TIMELINE

The grand challenge to decipher the cancer proteome

Samir Hanash and Ayumu Taguchi

Abstract | The quest to decipher protein alterations in cancer has spanned well over half a century. The vast dynamic range of protein abundance coupled with a plethora of isoforms and disease heterogeneity have been formidable challenges. Progress in cancer proteomics has substantially paralleled technological developments. Advances in analytical techniques and the implementation of strategies to de-complex the proteome into manageable components have allowed proteins across a wide dynamic range to be explored. The massive amounts of data that can currently be collected through proteomics allow the near-complete definition of cancer subproteomes, which reveals the alterations in signalling and developmental pathways. This allows the discovery of predictive biomarkers and the annotation of the cancer genome based on proteomic findings. There remains a considerable need for infrastructure development and the organized collaborative efforts to efficiently mine the cancer proteome.

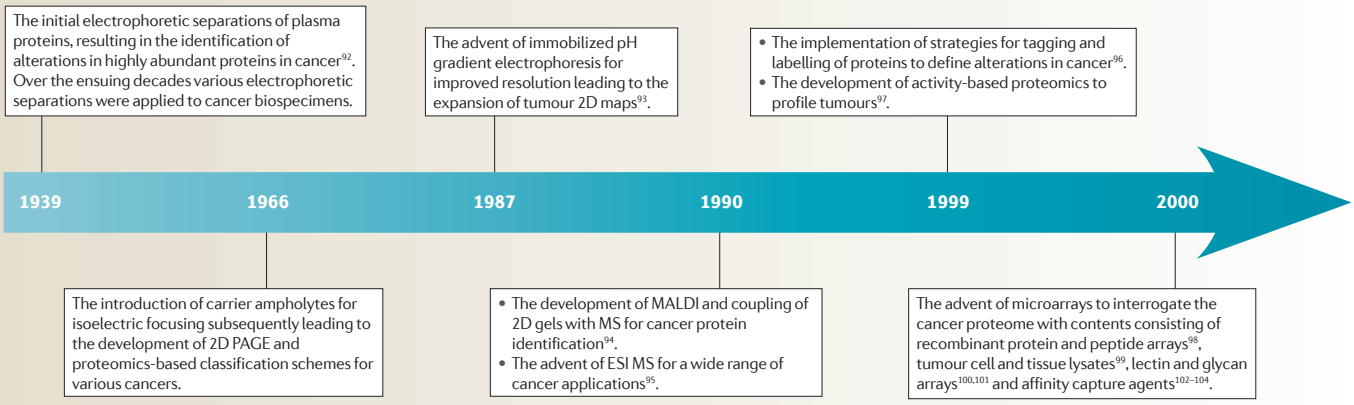
Illustrative of protein profiling to detect alterations in cancer half a century ago is a publication in *Nature* that defined patterns of globulin expression in serum among subjects with myeloma and macroglobulinemia[1] (TIMELINE). The study relied on what was then a new technique of starch gel electrophoresis, which revealed a few major bands in serum and plasma (FIG. 1a). Over the past several decades, the quest has been to profile an ever-increasing number of protein constituents in cells, tissues and biological fluids to determine their alterations in cancer. The introduction of polyacrylamide gels and isoelectric focusing using carrier ampholytes in the 1960s set the stage for the development of two-dimensional PAGE (2D PAGE) for separating proteins under denaturing conditions (FIG. 1b). This became a major tool for investigating complex proteomes and initially defined the field of proteomics.

The 2D PAGE era

The development of 2D PAGE, which allowed proteins to be resolved on the basis of their isoelectric point and molecular mass, represented not only a fundamental shift in the approach to protein analysis but also heralded the era of simultaneously investigating a large number of molecular constituents and implementing a systems approach to biological problems. Several hundred proteins could be analysed simultaneously, which led to this technique being used extensively. 2D PAGE studies have covered nearly every cancer type and a wide range of applications. For example, a 2D PAGE study published around a quarter of a century ago reported the analysis of leukaemia cells from various subtypes of acute lymphoblastic leukaemia (ALL)[2]. The study involved the analysis of 413 protein spots, resulting in a subset that could distinguish between the major subgroups of ALL. This subset included a new marker for common ALL and markers for cells of B and T cell lineages. Analysis of the 2D patterns also allowed the determination of T cell lineage in cases with an otherwise undifferentiated non-T cell and non-B cell phenotype.

Cancer proteomic programmes have been initiated around the use of 2D PAGE to characterize protein expression patterns for particular cancers. One such programme focused on extensive studies of bladder cancer in which large numbers of specimens, including tumour and non-tumour tissues, have been systematically analysed using 2D PAGE, resulting in the establishment of 2D protein expression databases and the identification of protein markers for bladder cancer[3]. Another ongoing effort in cancer proteomics using 2D PAGE has resulted in the establishment of a cancer proteome-expression database, The Genome Medicine Database of Japan (GeMDBJ; see the GeMDBJ website; Further information). The database contains proteomic data from surgically resected tissues and cultured cells of various malignancies, as well as the corresponding biological and clinicopathological data. This effort has resulted in the identification of several proteins that are associated with particular cancers as prognostic markers[4].

Timeline | **Progress in proteomics that has had an impact on cancer research**



2D, two dimensional; ESI, electrospray ionization; MALDI, matrix-assisted laser desorption ionization; MS, spectrometry.

Improvements have been made to the 2D PAGE approach, notably enhancing reproducibility and expanding the pH range for protein separations using immobilized pH gradients (IPG)[5]. Enhanced quantification of protein spots and the availability of techniques for in-gel comparative proteomic analysis have resulted from the introduction of fluorescence difference gel electrophoresis (DIGE)[6]. These developments notwithstanding, 2D PAGE has remained a challenging technique to master as it has a limited dynamic range, making it difficult to profile low-abundance proteins and to extract proteins from gels for their identification. The major advantages of 2D PAGE that have justified its continued use include the ability to resolve and visualize proteins into their diverse isoforms on the basis of differences in isoelectric point and/or molecular mass.

The advent of MS for protein analysis

Interest in the application of mass spectrometry (MS) for protein analysis has been longstanding, and progress has primarily depended on developments in instrumentation, which has been chronicled in numerous reviews[7–14]. An early application of MS is illustrated in the identification of proteins in 2D gels of extracts from leukaemia cells based on the analysis of tryptic digests by fast atom bombardment–collisionally-activated dissociation–tandem mass spectrometry (MS/MS). This was carried out, for example, for polypeptide OP18 (also known as stathmin 1), which is associated with acute lymphoid leukaemia[15].

The development of matrix-assisted laser desorption ionization (MALDI) MS provided a 'shot in the arm' for gel-based proteomics efforts, as protein spots or bands could be cut out from gels and their protein content digested and subjected to MALDI MS for identification[16]. As a result, 2D maps were developed for different cancer types. Illustrative of this effort is the construction of protein expression maps of lymphoid neoplasms[17]. A quantitative 2D database that includes 309 proteins corresponding to 389 protein spots across 42 lymphoid neoplasm cell lines has been constructed. Proteins separated by 2D PAGE were identified by MS and assigned expression levels on the basis of DIGE quantification. More than a twofold difference between various lineages was exhibited by 28 proteins, and decision tree classification identified proteins that could be used to classify the 42 cell lines according to their differentiation states.

