TUMORIGENESIS

To the death!

In the early stages of tumour initiation in the epithelium, when only a few cells are transformed, evidence suggests that normal and transformed cells that are in close proximity compete for space in which to grow, causing the 'loser' cells to undergo JNK-dependent apoptosis. This phenomenon is known as cell competition and can be tumour suppressive if the losers are transformed cells. Two papers now provide further insight into how this process might function in tumorigenesis.

Lethal (2) giant larvae (*l(2)gl*) is a tumour suppressor in *Drosophila melanogaster* that, when mutated in a mosaic tissue (surrounded by normal cells), causes these $l(2)gl^-$ cells to become the losers in cell competition. L(2)GL is a scaffold protein, and Tamori and colleagues identified HIV-1 Vpr-binding protein (VPRBP) as a new binding protein of L(2)GL2 by immunoprecipitation in human MCF7 breast cancer cells and MDCK canine epithelial cells. They also identified a homologue in *D. melanogaster* that they named Mahjong (MAHJ) that bound L(2)GL. Using the FLP–FRT system the authors generated mosaic flies so that equal numbers of $mahj^{-/-}$ and normal cells were produced. The $mahj^{-/-}$ cells were eliminated from wing discs and further analysis revealed that $mahj^{-/-}$ cells adjacent to normal cells had activated caspase 3 and were undergoing JNK-dependent apoptosis. This suggests that loss of *mahj*, like loss of *l(2)gl*, causes such cells to be the losers in cell competition with normal cells. Similarly, co-culture of VPRBP-deficient MDCK cells with normal MDCK cells induced caspase-dependent apoptosis in 45% of the VPRBP-deficient cells, which were subsequently extruded from the apical surface, demonstrating that cell competition also occurs in mammalian cells.

> demonstrating that cell competition also occurs in mammalian cells.

So, how could cell competition function in tumorigenesis? Menéndez and colleagues showed that $l(2)gl^-$ cells that are induced in normal wing discs undergo JNK-dependent apoptosis. The overexpresssion of the oncogenic ras^{v12} mutant in $l(2)gl^-$ cells allowed the overgrowth of $l(2)gl^-$ cells in normal wing discs. Furthermore, the authors found that the Hippo pathway was suppressed, and that yorkie (YKI) — a transcription factor that is suppressed by the Hippo pathway — was consequently activated in these cells. Furthermore, the authors found that overexpression of *yki* in $l(2)gl^-$ cells also caused overgrowth in mosaic wing discs, suggesting that suppression of the Hippo pathway is sufficient to allow $l(2)gl^-$ cells to form tumours in wing discs.

It has been proposed that the losers in cell competition are determined by a slower growth rate than the competing cells. However, although the expression of ras^{v12} or *yki* increased the growth rate of the $l(2)gl^-$ cells, apoptosis was still apparent in the patches of growing $l(2)gl^-$ ras^{v12} or $l(2)gl^-$ *yki* cells at the regions adjacent to normal cells. Indeed, the authors found that more than half of the $l(2)gl^-$ ras^{v12} cells were eliminated from the mosaic wing discs, indicating that growth rate does not determine the losers. Further analyses suggest that peripheral apoptosis in patches of transformed cells is a general feature of cell competition and that cells evade apoptosis from competition with adjacent normal cells by growing together, which reduces the number of cells that are exposed to the surrounding normal cells.

Together these papers have further defined the interesting phenomenon of cell competition, and additional understanding of this process could identify possible targets for prevention and therapy.

Gemma K. Alderton

ORIGINAL RESEARCH PAPERS Tamori, Y. *et al.* Involvement of Lgl and Mahjong/VprBP in cell competition. *PLoS Biol.* **8**, e1000422 (2010) | Menéndez, J. *et al.* A tumor-suppressing mehanism in *Drosophila* involving cell competition and the Hippo pathway. *Proc. Natl Acad. Sci. USA* 2 Aug 2010 (doi:10.1073/pnas.1009376107)

CORBIS

© 2010 Macmillan Publishers Limited. All rights reserved

In the news

TOO MUCH COTTON WOOL?

Protection is a natural instinct and the loss of damaged cells, which undergo apoptosis to protect long-lived organisms from the development of diseases such as cancer, is often used as an example of this. However, two papers published in *Genes and Development* suggest that too many cells 'falling on their swords' can also lead to cancer.

Exposure to high levels of radiation can induce tumours through the induction of DNA damage. Under these conditions, the tumour suppressor p53 is activated to ensure the death of cells with damaged DNA. "Until now everyone believed that a failure of damaged cells to undergo suicide allowed mutated cells to proliferate … we discovered that in certain settings … the body's natural cell suicide programme can fuel tumour development.", stated Andreas Strasser, a lead author on one of the papers that shows that in mice lacking PUMA, an essential protein downstream of p53, radiation-induced thymic lymphoma does not occur. (*TopNews*, 3 Aug 2010).

As Andreas Villunger, the lead investigator of the second paper explains, the function of stem cells is essential here "…they have to expand and regenerate like crazy, making new blood cells to prevent the animal dying from anaemia." (*U.S.News*, 2 Aug 2010). This rapid expansion from a stem cell population that probably has low level mutations as a result of the radiation produces a large population of cells with mutations that could lead to cancer development. In the absence of PUMA, high levels of cell death are absent, so rapid proliferation of the stem cells is not required.

Gerard Zambetti, a researcher independent of these studies, described them as "…a keystone finding" adding, "I imagine that there will be a lot of follow up." (*Nature*, 31 Jul 2010).

Nicola McCarthy

© 2010 Macmillan Publishers Limited. All rights reserved

IN BRIEF

SMALL RNAS

OncomiR addiction in an *in vivo* model of microRNA-21-induced pre-B-cell lymphoma

Medina, P. *et al. Nature* 8 Aug 2010 (doi:10.1038/nature09284)

MicroRNA-21 (miR-21) is overexpressed in most human tumours. To further explore its role in tumour progression, Medina *et al.* have developed a conditional mouse model in which *miR-21* expression can be suppressed by doxycycline administration. Overexpression of miR-21 in the absence of doxycycline in the haematopoietic system led to the appearance of invasive lymphoma, indicating a role for miR-21 in cancer progression. When cells from these tumours were injected into immunodeficient mice, this resulted in the development of solid tumours, demonstrating that miR-21 is also involved in tumour initiation. Switching off *miR-21* expression by incorporating doxycycline into the diet of the mice caused the elimination of the tumour cells by apoptosis and the complete regression of the lymphomas in a few days. These results show the importance of the oncogenic role of miR-21 and suggest that its inhibition could be a new and promising option for cancer therapy.

BREAST CANCER

Rb deletion in mouse mammary progenitors induces luminal-B or basal-like/EMT tumor subtypes depending on p53 status

Jiang, Z. *et al. J. Clin. Invest.* 2 Aug 2010 (doi:10.1172/JCI41490)

The tumour suppressor genes *RB1* and *TP53* are both commonly lost in cancer. Eldad Zacksenhaus and collaborators have shown that *Rb1* deletion in mouse mammary progenitor cells induces changes in the epithelium that can progress to histologically diverse tumours with luminal B or triple-negative tumour (TNT; tumours that do not express oestrogen receptor, progesterone receptor or ERBB2) features. These TNTs further evolved into basal-like tumours or tumours that showed epithelial-to-mesenchymal transition (EMT) and were highly aggressive. Gene expression analysis revealed that EMT tumours and a subset of basal-like tumours had mutations in *Trp53*. Reintroducing expression of RB inhibited the growth of TNTs. These results establish key roles for RB in breast cancer progression and for p53 in determining tumour subtype after the loss of RB.

DNA DAMAGE

6-Thioguanine selectively kills BRCA2-defective tumors and overcomes PARP inhibitor resistance

Issaeva, N. *et al. Cancer Res.* **70**, 6268–6276 (2010)

Defects in homologous-recombination (HR) DNA repair — caused by mutations in *BRCA1* and *BRCA2*, for example — are associated with familial and sporadic breast and ovarian cancer. Cisplatin or inhibitors of poly(ADP-ribose) polymerase (PARP) selectively kill HR-defective tumours; however, resistance has already been observed in clinical trials, as cells partially restore HR. After a screen of drugs that selectively kill BRCA2-defective cells, Helleday and colleagues found that 6-thioguanine (6TG) selectively kills BRCA2-defective xenograft tumours as efficiently as PARP inhibitors. They also showed that 6TG treatment kills BRCA1- and BRCA2-defective tumours that are resistant to PARP inhibitors and cisplatin, suggesting that 6TG could be used for the treatment of advanced tumours.

© 2010 Macmillan Publishers Limited. All rights reserved

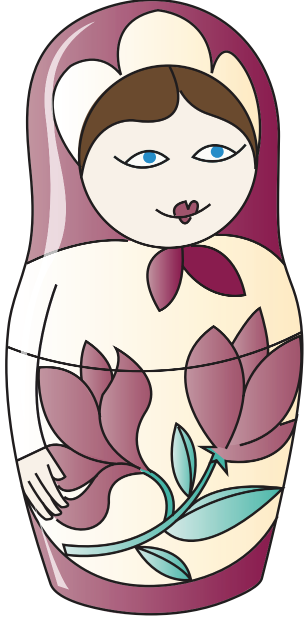
MACROENVIRONMENT

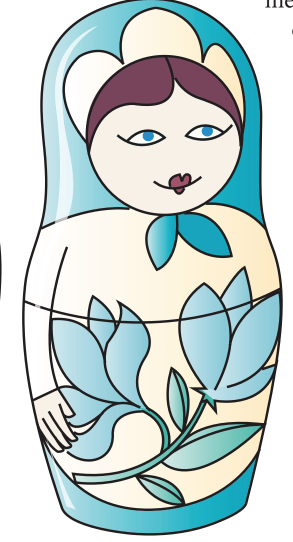
Stimulating resistance

A link between a positive environment and a better outcome for cancer patients has long been inferred, but it has been difficult to prove in the absence of concrete molecular pathways to underpin such an effect. A paper recently published in *Cell* may go some way to addressing this, as it found that mice living in an enriched environment (EE) are more resistant to tumour growth through a pathway involving brain-derived neurotrophic factor (BDNF), leptin and adiponectin.

An EE for mice consists of complex housing that contains sensory, cognitive, motor and social stimulation, as opposed to a standard cage. Lei Cao and colleagues found that the growth rate of transplanted malignant melanoma cells in mice housed in an EE was significantly reduced compared with mice housed in standard (control) conditions. Mice in an EE weighed less than control mice despite identical diets, prompting the authors to examine systemic metabolic changes. Levels of insulin-like growth factor 1 and leptin were reduced in mice housed in EE cages, and serum from these mice reduced the growth of melanoma cells *in vitro*. Increased levels of both of these factors have been associated with an increased risk of cancer development and progression, as they can stimulate the growth of cancer cells. Conversely, levels of adiponectin were increased.

Leptin is a hormone that relays information back to the hypothalamus, which is involved in regulating energy balance along with the production of neuroendocrine factors and the regulation of the immune system through the hypothalamic–pituitary–adrenal axis. Expression of BDNF by the hypothalamus was increased early on when mice were placed in an EE. Overexpression of BDNF through the infection of mice with viruses expressing this gene had a similar effect on melanoma growth to housing mice in an EE. In addition, the effect of an EE on melanoma growth was lost in mice in which *Bdnf* levels were reduced through microRNA-mediated knock down or *Bdnf* heterozygosity.





How does BDNF influence the levels of leptin and adiponectin? Both leptin and adiponectin are predominantly synthesized by white adipose tissue, and suppression of leptin expression is thought to be influenced by the baseline activity of the sympathetic nervous system affecting the expression of β-adrenergic receptors (β-ARs). The expression of β-ARs and one of their ligands, noradrenaline, was increased in white adipose tissue from mice housed in EE cages. Use of a β-blocker prevented the EE-mediated changes in leptin and adiponectin levels as well as the reduced growth of the transplanted melanoma cells. Moreover, leptin-deficient (*ob/ob*) mice did not show reduced melanoma growth when housed in an EE.

These results were not restricted to melanoma; injection of mice with colon cancer cells, or transgenic mice that spontaneously develop colon cancer (*Apc*$^{Min/+}$ mice), had reduced tumour burden in an EE compared with mice housed in control conditions. Moreover, exposure of mice with established tumours to an EE led to prolonged survival.

Overall, these data indicate that exposure to an EE induces increased expression of BDNF in the hypothalamus, which induces activation of the sympathetic nervous system resulting in the production of noradrenaline and increased expression of β-ARs in white adipose tissue, which suppresses leptin and increases adiponectin levels. This, combined with an increase in the immune response in mice housed in an EE, suggests that exposure to 'positive stress' reduces the growth of tumours in mice.

Nicola McCarthy

ORIGINAL RESEARCH PAPER Cao, L. *et al.* Environmental and genetic activation of a brain-adipocyte BDNF/leptin axis causes cancer remission and inhibition. *Cell* **142**, 52–64 (2010)

© 2010 Macmillan Publishers Limited. All rights reserved

IN BRIEF

BREAST CANCER

Ferroportin and iron regulation in breast cancer progression and prognosis

Pinnix, Z. K. *et al. Science Trans. Med.* **2**, 43ra56 (2010)

Some cancers have been shown to have an increased requirement for iron, and a new study provides insights into the potential underlying mechanism by showing that the ferroportin iron transport system is important in breast cancer progression and prognosis. The authors showed that ferroportin and the hormone hepcidin, which regulates ferroportin stability, are expressed in cultured breast epithelial cells. However, compared with these cells, ferroportin expression was reduced in breast cancer cell lines, and this correlated with increased iron availability. Breast cancer cells transfected with ferroportin showed reduced growth in a xenograft mouse model. In addition, human breast tumour samples showed reduced ferroportin expression, which correlated with increased anaplasia, and increased ferroportin and decreased hepcidin expression predicted a favourable prognosis in patients with breast cancer, showing that the *in vitro* and mouse results can be extended to human breast cancer.

THERAPY

NF1 is a tumor suppressor in neuroblastoma that determines retinoic acid response and disease outcome

Hölzel, M. *et al. Cell* **142**, 218–229 (2010)

Retinoic acid therapy of neuroblastoma shows a variable clinical response, but the reasons for this have been unknown. Using a large-scale RNA interference (RNAi) screen, Bernards and colleagues have now identified crosstalk between the tumour suppressor *NF1* and the effects of retinoic acid on differentiation in neuroblastomas. They show that loss of NF1 in neuroblastoma activates Ras–MEK signalling, which represses ZNF423, a key transcriptional regulator of retinoic acid receptors. Mutations in *NF1* occur in primary neuroblastomas, and low levels of *NF1* and *ZNF423* in neuroblastoma are associated with a poor outcome. Treatment of *NF1*-deficient neuroblastoma cells with MEK inhibitors restores responsiveness to retinoic acid, suggesting that this might be a viable combination therapy for retinoic acid-resistant neuroblastomas that are deficient for *NF1*.

DNA DAMAGE

DNA damage signaling in response to double-strand breaks during mitosis

Giunta, S. *et al. J. Cell Biol.* **190**, 197–207 (2010)

The activation of DNA damage response signalling pathways by DNA double-strand breaks (DSBs) has been well characterized in interphase cells and leads to activation of cell cycle checkpoint arrest and the appropriate DNA repair pathway(s). However, DSBs that occur in mitotic cells do not induce checkpoint activation or repair, but is DNA damage response signalling activated? Jackson and colleagues showed that DSBs induced in mitotic cells activate upstream (apical) DSB signalling, including phosphorylation of histone H2AX at DSBs. Inactivation of the apical kinases ATM and DNA-PK hypersensitized cells to drugs that induce DSBs, suggesting that the activation of the apical pathway marks damage to be repaired in the following G1 phase.

© 2010 Macmillan Publishers Limited. All rights reserved

Antibody-based proteomics: fast-tracking molecular diagnostics in oncology

Donal J. Brennan§, Darran P. O'Connor*§, Elton Rexhepaj*, Fredrik Ponten‡ and William M. Gallagher**

Abstract | The effective implementation of personalized cancer therapeutic regimens depends on the successful identification and translation of informative biomarkers to aid clinical decision making. Antibody-based proteomics occupies a pivotal space in the cancer biomarker discovery and validation pipeline, facilitating the high-throughput evaluation of candidate markers. Although the clinical utility of these emerging technologies remains to be established, the traditional use of antibodies as affinity reagents in clinical diagnostic and predictive assays suggests that the rapid translation of such approaches is an achievable goal. Furthermore, in combination with, or as alternatives to, genomic and transcriptomic methods for patient stratification, antibody-based proteomics approaches offer the promise of additional insight into cancer disease states. In this Review, we discuss the current status of antibody-based proteomics and its contribution to the development of new assays that are crucial for the realization of individualized cancer therapy.

*UCD School of Biomolecular and Biomedical Science, UCD Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland.
‡Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, SE-751 85 Uppsala, Sweden.
Correspondence to W.M.G. e-mail: william.gallagher@ucd.ie
§These authors contributed equally to this work.
doi:10.1038/nrc2902
Published online 19 August 2010

Personalized medicine requires the discovery and application of unambiguous prognostic, predictive and pharmacodynamic biomarkers to inform therapeutic decisions. High-throughput screening methods, particularly genomic and transcriptomic profiling, have vastly improved knowledge of the molecular basis of tumorigenesis, disease progression and therapeutic response[1,2]. As a result, individualized treatment regimens are now seen as an achievable goal. The role of antibodies in this arena is most likely to involve predictive biomarker development, as highlighted by the success of detecting both oestrogen receptor (ER) and ERBB2 (also known as HER2) expression in breast cancer. In the post-genomic era, however, progress in the development of clinically implemented assays has not kept pace with the rate of biomarker discovery[3]. As such, a pressing need exists for improved and innovative strategies to expedite the translation of cancer biomarkers into the clinical arena.

Antibody-based proteomics provides a logical strategy for the systematic generation and use of specific antibodies to explore the proteome[4,5]. The human proteome consists of approximately 20,500 non-redundant proteins[6], which are defined as a representative isoform from each gene locus[5]. Unlike DNA, which is subject to one major form of modification (methylation), proteins can be post-translationally altered in a myriad of ways; for example, by phosphorylation, acetylation and glycosylation — each of which is capable of producing a functional shift that potentially affects disease development, progression and therapeutic response. Despite this instrinic complexity, endeavours to describe the proteome and provide a comprehensive map of protein expression patterns in cells and tissues, such as the Human Protein Atlas (see the Human Protein Atlas (HPA) website; Further information) (BOX 1) and the complementary resource, Clinical Proteomic Technologies for Cancer (CPTAC; see the CPTAC website; Further information), are ongoing[7]. The improved understanding of antibody–antigen relationships provided by such efforts[8] will greatly expedite the development of new assays using affinity reagents to profile cancer proteomes and enable the exploitation of the specificity and sensitivity that is afforded by antibody-based approaches.

In this Review, we describe current antibody-based methods to identify and validate new cancer biomarkers and therapeutic targets, and discuss the systematic generation and evaluation of specific antibodies for the functional exploration of the cancer proteome. The integration of antibody-based approaches with existing

© 2010 Macmillan Publishers Limited. All rights reserved

At a glance

- Personalization of cancer therapy requires the identification of unambiguous diagnostic, prognostic and predictive biomarkers to facilitate the accurate stratification of patients and the monitoring of responses to targeted therapies.
- The systematic generation and validation of specific antibodies offers a high-throughput mechanism for the functional exploration of the proteome and a logical approach for fast-tracking the translation of identified biomarkers.
- Multiple approaches exist, each with specific characteristics and advantages that are suitable for a wide range of applications, which capitalize on the inherent specificity and sensitivity of antibodies as affinity reagents.
- The integration of antibody-based approaches with existing genomic and transcriptomic methods offers huge potential, and the clinical implementation of new high-throughput antibody-based approaches will depend on the integration of data across various platforms.
- The clinical application of new antibody-based assays demonstrates their utility as accurate, sensitive and robust diagnostic and prognostic tests and has led to the development of a new approach, known as pathway diagnostics, which is likely to have a crucial role in the design of future molecular therapeutic trials.

genomic and transcriptomic methods is considered and the clinical implementation of new high-throughput antibody-based approaches is examined, particularly in the context of enabling the personalization of cancer therapy.

Enabling technologies

The generation and use of antibodies for protein profiling on a global scale is an intuitive approach that enables the systematic examination of the human proteome using a wide range of high-throughput assays, including immunohistochemistry (IHC) on tissue microarrays (TMAs), pathway analysis using reverse phase protein arrays (RPPAs) and serum-based diagnostic assays using antibody arrays (BOX 2). Founded on the ability to systematically generate and validate specific and sensitive antibodies (BOX 3), these versatile assays are at the forefront of efforts to generate the molecular diagnostic and predictive assays that are required to facilitate the personalization of cancer therapy[9,10].

Antibody specificity is the foundation of antibody-based proteomics, and although the ideal approach to confirm specificity is the high-throughput production of paired antibodies directed towards separate and non-overlapping target protein epitopes (BOX 3), alternative approaches are also required. Several laboratories combine western blotting and IHC on identical cell lines (ideally using a non-expressing cell line as a negative control[11,12]) that are formatted as cell line microarrays to facilitate high-throughput validation when used in tandem with automated image analysis solutions. If a negative control cell line is not available, small interfering RNA (siRNA)-mediated knock down followed by both western blotting and IHC can also be used[11], and staining reproducibility should be confirmed using a TMA that is stained with each new antibody[12].

Reverse phase protein array
RPPA. Protein lysate dot blot in a high-density format on a solid surface that allows for multiple samples to be probed with the same antibody, or other affinity reagent, simultaneously.

Automated analysis of protein expression in tissue. TMAs developed by Kononen *et al.* are a high-throughput platform for the simultaneous investigation of protein expression in multiple tissue specimens, principally using IHC[13] (BOX 2). TMAs were preceded by a related technology known as a 'sausage block', which was constructed by the assembly of larger tissue fragments in a recipient block in a less organized manner[14], an approach that was subsequently modified by Wan *et al.*[15].

Over the past decade, TMAs have become an established and crucial component of the cancer biomarker discovery and validation pipeline[16,17]. Although TMAs have undoubtedly enabled the acceleration of translational pathology, new demands have been placed on the quality, reproducibility and accuracy of IHC assays. Variability in tissue collection, fixation and processing, antigen retrieval, titration of the antibody, application of secondary antibodies and multiple detection systems results in a multi-parameter assay that requires thorough optimization[18].

Historically, IHC assay development has been hampered by the lack of specific antibodies; however, the development of comprehensive antibody resources (BOX 1) promises to help overcome this obstacle. Traditional IHC, although ubiquitous in both clinical and research settings, has been criticized as a semi-quantitative approach. In particular, attention has centred on the intrinsic lack of reproducibility of manual IHC scoring, which remains a time-consuming and subjective process to which only limited statistical confidence can be assigned owing to inherent inter-observer and intra-observer variability and the semi-quantitative nature of the data[19,20]. For example, a large study of inter-laboratory variance in the IHC-based detection of ER in breast cancer samples across 200 laboratories in 26 countries demonstrated a false-negative rate of 30–60%[21]. As ER is routinely used to determine the need for adjuvant hormonal therapy in patients with breast cancer, this level of discordance is all the more alarming and highlights the urgent need for the standardization of molecular diagnostic approaches.

