thermo scientific

WHITE PAPER 64994

Global workflows for clinical proteomics research

Keywords
Precision medicine, biomarker, clinical research, BCA analysis, NanoDrop analysis, discovery, proteome characterization, protein panel

Summary

Precision medicine relies on routine assays to determine biological condition or state, and optimal response, including additional or alternative courses of action. The effectiveness of precision medicine is tied to the availability, selectivity, and sensitivity of tests when applied to specific diseases. To expand their testing options, clinical research laboratories are evolving to better develop, refine, and expand assays applicable to specific targets.

This white paper describes liquid chromatography coupled to mass spectrometry (LC-MS) and supporting consumables, and software applied in an integrated clinical proteomics research workflow designed to bridge the gap from discovery to routine methods. Specific workflow parameters, concepts, and routines are described. Evaluation points for assessment of the experimental hypothesis to maximize efficiency and quality are highlighted.

ThermoFisher SCIENTIFIC

Introduction

The drive to develop robust biological tests supporting precision medicine has created renewed emphasis on protein- and/or peptide-based biomarker panel development. New methods have created a paradigm shift from preparing and analyzing samples based exclusively on a narrow set of one to three of markers, to acquiring a comprehensive digitized record of each sample where the stored data can potentially be used to screen numerous panels.

There are two primary experimental steps needed to produce and validate a robust protein/peptide panel that may be used as a clinical research assay. The first is to use global compound profiling to identify a specific, small protein panel that can be incorporated into a routine, targeted data acquisition method. The second transitions generic, global profiling methods into robust and reproducible methods for comprehensive sample digitization across either a longitudinal study or initial trial studies that may lead to target panel selection and incorporation into a final, routine assay. Both use a common workflow of well-defined, integrated experimental steps designed to evaluate the experimental hypothesis and to determine the experimental parameters for the subsequent steps.

Due to large sample sizes and information content, and the need to convert large-scale data generation into actionable methods that can be validated in clinical settings, clinical research presents many challenges. To address these challenges, a standard experimental workflow designed to translate global proteome profiling into putative protein panels for verification and validation is presented here. Each step in the workflow is used to expand the sample-specific knowledge base to more effectively mine data and evaluate results toward the experimental hypothesis. The term “knowledge base” refers to sample preparation conditions, loading amounts, spectral libraries, and proteome coverage.

Figure 1 shows the generic workflow used to perform clinical research. The goal of the workflow is to differentiate the steps taken to 1) identify a set of putative markers routinely and reproducibly as measured by a well-characterized targeted experimental method, and 2) to optimize the standard operating procedures needed to perform comprehensive sample digitization for trial or longitudinal studies. The workflow steps are tightly linked. For example, the initial discovery workflow steps chosen are extremely important in determining the effectiveness and efficiency of the pilot study workflow.

Standards and controls

Because a study may continue over many months or years, well-defined standards are used at each step so initial studies can be linked to those performed towards the end and for intra- and inter-laboratory method transfer and comparison. In addition, standard and/or quality control compounds are used to perform real-time instrument evaluation and assess method robustness to determine when performance falls outside of system suitability thresholds. The objective is to minimize donor sample loss by immediately identifying when the LC-MS system requires maintenance or new UHPLC columns, prior to injecting donor sample.

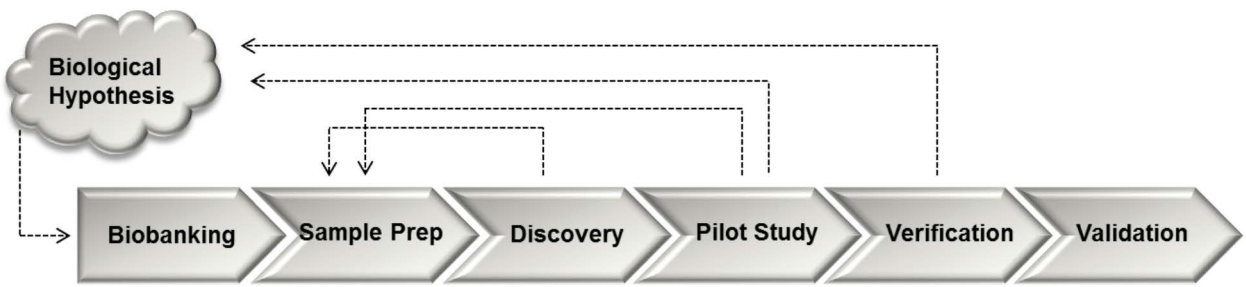

Figure 1. Generic clinical research workflow. Moving from left to right, the results obtained from each step are used to create and add to the sample-specific knowledge base. The cumulative results obtained after each step are used to evaluate the experimental hypothesis before progressing to the next step.