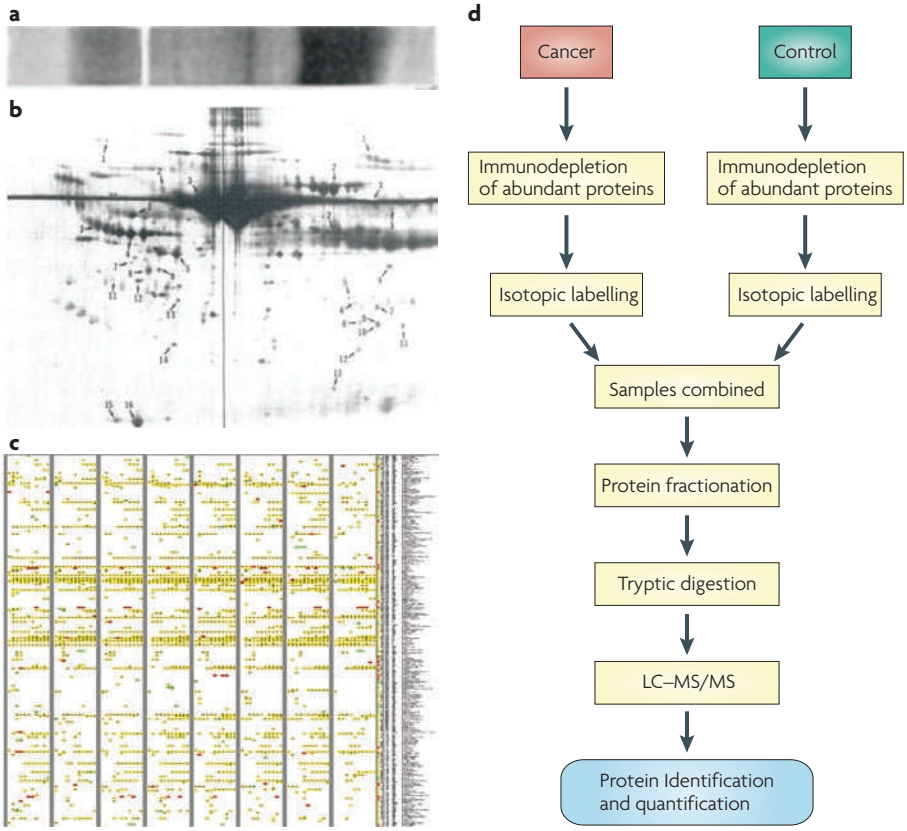


Figure 1 | **Increased resolving power of proteomics technologies as applied to serum and plasma. a** | Detection of a few major bands using one-dimensional (1D) electrophoresis as shown in a study of myeloma and macroglobulinemia[1]. **b** | Detection of a few hundred protein spots by 2D PAGE and the identification of some spots (indicated by numbers) by protein sequencing as shown in a study of plasma proteins among healthy subjects[88]. **c** | Identification of more than 1,500 proteins in plasma by liquid chromatography–tandem mass spectrometry (LC–MS/MS), based on a strategy for in-depth quantitative proteome profiling[89–90]. Columns represent individual fractions, rows represent individual proteins identified in particular fractions by LC–MS/MS. The colour scheme is indicative of case to control concentration ratios (red is increased, yellow is no change and green is decreased) for individual proteins based on differential isotope labelling of plasma from a lung cancer case and a control. **d** | Flow scheme of sample processing for in-depth quantitative proteome profiling.

The parallel development of electrospray ionization (ESI) MS for protein identification[18] coupled with various pre-fractionation and separation schemes and protein labelling has allowed the quantitative analysis of an ever-increasing number of proteins from cells, tissues and biological fluids (FIG 1c,d). The next generation of mass spectrometers currently available have significantly increased sensitivity and scan speed[19]. As a result, the identification of almost all the proteins translated from expressed genes in a cancer cell population has become achievable. Likewise, the exhaustive identification of proteins that are associated with particular cell compartments, such as proteins that are expressed on the cell surface, are secreted or are otherwise released into the extracellular space, is currently feasible. A study of the cell surface proteome of closely related metastatic and non-metastatic teratocarcinoma tumour cells on the basis of surface protein biotinylation followed by capture of biotinylated proteins using streptavidin-coated resin and the quantification of tryptic peptides by MS resulted in the identification of 998 proteins[20]. The study identified proteins that had been previously associated with metastatic spread as well as a large number of proteins that were not previously known to be expressed on the cell surface.

Concerns regarding the limited depth of the analysis and the reproducibility of some of the earlier MS technologies used for global profiling of tissues and biological fluids have led to collaborative studies to assess the limitations of MS for biological studies. A Human Proteome Organization-sponsored collaborative study examined the

sources of irreproducibility of MS-based proteomics[21]. A test sample, comprising 20 highly purified recombinant human proteins, was distributed to 27 laboratories. Although some of the laboratories reported misidentified proteins, centralized analysis of the raw data revealed that all the proteins had been detected by all the laboratories, which indicated that problems in data processing can be remedied. Another multi-laboratory study sponsored by the US National Cancer Institute demonstrated high reproducibility across laboratories and instrument platforms of MS-based multiple reaction monitoring (MRM) assays of proteins in low μg per ml concentrations[22]. Currently, in-depth profiling of plasma and other biological fluids allows proteins to be identified that span six or more logs of protein abundance (FIG. 2). At present, MS-based studies allow massive amounts of data to be produced and interrogated for a multitude of cancer applications.

The cancer proteome using microarrays

Microarray-based strategies have provided a high-throughput alternative to MS for interrogating particular aspects and components of the cancer proteome (TIMELINE). These include the determination of levels of proteins and their particular modifications in biological samples, as well as the determination of protein and peptide interactions with drugs, small ligands and other biomolecules, such as other proteins and autoantibodies to tumour antigens. Peptides and proteins can be synthesized in large numbers directly on the chip[23]. Alternatively, recombinant proteins can be arrayed, and efforts are underway to assemble large sets of purified recombinant proteins for microarrays and other applications[24]. However, other alternatives include arraying of tumour and tumour cell lysates and natural proteins derived from such lysates[25].

Integrating glycomics and proteomics

Post-translational modifications, notably glycosylation, are an important source of cancer biomarkers. Strategies for integrating proteomics and glycomics are under development[26]. The enrichment of glycoproteins can be accomplished by taking advantage of their affinity for lectins, which bind to specific glycans. Captured glycoproteins are subsequently analysed by MS. Improved glycoproteomic technologies and methodologies enable detailed glycan structure analysis and the sequencing of glycopeptide backbones.

Glycoproteins are particularly rich in potential diagnostic cancer markers[27]. A comparison of glycoproteins isolated from the serum of healthy subjects with those from patients with lung adenocarcinoma using multilectin affinity chromatography uncovered a large number of cancer-selective proteins, which included kallikrein N plasma 1 (KLKB1) and inter-α-trypsin inhibitor heavy chain 3 (ITIH3). A glycoproteomic approach based on Con A lectin-affinity chromatography, SDS–PAGE and MS analysis uncovered glycosylation changes that were associated with the differentiation of HT-29 colon cancer cells[28]. In another study, 30 target proteins of *N*-acetylglucosaminyl-transferase V (GNTV; also known as MGAT5) were identified by lectin blot analysis with L-phytohaemagglutinin (LPHA) and ESI–MS of colon cancer cells. Aberrant glycosylation of TIMP metallopeptidase inhibitor 1 (TIMP1), one of the GNTV target proteins, reduced the inhibition of matrix metallopeptidase 2 (MMP2) and MMP9, and correlated with cancer cell invasion and metastasis *in vivo* and *in vitro*[29]. Therefore, the elucidation of the contribution of altered glycosylation to cancer-related processes and the identification of aberrant glycan structures in proteins associated with cancer as a source of biomarkers are promising areas of research in oncoproteomics.

