British Journal of Medicine & Medical Research
4(1): 1-33, 2014

SCIENCEDOMAIN *international*
www.sciencedomain.org

Proteomic Biomarker Discovery: Review of the Science and Its Challenges with Clinical Samples

T. Hussain[1*], L. Scaife[2], V. Hodgkinson[2], J. P. Kneeshaw[3], M. Lind[2,4], P. Drew[2] and L. Cawkwell[2]

[1]University of Hull-Hull, York Medical School and Hull and East Yorkshire Hospitals, NHS Trust, Castle Hill Hospital, Hull, UK.
[2]Postgraduate Medical Institute of the University of Hull, Hull York Medical School, Hull, UK.
[3]Breast Unit, Hull and East Yorkshire Hospitals NHS Trust, Hull, UK.
[4]Queens Centre for Oncology and Haematology, Hull and East Yorkshire Hospitals NHS Trust, Hull, UK.

Authors' contributions

This work was carried out in collaboration between all authors. Author TH managed the literature searches and wrote the draft of the manuscript. Authors LC and VH contributed to the proteomics aspect of the paper. Authors JPK, ML and PD contributed towards the idea on the clinical application of the proteomic biomarkers and author LC supervised the manuscript preparation. All authors read and approved the final manuscript.

Review Article

Received 12th June 2013
Accepted 19th August 2013
Published 14th September 2013

ABSTRACT

Proteomics is a science that facilitates global analysis of protein expression, molecular interactions and functional states of protein in a cell, organ or organism under consideration. Protein at the molecular level represents the biological endpoint and therefore more suitable to assess cellular responses. Proteins biomarkers are routinely used in cancers for different purposes such as diagnosis, screening and predicting therapy responses. Proteomic science utilise both cell line models and clinical samples to make novel discoveries. Compared to cell lines, clinical samples provide more accurate representation of the tumour micro-environment, explore and identify stromal molecular targets. However, using clinical samples for biomarker discovery has its own challenges. In this review, we aim to provide clinicians an overview of the proteomic

**Corresponding author: Email: tasadooq.hussain@gmail.com;*

biomarker discovery pathway, examine the use of different clinical samples in proteomic research and highlight the limitations and challenges of using clinical samples with proteomic methods.

Keywords: Genomics; transcriptomics; proteomics; biomarker; tumour samples.

ABBREVIATIONS

CSF: Cerebro-Spinal Fluid; 1D-PAGE: One dimensional polyacramide gel electrophoresis; 2D-PAGE: Two dimensional polyacramide gel electrophoresis; DIGE: Differential in gel electrophoresis; DEP: Differentially expressed protein; ELISA: Enzyme linked immuno sorbent assay; GIST: Global internal standard technology; MS: Mass spectrometry; FFPE: Formalin fixed paraffin embedded; iTRAQ: Isobaric tag for relative and absolute quantification; ICAT: Isotope coded affinity tagging; IPA: Ingenuity pathway analysis; ICPL: Isotope coded protein labelling; LCM: Laser capture microdissesction; MALDI-TOF/TOF: Matrix assisted laser desorption ionisation and time of flight mass spectrometry; ESI: Electron spray ionisation; MRM: Multi chain monitoring; PTM: Post translational modification; STR: Short tandem repeat; SELDI: Surface enhanced laser desorption ionisation; TIF: Tumour interstitial fluid; RPA: Reverse Phase Assay.

1. INTRODUCTION

In the modern era, cancer is genetically the most complex and challenging condition and a predominant cause of premature death. In the UK alone, 28% of all deaths were due to cancers in 2010 [1]. However, with early diagnosis and appropriate treatment strategies improvements in the patients' prognosis and reductions in cancer mortalities can be achieved. Clinically biomarkers are utilised in the fight against cancers for purposes of cancer screening [2,3], diagnosis [4-7], therapy monitoring [8-10] and developing targeted cancer treatments. A clinical biomarker is defined, as a substance that indicates a biological or pathological process or a response to therapeutic intervention. These responses can be objectively measured and evaluated and occur as a result of specific changes at molecular level starting from the genome and involving any point up to the stage of protein synthesis [11]. Therefore, a biomarker can be a mutated gene signature, altered mRNA or a differentially expressed protein which can be broadly studied under the *"omic"* headings of genomics, transcriptomics, and proteomics.

In the past decade, the discovery of novel biomarkers in cancer research have risen exponentially with advancements in the DNA sequencing, gene expression profiling, protein separation, enrichment and identification methods [12,13]. At the same time, discovery studies, have also progressed from early years of cell line work [14-17] to utilise biological samples such as plasma, urine, CSF and fresh tumour samples [18-21] for novel discoveries. However, despite the major advancements, the transition of biomarkers from the scientific community towards a diagnostic test that can be used in daily clinical practice has been far from ideal. The challenges of using clinical samples (closet biological replicates) with omic studies and issues of clinical validations have all contributed to the delay in the transfer of predictive biomarkers described in the literature from lab to the clinical bed [22,23].

2. AIMS AND METHODS

The aim of this review is to provide an overview of proteomic workflow for biomarker discovery to clinicians and to discuss the challenges of using different clinical samples with proteomics. A literature search was conducted on pubmed using the term “challenges of clinical samples with proteomics” and 22 articles published in human studies in last 10 years were selected and reviewed.

3. BIOMARKER DISCOVERY: PROTEOMICS vs OTHER OMIC STUDIES

‘Omic' technologies include genomics, transcriptomics, proteomics and metabolomics all of which have a significant potential in generating novel biomarkers of exposure, susceptibility and response. However, for this review, the discussion as to why proteomics should be preferred over other omic studies for the cancer biomarker discoveries will be largely limited to comparing proteomics to genomics and transcriptomics studies only.

The genomic study involves analysing a change in the genome structure as a marker of response which is usually a DNA sequence. Genome based research in the last decade has transformed with the availability of information on gene sequencing and development of high-density DNA microarrays technique. As a result, functional genomics and genotype-phenotype studies have emerged in the forefront of genomic research [24,25]. The ability to analyse thousands of gene expressions has opened new avenues for research in field of therapeutics and medicine [26]. However, genomics compared to the other two *‘omic’* studies, due to the static nature of the human genome and very few mutations altering its functional status, may have a limited role in cancer biomarker studies. Further, as genomic based research rely on classification methods (e.g. prediction strength, support vector machine and naive Bayes) [27,28] to analyse differential gene expressions [29]; discrepancies between a higher differential gene expression and its classification method limits the usefulness of the method for data analysis [13]. Also, genomic complexities generated by alternative splicing of mRNA transcripts affect the gene function by adding or deleting functional domains and changing affinities, make genome based study alone less specific [30]. Furthermore, a lack of comprehensive analytical techniques to evaluate all splice variations, individual patient differences in larger population studies and tissue complexities all together make genomics less popular [28].

The transcriptome by definition is a subset of genes transcribed in an organism that link the genome, proteome and the cellular phenotype. With the advent of the DNA microarrays platform in transcriptomics, measuring the mRNA expression levels is now possible [31]. In a normal cell state, besides the post-translational modifications, frequency of mRNA expressions determines the cell protein content. However, in cancer cells, as mRNA turnovers can be excessive, a mismatch between the peak mRNA expressions and its protein translations is a possibility. The differences can run up to as high as 20 folds thereby making protein expression analysis from a quantitative mRNA data in transcriptomics based studies less reliable [32]. Therefore, to conclude, in isolation, both genomic and transcriptomic studies from all of the above reasons are insufficient to study the complex pathways of carcinogenesis.

A global analysis of protein expression, interaction and functional status can be studied using proteomics. Proteins via their post-translation modifications acquire stability and functional variability [33] and therefore give an accurate reflection of cell functionality and

response. Currently, proteomic technology has been used in two main areas of cancer research: early diagnosis and treatment (including prediction of response to treatment and targeting novel cancer agents). The initial results from both *in vitro* and *in vivo* studies are impressive [34]. Conventional proteomics employ gel-based (1D-PAGE; 2D-PAGE) techniques with mass spectrometry (MS) for protein separation, identification and analysis of differential expressions. With the technological advancements, new methods for protein enrichment, subcellular organelle fractionation [35], analysis of biological complexes [36] and specific post-translational modifications (*e.g.* phosphorylated proteins) [37] are now available to perform a more in-depth analysis of the proteome.

However, in order to understand the pathogenesis and progression of cancer, the altered genetic and metabolic factors essential to the development, growth, and proliferation of the malignant cells has to be fully established. Recognition of this important relationship is essential and important as cellular intermediary metabolism provide the bioenergetic/synthetic and catabolic requirements of the cell. Studying molecular genetics and proteomics along with the regulation of cellular intermediary metabolism can therefore provide novel revelations that will further facilitate our understanding of malignancy. In the absence of such an alliance, possibilities exist that one will readily peruse the contemporary literature and find innumerable instances in which evidence of change in the expression levels from gene expression studies (e.g. RT-PCR) and protein abundance studies (e.g. Western blot analysis) are taken as the corresponding changes in the cellular enzyme activity and associated pathway. Conversely, the absence of changes in expression could also be perceived as the enzyme-associated activity and pathway may not be involved in altered metabolism in a tumour cell or a disease process. In order to address this critical issue, one must therefore, integrate genetic, proteomic, and metabolic relationships and study them together. If the contemporary focus on evaluating genetic/proteomic relationships in the absence of essential metabolic studies is undertaken it can and will result in misleading conclusions and unwarranted interpretations and extrapolations that are essential to the development, growth, and proliferation of the malignant cells [38].

4. THE PROTEOMIC BIOMARKER DISCOVERY PATHWAY

The proteomic based biomarker discovery pipeline consists of four stages: the discovery phase, the data mining phase, the confirmation phase and the validation phase (Fig. 1). In the '*discovery stage'* proteins are separated and identified using gel-based or gel-free mass spectrometry approaches. Also, protein identification and differential expression can be achieved in a mass spectrometry-free proteomic approach using microarrays (forward and reverse-phase). In the '*data mining'* stage, proteins identified at the discovery stage are analysed by molecular mapping using a software programme such as Ingenuity Pathway Analysis (Ingenuity Systems). Interesting differentially expressed proteins (DEP) from this stage are then taken forward to the '*confirmation'* stage using an independent technique such as immunoblotting to check for ambiguities. Finally, DEPs are assessed for their clinical relevance at the final stage of '*clinical validation'* using either a tissue based-immunohistochemical (IHC) or a non-tissue based Enzyme Linked Immno Sorbent Assay (ELISA) approach.