Generic quality control (QC) compounds are used to normalize data and determine system suitability. Generic QC compounds increase data processing rigor, reduce cost, and can be incorporated into many different studies. As research is transitioned and study size is increased, specific QC targets are identified and incorporated into verification and validation steps that can be used for targeted protein/peptide quantitation and global profiling.

Checkpoints

Experimental checkpoints are used to evaluate the data and how it relates to the experimental hypothesis. If the experimental method and results support the hypothesis, then continuation of the study is warranted. If not, then the method is refined, or the hypothesis is re-stated and the experiment re-started. The primary research checkpoint is generally defined as completing a pilot study designed to confidently determine the effectiveness of the experimental method for evaluating the biological question, and the probability of successful completion of larger validation studies and clinical trials, because these require investing significantly more time and capital.

Experimental effectiveness

The effectiveness of the experimental workflow is defined as the efficiency with which the hypothesis can be confidently evaluated and decisions made. Effectiveness is affected by three main factors: 1) the background information used to formulate the hypothesis, 2) the access to well-characterized donor samples (biobanking), and 3) the instrumentation and consumables used to perform the experimental studies.

Biological hypothesis: experiment definition

The biological hypothesis defines all aspects of the overall experiment and each individual experimental step. The hypothesis states the focus of the study in terms of the specific disease or perturbation, and the targets that will be used to test and classify the biological state. The targets can be defined as proteins, specific peptides (covering N- or C-termini), the phosphor- or glycoproteome, or specific cell or vesicle types (e.g., stem cells, exosomes, HDL, etc.). The basis of the perturbation and target class defines the biological matrix from which the samples will be extracted from a biobanking source and compared. At study onset, the hypothesis is also used to stratify the donor pool (e.g., healthy, asymptomatic, acute) and to define the sample collection and sample preparation methods.

Sample preparation and pooling

Sample preparation converts raw biological sample to a state that can be introduced into a mass spectrometer for qualitative and quantitative analysis. For example, studies on Parkinson's disease or myelofibrosis extract peripheral blood mononuclear cells (PBMCs) from whole blood prior to enzymatic digestion and subsequent LC-MS analysis. The study goals dictate the complexity of the sample preparation steps. Several commercially available products can be used for routine sample preparation.

Performed prior to LC-MS analysis, sample pooling (Figure 2) provides important early-stage workflow advantages. First, it conserves sample by reducing the amount extracted from any one sample, while creating test samples representative of the donor class.

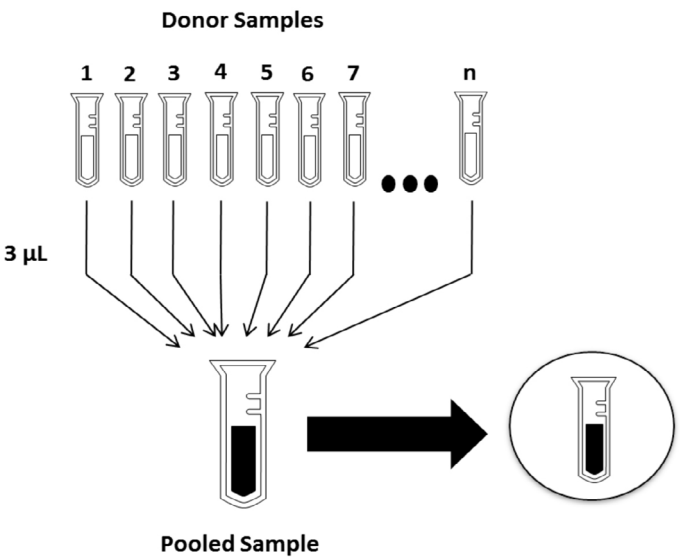

Figure 2. Creating a pooled sample for discovery and quantitation analyses.

Secondly, replicate, pooled sample analysis produces significantly lower biological variance, thus enabling evaluation of the reproducibility of the instrumental analysis. When analyzed during global profiling, the pooled sample plays an important role as a real-time instrument quality control and system suitability test and as a quality control metric for automated data processing.