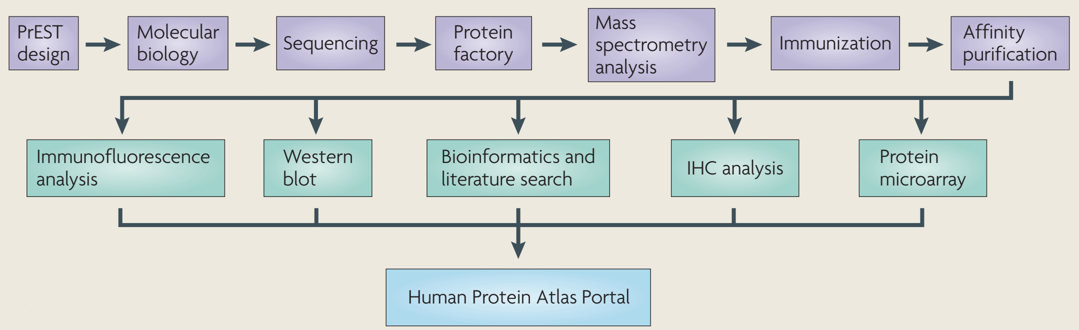
Automated IHC scoring systems offer the opportunity to further advance a well-established and clinically useful assay to accurately quantify both staining intensity and the subcellular localization of protein expression in a reproducible fashion[22,23]. In addition, automated analytical approaches provide quantitative data that can be subjected to more robust statistical analysis than the qualitative or semi-quantitative data that are produced from manual analysis[22,24]. Various automated image analysis solutions are currently available, some of which have received US Food and Drug Administration clearance for cancer-specific biomarker applications (regulatory issues are outlined in BOX 4). Most platforms comprise both image analysis software and scanning hardware[23,25], and focus on the quantification and subcellular localization (cytoplasmic, nuclear or membranous) of the relevant antigen (FIGS 1a,b). Several groups, including our own, have demonstrated that automated algorithms can be used to accurately quantify IHC staining in cell lines[26] and have also linked automated analysis of IHC to clinical outcome in a selection of different tumour types[22,24,27–30]. Although various automated image analysis approaches have been approved at a regulatory level, as indicated above, issues still remain regarding the integration of image analysis technologies into routine clinical practice.

© 2010 Macmillan Publishers Limited. All rights reserved

Box 1 | The Human Protein Atlas Program: a multidisciplinary antibody-based proteomics initiative

The Human Protein Atlas (HPA) is a gene-centric database[126] using the human genome sequence (Ensembl) as a template to select coding sequences corresponding to 50–150 amino acids, denoted Protein Epitope Signature Tags (PrESTs; see the figure). Based on the selection of regions with low similarity to other human genes, up to four different PrESTs are defined for each gene to enhance the probability of generating unique, specific antibodies[127]. Recombinant PrEST protein fragments are produced and then used as an antigen to develop polyclonal antibodies, which are affinity purified to generate unique, oligoclonal monospecific antibodies[128]. Binding specificity is tested on protein arrays containing various PrESTs, and all approved monospecific antibodies are subsequently used for western blotting using a standardized protocol. Protein expression patterns are visualized using immunohistochemistry (IHC) on tissue microarrays, representing 48 types of normal tissues and 216 human tumours corresponding to the 20 most common forms of human cancer and 47 cell lines[129,130]. High-resolution images are acquired from all immunostained tissue and cell microarray sections. All histology images are manually annotated and curated by certified pathologists, and cell images are analysed using image analysis-based algorithms[26]. Furthermore, all antibodies are applied to three human cancer cell lines using immunofluorescence and confocal microscopy to determine the subcellular localization of each protein[131].

Of the 20 new antibodies that are processed every day approximately 50% are approved for IHC. Therefore, protein expression data are generated for approximately 3,000 new antibodies per year. Specific emphasis is placed on the confirmation of antibody specificity, which is a key issue for any antibody-based proteomics initiative (BOX 3). All protein expression data, including images, antibody validation and immunized PrEST sequences, are published on the HPA web portal (see Further information), to provide a knowledge base for functional studies and biomarker discovery efforts. The portal contains data and images from both in-house generated monospecific antibodies and external, commercially available antibodies. To allow for searches and queries regarding protein expression profiles in normal tissues and cancer, a web-based analysis tool is also available[132]. The database is updated annually and the current release (version 6.0) contains more than 9 million images and protein profiles based on greater than 11,200 antibodies directed towards 8,489 unique proteins corresponding to 42% of all human protein-encoding genes.



A key potential advantage of such approaches is the provision of an unbiased and cost-effective method of data standardization, which is often a difficult issue with current pathological assessment. The technology itself is perhaps better seen as a complement to routine pathological assessment rather than as a replacement.

Although IHC is a well-validated, clinically applicable assay, immunofluorescence provides a complementary approach (FIG. 1c), particularly for the identification and quantification of co-localized proteins. Immunofluorescence also demonstrates a greater dynamic range than colorimetric IHC-based detection. Given the ongoing trend towards multiplex biomarker assay development, immunofluorescence may gain greater clinical use as a routine diagnostic and prognostic assay. In this context, immunofluorescence quantification is supported by various digital pathology solutions, allowing for the automated definition of regions of interest, cellular compartmentalization and fluorescent signal co-localization.

In translational medicine, the fluorescent-based AQUA (HistoRx) platform is one of the more established automated solutions for cancer biomarker assessment[31]. However, despite a large body of literature describing the approach, doubts remain about its reproducibility and clinical applicability, principally owing to a lack of independent validation studies. The AQUA platform identifies tumour cells using cytokeratin expression, and so creates a region of interest that allows for the definition of subcellular compartments and the accurate quantification of protein expression in formalin-fixed paraffin-embedded (FFPE) tissue samples. Several studies have demonstrated that AQUA can measure protein expression on histological specimens from various tumour types with good accuracy and reproducibility, and this can then be linked to clinical outcome[32–36]. As such, both colorimetric and fluorescent-based detection approaches, coupled with associated automated image analysis solutions, are viable, complementary techniques to accurately and reproducibly evaluate tissue-based cancer biomarkers.

Antibody arrays. Multiplex protein analysis, particularly of the serum proteome, offers great promise for the development of less invasive and more cost-effective diagnostic assays. However, the complexity of the serum proteome, which contains approximately 10,000 proteins with a dynamic range of at least 10 orders of magnitude, presents several technical challenges[37]. The serum

© 2010 Macmillan Publishers Limited. All rights reserved

Two-dimensional electrophoresis
2DE. Gel-based technique for the separation of proteins by isoelectric point in the first dimension (achieved by isoelectric focusing), followed by mass in the second dimension (achieved by SDS–PAGE). A higher resolution of protein separation is achieved compared with single dimension approaches.

Multi-dimensional liquid chromatography
Chomatographic separation in at least two dimensions, for example, reverse-phase chromatography followed by ion-exchange chromatography. Using additional dimensions increases the resolution of separation.

Tandem mass spectrometry
Often referred to as MS/MS, it uses two linked mass spectrometers to measure small amounts of proteins. Analytes are separated according to their mass and charge, with samples sorted and weighed in the first mass spectrometer, then fragmented in a collision cell, and fragments sorted and weighed in the second mass spectrometer.

proteome is dominated by 22 high-abundance proteins that constitute 99% of the total protein mass of serum (for example, albumin, immunoglobulins and transferrin) and effectively mask lower abundance proteins. This is one of the principle obstacles to the identification of new biomarkers, which are frequently present at low concentrations (pg per ml or lower compared with upper limits of mg per ml for high-abundance proteins)[37,38].

Several approaches have been used to examine the serum proteome, including classic separation techniques, such as two-dimensional electrophoresis (2DE) and multi-dimensional liquid chromatography, coupled with single or tandem mass spectrometry (MS/MS)[37–39]. Although these approaches have certain advantages, antibody arrays and RPPAs (BOX 2) have emerged as versatile platforms in serum-based proteomics, offering the opportunity to carry out multiplexed, rapid and sensitive (fM range) profiling of samples (reviewed in REFS 40–42) without fractionation or the depletion of high-abundance proteins. Multiple formats exist, facilitating multiplexed analysis of samples with differential labelling. Further modifications involve the selective identification of post-translational modifications, such as the use of lectin-antibody arrays to profile glycan variation on cancer-associated antigens (reviewed in REFS 43,44).

ELISA. Enzyme-linked immunosorbent assay (ELISA) remains the gold standard for measuring protein concentration in human body fluids, particularly blood. Such an approach allows for the accurate and sensitive detection of the antigen of interest; however, it is limited by the fact that classic approaches allow for single antigen detection only and often require relatively large volumes of sample material compared with newer methods.

Multiplex assays have been developed from traditional ELISA assays to quantify multiple antigens in a single sample simultaneously (BOX 2). Multiple proteins in a biological fluid sample can therefore be measured, and although commercial kits are limited to approximately 25 capture antibodies, custom design approaches could potentially scale up to 100 (REF. 45).

Variations on the ELISA approach, such as the Meso-Scale Discovery (MSD) platform that uses electrochemiluminescent detection on patterned arrays to quantify multiplexed biomarkers, offer further promise for patient stratification and monitoring of therapeutic responses. Preclinical studies examining the effect of the PI3K inhibitor LY294002 on human xenografts demonstrated the effectiveness of using multiplexed MSD assays to measure pharmacodynamic responses to accurately monitor the provision of therapy[46].

Compared with traditional ELISA, multiplex arrays have several advantages, including their high-throughput nature, requirement for smaller sample volume, the ability to evaluate one antigen in the context of multiple others and the ability to reliably detect different proteins across a broad dynamic range[47,48]. Although good correlations between ELISA and multiplex assays have been

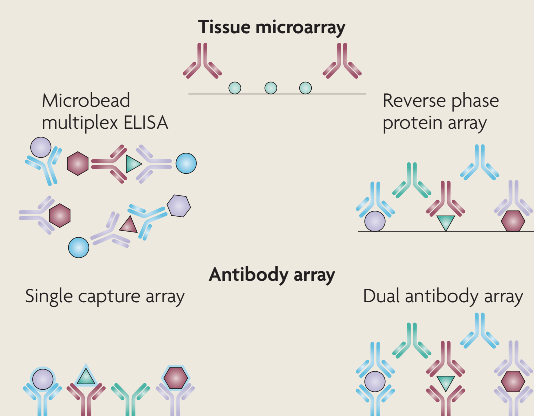
Box 2 | Antibody-based proteomics technology formats

Developed by Kononen *et al.*[13], tissue microarrays (TMAs) are a high-throughput platform for the simultaneous investigation of biomarkers in multiple tissue specimens. TMAs are assembled by acquiring cylindrical cores (0.6–2.0 mm in diameter) from donor paraffin-embedded tissues and re-embedding them in a single recipient block. The resultant TMA block is then sectioned, and immunohistochemistry and other assays such as immunoblotting[133] are carried out on individual sections (see the figure).

In the typical double antibody sandwich enzyme-linked immunosorbent assay (ELISA), antibody attached to the bottom of a well provides both antigen capture and immune specificity, and another antibody linked to an enzyme provides detection and an amplification factor. Several different multiplex formats are available. Bead-based multiplex assays are probably the most commonly used format[47,48]. Each bead set is coated with a specific capture antibody, and fluorescence- or streptavidin-labelled detection antibodies bind to the specific capture antibody complex on the bead set, which can be detected using flow cytometry.

Reverse phase protein array (RPPA) technology is a microproteomic approach in which protein lysate is immobilized on a solid surface and subsequently probed with antibodies. RPPAs allow the examination of the activation state of crucial cellular pathways using antibodies directed against total and phosphorylated protein[134,135]. In contrast to TMA or antibody array-based methods, RPPAs can use denatured protein lysates, thus removing the need for antigen retrieval, and the use of non-denatured protein lysates allows for protein–protein or protein–DNA interactions to be probed[135].

Antibody arrays are produced by printing antibodies onto a solid surface that is analogous to a DNA microarray. Two categories of antibody microarray formats have been described, namely direct labelling single-capture antibody arrays[136] and dual antibody (capture and read-out antibody) sandwich arrays[137]. In the direct labelling method, all proteins in a sample are tagged, so that bound proteins can be identified following incubation on a microarray. In the dual antibody format, proteins captured on the microarray are detected using a cocktail of labelled detection antibodies, with each antibody matched to one of the spotted antibodies[41].



© 2010 Macmillan Publishers Limited. All rights reserved

Box 3 | **Antibody specificity**

A key issue for all antibody-based assays is the confirmation of antibody specificity towards its antigen. Validation approaches include not only generic binding assays such as protein arrays, which use the antigen as a ligand for the assay, but also various kinetic binding assays that establish affinity and determine binding sites using epitope mapping[138]. Epitope mapping using X-ray crystallography or NMR, which can determine the three-dimensional structure of the binding complex, can be comprehensive approaches but are laborious. Other approaches for epitope mapping include scanning with peptides[139], which are either chemically synthesized or expressed on the surface of microorganisms[140,141]; however, these approaches can also be cumbersome.

One of the major challenges in generating reliable antibodies is high-throughput validation of protein-specific binding in different antibody-based assays[9,10]. This becomes particularly important when generating antibodies to proteins lacking independent experimental validation. Western blotting is often regarded as a gold standard for antibody specificity; however, post-translational modifications can make the interpretation of experimental results ambiguous. Likewise, antibodies that function well in western blotting using denatured proteins might not function in another assay, such as immunohistochemistry or immunofluorescence, in which proteins retain a degree of native conformation[142,143]. Validation of antibodies, therefore, remains a challenge, in particular for antibodies directed towards uncharacterized proteins.

The ideal approach to confirming antibody specificity is the high-throughput production of paired antibodies directed towards separate and non-overlapping target protein epitopes to allow sandwich-based assays and to facilitate the validation of the affinity reagents across various assay platforms, including immunofluorescence, immunohistochemistry and western blotting[138].

reported[49,50], experience with multiplex arrays remains limited and careful side-by-side comparisons are rare. Although concordance between ELISA and the multiplex assay is generally good when using tissue culture supernatant samples, it is much less robust when using serum or plasma samples[51]. Therefore, further direct comparisons in the clinical trial setting are warranted.

RPPAs. Cell signalling pathways are regulated in many instances by the post-translational modification of proteins, and in-depth analysis of deregulated cellular circuitry in cancer requires specialized technologies. As a result, innovative high-throughput proteomic approaches such as RPPAs (BOX 2) have been developed to examine pathway activation using phospho-specific antibodies in large panels of patient samples simultaneously[52]. One of the key advantages of RPPAs is the discrete amounts of patient material required, traditionally using frozen specimens, but more recently successfully demonstrated using FFPE samples[53]. This approach is dependent on the standardization of tissue collection to preserve the state of post-translational modifications[54]. In this context, pre-analytical fluctuations in phosphoproteins have been noted, reflecting the dynamic nature of kinase and phosphatase activity in excised tissue[54]; this reiterates the requirement for standardized sample procurement procedures to allow accurate inter-institutional data comparison.

Additional factors influencing the use of RPPAs focus on the sensitivity and dynamic range of detection of the assay. Traditional colorimetric and fluorimetric detection strategies offer excellent sensitivity, although they are hampered by poor dynamic range, requiring the serial dilution of samples. However, the sensitivity of the assay can be extended to the aM and zM range using electro-chemiluminescent[55] and evanescent field fluorescent detection[56] with a dynamic range of up to 5–6 logs. Similar dynamic ranges and sensitivity can also be achieved using near infrared-based detection, which removes the need for specialized, proprietary detection systems[57].

Clinical applications

In the past decade, vast amounts of data accrued from various molecular profiling platforms have facilitated a shift away from conventional broad therapeutic approaches to cancer, towards more tailored strategies. The expansion of personalized treatment protocols now depends on the development of robust, well-validated, informative predictive and pharmacodynamic assays. ER and ERRB2 were instrumental as early examples of predictive biomarkers in breast cancer and epitomize personalized medicine; however, more recently, transcriptomic approaches have led the way in the advancement towards individualized therapeutic protocols[2,58]. Although the ongoing trials will determine the clinical applicability of these gene expression assays, other reservations have been expressed regarding their prohibitive cost, reliance on frozen tissue and the advanced technical expertise required to use the technology[59].

The most clinically advanced gene expression signatures are MammaPrint[60,61] and OncotypeDx[62], which are currently the subject of large-scale prospective randomized control trials to assess their utility for the stratification of patients with breast cancer to determine the appropriate treatment approach[2,63,64]. Although several DNA microarray and reverse transcription (RT)-PCR-based prognostic and predictive assays have been proposed in other tumour types, such as lung[65–67], colon[68–70] and prostate[71,72] cancer, and lymphoma[73–75], these seem to be further from clinical application at present, mainly owing to a paucity of comprehensive validation studies. Here, we discuss clinically applicable antibody-based assays for diagnostics, incorporating serum- and tissue-based assays, prognostics, using IHC-based signatures, and molecular therapeutics, with a particular emphasis on pathway diagnostics using tissue-based assays (FIG. 2).

Diagnostics. Although tissue-based diagnosis ultimately remains the remit of the histopathologist, in the absence of screening programmes, a considerable number of tumours are diagnosed at an advanced stage, and so efforts have focused on developing more sensitive diagnostic assays. Such an approach should have a significant impact on cancer-related mortality rates.

Serum-based proteomics offer great hope, particularly in the development of more sensitive diagnostic assays. The current gold standard for validating putative biomarkers is ELISA. A high-throughput ELISA has extraordinary sensitivity and specificity for quantifying target analytes. However, ELISA development is costly (typically, US$100,000–US$2 million per biomarker candidate) and is associated with a long development lead time (>1 year) and a high failure rate[76], although the

Epitope mapping
Systematic identification and characterization of the minimum recognition domain for antibodies.

Sandwich-based assay
Antigen detection using surface-bound capture antibodies, followed by the application of the sample and subsequent detection using a second antibody raised against an alternative epitope on the same target protein.

© 2010 Macmillan Publishers Limited. All rights reserved

multiplexed approaches discussed above are likely to bring ELISA back to the forefront of biomarker identification and validation.

An important focus of serum-based proteomics has been placed on MS-based approaches; however, antibody-based approaches are viable alternatives and, in combination with MS-based techniques, form a highly complementary strategy for cancer biomarker enrichment to push MS-based detection into the ranges commonly achieved by ELISA. Given the long lead-time and high cost of ELISA development, an assay cannot be developed for every putative biomarker, and more affordable technologies with a shorter lead-time are required for biomarker validation. To this end, the combination of peptide enrichment with antibodies that are immobilised on affinity columns and MS offers great promise, particularly as antibody-based proteomic resources generate larger numbers of validated antibodies. This technology, known as stable isotope standards with capture by anti-peptide antibodies (SISCAPA)[77], has been used to quantify proteins in the physiologically relevant range (ng per ml)[76] and has recently been implemented in an automated multiplex (nine targets in one assay) format.

Serum-based proteomic screens can also be carried out using high-throughput antibody-based platforms, such as antibody arrays and RPPAs, both of which offer the opportunity of reduction to clinical utility in a timely fashion. Antibody arrays have been used to develop potential diagnostic assays for several different tumour types, although the focus has generally been on tumours with an insidious onset that are often diagnosed at a late stage, such as pancreatic cancer. Two groups have published serum signatures developed from antibody arrays, which seem to distinguish patients with pancreatic cancer from healthy controls[78,79]. Although such an assay would be hugely beneficial to the clinical arena, these studies have been hampered by low-density antibody arrays and small cohorts. Additionally, the simple comparison of patients with malignant pancreatic cancer to normal controls is probably an overly simplistic approach, and comparison between normal controls, those with a premalignant condition (such as chronic pancreatitis) and invasive disease may unearth more informative data.

The use of serum proteomics to monitor patients to detect early local recurrence or metastatic deposits and the application of antibody arrays to identify metastatic breast cancer[80] further illustrates the clinical potential of these approaches. The development of higher density antibody arrays, such as the recently published 810 'cancer related' antibody array, may also considerably advance this technology in the translational arena[81].

In addition to serum-based assays, tissue proteomics offer the opportunity to develop IHC-based assays to improve diagnostic sensitivity and disease classification. Kashani-Sabet *et al.*[82] combined gene expression data and IHC profiling using automated analysis to develop a highly sensitive (91%) and specific (95%) five marker diagnostic assay (comprised of actin-related protein 2/3 complex, subunit 2 (ARPC2), fibronectin 1 (FN1), regulator of G protein signalling 1 (RGS1), secreted phosphoprotein 1 (SPP1; also known as osteopontin) and WNT2) that distinguishes benign nevi from melanoma. The same group used a similar approach to develop a three marker prognostic assay (comprised of nuclear receptor coactivator 3 (NCOA3), SPP1 and RGS1) for melanoma[82]. This assay, initially developed in a cohort of 395 patients, and subsequently validated in an independent cohort of 141 patients, was an independent predictor of disease-specific survival in both cohorts. The integration of gene expression analysis and high-throughput IHC profiling using automated analysis offers great potential for the development of similar assays in other tumour types, particularly as more advanced mathematical models can be applied to quantitative automated IHC data.

Box 4 | **Regulatory issues pertaining to quantitative IHC and immunofluorescence analysis**

Within the US market, digital pathology and automated analysis algorithms for clinical use are regulated by the Center for Devices and Radiological Health of the Food and Drug Administration (FDA). There are two major processes by which medical devices come to the US market, namely FDA Clearance (sometimes referred to as 510(k) process) and FDA Approval. There are three FDA regulatory classifications of medical devices: Class I, Class II and Class III.

Class I medical devices present minimal potential harm to the user and are generally exempt from the pre-market notification process. Class II medical devices are devices for which existing methods, standards and guidance documents are available to provide assurances of safety and effectiveness. Class II devices typically require pre-market notification through the FDA Clearance process. Class III devices usually support or sustain human life, are of substantial importance in preventing the impairment of human health, and present a potential unreasonable risk of illness or injury to the patient. Typically, a Pre-Market Approval (PMA) submission to the FDA is required to allow marketing of a Class III medical device.

Currently, the FDA have classified whole slide imaging systems for viewing IHC within clinical context, as well as automated algorithms for IHC analysis, as Class II devices that require pre-market notification through FDA Clearance. Pre-market notification requires a new device to be compared for safety and effectiveness with another lawfully marketed model. In the case of an image analysis solution for a new biomarker, this would require a substantial equivalence study based on comparison of image analysis to conventional manual microscopy. In the case of a new approach to quantification of a marker such as ERBB2 (also known as HER2), for which several algorithms have been cleared, an equivalence study based on comparison to FDA-cleared algorithms would also be required.

Regulation of digital pathology and image analysis for primary diagnosis remains unclear, and the FDA has not yet decided whether a diagnostic application would be classified as a Class II or Class III device. This decision will have a major influence on the industry as the FDA charges a company with more than US$100 million in sales $4,000 to review a FDA Clearance application compared with $217,787 to review a PMA submission.