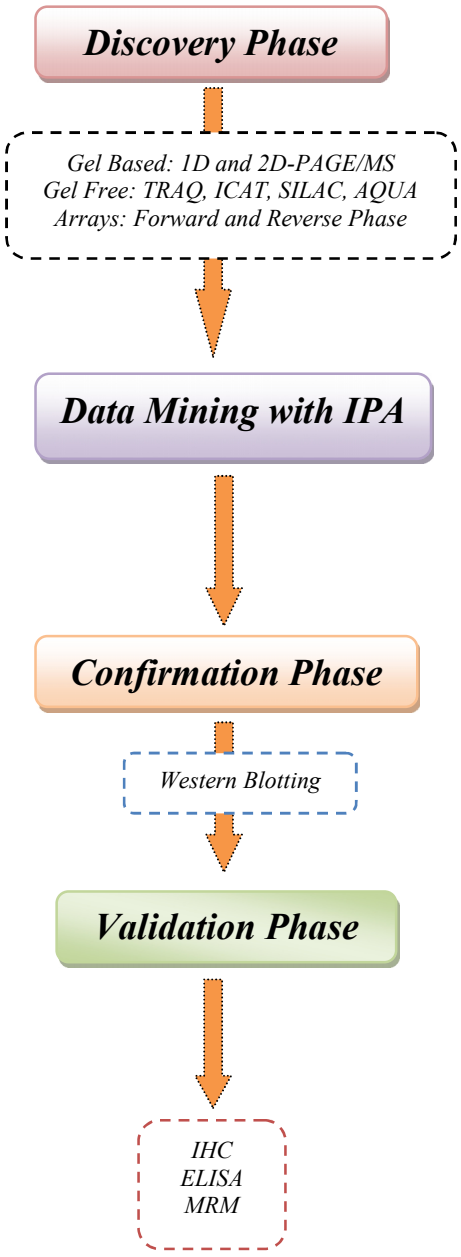

Fig. 1. The workflow of biomarker discovery pathway

1D/2D-PAGE/MS: *One and two dimensional polyacramide gel electrophoresis/Mass Spectrometry;* ***iTRAQ:*** *iso-baric tag for absolute and relative quantification;* ***SILAC:*** *Stable isotope labelling amino acid;* ***ICAT:*** *Isotope-coded affinity tagging;* ***AQUA:*** *Absolute quantification;* ***IPA:*** *Ingenuity Pathway Analysis;* ***IHC:*** *Immunohistochemistry;* ***ELISA****: Enzyme linked immuno sorbent assay;* ***MRM:*** *Multi chain reaction monitoring.*

4.1 Discovery Phase

This is the first stage in the proteomic biomarker workflow and involves protein separation using gel-based (1D-PAGE, 2D-PAGE and 2D-DIGE) or gel-free techniques (i-TRAQ, ICAT and SILAC) followed by protein identification using mass spectrometry (MS) [39-42]. However, protein identification and analysis of differential expressions can also be achieved using another approach which does not involve use of mass spectrometry called microarray. This method is usually preferred if proteins fail to generate a sufficient number of peptides with mass spectrometry assisted techniques to gain a significant identification, or if the specific form of protein is not represented in the database. Microarray-based proteomic methods can be employed in a forward (antibody immobilised) or reverse (protein lysate immobilised) phase [42] and offers a range of methods that compliment traditional mass spectrometry-based proteomic methods. However, none of the approaches when used on their own can fully analyse the whole proteome in a single experiment, therefore using them in combination increases the chances of wider proteome coverage. Therefore, a comparative proteomic approach is usually used to study differential protein expressions between different intervention groups (chemosensitive/resistant; radiosensitive/resistant).

4.1.1 MS-assisted gel based methods

This is a low-throughput method in which using principles of electrophoresis, proteins are separated in polyacramide gels based on molecular weight and iso-electric point following extraction from the sample in a suitable buffer (e.g. Laemmli buffer) prior to gel loading. The one-dimensional (1D-PAGE) method separates proteins based on molecular weight in vertical plane; separated proteins are visualized as bands and their molecular weight determined through use of a molecular marker. In the two-dimensional method, proteins separation takes place in horizontal and vertical planes based on the protein's iso-electric points and molecular weights. Isoelectric focusing (IEF) is used to separate proteins horizontally in the first dimension, according to their isoelectric point; the point at which the pH of the protein has no net electrical charge. First dimensional separation utilizes immobilised pH gradient (IPG) strips which function to provide a stable pH gradient in which to focus proteins of interest. Following horizontal separation by IEF, proteins are further separated vertically by their corresponding molecular weights, using sodium dodecyl sulphate (SDS) polyacramide gel electrophoresis (PAGE). IPG strips are firstly equilibrated whereby SDS is used to denature proteins and give them a net negative charge, ensuring that they travel towards the anode during electrophoresis [43]. Following equilibration, IPG strips are placed at the top of each gel and embedded, allowing proteins to migrate through the gel and achieve separation according to their individual molecular weight [44]. Proteins are then visualised through staining, typically using Coomassie blue staining due to its compatibility with downstream mass spectrometry. Differentially expressed proteins (DEPs) between treatment sensitive and resistant samples are then identified and analysed using commercially available software packages such as PDQuest [45]. The identified DEPs are then excised from the gel and partially digested into peptides and analysed using mass spectrometry.

4.1.1.1 Mass Spectrometry- MALDI and ESI

The overall aim of mass spectrometry based proteomics is to identify protein targets through the production and subsequent separation of ions according to their mass-to-charge ratio (*m/z*). Every mass spectrometer consists of an ion source, in order to produce ions from the complex protein sample; at least one mass analyser to separate ions according to their

individual *m/z* ratio; a detector, to register the number of emerging ions from the protein sample; and a computer, to process produce mass spectra for interpretation. In order for peptides to be separated, they must first be converted into ions and subsequently transferred into the gas phase. This is carried out using an ionisation source, the most common of which include electrospray ionisation (ESI) and matrix-assisted laser desorption/ionisation (MALDI) [45]. MALDI is the most common ionisation source employed when analysing the differential expression of protein spots identified from 2D-PAGE. This method relies on the utilisation of an excess of matrix, such as α-cyano-4-hydroxycinnamic acid (CHCA) and 2,5-dihydroxybenzoic acid (DHB), to ionise the analyte by absorbing energy from the laser. As ions exit the ion source, they pass through the mass analyser, which functions to separate ions according to their *m/z* ratio. Several different mass analysers exist including the ion trap and quadrupole however, the time-of-flight (TOF) mass analyser is the most frequently used in combination with a MALDI ionisation source. The TOF mass analyser consists of a flight tube in high vacuum to ensure collisions do not occur before ions reach the detector. Ions of different mass are subsequently separated based on the time it takes to transverse the length of the flight tube and strike the detector. Once ions collide with the detector a peptide mass fingerprint (PMF) is produced. Probable protein identification is made from the tryptic PMF by comparing with PMFs of theoretical protein digests in a protein database search (e.g. MASCOT search engine). A theoretical protein match is produced with an ion score based on the percentage of sequences matched with the computed database. All scores are then analysed against Mowse algorithm (p value of <0.05) in the database and the theoretical pH and molecular weight of statistically significant protein matches are then matched with the location of DEP on the 2D-gel [40]. Whilst PMF analysis is currently the most popular method for protein identification, a number of drawbacks associated with its use may require the use of tandem mass spectrometry (e.g TOF/TOF) to determine the amino acid sequence of a peptide and subsequently provide a greater level of confidence in the target protein identification.

ESI couple with MS/MS for peptide sequencing is applied to analyse samples in liquid phase. It is coupled with high performance liquid chromatography (HPLC) using a steady stream of solvents to produce droplets containing ions against a capillary anode and mass spectrometer inlet cathode [47]. Doubly charged ions from tryptic digests generated by a fine spray allows determination of an ions *m/z* value; as the size of the droplet decreases electric charge density increases on the surface making ions leave the droplets following mutual repulsions between similar charge ions; released ions are then directed into the mass analyser to generate peptide sequence data [40].

4.1.1.2 MS-assisted gel free methods

The MS-assisted gel-free approaches such as i-TRAQ; ICAT; SILAC; AQUA and SRM have been introduced to counter some of the limitations of gel-based methods such as gel to gel variability, dye sensitivities, limitations at detecting low-abundant proteins, background variations with structural proteins and quantification difficulties. Protein identification from these methods is based on high-throughput *'shotgun'* analysis of peptides from a complex liquid mixture using high performance liquid chromatography (HPLC). This follows MS identification and subsequent protein quantification using a number of quantitative proteomic approaches using isotope labelling or label free methods [48].

4.1.2 MS-*free* approach: microarrays-based proteomics

Microarrays provide a high-throughput proteomic method that allows screening of multiple proteins in the forward-phase approach. However, microarrays with antigen immobilised in a reverse-phase can screen only one protein at a given time. In the forward-phase approach, protein lysate is screened using multiple immobilised antibodies spotted onto a nitrocellulose coated microscope slide. This method is good to compare protein expressions of two samples (e.g. therapy sensitive vs therapy resistance) as it allows comparing differential protein expressions between the samples as fold changes [49]. Following protein extraction and quantification, proteins from two samples are labeled with different fluorescent dyes. The treatment sensitive sample is labeled with Cy3 fluorescent dye and treatment-resistant sample with Cy5 dye. Labeled samples are then incubated with the nitrocellulose slide containing immobilized antibodies. There is a range of commercially available microarray kits available covering up to 700 antibodies per slide. Differential protein expressions are detected by studying the fluorescent signal intensity of the antigen-antibody binding complexes for each sample at the wavelength corresponding to dye label. Protein detection using the forward-phase microarray method can be used complimentarily with the MS based approaches to enhance the credibility of discovered biomarkers. The reverse-phase microarray approach employs commercially available antibodies to measure protein expression levels in a large number of biological samples simultaneously in a quantitative manner [50].

4.2 Data Mining Phase

Although proteomics has great potential in providing deeper understanding of the role of individual proteins and protein networks in disease and in unveiling the underlying disease mechanisms, challenges arise in transforming the large-scale experimental data into biomedical knowledge for clinical practice. Therefore, in order to understand how proteins relate to, or interact with each other in a biological context, enhanced interpretation of the generated data through the use of data mining tools is essential. The overall goal is to extract useful information that leads to the identification of protein biomarker candidates. There are several widely-used bioinformatics software systems available to perform bioinformatics analysis and interpret the analysis result. Examples of some of the commonly used software programmes include R/Bioconductor, GALAXY, DAVID, KEGG, Panther, Gene Set Enrichment Analysis, PPI spider, Reactome and Ingenuity Pathway Analysis (Ingenuity Systems) [51]. IPA contains a library of approximately 3,000,000 biological and chemical interactions and functional annotations in the Ingenuity knowledge base. Each molecular pathway in knowledge base is manually curated by expert biologists and has definite scientific evidence. By uploading the list of Gene IDs and expression values into the Core Analysis tool, the network-generation algorithm identifies focused genes integrated in a global molecular network which can then be selected for further investigation. IPA also calculates the score p-value that reflects the statistical significance of association between the genes and the networks by the Fisher's exact test. An IPA identified phospho-inositol-triphosphate/phosphokinase-B (PIP3/AKT) signaling molecular pathway is illustrated as an example in Fig. 2 to facilitate the understanding of how information on molecular targets (upstream and downstream) can be obtained by mapping the identified DEPs onto a specific molecular pathway.

