Label-free mass spectrometry-based protein quantification technologies in proteomic analysis


Abstract

Major technological advances have made proteomics an extremely active field for biomarker discovery and validation in recent years. These improvements have lead to an increased emphasis on larger scale, faster and more efficient methods for protein biomarker discoveries in human tissues, cells and biofluids. However, most current proteomic methodologies for biomarker discovery and validation are not highly automated and generally labour intensive and expensive. Improved automation as well as software programs capable of handling a large amount of data are essential in order to reduce the cost of discovery and increase the throughput. In this review, we will discuss and describe the label-free mass spectrometry-based protein quantification technologies and a case study utilizing one of these methods for biomarker discovery.

Keywords: label-free quantitative analysis; biomarkers; proteomics; mass spectrometry

INTRODUCTION

Quantitative proteomics has become a widely applied approach for protein identification and quantification in complex biological samples [1–6]. One of the goals of proteomics is to measure protein expression and characterize the protein expression profile in specific tissues and biofluids for potential biomarker discovery. Even though tremendous efforts have been made to improve the proteomic technologies, there are still numerous challenges in analysing global protein expression due to the inherent complexity of biological samples. These challenges include: (i) sensitivity of the instrument and ability to identify novel proteins; (ii) low-to-moderate throughput of the system from sample preparation to data analysis; (iii) wide coverage of protein mass and abundance (dynamic range); and (iv) the ability to quantitatively analyse protein expression. At the present time, there is a significant doubt within the field of proteomics that any one technology can attain a complete and quantitative protein coverage of all proteins in a given tissue or biofluid. The most commonly used proteomic analysis approach is accomplished by a combination of either two-dimensional gel electrophoresis (2DE) or liquid chromatography (LC) to separate and visualize proteins/peptides and mass spectrometry (MS) to identify, characterize and quantify them.
While 2DE has been the workhorse in proteomics research efforts in the past decade, its lack of ability to widen the protein dynamic range as well as the inability to analyse hydrophobic or very high or low molecular weight proteins, still remains as its biggest disadvantage. One alternative approach to 2DE is the non-gel, ‘bottom-up’ LC-coupled MS-based shotgun proteomics [7–10]. Although some successes using stable isotopic labelling technology for protein quantification have been reported [11], it remains technically difficult to comprehensively characterize the global proteome due to the high costs of the labelling reagents and the nature of the methodology (i.e. in iTRAQ or ICAT approach it depends on labelling efficiency, [12–14]). Furthermore, simultaneous quantification of proteins from multiple samples is problematic due to limitation of available labelling reagents. Recently, the ion intensity-based, label-free quantitative proteomics (LFQP) approach has gradually gained more popularity as mass spectrometer performance has improved significantly [15–18]. LFQP provides a powerful tool to resolve and identify thousands of proteins from a complex biological sample. In this approach, proteins are first digested with a protease into a peptide mixture and subsequently analysed by tandem MS (MS/MS) and identified by database searching. Relative protein abundance is determined by either spectral counting or chromatographic peak intensity measurements (Figure 1). The LFQP approach is rapid and more sensitive than many other proteomic methods, and it increases the protein dynamic range 3- to 4-fold as compared to 2DE [19]. This method can also be automated and has the ability for large-scale proteome analysis.

**OVERVIEW**

Label-free and signal intensity-based relative protein quantification (shotgun proteomics) approaches are promising alternatives to stable isotope labelling approaches, and have been applied to quantify differentially expressed proteins from complex biological samples [20–23]. Currently there are two major widely used but fundamentally different label-free protein quantification strategies: (i) spectral counting—counting and comparing the number of fragment spectra identifying peptides of a given protein to assess relative protein abundance [10, 24, 25]; and (ii) peptide chromatographic peak intensity measurements—measuring and comparing the chromatographic peaks of peptide precursor ions belonging to a specific protein [17, 21–25]. These strategies are simple and cost-effective and have demonstrated high reproducibility and linearity at both peptide and protein levels [19]. Inclusion of statistical analysis with these approaches allows detection of small significant changes that are biologically meaningful. Several studies have demonstrated that both spectral counting and extracted ion chromatograms (XICs) of selected peptide ions correlate well with protein abundances in complex biological samples [17, 18, 23, 26, 27]. The application of these methods, however, especially to mammalian systems, requires more robust computing power and algorithm capable of handling chromatographic peak alignment, spectral counting and peptide ion intensity measurements for analysing changes in protein abundances in complex biological samples.