Discovery

Discovery experiments are designed to fully characterize all aspects of the study, including donor samples and spiking compounds used for quality control, normalization, and quantitative references. They define all of the workflow parameters that will be used for subsequent pilot studies and workflow validation. The discovery-stage results are also important because they are the first results used to evaluate the hypothesis.

Target proteome definition

Sample-specific characterization uses traditional discovery-based proteomics methods to identify and characterize peptide content resulting from enzymatic digestion and to associate identified peptides with their respective protein of origin. Sample-specific content is generated in the form of mass spectral libraries that define which compounds are detectable and their relative abundances based on the sample preparation methods. Defining the target proteome generally focuses on either a small subset of the donor sample set or replicate/fractionation analysis of the pooled sample. This devotes instrument time to acquiring the highest-quality MS and MS/MS data to sample as much of the proteome as possible. Because spectral matching significantly increases the speed and confidence of protein and peptide identification, the spectral libraries generated are used to perform routine data processing in all subsequent steps.

The initial sample protein amount or potential matrix effects may not be known. To establish initial digestion protocols, the protein amount can be determined through non-MS based methods such as BCA analysis or Thermo Scientific™ NanoDrop™ analysis. Replicate analyses of either a small set of samples, or a pooled sample, are processed to evaluate and optimize the sample preparation method.

The growth of clinical research has augmented the approaches to establishing the discovery knowledge base used for subsequent studies. The emergence of cloud-based tools for uploading local large-scale discovery results, with reorganizing and downloading capabilities, allows other research groups to bypass aspects of their own local discovery experiments to more quickly reach the pilot study stage. Additionally, recent developments in *in silico* predictions have added to the global repositories used to process large-scale data. However, knowledge base creation from local discovery-based empirical data provides the greatest confidence in results because all aspects of the sample preparation method are directly linked to spectral library creation. For this reason, this white paper focuses on local sample-specific knowledge base generation.

Spectral library creation

Protein and peptide characterization is performed using high-quality data-dependent acquisition (DDA) methods. These methods incorporate narrow precursor isolation settings (1 to 1.5 Da) and longer maximum ion fill times (80 to 100 ms as compared to the normal setting of 50 ms) to ensure the highest-quality product ion spectrum per peptide.

The two primary methods used to perform comprehensive proteome characterization are shown in Figure 3. The top path shows replicate analyses of pooled sample where different loading amounts are used. Replicate analyses overcome the stochastic sampling error associated with DDA. Replicate analyses also increase the number of fragment spectra acquired and matched per peptide subsequently used to evaluate the statistical representation of peptides and to increase confidence levels for the spectral library. In some instances, increasing the amount loaded on column results in a proportional increase in the number of peptides matched and the spectral quality obtained.

The second approach shown in Figure 3 uses peptide fractionation prior to LC-MS analysis. The number of fractions collected depends on the amount of material available and the complexity of the matrix. Automated fraction collection methods generally set the number of fractions to 12, 24, 48, and 96, where the first and last sets of fractions typically contain few peptides. To ensure the highest degree of confidence in the peptide-specific attributes collected in the spectral library, all fractions are analyzed using the same chromatographic gradient—the gradient that will be used in subsequent routine methods. Increasing the number of fractions will proportionally increase the proteome coverage and the quality of the spectral library entry for each peptide. Fraction collection methods capture and sequence medium- to high-abundance peptides in multiple fractions, and thus increase confidence in retention times, optimal precursor charge states, and product ion distributions. A library produced in this manner provides significantly more coverage and greater sample-specific content than a library produced by replicate analysis.

The spectral library is compiled and then used to establish proteome coverage, relative protein expression levels, and individual protein coverage based on the number of unique peptides confidently and reproducibly measured. The spectral library also helps in evaluating chromatographic performance by overlapping the fraction library with the replicate analysis. The increased number of qualitative metrics provided by spectral matching between the spectra obtained from the sample analysis and the spectral library facilitates more confident evaluation of the original hypothesis.

By the end of the discovery step, all the parameters needed to conduct an effective pilot study are generated. The experimental variance for the sample preparation and data acquisition methods will have also been established.