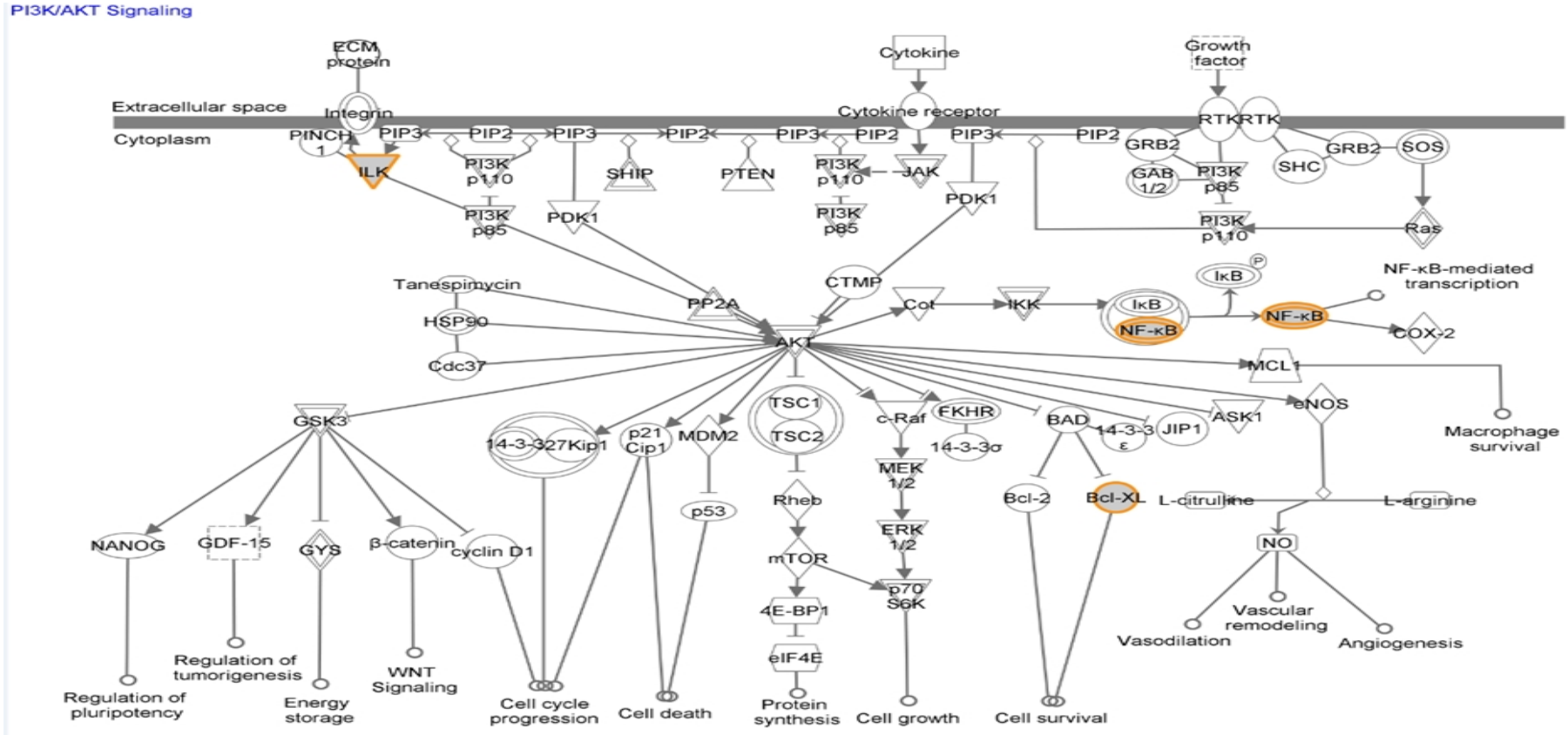

Fig. 2. IPA Identified PIP3/AKT Cell Signaling Molecular Pathway

The above figure is a general sketch of the PKB/AKT pathway identified from the Igenuity Pathway Analysis (Ingenuity Systems). PIP3 is activated by PI3K and recruits AKT to the cell membrane via activation of PDK1. AKT activation stimulates cell cycle progression, survival, metabolism and migration through phosphorylation of many physiological substrates. Using the pathway information on acquired chemotherapy resistance in breast cancers can be obtained by target activation of PIP3/AKT protein in the pathway and inhibition of cellular apoptosis and promotion of cell survival. The identified DEPs responsible for this are mapped on to the pathway seen in orange colours

4.3 Confirmatory and Validation Phases

The identified DEPs in the discovery phase are confirmed using an independent technique [52] in order to eliminate false discoveries. Immunoblotting is a commonly used confirmatory technique in translational research. It works on the principle of separating proteins in one dimension based on their molecular weight and transferring them to a nitrocellulose membrane or polyvinylidene fluoride (PVDF) to be immobilized. Proteins under investigation are probed using specific primary antibodies after blocking the non-specific background binding sites using a blocking solution (non-fat dried milk powder or bovine serum albumin). Following incubation with the primary antibody, one commonly used method for protein visualisation is the use of chemiluminescence, employing a horseradish peroxidase (HRP) conjugated secondary antibody. Via this method, the relative amount of protein can then be visualised by photographic film. The presence of an exposed band indicates the presence of that particular target protein within the sample. Band intensity is directly proportional to the amount of protein present. The photographic film can then undergo quantification using densitometry. During this process, target proteins are normalised against loading controls of anti-'housekeeping' antibodies (e.g. anti-alpha tubulin, ,anti- GAPDH and anti-beta actin) [39] which should demonstrate constant levels of expression within the protein sample. Through the use of loading controls, comparisons between band intensity produced by the primary antibody can be made, enabling a quantitation of fold change in expression to be determined. The success for western blotting heavily relies on the availability of a reliable antibody, specific to the protein of interest. Where suitable antibodies do not exist, additional analysis at the mRNA level using reverse transcriptase polymerase chain reaction (RT-PCR) or real time quantitative PCR (RTq-PCR) can be employed for confirmation of differential transcript expression. In the *in vitro* setting RNA interference (RNAi) using small interfering RNA (siRNA) molecules which result in gene silencing can be used to confirm protein expression [53].

Putative protein biomarkers that successfully pass through the validation phase of the biomarker discovery pipeline are assessed for their clinical relevance in the validation phase. Clinical significance of a particular protein biomarker is typically carried out using immunohistochemistry (IHC) using a series of retrospective archival tumour samples with the relevant clinical information. Such a method enables the validation of both expression and localisation of proteins within whole tissue sections which have been formalin-fixed, paraffin embedded (FFPE) and mounted onto glass microscope slides. Due to IHC being a relatively low throughput method, alternative validatory approaches also include tissue microarrays (TMA) which use tissue cores from hundreds of FFPE a tissue sample embedded into a new TMA allowing a single slide to be screened simultaneously for the expression of one particular protein using IHC. Alternatively, clinical validation can also be carried out using an ELISA based method, MS-based multireaction monitoring (MRM) or reverse phase arrays (RPA) [39,54].

The MS assisted peptide-Multi Reaction Monitoring (MRM) assay approach can be applied for measuring specific peptides in complex mixtures (e.g. tryptic digests of plasma) [55]. In this approach, selected cleaved peptide from protein is quantified against an isotope labelled standard internal control peptide. This method applied to quantify major plasma proteins can also be extended to enrich low end proteins thus providing platform for biomarker validations [56]. MRM technique provides a high-throughput quantitative platform and can be useful to detect inborn errors of metabolism, drug metabolites, pesticide analysis with no issues with protein folding and complex data processing software etc [56]. However, limitations using the technique come when dealing with short proteins (producing fewer peptides); genetic

variations of single amino acids and correct identifications of post-translational protein modifications [57].

5. SAMPLES IN BIOMARKER RESEARCH: CURRENT STATUS

Cancer biomarker studies in the past heavily depended on cell line models for novel discoveries. However, more recently, different biological samples such as plasma, serum, urine, CSF, tumour interstitial fluid, circulating tumour cells, and fresh tumour samples have all been used in biomarker discovery research. Using cell lines models and clinical samples have their own advantages and disadvantages; however, clinical samples are more preferable over cell lines because of their ability to reflect the true tumour environment and attached clinical and pathological data with them. However, despite this advantage, a more routine use of clinical samples in cancer biomarker studies remain largely restricted due to the issues of tissue heterogeneity and a lack of guidance on standardized methods for sample collection, transfer and storage [58-59]. In contrast, cell lines are utilised extensively in various translational research studies (breast, lung and rectal cancer studies) as they are easy available and can be expanded to large numbers by cell culture.

5.1 Cell Lines Proteomics

Cancer cell lines in biomarker research are a popular resource as they provide vital information on tumour biology and molecular characteristics such as cell-cell and cell-matrix interactions, gene expressions and cancer therapeutics [60-61]. Although cell lines as models of cancer cannot accurately mimic the tumour in its biological microenvironment, research using cell lines has been prominent in the proteomics literature since they are much easier to handle within controlled conditions [62]. The majority of published proteomics-based studies for the identification of biomarkers of chemotherapy resistance are based upon research using cell lines, used as *in vitro* models to simulate a clinical scenario. For this, an *in vitro* establishment of chemotherapy-resistant cell sub-lines using a choice of strategies is required in order to analyse the proteome and compare to that of the parental cell line [63]. Many DEPs have been identified from chemotherapy-resistant breast cancer cell sub-lines, mainly derived from luminal-type (ER-positive) MCF-7 parent cells [62-63], using 2D-PAGE/MS as a global proteome screening technique. However, despite the utility and handling superiority, the role of cell lines in *in-vitro* studies remains restricted mostly due to the inability to replicate the classical tumour morphology and tumour enzyme profile [62]. As a result, analysation of tumour microecology, study of tumour-stromal interactions and tumour-stromal receptors which are vital to our understanding of mechanisms of cancer dissemination, metastasis and targeted treatments becomes confined with cell lines. Specifically with proteomics (2D-PAGE and 2D-DIGE) cell lines provide a limited analytic ability for essential transmembrane proteins such as receptors, transporter proteins and channel proteins due to issues of membrane protein enrichment [64]. Further, with cell lines change of the molecular phenotype from repeated cultures is well known and can cause loss of tissue-specific functions (cell-cell interaction and secretion) to become different from *in-vivo* cells further limiting an accurate representation of the original tumour biology. A comparative study on proteomic phenotyping of cells lines by Cuiping et al. using Hepa 1-6 cell line has confirmed this finding [65]. Similar findings of variations in cell growth, hormone receptor content, clonogenicity and karotype was also demonstrated in MCF-7 breast cancer cell lines obtained between different laboratories by Osborne and colleagues [66]. An another limitation associated with the use of cell lines is the concept of '*false cell lines'* as highlighted from the demonstration of contamination of large number of cell lines with HeLa

cells [67]. These cells developed from glandular cancer of cervix in 1952 [68] has potential to invalidate research results by cross contamination. Therefore, cancer cell lines if not bought from reputable repositories after proper screening and validation (HLA typing; karyotyping; isoenzyme typing etc.) has potential to contaminate research finding [69].

Cell lines used in cancer research can be derived from primary tumour or tumour metastasis (e.g. aspirates or pleural effusion). Cell lines derived from advanced cancer stage (metastatic source) does not usually represent tumour diversity and progression [69]. Analysing breast cancer cell lines as an example reveal three different sources (pleural effusion; ascites; breast) for 8 commercially available cell lines (BT20, MDA-MB231; MDA-MB435; MDA-MB468; T47D; ZR75.1; SkBr3; MCF-7). These cell lines represent two pathological sub-types (invasive ductal carcinoma; adenocarcinoma) in late-progressive form (except BT20 which is derived from primary tumour). Therefore, using these cell lines in translational studies can be a source of bias. Further, as most of the anti-cancer therapies are designed against primary tumour, results from therapy research obtained using cell lines derived from metastatic cells may not have a stronger clinical relevance [70].

