See Note 1 for other technical specifications.
2. 0.22 µm Eppendorf spin filters.
3. A depletion column, e.g. MARS Hu-14 column (4.6 mm ID X 50 mm; capcity 20 µl, Agilent Technologies, Santa Clara, California, USA) and operating buffers A & B (see Note 2)
4. Spin concentrators with 5 kDa MW cutoff (Sartorius-Stedim, Vivaspin, 4ml, 5kDa cutoff)
5. For peptide concentration estimation, a NanoDrop 2000 Spectrophotometer (Thermo Scientific) or similar instrument.
6. A centrifugal evaporator, e.g. a SpeedVac™ (Thermo Scientific)
7. SepPak C18 cartridges, 50 mg (Waters).
8. LC-MS/MS system, i.e. for chromatography and electrospray tandem mass spectrometry with a nanospray interface, e.g. LTQ Orbitrap Velos Pro mass spectrometry coupled with Easy nLC-II (Thermo Scientific)

Chemicals:

1. Dissolution buffer: Dissolve 395 mg ammonium bicarbonate (ABC, i.e.NH_4HCO_3) in 100 ml MilliQ water to give 50 mM ABC.
2. Denaturation buffer: 8 M urea: Dissolve 24 g of urea in 20.0 ml of 50 mM ABC. When dissolved make volume to 50 ml.
3. Dithiothreitol (DTT) stock solution: Dissolve the weighed amount of dithiothreitol in 50 mM ABC to give 1.0 M final concentration.
4. Iodoacetamide stock solution: Dissolve weighed iodoacetamide in 50 mM ABC to give 1.0 M final concentration.
5. Sequencing grade modified trypsin (Promega).
6. Indexed Retention Time (iRT) peptides (Biognosys).

3. Procedure:

3.1. Discovery phase:

3.1.1. Depleted serum sample workflow:

Proteins targeted for depletion using an Agilent Hu14 MARS column:

α1-acid glycoprotein (P02763), Fibrinogen (P02761), α 1-antitrypsin(P01009), Haptoglobin (P00738), α 2-macroglobulin (P01023), IgA (P01876), Albumin (P02768), IgG (P01857, P01859-61; all major subclasses of gamma globulin), Apolipoprotein A-I (P02647), IgM (P01871), Apolipoprotein A-II (P02652), Transferrin (P02787), Complement 3 (P01024) & transthyretin (P02766).

3.1.1.1. LC-Based Immunodepletion of high abundance proteins:

1) To ensure sample homogeity vortex the sample and spin briefly.

2) Dilute the serum aliquot with three volumes of the dilution Buffer A, e.g. 15 µl plus 45 µl

3) Filter the sample through 0.22 µm spin filters to remove particulates.

4) The diluted sample is introudced onto the MARS column with a mobile phase compostion of 100 % Buffer A, at a flow rate of 0.125 ml/min. A generic gradient program is indicated as follows. (**Table 1**)

5) The autosampler method is set-up to deliver the flow-through (depleted serum) to the collection tube with sufficient allowance for delay from detection to collection (see **Note 3**).

6) Elute the bound fraction, collect if required for further analyses (100 % Buffer B at flow rate of 1.0 ml/min for 6 min).

7) Column regenerated and equilibration in 100 % buffer A for 7 min.

8) If there is concern about carry over, system blanks can be injected between runs

3.1.1.2. Buffer exchange:

Following depletion it will be necessary to concentrate the protein solution and change to a denaturing buffer. A method for buffer exchange is presented, see **Note 4** for other examples.

1) Pre-rinse the Ultrafiltration spin columns for the samples with 1 ml of the dilution buffer: +4 ºC, 3000 x g, ~20 minutes. At this stage it is possible to identify any spin columns that vary in performance (i.e. slow filtration), include sufficient columns to substitute these.

2) Concentrate the collected depleted serum fractions to a volume of ~100µl using the washed Ultrafiltration spin columns (+4 ºC, 3000 x g, ~20 min).

3) Perform buffer exchange with 8M Urea in 50 mM ABC.

a) 1200 µl of 8M Urea: +4 ºC, 3000 x g, 35 min.

b) 500 µl of 8M Urea: +4 ºC, 3000 x g, 35 min.

c) 500 µl of 8M Urea: +4 ºC, 3000 x g, 30 min.