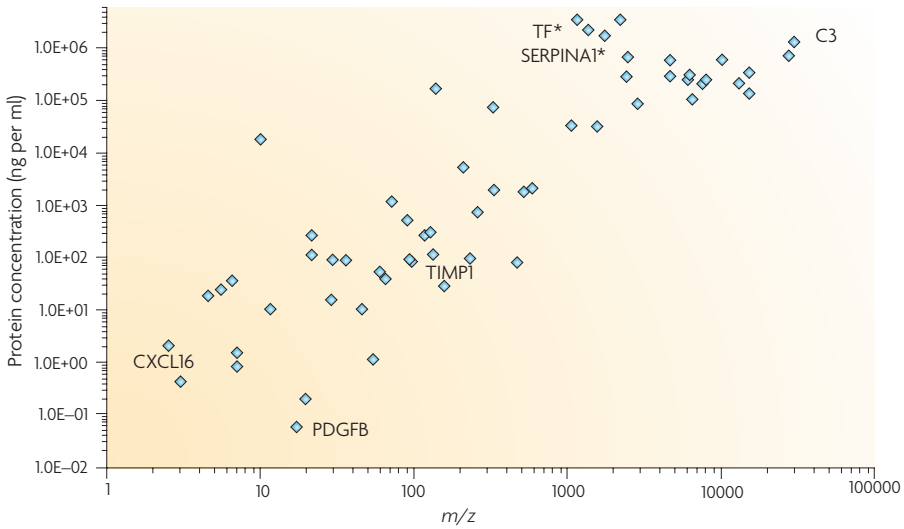


Figure 2 | **Depth of analysis of the plasma proteome.** Identification of 1,442 proteins in plasma that span more than 6 logs of protein abundance as determined by mass spectral counts and independent assays of a subset of proteins identified in a study of plasma from a pancreatic cancer mouse model[91]. PDGFB, platelet-derived growth factor B.

Profiling the cancer proteome in tumours

Illustrative of the progress made so far in cancer proteomics are studies of lung cancer and other epithelial tumours, which have encompassed tumour tissues, cells and biological fluids. The goals of such studies have included deciphering signalling pathways, identifying signatures related to tumour initiation, invasion and metastasis, and the discovery of diagnostic, predictive and prognostic biomarkers. Quantitative MS allows large-scale proteomics-based analysis of signalling events[30]. Metabolic stable isotope labelling by amino acids in cell culture (SILAC) provides a means for quantitative proteomic analysis. A study of colon cancer cells based on SILAC labelling followed by enrichment in phosphotyrosine proteins in combination with LC–MS/MS yielded 136 proteins that had a significant increase in tyrosine phosphorylation on SRC expression. A cluster of tyrosine kinases, SYK, EPH receptor A2 (EPHA2), SGK223, focal adhesion kinase (FAK) and MET, were found to be phosphoryated by SRC and required for invasive activity of SRC[31].

The use of SILAC has been extended to tissue analysis (FIG. 3). Its application to breast tumour tissue resulted in the identification of more than 5,000 proteins, including 100 protein kinases and 100 phosphopeptides, without the need for enrichment[32]. In addition, lung cancer has been the subject of multiple proteomic strategies to characterize signalling. To explore signalling downstream of oncogenic KRAS and epidermal growth factor receptor (EGFR) — the two most commonly mutated oncoproteins in non-small-cell lung cancer (NSCLC) — tyrosine phosphorylated proteins were extensively profiled by SILAC labelling and quantitative MS in isogenic human bronchial epithelial cells

(HBECs) and human lung adenocarcinoma cell lines. These cells expressed either one of two mutant alleles of EGFR (L858R and Del E746-750) or a mutant KRAS allele[33]. Tyrosine phosphorylation of some key signalling proteins was more predominant in HBECs expressing mutant EGFR than in HBECs expressing wild-type EGFR or mutant KRAS. Interestingly, ErbB receptor family proteins exhibited differences in phosphorylation at individual tyrosine residues between HBECs expressing EGFR-L858R and HBECs expressing the deletion mutant (EGFR-DelE746-750). Lysate array studies of proteins in signalling pathways relevant to EGFR that relied on laser-capture-microdissected NSCLC tissue led to the observation of differential phosphorylation of EGFR at residues Tyr1148, Tyr1068 and Tyr1045 between wild-type and mutant EGFR-expressing lung tumour tissue[34].

A Phosphoscan approach was used to study phosphotyrosine signalling downstream of EGFR in EGFR inhibitor-sensitive and EGFR inhibitor-resistant NSCLC cell lines[35]. Differential protein phosphorylation was observed between gefitinib-sensitive and gefitinib-resistant cell lines. A phosphoproteomic approach was applied to survey tyrosine kinase signalling across 41 NSCLC cell lines and tumours, which provided evidence of activated tyrosine kinases that included platelet-derived growth factor receptor-α (PDGFRα) and discoidin domain receptor tyrosine kinase 1 (DDR1), which had not been previously implicated in NSCLC[36]. Moreover, an MS strategy that quantified the temporal phosphorylation profiles of 222 tyrosine phosphorylated peptides across 7 time points following EGF treatment uncovered 31 tyrosine phosphorylation sites not previously known to be regulated by EGF stimulation[37]. These studies illustrate the depth of analysis of signalling processes in cancer that is currently achievable using proteomic methodologies. Although the emphasis has been on kinase activation and phosphorylation, there is also a need to develop methodologies that allow the similar assessment of the contribution of other post-translational modifications to signalling.

Several studies have sought to identify protein signatures that are related to tumour initiation, invasion and metastasis through the analysis of whole-cell and tissue lysates as well as through the analysis of individual compartments. Membrane purification of breast cancer cells followed by LC–MS/MS led to the identification of

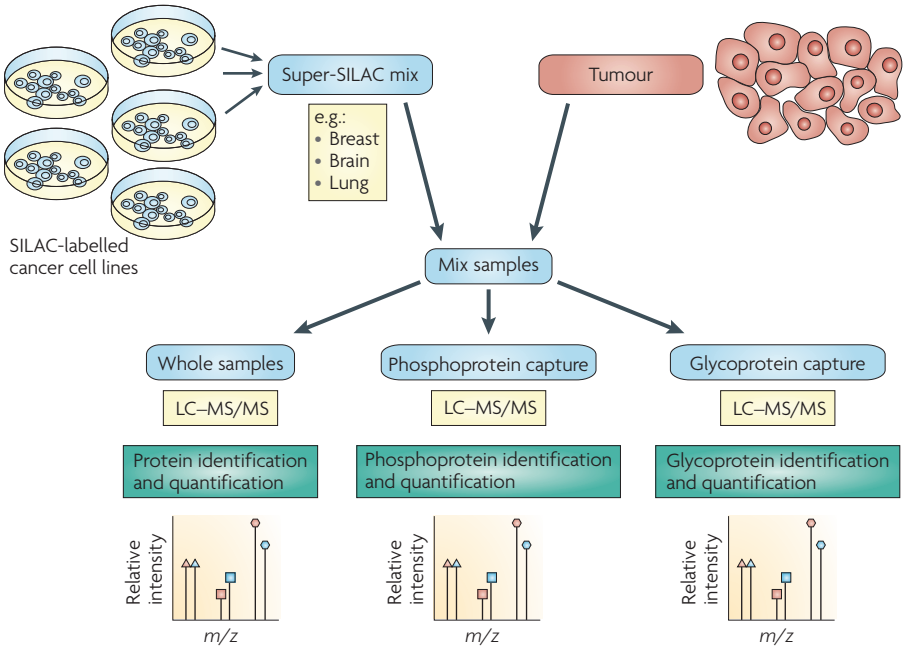


Figure 3 | **Quantitative profiling of proteins in tumour tissue using a modified SILAC strategy[32].** Stable isotope labelling by amino acids in cell culture (SILAC)-labelled proteins are extracted from cancer cell lines, mixed with tumour tissue lysates that are subjected to mass spectrometry (MS)-based analysis for protein identification and quantification of concentrations in tissue relative to cell lines for standardization. Alternatively, phosphoproteins or glycoproteins can be targeted for analysis through enrichment procedures. LC, liquid chromatography; *m/z*, mass to charge ratio.