5.2 Clinical Samples

5.2.1 Bio-fluids: plasma and serum proteomics

Human plasma is one of the most commonly analysed bio-fluid for purposes of diagnostics due to its easy accessibility, cellular versatility, and highest protein content [23]. Putnam classified plasma proteins into seven categories based on their functional characteristics comprehensively representing the whole human genome [71]. These include protein secreted by solid organs, immunoglobulins, 'long distance'-receptor ligands, 'local'-receptor ligands, tissue leakage proteins, temporary passengers, foreign proteins and aberrant secretions. Of all these different classes, aberrant secretions include proteins released from tumors and other diseased tissues, presumably not as a result of a functional requirement of the organism and include cancer markers, which may be normal, non-plasma-accessible proteins expressed, secreted, or released into plasma by tumour. Some of the commonly utilised plasma based biomarkers in day to day clinical practice and their clinical significances are summarised in the Table 1.

Plasma proteins show variations in their level of abundance and molecular weight in plasma and can be classified as high and low abundance proteins having high and low molecular weights respectively. Changes to the plasma protein expressions usually reflect the state of originating tissues at the molecular level as tissues are constantly bathed in plasma. Therefore, studying a change in the expression level of a protein usually provide vital information on the cell functionality and its physiological state. Generally, the pathologically expressed biomarkers are the low-abundant, low-molecular proteins which protein biomarker studies aim to assess. However, the isolation of low-abundant low molecular weight proteins for analysis remains a challenge due to the masking effect of the high-abundant, high-molecular weight proteins (albumin, transferrin, and immunoglobulins) existing in a dynamic range of 8-10 magnitude in the plasma over the low abundant proteins. Therefore, in order to study plasma for clinically important biomarkers (low-abundant proteins), high-abundant plasma proteins should be depleted and the low-abundant proteins should be enriched [23,72,73].

Table 1. Biomarker proteins in common use

Biomarker	Use	References
Non-Cancer markers		
Serum/Plasma Based Markers		
Prion	Creutzfeldt–Jakob screening	[119]
Amyloid β protein	Alzheimer's screening	[119]
Tau protein	Lewy body dementia screening	[119]
Interleukin 18	Acute Kidney Injury	[3]
Kidney Injury Molecule-1	Post Tubular Necrosis	[3]
Rheumatoid factors and anti-citrullinated protein/peptide	Rheumatoid arthritis screening	[4]
C-Reactive Protein	Inflammatory and Infective disorders	[120]
Cancer markers		
Serum/Plasma Based Markers		
CA 15-3/CA 27.29	Post operative surveillance in patients with no evidence of disease	[121]
CEA	Post operative surveillance in patients with no evidence of colonic cancer	[122]
PSA	Post operative surveillance in patients with no evidence of prostate cancer	[123]
CA125	Diagnosis and post operative surveillance in patients with no evidence of ovarian cancer	[124]
CA19-9	Diagnosis and post operative surveillance in patients with no evidence of pancreatic cancer	[125]
Ki-67	Surveillance in patients with no evidence of breast and prostate cancers	[121]
α-Feto protein	Diagnosis of germ cell testicular cancer and hepatocellular cancer	[5]
Tissue Based Markers		
ER	Predicting response to hormone therapy and assessing prognosis in breast cancer	[9]
PR	Combined with ER for predicting response to hormone therapy	[121]
HER2	Determining prognosis, most useful with node positive patients	[121]
uPA	Determining prognosis in breast ancer, including the subgroup with axillary node-negative disease	[121]
PAI-1	Usually assayed in combination with uPA for determining prognosis in breast cancer including subgroup of node-ve patients	[121]

Plasma profiling for clinical biomarkers can be carried out using enzyme assays, genome based analysis and proteomic methods with or without fractionization [74]. Among these methods, the application of conventional gel-based proteomics techniques such as 2D-PAGE-MALDI/TOF-MS is easier and more sensitive for plasma protein separation and identification [75]. As the difference in magnitude of protein concentration between low-abundance proteins and the rest of the constituents is higher, fractionization methods such as recycling immunosubtraction coupled with sequential anion-exchange and size-exclusion chromatography can also be employed in combination with 2D-PAGE/MS to separate proteins into fractionized columns on 2D-gels [74,75]. Occasionally, the low-abundant plasma proteins may also be available as bounded proteins and difficult to separate due to strong protein-protein interactions. One of the early method described using non-denaturisation buffer solvent to deplete plasma and serum of primary large proteins and isolate low abundant proteins is the immuno-affinity method [76,77]. A more complex method described by Tirumalai and Radhakrishnan et al. called the micro-capillary reversed-phase high performance liquid chromatography coupled online-tandem mass spectrometry (μHPLC-MS/MS) with 20% acetonitrile solvent [75] also achieves a full characterization of low-molecular weight proteome, identifying proteins belonging to different classes, transcription factors, receptors and nuclear proteins released into the plasma during necrosis, apoptosis or haemolysis isolation. However, analysing plasma multiprotein complexes and microdomains without pre-fractionation of lysates into subcellular components continue to remain a challenge [72]. In a more recently described novel approach by Zhou et al. using specific antibodies to capture high molecular weight proteins (e.g. albumin; IgA; IgM; IgA; apolipoprotein, transferrin etc.) analysed small bounded proteins in '*sub-biofluid'* fractionization method [78]. A total of 209 unique proteins have been identified using this approach of which 12 are currently used as clinical biomarkers (Table 2) [74]. However, gel-based proteomic approaches due to issues of gel to gel reproducibility and detection sensitivities can be a hindrance for plasma analysis [79]. To overcome the problem, a pre-label florescent labelling system called 2D-DIGE which reduces spot pattern variability and the gel number in an experiment, allowing a simple and accurate spot matching can be used [80]. Further, one another factor that limits the use of plasma for biomarker discoveries is the difficulty of assessing post-translational modifications (PTM). Most PTMs occur *in vivo* and are transient in a small fraction of proteins (less than 1%); principally because they are present in substoichiometric amounts on protein molecules there isolation in a sufficiently large amount of modified proteins for biochemical studies is not easily attainable [81]. Further, a lack of specific antibodies to detect PTMs despite their common occurrence adds to the difficulties at analysing them [82]. In conclusion, although plasma is an ideal fluid for cancer diagnostics and biomarker research, it remains under utilised due to the challenges of protein separation (low-abundant protein isolation) and/or difficulties of protein-protein interactions.

5.2.2 Tumour interstitial fluid proteomics

The tumour interstitial fluid (TIF) is the fluid of a solid tumour, consisting of the liquid phase interposed between the newly formed vascular walls of the tumour and the plasma membrane of the neoplastic cells [83]. The TIF in addition to the set of blood soluble phase-borne proteins, holds a subset of aberrantly externalized components, mainly proteins, released by tumour cells and tumour microenvironment through various mechanisms and include classical secretory (exosomes and membrane protein shedding) and non-classical secretory proteins [84]. These proteins studied under the heading of *'secretome'* analysis include growth factors, extracellular matrix-degrading proteases, cell motility factors, immunoregulatory cytokines or other bioactive molecules. The main function these proteins

is to regulate cell-to-cell and cell-to-extracellular matrix interactions and drive the processes of differentiation, invasion, metastasis and angiogenesis of cancers. Secretome analysis for biomarkers is usually carried using genomic-based computational prediction and proteomic methods. However, genomics based methods due to inconsistent expression pattern between mRNA and proteins from post translation modifications and/or absence of peptide or cell retention signals from secreted proteins may have a limited role in full secretome characterization [84]. In comparison, secretome analysis with proteomics is largely confined to '*in-vitro'* models due to difficulties in obtaining a pure secretome. However, Celis et al. and Varnum et al. analysed TIF '*in vivo'* environment using small pieces of freshly dissected invasive breast cancer and nipple aspirate fluid (NAF) containing proteins directly secreted by the ductal and lobular epithelium in women with breast cancers [85-86]. Huang and colleagues also investigated '*in-vivo'* cancer secretome in a novel approach using capillary ultrafiltration probe implantated into tumour masses of a live mouse in progressive and regressive stages [86]. Using MS proteomics, five secreted proteins (cyclophilin-A, S100A4, profilin-1, thymosin beta 4 and 10) were identified and associated with tumour progression and five (fetuin-A, alpha-1-antitrypsin 1–6, and contrapsin) with tumour regression.

Table 2. Biomarkers associated with high abundance proteins identified by sub-biofluid fractionisation method

Protein	Association	Marker Specificity	Reference
Glycosylasparaginase	Albumin	I-cell disease	[74]
Paraneoplastic antigen MA1	IgA	Paraneoplastic neurological syndrome	[74]
Meningioma-expressed antigen 6/11	Albumin	Meningioma	[74]
Dihydropteridine reductase	Apolipoprotein	Tetrahydrobiopterin deficiency	[74]
Coagulation factor 7 precursor	Albumin	Liver Cirrhosis	[74]
ASM-like phosphodiesterase 3a	Apolipoprotein	Bladder tumours	[74]
Prostate transgluminase	IgA, IgM	Prostate cancer	[74]
Pregnancy-plasma protein-A	IgG, IgM	Down syndrome	[74]
Hsc70-interacting protein	Albumin	Cell growth	[74]
Ryanodine receptor 2	Albumin	Diabetes	[74]
Bone morphogenic protein 3b	IgA	Osteophytosis	[74]
PSA	Albumin, IgG	Prostate cancer	[74]

5.2.3 Tumour samples proteomics

Clinical tissue samples (biopsy or resected tumour specimens) provide a more accurate representation of tumour micro-environment and come with the patient clinical and pathological data attached. Translational studies have used different clinical samples with various '*omic*' platforms to study therapy responses [10], classify tumours [88] and investigate tumour biology [89]. Biomarker studies of the early years used formalin fixed paraffin embedded (FFPE) tissues with gene expression profiling for making novel discoveries [89]. Since then, using smaller amount of protein with high resolution MS, global analysis of clinical samples with proteomic methods has been achieved [90,91]. More

recently proteomic studies have utilised tumour samples with MS to monitor and assess responses to cytotoxic therapies [92,93]. Our research group have recently published a pilot study using post-treatment resected fresh breast tumour samples for the first time with proteomics (antibody microarray and 2D-PAGE-MALDI/TOF/TOF) to discover biomarkers of anthracycline-taxane breast chemotherapy resistance [10]. However, clinical tissues by nature are complex and heterogeneous and therefore utilising them for biomarker discovery studies involves overcoming the obstacles of tissue heterogeneity. Also, other issues such as tissue licensing, storage, harvesting, collection, transport, patient consent and ethics etc would need addressing before a more routine use of clinical tissues for biomarker research is advocated. Clinical tissues commonly employed for biomarker research include both fresh and frozen-formalin fixed samples. Analysing formalin fixed paraffin-embedded specimens over fresh samples has an inherent advantage as large libraries of such specimens with long-term follow-up data are widely available. However, there are limitations for using FFPE tissues and include issues of formalin linked nuclear acids and cellular protein cross-linkages [94] tissue fragmentation and chemical modifications of the RNA [95].

One of the major factors that preclude data interpretation from biomarker studies utilising clinical tissues include tissue heterogeneity and cellular diversity [39]. In order to recapitulate the *in-vivo* molecular interactions that drive disease at the micro-environment level, analysation of sub-populations of cells from the heterogeneous microecology is required. However, a selective seclusion and utilisation of tumour cells from a heterogeneous tumour environment can be a challenge. Laser Capture Micro-Dissection (LCM) is one technique that isolates histologically pure cancer cells using laser-assisted micro-dissection from complex heterogeneous tissues and micro-environments [96]. Two types of LCM based technologies are known to date: the thermoplastic film contact based Arcturus system and the non-contact laser pressure catapulting PALM system. Since its advent, genomic and proteomic biomarker studies have witnessed an important step forward. Using comparative proteomics (2D-PAGE/MS and LC-MS/MS) with LCM on isolated breast cancer cells, proteins of breast cancer metastasis and prognosis have been identified [96]. However, as protein yields from the cancer cells using LCM technique are usually low limiting its use; the problem has been partly addressed by using nono-LC-FTICR/MS in combination accurate mass and time (AMT) tag database searching [96].