Final volume of sample should be ~100 µl.

4) To ensure that the concentrated proteins are in solution and reduce losses to the filter, ultrasonicate the spin columns for 5 min on ice and withdraw the liquid with a pipette.

5) To further reduce the losses from transfer of the concentrate, wash the spin columns with 50 µl of 8M Urea by sonicating 5 min on ice. Combine this with the concentrated sample.

3.1.1.3. Trypsin digestion**:**

1) Add 1 µl 1 M dithiothretiol (DTT) and incubate for 1 hr at 37^0C

2) Add 1 µl 1 M iodoacetamide, maintain in darkness at room temperature for 30 min.

3) Dilute the samples to reduce the urea concentration < 1M.

4) Add 10 µl Sequencing Grade Modified trypsin at 37^0C at a ratio of 1:30 (Trypsin:Protein) for overnight (16-18 hr).

3.1.1.4. Desalting and peptide concentration:

The digested samples are acidified using 10% trifluoroacetic acid (TFA), then desalted using Sep-Pak cartridge (WAT054955, Vac 1 cc 50mg, Waters).

1) Acidify the digested peptides with 50 µl of 10% TFA. Check the pH (<2.5).
2) 'Wet' the Sep-Pak column with 1 ml of 100 % methanol
3) Equilibrate with 1 ml of 80% Acetonitrile + 0.1% TFA
4) Equilibration with 2 x 1 ml of 0.1% TFA
5) Collect the flow-through as you pass the sample steadily through the column (not to dryness), then pass it through once again.
6) Wash the cartridge with 3 x 1 ml of 2%ACN + 0.1% FA (notice the change to formic acid)
7) Elute the peptides with 1 ml of 80% Acetonitrile + 0.1% FA (see **Note 5**).
8) Speed vac to dry
9) Reconstitute with 2% formic acid + 2% acetonitrile (use 60 µl for a sample from 15 µl of serum depleted of the 14 most abundant proteins).

3.1.1.5. Dilution for Sample injection.

1) Estimate the peptide content using a NanoDrop spectrophotometer with the Protein method at 280 nm.Use the 260/280 ratio as a sample integrity, a values in the order of 0.6 - 1 should be expected for a typical sample from 15 µl of serum depleted of the 14 most abundant proteins dissolved in 60 µl.

2) In order to inject 200 ng from a volume of 5 µl injection, the sample should be further diluted to a concentration of 0.04 µg/µl.

3) To assist in retention time mapping and SRM assay development include a spiked amount of synthetic retention time standards, indexed retention time (iRT) peptides (Biognosys). Spike 1 µl of iRT peptides (50 µl stock) per 20 µl of sample.

4) Perform LC-MS/MS analysis of the samples in triplicate as randomized batches.

3.1.2. Undepleted serum sample workflow:

1) Dilute a 2 µl raw serum sample with 100 µl denaturation buffer (Note 6).

2) Reduce the proteins disulphide bond with DTT (final concentration 10 mM) for 1 hr at 37 ^{0}C. Add 1 µl of DTT stock solution.

3) After one hour incubation, alkylate the disulphide bridges with iodoacetamide (final concentration 13 mM) incubate for 30 min at room temperature in dark. Add 1.4 µl of iodoacetamide stock solution.

4) Dilute the samples to 900 µl with 50 mM ABC (to reduce urea <1M).

5) Trypsin reconstitution: 20 µg trypsin vial (Promega sequencing grade) in 70 µl MilliQ water (0.29 µg/ µl).

6) Add in 1:30 ratio (Trypsin:Protein). For 2 µl undepleted serum (approx. conc ~65-70 µg/µl), added 15 µl of trypsin (4.35 µg). Incubate at 37 ^{0}C overnight.

7) Desalting as discussed in the depleted serum sample workflow.