1,919 proteins, of which 13 were differentially expressed in metastatic and non-metastatic cells[38]. Among these, overexpression of ecto-5′-nucleotidase (CD73) and integrin β1 by immunohistochemistry was found to be significantly associated with poor outcome. A comparison of normal bronchial epithelium, hyperplasia, squamous metaplasia, dysplasia, squamous cell carcinoma, atypical adenomatous hyperplasia and adenocarcinoma from 144 patients yielded a substantial number of differentially expressed proteins[39]. Comparative analysis of microdissected tumour tissue of lung squamous cell carcinoma that metastasized to lymph nodes with non-metastatic tumours by 2D DIGE identified proteins that were differentially expressed in the two groups, including increased expression of annexin A2 (ANXA2), heat shock protein 27 and cytokeratin 19 (KRT19) in the metastatic group[40]. Reduced levels of 14-3-3σ were found in the metastatic group and 14-3-3σ was shown to have an inhibitory role in invasion[41]. In a study of lung adenocarcinomas, integration of 2D PAGE and transcriptomic lung tumour data resulted in complementary findings: upregulated levels of RNAs and proteins that were associated with the glycolytic pathway were predictive of poor outcome in early-stage disease[42]. In another study, a proteomic signature in lung tumours was found to be associated with relapse-free survival and overall survival[43].

In studies of colon cancer, an LC–MS/MS-based comparison of the secretomes from the lymph-node metastatic SW620 colon cancer cell line with its primary counterpart yielded 910 proteins, of which 145 were differentially expressed. Serum levels of two of the differentially expressed proteins, trefoil factor 3 (TFF3) and growth differentiation factor 15 (GDF15), were found to be significantly higher in cases of metastatic colorectal cancer compared with non-metastatic controls[44]. A comparison of cell surface proteins of two colon cancer cell lines, KM12C and KM12SM, that differed in their metastatic potential yielded 291 membrane proteins, of which 60 proteins were differentially expressed[45]. Tissue microarrays provided evidence that the expression of two of the differentially expressed proteins, junction plakoglobin (JUP) and hydroxysteroid (17-β) dehydrogenase 8 (HSD17B8), correlated with disease progression. The associations reported in these studies are correlative in nature. Integration of findings with genomic data and with functional analysis will probably lead to a mechanistic understanding of the role of proteomic alterations in tumour development and metastasis.

Circulating cancer biomarkers

The rich content in proteins of serum and plasma that reflect diverse physiological or pathological states, and the ease with which this compartment can be sampled, make it a favourite choice for biomarker applications. Because of the complexity of plasma and the substantial dynamic range in abundance of its protein constituents, indirect approaches have been used to search for cancer-associated proteins that might be found in this compartment. One approach is to develop a catalogue of secreted proteins that are expressed by cancer cells. Analysis of the conditioned media of lung cancer cell lines of different histological backgrounds yielded 1,830 proteins that were either unique or common across lineages[46]. Analysis of ovarian cancer cells in culture yielded 6,400 proteins, including a substantial proportion of proteins that were secreted or shed into the media. The proteins shed into the media were found to be particularly related to processes of cell adhesion[47]. SILAC was used to compare the secreted proteins from pancreatic cancer-derived cells with proteins from non-neoplastic pancreatic duct epithelial cells[48]. Of the 145 differentially secreted proteins that were identified, some had been previously described as associated with pancreatic cancer and some were novel. Such catalogues of secreted proteins have a somewhat limited utility without knowledge of the secretion rates, clearance from the circulation and differential secretion between cancer cells compared with other normal cell types that also secrete these proteins, and so this might limit the contribution of cancer cells to protein circulating levels[49].

As an alternative, proximal biological fluids enriched with proteins from tumour tissue have been investigated as a source of circulating cancer biomarkers. For example, RBAP46 (also known as retinoblastoma binding protein 7 (RBBP7)) has been found to be upregulated in both conditioned media of NSCLC cell lines and pleural effusions from patients with lung cancer on the basis of LC–MS/MS analysis[50]. RBAP46 was also demonstrated to be upregulated in the tumour tissue and serum of patients with lung cancer. In another study, comparison of plasma and pleural effusions from patients with lung cancer with those from subjects with inflammatory pleuritis using narrow-range ampholytes and LC–MS/MS yielded a large number of tumour tissue proteins that were enriched in pleural fluid from patients with lung cancer[51]. Quantitative proteomics was applied to pancreatic juice from subjects with pancreatic intraepithelial neoplasias (representing precursor lesions for pancreatic cancer) and from patients with pancreatic cancer. This led to the identification of proteins with increased levels in pancreatic juice from precursor and/or cancer cases compared with controls[52]. The protein anterior gradient homologue 2 (AGR2) seemed promising on the basis of pancreatic juice findings[53]. However, there was no correlation observed between AGR2 levels in pancreatic juice and serum; this is illustrative of the challenges of predicting disease progression from protein analysis in one compartment and protein levels in the compartment of interest (that is, in serum and plasma).

The success of direct profiling of serum and plasma for circulating cancer biomarkers has depended on the depth of analysis and rigor in experimental design for the choice of cases and controls[54]. In studies of NSCLC, 1D LC–MS/MS analysis of serum samples from patients and healthy controls yielded 931 proteins that were identified with at least two peptides, and 62 proteins that were differentially expressed between the two comparison groups[55]. The usefulness of these candidates remains to be determined. A new promising approach to lung cancer biomarker discovery is based on the proteomic profiling of pulmonary venous effluent that drains the tumour vascular bed as well as the proteomic profiling of systemic arterial blood obtained from the same subjects — this is based on the concept that effluent blood contains higher concentrations of potential biomarkers compared with more distal blood[56]. Connective tissue-activating peptide III (CTAPIII; also known as NAP2 or pro-platelet basic protein (PPBP)) was found to occur at significantly higher concentrations in effluent blood than in peripheral blood, and blood levels of CTAPIII were found to be increased in patients with lung cancer in two independent population cohorts. Levels of this protein were found to be decreased after tumour resection.

The identification of biomarkers of drug resistance in the treatment of lung cancer has been the subject of numerous proteomic studies using either comparative analysis of drug-sensitive and drug-resistant cell lines[57–60] or clinical samples. Comparative analysis of lung adenocarcinoma tissue protein expression profiles using 2D DIGE resulted in a set of proteins that could distinguish responders to gefitinib treatment from non-responders[61]. An algorithm was developed to predict outcome after the treatment of patients with NSCLC using an EGFR inhibitor based on proteomic analysis of serum. The algorithm was tested with data from several independent cohorts and was demonstrated to reliably classify patients according to their outcome[62]. This predictor was also associated with survival after treatment with erlotinib in patients undergoing first-line therapy for NSCLC[63]. A predictive algorithm based on 11 distinct mass to charge ratio (m/z) features could predict overall survival and progression-free survival in a blinded test set of patients treated with erlotinib[64]. The elucidation of the identity of related proteins and the basis for their correlation with outcome would add more importance to the findings.

Harnessing the immune response to cancer

Numerous proteomic studies have sought to identify cancer biomarkers by harnessing the immune response that occurs with tumour development[65]. Autoantibodies against

Glossary

2D PAGE
A process of separating proteins in gels based on their charge and molecular mass.

Electrospray ionization
ESI. A mass spectrometry method to ionize macromolecules or peptides by electrospray leading to their identification.