Other important issue related to the use of clinical tissues in translational research concerns to patient consent and ethics approvals. Clinical practice and medical research is bounded by an ethical code to safeguards patient rights and research misuse (e.g. market penetrations by companies). The UK government following '*The Kennedy and Redfern'* enquires [98] (The Report of The Royal Liverpool Children's Inquiry; 2001) legislated 'Human Tissue Act 2004'. This act strictly controls the whole body donations and translational research using human tissues. Local and institutional bodies (e.g. ethics committees) are responsible for the implementation of the regulations and overlooking the sample use in accordance with the approved purpose. However, the issues relating to the NHS ethical approval and governance can be varied and marred by challenges of operational variability, lack of integration between R & D departments, local management issues and multi-level bureaucracy. All these factors in isolation and/or in combination threaten to make translational studies using human tissues a complex and unattractive option [99]. A brief overview of the benefits and limitations using tissue samples with proteomic-based biomarker research are highlighted in Table 3. However, some of the above challenges can be met and overcome with the concept of '*tissue banking.*' Using tissue banks, investigators can use clinical samples in a more accessible manner by passing the need for local ethical approvals and patient consent. Bio-banks working under strict regulations use standard

operating procedures to collect and store high quality clinical samples (live and pathological archival) that can be used for translational research [100, 101]. However, in the absence of universal validity of informed consent covering all patient groups (e.g. pediatric patients), limitations can still exist with regards to the use of surplus clinical tissue and the necessity to access patients clinical data (critical for proteomic research) when using samples from bio-banks [102]. Some of these issues addressed by strengthening system of information delivery, educational awareness, patient specific consent forms, training in methods to obtain consent and setting up statutory monitoring bodies [103].

Table 3. Benefits & challenges of using clinical samples with proteomics

Clinical Tissue	Benefits	Limitations
Fresh tissue; biopsy samples	Pre-treatment sample High tumour percentage Good sample quality May be helpful for predictive biomarker discovery	Requires ethic approval Requires patient consent Optimization of techniques of protein extraction and quantification Sample size variations
Formalin fixed paraffin embedded tissues (FFPE)	Multiple FFPE cores can be assembled in a single block and used in tissue microarrays- MS analysis to collate large information for further proteomic analysis Helpful for predictive biomarker discoveries	Issues of protein solubilisation Optimization of techniques of protein extraction Protein modifications from formaldehyde fixation limiting antigen detection Ethical and consent issues
Fresh tissue; resected samples	Helpful in assessing therapy responses Provides markers of therapy monitoring Provides prognostic markers of disease	Tissue heterogeneity Variable sample quality and amount Optimization of techniques of protein extraction and identification Background variations with structural proteins

In Table 4 a list of 69 unique protein biomarkers are summarised with their proposed end-clinical applications. These biomarkers have been discovered in MS dependant and MS-free proteomic approaches using different tumour samples and cancer cell lines. To date, even though many cancer biomarkers have been discovered with proteomics, the translation of discovered biomarkers to the clinical bed side has been rather slow and tedious. The main reason to this includes a lack of evidence on the clinical effectiveness of many discovered markers in large populations from randomised controlled trials.

Table 4. Cancer biomarkers discovered using cell lines and clinical samples with proteomic methods

Biomarkers identified using breast clinical samples	Clinical Sample used	Proteomic Method	Proposed Clinical Application	Clinically validated	References
tBID	Breast tumour samples, ER+ ductal type	Antibody Microarray	Putative biomarkers of breast chemoresistance	Yes	[127]
BcL-XL	Breast tumour samples, ER+ ductal type	Antibody Microarray	Putative biomarkers of breast chemoresistance	No	[127]
14 3 3 theta/tau	Breast tumour samples, ER+ ductal type	2D-PAGE/MS	Putative biomarkers of breast chemoresistance	Yes	[127, 128]
Ubiquitin carboxyl-terminal hydrolase 5 (UCTH 5)	Breast TIF	LC–MS/MS	Putative biomarkers of breast cancer detection	Yes	[83]
Protein disulfide-isomerase (PDI)	Breast TIF	LC–MS/MS	Putative biomarkers of breast cancer detection	Yes	[83]
Platelet-derived endothelial cell growth factor (PD-ECGF)	Breast TIF	LC–MS/MS	Putative biomarkers of breast cancer detection	Yes	[83]
Peroxiredoxin-2 (PRDX2)	Breast TIF	LC–MS/MS	Putative biomarkers of breast cancer detection	Yes	[83]
Galectin-1	Breast TIF	LC–MS/MS	Putative biomarkers of breast cancer detection	Yes	[83]
Elongation factor 1-beta (EF-1-beta)	Breast TIF	LC–MS/MS	Putative biomarkers of breast cancer detection	Yes	[83]
Chloride intracellular channel protein 1 (CLIC1)	Breast TIF	LC–MS/MS	Putative biomarkers of breast cancer detection	Yes	[83]
Cellular retinoic acid-binding protein 2 (CRABP-II)	Breast TIF	LC–MS/MS	Putative biomarkers of breast cancer detection	Yes	[83]
Calreticulin	Breast TIF	LC–MS/MS	Putative biomarkers of breast cancer detection	Yes	[83]
Osteopontin	Breast nipple aspirate fluid	LC–MS/MS	Putative biomarkers of breast cancer detection	No	[86]
Cathepsin D	Breast nipple aspirate fluid	LC–MS/MS	Putative biomarkers of breast cancer detection	No	[86]

Biomarkers identified using colorectal cancer tissue samples					
Translationally controlled tumour protein	Colorectal cancer tissue samples	2D-PAGE/HPLC/MS	Putative biomarkers of colorectal cancer detection	Yes	[129]
S100A9	Colorectal cancer tissue samples	2D-PAGE/HPLC/MS	Putative biomarkers of colorectal cancer detection	Yes	[129]
S100A11	Colorectal cancer FFPE samples	2D-PAGE/MALDI	Putative biomarkers of colorectal cancer metastasis	Yes	[130]
Heat shock protein 60	Colorectal cancer tissue samples	2D-PAGE/HPLC/MS	Putative biomarkers of colorectal cancer detection	Yes	[129]
GSTM3	Colorectal cancer FFPE samples	2D-PAGE/MALDI	Putative biomarkers of colorectal cancer metastasis	Yes	[130]
FXYD3	Colorectal cancer FFPE samples	2D-PAGE/MALDI	Putative biomarkers of colorectal cancer metastasis	Yes	[130]
Aldehyde dehydrogenase 1	Colorectal cancer tissue samples	2D-PAGE/HPLC/MS	Prognostic biomarker of colorectal cancer	Yes	[129]
14-3-3β	Colorectal cancer tissue samples	2D-PAGE/HPLC/MS	Prognostic biomarker of colorectal cancer	Yes	[129]
Biomarkers identified using gastric cancer tissue samples					
Selenium-binding protein 1	Gastric cancer tissue samples	2D-PAGE/MALDI/TOF/MS	Putative biomarkers of cancer prognosis and outcome	Yes	[131]
ENO1, GRP78, GRP94, PPIA, PRDX1, and PTEN	Gastric cancer tissue samples	2D-PAGE/MALDI/TOF/MS	Putative biomarkers of gastric cancer detection	Yes	[132]
MAWD-binding protein	Gastric cancer tissue samples	2D-PAGE/MALDI/TOF/MS	Putative biomarkers of gastric cancer detection	Yes	[133]
Cathepsin B	Gastric cancer tissue samples	2D-PAGE/MALDI/TOF/MS	Putative biomarkers of cancer prognosis	Yes	[134]

CLIC1	Gastric cancer tissue samples	2D-PAGE/MALDI/TOF/MS	Putative biomarkers of cancer prognosis and outcome	Yes	[135]
AMP-18	Gastric cancer tissue samples	2D-PAGE/MALDI/TOF/MS	Putative biomarkers of gastric cancer detection	Yes	[136]
HSP27	Gastric cancer tissue samples	2D-PAGE/MALDI/TOF/MS	Putative biomarkers of cancer metastasis	Yes	[137]
Biomarkers identified for liver cancers using plasma samples					
A1AG1	Plasma	Lectin coupled-MRM/MS	Putative biomarkers of liver cancer detection	Yes	[138]
AACT	Plasma	Lectin coupled-MRM/MS	Putative biomarkers of liver cancer detection	Yes	[138]
A1AT	Plasma	Lectin coupled-MRM/MS	Putative biomarkers of liver cancer detection	Yes	[138]
CERU	Plasma	Lectin coupled-MRM/MS	Putative biomarkers of liver cancer detection	Yes	[138]
Biomarkers identified using ovarian tumour samples					
Rho G-protein dissociation inhibitor (RhoGDI)	Microdissected ovarian tumour samples (invasive, non-invasive and low malignant potential ovarian tumours)	LCM/2D-PAGE/MS	Putative biomarkers of ovarian cancer detection		
Glyoxalase I	Microdissected ovarian tumour samples (invasive, non-invasive and low malignant potential ovarian tumours)	LCM/2D-PAGE/MS	Putative biomarkers of ovarian cancer detection		

Biomarkers identified using pancreatic FFPE tissue samples					
STML2	FFPE tissue sections	LC-tandem MS (MS/MS).	Prognostic biomarker of pancreatic ductal adenocarcinoma	Yes	[140]
OLFM4	FFPE tissue sections	LC-tandem MS (MS/MS).	Prognostic biomarker of pancreatic ductal adenocarcinoma	Yes	[140]
ECH1	FFPE tissue sections	LC-tandem MS (MS/MS).	Prognostic biomarker of pancreatic ductal adenocarcinoma	Yes	[140]
Biomarkers identified using renal tumour samples					
Vimentin	Renal cell carcinoma tissue samples	2D-PAGE/ MALDI-TOF/ MS	Putative biomarkers of renal cancer detection	No	[141]
Thioredoxin-dependent peroxide reductase	Renal cell carcinoma tissue samples	2D-PAGE/ MALDI-TOF/ MS	Putative biomarkers of renal cancer detection	No	[141]
Retinal dehydrogenase 1	Renal cell carcinoma tissue samples	2D-PAGE/ MALDI-TOF/ MS	Putative biomarkers of renal cancer detection	No	[141]
Peroxiredoxin-6	Renal cell carcinoma tissue samples	2D-PAGE/ MALDI-TOF/ MS	Putative biomarkers of renal cancer detection	No	[141]
Hypoxia inducible domain family member 1A	Renal cell carcinoma tissue samples	2D-PAGE/ MALDI-TOF/ MS	Putative biomarkers of renal cancer detection	No	[141]
Glutathione S-transferase P	Renal cell carcinoma tissue samples	2D-PAGE/ MALDI-TOF/ MS	Putative biomarkers of renal cancer detection	No	[141]
CD2-associated protein	Renal cell carcinoma tissue samples	2D-PAGE/ MALDI-TOF/ MS	Putative biomarkers of renal cancer detection	No	[141]
Annexin A5	Renal cell	2D-PAGE/ MALDI-TOF/	Putative biomarkers of	No	[141]