© 2010 Macmillan Publishers Limited. All rights reserved

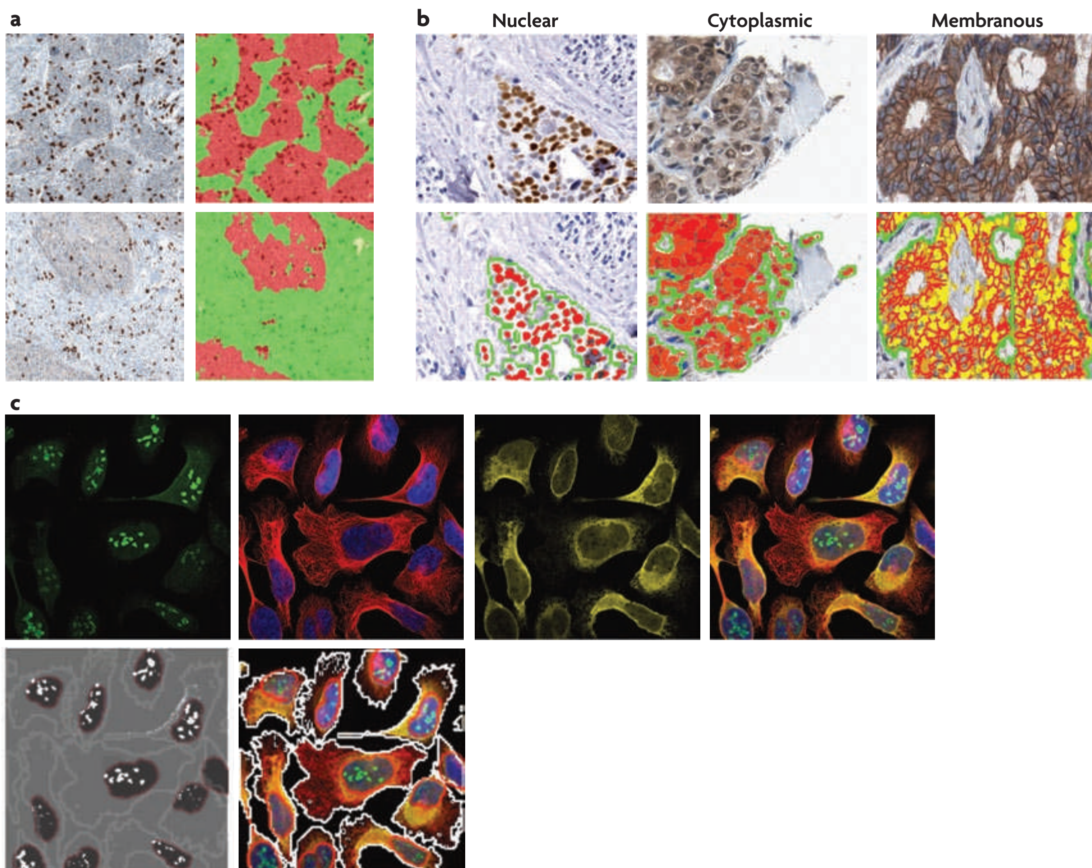


Figure 1 | **Automated quantification of protein expression using immunohistochemistry and immunofluorescence.** **a** | One of the major challenges facing immunohistochemistry (IHC) automated analysis software is distinguishing tumour epithelium from stromal cells, such as fibroblasts and lymphocytes. Advanced pattern recognition software can be used in a supervised manner to train representative images of different tissue patterns such as tumour and stroma, and once trained the algorithm can automatically distinguish between the two compartments. In this example, Genie (Aperio) was trained to distinguish between tumour epithelial cells (shown in red) and stroma (shown in green). **b** | Other automated IHC quantification packages can be used to quantify protein expression in different subcellular compartments in tissues (that is, the membrane, cytoplasm and nuclear compartments). These packages, such as the example provided here through the use of *IHC-MARK* (patent pending; Oncomark Ltd), provide quantitative data, such as positive and negative cell counts and staining intensity. **c** | Alternatively, immunofluorescence microscopy can provide accurate information about subcellular localization in terms of antigen expression, as demonstrated by Ki67 staining of tumour cells (green), the nucleus (blue), microtubules (red) and endoplasmic reticulum (yellow). Immunofluorescence-based image analysis packages, such as Definiens Tissue Recognition technology, can be used to subtract the subcellular compartment signal and provide a read out of the overall staining for the total number of cells (dark black shaded areas in final output), the number of nucleoli (white spots in final output), as well as the boundaries of each subcellular area.

This approach was recently highlighted by Gould Rothberg *et al.*[36] who used a fully automated quantitative immunofluorescence approach combined with genetic algorithms to identify a five marker prognostic assay for melanoma.

In another example, Ring *et al.*[83] recently described a five marker diagnostic assay (comprised of tripartite motif-containing 29 (TRIM29), carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5), SLC7A5, mucin 1 (MUC1), cytokeratin 5 (CK5) and CK6)) called PulmoStrat that distinguishes between squamous cell carcinoma and adenocarcinoma of the lung. The assay uses a weighted algorithm trained to discriminate adenocarcinoma from squamous cell carcinoma and was assessed in three independent cohorts comprising more than 1,000 patients. PulmoStrat could have considerable clinical ramifications, as epidermal growth factor receptor (EGFR) mutations, which predict response to EGFR inhibitors such as gefitinib and erlotinib are much more prevalent in adenocarcinoma[84,85].

Prognostics. The advent of TMAs and high-throughput pathology has provided an ideal platform for the development of IHC-based surrogates of gene expression profiles and has enabled the production of simple, effective and reproducible assays that are readily translatable to the clinic[59]. Several groups have used IHC-based surrogates to validate breast cancer molecular subtypes, including Neilson *et al.*[86] and Carey *et al.*[87] who demonstrated that predefined breast cancer molecular subtypes could be identified using a small number of IHC markers. Additionally, it has been demonstrated that several

© 2010 Macmillan Publishers Limited. All rights reserved

Unsupervised analysis
A form of gene expression analysis that involves the discovery of empirical structure (patterns) in a given data set without taking into account any prior knowledge of the underlying biology. Gene expression patterns that are discovered in this manner should be unbiased.

different prognostic gene expression signatures developed for breast cancer over the past decade show significant agreement in outcome prediction in a single cohort of patients[88]. As these signatures demonstrated minimal overlap of genes, these findings suggest that they are all tracking a similar phenotype, which may be identifiable in the future by using a panel of IHC markers.

Another approach is to integrate gene expression data and antibody-based tissue profiling to generate new IHC-based signatures. Ring *et al.*[89] generated polyclonal affinity-purified antibodies against 700 targets that were identified from a range of published breast cancer gene expression data sets and developed a five marker signature that measures p53, N-myc downstream regulated 1 (NDRG1), CEACAM5, SLC7A5 and HTF9C (also known as TRMT2A), and predicts disease-free and overall survival in ER-positive patients with breast cancer. Although the five markers are associated with various intracellular pathways, all can be linked to cellular proliferation and differentiation[89]. A subsequent validation study suggested that MammaStrat may predict outcome in ER-positive, lymph node-negative breast tumours and so identify a group of patients who would benefit from adjuvant chemotherapy[90].

Although several other IHC signatures have been published in breast[91,92], colorectal[68] and renal cell[93,94] carcinomas, these have generally been studied in small patient cohorts and have been generated from a limited number of preselected proteins. Such studies require extensive validation and the preselection of antibodies based on expression patterns in a single cohort may hamper future validation studies. The integration of gene expression and proteomic data, combined with the high-throughput generation and/or screening of comprehensive antibody panels, allows for unsupervised analysis of protein expression in large patient cohorts, which is more likely to produce more robust assays, as demonstrated by Ring *et al.*[89].

Molecular therapeutics. The advent of molecularly targeted therapy has led to a shift of emphasis away from prognostic signatures. Although intrinsic signatures provide a global view of tumour phenotype, the heterogeneous nature of cancer suggests that more subtle approaches, which are based on the profiling of specific intracellular pathways to personalize treatment regimens, could ultimately prove to be more beneficial. This approach, known as pathway diagnostics, is already in practice, as demonstrated by the use of *KRAS* mutation status (assessed using various assays, although nested PCR followed by direct sequencing and allele-specific real-time PCR is most widely used at present[95]) to predict response to therapeutic EGFR-specific antibodies in metastatic colorectal cancer[96,97].

Clinical application of pathway diagnostics involves a key shift towards monitoring upstream and downstream indicators of pathway function before, during and following treatment. Lessons learned from the first decade of molecular therapeutics suggest that two key elements will predict their successful translation into the clinic. First, the identification of the correct patient subgroup is paramount, as demonstrated by the aforementioned examples of EGFR inhibitors in lung cancer[84,85]. Present understanding suggests that the identification of the correct patient cohort depends on two important factors: namely, the activity of the pathway being targeted and the molecular lesion leading to target activation[98]. Imatinib, which inhibits the constitutive kinase activity of the breakpoint cluster region (BCR)–ABL1 oncogenic fusion protein, which is the product of a chromosomal translocation in patients with chronic myeloid leukaemia[99], is the most obvious example of this phenomenon. Second, measuring inhibition of the targeted signalling pathway is required to guide dose selection and scheduling, and may also monitor off-target effects of potential drugs, thus predicting side effect profiles at an earlier stage[100]. Such an approach also offers the opportunity to identify patterns of both intrinsic and acquired resistance to therapy.

Another example of the importance of identification of the correct patient subgroup is the use of Raf inhibitors in melanomas that harbour a BRAF mutation, which is present in approximately 50% of cases[101]. BRAF mutations (particularly the V600E mutation) result in increased basal kinase activity and hyperactivity of the MAPK pathway, thus promoting tumorigenesis[102]. Targeting BRAF or its downstream effectors may have potential benefits, and early clinical trials examining non-selective Raf kinase inhibitors, such as sorafenib[103–105], and selective compounds, such as PLX4032 (REF. 106), which specifically targets BRAF-V600E, have

Figure 2 | **Antibody-based proteomics and personalized cancer medicine.** The integration of data from high-throughput screening methodologies, such as DNA microarrays, mass spectrometry and functional genomic screens, with antibody-based proteomics, offers a great opportunity to identify new, robust cancer biomarkers. Clinically applicable assays may be developed using various approaches, including immunohistochemistry (IHC)-based prognostic and predictive signatures and pathway analysis using reverse phase protein arrays (RPPAs) and/or IHC. Immunohistochemical analysis offers the potential to identify patient subgroups for targeted therapy while also monitoring therapeutic response. Serum-based proteomics using antibody arrays and/or RPPAs offer less invasive approaches for diagnosis, particularly for screening high-risk populations and the early detection of recurrence, although molecular imaging has the potential to become the major platform for diagnostic, prognostic and predictive tests in the future. TMA, tissue microarray.

© 2010 Macmillan Publishers Limited. All rights reserved

shown promise. Such optimism should be tempered by recent findings that Raf inhibitors promote tumour growth and MAPK activation in tumours expressing wild-type Ras and Raf as well as mutant KRAS[107], which further highlights the requirement for strict patient selection and the development of companion molecular diagnostics for the future trials of such agents.

Such assays are most likely to focus on monitoring pathway activity by either definitive quantification of phosphorylated proteins or the identification of surrogate markers of pathway activity. Antibody-based proteomics offers two approaches for monitoring such pathway activity in tumour samples, namely multi-marker IHC-based assays and RPPAs. Both require a large library of phospho-specific antibodies, and although RPPA may be an excellent platform in the discovery phase, particularly in clinical trials in which small amounts of tissue can be profiled with a large number of antibodies, an IHC-based assay may arguably be a more clinically applicable assay. To date, the optimization of phospho-specific antibodies for IHC has been difficult, suggesting that standardization of such an assay is challenging and that surrogate markers of pathway activation could provide an alternative and more robust approach.

Several groups have published IHC-based studies outlining the activation of different intracellular pathways in breast cancer, oesophageal cancer, renal cell carcinoma, soft tissue sarcoma and prostate cancer[100,108–113]. In general, these studies have been limited by small patient cohorts; however, Dahinden *et al.*[114] recently used 15 antibodies to examine both the von Hippel-Lindau tumour suppressor (VHL) and PTEN pathways in 800 clear cell renal cell carcinomas and were able to refine tumour grading and staging accordingly. Likewise, Yoshizawa *et al.*[111] demonstrated that the activation of the AKT pathway was associated with a poor prognosis in 300 cases of non-small-cell lung cancer. Although these studies used large cohorts, they will require validation in independent cohorts from multiple institutions. Both studies used manual IHC analysis, and the development of robust algorithms for IHC quantification will potentially allow the application of more complex mathematical models to IHC data, thus allowing more complex pathway analysis in large cohorts of patients.

The above-mentioned studies were carried out using retrospective cohorts; however, it is likely that reduction to clinical utility will depend on using samples from prospective trials of molecularly targeted agents. This idea is now being applied to colorectal cancer, and two large Phase III trials examining the addition of EGFR monoclonal antibodies to current therapy regimens have now finished the recruitment of more than 2,000 patients[96]. All samples will be profiled in an attempt to prospectively identify predictive biomarkers. An appropriate approach that could be used in this setting would be to use RPPAs to profile a larger number of proteins both upstream and downstream of the inhibitor target to identify potential predictive markers and to then reduce the assay to a clinically applicable multiplex IHC assay or ELISA. Pernas *et al.*[115] recently used RPPAs to profile head and neck squamous cell carcinoma cell lines to pathways downstream of EGFR, such as ERK, AKT, signal transducer and activator of transcription 3 (STAT3) and nuclear factor-κB (NF-κB), to identify markers of response to the EGFR inhibitor gefitinib. Interestingly, both STAT3 and phospho-STAT3 were associated with gefitinib response in cell lines and tumour samples using both IHC and RPPAs[115]. Although this study included a small number of patients, and the role of STAT3 as a predictive biomarker requires further validation, it highlights the approach that could be used in the translational group of larger clinical trials in which pretreatment and post-treatment analysis of protein expression can be used to identify predictive biomaker panels (FIG. 3). Given the precious nature of the material available from prospective trials and the small size of pretreatment biopsy samples, RPPAs may be an ideal platform for initial discovery with predictive assays that are later reduced in complexity to IHC or ELISA to ensure clinical applicability.

Such an approach would be a field shift from current practice, in which biomarker discovery is predominantly based on retrospective cohorts, and prospective cohorts are generally used for validation. It is plausible that this could be one of the major reasons why so few biomarkers survive independent validation[3,116]. A need to fast-track predictive biomarker development has been highlighted by several groups[2,117], and it is likely that the combination of high-throughput *in vitro* assays, combining gene expression analysis and functional genomic screens, and proteomic profiling in smaller numbers of patients participating in prospective randomized control trials may be a more successful strategy (FIG. 3).

Although this approach is suitable for newly identified compounds, it is also important to focus on established molecular therapeutics. Anti-angiogenic therapy using monoclonal antibodies and tyrosine kinase inhibitors targeting vascular endothelial growth factor (VEGF) has become the standard of care in several solid tumours, including colorectal cancer, renal cell carcinoma, breast cancer, non-small-cell lung cancer and glioblastoma either alone or in combination with chemotherapy[118]. Despite promising results in various tumour types, anti-angiogenic therapy is still limited by a lack of predictive biomarkers, particularly as innate and acquired resistance is an ever increasing clinical dilemma[119,120]. To examine this, several groups have recently used multiplex bead assays to profile various circulating cytokines and angiogenic factors (CAFs) in patients treated with anti-angiogenic agents[121–123].

In a Phase II trial of untreated metastatic colorectal cancer comparing the addition of the monocolonal VEGF-specific antibody bevacizumab to a chemotherapeutic regimen combining fluorouracil, leucovorin and irinotecan, Kopetz *et al.*[122] used a multiplex bead assay to demonstrate that several CAFs, including basic fibroblast growth factor (also known as FGF2), hepatocyte growth factor (HGF), placental growth factor (PGF), stromal-derived factor 1 (SDF1) and monocyte chemoattractant protein 3 (MCP3), were significantly increased from baseline pretreatment levels before any radiological evidence of progressive disease. These results highlight how this approach can be used to monitor

Retrospective cohort
A study in which the medical records and possibly also the previous tissue specimens of groups of patients with a specific diagnosis (for example, breast cancer) are collected.

Prospective trial
A trial in which the participants or patients are identified, followed over time and the effects of different conditions on their eventual outcome are measured.

© 2010 Macmillan Publishers Limited. All rights reserved

patients on therapy. Using a complementary approach, Nikolinakos *et al.*[123] used multiplex bead assay profiling to identify a CAF signature consisting of HGF and interleukin-12 (IL-12), which predicts response to the anti-angiogenic tyrosine kinase inhibitor pazopanib in early-stage non-small-cell lung cancer. Although these data require validation, they demonstrate the application of multiplex bead assays to identify new predictive biomarkers for angiogenesis inhibitors.

Translating assays to the clinic

Despite great advances in the preclinical arena, the translation of new assays to the clinic has been slow. This problem is likely to be multifactorial; however, several issues have come to the fore over the past decade. It is now widely accepted that predictive, rather than prognostic, markers will maximally benefit personalized therapeutic regimens. This poses a considerable problem for solid tissue malignancies.

As technology platforms continue to improve it is likely that high-throughput clinical translation will require a substantial change in the scientific and clinical approach to diagnostic and prognostic assay development. Although descriptive studies demonstrating a technology are helpful, they rarely answer a pertinent clinical question, and it is therefore imperative that translational oncology moves back towards a hypothesis-driven approach in which studies are designed to answer specific, predefined clinical questions similar to those addressed above with respect to angiogenesis inhibitors[122,123].

As mentioned previously, serum-based diagnostics have not delivered on their initial promise, which might be due to poor study design, particularly as most studies have simply compared disease to normal controls in small, underpowered cohorts. A move away from population-based assays to specific tests for high-risk groups could allow for easier clinical translation. A diagnostic serum test for screening high-risk individuals with premalignant conditions such as Barrett's oesophagus, ulcerative colitis, ductal carcinoma *in situ* and atypical endometrial hyperplasia would be particularly beneficial, as patients would avoid recurrent invasive investigations such as endoscopy. Diagnostic serum assays for high-risk individuals are likely to involve different proteins and pathways from those that could be developed for population-based screening, as the underlying biology is likely to be distinct and this group would require a highly specific test, compared with a sensitive test that is required for population-based screening.

Although predictive biomarkers are increasingly used in leukaemia trials[117], efforts to apply such markers in clinical trials for treating solid tumours have not been particularly successful, as it is challenging to gain access to tumour tissue during treatment so that predictive biomarkers can be measured. Unlike leukaemia, in which large numbers of tumour cells are present in the peripheral blood, solid tumour tissue is usually only accessed at diagnosis, by either biopsy or resection. Although this approach might be sufficient to study prognostic biomarkers, it severely limits the application of predictive and pharmacodynamic biomarkers because these measurements are ideally carried out concurrently with treatment[117]. In addition, experimental drugs are typically evaluated in patients with late-stage disease who do not routinely undergo additional tumour biopsies. It is hoped that serum-based proteomics, or proteomic strategies that evaluate predictive markers in circulating

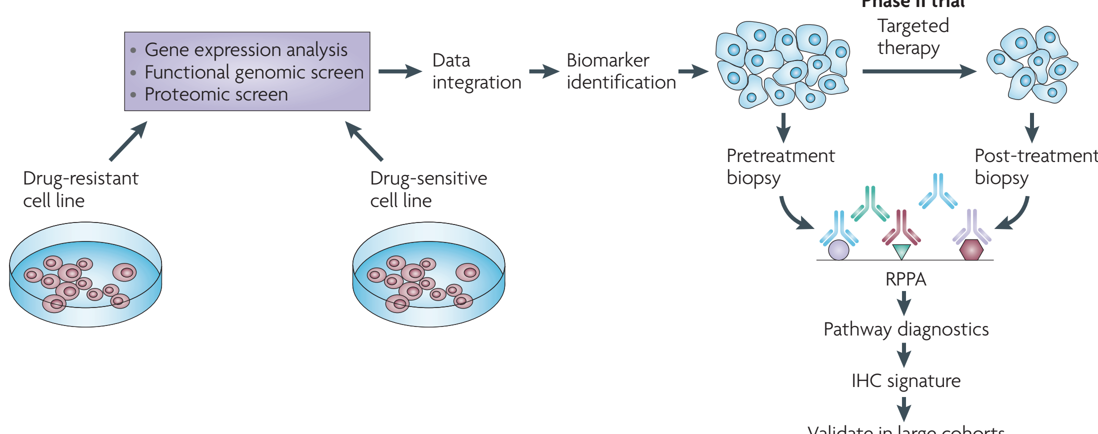


Figure 3 | **Translating antibody-based assays into the clinic.** Predictive biomarkers for targeted therapies will be crucial for identifying correct patient subgroups. The development of such biomarker panels is likely to be based on a systems biology approach, in which high-throughput screening using various methodologies can be carried out *in vitro*. The integration of gene expression and proteomic data will be crucial to the identification of new biomarkers. It will also be necessary to institute biomarker testing in patients at an earlier stage than is currently practised, particularly in Phase I and II trials. Analysis of pathway activation is likely to be a key predictor of response. Reverse phase protein arrays (RPPAs) offer the opportunity to institute pathway diagnostics in early trials, particularly if pretreatment and post-treatment tissue samples can be obtained. Given the precious nature of the material available from prospective studies and the small size of pretreatment biopsy samples, RPPAs may be an ideal platform for initial discovery with predictive assays reduced to clinical applicability through immunohistochemistry (IHC) or enzyme-linked immunosorbent assay (ELISA) over time.

© 2010 Macmillan Publishers Limited. All rights reserved

tumour cells, might help overcome this hurdle; however, trial designs incorporating multiple biopsies during treatment could also be necessary.

In addition, it is now obvious that extensive heterogeneity exists between patients and tumour samples, and investigations of biomarkers may need to be evaluated in the context of specific tumour subtypes. This was demonstrated by Kobel *et al.*[124] who examined 20 tissue markers in a population-based cohort of 500 epithelial ovarian carcinomas, and demonstrated that the association between biomarker expression and survival varied substantially between histological subtypes and it could be easily overlooked in whole-cohort analyses. As the prevalence of certain tumour subtypes is low (such as clear cell carcinoma of the ovary), it is likely that large research consortia will be required to build collaborative efforts in this field. Such consortia will rely on standardization of tissue fixation and processing, as well as standardized protocols for the collection, processing and storage of serum samples. Although this has been highlighted as a particularly important issue with regard to the reproducibility of various assays, it remains an important consideration when developing a research consortium.

The other obvious bottleneck in transfer from biomarker discovery to clinical application is primarily related to a lack of rigorous validation of emerging biomarkers. In 2004, a standards template was developed: the standards for reporting of diagnostic accuracy (STARD) initiative[125]. Unfortunately, many of the studies published regarding new candidate biomarkers fail to meet these standards. In particular, the studies are often carried out on small retrospective cohorts and lack statistical power. Additionally, many biomarker studies fail to include an independent validation stage, whereby the biomarker is evaluated using a second independent cohort of patients[59].

Conclusion

The past decade has witnessed considerable progress in the development and advancement of affinity techniques, methodologies and concepts. Using the technologies described above, antibody-based proteomics offers the opportunity to exploit the specificity and sensitivity associated with antibody-based assays to functionally interrogate tumour biology on a proteome-wide level. Such an approach has the potential to identify new cancer biomarker panels, which can be reduced to clinically applicable assays, including IHC and ELISA, thus providing a high-throughput approach for biomarker development, validation and clinical implementation. It is hoped that such an approach will accelerate the development of personalized therapeutic regimens for cancer patients.