Fluorescence difference gel electrophoresis
DIGE. A method that labels protein samples with fluorescent dyes before electrophoresis.

Immobilized pH gradients
IPG. A process of generating a pH gradient by immobilizing gradient chemicals (immobilines) in the acrylamide matrix.

Lectin
A sugar-binding protein that is specific for the sugar moieties it binds.

Matrix-assisted laser desorption ionization
MALDI. A mass spectrometry method to ionize proteins and peptides deposited in a matrix leading to their identification.

Multiple reaction monitoring
MRM. A technique that targets multiple specific peptides for their quantification by mass spectrometry.

Secretome
The ensemble of proteins released by cells into the extracellular environment.

Stable isotope labelling by amino acids in cell culture
SILAC. A method for non-radioactive labelling of proteins in culture based on the uptake of labelled amino acids.

Tandem mass spectrometry
MS/MS. A two-stage separation process in mass spectrometry with fragmentation in-between allowing identification of the precursor.

autologous tumour-associated antigens have been detected before the onset of symptoms and so could be biomarkers for early cancer diagnosis. Panels of antigens have been developed that are recognized by circulating autoantibodies among patients with epithelial tumours. Using a protein microarray approach, autoantibodies directed against annexin A1, 14-3-3σ and laminin receptor (LAMR1; also known as ribosomal protein SA (RPSA)) have been demonstrated in the sera of newly diagnosed patients with lung cancer, as well as in pre-diagnostic sera[66–68]. Another study demonstrated the occurrence of autoantibodies to cytokeratin 18 and villin 1 (REF. 69). An autoantibody signature was uncovered in sera from patients with different subtypes of lung cancer by means of protein arrays from a cDNA expression library, which allowed the discrimination of lung cancer cases from controls[70]. A phage cDNA expression library constructed with tumour tissue from 30 patients with lung cancer and biopanned using serum pools of patients with NSCLC and healthy controls resulted in the identification of 6 phage peptide clones with high seropositivity among patients with NSCLC[71].

In another study, ten tumour-associated antigens that were selected on the basis of prior studies were assayed in sera from newly diagnosed patients with lung cancer, smokers with ground-glass opacities or benign solid nodules and smoker controls. This resulted in a panel consisting of MYC, cyclin A, cyclin B1, cyclin D1, cyclin-dependent kinase 2 (CDK2) and survivin that had an 81% sensitivity at 97% specificity in distinguishing between cancer cases and smoker controls[72]. Similar approaches used in the search of autoantibodies to tumour antigens in pancreatic cancer have uncovered immunoglobulin-based reactivity against a large number of metabolic enzymes and cytoskeletal proteins[73]. In another study 2D PAGE western blots, revealed six different isoforms of enolase A1. Serum from patients with pancreatic cancer exhibited autoantibody reactivity directed specifically against the two most acidic and phosphorylated isoforms[74]. Proteomic analysis also led to the identification of specific calreticulin isoforms that were associated with autoantibodies in pancreatic cancer[75]. Likewise, proteomic approaches have led to the identification of multiple tumour antigens that induce autoantibodies in breast cancer. Reactivity to a panel of five antigens resulted in the substantial discrimination between breast cancer cases, carcinoma *in situ* cases and healthy controls[76].

Cancer proteomics grand challenges

Given the multitude of approaches to profile the cancer proteome in its various dimensions, as presented in the preceding sections, there is a compelling need for large-scale, integrative and collaborative efforts to elucidate the range of proteome alterations in cancer, similar to the current efforts to define the range of genomic alterations in cancer. Although current technologies allow unprecedented depths of analyses,
it is impractical to conceptualize an all-encompassing single human cancer proteome project, given the inherent complexity of the proteome. An alternative would be to conceptualize several targeted cancer proteome projects and initiatives with clearly defined objectives and milestones. Two such types of 'cancer proteomics grand challenges' would address the need for cancer biomarkers that have diagnostic relevance on the one hand and the need to define altered signalling pathways in cancer that have therapeutic
relevance on the other hand.

> "there is a compelling need for large-scale, integrative and collaborative efforts to elucidate the range of proteome alterations in cancer"

Advances in the deep profiling of the plasma proteome and glycoproteome offer an opportunity for a major collaborative initiative to profile the plasma oncoproteome. The driving objectives of such an initiative would include the discovery and validation of biomarkers that inform of the risk of developing common cancer types, as well as biomarkers for the early detection of cancer and the diagnosis and monitoring of cancer for regression and progression (BOX 1). An initiative to define altered signalling pathways associated with tumorigenesis would include objectives to define post-translational modifications in cancer, to elucidate the cancer interactome, and to define the altered localization of proteins in various intracellular and extracellular compartments. The driving objectives of a signalling initiative would be to develop new targets for molecular imaging and therapeutics as well as novel classification schemes that are predictive of response to targeted therapy. A notable development in this regard is a request for proposals issued by the Clinical Proteomic Technologies for Cancer (CPTC) programme at the National Cancer Institute (RFA-CA-10-016 (see Further information)). The goal is to systematically define the functional cancer proteome that derives from alterations in cancer genomes. These types of projects would further drive the development of resources and improvements in technology that are needed for their execution.

Some pilot projects have been initiated that support the feasibility of engaging in large-scale proteomic initiatives. The first proteome project to be conceived through the activity of the Human Proteome Organization (HUPO; see Further information) is a plasma proteome project, which has completed its pilot phase. There was a consensus among leaders in industry, government and academia that the plasma is perhaps the most crucial compartment for information about the status of health and disease of an individual through profiling of its protein constituents. The pilot phase of this project had objectives that included the comparison of a broad range of technology platforms for the characterization of proteins in human plasma and serum, as well as the assessment of the influence of various technical variables on specimen collection, handling and storage. Standardized samples were distributed to 18 participating laboratories and an integrated analysis of the resulting data was carried out[77]. An initial integration of data resulted in 3,020 proteins identified with 2 or more peptides. Application of rigorous statistical methodology, taking into account multiple hypothesis-testing, resulted in a reduced set of 889 proteins identified with high confidence[78]. This pilot phase, though modest in scale, is illustrative of the merits of an organized, collaborative effort around a well-defined study.

A notable project that illustrates the merits of developing resources for proteomics and that has relevance to cancer is the Human Protein Atlas project (see the Human Protein Atlas website; Further information)[79]. This project aims to experimentally annotate the human protein complement of the genome in a gene-centric manner using antibodies for the systematic analysis of the cellular distribution and the subcellular localization of proteins in normal and disease tissues. Antibodies are validated on the basis of protein array assays, western blot analysis, immunohistochemistry and immunofluorescence-based confocal

microscopy. An application of Protein Atlas is the development of a web-based tool for *in silico* biomarker discovery for cancer[80]. Search queries are based on the human tissue profiles in normal and cancer cells in the Human Protein Atlas portal and rely on annotations, carried out by pathologists, of images representing immunohistochemically stained tissue sections. Search tools allow the exploration of the Human Protein Atlas to discover potential tissue-specific, cell type-specific and tumour type-specific markers. Another illustrative project that is directly relevant to proteomics-based cancer investigations has been focused on the development of a library of human cancer-specific peptides for use in multiple reaction monitoring for biomarker discovery and validation[81].