3.1.3. Sample Analysis and Data Acquisition:

LC-MS/MS Conditions

1. EasyNano-LC: Vented pre-column configuration, a 20 x 0.1 mm i.d. pre-column packed with ReproSil-Pur 5 µm 200 Å C18-AQ, (Dr. Maisch HPLC GmbH, Ammerbuch-Entringen, Germany) connected by a New objective two-way union together with a 75 µm x 150 mm analytical column packed with same packing material.
2. A separation gradient from 2 to 35% B in 65 min. at a flow rate of 300 nl/min
3. Autosampler set-up: 20 µl sample loop for 5 µl injections
4. Sample randomization: The samples are randomized to remove the influence of injection order. Batches of single injections (each samples separated by a 15 min blank method), with three/four replicate injections in total (3 or 4 batches) with the system performance monitored between batches using a lab standard. A pool of the samples in the batch is analysed at the start and finish of each batch. The maintenance of constant/accountable instrument performance is essential for a successful LFQ experiment.
5. A) Orbitrap-Velos Pro: Data dependent MS/MS data acquisition, with ionization in positive ion mode with CID of the 15 most intense ions (m/z 300-2000, charge states > 1+). Dynamic exclusion 60 s. Orbitrap precursor ion scan resolution 60,000 (at m/z 400), with a target value of 1,000,000 ions and a maximum injection time of 100 ms. For the ion trap the target values and maximum injection time values are set to 50,000 and 50 ms. When making the selection of the top "n" most intense ions it is important to consider the associated duty cycle and the width of the *chromatographic peak, i.e. ensuring that there are sufficient MS1 data points to describe the peak elution profile.

   B) Q Exactive: The method employed Data dependent MS/MS data acquisition, with ionization in positive ion mode with HCD of the 10 most intense ions (m/z 300-2000,

charge states > 1+). Dynamic exclusion 20 s, a resolution of 140,000 (at m/z 400), with a target automatic gain control (AGC) values of 3,000,000 ions and 100 ms maximum injection time. For the MS2 level, the target values and maximum injection time values are set to 50,000 and 250 ms respectively.

3.1.4. Data Analysis:

Currently there are two popular software packages available for analysing label free data, the free ware, MaxQuant (15), and the commercial package Progenesis from Nonlinear (16). A description of MaxQuant is presented. Details of the implementation of Progenesis for label free quantitation are described elsewhere (14).

3.1.4.1. MaxQuant:

MaxQuant is a proteomics mass spectrometry data analysis package that includes its own database search algorithm, Andromeda (17). It can be used to perform the complete data analysis work flow on a single or groups of raw MS data files to determine peptide and protein identification along with estimations of abundance. For usage of the software and additional information concerning its installation, the reader is referred to www.maxquant.org. Here MaxQuant v.1.4.1.2 is described together with Perseus v.1.4.0.20, which is an open source platform that was developed to process MaxQuant data (18).

1. Open the MaxQuant.exe file.
2. Load the acquired LC-MS/MS raw files into the MaxQuant environment (19).
3. The key points in the context of label free quantification (20) are as follows.

4) In the “Group-specific parameters” tab specify following.

    a) Type>Standard

    b) Multiplicity>1

    c) Label-free quantification>LFQ

Other parameters like variable modifications, digestion mode, instrument type and max missed cleavages are user specific.

5) In “Global parameters” tab select following
    a) FASTA files>Navigate to the FASTA file, uploaded into the Andromeda interface
    b) Match between runs>Select this to compare multiple files
6) For general usage leave the other parameters with default settings.
    a) Number of threads: This depends on your computer configuration, the number of core. The general recommendation is 1 thread/2GB derived from MaxQuant Google group discussions.
7) Select start:

    The progress can be monitored in the “Performance” tab, and once the analysis is completed the status will be displayed in the notification window. The results can be found in the ~*combined*\ directory as the *proteinGroups.txt* file. See Note 7.
8) The reproducibility across the analytical replicates can be assessed by plotting the scatter plot using Perseus as shown in **Figure 3**.
9) For downstream bioinformatics analysis Perseus, which was tailored for use with MaxQuant, provides many options. Additional analyses can be made with R (21), or statistical packages like SPSS (22).

3.2) Validation phase:

In this protocol the use of Skyline (27) is described. Skyline is a vendor independent software developed for SRM data analysis (see Note 8). The discovery data and search results can be imported into Skyline and the availability of proteotypic peptides and stable interference-free

SRM transitions established. Similarly the data from the analysis of heavy-labelled synthetic peptides can be used.