1. Chin, L. & Gray, J. W. Translating insights from the cancer genome into clinical practice. *Nature* **452**, 553–563 (2008).
2. van't Veer, L. J. & Bernards, R. Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature* **452**, 564–570 (2008).
3. Hartwell, L., Mankoff, D., Paulovich, A., Ramsey, S. & Swisher, E. Cancer biomarkers: a systems approach. *Nature Biotechnol.* **24**, 905–908 (2006).
4. Hanash, S. HUPO initiatives relevant to clinical proteomics. *Mol. Cell. Proteomics* **3**, 298–301 (2004).
5. Uhlen, M. & Ponten, F. Antibody-based proteomics for human tissue profiling. *Mol. Cell. Proteomics* **4**, 384–393 (2005).
6. Clamp, M. *et al.* Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl Acad. Sci. USA* **104**, 19428–19433 (2007).
7. Ponten, F., Jirstrom, K. & Uhlen, M. The Human Protein Atlas - a tool for pathology. *J. Pathol.* **216**, 387–393 (2008).
8. Taussig, M. J. *et al.* ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nature Methods* **4**, 13–17 (2007).
9. Uhlén, M. *et al.* A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteomics* **4**, 1920–1932 (2005).
This paper describes the Human Protein Atlas as an antibody-based proteomics initiative.
10. Warford, A., Flack, G., Conquer, J. S., Zola, H. & McCafferty, J. Assessing the potential of immunohistochemistry for systematic gene expression profiling. *J. Immunol. Methods* **318**, 125–137 (2007).
11. Jogi, A. *et al.* Nuclear expression of the RNA-binding protein *RBM3* is associated with an improved clinical outcome in breast cancer. *Mod. Pathol.* **22**, 1564–1574 (2009).
12. Bordeaux, J. *et al.* Antibody validation. *Biotechniques* **48**, 197–209 (2010).
13. Kononen, J. *et al.* Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nature Med.* **4**, 844–847 (1998).
This seminal paper describes TMA technology.
14. Battifora, H. The multitumor (sausage) tissue block: novel method for immunohistochemical antibody testing. *Lab. Invest.* **55**, 244–248 (1986).
15. Wan, W. H., Fortuna, M. B. & Furmanski, P. A rapid and efficient method for testing immunohistochemical reactivity of monoclonal antibodies against multiple tissue samples simultaneously. *J. Immunol. Methods* **103**, 121–129 (1987).
16. Camp, R. L., Neumeister, V. & Rimm, D. L. A decade of tissue microarrays: progress in the discovery and validation of cancer biomarkers. *J. Clin. Oncol.* **26**, 5630–5637 (2008).
17. Brennan, D. J. *et al.* Contribution of DNA and tissue microarray technology to the identification and validation of biomarkers and personalised medicine in breast cancer. *Cancer Genom. Proteom.* **4**, 3–16 (2007).
18. Hewitt, S. M. The application of tissue microarrays in the validation of microarray results. *Meth.Enzymol.* **410**, 400–415 (2006).
19. Fernandez, D. C., Bhargava, R., Hewitt, S. M. & Levin, I. W. Infrared spectroscopic imaging for histopathologic recognition. *Nature Biotech.* **23**, 469–474 (2005).
20. Rimm, D. What brown cannot do for you. *Nature Biotech.* **24**, 914–916 (2006).
21. Rhodes, A., Jasani, B., Barnes, D. M., Bobrow, L. G. & Miller, K. D. Reliability of immunohistochemical demonstration of oestrogen receptors in routine practice: interlaboratory variance in the sensitivity of detection and evaluation of scoring systems. *J. Clin. Pathol.* **53**, 125–130 (2000).
22. Brennan, D. J. *et al.* Altered cytoplasmic-nuclear ratio of survivin is a prognostic indicator in breast cancer. *Clin. Cancer Res.* **14**, 2681–2689 (2008).
23. Mulrane, L., Rexhepaj, E., Penney, S., Callanan, J. J. & Gallagher, W. M. Automated image analysis in histopathology: a valuable tool in medical diagnostics. *Expert Rev. Mol. Diagn.* **8**, 707–725 (2008).
24. Rexhepaj, E. *et al.* Novel image analysis approach for quantifying expression of nuclear proteins assessed by immunohistochemistry: application to measurement of estrogen and progesterone receptor levels in breast cancer. *Breast Cancer Res.* **10**, R89 (2008).
25. Rojo, M. G., Garcia, G. B., Mateos, C. P., Garcia, J. G. & Vicente, M. C. Critical comparison of 31 commercially available digital slide systems in pathology. *Int. J. Surg. Pathol.* **14**, 285–305 (2006).
26. Strömberg, S. *et al.* A high-throughput strategy for protein profiling in cell microarrays using automated image analysis. *Proteomics* **7**, 2142–2150 (2007).
27. Brennan, D. J. *et al.* The transcription factor *Sox11* is a prognostic factor for improved recurrence-free survival in epithelial ovarian cancer. *Eur. J. Cancer* **45**, 1510–1517 (2009).
28. Pages, F. *et al.* Effector memory T cells, early metastasis, and survival in colorectal cancer. *N. Engl. J. Med.* **353**, 2654–2666 (2005).
29. Galon, J. *et al.* Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science.* **313**, 1960–1964 (2006).
30. Brennan, D. J. *et al.* Tumour-specific HMG-CoAR is an independent predictor of recurrence free survival in epithelial ovarian cancer. *BMC Cancer* **10**, 125 (2010).
31. Camp, R., Chung, G. & Rimm, D. Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nature Med.* **8**, 1323–1327 (2002).
32. Dolled-Filhart, M. *et al.* Classification of breast cancer using genetic algorithms and tissue microarrays. *Clin. Cancer Res.* **12**, 6459–6468 (2006).
33. Giltnane, J. *et al.* Quantitative measurement of epidermal growth factor receptor is a negative predictive factor for tamoxifen response in hormone receptor positive premenopausal breast cancer. *J. Clin. Oncol.* **25**, 3007–3014 (2007).
34. McCabe, A., Dolled-Filhart, M., Camp, R. L. & Rimm, D. L. Automated quantitative analysis (AQUA) of *in situ* protein expression, antibody concentration, and prognosis. *J. Natl Cancer Inst.* **97**, 1808–1815 (2005).
This paper reported the demonstration of automated analysis of immunofluorescence.
35. Camp, R. L., Dolled-Filhart, M., King, B. L. & Rimm, D. L. Quantitative analysis of breast cancer tissue microarrays shows that both high and normal levels of *HER2* expression are associated with poor outcome. *Cancer Res.* **63**, 1445–1448 (2003).
36. Gould Rothberg, B. E. *et al.* Melanoma prognostic model using tissue microarrays and genetic algorithms. *J. Clin. Oncol.* **27**, 5772–5780 (2009).
37. Hu, S., Loo, J. A. & Wong, D. T. Human body fluid proteome analysis. *Proteomics* **6**, 6326–6353 (2006).
38. Anderson, N. L. *et al.* The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* **3**, 311–326 (2004).
39. Hanash, S. Disease proteomics. *Nature.* **422**, 226–232 (2003).

© 2010 Macmillan Publishers Limited. All rights reserved

40. Borrebaeck, C. A. & Wingren, C. High-throughput proteomics using antibody microarrays: an update. *Expert Rev. Mol. Diagn.* **7**, 673–686 (2007).
41. Kingsmore, S. F. Multiplexed protein measurement: technologies and applications of protein and antibody arrays. *Nature Rev. Drug Discov.* **5**, 310–320 (2006).
42. Wingren, C. & Borrebaeck, C. A. Antibody microarray analysis of directly labelled complex proteomes. *Curr. Opin. Biotechnol.* **19**, 55–61 (2008).
43. Chen, S. & Haab, B. B. Analysis of glycans on serum proteins using antibody microarrays. *Methods Mol. Biol.* **520**, 39–58 (2009).
44. Narimatsu, H. *et al.* A strategy for discovery of cancer glyco-biomarkers in serum using newly developed technologies for glycoproteomics. *FEBS J.* **277**, 95–105 (2010).
45. Schwenk, J. M., Gry, M., Rimini, R., Uhlen, M. & Nilsson, P. Antibody suspension bead arrays within serum proteomics. *J. Proteome Res.* **7**, 3168–3179 (2008).
46. Gowan, S. M. *et al.* Application of meso scale technology for the measurement of phosphoproteins in human tumor xenografts. *Assay Drug Dev. Technol.* **5**, 391–401 (2007).
47. Leng, S. X. *et al.* ELISA and multiplex technologies for cytokine measurement in inflammation and aging research. *J. Gerontol. A Biol. Sci. Med. Sci.* **63**, 879–884 (2008).
48. Morgan, E. *et al.* Cytometric bead array: a multiplexed assay platform with applications in various areas of biology. *Clin. Immunol.* **110**, 252–266 (2004).
49. Elshal, M. F. & McCoy, J. P. Multiplex bead array assays: performance evaluation and comparison of sensitivity to ELISA. *Methods* **38**, 317–323 (2006).
50. Khan, S. S., Smith, M. S., Reda, D., Suffredini, A. F. & McCoy, J. P. Jr. Multiplex bead array assays for detection of soluble cytokines: comparisons of sensitivity and quantitative values among kits from multiple manufacturers. *Cytometry B Clin. Cytom.* **61**, 35–39 (2004).
51. Prabhakar, U. *et al.* Validation and comparative analysis of a multiplexed assay for the simultaneous quantitative measurement of Th1/Th2 cytokines in human serum and human peripheral blood mononuclear cell culture supernatants. *J. Immunol. Methods* **291**, 27–38 (2004).
52. Spurrier, B., Ramalingam, S. & Nishizuka, S. Reverse-phase protein lysate microarrays for cell signaling analysis. *Nature Protoc.* **3**, 1796–1808 (2008).
This paper describes the basis of how RPPAs can be used for pathway diagnostics.
53. Berg, D., Hipp, S., Malinowsky, K., Bollner, C. & Becker, K. F. Molecular profiling of signalling pathways in formalin-fixed and paraffin-embedded cancer tissues. *Eur. J. Cancer* **46**, 47–55 (2010).
54. Espina, V. *et al.* A portrait of tissue phosphoprotein stability in the clinical tissue procurement process. *Mol. Cell. Proteomics* **7**, 1998–2018 (2008).
55. Chung, J. L., Lee, S. J., Kris, Y., Braunschweig, T. & Hewitt, S. M. A well-based reverse-phase protein array applicable to extracts from formalin-fixed paraffin-embedded tissue. *Proteomics Clin. Appl.* **2**, 1539–1547 (2008).
56. Pawlak, M. *et al.* Zeptosens' protein microarrays: a novel high performance microarray platform for low abundance protein analysis. *Proteomics* **2**, 383–393 (2002).
57. Dupuy, L. *et al.* A highly sensitive near-infrared fluorescent detection method to analyze signalling pathways by reverse-phase protein array. *Proteomics* **9**, 5446–5454 (2009).
58. Brennan, D. J. *et al.* Application of DNA microarray technology in determining breast cancer prognosis and therapeutic response. *Expert Opin. Biol. Ther.* **5**, 1069–1083 (2005).
59. Brennan, D. J. & Gallagher, W. M. Prognostic ability of a panel of immunohistochemistry markers - retailoring of an 'old solution'. *Breast Cancer Res.* **10**, 102 (2008).
60. van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
This seminal paper describes the 70 gene prognostic signature for breast cancer that is now the basis of the MammaPrint assay.
61. Glas, A. M. *et al.* Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* **7**, 278 (2006).
62. Paik, S. *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351**, 2817–2826 (2004).
This paper describes the Oncotype Dx assay.
63. Sparano, J. TAILORx: trial assigning individualized options for treatment (Rx). *Clin. Breast Cancer* **7**, 347–350 (2006).
64. Buyse, M. *et al.* Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J. Natl Cancer Inst.* **98**, 1183–1192 (2006).
65. Tan, E. H. *et al.* A multicentre phase II gene expression profiling study of putative relationships between tumour biomarkers and clinical response with erlotinib in non-small-cell lung cancer. *Ann. Oncol.* **21**, 217–222 (2010).
66. Chen, H. Y. *et al.* A five-gene signature and clinical outcome in non-small-cell lung cancer. *N. Engl. J. Med.* **356**, 11–20 (2007).
67. Guo, N. L. *et al.* Confirmation of gene expression-based prediction of survival in non-small cell lung cancer. *Clin. Cancer Res.* **14**, 8213–8220 (2008).
68. Hao, J. M. *et al.* A five-gene signature as a potential predictor of metastasis and survival in colorectal cancer. *J. Pathol.* **220**, 475–489 (2010).
69. Garman, K. S. *et al.* A genomic approach to colon cancer risk stratification yields biologic insights into therapeutic opportunities. *Proc. Natl Acad. Sci. USA.* **105**, 19432–19437 (2008).
70. Barrier, A. *et al.* Stage II colon cancer prognosis prediction by tumor gene expression profiling. *J. Clin. Oncol.* **24**, 4685–4691 (2006).
71. Cheville, J. C. *et al.* Gene panel model predictive of outcome in men at high-risk of systemic progression and death from prostate cancer after radical retropubic prostatectomy. *J. Clin. Oncol.* **26**, 3930–3936 (2008).
72. Kosari, F. *et al.* Identification of prognostic biomarkers for prostate cancer. *Clin. Cancer Res.* **14**, 1734–1743 (2008).
73. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
74. Rosenwald, A. *et al.* The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N. Engl. J. Med.* **346**, 1937–1947 (2002).
This seminal paper describes the classification of B cell lymphoma.
75. Lenz, G. *et al.* Stromal gene signatures in large-B-cell lymphomas. *N. Engl. J. Med.* **359**, 2313–2323 (2008).
76. Whiteaker, J. R. *et al.* Antibody-based enrichment of peptides on magnetic beads for mass-spectrometry-based quantification of serum biomarkers. *Anal. Biochem.* **362**, 44–54 (2007).
77. Anderson, N. L. *et al.* Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J. Proteome Res.* **3**, 235–244 (2004).
78. Ingvarsson, J. *et al.* Detection of pancreatic cancer using antibody microarray-based serum protein profiling. *Proteomics* **8**, 2211–2219 (2008).
79. Orchekowski, R. *et al.* Antibody microarray profiling reveals individual and combined serum proteins associated with pancreatic cancer. *Cancer Res.* **65**, 11193–11202 (2005).
80. Carlsson, A. *et al.* Serum proteome profiling of metastatic breast cancer using recombinant antibody microarrays. *Eur. J. Cancer.* **44**, 472–480 (2008).
81. Schroder, C. *et al.* Dual-color proteomic profiling of complex samples with a microarray of 810 cancer-related antibodies. *Mol. Cell. Proteomics* **9**, 1271–1280 (2010).
82. Kashani-Sabet, M. *et al.* A multi-marker assay to distinguish malignant melanomas from benign nevi. *Proc. Natl Acad. Sci. USA* **106**, 6268–6272 (2009).
This paper describes an IHC assay for the diagnosis of malignant melanoma.
83. Ring, B. Z. *et al.* A novel five-antibody immunohistochemical test for subclassification of lung carcinoma. *Mod. Pathol.* **22**, 1032–1043 (2009).
84. Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**, 2129–2139 (2004).
85. Paez, J. G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science.* **304**, 1497–1500 (2004).
86. Nielsen, T. *et al.* Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin. Cancer Res.* **10**, 5367–5374 (2004).
87. Carey, L. *et al.* Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* **295**, 2492–2502 (2006).
88. Fan, C. *et al.* Concordance among gene-expression-based predictors for breast cancer. *N. Engl. J. Med.* **355**, 560–569 (2006).
89. Ring, B. *et al.* Novel prognostic immunohistochemical biomarker panel for estrogen receptor-positive breast cancer. *J. Clin. Oncol.* **24**, 3039–3047 (2006).
This paper describes a five marker IHC prognostic assay for breast cancer, which is an example of integrating transcriptomic and proteomic data.
90. Ross, D. T. *et al.* Chemosensitivity and stratification by a five monoclonal antibody immunohistochemistry test in the NSABP B14 and B20 trials. *Clin. Cancer Res.* **14**, 6602–6609 (2008).
91. Crabb, S. J. *et al.* Can clinically relevant prognostic subsets of breast cancer patients with four or more involved axillary lymph nodes be identified through immunohistochemical biomarkers: a tissue microarray feasibility study. *Breast Cancer Res.* **10**, R6 (2008).
92. Charpin, C. *et al.* A signature predictive of disease outcome in breast carcinomas, identified by quantitative immunocytochemical assays. *Int. J. Cancer* **124**, 2124–2134 (2009).
93. Kim, H. L. *et al.* Using protein expressions to predict survival in clear cell renal carcinoma. *Clin. Cancer Res.* **10**, 5464–5471 (2004).
94. Klatte, T. *et al.* Molecular signatures of localized clear cell renal cell carcinoma to predict disease-free survival after nephrectomy. *Cancer Epidemiol. Biomarkers Prev.* **18**, 894–900 (2009).
95. Kobunai, T., Watanabe, T., Yamamoto, Y. & Eshima, K. The frequency of KRAS mutation detection in human colon carcinoma is influenced by the sensitivity of assay methodology: a comparison between direct sequencing and real-time PCR. *Biochem. Biophys. Res. Commun.* **395**, 158–162 (2010).
96. Banck, M. S. & Grothey, A. Biomarkers of resistance to epidermal growth factor receptor monoclonal antibodies in patients with metastatic colorectal cancer. *Clin. Cancer Res.* **15**, 7492–7501 (2009).
97. Siena, S., Sartore-Bianchi, A., Di Nicolantonio, F., Balfour, J. & Bardelli, A. Biomarkers predicting clinical outcome of epidermal growth factor receptor-targeted therapy in metastatic colorectal cancer. *J. Natl Cancer Inst.* **101**, 1308–1324 (2009).
98. Weinstein, I. B. Cancer. Addiction to oncogenes - the Achilles heal of cancer. *Science* **297**, 63–64 (2002).
99. Sawyers, C. L. Shifting paradigms: the seeds of oncogene addiction. *Nature Med.* **15**, 1158–1161 (2009).
100. Thomas, G. V. *et al.* Antibody-based profiling of the phosphoinositide 3-kinase pathway in clinical prostate cancer. *Clin. Cancer Res.* **10**, 8351–8356 (2004).
101. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954 (2002).
This is a key paper outlining why the identification of the correct patient cohort for molecular therapeutics is of utmost importance.
102. Wan, P. T. *et al.* Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* **116**, 855–867 (2004).
103. Amaravadi, R. K. *et al.* Phase II trial of temozolomide and sorafenib in advanced melanoma patients with or without brain metastases. *Clin. Cancer Res.* **15**, 7711–7718 (2009).
104. Hauschild, A. *et al.* Results of a phase III, randomized, placebo-controlled study of sorafenib in combination with carboplatin and paclitaxel as second-line treatment in patients with unresectable stage III or stage IV melanoma. *J. Clin. Oncol.* **27**, 2823–2830 (2009).
105. McDermott, D. F. *et al.* Double-blind randomized phase II study of the combination of sorafenib and dacarbazine in patients with advanced melanoma: a report from the 11715 Study Group. *J. Clin. Oncol.* **26**, 2178–2185 (2008).
106. Flaherty, K. T. *et al.* Phase I study of PLX4032: proof of concept for V600E BRAF mutation as a therapeutic target in human cancer. *J. Clin. Oncol. Abstr.* **27**, 9000 (2009).
107. Hatzivassiliou, G. *et al.* RAF inhibitors prime wild-type RAF to activate the MAPK pathway and enhance growth. *Nature* **464**, 431–435 (2010).
108. Tomita, Y. *et al.* Prognostic significance of activated AKT expression in soft-tissue sarcoma. *Clin. Cancer Res.* **12**, 3070–3077 (2006).
109. Capodanno, A. *et al.* Dysregulated PI3K/Akt/PTEN pathway is a marker of a short disease-free survival in node-negative breast carcinoma. *Hum. Pathol.* **40**, 1408–1417 (2009).
110. Abdeen, A. *et al.* Correlation between clinical outcome and growth factor pathway expression in osteogenic sarcoma. *Cancer.* **115**, 5243–5250 (2009).

© 2010 Macmillan Publishers Limited. All rights reserved

111. Yoshizawa, A. *et al.* Overexpression of phospho-eIF4E is associated with survival through AKT pathway in non-small cell lung cancer. *Clin. Cancer Res.* **16**, 240–248 (2010).
This paper describes an example of IHC-based pathway analysis.
112. Tsavachidou-Fenner, D. *et al.* Gene and protein expression markers of response to combined antiangiogenic and epidermal growth factor targeted therapy in renal cell carcinoma. *Ann. Oncol.* 20 Jan 2010 (doi:10.1093/annonc/mdp600).
113. Uemura, N. *et al.* Antibody-based proteomics for esophageal cancer: identification of proteins in the nuclear factor-κB pathway and mitotic checkpoint. *Cancer science.* **100**, 1612–1622 (2009).
114. Dahinden, C. *et al.* Mining tissue microarray data to uncover combinations of biomarker expression patterns that improve intermediate staging and grading of clear cell renal cell cancer. *Clin. Cancer Res.* **16**, 88–98 (2010).
115. Pernas, F. G. *et al.* Proteomic signatures of epidermal growth factor receptor and survival signal pathways correspond to gefitinib sensitivity in head and neck cancer. *Clin. Cancer Res.* **15**, 2361–2372 (2009).
This paper describes an example of combining RPPAs and IHC to carry out pathway diagnostics and identify predictive biomarkers.
116. Gutman, S. & Kessler, L. The US Food and Drug Administration perspective on cancer biomarker development. *Nature Rev. Cancer* **6**, 565–571 (2006).
117. Sawyers, C. L. The cancer biomarker problem. *Nature* **452**, 548–552 (2008).
118. Bose, D. *et al.* Vascular endothelial growth factor targeted therapy in the perioperative setting: implications for patient care. *Lancet Oncol.* **11**, 373–382 (2010).
119. Bergers, G. & Hanahan, D. Modes of resistance to anti-angiogenic therapy. *Nature Rev. Cancer* **8**, 592–603 (2008).
120. Azam, F., Mehta, S. & Harris, A. L. Mechanisms of resistance to antiangiogenesis therapy. *Eur. J. Cancer* **46**, 1323–1332 (2010).
121. Byers, L. A. *et al.* Serum signature of hypoxia-regulated factors is associated with progression after induction therapy in head and neck squamous cell cancer. *Mol. Cancer Ther.* **9**, 1755–1763 (2010).
122. Kopetz, S. *et al.* Phase II trial of infusional fluorouracil, irinotecan, and bevacizumab for metastatic colorectal cancer: efficacy and circulating angiogenic biomarkers associated with therapeutic resistance. *J. Clin. Oncol.* **28**, 453–459 (2010).
This paper describes the utility of multiplex bead assays to predictive markers for anti-angiogenic treatment.
123. Nikolinakos, P. G. *et al.* Plasma cytokine and angiogenic factor profiling identifies markers associated with tumor shrinkage in early-stage non-small cell lung cancer patients treated with pazopanib. *Cancer Res.* **70**, 2171–2179 (2010).
124. Kobel, M. *et al.* Ovarian carcinoma subtypes are different diseases: implications for biomarker studies. *PLoS Med.* **5**, e232 (2008).
125. Bossuyt, P. M. *et al.* The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann. Intern. Med.* **138**, W1–W12 (2003).
126. Berglund, L. *et al.* A genecentric Human Protein Atlas for expression profiles based on antibodies. *Mol. Cell. Proteomics* **7**, 2019–2027 (2008).
127. Berglund, L. *et al.* A whole-genome bioinformatics approach to selection of antigens for systematic antibody generation. *Proteomics* **8**, 2832–2839 (2008).
128. Nilsson, P. *et al.* Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics* **5**, 4327–4337 (2005).
129. Kampf, C. *et al.* Antibody-based tissue profiling as a tool for clinical proteomics. *Clin. Proteomics* **1**, 285–299 (2004).
130. Andersson, A. C. *et al.* Analysis of protein expression in cell microarrays: a tool for antibody-based proteomics. *J. Histochem. Cytochem.* **54**, 1413–1423 (2006).
131. Barbe, L. *et al.* Toward a confocal subcellular atlas of the human proteome. *Mol. Cell. Proteomics* **7**, 499–508 (2008).
132. Björling, 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).
133. Chung, J. *et al.* Transfer and multiplex immunoblotting of a paraffin embedded tissue. *Proteomics* **6**, 767–774 (2006).
134. 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).
135. Petricoin, E. F., Zoon, K. C., Kohn, E. C., Barrett, J. C. & Liotta, L. A. Clinical proteomics: translating benchside promise into bedside reality. *Nature Rev. Drug Discov.* **1**, 683–695 (2002).
136. Haab, B. B., Dunham, M. J. & Brown, P. O. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol.* **2**, research0004.1–research0004.13 (2001).
137. Lizardi, P. M. *et al.* Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nature Genet.* **19**, 225–232 (1998).
138. Rockberg, J., Lofblom, J., Hjelm, B., Uhlen, M. & Stahl, S. Epitope mapping of antibodies using bacterial surface display. *Nature Methods* **5**, 1039–1045 (2008).
139. van Zonneveld, A. J., van den Berg, B. M., van Meijer, M. & Pannekoek, H. Identification of functional interaction sites on proteins using bacteriophage-displayed random epitope libraries. *Gene* **167**, 49–52 (1995).
140. Christmann, A., Wentzel, A., Meyer, C., Meyers, G. & Kolmar, H. Epitope mapping and affinity purification of monospecific antibodies by *Escherichia coli* cell surface display of gene-derived random peptide libraries. *J. Immunol. Methods* **257**, 163–173 (2001).
141. Petersen, G., Song, D., Hugle-Dorr, B., Oldenburg, I. & Bautz, E. K. Mapping of linear epitopes recognized by monoclonal antibodies with gene-fragment phage display libraries. *Mol. Gen. Genet.* **249**, 425–431 (1995).
142. Stadler, C., Skogs, M., Brismar, H., Uhlen, M. & Lundberg, E. A single fixation protocol for proteome-wide immunofluorescence localization studies. *J. Proteomics* **73**, 1067–1078 (2010).
143. Paavilainen, L. *et al.* The impact of tissue fixatives on morphology and antibody-based protein profiling in tissues and cells. *J. Histochem. Cytochem.* **58**, 237–246 (2010).