	carcinoma tissue samples	MS	renal cancer detection		
Biomarkers identified using MCF-7 breast cancer cell lines					
Stathmin	MCF-7 cell sub-lines	2D-PAGE/MS	Putative biomarkers of breast chemoresistance	Yes	[34]
Sorcin	MCF-7 cell sub-lines	2D-PAGE/MS	Putative biomarkers of breast chemoresistance	Yes	[34]
Peroxiredoxin-6	MCF-7 cell sub-lines	2D-PAGE/MS	Putative biomarkers of breast chemoresistance	Yes	[34]
HSP 27	MCF-7 cell sub-lines	2D-PAGE/MS	Putative biomarkers of breast chemoresistance	Yes	[34]
Cytokeratin 19	MCF-7 cell sub-lines	2D-PAGE/MS	Putative biomarkers of breast chemoresistance	Yes	[34]
Cathepsin D Chain B	MCF-7 cell sub-lines	2D-PAGE/MS	Putative biomarkers of breast chemoresistance	Yes	[34]
ATP synthase β	MCF-7 cell sub-lines	2D-PAGE/MS	Putative biomarkers of breast chemoresistance	Yes	[34]
Alpha tropomyosin	MCF-7 cell sub-lines	2D-PAGE/MS	Putative biomarkers of breast chemoresistance	Yes	[142]
14 3 3 epsilon	MCF-7 cell sub-lines	2D-PAGE/MS	Putative biomarkers of breast chemoresistance	Yes	[34]
14 3 3 sigma (Stratifin)	MCF-7 cell sub-lines	2D-PAGE/MS	Putative biomarkers of breast chemoresistance	Yes	[142]
Biomarkers identified using HCC cancer cell lines					
HSP60	HepG2 cell lines	2D-PAGE/MS	Putative biomarkers of hepatocellular carcinoma detection	Yes	[143]
HSP70	HepG2 cell lines	2D-PAGE/MS	Putative biomarkers of hepatocellular carcinoma detection	Yes	[143]
Biomarkers identified using leukemia cell lines					
Tromomyosin 3	Acute myeloid leukemia cells	2D-PAGE/ MALDI-TOF/ MS	Putative biomarkers of cancer prognosis and	Yes	[144]

			outcome		
Lipocortin 1	Acute myeloid leukemia cells	2D-PAGE/ MALDI-TOF/ MS	Putative biomarkers of cancer prognosis and outcome	No	[144]
Catalase	Acute myeloid leukemia cells	2D-PAGE/ MALDI-TOF/ MS	Putative biomarkers of cancer prognosis and outcome	No	[144]
Annexin A10	Acute myeloid leukemia cells	2D-PAGE/ MALDI-TOF/ MS	Putative biomarkers of cancer prognosis and outcome	No	[144]
Alpha-enolase	Acute myeloid leukemia cells	2D-PAGE/ MALDI-TOF/ MS	Putative biomarkers of cancer prognosis and outcome	No	[144]
Biomarkers identified using pancreatic cancer cell lines					
VIM	Clone #3/#8 cell sub-lines	2D-DIGE/ MALDI-TOF/ MS	Pancreatic cancer invasion-related biomarker proteins	Yes	[145]
STIP1	Clone #3/#8 cell sub-lines	2D-DIGE/ MALDI-TOF/ MS	Pancreatic cancer invasion-related proteins	Yes	[145]
KRT18	Clone #3/#8 cell sub-lines	2D-DIGE/ MALDI-TOF/ MS	Pancreatic cancer invasion-related biomarker proteins	Yes	[145]
GAPDH	Clone #3/#8 cell sub-lines	2D-DIGE/ MALDI-TOF/ MS	Pancreatic cancer invasion-related biomarker proteins	Yes	[145]
ALDH1A1	Clone #3/#8 cell sub-lines	2D-DIGE/ MALDI-TOF/ MS	Pancreatic cancer invasion-related biomarker proteins	Yes	[145]
Biomarkers identified using prostate cell lines & serum					
Annexin 1	DU145 cells	2D-PAGE/MS	Prognostic biomarker of prostate cancer	Yes	[146]
Prostate membrane-specific antigen	Serum	SELDI-TOF/MS	Diagnostic biomarker of prostate cancer	Yes	[147]

6. BIOMARKER VALIDATION

Exploratory biomarkers become valid biomarkers by the process of validation. A valid biomarker by definition should be measured in an analytic system with established performance status with wider agreement on the physiologic, pharmacologic, toxicologic or clinical significance in medical and scientific communities [103]. Validation maintains quality assurance through technology integration, assay development and regulatory pathways and establishes standardized guidelines on analytic methods used for biomarker measurement [104]. This process can be divided into three distinct categories: method validation, pharmacokinetic validation and laboratory validation covering range of issues pertaining to clinical use of biomarkers.

6.1 Method Validation and Related Issues

Qualification of biomarkers depends on characterization and validation of assays in the clinical context and biomarkers usually fail due to poor choice of assay and lack of clinical validation [104,105]. The regulatory guidance on assay validation was very primitive until 2003 when the definition of International Organisation for Standardisation a '*fit-for-purpose'* approach was described to [104,106]. This definition incorporates assay validations in three stage of: pre-validation, exploratory/advance validation and in-study validation. The pre-validation involves defining the purpose for biomarker use and selecting the assay; the exploratory phase formulates assay methodology and acquisition of assay reagents and the actual assay performance is carried out at the in-study phase. Using a range of technology spectrum (low-end to high- end) analytic platforms for various validation assays (e.g. immunoassays; reverse-transcription-PCR; multiplex ligand- binding assays) have been developed [130]. However, issues like quality control, assay sensitivities, standardization of assay conditions, lack of comprehensive database and reproducibility all preclude a straightforward clinical application of these assays.

Exploratory phase of biomarker research has undergone a major transformation with technological advances [57,64-69] which has superseded the assay developments by a long way to create a bottleneck between discoveries and validation [107]. Immunoassay formats (e.g. ELISA) which has remained the 'workhorse' for clinical validations in last four decades still continue to do so [108]. However, with improvements in the immunoassay field (e.g. multiplex assays; robotics; automated analyzer) and issues of working with large sample volumes and reproducibility addressed; one hopes to see in future a larger series of discovered biomarkers getting validated using these novel techniques.

6.2 Laboratory Validation and Related Issues

The performance status and the predictive ability of a discovered biomarker are characterized by the process of clinical validation in the same patient group or independent patient group (cross-validations) [95]. Clinical validations require information on clinical outcomes (e.g. therapy response vs therapy resistance); defined surrogate endpoint (e.g. radiological cancer size measurements); adequate patient sample size for meaningful conclusions [109]. Pathological diagnosis is arrived using biopsy samples, for this, samples are processed and stored as FFPE blocks in histological archives. Various cancer archives are globally available and are efficient and an economical source for the biomarker validation in translational studies [95,110]. However, FFPE samples can come with their own problems (discussed in section 5.2.3) and pose a hindrance affecting the study outcomes. Secondly,

defining a valid surrogate endpoint for clinical validation can an issue especially with non-cancerous conditions (e.g. systemic lupus erythematosis) which are more of a 'concept disease' with very little direct observation involved [109]. However, with cancerous diseases this could be less complicated as objectively measured surrogates (e.g. final pathological size; radiological sizes; gene classifiers) are in use to assess tumour responses following therapy completion or whilst on-treatment [111,112]. Using tissue samples (e.g. FFPE blocks) for clinical validations also require a prior ethical approval and patient consent for sample storage and data access; all this can be an additional burden and may come in the way of using tissue samples more in a more free approach. Further, in order to establish putative or predictive links between the biomarker and its observed clinical outcome, validation studies has to be of adequate sample size and study power (80%); in this way validity of the results and maximum efficiency of analysis can be ensured [113]. However, achieving adequate sample sizes can be a formidable task as not all developmental laboratories have resources to conduct large studies. An useful alternative to the above problem involves an establishment of independent databases on the lines of Early Detection Research Network of National Cancer Institute [114] which will help expand informatics framework and provide assistance to facilitate early validation of discovered cancer biomarkers.

Furthermore, limitations to laboratory validations can also occur from the technique employed for clinical validations. In general, laboratory validations are carried out using immuno-histo-chemistry, ELISA and multi-reaction monitoring (MRM) [56,115]. Of these, immuno-histo-chemical studies are most commonly applied method for clinical validations. However, limitations to the technique mostly from the interpretative pitfalls of tumour staining, myofibroblast staining, interdigitated benign cells and transmembrane localization can all preclude a more generalised usage of the technique with different tissue samples [142]. Both FFPE and frozen tissues can be used to detect protein antigens using the IHC technique. However, the caveat using tissue based techniques as opposed to the solution-based immunoassays come from the variability in the antigen preservation, unpredictability and poor detection rates of relatively more abundant native proteins [116]. Further, issues such as non-specific tissue-binding (especially with avidin), background staining variations (due to endogenous biotin) and signal amplifications can all affect the standards of the method further limiting its use in clinical validations. With the advent of newer polymer-based IHC methods like EPOS (Enhanced Polymer One Step), Envision using multiple enzyme and primary antibody conjugates against the dextran polymer background enhanced signal amplifications (e.g. Tyramide, Fluorescyl-Tyramide, Rolling Circle) the detection sensitivity of IHC can be improved [117,118].However, financial implications, logistic limitations (collecting and compiling biomarker data), selective availability, issues with of expansions of validated databases and faulty designs has to be addressed before a generalised use of the technology can be promoted.

7. CONCLUSION

Proteomics is a relatively new field in cancer research and currently developing. Despite the challenges of working with clinical samples, proteomic science has successfully explored different clinical samples and identified biomarkers for a range of clinical purposes. However, with the challenges of clinical validations, sample acquisition and processing still to be fully addressed, *in-vitro* cell line models may continue to dominate the proteomic cancer research for quite some time to come.

CONSENT

Not applicable.

ETHICAL APPROVAL

Not applicable.

FUNDING

Authors declare there was no funding received for the production of this manuscript.

COMPETING INTERESTS

Authors declare no conflict of interest.

REFERENCES

1. Statistics OFN. Mortality Statistics: Deaths registered in 2010, England and Wales; 2011.
2. Galvin JE. Dementia screening, biomarkers and protein misfolding: Implications for public health and diagnosis. Prion. 2011;5(1):16-21.
3. Wu I, Parikh CR. Screening for kidney diseases: older measures versus novel biomarkers. Clin J Am Soc Nephrol. 2008;3(6):1895-901.
4. Nicaise Roland P, et al. Antibodies to mutated citrullinated vimentin for diagnosing rheumatoid arthritis in anti-CCP-negative patients and for monitoring infliximab therapy. Arthritis Res Ther. 2008;10(6):R142.
5. Sadek A, et al. Alpha feto-protein and albumin in ascitic fluid in hepatocellular carcinoma patients. J Egypt Soc Parasitol, 1997;27(2):455-64.
6. Kurman RJ, et al. Cellular localization of alpha-fetoprotein and human chorionic gonadotropin in germ cell tumors of the testis using and indirect immunoperoxidase technique. Cancer. 1977;40(5):2136-51.
7. Ravishankaran P, Shah AM, Bhat R. Correlation of interleukin-6, serum lactate, and C-reactive protein to inflammation, complication, and outcome during the surgical course of patients with acute abdomen. J Interferon Cytokine Res. 2011;31(9):685-90.
8. Chorn N. Accurate identification of HER2-positive patients is essential for superior outcomes with trastuzumab therapy. Oncol Nurs Forum. 2006;33(2):265-72.
9. Allegra JC, Lippman ME. Estrogen receptor determination predicts response to tamoxifen therapy. Recent Results Cancer Res. 1980;71:16-9.
10. Hodgkinson VC, et al. Proteomic identification of predictive biomarkers of resistance to neoadjuvant chemotherapy in luminal breast cancer: a possible role for 14-3-3 theta/tau and tBID? J Proteomics. 2012;75(4):1276-83.
11. Atkinson AJ. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69(3):89-95.
12. Anderson NL. The roles of multiple proteomic platforms in a pipeline for new diagnostics. Mol Cell Proteomics. 2005;4(10):1441-4.
13. Xiong M, Fang X, Zhao J. Biomarker identification by feature wrappers. Genome Res. 2001;11(11):1878-87.
14. James CR, et al. BRCA1, a potential predictive biomarker in the treatment of breast cancer. Oncologist. 2007;12(2):142-50.