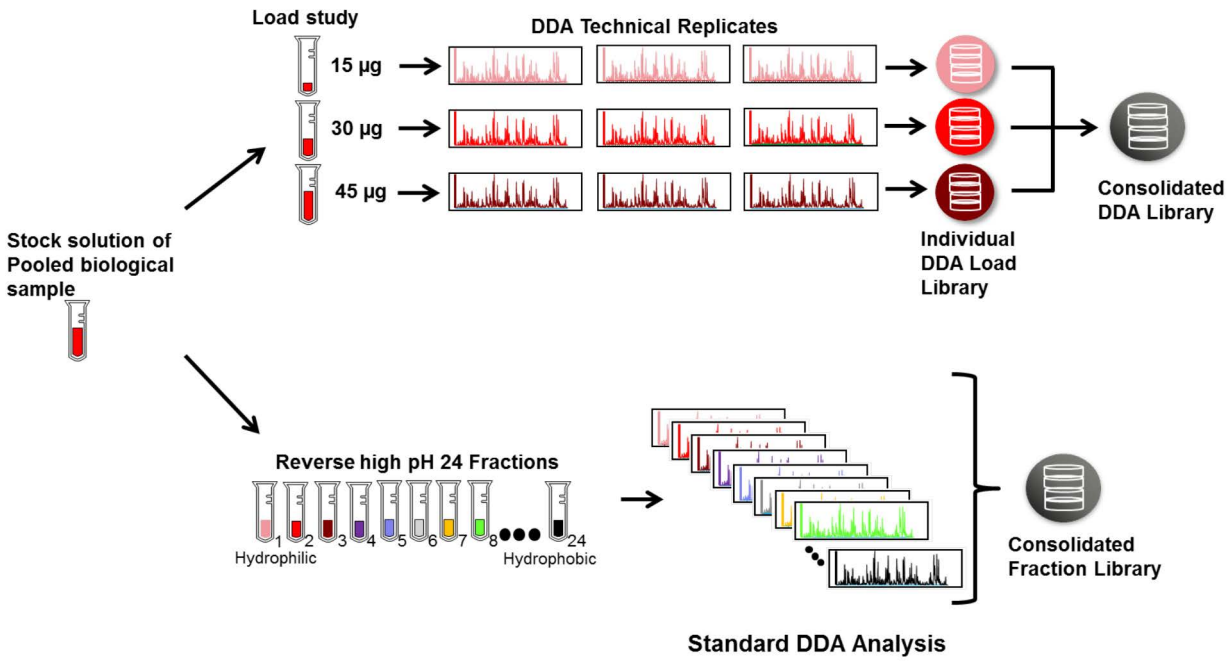

Figure 3. Two methods of creating spectral libraries to establish pilot study data acquisition parameters. Two methods are used to acquire high-quality DDA spectra to reduce the probability of chimeric spectral acquisition. The resulting data can be searched using the sequencing matching routines provided in Thermo Scientific™ Proteome Discoverer™ software or by using de novo sequencing routines.

Pilot study

The pilot study acquires large-scale data sets to profile the global proteome. A set of 20 to 100 samples per donor group is usually targeted for data acquisition. The number of samples in the group is determined by balancing study cost and instrument time against the statistical power needed to accurately determine the global differential expression. Global profiling is performed for two basic reasons. The first is to avoid experimental bias associated with selected reaction monitoring (SRM) or parallel reaction monitoring (PRM) of a select panel at the outset of the pilot study. The second is to create a well-characterized experimental method for the simultaneous detection and quantification of hundreds or even thousands of proteins per sample. Post-acquisition data extraction, integration, and quantification are used to fully digitize sample contents.

QC metrics

To increase the large-scale data quality and longitudinal relevancy, researchers spike equivalent amounts of peptides and proteins into each sample at different stages of the sample preparation protocol (Figure 4). These external samples or normalizing agents provide internal QC metrics to continually evaluate sample preparation, chromatography, and mass spectral performance prior to the injection of the next donor sample. In addition, the spiked QC compounds provide landmarks from which all peptide metrics can be converted from absolute to relative values. Relative values allow researchers to maintain data relevancy throughout the lifetime of the study, which may cover many months or years. This process is critical as the pilot study data acquisition may bridge scheduled and non-scheduled interruptions due to routine instrument cleaning or exchanging chromatographic columns.