Acknowledgements
The authors wish to acknowledge funding from Enterprise Ireland, the Health Research Board of Ireland (Programme Grant: Breast Cancer Metastasis: Biomarkers and Functional Mediators and a HRB Career Development Fellowship awarded to D.P.O'C.), the European Commission (in the context of the Marie Curie Industry-Academic Partnership and Pathways programme, Target-Melanoma), Science Foundation Ireland (in the context of the Strategic Research Cluster, Molecular Therapeutics for Cancer Ireland) and the Knut and Alice Wallenberg Foundation. The UCD Conway Institute is funded by the Programme for Research in Third Level Institutions, as administered by the Higher Education Authority of Ireland.

Competing interests statement
The authors declare competing financial interests; see web version for details.

DATABASES
National Cancer Institute Drug Dictionary: http://www.cancer.gov/drugdictionary/
bevacizumab | erlotinib | fluorouracil | gefitinib | imatinib | irinotecan | leucovorin | pazopanib | PLX4032 | sorafenib
Pathway Interaction Database: http://pid.nci.nih.gov/
AKT | EGFR | ERK | PTEN | STAT3

FURTHER INFORMATION
William M. Gallagher's homepage: http://www.cbtlab.ie
CPTAC: http://antibodies.cancer.gov/
Human Protein Atlas: http://www.proteinatlas.org/

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2010 Macmillan Publishers Limited. All rights reserved

Functional proteomics to dissect tyrosine kinase signalling pathways in cancer

Walter Kolch‡ and Andrew Pitt*§

Abstract | Advances in the generation and interpretation of proteomics data have spurred a transition from focusing on protein identification to functional analysis. Here we review recent proteomics results that have elucidated new aspects of the roles and regulation of signal transduction pathways in cancer using the epidermal growth factor receptor (EGFR), ERK and breakpoint cluster region (BCR)–ABL1 networks as examples. The emerging theme is to understand cancer signalling as networks of multiprotein machines which process information in a highly dynamic environment that is shaped by changing protein interactions and post-translational modifications (PTMs). Cancerous genetic mutations derange these protein networks in complex ways that are tractable by proteomics.

Chemical biology
Use of chemicals, usually drugs or drug-like compounds, to probe biological systems to measure the response of biological systems to perturbations. In proteomics, this term also increasingly refers to the use of affinity reagents to enrich classes of proteins for further analysis.

Chemical genetics
A part of chemical biology that focuses on the use of chemicals to explore genetic systems and genetic factors that determine drug sensitivity.

*Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland.
‡Systems Biology Ireland, University College Dublin, Belfield, Dublin 4, Ireland.
§Sir Henry Wellcome Functional Genomics Facility, University of Glasgow, Glasgow G12 8QQ, UK.
Correspondence to W.K.: e-mail: walter.kolch@ucd.ie
doi:10.1038/nrc2900
Published online 19 August 2010

Although the causes of cancer lie in mutations or epigenetic changes at the genetic level, their molecular manifestation is the dysfunction of biochemical pathways at the protein level. Therefore, only by studying the proteome can we gain functional understanding of the pathways that are deranged in cancer. Rather than just identifying proteins, functional proteomics focuses on the generation of information about proteins, such as expression levels, post-translational modifications (PTMs) and activity, which directly contribute to a functional understanding of a biological system. Functional proteomics feeds directly into the systematic analysis of biochemical networks, often using mathematical modelling or other systems biology tools. It also provides readouts for chemical biology and chemical genetics that are necessary to interpret the action of drugs.

The growing interest in functional proteomics is not only fuelled by the prospect of a true functional understanding but also by substantial improvements in technology and methodology. Advances in mass spectrometry (MS) have extended the sensitivity, accuracy and speed of analysis to now routinely enable the identification of several thousand proteins per experiment. The introduction of MS methods for accurate relative and absolute protein quantification and the large-scale analysis of PTMs, such as phosphorylation and ubiquitylation, have allowed truly functional proteomics to be carried out. MS is now joined by antibody and protein–protein interaction arrays[1], fluorescence- and flow cytometry-based detection of proteins and PTMs[2], and optical spectroscopic methods of proteome analysis[3,4]. These latter techniques are promoted by an ever-increasing repertoire of specific antibodies against proteins and PTMs, and bring single-cell proteomics into reach.

Proteomics has been widely applied to cancer research and the mapping of cancerous signalling pathways and the vast literature exceeds the scope of a single Review. Therefore, we will use three main cancer pathways, epidermal growth factor receptor (EGFR, also known as ERBB1), breakpoint cluster region (BCR)–ABL1 and ERK, as examples to describe some of the new paradigms in cancer signalling that owe their discovery to proteomics. The EGFR network is frequently altered in various human cancers and has been extensively studied using proteomics both at the level of the receptor and the downstream pathways. *BCR–ABL1* stands out among oncogenes, as chronic myelogenous leukaemia (CML) — which is caused by this fusion gene — is with few exceptions strictly dependent on BCR–ABL1 function[5]. Thus, BCR–ABL1 signalling could reveal the minimum assortment of pathways required for transforming a human cell. The ERK pathway is a crucial effector of the EGFR and BCR–ABL1 pathways. Hence, although other important pathways — such as the mTOR pathway — are not discussed here, our selection of signalling pathways allows us to describe general principles and representative new insights into cancer signalling pathways that we have gained through proteomics. As technological and methodological advances

© 2010 Macmillan Publishers Limited. All rights reserved

At a glance

- Signalling pathways are commonly deranged in cancer and quantitative proteomics offers powerful approaches to map these pathways and their aberrations in cancer.
- Hubs in signalling pathways feature multiple protein interactions, which are involved in information processing and specification of the biological responses. These networks can be mapped by interaction proteomics to reveal molecular mechanisms of transformation and potential targets for therapeutic interventions.
- The oncogenic actions of the epidermal growth factor receptor (EGFR) network and the breakpoint cluster region (*BCR*)–*ABL1* oncogene rely on the dynamic assembly of multiprotein complexes, which activate multiple downstream pathways that cooperate in transformation. In EGFR networks, the oncogenic potential increases with the number of downstream pathways being activated.
- The dynamic assembly of protein complexes is regulated by post-translational modifications (PTMs) such as phosphorylation. Advances in phosphoproteomics allow the targeted and global mapping of phosphorylation networks, confirming that kinase networks play major parts in cancer and offer numerous new targets for therapeutic intervention.
- In addition to phosphorylation, a role for PTMs in the regulation of cancer cell biology is becoming increasingly recognized. For example, proteomic studies of ubiquitylation are beginning to unravel extensive alterations that contribute to cancer, such as growth factor receptor activation, transcription factor function, protein localization and degradation.
- Dynamic changes in protein abundance and PTMs may also contribute to cancer cell heterogeneity, and new proteomics technologies based on optical, spectroscopic and microarray methods are being developed to analyse individual cells.

Scaffold protein
A protein that can simultaneously bind two or more other proteins, and thereby facilitate physical and functional interactions between the client proteins that bind to it.

Matrix management
A flexible management approach that assigns people with the required skill sets to projects, typically drawing expertise from different departments. This comparison is used to illustrate that a protein with a defined molecular function, such as a kinase, can be used by several different pathways.

Modularization
The grouping of different functions into a single unit (module) so that the output of the module can be treated as a single functional entity, such as the ability of different combinations of components of protein complexes to achieve the same output.

Non-oncogene addiction
This occurs when the action of oncogenes needs to be supported by apparently normally functioning signalling pathways that allow the mutated oncogene to develop its transforming activity.

are integral to functional proteomics, we briefly discuss the experimental approaches, although the emphasis of this Review is on the biological findings.

We begin by discussing protein–protein interactions. Most cellular processes, including signal transduction, require the coordinated formation of multiprotein complexes. A main focus in functional proteomics has been to analyse the formation of specific protein–protein interactions and the consequent assembly of macromolecular protein complexes, particularly how these assemblies are regulated by scaffold proteins and PTMs and how they affect pathway function. We then discuss selected PTMs associated with cancer pathways. Although phosphorylation and glycosylation have dominated the post-translational area of research, we increasingly realize that a large range of modifications, including ubiquitylation, play crucial functions in cellular regulation. We conclude with an outlook towards single-cell proteomics and the clinical relevance of functional cancer proteomics.

Protein–protein interactions

Protein–protein interactions have important roles in cellular processes. However, deconvoluting the complex nature of these interactions has proved to be difficult. Stable protein assemblies function as 'molecular machines' in all cellular processes, from transcription machineries in the nucleus, to ribosomes that translate mRNA into proteins and to molecular motors that generate the forces for intracellular transport and cell motility. These stable protein complexes are highly amenable to proteomic analysis and powerful isolation procedures have been developed to purify stable protein complexes for MS analysis. They usually use tandem affinity purification (TAP) tags, which allow the efficient purification of stable protein complexes by a sequential pulldown–elution–pulldown protocol[6]. Genome-wide interaction studies[7,8] using this methodology showed that the yeast proteome is organized in 500–550 protein complexes, thus experimentally confirming the concept that organization based on protein–protein interactions can specify biological functions. Does this concept also apply to molecular information processors in mammalian cells?

In principle, dynamically changing protein interactions could provide the speed, plasticity and compartmentalization needed to process signals that encode temporal and spatial information. However, these associations may be transient and feature complex dynamics and stoichiometries between different complexes that compete for the same component. Thus, these interactions are often regulated through transient PTMs such as phosphorylation. Specific phosphorylation-dependent-binding motifs have been identified on many signalling proteins. For example, activated receptor tyrosine kinases (RTKs) phosphorylate tyrosines in their cytoplasmic kinase domains which provide binding sites for downstream effectors that direct the assembly of transient signalling complexes[9–11]. RTKs, especially those of the ErbB family, are often overexpressed or mutated in cancer, which results in the assembly of aberrant signalling complexes[9,12].

The technical aspects of mapping protein interactions in signalling pathways have been covered in recent reviews[13–15]. The methodologies rely on the application of quantitative MS techniques[16] (BOX 1) and often require the generation of dynamic interaction maps[17,18]. One of the main advantages is that the quantitative techniques can separate true interactors from contamination[19]. However, substantial challenges remain. Methodologies for measuring the specific occupancy of PTM sites on individual protein molecules are limited. Similarly, in the currently used proteomic pulldown experiments, we cannot distinguish between different subcomplexes, which is akin to having a corporate organization chart with all the names but no affiliations to departments. However, emerging evidence suggests that signal transduction uses a matrix management approach, in which one component fulfils different tasks in different complexes. Mapping this matrix is further complicated by the fact that the structure of a complex can change dynamically over time, and certain complexes may only be necessary for a limited time to activate their downstream biological pathways.

This provides a major challenge, as it may be failures in matrix management that drive cancer rather than individual mutations altering one function of a protein. The crosstalk between protein complexes also contributes to the robustness of cancer signalling by allowing modularization[20] and generating molecular heterogeneity[21], which together allow rapid and versatile adaptation. This view is encapsulated in the concept of non-oncogene addiction, which highlights the importance of normal pathways in supporting the biological effects of mutated pathways[22]. This concept has wide ramifications for the design of rational cancer therapies and the identification of promising targets. The hope is that the common phenomenon of

© 2010 Macmillan Publishers Limited. All rights reserved

Box 1 | **Quantitative proteomics**

Mass spectrometry (MS) is not an *a priori* quantitative method but several approaches have been developed to circumvent this problem. Most modern MS approaches are (semi)quantitative, and this has proven key to the success of proteomics to unravel signalling networks. The most used methods are listed below[110].

Two-dimensional polyacrylamide gel electrophoresis (PAGE)
The sensitivity, linear dynamic range and robustness of this method have been greatly improved by the use of fluorescent stains and tags[111], and the analysis has been substantially simplified by several automated software packages. Proteins are identified by MS.

Isotopic labelling
The ability of MS to resolve heavy and light isotopically labelled molecules has been extensively used to design approaches for quantitation[112–115]. The common labelling methods are metabolic, typified by stable isotope labelling with amino acids in cell culture (SILAC)[116–118]. Samples are grown on media containing a normal or isotopically labelled form of an amino acid or food source[119]. Comparison of protein abundance uses the light and heavy peaks observed in the mass spectrum. SILAC is best suited for cultured cells but has also been used to label whole animals[120]. Chemically reactive, isotopically labelled reagents are used to tag the protein or peptides (such as isotope-coded affinity tag (ICAT)[121] and isobaric tag for relative and absolute quantitation (iTRAQ)[122]). For absolute quantification (AQUA), isotopically labelled peptides are added to the sample as an internal standard for quantification[123].

Label-free
Improvements in the robustness and reproducibility of both MS and chromatography paved the way for direct comparisons between MS data sets using retention time, mass and ion intensity[124,125]. As they are uncomplicated and reasonably reliable, label-free approaches may become the methods of choice for proteomics experiments.

Antibody arrays
The repertoire of antibodies that have highly selective binding to native or post-translationally modified protein epitopes has expanded rapidly and the arrays based on these antibodies have been used extensively to quantify antigens[126,127]. This approach has now developed into a thriving business and many companies offer commercial and bespoke custom arrays.

***In situ* fluorescence**
The use of a fusion protein approach, in which the protein of interest is genetically fused to green fluorescent protein (GFP) or the range of GFP analogues that have now been developed, can allow determination of the cellular localization and abundance of a protein using fluorescence microscopy[128]. Automated methods for tracking cells and determining abundance and localization are only just becoming available.

SH2 domain
This domain was first discovered as a conserved domain in the Src kinase family. It recognizes short peptide motifs that contain a phosphorylated tyrosine residue and thus function as phosphotyrosine-dependent protein interaction sites.

cancer drug resistance can be conquered by the parallel targeting of oncogenic pathways and non-oncogene addictive alterations.

Protein complexes. The EGFR pathway is one of the most extensively studied signalling pathways relevant to cancer[23] and the EGFR interactome is one of the most well-described interactomes, both biochemically[24] and theoretically[25,26] (FIG. 1). Major questions are how EGFR coordinates diverse biological responses and how response specificity is generated. Functional proteomics has deciphered the components of protein complexes formed by EGFR and has also contributed to an understanding of the topological organization of the downstream signalling pathways.

MS-based quantitative proteomics to analyse proteins that interact with phosphopeptides that are phosphorylated in response to ligand binding showed that the four members of the ErbB family have differential preferences for interaction partners[27]. Interestingly, growth factor receptor-bound protein 2 (GRB2), which initiates activation of the ERK pathway, and PI3K, which activates the AKT pathway, have several binding sites on one receptor. Multimeric binding could be a simple mechanism to transmit the strength of the input into these two pathways. A complementary approach used protein microarrays comprising most known SH2 domains and phosphotyrosine-binding (PTB) domains in the human genome to measure the dissociation constants of each domain for 61 peptides that are physiologically tyrosine phosphorylated when EGFR and ERBB2–4 are activated[28]. The resulting systematic interaction maps showed that the binding complement differs between receptor dimers, with ERBB2 (also known as HER2)–ERBB3 having the most interactions, followed by EGFR–ERBB3>EGFR–ERBB2>EGFR–EGFR. This distribution correlates with the transforming potencies of ErbB dimers[29,30], indicating that an increase in interaction partners enhances oncogenicity. Interestingly, EGFR and ERBB2 but not ERBB3 became increasingly promiscuous when their expression level or activation state was increased, resulting in the binding of more proteins and the activation of more signalling pathways at higher concentrations of these receptors[28]. This result may explain why these receptors are often overexpressed in cancer. Comparison of the interactors identified in these two studies showed overlaps as well as differences; for example, signal transducer and activator of transcription 5 (STAT5) was found to bind only to EGFR and ERBB4 in the MS study. These differences could be due to different experimental systems or reflect differences in the technologies used in the two studies.

These studies also suggested that the biological specificity of receptor signalling is determined by the composition of the signalling complexes assembled by the receptor and the ligand-induced changes in its interactome. Implicit in this hypothesis is the assumption that changes in the receptor interactome are propagated throughout the network to affect downstream protein interactions and thereby dynamically shape the network topology used to transduce a signal. This theory was recently confirmed by analysis of the dynamic ERK interactome in the control of differentiation versus proliferation in PC12 cells[31]. The ERK pathway consists of a three-tiered kinase cascade, Raf–MEK–ERK and is a main effector of EGFR signalling (FIG. 1) which mediates many responses to EGF that are subverted in cancer, such as cell proliferation, transformation, differentiation, migration and survival. Stable isotope labelling with amino acids in cell culture (SILAC)-based quantitative proteomics was used to determine changes in the ERK interactome in response to nerve growth factor (NGF; which induces differentiation) and EGF (which induces proliferation). The results of this study increased the number of proteins identified in the ERK interactome from 170 proteins[32] to >280. Importantly, 60 protein interactions changed in a differentiation-dependent manner. Detailed analysis of a subset of ERK interactors showed that these interactors regulate the pathway at different steps, including the activation kinetics of ERK, crosstalk with other pathways and phosphorylation of transcription factors. These biochemical data were incorporated into a mathematical

© 2010 Macmillan Publishers Limited. All rights reserved

model, which revealed that the EGF and NGF pathways are under a distributed control mechanism rather than governed by a single master switch.

These results show that signalling pathways rely on protein interactions, which distribute control throughout the network, and that quantitative interaction proteomics is a formidable tool to map these pathways and their topologies. Furthermore, these findings predict that it is difficult to disrupt signalling networks by interventions targeting a single node, adding a mechanistic explanation to the clinical experience that the efficacy of single-agent cancer therapies is usually limited. These studies are complemented by the analysis of downstream phosphorylation events that we discuss further below.

The proteomic studies of the EGFR pathway demonstrate how signalling specificity is achieved through regulated protein interactions and how receptor overexpression can expand the number of interaction partners sufficiently to drive oncogenesis. This theme of oncogenesis, which results from the unwanted spread of signalling into many pathways owing to an expansion of the number of protein complexes assembled, is more clearly demonstrated when signal regulation is usurped by an oncoprotein. BCR–ABL1 is an oncogenic fusion protein that results from the chromosomal translocation that causes CML and inhibition of its kinase activity can reverse disease[33,34]. The translocation that results in *BCR–ABL1* is a rare case in which human tumorigenesis can be tied to a single genetic event based on molecular evidence and the unprecedented clinical success of the BCR–ABL1 inhibitor imatinib (Gleevec; Novartis)[5].

What causes this exception? Classical biochemical studies suggested that BCR–ABL1 is a constitutively active tyrosine kinase and 'super adaptor' that interacted with almost every pathway implicated in the oncogenic transformation of haematopoietic cells and fibroblasts[35] (FIG. 2). The challenge was to define which of these pathways is required for BCR–ABL1-induced transformation. An interaction proteome screen of BCR–ABL1-associated proteins revealed a network of several hundred proteins that act as direct or indirect interactors[36]. Evaluation of the highly connected hubs revealed only seven core components: GRB2, Src homology 2 domain-containing transforming protein 1 (SHC1), CRKL, CBL, p85 (also known as PIK3R2), suppressor of T cell receptor signalling 1 (STS1, also known as UBASH3B) and SH2 domain-containing inositol phosphatase 2 (SHIP2, also known as INPPL1), which coordinate a signalling network comprising several hundred proteins (FIG. 2). Interestingly, many of the core components of this network are scaffolds, suggesting that BCR–ABL1 signalling is orchestrated through propagating layers of protein complexes. BCR–ABL1 kinase inhibitors disrupted this network, leading the authors to conclude that the action of these drugs should be considered to interfere with the equilibrium state of an intricate network of protein complexes rather than just inhibit a single component. Given that several proteins interact with more than one of the seven core components, even the perturbation of a single interaction should have repercussions throughout the network, as it could shift the equilibrium composition of other complexes. This view may become important in the quest for alternative strategies that can circumvent resistance to tyrosine kinase inhibitors.