In setting up the Skyline file and associated spectral library and creating peptide retention time indices you will need:

a) Representative *.raw files in which the targets were identified,
b) The search results, e.g. from MaxQuant (msms.txt & modifications.xml) or Proteome Discover (*.msf).
c) A FASTA file of the proteome.
d) iRT peptides to spike the analysed samples

3.2.1. Workflow for SRM analysis using Skyline:

1. Create a library in Skyline using the search results
    a) In the peptide settings select a background proteome (e.g. HUMAN SwissProt), the number of missed cleavages and enzyme.
    b) Follow the tabs through to select the filters for peptide length (e.g. 7 to 25) and unwanted modifications.
    c) Build the library by importing the search results.
    d) Define the permitted modifications
    e) Once the library is assigned you can query the library for spectra of the protein targets.

2. Use Skyline to analyse the discovery data and to choose proteotypic peptides identified in the discovery phase (Note 9). At this stage you should have a reasonable overview of the peptides useful for quantifying your protein targets. Transition list for the native peptides can be made although it will become important to validate these with heavy-labelled synthetic equivalents.

3. Synthetic peptide analysis

a) Pool synthetic peptides for transition generation: (See Note 10).

b) LC-MS/MS analysis with a Q Exactive: Analyze the pooled sample together with indexed retention time peptides (iRT) on a Q Exactive MS (Note 11 & 12).

c) Database search: For the database search, use a FASTA file where the synthetic peptides are concatenated as one protein, the sequence coverage instantly provides an indication of how successful the choice of peptides and their analysis was (see Note 13).

4. Use Skyline to establish the best co-eluting, interference free transitions for the assays using a triple quadrupole mass spectrometer.
    a) From the synthetic peptide analysis results the library can be recreated (see above).
    b) Import the *.Raw file and build an iRT retention time calculator using the iRT peptides as retention time standards.
    c) Check the peak detection, suitability of transition and aim to include five good transitions per peptide, deselecting inappropriate choices and selecting the correct peak apex.
    d) Export the transition list. For the TSQ mass spectrometer. The *.csv output includes such parameters as the precursor and fragment mass, the collision energy, the retention time window (start and stop), the peptide sequence, the protein identity, the fragment ion and its rank. With the current version of Skyline manually editing (see **Table 2)** is required so that the table is in the appropriate format for the TSQ triple quadrupole**,** for some AB Sciex instruments the output may directly be used.

The polarity is 1 for positive electrospray, the trigger an ion intensity threshold (100 by default) and “Reference” is used to indicate a standard that can be used to modify the scheduling on the fly (a default value of 0 removes this option).

5. SRM analysis:

a) Perform a non-scheduled run from your sample where only the iRT peptides are monitored. Measurement of the retention times together with known iRT values allows calculation of retention time indices for the target peptides.

b) Import the non-scheduled run*.Raw file into Skyline interface for calculation of actual retention times for the target peptides. This is achieved by utilizing calculated iRT values for the target peptides and measured retention times for iRT peptides.

c) Assign the retention time indices and permitted tolerance for the SRM method together with all collision energies (Note 13).

d) Perform a scheduled run where proteotypic peptides (protein target) transitions are monitored along with its heavy labelled counterpart and iRT. Then analyse the scheduled *.Raw file with Skyline, and evaluate the transitions for all targets. A few scheduled runs can be submitted for transition optimization and validation (Note 14).

e) Finally, analyse the synthetic peptide pool together with iRT peptides in sample matrix using optimized, scheduled method to monitor the transitions (protein targets). Transitions affected by sample background interference can be used for identification but should not be used for quantification.

f) In order to maintain the mass spectrometry performance stable throughout the batch, the inclusion of short run blanks runs (i.e. no sample) in between samples is

beneficial. Additionally, the inclusion of a pooled sample to the batch, for example, at the start and end of the batch provides a useful overview of the instrument performance (Note 15).