Informatics to mine the oncoproteome

Massive amounts of complex and heterogeneous proteomic data are currently being generated. It can be argued that proteomic data are currently being generated faster than they can be fully exploited, and the pace is likely to continue to accelerate. Furthermore, effective mining of the cancer proteome would require the integration of proteomic data from multiple sources, including primary tissue, cells, biological fluids and animal models, as well as the integration of proteomic data with other types of data. Databases have been developed for depositing and retrieving proteomic data sets, including PRIDE[82], PeptideAtlas[83], UniPep[84], the Global Proteome Machine[85], Proteopedia[86], and Proteome Commons and its Tranche file-sharing system (see the PRIDE, PeptideAtlas, UniPep, Global Proteome Machine, Proteopedia and the Proteome Commons websites; Further information). However, there remains a need to organize such data and make accessible detailed characteristics of samples and experimental conditions associated with the data in a readily retrievable manner. Therefore, an oncoproteome database that encompasses cancer proteomic data derived from cell, tissue and biological fluid profiling would provide a valuable resource in the quest to decipher protein alterations in cancer and translate the findings into clinical applications.

Perspective

Advances in proteomics technologies have allowed the profiling of the proteome with substantial reliability and depth of analysis, making it possible to interrogate the proteome for applications to cancer. However, such developments have not been widely appreciated. A case in point is the frequently expressed requirement that proteomic data need to be validated using old-fashioned criteria such as western blots, which may be less quantitative, less specific and therefore less informative[87]. The multiple facets of the proteome and its complexity and the potential for alterations at multiple levels in cancer complicate the task of mining the cancer proteome. As a result, there is substantial benefit from the development of a 'road map' with clearly defined objectives and milestones to move forwards in cancer proteomics. Although challenging to implement, given the need to engage policy makers, scientists and both the private and public sectors on an international scale, much can be learned from the success of genomics in meeting this challenge.

Box 1 | **Objectives of a cancer plasma proteome project**

A grand challenge for proteomics would be to develop new cancer diagnostics based on the application of proteomics technologies to the plasma proteome. Such a challenge would comprise:

- Comprehensive quantitative analysis of plasma protein constituents among subjects that were later diagnosed with cancer that were part of large population cohorts to identify proteins that might predict the onset or risk of major common cancers.
- Comprehensive quantitative analysis of plasma protein constituents among newly diagnosed subjects with cancer to define subsets that provide information about particular lineages and signalling pathways that drive tumour development, progression and response to targeted therapies.
- Elucidation of protein subsets that vary with biological parameters including age, sex and ethnicity.
- Elucidation of protein subsets that vary with diet and with common exposures, notably tobacco smoke.
- Development of a knowledge base of plasma protein alterations in cancer.
- Development of affinity capture agents for plasma proteins with cancer relevance.
- Development of proteotypic peptides for cleaved and modified plasma proteins associated with cancer.
- Development of standardized specimen reference sets to be made available to investigators.

Samir Hanash and Ayumu Taguchi are at the Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, M5-C800, PO BOX 19024, Seattle, Washington 98109, USA
Correspondence to S.H.
e-mail: shanash@fhcrc.org

doi:10.1038/nrc2918

1. Fine, J. M. & Creyssel, R. Starch gel electrophoresis studies on abnormal proteins in myeloma and macroglobulinaemia. *Nature* **183**, 392 (1959).
2. Hanash, S. M., Baier, L. J., McCurry, L. & Schwartz, S. Lineage related polypeptide markers in acute lymphoblastic leukemia detected by two-dimensional electrophoresis. *Proc. Natl Acad. Sci. USA* **83**, 807–811 (1986).
3. Moreira, J. M. *et al.* Bladder cancer-associated protein, a potential prognostic biomarker in human bladder cancer. *Mol. Cell. Proteomics* **9**, 161–177 (2010).
4. Kondo, T. Cancer proteome-expression database: Genome Medicine Database of Japan Proteomics. *Expert Rev. Proteomics* **7**, 21–27 (2010).
5. Strahler, J. R. *et al.* High resolution two-dimensional polyacrylamide gel electrophoresis of basic polypeptides: use of immobilized pH gradients in the first dimension. *Electrophoresis* **8**, 165–173 (1987).
6. Timms, J. F. & Cramer, R. Difference gel electrophoresis. *Proteomics* **8**, 4886–4897 (2008).
7. Hanash, S., Pitteri, S. & Faca, V. Mining the plasma proteome for cancer biomarkers. *Nature* **452**, 571–579 (2008).
8. Cravatt, B. F., Simon, G. M. & Yates, J. R. The biological impact of mass-spectrometry-based proteomics. *Nature* **450**, 991–1000 (2007).
9. Siuti, N. & Kelleher, N. L. Decoding protein modifications using top-down mass spectrometry. *Nature Methods* **4**, 817–821 (2007).
10. Nesvizhskii, A. I., Vitek, O. & Aebersold, R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature Methods* **4**, 787–797 (2007).
11. Cox, J. & Mann, M. Is proteomics the new genomics? *Cell* **130**, 395–398 (2007).
12. Petricoin, E. F., Belluco, C., Araujo, R. P. & Liotta, L. A. The blood peptidome: a higher dimension of information content for cancer biomarker discovery. *Nature Rev. Cancer* **6**, 961–967 (2006).
13. Hanash, S. Disease proteomics. *Nature* **422**, 226–232 (2003).
14. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
15. Hanash, S. M. *et al.* Highly resolving two-dimensional gels for protein sequencing. *Proc. Natl Acad. Sci. USA* **88**, 5709–5713 (1991).
16. Constans, A. MALDI - Pioneering ionization technique paved the way for proteomics. *The Scientist* **19**, 37 (2005).
17. Fujii, K., Kondo, T., Yamada, M., Iwatsuki, K. & Hirohashi, S. Toward a comprehensive quantitative proteome database: protein expression map of lymphoid neoplasms by 2-D DIGE and MS. *Proteomics* **6**, 4856–4876 (2006).
18. Fenn., J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64–71 (1989).
19. Olsen, J. V. *et al.* A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* **8**, 2759–2769 (2009).
20. Roesli, C. *et al.* Comparative analysis of the membrane proteome of closely related metastatic and nonmetastatic tumor cells. *Cancer Res.* **69**, 5406–5414 (2009).
21. Bell, A. W. *et al.* A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nature Methods* **6**, 423–430 (2009).