The big question is whether these findings from cultured cells translate to the clinic? This question was addressed mainly by using reverse-phase protein microarrays, on which lysates of tumour samples are arrayed and interrogated with antibodies. This strategy only allows targeted studies but can examine large numbers of patients and functions using the limited amounts

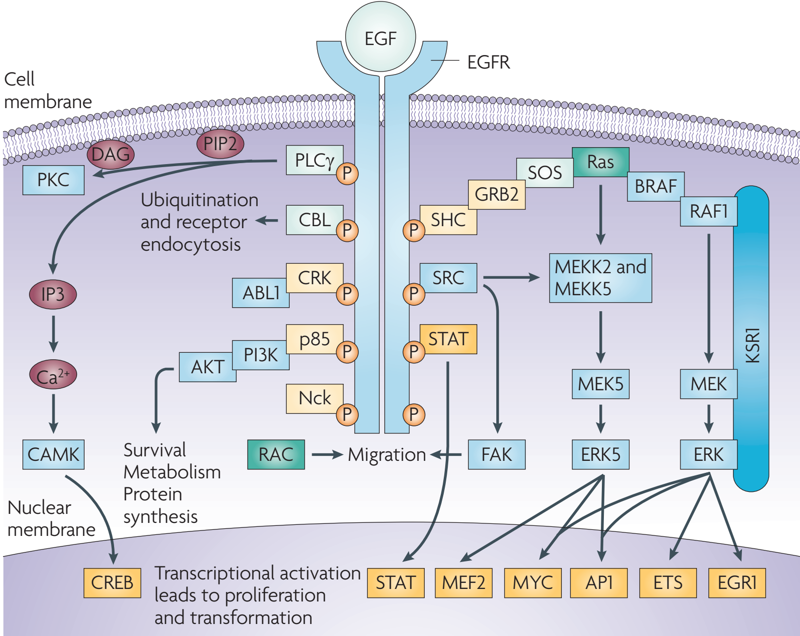


Figure 1 | **The epidermal growth factor receptor signalling network.** On activation (by ligand binding), the epidermal growth factor receptor (EGFR, also known as ERBB1) autophosphorylates tyrosine residues in its cytoplasmic domain, which serve as docking sites for the assembly of protein complexes that transduce EGF signals to generate specific biological responses. It is unlikely that the full complement of binding partners is recruited to each receptor and more plausible that individual EGFRs form complexes that have different compositions. In addition, all four members of the ErbB family (EGFR, ERBB2 (also known as HER2), ERBB3 and ERBB4) can form heterodimers (not shown), which share common binding proteins but, owing to the different representation of docking sites, assemble receptor complexes with different signalling properties[27]. The EGFR network has been reviewed in detail elsewhere[145,146] and only the better characterized downstream pathways are shown here. Kinases are light blue, scaffolds are dark blue, adaptor proteins are yellow, G proteins are green and transcription factors are orange. Phospholipase Cγ (PLCγ) is a phospholipase, CBL is an ubiquitin ligase, son of sevenless homologue (SOS) is a guanine nucleotide exchange factor. Small molecule second messengers are red: phosphatidylinositol-4,5-bisphosphate (PIP2), diacylglycerol (DAG), inositol-1,4,5-triphosphate (IP3) and Ca^{2+}. The downstream pathway discussed in this Review is the ERK pathway, which is initiated through the recruitment of a protein complex that contains the Src homology 2 domain-containing (SHC) and growth factor receptor-bound protein 2 (GRB2) adaptor proteins and the exchange factor SOS to activated EGFR. SOS exchanges GDP for GTP on Ras, which induces a conformational change that enables Ras to bind Raf kinases with high affinity. Raf activation is a complex process, involving dephosphorylation, phosphorylation, homo- and heterodimerization and binding to the kinase suppressor of ras 1 (KSR1) scaffold[147–149]. Activated Raf phosphorylates and activates MEK, which in turn phosphorylates and activates ERK. ERK exerts its different biological effects through many substrates[32], including transcription factors. AP1, activator protein 1; CAMK, calcium/calmodulin-dependent protein kinase; CREB, cyclic AMP-responsive element-binding protein; EGR1, early growth response 1; FAK, focal adhesion kinase; MEF2, myocyte enhancer factor 2; PKC, protein kinase C; STAT, signal transducer and activator of transcription.

© 2010 Macmillan Publishers Limited. All rights reserved

Stable isotope labelling with amino acids in cell culture (SILAC)
This method involves the *in vivo* metabolic labelling of samples with amino acids that carry stable (non-radioactive) heavy isotope substitutions of atoms which, when analysed by MS, produce 'conjugated' peptide peaks. These peaks originate from the same protein but show a characteristic mass shift which corresponds to the mass difference between the light and heavy label. The relative intensity of conjugated peak pairs provides the relative abundance of a protein in the two samples.

Node
This describes an object in graph form, and the connections between objects are termed edges. In signalling networks, nodes represent proteins (or genes, if they are based on genetic information) and edges represent the relationship between the nodes, such as binding, regulation or modification.

of tissue, such as tumour biopsies, that are typically available in clinical settings. Such studies have yielded clinically useful information and are discussed in detail by Brennan *et al.* in this Focus issue[37].

Protein scaffolds. Having made the point that protein interactions are important, the question is what holds them together? Through simultaneously binding two or more client proteins, scaffold proteins generate platforms that alter signalling kinetics, control crosstalk between signalling pathways and insulate elements from each other. Although the role of scaffolds in cancer still needs to be determined, we briefly discuss them as they play fundamental parts in regulating both the steady-state and acute-response kinetics of many pathways involved in cancer. Scaffolds seem to be most relevant to mechanisms of non-oncogene addiction[22], whereby normal components of signalling networks support the aberrant components. Aberrations in ERK signalling are common in cancer[38] and hence we discuss two ERK pathway scaffolds.

The kinase suppressor of ras (KSR) proteins scaffold the three-tiered Raf–MEK–ERK module and contribute to the regulation of its signalling dynamics and spatio-temporal control. KSR1 is not essential for, but enhances ERK activation[39]. Importantly, knocking out *Ksr1* restrains transformation by oncogenic HRAS-G12V in cultured cells[40] and HRAS-G12V-driven skin carcinogenesis in mice[41]. KSR1 constitutively interacts with MEK and the KSR1–MEK complex translocates to the plasma membrane on mitogen stimulation[42]. MEK bound to KSR1 also promotes the recruitment of BRAF, which enhances MEK phosphorylation by BRAF. Interestingly, ERK binding induces feedback phosphorylation of KSR1 and BRAF by ERK, which promotes the dissociation of BRAF and KSR1 and their release from the plasma membrane into the cytosol. It is unclear whether this is a purely negative feedback pathway that limits ERK activation, or whether this feedback also has a role in ensuring the turnover of client proteins on the scaffold and in signal localization. KSR1 also directly allosterically activates Raf kinases by forming side-to-side dimers[43]. KSR1 selectively interacts with ERK dimers and confines ERK signalling to cytosolic substrates, and the expression of ERK point mutants that prevent ERK dimerization counteracts cellular transformation *in vitro* and tumour development in mouse xenograft models[44]. Thus, KSR1 can control several aspects of ERK signalling, including substrate specificity and spatio-temporal activation kinetics.

Recently, the KSR2 interactome was studied using quantitative MS-based proteomics to compare proteins that bind to KSR2 in the presence and absence of tumour necrosis factor-α (TNFα) stimulation[45]. Of >100 proteins potentially in the complex, approximately 40 were recruited on TNFα stimulation, highlighting the highly dynamic properties of this signalling complex. Interestingly, KSR2 recruited ARAF rather than RAF1 or BRAF. ARAF is a poor activator of MEK and its main, kinase-independent, function is as an inhibitor of the pro-apoptotic mammalian STE20-like protein kinase 2 (MST2, also known as STK3) pathway[46], indicating that

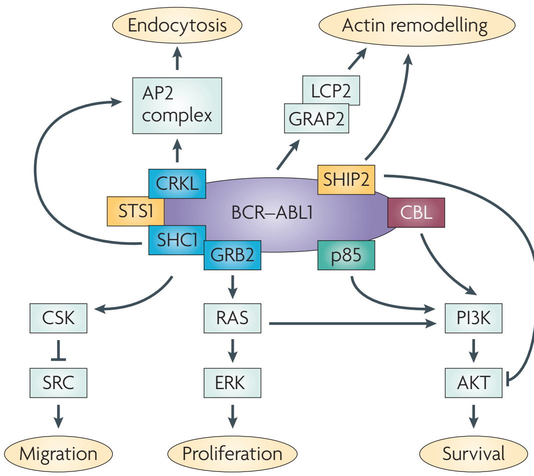


Figure 2 | **The BCR–ABL1 signalling network.** Breakpoint cluster region (BCR)–ABL1 activates downstream signalling pathways through a core network of seven closely associated proteins[36]. They comprise adaptor proteins (blue), phosphatases (orange) and the regulatory p85 subunit of PI3K (green). Selected downstream signalling pathways are depicted in light green and biological outcomes in yellow. Most biological effects are regulated through multiple pathways; for example, actin polymerization is regulated through the growth factor receptor-bound protein 2 (GRB2)-related adaptor protein 2 (GRAP2, also known as GADS)–lymphocyte cytosolic protein 2 (LCP2, also known as SLP76)–Nck adaptor pathway[150] or through SH2 domain-containing inositol phosphatase 2 (SHIP2) regulation of phosphatidylinositol-3,4,5-trisphosphate levels[151]. The scheme is highly simplified for clarity. Arrows do not imply direct connections and may summarize several steps. The figure is based on data from Brehme *et al.*[36]. AP2, adaptor protein complex 2; CSK, c-src tyrosine kinase; SHC1, Src homology 2 domain-containing 1; STS1, suppressor of T cell receptor signalling 1.

KSR2 may have a role in redirecting ARAF to the ERK pathway. As scaffolding vastly accelerates reaction rates by co-localizing enzymes and substrates[47], KSR2-bound ARAF is expected to be an effective MEK activator. Thus, this shift of ARAF from its anti-apoptotic function in the MST2 pathway to the mitogenic ERK pathway may contribute to oncogenic transformation.

A detailed comparison of the KSR1 and KSR2 interactomes showed the similarity of these scaffolds in promoting RTK-mediated ERK signalling but also identified specific interactions. For example, the calcium-dependent phosphatase calcineurin only binds to KSR2 (REF. 48). In response to calcium signals, calcineurin dephosphorylates the 14-3-3 binding sites on KSR2, which allows KSR2 to be recruited to the plasma membrane and activation of the ERK pathway[48]. Another proteomic study showed that KSR2 binds to AMP kinase (AMPK, also known as PRKAB1), mediating its stimulatory effects on glucose uptake and fatty acid oxidation[49]. This observation raises an interesting hypothesis: that KSR2 may coordinate links between mitogenic signalling and metabolic pathways by binding and coordinating the functions of ARAF and AMPK. ARAF can bind and inhibit

© 2010 Macmillan Publishers Limited. All rights reserved

Warburg effect
This was named after a discovery made by the German biochemist Otto Warburg in the 1920s that cancer cells predominantly use anaerobic glycolysis rather than oxidative phosphorylation, even when oxygen is abundant. As a result, pyruvate is converted to lactate instead of being oxidized by the mitochondria of cancer cells.

O-linked N-acetylglucosamine acylation (O-GlcNAcylation)
A form of glycosylation found in nuclear and cytosolic proteins in which O-GlcNAc is added to the hydroxyl groups of serine and threonine residues that can also serve as phosphorylation sites.

Isobaric tag for relative and absolute quantitation (iTRAQ)
A stable isotope labelling method for the quantitation of peptides by MS, in which a molecule containing normal or heavy isotopes is used to chemically modify the proteins or peptides from each individual sample. The fragmentation of the labelled molecules then gives rise to specific reporter ions that can be used to measure the relative amounts of each protein present in each sample.

pyruvate kinase M2 (PKM2)[50]. In many cancers, PKM2 is aberrantly expressed in its low-activity dimeric form, which favours aerobic glycolysis — the hallmark of the Warburg effect[51]. Thus, KSR2 could enhance AMPK-mediated glucose uptake and PKM2 inhibition by ARAF, thereby promoting the Warburg effect and tumour growth. Consequently, KSR2 downregulation should ameliorate the Warburg effect and slow tumour growth, a prediction that is easily testable as *Ksr2*$^{-/-}$ mice are available[49].

The above examples illustrate the key role of protein complexes in signalling and equally highlight the importance of dynamic changes in the interactions that ensure the proper interpretation of signals and responses to signals. Conceptually, this view has important implications. They include the notion that connections in signalling pathways are not hardwired, that a protein with a defined biochemical function can adopt different biological functionalities as part of different protein complexes and that the kinetics of signalling specificity intimately involve dynamic changes in protein interactions. How are connections in signalling pathways brought about? Undoubtedly, PTMs are a salient part of the answer. The specific roles of PTMs in determining protein complex formation are just emerging but many proteomic techniques for PTM analysis are available. In the following section, we briefly review how these technologies have been applied to study dynamic processes in cell signalling.

Post-translational modifications

PTMs are widely studied dynamic processes in cell signalling, which is not surprising considering the fundamental roles they have in regulating signalling pathways, and many studies have mapped cancer-associated PTMs. Phosphorylation of serine, threonine and tyrosine residues have received the most attention from the proteomics community, although ubiquitylation, acylation and glycosylation are coming into the limelight. O-linked N-acetylglucosamine acylation (O-GlcNAcylation) is also receiving considerable interest, mainly owing to its competing role in regulating phosphorylation by blocking potential phosphorylation sites.

Phosphorylation. Quantitative methods for the study of protein phosphorylation have been described elsewhere[13,52–56]. The first phosphoproteome to be extensively studied was that of EGFR activation (reviewed by Blagoev[24]) and several other RTK pathways have now been described. Global tyrosine phosphorylation is usually mapped using enrichment with anti-phosphotyrosine antibodies and subsequent MS analysis. Using this method, time-course studies of EGF-induced tyrosine phosphorylation and comparisons of signalling between different RTKs were performed[57]. For example, SILAC-based quantitative proteomics was used to compare the tyrosine phosphorylation induced by EGF versus that induced by platelet-derived growth factor (PDGF) in human mesenchymal stem cells[58]. EGF induces osteogenic differentiation whereas PDGF sustains proliferation and migration[59]. Most tyrosine phosphorylation events were common to both growth factors, except the PI3K pathway, which was exclusively stimulated by PDGF. Inhibiting PI3K converted PDGF to a differentiation factor, elegantly demonstrating that biological specificity of RTK signalling is encoded by distinct biochemical differences in downstream signalling pathways that arise from the protein complexes assembled at RTKs. A more recent study that analysed the tyrosine phosphorylation induced by EGF in human mammary epithelial cells revealed significantly different stoichiometries of phosphorylation at different sites in the same protein[60], suggesting intricate temporal and dynamic regulation of phosphorylation.

Quantitative proteomics was also successfully used to map the tyrosine phosphorylation network induced by cancer-associated EGFR mutants. Glioblastoma (or astrocytoma grade IV) is a highly malignant brain tumour and often expresses a truncated EGFR mutant, EGFRvIII, which is expressed from an amplified gene locus and enhances malignant behaviour in a dose-dependent manner[61]. Isobaric tag for relative and absolute quantitation (iTRAQ) analysis of glioblastoma cell lines that expressed different levels of EGFRvIII or a kinase-dead EGFRvIII as a control revealed that increasing EGFRvIII expression shifts signalling from ERK and STAT3 towards the PI3K pathway and also induces phosphorylation and transactivation of the hepatocyte growth factor receptor MET[62]. Consequently, the combination of EGF and MET inhibitors synergized to kill glioblastoma cells *in vitro*, producing a rational drug combination based on phosphoproteomics data.

Another SILAC-based proteomics study in non-small-cell lung cancer (NSCLC) cell lines compared phosphotyrosine signalling induced by wild-type EGFR, mutation-activated EGFR and overexpressed MET[63]. It revealed networks that extensively overlapped in the regulation of cell adhesion, motility, proliferation and survival. Interestingly, mutant EGFRs preferentially induced phosphorylation of cytoskeletal proteins, scaffolds implicated in cytoskeletal regulation and motility, and negative feedback inhibitors of EGFR signalling, such as sprouty 1 (SPRY1), SPRY2, SPRY4 and mitogen-inducible gene 6 (MIG6, also known as ERRFI1). The functional effects of these phosphorylation events are largely unknown but it is tempting to speculate that mutant EGFRs can sabotage the function of physiological feedback inhibitors. This hypothesis is consistent with observations that EGFR feedback inhibitors are often downregulated in cancers[64]. Crosstalk between MET and mutant EGFRs provided mutual activation of the respective receptors and downstream signalling pathways[63]. Interestingly, the sensitivity to EGFR and MET inhibitors as measured by biochemical outputs correlated with the extent of control of the respective receptors over the downstream network. This puts an interesting twist on the observation discussed above that EGFR oncogenicity increases with the number of pathways it engages[28–30], as such functional expansion should also increase the vulnerability of the downstream network to receptor inhibition. The available clinical experience with EGFR inhibitors shows that spectacular regressions are achievable in patients with NSCLC but this only occurs in a small subset of patients and does not have lasting effects[65].

Does this reflect an arms race between the increasing fragility and adaptive potential of RTK networks, which is afforded by the recruitment of an increasing number

© 2010 Macmillan Publishers Limited. All rights reserved

Immobilized metal ion affinity chromatography (IMAC)
A method for the enrichment of phosphopeptides that exploits the propensity of metal ions such as iron and gallium to bind phosphate groups.

of downstream effectors? This hypothesis suggests a counterintuitive approach of combining EGFR inhibitors with inhibitors of downstream pathways to boost drug efficacy and prevent resistance. The combined inhibition of a receptor and a downstream effector could offer simple, general and hence attractive guidelines for the design of combination therapies and it seems worthwhile to explore this possibility further.

A large-scale phosphotyrosine proteomic screen using 41 NSCLC cell lines and >150 NSCLC tumours found many RTKs, including EGFR and MET, and non-RTKs were activated in NSCLC[66]. The contribution of most of these kinases to NSCLC pathogenesis is unknown but an important prediction of this study is that the drug responsiveness of individual tumours should be highly dependent on whether these activated kinases function as part of hierarchical networks or in independent pathways. If they form hierarchical networks, inhibition of the master kinase will be therapeutically effective, as demonstrated by the spectacular efficacies of BCR–ABL1 inhibitors in CML. If these kinases function in independent networks or as part of redundant topologies, combination therapy seems necessary. Therefore, for personalized therapy, it will be important to map kinase activation profiles to network connectivity maps that control specific biological responses. This idea has also been adopted by the pharmaceutical industry, especially for kinase inhibitor screening (BOX 2).

The idea of correlating phosphotryosine networks with specific biological behaviours was also borne out in studies on the effects of human ERBB2 in human mammary epithelial cells (HMECs)[67,68]. ERBB2 is overexpressed in ~25% of breast cancers and strongly correlates with poor prognosis. Although ERBB2 lacks a known ligand, it can heterodimerize with EGFR and alter EGFR signalling. Quantitative MS based on iTRAQ labelling was used to map phosphotyrosine networks in parental HMECs and HMECs engineered to overexpress ERBB2 that were stimulated with EGF or heregulin (also known as neuregulin 1)[67]. The results show that in ERBB2-overexpressing cells, EGF stimulates migration through multiple signalling pathways, whereas heregulin uses a subset of the migration network. The application of regression-based computational modelling identified combinations of phosphorylation sites that correlate with proliferation and migration. Subsequent refinement of the model[68] allowed the identification of phosphotyrosine network elements that differentially control migration and proliferation. This culminated in the description of nine phosphorylation sites on six proteins that were associated with the PI3K pathway and endocytosis. These sites served as a 'network gauge' that captured the predictive capability of the full model. This result also elegantly demonstrates the analytical power added by computational modelling to the interpretation of complex dynamic data sets.

Phosphoserine and phosphothreonine antibodies are less well developed than phosphotyrosine antibodies, and most global phosphoproteome studies use chromatographic enrichment technologies. These usually exploit the propensity of TiO_2 or Fe^{3+} to interact with phosphate groups and are often combined with a preceding separation step that counterselects phosphopeptides such as strong cation exchange columns[69]. Chromatographic enrichment approaches are available in several variations, which enrich overlapping but distinct fractions of the phosphoproteome[70]. It is now possible to identify 10,000–20,000 phosphorylation sites. For example, proteomic analysis of phosphorylation induced by the fms-related tyrosine kinase 3 (FLT3)-internal tandem duplication (ITD) mutant tyrosine kinase quantified >12,000 phosphorylation sites and demonstrated subcellular localization-specific signalling of FLT3 (REF. 71). FLT3-ITD consists of in-frame tandem duplications of the juxtamembrane domain of FLT3 and is the most frequent oncogenic FLT3 mutation in acute myeloid leukaemia (AML), causing constitutive activation of FLT3 and retardation of trafficking through the endoplasmic reticulum (ER). When localized at the ER, FLT3-ITD aberrantly activates STAT5 signalling and fails to activate the PI3K and ERK pathways that are targets of ligand-activated wild-type FLT3. By contrast, when located at the membrane, FLT3-ITD preferentially activates PI3K and ERK over STAT5. These pathway-specific effects were reflected in the global phosphoproteome and also helped identify putative downstream targets of AKT and proviral integration site (PIM) kinases.

An interesting approach to functional studies is to combine global phosphoproteomics with small interfering RNA (siRNA) knockdown studies. An example is the quantitative proteomic assessment of early signalling events in integrin signalling[72]. Integrin-mediated cell adhesion and anti-apoptotic signalling is essential for cancer cell spread and invasion. SILAC combined with immobilized metal ion affinity chromatography (IMAC)

Box 2 | Kinase profiling and kinomics

Kinase profiling refers to the process of screening kinase inhibitors for activity and selectivity[129]. Given the part that aberrant kinase activity plays in cancer and the huge efforts to study kinases, their substrates and inhibitors, the increase in the number of kinase inhibitors being translated into anticancer drugs in recent years is not surprising (it is estimated that around one-third of all major clinical trials currently involve kinase inhibitors). However, the first kinase inhibitor (imatinib) only received clinical approval in 2001. Analysis of the human genome initially predicted the existence of 518 kinases[130], most of which have now been validated and >400 of these are currently available in various screens. However, this is unlikely to be the full story, as new kinases are being identified and tissue-specific roles of kinases are being elucidated[131]. Methods to identify kinases generally rely on the use of peptide substrates in microtitre plate, bead or array format that can be used to measure the activity of a set of purified kinases, although with recent improvements[132–135], methods are getting closer to native *in vivo* conditions. Miniaturization through microfluidics and new detection methods are greatly improving sensitivity and reducing costs. The benefits of these approaches are obvious: they are highly parallel, so many compounds and kinases can be tested, and quantitative analysis allows the crucial question of off-target activity to be addressed.

Kinomics[130] is formally the identification of kinases at the genetic, or preferably proteomic, level. However, this term should be expanded to include the challenging task of identifying kinase targets and elucidating the cellular pathways and networks in which they are involved and the complex regulation of these kinases. Proteomics approaches coupled to modelling through systems biology methods are in the forefront of the techniques being applied. A substantial remaining challenge is the identification of the downstream effector kinases that are the best drug targets. Furthermore, chemical proteomics is a valuable tool to identify on- and off-target interactions of individual kinase inhibitors[136].

© 2010 Macmillan Publishers Limited. All rights reserved

Electron transfer dissociation (ETD)
A recently introduced MS method for the fragmentation of molecules by transferring electrons from anion radicals to positively charged ions. It is a non-ergodic (rapid, kinetically controlled) process, so energy is not redistributed and many bonds are broken in the molecule, not just the weakest ones as seen in collision-induced dissociation.