**LABEL-FREE PROTEIN QUANTIFICATION METHOD**

All LFPQ methods discussed previously include four fundamental steps as depicted in Figure 1: (i) sample preparation—protein extraction, reduction, alkylation and digestion; (ii) sample separation and analysis—LC and MS/MS; (iii) data analysis—identification, quantification and statistical analysis; and (iv) data interpretation—protein–protein interaction network and pathway analysis.

**Protein extraction, reduction, alkylation and digestion**

In general, proteins are extracted from tissues, cultured cells or biofluids in freshly made lysis buffer containing 8 M urea and 10 mM DTT. In order to take the same amount of proteins from each
sample, protein concentrations are determined by Bradford assay [28]. The same lysis buffer is normally used as the background reference for protein assay in order to obtain a relatively accurate measurement among all samples (due to the presence of urea in lysis buffer).

Resulting protein extracts are subsequently reduced and alkylated with DTT and iodoacetamide to block sulfhydryl groups in proteins. Alternatively, the volatile reagents triethylphosphine and iodoethanol are also used to minimize sample preparation variations [29]. Protein mixtures are then digested by trypsin and tryptic digests are generally filtered with 0.45 μm spin filters before applied to HPLC to avoid column clogging.

**Mass spectrometric analysis**

All digested samples should be randomized for injection order to remove systematic bias from data acquisition. Typically <20 μg of the tryptic peptides are required for each injection onto a C18 microbore column (i.d. = 1 mm, length = 5 cm) or nano/capillary column (i.d. = 100 μm, length = 5 cm). Peptides are typically eluted with a linear gradient from 5% to 45% acetonitrile developed over 120 min at a flow rate of 50 μl/min (or 500 nl/min for nano system) and effluent is electro-sprayed into a mass spectrometer (LTQ’s linear ion trap from ThermoFisher Scientific or Q-TOFs from Waters). In spectral counting approach, the more of a particular protein is present in a sample, the more MS/MS spectra are collected for peptides of that protein, thus relative abundance of that protein can be obtained by comparing the number of MS/MS spectra among a set of experiments. The advantage of this approach is that both protein identification and protein quantification can be achieved by extensive MS/MS data collection across the chromatographic time scale simultaneously. The disadvantage of this method, however, is its dependency on the quality of MS/MS peptide identification, because errors in peptide identification can lead to inaccurate protein quantification [19, 30]. Nonetheless, in practice, the correlation between the amount of protein and number of MS/MS spectra indeed holds and many investigators are using it for quantitative protein profiling of complex biological samples [10, 24, 25, 31]. The other approach is based on peptide ion intensity measurements by integrating area under the curve (AUC) and comparing them for their relative abundance [27]. It requires that all the MS data be collected in the data-dependent ‘Triple-Play’ mode as shown in Figure 2 (MS scan, Zoom scan and MS/MS scan). These three important experimental parameters plus chromatographic retention time determine the analytical accuracy of protein identification and protein quantification by ion intensity [27]. The advantage of this approach is the influence of interfering signals (from peptides or other contaminants) of similar mass but distinct peptide sequences can be distinguished based on zoom scan and MS/MS data, minimizing false positives both in protein identification and protein quantification. However, because it requires peak alignment by a special algorithm for accurate protein quantification, it is not trivial to perform LC-MS runs reproducibly, especially for an experiment involving a large set of samples (i.e. n > 40).

**Protein identification**

In general, MASCOT, SEQUEST and X! Tandem database search algorithms are used for peptide sequence identification. Each algorithm compares the observed peptide MS/MS spectra and theoretically derived spectra from the database to assign quality scores. These quality scores and other important predictors are combined in the algorithm that assigns an overall score to each peptide.

The confidence in protein identification is increased with the number of distinct amino acid sequences identified. Therefore, proteins are normally categorized into different priority group depending on whether they have only one or multiple unique sequences of the required peptide identification confidence [32]. For example, in the method developed by Higgs et al. [27], a protein is classified in the higher priority group if it has at least two distinct amino acid sequences with the required ID confidence (Table 1); otherwise it is classified in the lower priority group. Many would view any protein identification with a ‘single-hit-wonder’ as questionable [33].