The two most common standards used for normalization and QC checks are protein(s) and peptides. Each should be commercially available and contain unique sequences relative to the studies performed. For example, non-human protein sources are spiked into human tissue studies to eliminate potential interference, enabling system-wide evaluation. Protein standards are spiked at the onset of sample preparation so the standards progress through all experimental steps, including digestion and desalting. The resulting peptides can be used to evaluate digestion efficiency, determine the rank order, and calibrate retention times. Peptide standards often contain isotopically enriched amino acids that are used to establish retention time markers to convert absolute retention time measurements of endogenous

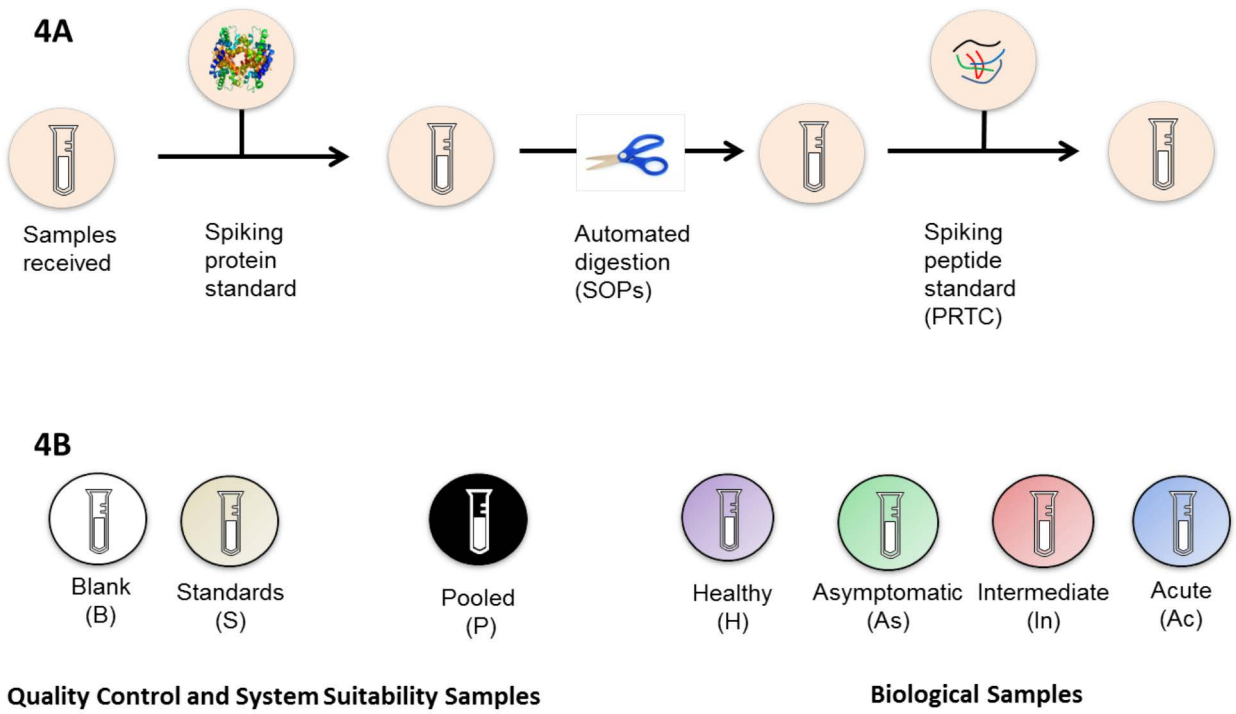

Figure 4. External standard addition at sample preparation. External standards are used for QC and normalization routines during automated data processing. Figure 4A shows external standards spiked before enzymatic digestion and before analysis. Figure 4B lists the different types of samples analyzed in pilot studies.

peptides to their relative values for the purposes of data processing and correlation across longitudinal studies. Peptide kits generally contain between 9 and 15 peptides to reduce the potential for background interference while providing enough markers for accurate retention time modeling.

Data acquisition

After the samples are prepared, pilot study data acquisition is performed. Figure 4B shows the different sample types generally included in a pilot study. Standard and donor samples both play an important role in defining the success of the pilot study. The standards are comprised of blanks (usually containing the peptide standards), and the protein standard mixture. If only the protein standard is used to serve both functions, then one blank is used. Generally analyzed using a ballistic gradient to increase throughput, blanks are used to determine carryover and clean the injector port and trapping column. Containing the same complexity and analyzed using the same experimental conditions as the donor samples, the pooled sample is critical in determining system suitability and in providing a reliable standard for systematic evaluation of the automated software routines used. Replicate injections are extracted from the same vial to determine systematic and biological variance.

The donor sources, samples, and color-coding and labels shown in Figure 4B will be used in the following theoretical example. Typically, equivalent numbers of samples from donors classified as “healthy/normal” are used as references to which samples from all other donor groups are compared. “Asymptomatic,” “independent,” and “acute” donors help to further stratify disease progression and may also have subgroups within each donor class. It is important to study well-characterized donors to reduce biological variance when performing differential expression analysis.