WD40 domain
A protein domain consisting of 4–16 repeats of an approximately 40 amino acid-long motif that ends with a W–D (tryptophan–aspartic acid) dipeptide. The WD40 domains form a circular β-sheet propeller structure that serves as a structural platform for protein interactions and the specificity of the interactions is determined by sequences outside the WD repeats.

enrichment showed that >500 phosphorylation sites in 357 proteins changed when integrins interacted with collagen, a common integrin ligand in the extracellular matrix. siRNA screens against 33 of these proteins with kinase or phosphatase activity identified three integrin-regulated kinases, p21-activated kinase 2 (PAK2), G protein-coupled receptor kinase 6 (GRK6) and DBF4, which are important in cell migration. Recent evidence suggests that these proteins are deregulated in cancer. PAK2 is hyperphosphorylated in ovarian cancer[73], and a siRNA screen identified that GRK6 was required for the viability of myeloma cells but not normal cells[74]. DBF4 is a regulatory subunit of the cell cycle kinase CDC7 and this complex is overexpressed in various cancer cell lines and human cancers[75].

The focus on phosphoproteomics in recent years and the application of functional proteomics methods[76] led to the identification of O-GlcNAcylation as an important PTM in signalling. As O-GlcNAcylation occurs at serine and threonine residues that are also targets for phosphorylation, they mutually exclude each other, resulting in dynamic crosstalk between phosphorylation and O-GlcNAcylation[77]. Although many of the (patho)physiological consequences of this crosstalk remain to be explored, it has been implicated in several diseases, including cancer[78], and demonstrated to be crucial for cellular processes that are important in cancer biology such as cytokinesis[79]. The application of improved MS techniques (such as electron transfer dissociation (ETD)) for detecting O-GlcNAcylation[80] or the parallel detection of phosphorylation and glycosylation[81] will likely increase attention to this field.

Figure 3 | **The ubiquitin system.** Ubiquitin (Ub) is a small protein that is transferred as a versatile regulatory post-translational modification (PTM) to target proteins in a three-step process[152]. In the first step, an E1 ubiquitin-activating enzyme uses ATP to form a covalent thioester with ubiquitin. E1 then transfers the ubiquitin moiety onto an ubiquitin-conjugating enzyme E2. E2 associates with one of several hundred E3 ubiquitin ligases which provide specificity by also binding the target protein. The E2-binding and target recognition domains can reside in one protein (as shown) or in different proteins that are part of a ligase complex. Finally, the ubiquitin moiety is transferred to a lysine residue in the target protein directly or through an E3 ubiquitin ligase. Further ubiquitins can be added to lysines in ubiquitin itself. Depending on the number of ubiquitin moieties transferred and the linkage type, ubiquitylation can have different functions that are often derailed in cancer[83] and many other diseases[153].

Ubiquitylation. Ubiquitylation and the proteasomal pathway have key functions in oncogenesis and cancer[82–88] (FIG. 3), typified by the tumour suppressor p53 (REFS 89,90), and are being investigated as pharmaceutical targets[91–94]. Since the early use of functional proteomics to identify the structure and substrate specificity for the S phase kinase-associated protein 1 (SKP1)–cullin 1 (CUL1)–F-box (SCF) complex[95,96], this method has continued to have an important role in the study of ubiquitylation, and in identifying the components of ligase complexes and those involved in the ubiquitylation of proteins[97]. Recent studies have also focused on the non-classical ligases such as the damage-specific DNA-binding protein 1 (DDB1)–CUL4A–regulator of cullins 1 (ROC1, also known as RBX1) E3 ubiquitin ligase, a complex that lacks a SKP1-like adaptor and is involved in the regulation of DNA repair[98]. TAP affinity purification of the complex and subsequent MS analysis were used to identify a novel family of 16 WD40 domain-containing proteins that recruit the substrate to the complex[99]. This novel family of proteins was called the DDB1–CUL4A-associated WD40 domain proteins (DCAF proteins) and includes DDB2, a protein mutated in the cancer-predisposing syndrome xeroderma pigmentosum. However, given the connection of this complex with DNA repair, it seems likely that it has a much broader role in cancer and in sensitivity of cells to DNA-damaging treatments. Methods are continually being improved for the global identification of ubiquitylation sites[100] and to identify substrates in specific pathways[101]. In addition to MS approaches, protein arrays are valuable tools for identifying substrates for PTMs. As these assays are done *in vitro*, versatile manipulations and comparisons are possible. Protein microarrays displaying >8,000 human proteins were exposed to cell extracts that replicate the mitosis checkpoint and anaphase release to identify targets of the anaphase-promoting complex (APC)[102]. The APC is a multiprotein complex with E3 ubiquitin ligase activity that promotes cell cycle progression during mitosis. Its substrates include mitotic cyclins and many other proteins that orchestrate this cell cycle transition. The array-based ubiquitylation assay compared extracts that contained inhibited and activated APC, detecting most of the known substrates and seven potential new substrates. A similar experiment was used to identify and compare substrates for NEDD4 and NEDD4L[103]. In addition to the shared substrates identified, NEDD4 showed a preference for tyrosine kinases, and NEDD4 knockdown sustained tyrosine kinase signalling. NEDD4 family members regulate various pathways and cellular functions that are implicated in cancer, such as many RTK pathways and transcription factors, and are found overexpressed in breast, prostate and pancreatic cancers[104].

This brief description of PTM aberrations in cancer is far from complete and there are many other PTMs that are altered in cancer (BOX 3). However, PTMs are

© 2010 Macmillan Publishers Limited. All rights reserved

emerging as important drug targets. The high hopes and resources invested in kinase inhibitor drugs are inextricably linked to their ability to prevent regulatory phosphorylations of crucial substrates. We soon may witness similar prolific efforts to target other PTMs for cancer therapy such as ubiquitylation[92]. Notably, the validation of PTM targets is enabled by proteomics. PTMs are largely undetectable in genomic studies and to fully exploit their therapeutic potential, it will be necessary to understand the dynamics of PTM turnover. It is tempting to speculate that PTM dynamics also contribute to the well-known heterogeneity in cancer. Dynamic changes in protein expression and localization can determine the drug sensitivity of cancer cells[105], and dynamic changes in PTMs could vastly expand the window of opportunities of cancer cells to escape adverse conditions and therapy.

Towards single-cell proteomics

New techniques enabling the functional analysis of individual cells are increasingly being used to address the heterogeneity of cancer cells, especially in terms of drug resistance. One of the first methods was Phospho-Flow, invented by the Nolan group[98]. This technique uses phospho-specific antibodies to label antigens in fixed cells. Subsequent analysis by flow cytometry provides data at the single-cell and population levels that allow the investigation of heterogeneity and are applicable to the modelling of cancer signalling pathways[106]. More recently, Cohen *et al.* tagged ~1,000 individual endogenous proteins in lung cancer cells with a fluorophore and monitored their expression and subcellular location by automated time-lapse imaging[105]. Treatment with the topoisomerase 1 inhibitor camptothecin caused the rapid relocation of proteins that are associated with the mechanism of drug action and slower changes in protein abundance. Interestingly, although most drug-induced responses were similar in individual cells across the population, a subset of 24 proteins showed high cell-to-cell variability 1 day after drug exposure. The upregulation of two of these proteins, the RNA helicase DEAD box polypeptide 5 (DDX5) and replication factor C1 (RFC1), correlated with drug resistance and cell survival, showing that escape mechanisms can result from changes in protein dynamics that are unique to a subset of a cell population. This is an intriguing proposition for the development of drug resistance that warrants further investigation in preclinical cancer models.

A different approach used a microfluidic platform that traps ~1,000 individual cells and analyses their behaviour using imaging, optical indicator dyes and fluorescent antibodies[4]. This platform was used to assess the signalling dynamics of normal human haematopoietic stem cells (HSCs) and CML stem cells (CML-SCs) in response to the second-generation BCR–ABL1 inhibitor dasatinib (Sprycel; Bristol–Myers Squibb). Although significant differences in the responses of individual cells were observed, at the population level dasatinib was more cytotoxic to HSCs than CML-SCs, and strongly and selectively inhibited migration of CML-SCs. Thus, single-cell analysis can reveal important information on divergent and synchronized behaviour within cell populations, especially in regard to drug action. It would be fascinating to combine these approaches to collect more detailed information on cancer cell heterogeneity.

Perspective

Although this Review discusses only a small fraction of studies, it illustrates the power of functional proteomics in mapping cancer signalling pathways. The biggest

Box 3 | **Other post-translational modifications**

Protein function is crucially regulated by post-translational modifications (PTMs) and the role of PTM aberrations in the faulty regulation of cancer signalling networks is becoming increasingly appreciated. Proteomics is currently expanding its scope beyond phosphorylation, delivering exciting results on other PTMs, which are described below.

Acetylation and methylation

The regulation of gene expression by the acetylation and methylation of lysines and arginines in histones is well established, and the term the 'histone code' highlights the importance of this epigenetic control[137]. Silencing of tumour suppressor gene expression owing to subversion of the histone code is well known[137]. However, proteomics has enabled a systematic study of these histone modifications[138]. Importantly, proteomics identified many other chromatin-independent targets. For example, stable isotope labelling with amino acids in cell culture (SILAC) analysis of breast cancer cells treated with vorinostat (Zolinza; Merck), a histone deacetylase (HDAC) inhibitor, revealed changes in the expression of transcription factors and metabolic, structural, chaperone and cell cycle proteins[139]. HDAC6 was identified as an inhibitor of EGFR endocytosis by deacetylating microtubules[140].

Acylation

A well-characterized example of acylation that is relevant to cancer is the modification of Ras proteins by farnesylation, geranylation and palmitoylation, which localizes Ras to distinct membrane compartments[141]. As Ras mutations are frequent in human cancers, the Ras-modifying acylases became major drug targets, unfortunately with disappointing clinical results[142]. Much effort was invested to understand the opportunities and failures of clinically targeting Ras acylases, but quantitative proteomic techniques were only recently developed to identify palmitoylated proteins[143].

Oxidative modifications and cysteine nitrosylation

These modifications have also been shown to play an important part in signalling[144]. They are likely to have a role in cancer owing to the hypoxic and inflammatory environment associated with cancer, but systematic proteomics studies are lagging behind.

Cyclin
A regulatory subunit that is essential for the activity of cell cycle-dependent kinases (CDKs). Its name derives from the periodic expression of cyclins during the cell cycle, which is due to the regulated degradation by the ubiquitin–proteasome system that is thought to drive the cell cycle.

© 2010 Macmillan Publishers Limited. All rights reserved

Micro-engineering
The use of micro-fabricated devices that have small (micron)-scale features (such as channels, wells and vessels) to allow the processing of small volumes of fluid.

challenge is to bring proteomics technologies to clinical applications. Despite impressive progress, especially in biomarker discovery, substantial hurdles remain. Most current proteomics technologies are too slow, too complex and too expensive to be used in the clinical laboratory, and the existence of many different experimental approaches leaves a deficit in standardization. However, further innovations that may overcome these barriers are on the horizon. Although MS is dominating the field, alternative technologies are appearing. Great progress is to be expected from array-based methods, optical methods and micro-engineering approaches. Although these methods cannot identify proteins *de novo*, their attractiveness lies in their multiplexing capabilities, throughput and prospect of reducing analytical costs. In particular, methods based on flow cytometry, protein and tissue microarrays and micro-engineered devices are amenable to adaptation to the clinical laboratory.

Our capability to routinely identify or test thousands of proteins per analysis has shifted the bottleneck in proteomics towards data analysis and interpretation. The increasing use of advanced bioinformatics and systems biology tools is beginning to unlock network properties[107]. Recent examples are the development of a predictor for breast cancer prognosis based on the modularity of protein interaction networks[108] and the identification of cancer-associated phosphorylation networks through the combined alignment of conserved phosphorylation sites and kinase–substrate networks[109]. This latter study also showed that we still miss large parts of phosphorylation networks.

As we only begin to apply proteomics to unravel the role of other PTMs, such as ubiquitylation, sumoylation and glycosylation, in signalling we can expect a further stream of data. It will be a challenging and informative task to understand the regulatory interplay between the different PTMs both at the level of individual proteins and in networks. In addition, these investigations may provide molecular alterations that can serve as biomarkers for cancer and thereby bridge the current gap between mechanistic understanding and mainly phenomenological markers.

1. Wolf-Yadlin, A., Sevecka, M. & MacBeath, G. Dissecting protein function and signaling using protein microarrays. *Curr. Opin. Chem. Biol.* **13**, 398–405 (2009).
2. Schulz, K. R., Danna, E. A., Krutzik, P. O. & Nolan, G. P. Single-cell phospho-protein analysis by flow cytometry. *Curr. Protoc. Immunol.* **78**, 8.17.1–8.17.20 (2007).
3. Fournier, F. *et al.* Biological and biomedical applications of two-dimensional vibrational spectroscopy: proteomics, imaging, and structural analysis. *Acc. Chem. Res.* **42**, 1322–1231 (2009).
4. Faley, S. L. *et al.* Microfluidic single cell arrays to interrogate signalling dynamics of individual, patient-derived hematopoietic stem cells. *Lab Chip* **9**, 2659–2664 (2009).
5. Melo, J. V. & Barnes, D. J. Chronic myeloid leukaemia as a model of disease evolution in human cancer. *Nature Rev. Cancer* **7**, 441–453 (2007).
6. Burckstummer, T. *et al.* An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nature Methods* **3**, 1013–1019 (2006).
7. Gavin, A. C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
8. Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
References 7 and 8 are landmark papers showing the mapping of the yeast interactome by MS-based proteomics.
9. Bublil, E. M. & Yarden, Y. The EGF receptor family: spearheading a merger of signaling and therapeutics. *Curr. Opin. Cell Biol.* **19**, 124–134 (2007).
10. Hammond, D. E. *et al.* Quantitative analysis of HGF and EGF-dependent phosphotyrosine signaling networks. *J. Proteome Res.* **9**, 2734–2742 (2010).
11. Gordus, A. *et al.* Linear combinations of docking affinities explain quantitative differences in RTK signaling. *Mol. Syst. Biol.* **5**, 235 (2009).
12. Mitsudomi, T. & Yatabe, Y. Epidermal growth factor receptor in relation to tumor development: *EGFR* gene and cancer. *FEBS J.* **277**, 301–308 (2010).
13. Dengjel, J., Kratchmarova, I. & Blagoev, B. Receptor tyrosine kinase signaling: a view from quantitative proteomics. *Mol. Biosyst.* **5**, 1112–1121 (2009).
14. von Kriegsheim, A., Preisinger, C. & Kolch, W. Mapping of signaling pathways by functional interaction proteomics. *Methods Mol. Biol.* **484**, 177–192 (2008).
15. Preisinger, C., von Kriegsheim, A., Matallanas, D. & Kolch, W. Proteomics and phosphoproteomics for the mapping of cellular signalling networks. *Proteomics* **8**, 4402–4415 (2008).
16. Ong, S.-E. & Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nature Chem. Biol.* **1**, 252–262 (2005).
17. Blagoev, B. & Mann, M. Quantitative proteomics to study mitogen-activated protein kinases. *Methods* **40**, 243–250 (2006).
18. Cheng, X. Understanding signal transduction through functional proteomics. *Expert Rev. Proteomics* **2**, 103–116 (2005).
19. Andersen, J. S. *et al.* Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* **426**, 570–574 (2003).
20. Amit, I., Wides, R. & Yarden, Y. Evolvable signaling networks of receptor tyrosine kinases: relevance of robustness to malignancy and to cancer therapy. *Mol. Syst. Biol.* **3**, 151 (2007).
21. Papanikolaou, N. A. & Papavassiliou, A. G. Protein complex, gene, and regulatory modules in cancer heterogeneity. *Mol. Med.* **14**, 543–545 (2008).
22. Luo, J., Solimini, N. L. & Elledge, S. J. Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell* **136**, 823–837 (2009).
23. Henson, E. S. & Gibson, S. B. in *Signal Transduction: Pathways, Mechanisms and Diseases* (ed. Sitaramayya, A.) 119–141 (Springer-Verlag, New York, 2010).
24. Blagoev, B., Kratchmarova, I., Olsen, J. V. & Mann, M. in *EGFR Signaling Networks in Cancer Therapy* (eds Haley, J. D. & Gullick, W. J.) 190–198 (Springer-Verlag, New York, 2008).
25. Kholodenko, B. N. Cell-signalling dynamics in time and space. *Nature Rev. Mol. Cell Biol.* **7**, 165–176 (2006).
26. Wiley, H. S., Shvartsman, S. Y. & Lauffenburger, D. A. Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell Biol.* **13**, 43–50 (2003).
27. Schulze, W. X., Deng, L. & Mann, M. Phosphotyrosine interactome of the ErbB-receptor kinase family. *Mol. Syst. Biol.* **1**, 2005.0008 (2005).
This paper describes an exhaustive, quantitative MS-based proteomics strategy to functionally map the protein interactions that mediate ErbB family signalling.
28. Jones, R. B., Gordus, A., Krall, J. A. & MacBeath, G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* **439**, 168–174 (2006).
Similar to Reference 27, this paper describes the use of protein microarrays for the functional mapping of protein interactions in the ErbB family.
29. Alroy, I. & Yarden, Y. The ErbB signaling network in embryogenesis and oncogenesis: signal diversification through combinatorial ligand–receptor interactions. *FEBS Lett.* **410**, 83–86 (1997).
30. Normanno, N. *et al.* Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene* **366**, 2–16 (2006).
31. von Kriegsheim, A. *et al.* Cell fate decisions are specified by the dynamic ERK interactome. *Nature Cell Biol.* **11**, 1458-U172 (2009).
This paper shows how dynamic changes in ERK interactions with other proteins can specify cell fate decisions in response to different growth factors. An important conclusion is that differentiation is not controlled by a master regulator but that control is functionally distributed over an entire protein network.
32. Yoon, S. & Seger, R. The extracellular signal-regulated kinase: multiple substrates regulate diverse cellular functions. *Growth Factors* **24**, 21–44 (2006).
33. Irvine, D. A., Heaney, N. B. & Holyoake, T. L. Optimising chronic myeloid leukaemia therapy in the face of resistance to tyrosine kinase inhibitors — a synthesis of clinical and laboratory data. *Blood Rev.* **24**, 1–9 (2010).
34. Nicholson, E. & Holyoake, T. The chronic myeloid leukemia stem cell. *Clin. Lymphoma Myeloma* **9** (Suppl. 4), 376–381 (2009).
35. Sawyers, C. L. Signal transduction pathways involved in BCR–ABL transformation. *Baillieres Clin. Haematol.* **10**, 223–231 (1997).
36. Brehme, M. *et al.* Charting the molecular network of the drug target Bcr—Abl. *Proc. Natl Acad. Sci. USA* **106**, 7414–7419 (2009).
This paper uses interaction proteomics to map the core protein network involved in signalling by the *BCR–ABL1* oncogene and also examines the effects of the BCR–ABL1 kinase inhibitor imatinib on the network.
37. Brennan, D. J., O'Connor, D. P., Rexhepaj, E., Ponten, F. & Gallagher, W. M. Antibody-based proteomics: towards the implementation of personalised diagnostic and predictive protocols for cancer patients. *Nature Rev. Cancer* **10**, 605–617 (2010).
38. Dhillon, A. S., von Kriegsheim, A., Grindlay, J. & Kolch, W. Phosphatase and feedback regulation of Raf-1 signaling. *Cell Cycle* **6**, 3–7 (2007).
39. Nguyen, A. *et al.* Kinase suppressor of Ras (KSR) is a scaffold which facilitates mitogen-activated protein kinase activation *in vivo*. *Mol. Cell Biol.* **22**, 3035–3045 (2002).
40. Kortum, R. L. & Lewis, R. E. The molecular scaffold KSR1 regulates the proliferative and oncogenic potential of cells. *Mol. Cell Biol.* **24**, 4407–4416 (2004).
41. Lozano, J. *et al.* Deficiency of kinase suppressor of Ras1 prevents oncogenic Ras signaling in mice. *Cancer Res.* **63**, 4232–4238 (2003).
42. McKay, M. M., Ritt, D. A. & Morrison, D. K. Signaling dynamics of the KSR1 scaffold complex. *Proc. Natl Acad. Sci. USA* **106**, 11022–11027 (2009).
43. Rajakulendran, T., Sahmi, M., Lefrancois, M., Sicheri, F. & Therrien, M. A dimerization-dependent mechanism drives RAF catalytic activation. *Nature* **461**, 542–545 (2009).