**Protein quantification**

In spectral counting, protein quantification is achieved by comparing the number of MS/MS spectra from the same protein as described earlier [10, 24, 25]. In the method by peak intensity measurements, the key is the chromatographic peak alignment. Due to the fact that large biomarker studies can produce chromatographic shifts as a result of multiple injections of the samples onto the same
Figure 2: ‘Triple-play’ MS method. After the effluent is electro-sprayed into a mass spectrometer, primary MS scans are acquired for peptide mass determination, followed by zoom scans for charge state determination, MS/MS scans for peptide sequence determination and database searching for protein identification. Every peak above the ion intensity threshold will be selected for ‘triple-play’. 
HPLC column, this peak alignment is critical in order to provide the most accurate comparative data. Un-aligned peak comparison will result in larger variability and inaccuracy in peptide quantification [27]. A graphical example of a comparison of peptide quantities is shown in Figure 3. To be qualified for protein quantification procedure, each aligned peak must match precursor ion (MS data), charge state (zoom scan data), fragment ions (MS/MS data) and retention time. After alignment, the AUCs for individually aligned peaks from identified peptides from each sample are computed; the AUCs are then compared for relative protein abundance. Another important and critical procedure in protein quantification is data normalization. Advanced data filtering and processing procedures are required to extract meaningful information in order to correct global errors in data quality. Normalization can correct global drifts occurring at various levels during multiple MS data acquisitions. For example, in the method described by Higgs et al. [27], the data are normalized using quantile normalization method [34]. Quantile normalization is a method of normalization that essentially ensures that every sample has a peptide intensity histogram of the same scale, location and shape. This normalization procedure removes trends introduced by sample handling, sample preparation, total protein differences and changes in instrument sensitivity while running multiple samples.

**Statistical analysis**

The number of significant changes between groups, the fold changes (FCs) and the variability (coefficient of variation or CV) can be summarized for each priority level. The threshold for significance is set to control the false discovery rate (FDR) for each comparison at an investigator-desired percentile, normally 5% [35]. The FDR is estimated by the $q$-value which is an adjusted $P$-value. The FDR is the proportion of significant changes that are false positives. If proteins with a $q$-value $\leq 0.05$ are declared significant, it is expected that 5% of the declared changes will be false positives. In the method described by Higgs et al. [27], the $P$-value to $q$-value adjustment is done separately for Priority 1, Priority 2 and the MODERATE confidence categories (Table 2).

**FC** is computed from the means on the AUC scale (antilog) as follows:

**FC** = \[ \frac{\text{mean treated group}}{\text{mean control group}} \]

**When** mean treated group $\geq$ mean control group (up-regulation)

**FC** = \[ -\frac{\text{mean control group}}{\text{mean treated group}} \]

**When** mean control group $>$ mean treated group (down-regulation)

---

**Table I:** Prioritization of protein identification

<table>
<thead>
<tr>
<th>Priority</th>
<th>Protein ID confidence</th>
<th>Multiple sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High (90–100%)</td>
<td>Yes (≥2 unique sequences)</td>
</tr>
<tr>
<td>2</td>
<td>High (90–100%)</td>
<td>No (single sequence)</td>
</tr>
<tr>
<td>3</td>
<td>Moderate (75–89%)</td>
<td>Yes (≥2 unique sequences)</td>
</tr>
<tr>
<td>4</td>
<td>Moderate (75–89%)</td>
<td>No (single sequence)</td>
</tr>
</tbody>
</table>

Protein ID confidence is based on peptide ID confidence assigned by an algorithm developed by Higgs et al. [27, 32].

**Figure 3:** Peptide quantification by extracted ion chromatograms (XIC). (A and B) XIC from control and treated samples, respectively; Arrows indicate the peak of interest. (C) Bar graph of the relative quantity of the peptide of interest [as indicated by arrows (A) and (B)] can be calculated and compared by integrating the AUC.
Absolute \( FC = |FC| = \) absolute or positive value of the FC

An FC of 1 means there is no change.

Also the median %CV for each priority level is determined. The %CV is the SD/mean on a % scale. The %CV is given both for the replicate variation as well as the combined replicate plus sample variation (Table 2).