Figure 5 shows an example acquisition sequence for a large-scale pilot study. The data acquisition order for the donor samples is randomized and a set of standards is acquired at defined time intervals. The acquisition loop count of donor samples to standards can be modulated based on the number of samples (length of study), sample cleanliness, and general expectation of instrument cleaning intervals. The number of injections of pooled sample and different donor group sample should be equivalent.

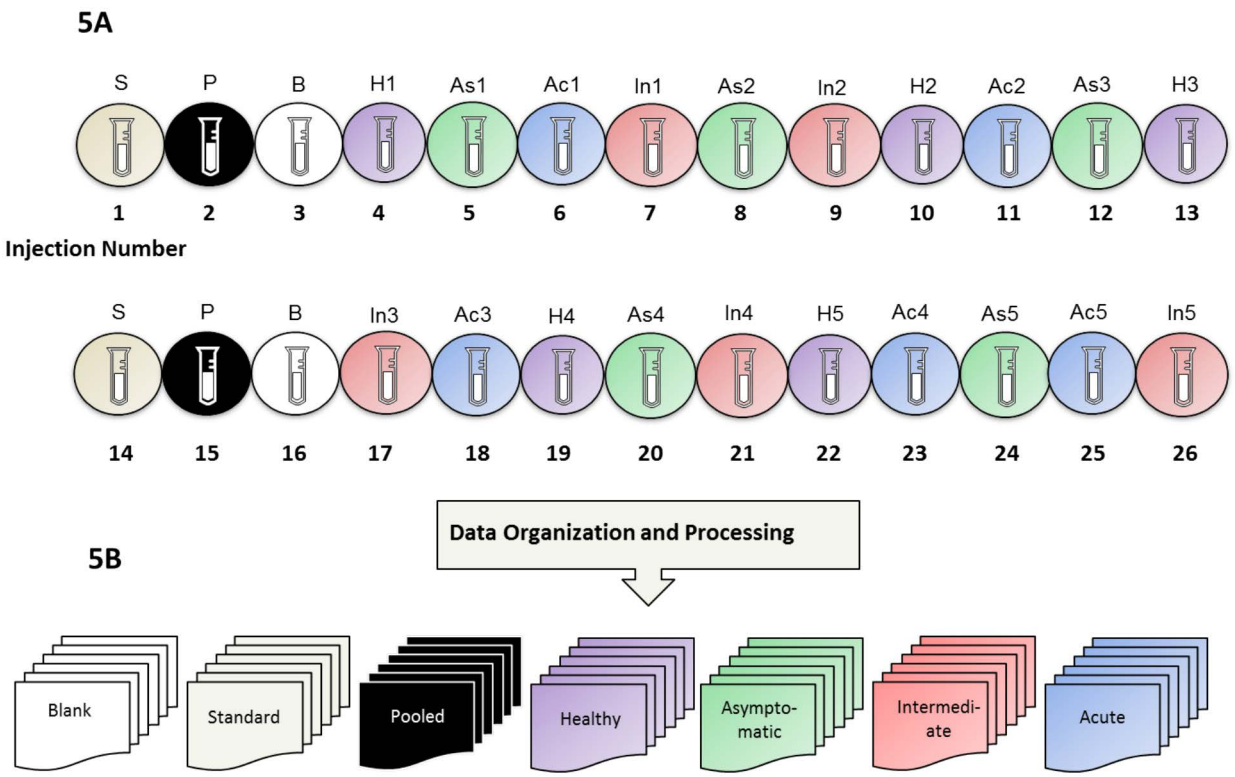

Figure 5. Example protocol for a random sample acquisition sequence for large-scale pilot studies. The acquisition sequence uses randomized donor sample data acquisition with a set of blanks and QC samples analyzed at a set frequency. Figure 5B shows the post-acquisition data organization for automated global proteome profiling.

The number of samples available seldom enables technical replicate acquisition for each sample. Therefore, standards spiked into each sample provide a means to evaluate the reproducibility of data acquisition across all samples. Analysis of blanks (usually spiked with a standard peptide kit) and/or protein standard digests, allow measurement of system suitability and reproducibility. In addition, periodic acquisition of pooled sample provides a QC sample for automated data processing and a robust method to determine experimental variance. The number of technical replicates of pooled sample should be similar to the number of donor samples analyzed.