4 Notes:

1. As the recommended procedures required to the dilution of the serum, it is important that the sample loop volume used with the LC system is larger enough, e.g. in the order of three times the injected volume (200-300 µl depending on the processed amount and the column type) to accommodate for the diluted sample and reduce losses. Also the sampler should preferably be configured such that after loading the sample would be back-flushed directly onto the column, i.e. does not need to pass through the whole loop.
2. Buffers: The identity of the Agilent buffers remain proprietary information. However, for the use of the IgY media, as has been supplied by Sigma and others, the buffers may be prepared according to the details provided. Dilution Buffer: Tris-Buffered Saline (TBS) - 10 mM Tris-HCl with 0.15 M NaCl, pH 7.4 Stripping Buffer: 1 M Glycine, pH 2.5 Neutralization Buffer: 1 M Tris-HCl, pH 8.0.
3. As the methods includes a flow increase, the volume of the collected fraction will change rapidly as its duration exceeds the flow change. Method development with test serum should be made to optimize the timing of the fraction collection.
4. In comparison to buffer exchange, commonly used actone or acetone/TCA precipitation approaches can be used to recover the proteins from depleted flow through. Which approach to follow is based on the yield (buffer exchange vs protein precipitation) however should be tested.

5. Elution with a lower acetonitrile concentration, e.g. can be sufficient and may help to remove large problematic peptides/proteins and unwanted contaminants.
6. For quantitative analysis it is important to aliquot reproducible quantities. Increasing the aliquot size could be used to improve the reproducibility, e.g. to 5 µl, although due to demands for larger quantities of trypsin this has its draw backs. Alternatively an aliquot of the diluted serum can be digested.
7. The processing times is dependent on the number and type of *.raw files as well as configuration of processing computers. In our experience, processing 60 *.raw files obtained from an LTQ Orbitrap Velos Pro takes one day using a computer with the following specifications.
    a. Processor – Intel(R)Xeon(R) CPU E5 – 26090 @ 24 GHz, RAM – 128 GB and 64-bit operating system.
8. The development of transition lists can be made using different software dedicated to SRM data analysis (24-26). Detailed step-by-step instructions for using Skyline are found on the web page (https://skyline.gs.washington.edu/labkey/project/home/software/Skyline/begin.view).
9. Ideally synthetic peptides should be 7 - 25 amino acids in length and unique. The absence of methionine as well as missed cleavages and ragged ends is preferred. Synthetic peptides with carbamidomethylation modification can also be obtained as well as other variants. Shorter or more diverse combinations may be necessary when there is a shortage of unique peptides for the target. Additional proteotypic peptides that were not identified might be considered for synthesis and subsequence confirmation, particularly if only 1-2 peptides have been successfully identified for given protein.
10. Usually the concentrations reported by the manufacturer are based on total concentration from peptide synthesis, which also includes synthesis by-products, such

as single amino acids and truncated peptides. Therefore, calculating equimolar amount of peptides into the pool is not feasible. In our case 10 µl from each peptide vial was used for initial analysis

11. For library creation HCD spectra from a Q Exactive compare favourably with the CID spectra generated by triple quapdrupole instruments, such as the TSQ Vantage. In the following workflow using heavy labelled synthetic peptides analysed with a Q Exactive and TSQ Vantage is described.
12. At this stage it is advantageous if you are using similar LC configuration and gradient as intended for the subsequent validation. Nevertheless differences in gradient and can be compensated by the iRT peptides if the column stationary phase/packing media is the same.
13. In the case of missing identifications, missing peptides can be pooled together or even injected individually to complement the results.
14. Validate the integrity of the transitions in a pooled sample spiked with heavy labelled and iRT peptides. The pooled sample should be representative of the sample matrix to be analysed in the validation step. (If possible, this sample should also be used in following step as a quality control sample, QC-sample. See step 5 point f). The sample matrix will generate some interference, which should be taken into account for when choosing the stable transitions for detection and quantification. Ideally the selected transitions should have clear and matching profiles and simple ions that are likely to be common to many proteins, e.g. y1 and y2, should be avoided.
15. After determining the frequency of washes, sample order should be randomized for data acquisition. When queuing the samples for data acquisition, QC-sample should be injected first, once a day and the last. Since it should be same across the sample batch, QC-sample data can be used to determine the stability of the system and also to monitor

any accumulation of the signal, which could affect the results. Even though the careful planning and scheduling of the washes, this might still happen particularly with clinical samples where heterogeneity between the samples might be vast.