**Software for protein quantification**

Many open-source, commercial and custom (in-house proprietary) software packages are developed for label-free protein quantification purposes. Examples of open-source algorithms include MzMine (VTT Finland) [36, 37], MSight (SIB) [38], PEPPer (Broad MIT) [39, 40] and SuperHirn (IMSB@ETH) [41]; commercial packages include SIEVE (ThermoFisher Scientific, Waltham, MA), PLGS Identity Expression Informatics (Waters, Milford, MA) [42–45], Rosetta Elucidator (Rosetta Biosoftware) and MS-Xelerator (MsMetrix). Custom proprietary algorithms are normally instrument dependent, i.e., the algorithm developed by Higgs et al. [27] extracts XICs from .RAW files of LCQ or LTQ MS data. All these software tools allow for data format conversion, LC-MS visualization, data filtering, peak alignment and normalization and data quality evaluation [46]. Almost all are still in development and have room for improvement. User interface and calculation efficiency and accuracy of the processing workflow as well as the ability to handle a large data set can be improved. With these improvements, comparative LC/MS-based label-free protein quantification platform will have a tremendous potential for applications in biomarker discovery and validation.

**Quality assurance and quality control**

To assess the stability of the HPLC system and MS instrument, a known purified standard protein is commonly spiked into every sample at a constant amount as an internal reference for assessment of technical variations before tryptic digestion. Several considerations should be given for the selection of the standard: (i) the protein must not come from the same species as the sample of interest; (ii) a series of signature peptides should be easily detected and identified by the instrument; and (iii) the amount of the standard protein spiked into each sample should be comparable to the amount of median abundant proteins in the sample. After global protein identification and quantification, these peptides and their relative quantities should be inspected for QA/QC purpose.

**Data interpretation—pathway analysis and protein classification**

To better understand the biological significance of the protein expression changes, the results (including protein IDs and FC from the LC/MS analysis) can then be analysed using protein–protein interaction and/or pathway analysis software. The software allows for creation of protein–protein interaction networks, biological pathways and gene regulation networks from a data set, which will help better understand specific biological processes that are involved in a particular study. An in-depth systems biology understanding must also be applied to the resulting data in order to develop sound follow-up biomarker experiments.

**A CASE STUDY—BIOMARKER DISCOVERY FOR CISPLATIN RESISTANCE IN HUMAN OVARIAN CANCER CELLS**

To illustrate the utility of this LC/MS-based method described previously, a practical example is provided that outlines how a proteomic experiment is

<table>
<thead>
<tr>
<th>Protein priority</th>
<th>Peptide ID confidence</th>
<th>Multiple sequences</th>
<th>Number of proteins</th>
<th>Number of significant changes</th>
<th>Max absolute fold change</th>
<th>Median %CV replicate</th>
<th>Median %CV replicate + sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High</td>
<td>Yes</td>
<td>855</td>
<td>95</td>
<td>5.51</td>
<td>11.71</td>
<td>12.87</td>
</tr>
<tr>
<td>2</td>
<td>High</td>
<td>No</td>
<td>583</td>
<td>25</td>
<td>18.80</td>
<td>21.58</td>
<td>23.65</td>
</tr>
<tr>
<td>3</td>
<td>Moderate</td>
<td>Yes</td>
<td>27</td>
<td>0</td>
<td>5.12</td>
<td>22.03</td>
<td>24.77</td>
</tr>
<tr>
<td>4</td>
<td>Moderate</td>
<td>No</td>
<td>652</td>
<td>16</td>
<td>78.77</td>
<td>28.70</td>
<td>31.95</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td>2117</td>
<td>136</td>
<td>78.77</td>
<td>1796</td>
<td>19.55</td>
</tr>
</tbody>
</table>
designed and what steps are involved in data interpretation. Prior to any experiment, study design must be completed with a statistician in order to ensure that the study answers the questions of interest and has sufficient technical and biological replicates to detect small but significant changes using appropriate statistical methods. A technical replicate is a replicate sample from the same biological sample. For example, split a single biological sample into two parts and run both replicates in the experiment. This will allow for assessment of instrument errors. Biological replicates are samples from independent experimental units (e.g. each of the 10 human plasma samples from 10 different individuals). While a technical replicate estimates the precision for the assay itself, biological replicates provide an estimation of biological variation [47, 48]. In general, biological replicates are more informative than technical replicates.