Data processing

Figure 5B shows the organization of data files prior to automated processing. The files are grouped based on content or description. Automated data processing software should facilitate both the pilot study and QC/system suitability methods. That is, the software should be able to support processing, interpretation, and scoring based on the origin of data. The QC and pooled RAW file data processing are used to evaluate overall instrument performance, establish experimental variance, and help determine biological variation following global profiling of all donor files.

Global proteome profiling of the large-scale data is used to stratify and credential the resulting protein list to help determine biological relevancy. The goal of large-scale data processing is to convert individual protein and peptide statistics to the global statistics needed to credential candidates and to stratify which proteins and peptides exhibit measured differential expression between the control/healthy and disease classes.

The data processing routine evaluates all peptides through spectral matching and performs qualitative and quantitative analysis of all data files. The individual results are then assembled into groups for variance analysis and to determine the absolute and/or relative area-under-curve (AUC) values and ratios between experimentally defined groups. Accompanying statistical information, such as %CV and p-values, and overall score are calculated. All peptides are then assembled under the protein of origin, and the protein statistics and scoring are determined. The integrated AUC values for each biological group are compared to determine the differential expression. The overall results are then presented as the large-scale data analysis.

Figure 6 shows the data processing workflow. First, the AUC ratios, receiver operator curves (ROCs)—provided there are enough samples per group to be statistically relevant—and/or p-values are determined (Figure 6B). After the protein- and peptide-specific ratios and corresponding p-values are determined for all groups, the p-values can be used to stratify the proteins (or peptides) from the most confident p-value to the least, based on the acute disease state, as shown in Figure 6C. The stratification for the acute state generates statistically relevant target proteins, which are then used to interrogate all other groups to determine the distinguishing potential for the stratified set of proteins. The hypothetic results displayed in Figure 6C show an ideal situation where the majority of proteins stratified in the acute group are not collectively stratified in other biological groups. To determine biological relevancy and evaluate the experimental hypothesis, the set of proteins and corresponding AUC ratios can be evaluated using pathway databases [e.g. Thermo Scientific™ ProteinCenter™ software, Ingenuity® Pathway Analysis (IPA®), or Cytoscape®].

Verification and validation

Verification and validation use the set of targets identified in the discovery step to refine the experimental method and to increase method throughput and quality. Even though the experimental procedure may still involve global profiling, incorporating target-specific peptide standards becomes essential for verification of chromatographic and mass spectral parameters such as retention time, precursor charge state distribution, and product ion distribution. Target-specific peptide standards are differentiated from endogenous analogs through isotopic enrichment at one or two residues, and can be introduced into each sample as incorporated into a recombinant protein or peptides. Stable isotope-labeled (SILs) peptides are spiked into the biological sample at levels estimated to be equivalent to endogenous levels to minimize adverse experimental effects. Each SIL can be spiked at different levels based on pilot study results. SILs can then also be used as QC and normalizing components.