22. Addona, T. A. *et al.* Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nature Biotech.* **27**, 633–641 (2009).
23. Stoevesandt, O., Taussig, M. J. & He, M. Protein microarrays: high-throughput tools for proteomics. *Expert Rev. Proteomics* **6**, 145–157 (2009).
24. Hu, S. *et al.* Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell* **139**, 610–622 (2009).
25. Anderson, T., Wulfkuhle, J., Liotta, L., Winslow, R. L. & Petricoin, E. Improved reproducibility of reverse-phase protein microarrays using array microenvironment normalization. *Proteomics* **9**, 5562–5566 (2009).
26. Taylor, A. D., Hancock, W. S., Hincapie, M., Taniguchi, N. & Hanash, S. M. Towards an integrated proteomic and glycomic approach to finding cancer biomarkers. *Genome Med.* **1**, 57 (2009).
27. Heo, S. H., Lee, S. J., Ryoo, H. M., Park, J. Y. & Cho, J. Y. Identification of putative serum glycoprotein biomarkers for human lung adenocarcinoma by multilectin affinity chromatography and LC-MS/MS. *Proteomics* **7**, 4292–4302 (2007).
28. Vercoutter-Edouart, A. S., Slomianny, M. C., Dekeyzer-Beseme, O., Haeuw, J. F. & Michalski, J. C. Glycoproteomics and glycomics investigation of membrane *N*-glycosylproteins from human colon carcinoma cells. *Proteomics* **8**, 3236–3256 (2008).
29. Kim, Y. S. *et al.* Functional proteomics study reveals that *N*-Acetylglucosaminyltransferase V reinforces the invasive/metastatic potential of colon cancer through aberrant glycosylation on tissue inhibitor of metalloproteinase-1. *Mol. Cell. Proteomics* **7**, 1–14 (2008).
30. Choudhary, C. & Mann, M. Decoding signalling networks by mass spectrometry-based proteomics. *Nature Rev. Mol. Cell Biol.* **11**, 427–439 (2010).
31. Leroy, C. *et al.* Quantitative phosphoproteomics reveals a cluster of tyrosine kinases that mediates SRC invasive activity in advanced colon carcinoma cells. *Cancer Res.* **69**, 2279–2286 (2009).
32. Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R. & Mann, M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature Methods* **7**, 383–385 (2010).
33. Guha, U. *et al.* Comparisons of tyrosine phosphorylated proteins in cells expressing lung cancer-specific alleles of EGFR and KRAS. *Proc. Natl Acad. Sci. USA* **105**, 14112–14117 (2008).
34. VanMeter, A. J. *et al.* Laser capture microdissection and protein microarray analysis of human non-small cell lung cancer: differential epidermal growth factor receptor (EGPR) phosphorylation events associated with mutated EGFR compared with wild type. *Mol. Cell. Proteomics* **7**, 1902–1924 (2008).
35. Guo, A. *et al.* Signaling networks assembled by oncogenic EGFR and c-Met. *Proc. Natl Acad. Sci. USA* **105**, 692–697 (2008).
36. Rikova, K. *et al.* Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131**, 1190–1203 (2007).
37. Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A. & White, F. M. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl Acad. Sci. USA* **104**, 5860–5865 (2007).
38. Leth-Larsen, R. *et al.* Metastasis-related plasma membrane proteins of human breast cancer cells identified by comparative quantitative mass spectrometry. *Mol. Cell. Proteomics* **8**, 1436–1449 (2009).
39. Shen, J. *et al.* Identification and validation of differences in protein levels in normal, premalignant, and malignant lung cells and tissues using high-throughput Western Array and immunohistochemistry. *Cancer Res.* **66**, 11194–11206 (2006).
40. Yao, H. *et al.* Identification of metastasis associated proteins in human lung squamous carcinoma using two-dimensional difference gel electrophoresis and laser capture microdissection. *Lung Cancer* **65**, 41–48 (2009).
41. Li, D. J. *et al.* Identificating 14-3-3 sigma as a lymph node metastasis-related protein in human lung squamous carcinoma. *Cancer Lett.* **279**, 65–73 (2009).
42. Chen, G. *et al.* Protein profiles associated with survival in lung adenocarcinoma. *Proc. Natl Acad. Sci. USA* **100**, 13537–13542 (2003).
43. Yanagisawa, K. *et al.* A 25-signal proteomic signature and outcome for patients with resected non-small-cell lung cancer. *J. Natl. Cancer Inst.* **99**, 858–867 (2007).
44. Xue, H. *et al.* Identification of serum biomarkers for colorectal cancer metastasis using a differential secretome approach. *J. Proteome Res.* **9**, 545–555.
45. Luque-Garcia, J. L. *et al.* Differential protein expression on the cell surface of colorectal cancer cells associated to tumor metastasis. *Proteomics* **10**, 940–952 (2010).
46. Planque, C. *et al.* Identification of five candidate lung cancer biomarkers by proteomics analysis of conditioned media of four lung cancer cell lines. *Mol. Cell. Proteomics* **8**, 2746–2758 (2009).
47. Faca, V. M. *et al.* Proteomic analysis of ovarian cancer cells reveals dynamic processes of protein secretion and shedding of extra-cellular domains. *PLoS ONE* **3**, e2425 (2008).
48. Gronborg, M. *et al.* Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. *Mol. Cell. Proteomics* **5**, 157–171 (2006).
49. Lutz, A. M., Willmann, J. K., Cochran, F. V., Ray, P. & Gambhir, S. S. Cancer screening: a mathematical model relating secreted blood biomarker levels to tumor sizes. *PLoS Med.* **5**, e170 (2008).
50. Wang, C. L. *et al.* Discovery of retinoblastoma-associated binding protein 46 as a novel prognostic marker for distant metastasis in nonsmall cell lung cancer by combined analysis of cancer cell secretome and pleural effusion proteome. *J. Proteome Res.* **8**, 4428–4440 (2009).
51. Pernemalm, M. *et al.* Use of narrow-range peptide IEF to improve detection of lung adenocarcinoma markers in plasma and pleural effusion. *Proteomics* **9**, 3414–3424 (2009).
52. Tian, M. *et al.* Proteomic analysis identifies MMP-9, DJ-1 and A1BG as overexpressed proteins in pancreatic juice from pancreatic ductal adenocarcinoma patients. *BMC Cancer* **8**, 241 (2008).
53. Chen, R. *et al.* Elevated level of anterior gradient-2 in pancreatic juice from patients with pre-malignant pancreatic neoplasia. *Mol. Cancer* **9**, 149 (2010).
54. Pepe, M. S., Feng, Z., Janes, H., Bossuyt, P. M. & Potter, J. D. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J. Natl Cancer Inst.* **100**, 1432–1438 (2008).
55. Pan, J., Chen, H. Q., Sun, Y. H., Zhang, J. H. & Luo, X. Y. Comparative proteomic analysis of non-small-cell lung cancer and normal controls using serum label-free quantitative shotgun technology. *Lung* **186**, 255–261 (2008).
56. Yee, J. *et al.* Connective tissue-activating peptide III: a novel blood biomarker for early lung cancer detection. *J. Clin. Oncol.* **27**, 2787–2792 (2009).
57. Patel, N. *et al.* Rescue of paclitaxel sensitivity by repression of Prohibitin1 in drug-resistant cancer cells. *Proc. Natl Acad. Sci. USA* **107**, 2503–2508.
58. Qu, Y., Yang, Y., Liu, B. & Xiao, W. Comparative proteomic profiling identified sorcin being associated with gemcitabine resistance in non-small cell lung cancer. *Med. Oncol.* 10 Dec 2009 (doi:10.1007/s12032-009-9379-5).
59. Eriksson, H. *et al.* Quantitative membrane proteomics applying narrow range peptide isoelectric focusing for studies of small cell lung cancer resistance mechanisms. *Proteomics* **8**, 3008–3018 (2008).
60. Keenan, J., Murphy, L., Henry, M., Meleady, P. & Clynes, M. Proteomic analysis of multidrug-resistance mechanisms in adriamycin-resistant variants of DLKP, a squamous lung cancer cell line. *Proteomics* **9**, 1556–1566 (2009).
61. Okano, T. *et al.* Proteomic signature corresponding to the response to gefitinib (Iressa, ZD1839), an epidermal growth factor receptor tyrosine kinase inhibitor in lung adenocarcinoma. *Clin. Cancer Res.* **13**, 799–805 (2007).
62. Taguchi, F. *et al.* Mass spectrometry to classify non-small-cell lung cancer patients for clinical outcome after treatment with epidermal growth factor receptor tyrosine kinase inhibitors: a multicohort cross-institutional study. *J. Natl Cancer Inst.* **99**, 838–846 (2007).
63. Amann, J. M. *et al.* Genetic and proteomic features associated with survival after treatment with erlotinib in first-line therapy of non-small cell lung cancer in Eastern Cooperative Oncology Group 3503.*J. Thorac. Oncol.* **5**, 169–178 (2010).
64. Salmon, S. *et al.* Classification by mass spectrometry can accurately and reliably predict outcome in patients with non-small cell lung cancer treated with erlotinib-containing regimen. *J. Thorac. Oncol.* **4**, 689–696 (2009).
65. Desmetz, C., Maudelonde, T., Mange, A. & Solassol, J. Identifying autoantibody signatures in cancer: a promising challenge. *Expert Rev. Proteomics* **6**, 377–386 (2009).
66. Madoz-Gurpide, J., Kuick, R., Wang, H., Misek, D. E. & Hanash, S. M. Integral protein microarrays for the identification of lung cancer antigens in sera that induce a humoral immune response. *Mol. Cell. Proteomics* **7**, 268–281 (2008).
67. Pereira-Faca, S. R. *et al.* Identification of 14-3-3 theta as an antigen that induces a humoral response in lung cancer. *Cancer Res.* **67**, 12000–12006 (2007).
68. Qiu, J. *et al.* Occurrence of autoantibodies to annexin I, 14-3-3 theta and LAMR1 in prediagnostic lung cancer sera. *J. Clin. Oncol.* **26**, 5060–5066 (2008).
69. Nagashio, R. *et al.* Detection of tumor-specific autoantibodies in sera of patients with lung cancer. *Lung Cancer* **62**, 364–373 (2008).
70. Leidinger, P. *et al.* Identification of lung cancer with high sensitivity and specificity by blood testing. *Respir. Res.* **11**, 18 (2010).
71. Wu, L. L. *et al.* Development of autoantibody signatures as novel diagnostic biomarkers of non-small cell lung cancer. *Clin. Cancer Res.* **16**, 3760–3768 (2010).
72. Rom, W. N. *et al.* Identification of an autoantibody panel to separate lung cancer from smokers and nonsmokers. *BMC Cancer* **10**, 234 (2010).
73. Tomaino, B. *et al.* Autoantibody signature in human ductal pancreatic adenocarcinoma. *J. Proteome Res.* **6**, 4025–4031 (2007).
74. Tomaino, B. *et al.* Circulating autoantibodies to phosphorylated α-enolase are a hallmark of pancreatic cancer. *J. Proteome Res.* 10 Jun 2010 (doi:10.1021/pr100213b).
75. Hong, S. H. *et al.* An autoantibody-mediated immune response to calreticulin isoforms in pancreatic center. *Cancer Res.* **64**, 5504–5510 (2004).
76. Desmetz, C. *et al.* Identification of a new panel of serum autoantibodies associated with the presence of *in situ* carcinoma of the breast in younger women. *Clin. Cancer Res.* **15**, 4733–4741 (2009).
77. Omenn, G. S. *et al.* Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **5**, 3226–3245 (2005).
78. States, D. J. *et al.* Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nature Biotech.* **24**, 333–338 (2006).
79. Berglund, L. *et al.* A genecentric Human Protein Atlas for expression profiles based on antibodies. *Mol. Cell. Proteomics* **7**, 2019–2027 (2008).
80. Bjorling, E. *et al.* A web-based tool for *in silico* biomarker discovery based on tissue-specific protein profiles in normal and cancer tissues. *Mol. Cell. Proteomics* **7**, 825–844 (2008).
81. Yang, X. & Lazar, I. M. MRM screening/biomarker discovery with linear ion trap MS: a library of human cancer-specific peptides. *BMC Cancer* **9**, 96 (2009).
82. Jones, P. *et al.* PRIDE: a public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Res.* **34**, D659–D663 (2006).
83. Deutsch, E. W. *et al.* Human Plasma PeptideAtlas. *Proteomics* **5**, 3497–3500 (2005).
84. Zhang, H. *et al.* UniPep-a database for human N-linked glycosites: a resource for biomarker discovery. *Genome Biol.* **7**, R73 (2006).
85. Craig, R., Cortens, J. P. & Beavis, R. C. Open source system for analyzing, validating, and storing protein identification data. *J. Proteome Res.* **3**, 1234–1242 (2004).
86. Mathivanan, S. *et al.* Human Proteinpedia enables sharing of human protein data. *Nature Biotech.* **26**, 164–167 (2008).
87. Mann, M. Can proteomics retire the western blot? *J. Proteome Res.* **7**, 3065–3065 (2008).
88. Rosenblum, B. B., Neel, J. V. & Hanash, S. M. Two-dimensional electrophoresis of plasma polypeptides reveals "high" heterozygosity indices. *Proc.Natl Acad. Sci. USA* **80**, 5002–5006 (1983).
89. Faca, V. *et al.* Contribution of protein fractionation to depth of analysis of the serum and plasma proteomes. *J. Proteome Res.* **6**, 3558–3565 (2007).