Group size determination depends on the size of the effect to be detected (FC) and the sample-to-sample biological variation expected (CV), and which error rates to be controlled. It is best to control the FDR instead of the false positive rate (FPR) when hundreds of proteins are analysed. The FDR can be large (e.g. >0.05) even if the FPR is small (e.g. <0.05). If control of FDR is chosen, then the proportion of proteins that will change (the prevalence) has to be estimated. With this information the group size required for given power (probability of determining a true change, i.e. the sensitivity) can be computed. Table 3 shows a suggested group size with given FC and % CV. As the percent of proteins expected to change varies, the group size required should be matched accordingly.

The goal of this case study was to discover potential protein biomarker candidates of cisplatin resistance in human ovarian cancer by quantitatively comparing the expression profiles of cisplatin-sensitive and cisplatin-resistant cell lines.

### Cell lines and study design

Two pairs of cell lines, A2780 and 2008 cisplatin-sensitive human ovarian cancer cell lines and their resistant counterparts, A2780/CP and 2008/C13, were used in this study. These cell lines have been used as model systems for drug-resistance studies [49–52]. All cell lines were handled under identical conditions and maintained at 37°C in a humidified incubator containing 5% CO₂ in RPMI-1640 media supplemented with 15% fetal bovine serum. Proteins were prepared and subjected to LC/MS/MS analysis as previously described [27].

Two sets of cell lines A2780 (cisplatin-sensitive) and its resistant counterpart, A2780CP, and 2008 (cisplatin-sensitive) and its resistant counterpart, 2008C13, were used for the study. Cell growing conditions for Groups 1 and 2 and Groups 5 and 6 were the same, while Groups 3 and 4 and Groups 7 and 8 included 10 and 20 μM of cisplatin in the media, respectively.

### Table 3: Group size determination (the power is fixed at 95%, the FDR is 5%)

<table>
<thead>
<tr>
<th>Fold change</th>
<th>%CV (%Error)</th>
<th>%Proteins changed</th>
<th>Group size</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>20</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>1.5</td>
<td>20</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>1.5</td>
<td>20</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>1.5</td>
<td>15</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>1.5</td>
<td>15</td>
<td>10</td>
<td>6</td>
</tr>
</tbody>
</table>

### Table 4: Experimental design for Biomarker discovery of Cisplatin resistance

<table>
<thead>
<tr>
<th>Group number</th>
<th>Group name</th>
<th>Group abbreviation</th>
<th>Cell type</th>
<th>Cisplatin treatment (μM)</th>
<th>Number of samples per group</th>
<th>Total number of replicate injections per group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A2780</td>
<td>CIS</td>
<td>Sensitive</td>
<td>0</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>A2780CP</td>
<td>CIR00</td>
<td>Resistant</td>
<td>0</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>A2780CP-10</td>
<td>CIR10</td>
<td>Resistant</td>
<td>20</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>A2780CP-20</td>
<td>CIR20</td>
<td>Resistant</td>
<td>0</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>2008</td>
<td>C25</td>
<td>Sensitive</td>
<td>0</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>2008C13</td>
<td>C2R00</td>
<td>Resistant</td>
<td>0</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>2008C13-10</td>
<td>C2R10</td>
<td>Resistant</td>
<td>10</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>2008C13-20</td>
<td>C2R20</td>
<td>Resistant</td>
<td>20</td>
<td>6</td>
<td>12</td>
</tr>
</tbody>
</table>
There were six samples per group except for Group 3, yielding a total of 93 randomized HPLC injections. Samples were run on a Surveyor HPLC system (ThermoFisher Scientific, Waltham, MA) with a C18 microbore column (Zorbax 300SB-C18, 1 × 5 cm).

**Data normalization and quality assessment**

In this study, all injections were performed using the same C18 microbore column. Chicken lysozyme was spiked into every sample at a constant amount before tryptic digestion. After tryptic digestion, nine chicken lysozyme peptides were identified and quantified (all peptides with ID confidence >75%).