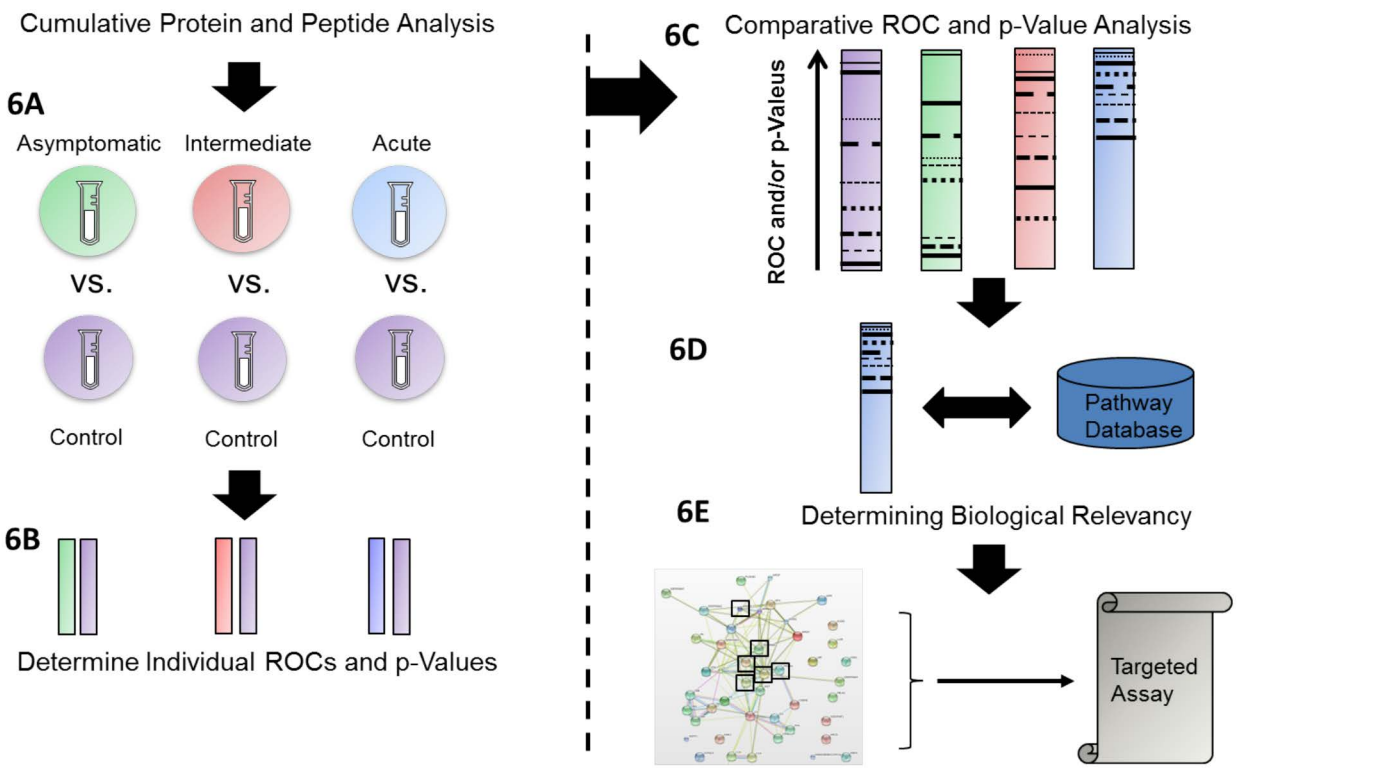

Figure 6. Credentialing proteins and/or peptides using large-scale data acquisition: workflow steps needed to convert the raw AUC values and scores at the protein and peptide levels to a potentially distinguishing protein panel. The measured protein and peptide AUC values in the individual groups are compared to the control/healthy results to determine the AUC expression ratios (Figure 6A) and to calculate the statistical significance as defined by either ROC scores, p-values, or both (Figure 6B). The proteins are then stratified based on statistical relevancy and compared across each donor group (Figure 6C). To determine biological relevancy, the set of proteins identified can be evaluated against pathway or gene ontology databases (Figure 6D). The list of identified proteins can then be transitioned into routine validation assays.

Validation studies define the instrument response factor for the compound amounts in the sample. In other words, validation ensures that instrument response accurately correlates to the endogenous levels across the expected biological expression level. Similar experimental procedures are used for both global profiling and targeted data acquisition methods. For targeted methods, additional precursor mass values associated with the SILs are included in the instrument method. Typically, quantitation curves are developed by spiking SILs into a static amount of biological sample at known levels. The spiking levels should cover the expected biological expression level to ensure accurate correlation between instrument response and levels. Key metrics evaluated in the validation step determine the optimal instrument response to the spiked levels (e.g., linear, weighted linear, or log-log), the overall goodness of fit collectively for all levels as represented by the regression value, the precision for each level measurement (coefficient of variance), and the accuracy for each level as defined by recovery (relative standard deviation).

If the results from the validation study continue to support the original hypothesis, the methods can be moved on to a clinical trial for further analysis.

Conclusion

Precision medicine relies on robust LC-MS assays. An integrated clinical proteomics LC-MS workflow designed to translate global proteome profiling into putative protein panels for verification and validation was presented. Each step in the workflow expands the sample-specific knowledge base to more effectively mine data and evaluate results toward confirming or rejecting the experimental hypothesis and ultimately leading clinical trials if appropriate.

Resources

Clinical & Translational Research Learning Center

Protein Analysis in Clinical & Translational Research

Mertins, P.; Mani, D.R.; Ruggles, K.V. et al. Proteogenomics connects somatic mutations to signaling in breast cancer. *Nature.* **2016,** *534*(7605), 55–62.

Stewart, P.; Fang, B.; Slebos, R. J. C. et al. Relative protein quantification and accessible biology in lung tumor proteomes from four LC-MS/MS discovery platforms. *Proteomics.* **2017,** *17,* 1600300.

For Research Use Only. Not for use in diagnostic procedures.

Find out more at **thermofisher.com/clinicalresearchsolutions**

ThermoFisher SCIENTIFIC