90. Faca, V. *et al.* Quantitative analysis of acrylamide labeled serum proteins by LC-MS/MS. *J. Proteome Res.* **5**, 2009–2018 (2006).
91. Faca, V. M. *et al.* A mouse to human search for plasma proteome changes associated with pancreatic tumor development. *PLoS Med.* **5**, e123 (2008).
92. Longsworth, L. G., Shedlovsky, T. & Macinnes, D. A. Electrophoretic patterns of normal and pathological human blood serum and plasma. *J. Exp. Med.* **70**, 399–413 (1939).
93. Hanash, S. M. & Strahler, J. R. Advances in two-dimensional electrophoresis. *Nature* **337**, 485–486 (1989).
94. Rasmussen, H. H., Mortz, E., Mann, M., Roepstorff, P. & Celis, J. E. Identification of transformation sensitive proteins recorded in human two-dimensional gel protein databases by mass spectrometric peptide mapping alone and in combination with microsequencing. *Electrophoresis* **15**, 406–416 (1994).
95. Mann, M., Hendrickson, R. C. & Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Ann. Rev. Biochem.* **70**, 437–473 (2001).
96. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**, 212–217 (2006).
97. Cravatt, B. F., Wright, A. T. & Kozarich, J. W. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Ann. Rev. Biochem.* **77**, 383–414 (2008).
98. Zhu, H., Bilgin, M. & Snyder, M. Proteomics. *Ann. Rev. Biochem.* **72**, 783–812 (2003).
99. Paweletz, C. P. *et al.* Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene* **20**, 1981–1989 (2001).
100. Nimrichter, L. *et al.* Intact cell adhesion to glycan microarrays. *Glycobiology* **14**, 197–203 (2004).
101. Tateno, H. *et al.* A novel strategy for mammalian cell surface glycome profiling using lectin microarray. *Glycobiology* **17**, 1138–1146 (2007).
102. Haab, B. B. Antibody arrays in cancer research. *Mol. Cell. Proteomics* **4**, 377–383 (2005).
103. Kirby, R. *et al.* Aptamer-based sensor arrays for the detection and quantitation of proteins. *Anal. Chem.* **76**, 4066–4075 (2004).
104. Gaster, R. S. *et al.* Matrix-insensitive protein assays push the limits of biosensors in medicine. *Nature Med.* **15**, 1327–1332 (2009).

Acknowledgements
The authors would like to thank colleagues in the Molecular Diagnostics Program at FHCRC for stimulating discussions.

Competing interests statement
The authors declare no competing financial interests.

DATABASES

National Cancer Institute Drug Dictionary: http://www.cancer.gov/drugdictionary/
erlotinib | gefitinib
Pathway Interaction Database: http://pid.nci.nih.gov/EGFR

FURTHER INFORMATION

Samir Hanash's homepage: http://hanash-lab.fhcrc.org/
Global Proteome Machine: http://www.thegpm.org/
Human Protein Atlas: http://www.proteinatlas.org/
HUPO: http://www.hupo.org/
Peptide atlas: http://www.peptideatlas.org/
PRIDE: http://www.ebi.ac.uk/pride/
Proteome Commons: http://www.proteomecommons.org/
Proteopedia: http://www.proteopedia.org/wiki/index.php/
RFA-CA-10-016: http://grants.nih.gov/grants/guide/rfa-files/RFA-CA-10-016.html
The Genome Medicine Database of Japan (GeMDBJ): http://gemdbj.nibio.go.jp
Unipep: http://www.unipep.org/

ALL LINKS ARE ACTIVE IN THE ONLINE PDF