15. Basbug M, et al. Prognostic value of preoperative CEA and CA 19-9 levels in patients with colorectal cancer. Hepatogastroenterology. 2011;58(106):400-5.
16. Cetintas VB, et al. Cisplatin resistance induced by decreased apoptotic activity in non-small-cell lung cancer cell lines. Cell Biol Int. 2012;36(3):261-5.
17. Gazdar AF, et al. Lung cancer cell lines as tools for biomedical discovery and research. J Natl Cancer Inst. 2010;102(17):1310-21.
18. Pitteri SJ, Hanash SM. Proteomic approaches for cancer biomarker discovery in plasma. Expert Rev Proteomics. 2007;4(5):589-90.
19. Shao C, Wang Y, Gao Y. Applications of urinary proteomics in biomarker discovery. Sci China Life Sci. 2011;54(5):409-17.
20. Olsson B, et al. Biomarker-based dissection of neurodegenerative diseases. Prog Neurobiol. 2011;95(4):520-34.
21. Sun YF, et al. Circulating tumor cells: advances in detection methods, biological issues, and clinical relevance. J Cancer Res Clin Oncol. 2011;137(8):1151-73.
22. Hodgkinson VC, et al. Biomarkers of chemotherapy resistance in breast cancer identified by proteomics: current status. Cancer Lett. 2010;294(1):13-24.
23. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics. 2002;1(11):845-67.
24. Lipshutz RJ, et al. High density synthetic oligonucleotide arrays. Nat Genet. 1999;21(1 Suppl):20-4.
25. Shou J, et al. Optimization and validation of small quantity RNA profiling for identifying TNF responses in cultured human vascular endothelial cells. J Pharmacol Toxicol Methods. 2006;53(2):152-9.
26. Hedenfalk I, et al. Gene-expression profiles in hereditary breast cancer. N Engl J Med, 2001;344(8):539-48.
27. Golub TR, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science. 1999;286(5439):531-7.
28. Furey TS, et al. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics. 2000;16(10):906-14.
29. Moler EJ, Chow ML, Mian IS. Analysis of molecular profile data using generative and discriminative methods. Physiol Genomics. 2000;4(2):109-126.
30. Gardina PJ, et al. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. BMC Genomics. 2006;7:325.
31. Leder EH, Merila J, Primmer CR. A flexible whole-genome microarray for transcriptomics in three-spine stickleback (Gasterosteus aculeatus). BMC Genomics. 2009;10:426.
32. Gygi SP, et al. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol. 1999;19(3):1720-30.
33. Johann DJ, Jr., et al. Clinical proteomics and biomarker discovery. Ann N Y Acad Sci. 2004;1022:295-305.
34. Chuthapisith S, et al. Proteomic profiling of MCF-7 breast cancer cells with chemoresistance to different types of anti-cancer drugs. Int J Oncol. 2007;30(6):1545-51.
35. Arnott D, Kishiyama A, Mohtashemi I, Shillinglaw W. Proceedings of the 50th ASMS Conference onMass Spectrometry and Allied Topics. American Society ofMass Spectrometry, Orlando, FL; 2002.
36. Simpson RJ, Connolly LM, Eddes JS, Pereira JJ. Proteomic analysis of the human colon carcinoma cell line (LIM 1215): development of a membrane protein database. Electrophoresis. 2000;21:1707–1732.
37. Zhou M, Lucas DA, et al. An investigation into the human serum "interactome". Electrophoresis. 2004;25(9):1289-1298.

38. Costello L, Franklin R. Tumor cell metabolism: the marriage of molecular genetics and proteomics with cellular intermediary metabolism; proceed with caution! Mol Cancer. 2006;5:59
39. Hodgkinson VC, et al. Biomarkers of chemotherapy resistance in breast cancer identified by proteomics: current status. Cancer Lett. 2010;294(1):13-24.
40. Gygi SP, et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol. 1999;17(10):994-9.
41. Ross PL, et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics. 2004;3(12):1154-69.
42. Zong Y, et al. Forward-phase and reverse-phase protein microarray. Methods Mol Biol. 2007;381:363-74.
43. Rabilloud T, Valette C, Lawrence JJ. Sample application by in-gel rehydration improves the resolution of two-dimensional electrophoresis with immobilized pH gradients in the first dimension. Electrophoresis. 1994;15(12):1552-8.
44. Weiss W, Gorg A. High-resolution two-dimensional electrophoresis. Methods Mol Biol. 2009;564:13-32.
45. Zhu C, Rawe I, Joyce NC. Differential protein expression in human corneal endothelial cells cultured from young and older donors. Mol Vis. 2008;14:1805-14.
46. Yates JR. Mass spectrometry. From genomics to proteomics. Trends Genet. 2000;16(1):5-8.
47. Lin D, Tabb DL, Yates JR. Large-scale protein identification using mass spectrometry. Biochim Biophys Acta. 2003;1646(1-2):1-10.
48. Chen EI, Yates JR. Cancer proteomics by quantitative shotgun proteomics. Mol Oncol, 2007;1(2):144-59.
49. Smith L, et al. The analysis of doxorubicin resistance in human breast cancer cells using antibody microarrays. Mol Cancer Ther. 2006;5(8):2115-20.
50. Spurrier B, Ramalingam S, Nishizuka S. Reverse-phase protein lysate microarrays for cell signaling analysis. Nat Protoc. 2008;3(11):1796-808.
51. Satoh J. Molecular network analysis of human microRNA targetome: from cancers to Alzheimer's disease. BioData Min. 2012;5(1):17.
52. Paulovich AG, et al. The interface between biomarker discovery and clinical validation: The tar pit of the protein biomarker pipeline. Proteomics Clin Appl. 2008;2(10-11):1386-1402.
53. Scaife L, et al. Differential proteomics in the search for biomarkers of radiotherapy resistance. Expert Rev Proteomics. 2011;8(4):535-52.
54. Kim K, Kim Y. Preparing multiple-reaction monitoring for quantitative clinical proteomics. Expert Rev Proteomics. 2009;6(3):225-9.
55. Kuhn E, et al. Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and 13C-labeled peptide standards. Proteomics. 2004;4(4):1175-86.
56. Anderson L, Hunter CL. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Mol Cell Proteomics. 2006;5(4):573-88.
57. Unwin RD, et al. Multiple reaction monitoring to identify sites of protein phosphorylation with high sensitivity. Mol Cell Proteomics. 2005;4(8):1134-44.
58. Ericsson C, Nister M. Blood plasma handling for protein analysis. Methods Mol Biol. 2011;675:333-41.
59. Dihazi H, Muller GA. Urinary proteomics: a tool to discover biomarkers of kidney diseases. Expert Rev Proteomics. 2007;4(1):39-50.
60. Strand DW, Hayward SW. Modeling stromal-epithelial interactions in disease progression. Discov Med. 2010;9(49):504-11.

61. Anderson AR, et al. Microenvironment driven invasion: a multiscale multimodel investigation. J Math Biol. 2009;58(4-5):579-624.
62. Lacroix M, Leclercq G. Relevance of breast cancer cell lines as models for breast tumours: an update. Breast Cancer Res Treat. 2004;83(3):249-89.
63. Watson MB, Lind MJ, Cawkwell L. Establishment of in-vitro models of chemotherapy resistance. Anticancer Drugs. 2007;18(7):749-54.
64. Zhang W, et al. Affinity enrichment of plasma membrane for proteomics analysis. Electrophoresis. 2003;24(16):2855-63.
65. Pan C, et al. Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions. Mol Cell Proteomics. 2009;8(3):443-50.
66. Osborne CK, Hobbs K, Trent JM. Biological differences among MCF-7 human breast cancer cell lines from different laboratories. Breast Cancer Res Treat. 1987;9(2):111-21.
67. Nelson-Rees WA, Daniels DW, Flandermeyer RR. Cross-contamination of cells in culture. Science. 1981;212(4493):446-52.
68. Lucey BP, Nelson-Rees WA, Hutchins GM. Henrietta Lacks, HeLa cells, and cell culture contamination. Arch Pathol Lab Med. 2009;133(9):1463-7.
69. Masters JR, et al. Short tandem repeat profiling provides an international reference standard for human cell lines. Proc Natl Acad Sci USA. 2001;98(14):8012-7.
70. Burdall SE, et al. Breast cancer cell lines: friend or foe? Breast Cancer Res. 2003;5(2):89-95.
71. Putnam. The Plasma Proteins Structure, Function, and Genetic Control. 2nd ed.; 2006.
72. Ahmed FE. Sample preparation and fractionation for proteome analysis and cancer biomarker discovery by mass spectrometry. J Sep Sci. 2009;32(5-6):771-98.
73. Luque-Garcia JL, Neubert TA. Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry. J Chromatogr A. 2007;1153(1-2):259-76.
74. Veenstra TD, et al. Biomarkers: mining the biofluid proteome. Mol Cell Proteomics. 2005;4(4):409-18.
75. Tirumalai RS, et al. Characterization of the low molecular weight human serum proteome. Mol Cell Proteomics. 2003;2(10):1096-103.
76. Adkins JN, et al. Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. Mol Cell Proteomics. 2002;1(12):947-55.
77. Sato AK, et al. Development of mammalian serum albumin affinity purification media by peptide phage display. Biotechnol Prog. 2002;18(2):182-92.
78. Zhou M, et al. An investigation into the human serum "interactome". Electrophoresis. 2004;25(9):1289-98.
79. Keller BO, et al. Interferences and contaminants encountered in modern mass spectrometry. Anal Chim Acta. 2008;627(1):71-81.
80. Bitarte N, et al. Moving forward in colorectal cancer research, what proteomics has to tell. World J Gastroenterol. 2007;13(44):5813-21.
81. Salzano AM, Crescenzi M. Mass spectrometry for protein identification and the study of post translational modifications. Ann Ist Super Sanita. 2005;41(4):443-50.
82. Gronborg M, et al. A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. Mol Cell Proteomics. 2002;1(7):517-27.