Acknowledgement:

The National Technology Agency of Finland (Finland Distinguished Professor Programme, grant 40398/11), the Academy of Finland (Centre of Excellence in Molecular Systems Immunology and Physiology Research, grant 250114), JDRF, The Sigrid Jusélius Foundation and Biocentre Finland (Turku Proteomics Facility) are thanked for their financial support.

References:

1. Veenstra TD, Conrads TP, Hood BL, Avellino AM, Ellenbogen RG, Morrison RS (2005) Biomarkers: mining the biofluid proteome. Mol Cell Proteomics. 4:409-418.
2. Lyons TJ, Basu A (2012) Biomarkers in diabetes: hemoglobin A1c, vascular and tissue markers, Transl Res 159:303-312
3. Crutchfield CA, ThomasSN, Sokoll LJ, Chan DW (2016) Advances in mass spectrometry-based clinical biomarker discovery. Clin Proteomics 13
4. Dayon L, Kussmann M (2013) Proteomics of human plasma: A critical comparison of analytical workflows in terms of effort, throughput and outcome. EuPA OPEN PROTEOMICS 8-16
5. Rifai N, Gillette MA, Carr SA (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nature Biotechnology 24:971 – 983
6. Surinova S, Schiess R, Hüttenhain R, Cerciello F, Wollscheid B, Aebersold R (2011) On the development of plasma protein biomarkers. J Proteome Res 10:5-16

7. Picotti P, Aebersold R (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Methods 9:555-566

8. Lange V, Picotti P, Domon B, and Aebersold R (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. Mol Syst Biol 4:222

9. Issaq HJ, Xiao Z and Veenstra TD (2007) Serum and Plasma Proteomics. Chem. Rev 107(8): 3601-3620

10. Tu C, Rudnick PA, Martinez MY, Cheek KL, Stein SE, Slebos RJ, Liebler DC (2010) Depletion of abundant plasma proteins and limitations of plasma proteomics. J. Proteome Res 9(10):4982-91

11. Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M (2016) Plasma Proteome Profiling to Assess Human Health and Disease. Cell Syst Mar 23;2(3):185-95

12. Ahmed N, Barker G, Oliva K, Garfin D, Talmadge K, Georgiou H, Quinn M, Rice G (2003) An approach to remove albumin for the proteomic analysis of low abundance biomarkers in human serum. Proteomics 3(10):1980-7

13. Smith MP, Wood SL, Zougman A, Ho JT, Peng J, Jackson D, Cairns DA, Lewington AJ, Selby PJ, Banks RE (2011) A systematic analysis of the effects of increasing degrees of serum immunodepletion in terms of depth of coverage and other key aspects in top-down and bottom-up proteomic analyses. Proteomics 11(11):2222-35

14. Moulder R, Goo YA, Goodlett DR (2016) Label-Free Quantitation for Clinical Proteomics. Methods Mol Biol 1410:65-76

15. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nature Biotechnology 26:1367 – 1372

16. Megger DA, Bracht T, Kohl M, Ahrens M, Naboulsi W, Weber F, Hoffmann AC, Stephan C, Kuhlmann K, Eisenacher M, Schlaak JF, Baba HA, Meyer HE, Sitek B. (2013) Proteomic differences between hepatocellular carcinoma and nontumorous liver tissue investigated by a combined gel-based and label-free quantitative proteomics study, Mol Cell Proteomics 12:2006-2020

17. Cox J, Neuhauser N, Michalski A, Scheltema RA., Olsen JV, Mann M (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment, J Proteome Res 10:794–1805

18. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data, Nat Methods Jun 27

19. Tyanova S, Mann M, and Cox J. (2014) MaxQuant for In-Depth Analysis of Large SILAC Datasets, Methods Mol Biol 1188:351-364

20. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 13:2513-2526

21. R Development Core Team (2011 R: A Language and Environment for Statistical Computing. Vienna, Austria.

22. IBM Corp. Released 2013. IBM SPSS Statistics for Windows Version 22.0. Armonk, NY: IBM Corp.

23. Escher C, Reiter L, MacLean B, Ossola R, Herzog F, Chilton J, MacCoss MJ, Rinner O (2012) Using iRT, a normalized retention time for more targeted measurement of peptides Proteomics 12(8): 1111-21