In the plot shown in Figure 4, the individual protein quantities (as presented by the peak intensities) are displayed for each injection. The overall mean for each group is displayed by the line across the plot. This plot presents a visual quality control assessment of the ability of the instrument to measure a constant amount of protein over sequential injections. Since a constant amount of chicken lysozyme was spiked into all the samples, it should show no significant change between groups. If there is a significant group effect (i.e., if q-value <0.05), then one should interpret significant changes in other proteins with smaller FCs as possibly due to spurious effects. In this experiment, the largest absolute FC for chicken lysozyme was 1.16 with a q-value of 0.0667. Even though this is not a significant change based on the q-value threshold of <0.05, any significant FCs of absolute magnitude <1.16 in this study should be interpreted with caution.

**Table 5:** Biological processes in which the differentially expressed proteins are involved

<table>
<thead>
<tr>
<th>Biological process</th>
<th>Number of proteins involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplification</td>
<td>6</td>
</tr>
<tr>
<td>Angiogenesis</td>
<td>2</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>24</td>
</tr>
<tr>
<td>Differentiation</td>
<td>16</td>
</tr>
<tr>
<td>Mitosis</td>
<td>9</td>
</tr>
<tr>
<td>Regulation of signal transduction</td>
<td>13</td>
</tr>
<tr>
<td>Cell survival</td>
<td>16</td>
</tr>
<tr>
<td>Damage</td>
<td>6</td>
</tr>
<tr>
<td>DNA damage recognition</td>
<td>2</td>
</tr>
<tr>
<td>DNA recombination</td>
<td>35</td>
</tr>
<tr>
<td>Transformation</td>
<td>8</td>
</tr>
</tbody>
</table>

*Protein classification based on the analysis by PathwayStudio® 4.0. All proteins (Priority 1–4) with significant changes (q < 0.05) are included.
2117 proteins were identified and quantified from all samples (summarized in Table 2). Among them, 855 proteins were in the Priority 1 group and 95 of them showed significant changes that are potentially associated with cisplatin resistance. The significance threshold is set to control the FDR at <5%. The replicate median % CV (technical variation) for the Priority 1 proteins was 11.71% and the combined replicate plus sample median % CV was 12.87% (technical plus biological variations). There were also 41 proteins that had significant changes among the 1262 proteins that were less confidently identified (Priorities 2–4) (136 total proteins as shown in Table 2). The cisplatin resistance-specific protein changes were extracted when the two types of cell lines were directly compared (A2780/A2780-CP versus 2008/2008-C13*/C35.25), suggesting that other protein expression level changes in the two cell lines may be cell line-specific.

Data interpretation
The results from the LC/MS analysis were then analysed by PathwayStudio® 4.0 (PathwayStudio® 4.0, Ariadne Genomics, Rockville, MD, USA). This software allows for placement of proteins onto biological pathways, gene regulation networks and protein–protein interaction diagrams from a data set. With a list of gene names derived from the proteomic results, more than 1700 connections were made from the protein list. A ‘Classification Table’ was thus created based on these results. Table 5 shows the biological processes that these proteins were found to be involved in when they were run against the software’s ResNet database. These processes included amplification, angiogenesis, apoptosis and more. Based on the experimental data and the information obtained from the pathway analysis and other literature searches, a hypothesis can be generated involving multiple specific pathways that may contribute to cisplatin drug resistance (Figure 5) [53].

CLOSING REMARKS
Although the need for the development of high-throughput, sensitive and cost-effective MS-based quantitative protein analysis methods for biomarker discovery remains a significant challenge in proteomics, the ability of emerging proteomic technologies to enhance biomarker discovery has been shown in this review. To reduce false-positive discoveries, significant development on bioinformatics and robust validation methods will be required. As the future of proteomics unfolds, it will certainly create many more opportunities for the development of novel drug targets, stratification of patient populations, earlier therapeutic development choices and improved understanding of proteomic data and underlying biology. Additional developments in label-free protein quantification technologies will no doubt help to meet the demands of both proteomics and clinical applications.

Key Points
- Label-free LC/MS-based protein quantification method is reviewed in detail including protein identification, quantification, statistical data analysis and software packages for data analysis.
- A case study is presented describing the application of the method.
Acknowledgements
The authors thank Ms. Guihong Qi for assisting the case study. The authors would like to thank support from the Indiana 21st Century Research and Technology Fund and Indiana Genomics Initiative (INGEN).

References