83. Gromov P, et al. Tumor interstitial fluid - A treasure trove of cancer biomarkers. Biochim Biophys Acta; 2013.
84. Xue H, Lu B, Lai M. The cancer secretome: a reservoir of biomarkers. J Transl Med. 2008;6:52.
85. Celis JE, et al. Proteomic characterization of the interstitial fluid perfusing the breast tumor microenvironment: a novel resource for biomarker and therapeutic target discovery. Mol Cell Proteomics 2004;3(4):327-44.
86. Varnum SM, et al. Proteomic characterization of nipple aspirate fluid: identification of potential biomarkers of breast cancer. Breast Cancer Res Treat. 2003;80(1):87-97.
87. Huang CM, et al. Mass spectrometric proteomics profiles of in vivo tumor secretomes: capillary ultrafiltration sampling of regressive tumor masses. Proteomics. 2006;6(22):6107-16.
88. Perou CM, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-52.
89. Zepeda-Castilla EJ, et al. [Molecular classification of breast cancer]. Cir Cir, 2008;76(1):87-93.
90. Ostasiewicz P, et al. Proteome, phosphoproteome, and N-glycoproteome are quantitatively preserved in formalin-fixed paraffin-embedded tissue and analyzable by high-resolution mass spectrometry. J Proteome Res. 2010;9(7):3688-700.
91. Wisniewski JR, Ostasiewicz P, Mann M. High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. J Proteome Res. 2011;10(7):3040-9.
92. Bauer JA, et al. Identification of markers of taxane sensitivity using proteomic and genomic analyses of breast tumors from patients receiving neoadjuvant paclitaxel and radiation. Clin Cancer Res. 2010;16(2):681-90.
93. Jensen LR, et al. Assessment of early docetaxel response in an experimental model of human breast cancer using DCE-MRI, ex vivo HR MAS, and in vivo 1H MRS. NMR Biomed. 2010;23(1):56-65.
94. Vincek V, et al. A tissue fixative that protects macromolecules (DNA, RNA, and protein) and histomorphology in clinical samples. Lab Invest. 2003;83(10):1427-35.
95. Sadi AM, et al. Clinical relevance of DNA microarray analyses using archival formalin-fixed paraffin-embedded breast cancer specimens. BMC Cancer. 2011;11:253:1-13.
96. Braakman RB, et al. Laser capture microdissection applications in breast cancer proteomics. Methods Mol Biol. 2011;755:143-54.
97. Tammen H, et al. Expression profiling of breast cancer cells by differential peptide display. Breast Cancer Res Treat. 2003;79(1):83-93.
98. The Report of The Royal Liverpool Children's Inquiry; 2001.
99. McDonach E, Barbour R, Williams B. Reflections on applying for NHS ethical approval and governance in a climate of rapid change: prioritizing process over principles? International Journal of Social Research Methodology and Theory. 2009;12(3):227-241.
100. Tschulik A, Zatloukal K. [The increasing importance of tumor and tissue banks in the light of genomic and proteomic research]. Pathologe. 2001;22(5):310-5.
101. Brisson AR, et al. Translational research in pediatrics: tissue sampling and biobanking. Pediatrics. 2012;129(1):153-62.
102. O'Neill O. Medical and scientific uses of human tissue. J Med Ethics. 1996;22(1):5-7.
103. Chau CH, et al. Validation of analytic methods for biomarkers used in drug development. Clin Cancer Res. 2008;14(19):5967-76.
104. Cummings J, et al. Fit-for-purpose biomarker method validation for application in clinical trials of anticancer drugs. Br J Cancer. 2010;103(9):1313-7.

105. Carden CP, et al. Can molecular biomarker-based patient selection in Phase I trials accelerate anticancer drug development? Drug Discov Today. 2010;15(3-4):88-97.
106. Lee JW, et al. Fit-for-purpose method development and validation for successful biomarker measurement. Pharm Res. 2006;23(2):312-28.
107. Diamandis EP. Cancer biomarkers: can we turn recent failures into success? J Natl Cancer Inst. 2010;102(19):1462-7.
108. Valentin MA, et al. Validation of immunoassay for protein biomarkers: bioanalytical study plan implementation to support pre-clinical and clinical studies. J Pharm Biomed Anal. 2011;55(5):869-77.
109. Liang MH, et al. Methodologic issues in the validation of putative biomarkers and surrogate endpoints in treatment evaluation for systemic lupus erythematosus. Endocr Metab Immune Disord Drug Targets. 2009;9(1):108-12.
110. Guo T, et al. Proteome analysis of microdissected formalin-fixed and paraffin-embedded tissue specimens. J Histochem Cytochem. 2007;55(7):763-72.
111. Lyou CY, et al. Computer-aided evaluation of breast MRI for the residual tumor extent and response monitoring in breast cancer patients receiving neoadjuvant chemotherapy. Korean J Radiol. 2011;12(1):34-43.
112. Huang CC, et al. Predicting relapse in favorable histology Wilms tumor using gene expression analysis: a report from the Renal Tumor Committee of the Children's Oncology Group. Clin Cancer Res. 200915(5):1770-8.
113. Leffondre K, Abrahamowicz M, Siemiatycki J. Evaluation of Cox's model and logistic regression for matched case-control data with time-dependent covariates: a simulation study. Stat Med. 2003;22(24):3781-94.
114. Srivastava S, Kramer BS. Early detection cancer research network. Lab Invest. 2000;80(8):1147-8.
115. Lerwill MF. Current practical applications of diagnostic immunohistochemistry in breast pathology. Am J Surg Pathol. 2004;28(8):1076-91.
116. Bobrow MN, et al. Catalyzed reporter deposition, a novel method of signal amplification. Application to immunoassays. J Immunol Methods. 1989;125(1-2):279-85.
117. Sabattini E, et al. The EnVision++ system: a new immunohistochemical method for diagnostics and research. Critical comparison with the APAAP, ChemMate, CSA, LABC, and SABC techniques. J Clin Pathol. 1998;51(7):506-11.
118. Gross AJ, Sizer IW. The oxidation of tyramine, tyrosine, and related compounds by peroxidase. J Biol Chem. 1959;234(6):1611-4.
119. Galvin JE. Dementia screening, biomarkers and protein misfolding: Implications for public health and diagnosis. Prion. 2011;5(1):16-21.
120. Ravishankaran P, Shah AM, Bhat R. Correlation of interleukin-6, serum lactate, and C-reactive protein to inflammation, complication, and outcome during the surgical course of patients with acute abdomen. J Interferon Cytokine Res. 2011;31(9):685-90.
121. Chung L, Baxter RC. Breast cancer biomarkers: proteomic discovery and translation to clinically relevant assays. Expert Rev Proteomics. 2012;9(6):599-614.
122. Lin JK, et al. Early postoperative CEA level is a better prognostic indicator than is preoperative CEA level in predicting prognosis of patients with curable colorectal cancer. Int J Colorectal Dis. 2011;26(9):1135-41.
123. Yamanaka Y, Takahashi S. Usefulness of PSA test for diagnosis of prostatic cancer. Nihon Rinsho. 2011;69(Suppl 5):233-7.
124. Tang A, et al. Prognostic value of elevated preoperative serum CA125 in ovarian tumors of low malignant potential: a multinational collaborative study (ANZGOG0801). Gynecol Oncol. 2012;126(1):36-40.

125. Mehta J, et al. Evaluating the efficacy of tumor markers CA 19-9 and CEA to predict operability and survival in pancreatic malignancies. Trop Gastroenterol. 2010;31(3):190-4.
126. Kashiwagi S, et al. Advantages of adjuvant chemotherapy for patients with triple-negative breast cancer at Stage II: usefulness of prognostic markers E-cadherin and Ki67. Breast Cancer Res. 2011;13(6):R122.
127. Hodgkinson VC, et al. Proteomic identification of predictive biomarkers of resistance to neoadjuvant chemotherapy in luminal breast cancer: a possible role for 14-3-3 theta/tau and tBID? J Proteomics. 2012;75(4):1276-83.
128. Hodgkinson VC, et al. Pilot and feasibility study: comparative proteomic analysis by 2-DE MALDI TOF/TOF MS reveals 14-3-3 proteins as putative biomarkers of response to neoadjuvant chemotherapy in ER-positive breast cancer. J Proteomics. 2012;75(9):2745-52.
129. O'Dwyer D, et al. The proteomics of colorectal cancer: identification of a protein signature associated with prognosis. PLoS One. 2011;6(11):e27718.
130. Meding S, et al. Tissue-based proteomics reveals FXYD3, S100A11 and GSTM3 as novel markers for regional lymph node metastasis in colon cancer. J Pathol; 2012.
131. Xia YJ, et al. Suppression of selenium-binding protein 1 in gastric cancer is associated with poor survival. Hum Pathol. 2011;42(11):1620-8.
132. Bai Z, et al. Proteomics-based identification of a group of apoptosis-related proteins and biomarkers in gastric cancer. Int J Oncol. 2011;38(2):375-83.
133. Zhang J, et al. Comparative analysis of the protein profiles from primary gastric tumors and their adjacent regions: MAWBP could be a new protein candidate involved in gastric cancer. J Proteome Res. 2007;6(11):4423-32.
134. Ebert MP, et al. Overexpression of cathepsin B in gastric cancer identified by proteome analysis. Proteomics. 2005;5(6):1693-704.
135. Chen CD, et al. Overexpression of CLIC1 in human gastric carcinoma and its clinicopathological significance. Proteomics. 2007;7(1):155-67.
136. He QY, et al. Diverse proteomic alterations in gastric adenocarcinoma. Proteomics. 2004;4(10):3276-87.
137. Chen J, et al. Proteome analysis of gastric cancer metastasis by two-dimensional gel electrophoresis and matrix assisted laser desorption/ionization-mass spectrometry for identification of metastasis-related proteins. J Proteome Res. 2004;3(5):1009-16.
138. Ahn YH, et al. A lectin-coupled, targeted proteomic mass spectrometry (MRM MS) platform for identification of multiple liver cancer biomarkers in human plasma. J Proteomics. 2012;75(17):5507-15.
139. Jones MB, et al. Proteomic analysis and identification of new biomarkers and therapeutic targets for invasive ovarian cancer. Proteomics. 2002;2(1):76-84.
140. Takadate T, et al. Novel prognostic protein markers of resectable pancreatic cancer identified by coupled shotgun and targeted proteomics using formalin-fixed paraffin-embedded tissues. Int J Cancer. 2013;132(6):1368-82.
141. Sun CY, et al. Proteomic analysis of clear cell renal cell carcinoma. Identification of potential tumor markers. Saudi Med J. 2010;31(5):525-32.
142. Liu Y, et al. Identification of 14-3-3sigma as a contributor to drug resistance in human breast cancer cells using functional proteomic analysis. Cancer Res. 2006;66(6):3248-55.
143. Looi KS, et al. Using proteomic approach to identify tumor-associated antigens as markers in hepatocellular carcinoma. J Proteome Res. 2008;7(9):4004-12.
144. Lopez-Pedrera C, et al. Proteomic analysis of acute myeloid leukemia: Identification of potential early biomarkers and therapeutic targets. Proteomics. 2006;6(Suppl 1):S293-9.

145. Walsh N, et al. Identification of pancreatic cancer invasion-related proteins by proteomic analysis. Proteome Sci. 2009;7:3.
146. Ornstein DK, Tyson DR. Proteomics for the identification of new prostate cancer biomarkers. Urol Oncol. 2006;24(3):231-6.
147. Ramirez ML, Nelson EC, Evans CP. Beyond prostate-specific antigen: alternate serum markers. Prostate Cancer Prostatic Dis. 2008;11(3):216-29.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=215&id=12&aid=1997