24. Brusniak MK, Kwok S, Christiansen M, Campbell D, Reiter L, Picotti P, Kusebauch U, Ramos H, Deutsch EW, Chen J, Moritz RL and Aebersold R (2011)

ATAQS: A computational software tool for high throughput transition optimization and validation for selected reaction monitoring mass spectrometry BMC Bioinformatics 12-78

25. Teleman J, Karlsson C, Waldemarson S, Hansson K, James P, Malmström J, Levander F (2012) Automated selected reaction monitoring software for accurate label-free protein quantification J Proteome Res 11(7):3766-73

26. Aiyetan P, Thomas SN, Zhang Z and Zhang H (2015) MRMPlus: an open source quality control and assessment tool for SRM/MRM assay development BMC Bioinformatics 16:411

27. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments Bioinformatics 26:966– 968

Table 1 LC Method for depletion with an Agilent 4.6 x 50 mm Hu14 Column

Cycle	**Time (min)**	**Dilution Buffer– A %**	**Stripping Buffer-B %**	**Flow rate (mL/min)**
Injection	0	100	0	0.125
Wash	9.5	100	0	0.125
Wash	9.6	100	0	1.0
Wash	11.5	100	0	1.0
Wash	11.6	0	100	1.0
Stripping	16.00	0	100	1.0
Re-equilibration	16.1	100	0	1.0
Stop	25.00	100	0	1.0

Table 2: For SRM analysis with a TSQ the exported transition list (**a**) should be reordered (**b**). The first five columns are identical, but the polarity, the intensity trigger for MS/MS and reference status added. The name can be derived from the protein ID, sequence, ion and ion rank.

a) Transition list Output: (note that the columns are not labelled in the created file)

Precursor mass	fragment mass	Collision energy	Start time	End time	Sequence	Protein ID	ion	ion rank

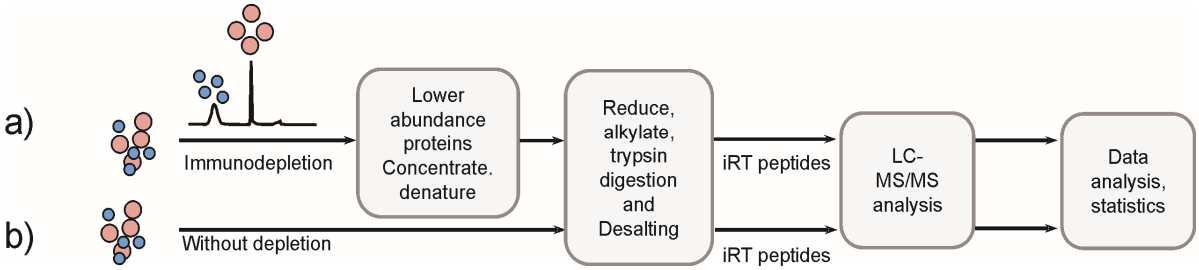
b) Transition list edited for use with a Thermo Scientific TSQ

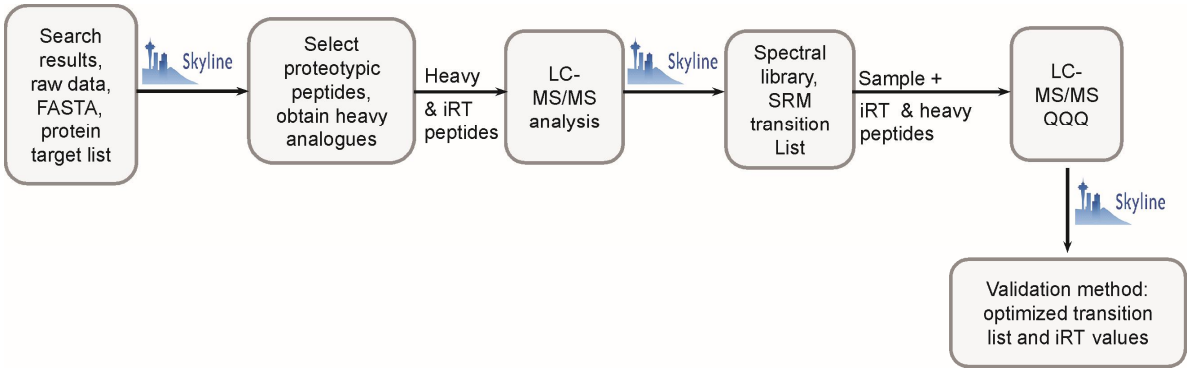
Precursor mass	fragment mass	Collision energy	Start time	End time	Polarity	Trigger	Reference	Name

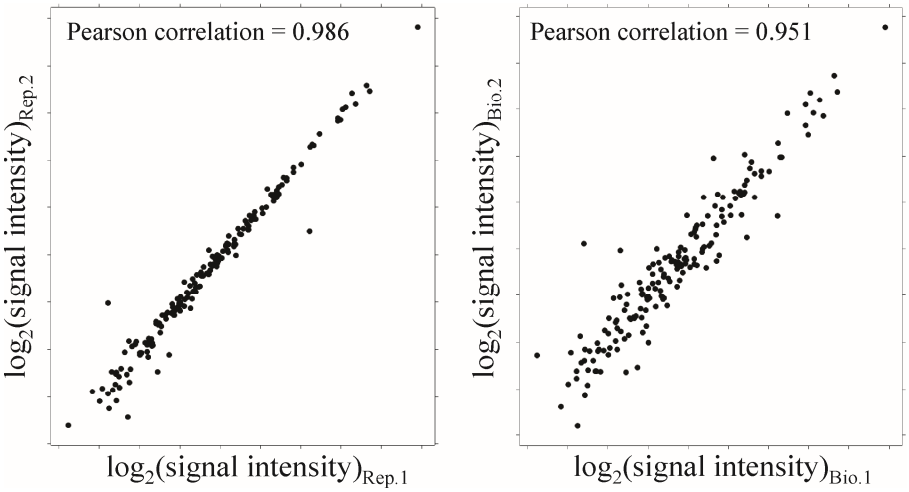
Figure 1. **Sample preparation for LC-MS/MS analysis. a)** From serum samples the high abundance proteins are removed by immunoaffinity depletion and the fraction containing the low abundant proteins concentrated and denatured prior to enzymatic digestion and desalting. **b)** Alternatively, for discovery or targeted validation, the serum is directly digested without depletion. Indexed retention time peptides are added to the samples prior to LC-MS/MS to assist in the development of targeted assays by SRM. Although both approaches (**a** & **b**) can be used in the SRM validation phase, the additional sensitivity provided frequently facilitates directed analysis from the undepleted sera.

Figure 2. Development of an SRM targeted assay. Skyline is used to identify the proteotypic peptides associated with the differentially abundant proteins found in the discovery data. To enable unambiguous confirmation of the tandem mass spectral identifications, isotopically labeled synthetic analogues are obtained. LC-MS/MS analysis is made of the heavy peptides together with indexed retention time (iRT) peptides, for example using a Q Exactive, and the data used to develop a spectral library of the targets. Unscheduled/scheduled analysis of the heavy peptides together with a serum sample using a triple quadrupole mass spectrometer (QQQ) is used to confirm the co-elution of the targets and their isotopes and choose the best transitions for the SRM method. The iRT peptides are used to provide retention time indices for the targets in the sample matrix, which are subsequently used for scheduling the SRM method.

Figure 3. **Scatter plot depicting the correlation between technical replicate (A) and biological replicate (B).** The **Figure 3A** and **B** shows high Pearson correlation coefficient values for analytical replicates reflecting the reproducibility of label free quantification.


a)
b)
Immunodepletion
Without depletion
Lower abundance proteins Concentrate. denature
Reduce, alkylate, trypsin digestion and Desalting
iRT peptides
iRT peptides
LC-MS/MS analysis
Data analysis, statistics


Search results, raw data, FASTA, protein target list
Skyline
Select proteotypic peptides, obtain heavy analogues
Heavy & iRT peptides
LC-MS/MS analysis
Skyline
Spectral library, SRM transition List
Sample + iRT & heavy peptides
LC-MS/MS QQQ
Skyline
Validation method: optimized transition list and iRT values


Pearson correlation = 0.986
log2(signal intensity)Rep.2
log2(signal intensity)Rep.1
Pearson correlation = 0.951
log2(signal intensity)Bio.2
log2(signal intensity)Bio.1