© 2010 Macmillan Publishers Limited. All rights reserved

44. Casar, B., Pinto, A. & Crespo, P. Essential role of ERK dimers in the activation of cytoplasmic but not nuclear substrates by ERK–scaffold complexes. *Mol. Cell* **31**, 708–721 (2008).
45. Liu, L. *et al.* Proteomic characterization of the dynamic KSR-2 interactome, a signaling scaffold complex in MAPK pathway. *Biochim.Biophys. Acta* **1794**, 1485–1495 (2009).
46. Rauch, J. *et al.* Heterogeneous nuclear ribonucleoprotein H blocks MST2-mediated apoptosis in cancer cells by regulating A-Raf transcription. *Cancer Res.* **70**, 1679–1688 (2010).
47. Levchenko, A., Bruck, J. & Sternberg, P. W. Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc. Natl Acad. Sci. USA* **97**, 5818–5823 (2000).
48. Dougherty, M. K. *et al.* KSR2 is a calcineurin substrate that promotes ERK cascade activation in response to calcium signals. *Mol. Cell* **34**, 652–662 (2009).
A proteomic comparison of binding partners of KSR1 and KSR2 revealed the selective regulation of KSR2 by the associated calcium-regulated phosphatase calcineurin, thus demonstrating the regulation of the function of a scaffold protein by a second messenger pathway.
49. Costanzo-Garvey, D. L. *et al.* KSR2 is an essential regulator of AMP kinase, energy expenditure, and insulin sensitivity. *Cell Metab.* **10**, 366–378 (2009).
50. Mazurek, S., Drexler, H. C., Troppmair, J., Eigenbrodt, E. & Rapp, U. R. Regulation of pyruvate kinase type M2 by A-Raf: a possible glycolytic stop or go mechanism. *Anticancer Res.* **27**, 3963–3971 (2007).
51. Christofk, H. R. *et al.* The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **452**, 230–233 (2008).
52. Macek, B., Mann, M. & Olsen, J. V. Global and site-specific quantitative phosphoproteomics: principles and applications. *Annu. Rev. Pharmacol. Toxicol.* **49**, 199–221 (2009).
53. Oppermann, F. S. *et al.* Large-scale proteomics analysis of the human kinome. *Mol. Cell Proteomics* **8**, 1751–1764 (2009).
54. Mayya, V. & Han, D. K. Phosphoproteomics by mass spectrometry: insights, implications, applications and limitations. *Expert Rev. Proteomics* **6**, 605–618 (2009).
55. Olsen, J. V. *et al.* Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648 (2006).
This paper reports the large-scale quantitative mapping of phosphorylation events triggered by EGF, revealing new insights into the role of phosphorylation in signalling.
56. Schmezle, K. & White, F. M. Phosphoproteomic approaches to elucidate cellular signaling networks. *Curr. Opin. Biotechnol.* **17**, 406–414 (2006).
57. Blagoev, B., Ong, S. E., Kratchmarova, I. & Mann, M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nature Biotechnol.* **22**, 1139–1145 (2004).
58. Kratchmarova, I., Blagoev, B., Haack-Sorensen, M., Kassem, M. & Mann, M. Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. *Science* **308**, 1472–1477 (2005).
59. Ball, S. G., Shuttleworth, C. A. & Kielty, C. M. Platelet-derived growth factor receptors regulate mesenchymal stem cell fate: implications for neovascularization. *Expert Opin. Biol. Ther.* **10**, 57–71 (2010).
60. Heibeck, T. H. *et al.* An extensive survey of tyrosine phosphorylation revealing new sites in human mammary epithelial cells. *J. Proteome Res.* **8**, 3852–3861 (2009).
61. Gan, H. K., Kaye, A. H. & Luwor, R. B. The EGFRvIII variant in glioblastoma multiforme. *J. Clin. Neurosci.* **16**, 748–754 (2009).
62. Huang, P. H. *et al.* Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc. Natl Acad. Sci. USA* **104**, 12867–12872 (2007).
63. Guo, A. *et al.* Signaling networks assembled by oncogenic EGFR and c-Met. *Proc. Natl Acad. Sci. USA* **105**, 692–697 (2008).
64. Amit, I. *et al.* A module of negative feedback regulators defines growth factor signaling. *Nature Genet.* **39**, 503–512 (2007).
65. Govindan, R. A review of epidermal growth factor receptor/HER2 inhibitors in the treatment of patients with non-small-cell lung cancer. *Clin. Lung Cancer* **11**, 8–12 (2010).
66. Rikova, K. *et al.* Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131**, 1190–203 (2007).
Proteomic profiling of tyrosine phosphorylation in a set of approximately 200 lung cancers identified characteristic phosphorylation signatures, as well as known and new tyrosine kinases that are involved in lung cancer.
67. Wolf-Yadlin, A. *et al.* Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. *Mol. Syst. Biol.* **2**, 54 (2006).
68. Kumar, N., Wolf-Yadlin, A., White, F. M. & Lauffenburger, D. A. Modeling HER2 effects on cell behavior from mass spectrometry phosphotyrosine data. *PLoS Comput. Biol.* **3**, e4 (2007).
69. Villen, J. & Gygi, S. P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nature Protoc.* **3**, 1630–1638 (2008).
70. Bodenmiller, B., Mueller, L. N., Mueller, M., Domon, B. & Aebersold, R. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nature Methods* **4**, 231–237 (2007).
71. Choudhary, C. *et al.* Mislocalized activation of oncogenic RTKs switches downstream signaling outcomes. *Mol. Cell* **36**, 326–339 (2009).
72. Chen, Y. *et al.* Combined integrin phosphoproteomic analyses and small interfering RNA-based functional screening identify key regulators for cancer cell adhesion and migration. *Cancer Res.* **69**, 3713–3720 (2009).
This study uses a combination of quantitative proteomics and functional siRNA experiments to identify proteins that mediate integrin-regulated cell adhesion and migration.
73. Siu, M. K. *et al.* Differential expression and phosphorylation of Pak1 and Pak2 in ovarian cancer: effects on prognosis and cell invasion. *Int. J. Cancer* **127**, 21–31 (2010).
74. Tiedemann, R. E. *et al.* Kinome-wide RNAi studies in human multiple myeloma identify vulnerable kinase targets, including a lymphoid-restricted kinase, GRK6. *Blood* **115**, 1594–1604 (2010).
75. Bonte, D. *et al.* Cdc7–Dbf4 kinase overexpression in multiple cancers and tumor cell lines is correlated with p53 inactivation. *Neoplasia* **10**, 920–931 (2008).
76. Hall, M. C. Proteomics modifies our understanding of cell cycle complexity. *Sci. Signal.* **3**, pe4 (2010).
77. Shimoji, S., Park, K. & Hart, G. W. Dynamic crosstalk between GlcNacylation and phosphorylation: roles in signaling, transcription and human disease. *Curr. Signal Transduction Therapy* **5** (Supp.), 25–40 (2010).
78. Hart, G. W., Housley, M. P. & Slawson, C. Cycling of O-linked β-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**, 1017–1022 (2007).
79. Wang, Z. *et al.* Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci. Signal.* **3**, ra2 (2010).
This paper shows that O-GlcNAcylation competes with phosphorylation and plays a major part in the regulation of mitotic spindle assembly and cytokinesis.
80. Chalkley, R. J., Thalhammer, A., Schoepfer, R. & Burlingame, A. L. Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proc. Natl Acad. Sci. USA* **106**, 8894–8899 (2009).
81. Wu, S. L. *et al.* Dynamic profiling of the post-translational modifications and interaction partners of epidermal growth factor receptor signaling after stimulation by epidermal growth factor using Extended Range Proteomic Analysis (ERPA). *Mol. Cell Proteomics* **5**, 1610–1627 (2006).
82. Yang, W. L., Wu, C. Y., Wu, J. & Lin, H. K. Regulation of Akt signaling activation by ubiquitination. *Cell Cycle* **9**, 486–497 (2010).
83. Kitagawa, K., Kotake, Y. & Kitagawa, M. Ubiquitin-mediated control of oncogene and tumor suppressor gene products. *Cancer Sci.* **100**, 1374–1381 (2009).
84. Ardley, H. C. Ring finger ubiquitin protein ligases and their implication to the pathogenesis of human diseases. *Curr. Pharm. Des.* **15**, 3697–3715 (2009).
85. Vlachostergios, P. J., Patrikidou, A., Daliani, D. D. & Papandreou, C. N. The ubiquitin-proteasome system in cancer, a major player in DNA repair. Part 1: post-translational regulation. *J. Cell. Mol. Med.* **13**, 3006–3018 (2009).
86. Vlachostergios, P. J., Patrikidou, A., Daliani, D. D. & Papandreou, C. N. The ubiquitin-proteasome system in cancer, a major player in DNA repair. Part 2: transcriptional regulation. *J. Cell. Mol. Med.* **13**, 3019–3031 (2009).
87. Nakayama, K. I. & Nakayama, K. Ubiquitin ligases: cell-cycle control and cancer. *Nature Rev. Cancer* **6**, 369–381 (2006).
88. Mani, A. & Gelmann, E. P. The ubiquitin-proteasome pathway and its role in cancer. *J. Clin. Oncol.* **23**, 4776–4789 (2005).
89. Coutts, A. S., Adams, C. J. & La Thangue, N. B. p53 ubiquitination by Mdm2: a never ending tail? *DNA Repair* **8**, 483–490 (2009).
90. Marine, J. C. & Lozano, G. Mdm2-mediated ubiquitylation: p53 and beyond. *Cell Death Diff.* **17**, 93–102 (2010).
91. Yang, Y. L., Kitagaki, J., Wang, H., Hou, D. X. & Perantoni, A. O. Targeting the ubiquitin-proteasome system for cancer therapy. *Cancer Sci.* **100**, 24–28 (2009).
92. Hoeller, D. & Dikic, I. Targeting the ubiquitin system in cancer therapy. *Nature* **458**, 438–444 (2009).
93. Ande, S. R., Chen, J. J. & Maddika, S. The ubiquitin pathway: an emerging drug target in cancer therapy. *Eur. J. Pharm.* **625**, 199–205 (2009).
94. Nalepa, G., Rolfe, M. & Harper, J. W. Drug discovery in the ubiquitin-proteasome system. *Nature Rev. Drug Discov.* **5**, 596–613 (2006).
95. Tang, X. *et al.* Genome-wide surveys for phosphorylation-dependent substrates of SCF ubiquitin ligases. *Methods Enzymol.* **399**, 433–458 (2005).
96. Bai, C. *et al.* SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-Box. *Cell* **86**, 263–274 (1996).
97. Radivojac, P. *et al.* Identification, analysis, and prediction of protein ubiquitination sites. *Proteins* **78**, 365–380 (2010).
98. Irish, J. M. *et al.* Single cell profiling of potentiated phospho-protein networks in cancer cells. *Cell* **118**, 217–228 (2004).
99. Angers, S. *et al.* Molecular architecture and assembly of the DDB1–CUL4A ubiquitin ligase machinery. *Nature* **443**, 590–593 (2006).
100. Jeram, S. M. *et al.* An improved SUMmOn-based methodology for the identification of ubiquitin and ubiquitin-like protein conjugation sites identifies novel ubiquitin-like protein chain linkages. *Proteomics* **10**, 254–265 (2010).
101. Burande, C. F. *et al.* A label-free quantitative proteomics strategy to identify E3 ubiquitin ligase substrates targeted to proteasome degradation. *Mol. Cell. Proteomics* **8**, 1719–1727 (2009).
102. Merbl, Y. & Kirschner, M. W. Large-scale detection of ubiquitination substrates using cell extracts and protein microarrays. *Proc. Natl Acad. Sci. USA* **106**, 2543–2548 (2009).
103. Persaud, A. *et al.* Comparison of substrate specificity of the ubiquitin ligases Nedd4 and Nedd4-2 using proteome arrays. *Mol. Syst. Biol.* **5**, 333 (2009).
References 102 and 103 demonstrate the use of protein microarrays to systematically identify substrates for ubiquitin ligases on a large scale.
104. Chen, C. & Matesic, L. E. The Nedd4-like family of E3 ubiquitin ligases and cancer. *Cancer Metastasis Rev.* **26**, 587–604 (2007).
105. Cohen, A. A. *et al.* Dynamic proteomics of individual cancer cells in response to a drug. *Science* **322**, 1511–1516 (2008).
106. Irish, J. M., Kotecha, N. & Nolan, G. P. Mapping normal and cancer cell signalling networks: towards single-cell proteomics. *Nature Rev. Cancer* **6**, 146–155 (2006).
107. Kreeger, P. K. & Lauffenburger, D. A. Cancer systems biology: a network modeling perspective. *Carcinogenesis* **31**, 2–8 (2010).
108. Taylor, I. W. *et al.* Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature Biotechnol.* **27**, 199–204 (2009).
This paper shows that dynamic changes in the organization of cellular protein–protein interactions differ between patients with breast cancer and can be used to predict prognosis.
109. Tan, C. S. *et al.* Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. *Sci. Signal.* **2**, ra39 (2009).
Here, the reconstruction of kinase–substrate networks based on phosphoproteomics data and evolutionary conservation of the phoshorylation sites revealed that several diseases, including cancer, affect conserved parts of the phosphorylation networks.
110. Fenselau, C. A review of quantitative methods for proteomic studies. *J. Chromatogr. B* **855**, 14–20 (2007).

© 2010 Macmillan Publishers Limited. All rights reserved

111. Van den Bergh, G. & Arckens, L. Recent advances in 2D electrophoresis: an array of possibilities. *Expert Rev. Proteomics* **2**, 243–252 (2005).
112. Pan, S. & Aebersold, R. Quantitative proteomics by stable isotope labeling and mass spectrometry. *Methods Mol. Biol.* **367**, 209–218 (2007).
113. Wepf, A., Glatter, T., Schmidt, A., Aebersold, R. & Gstaiger, M. Quantitative interaction proteomics using mass spectrometry. *Nature Methods* **6**, 203–205 (2009).
114. Linscheid, M. W., Ahrends, R., Pieper, S. & Kuhn, A. Liquid chromatography-mass spectrometry-based quantitative proteomics. *Methods Mol. Biol.* **564**, 189–205 (2009).
115. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **389**, 1017–1031 (2007).
116. Everley, P. A., Krijgsveld, J., Zetter, B. R. & Gygi, S. P. Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research. *Mol. Cell. Proteomics* **3**, 729–735 (2004).
117. Mann, M. Functional and quantitative proteomics using SILAC. *Nature Rev. Mol. Cell Biol.* **7**, 952–958 (2006).
118. Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
119. Beynon, R. J. & Pratt, J. M. Metabolic labeling of proteins for proteomics. *Mol. Cell. Proteomics* **4**, 857–872 (2005).
120. Kruger, M. *et al.* SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* **134**, 353–364 (2008).
121. Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotech.* **17**, 994–999 (1999).
122. Ross, P. L. *et al.* Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154–1169 (2004).
123. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl Acad. Sci. USA* **100**, 6940–6945 (2003).
124. Haqqani, A. S., Kelly, J. F. & Stanimirovic, D. B. Quantitative protein profiling by mass spectrometry using label-free proteomics. *Methods Mol. Biol.* **439**, 241–256 (2008).
125. Negishi, A. *et al.* Large-scale quantitative clinical proteomics by label-free liquid chromatography and mass spectrometry. *Cancer Sci.* **100**, 514–519 (2009).
126. Kopf, E. & Zharhary, D. Antibody arrays — an emerging tool in cancer proteomics. *Int. J. Biochem. Cell Biol.* **39**, 1305–1317 (2010).
127. Haab, B., Dunham, M. & Brown, P. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol.* **2**, research0004 (2001).
128. Rino, J., Braga, J., Henriques, R. & Carmo-Fonseca, M. Frontiers in fluorescence microscopy. *Int. J. Dev. Biol.* **53**, 1569–1579 (2009).
129. Goldstein, D. M., Gray, N. S. & Zarrinkar, P. P. High-throughput kinase profiling as a platform for drug discovery. *Nature Rev. Drug Discov.* **7**, 391–397 (2008).
130. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
131. Fedorov, O., Muller, S. & Knapp, S. The (un)targeted cancer kinome. *Nature Chem. Biol.* **6**, 166–169 (2010).
132. Kubota, K. *et al.* Sensitive multiplexed analysis of kinase activities and activity-based kinase identification. *Nature Biotechnol.* **27**, 933–940 (2009).
133. Khan, I. H. *et al.* Microbead arrays for the analysis of ErbB receptor tyrosine kinase activation and dimerization in breast cancer cells. *Assay Drug Dev. Technol.* **8**, 27–36 (2010).
134. Wu, D., Sylvester, J. E., Parker, L. L., Zhou, G. & Kron, S. J. Peptide reporters of kinase activity in whole cell lysates. *Biopolymers* **94**, 475–486 (2010).
135. Parikh, K., Peppelenbosch, M. P. & Ritsema, T. Kinome profiling using peptide arrays in eukaryotic cells. *Methods Mol. Biol.* **527**, 269–280 (2009).
136. Rix, U. & Superti-Furga, G. Target profiling of small molecules by chemical proteomics. *Nature Chem. Biol.* **5**, 616–624 (2009).
137. Chi, P., Allis, C. D. & Wang, G. G. Covalent histone modifications — miswritten, misinterpreted and mis-erased in human cancers. *Nature Rev. Cancer* **10**, 457–469 (2010).
138. Young, N. L. *et al.* High throughput characterization of combinatorial histone codes. *Mol. Cell. Proteomics* **8**, 2266–2284 (2009).
139. Zhou, Q. *et al.* Screening for therapeutic targets of vorinostat by SILAC-based proteomic analysis in human breast cancer cells. *Proteomics* **10**, 1029–1039 (2010).
140. Deribe, Y. L. *et al.* Regulation of epidermal growth factor receptor trafficking by lysine deacetylase HDAC6. *Sci. Signal.* **2**, ra84 (2009).
141. Rocks, O., Peyker, A. & Bastiaens, P. I. Spatio-temporal segregation of Ras signals: one ship, three anchors, many harbors. *Curr. Opin. Cell Biol.* **18**, 351–357 (2006).
142. Hao, D. & Rowinsky, E. K. Inhibiting signal transduction: recent advances in the development of receptor tyrosine kinase and Ras inhibitors. *Cancer Invest.* **20**, 387–404 (2002).
143. Yang, W., Di Vizio, D., Kirchner, M., Steen, H. & Freeman, M. R. Proteome scale characterization of human S-acylated proteins in lipid raft-enriched and non-raft membranes. *Mol. Cell. Proteomics* **9**, 54–70 (2010).
144. Spickett, C. M., Pitt, A. R., Morrice, N. & Kolch, W. Proteomic analysis of phosphorylation, oxidation and nitrosylation in signal transduction. *Biochim. Biophys. Acta* **1764**, 1823–1841 (2006).
145. Wilson, K. J., Gilmore, J. L., Foley, J., Lemmon, M. A. & Riese, D. J. 2nd. Functional selectivity of EGF family peptide growth factors: implications for cancer. *Pharmacol. Ther.* **122**, 1–8 (2009).
146. Citri, A. & Yarden, Y. EGF–ERBB signalling: towards the systems level. *Nature Rev. Mol. Cell Biol.* **7**, 505–516 (2006).
147. Yao, Z. & Seger, R. The ERK signaling cascade — views from different subcellular compartments. *Biofactors* **35**, 407–416 (2009).
148. McKay, M. M. & Morrison, D. K. Integrating signals from RTKs to ERK/MAPK. *Oncogene* **26**, 3113–3121 (2007).
149. Dhillon, A. S., Hagan, S., Rath, O. & Kolch, W. MAP kinase signalling pathways in cancer. *Oncogene* **26**, 3279–3290 (2007).
150. Preisinger, C. & Kolch, W. The Bcr–Abl kinase regulates the actin cytoskeleton via a GADS/Slp-76/Nck1 adaptor protein pathway. *Cell Signal.* **22**, 848–856 (2010).
151. Dyson, J. M. *et al.* The SH2-containing inositol polyphosphate 5-phosphatase, SHIP-2, binds filamin and regulates submembraneous actin. *J. Cell Biol.* **155**, 1065–1079 (2001).
152. Ikeda, F. & Dikic, I. Atypical ubiquitin chains: new molecular signals. *EMBO Rep.* **9**, 536–542 (2008).
153. Schwartz, A. L. & Ciechanover, A. Targeting proteins for destruction by the ubiquitin system: implications for human pathobiology. *Annu. Rev. Pharmacol. Toxicol.* **49**, 73–96 (2009).

Acknowledgements

We apologise for omitting many important contributions due to space constraints. We are grateful for funding from Science Foundation Ireland grant 06/CE/B1129 (W.K.) and the Biotechnology and Biological Sciences Research Council and the Engineering and Physical Sciences Research Council through the Radical Solutions for Researching the Proteome (RASOR) project BB/C511572/1 (A.P.).

Competing interests statement

The authors declare no competing financial interests.

DATABASES

National Cancer Institute Drug Dictionary: http://www.cancer.gov/drugdictionary
dasatinib | imatinib
Pathway Interaction Database: http://pid.nci.nih.gov
AKT | EGFR | EGFR network | ERK | MET | mTOR | p53 | PDGF

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2010 Macmillan Publishers Limited. All rights reserved

Activity-based protein profiling for biochemical pathway discovery in cancer

Daniel K. Nomura, Melissa M. Dix and Benjamin F. Cravatt

Abstract | Large-scale profiling methods have uncovered numerous gene and protein expression changes that correlate with tumorigenesis. However, determining the relevance of these expression changes and which biochemical pathways they affect has been hindered by our incomplete understanding of the proteome and its myriad functions and modes of regulation. Activity-based profiling platforms enable both the discovery of cancer-relevant enzymes and selective pharmacological probes to perturb and characterize these proteins in tumour cells. When integrated with other large-scale profiling methods, activity-based proteomics can provide insight into the metabolic and signalling pathways that support cancer pathogenesis and illuminate new strategies for disease diagnosis and treatment.

Bio-orthogonal chemical handle
A chemical handle that can be specifically modified with reporter tags within the confines of a biological environment.

Click chemistry
Chemistry that allows for quick and reliable joining of small units; the most commonly used click-chemistry reaction is the Huisgen azide-alkyne cycloaddition using a copper catalyst.

The Skaggs Institute for Chemical Biology and Department of Chemical Physiology, The Scripps Research Institute, La Jolla, California 92037, USA.
Correspondence to B.F.C.
e-mail: cravatt@scripps.edu
doi:10.1038/nrc2901
Published online
12 August 2010

Cells with fundamental metabolic alterations commonly arise during tumorigenesis, and it is these types of changes that help to establish a biochemical foundation for disease progression and malignancy[1,2]. A seminal example of this was discovered in the 1920s when Otto Warburg found that cancer cells consume higher levels of glucose and secrete most of the glucose carbon as lactate rather than oxidizing it completely[3,4]. Since then, studies by multiple groups have uncovered a diverse array of metabolic changes in cancer, including alterations in glycolytic pathways[3–6], the citric acid cycle[7], glutaminolysis[8,9], lipogenesis[10], lipolysis[11] and proteolysis[12]. These in turn modulate the levels of cellular building blocks (lipids, nucleic acids and amino acids), cellular energetics, oncogenic signalling molecules and the extracellular environment to confer pro-tumorigenic and malignant properties.

Despite these advances, our current understanding of cancer metabolism is far from complete and would probably benefit from experimental strategies that are capable of profiling enzymatic pathways on a global scale. To this end, conventional genomic[13,14] and proteomic[15–18] methods, which comparatively quantify the expression levels of transcripts and proteins, respectively, have yielded many useful insights. These platforms are, however, limited in their capacity to identify changes in protein activity that are caused by post-translational mechanisms[19]. Annotating biochemical pathways in cancer is further complicated by the potential for enzymes to carry out distinct metabolic activities in tumour cells that might not be mirrored in normal physiology. In addition, a substantial proportion of the human proteome remains functionally uncharacterized, and it is likely that at least some of these poorly understood proteins also have roles in tumorigenesis. These challenges require new proteomic technologies that can accelerate the assignment of protein function in complex biological systems, such as cancer cells and tumours. In this Review, we discuss one such proteomic platform, termed activity-based protein profiling (ABPP)[20–22], and its implementation in the discovery and functional characterization of deregulated enzymatic pathways in cancer. We discuss the evidence that, when coupled with other large-scale profiling methods, such as metabolomics[23,24] and proteomics[15–18], ABPP can provide a compelling, systems-level understanding of biochemical networks that are important for the development and progression of cancer.

ABPP for enzyme discovery in cancer

ABPP uses active site-directed chemical probes to directly assess the functional state of large numbers of enzymes in native biological samples (FIG. 1). Activity-based probes consist of at least two key elements: a reactive group for binding and covalently labelling the active sites of many members of a given enzyme class (or classes), and a reporter tag for the detection, enrichment and identification of probe-labelled enzymes in proteomes. Activity-based probes can be adapted for *in situ* or *in vivo* labelling by substituting the reporter tag with a bio-orthogonal chemical handle, such as an alkyne. Probe-labelled enzymes are then detected by subsequent click chemistry conjugation

© 2010 Macmillan Publishers Limited. All rights reserved

At a glance

- Activity-based protein profiling (ABPP) facilitates the discovery of deregulated enzymes in cancer.
- Competitive ABPP yields selective inhibitors for functional characterization of cancer enzymes.
- ABPP can be integrated with metabolomics to map deregulated enzymatic pathways in cancer.
- ABPP can be integrated with other proteomic methods to map proteolytic pathways in cancer.
- ABPP probes can be used to image tumour development in living animals.

to various azide-modified reporter tags[25,26]. There are currently activity-based probes for a multitude of enzyme classes, including many that have central roles in cancer, such as hydrolases and proteases[20,27–34], kinases[35–38], phosphatases[39], histone deacetylases[40,41], glycosidases[42,43] and various oxidoreductases[44,45]. ABPP can be applied to virtually any cell or tissue (assuming that the genome of the parental organism has been sequenced) and can be combined with a range of analytical methods for data acquisition, including gel- and mass spectrometry (MS)-based methods[21]. Although the specificity of ABPP probes is not absolute, and these probes can be toxic and disrupt biochemical pathways when applied to living systems, they are of great value for characterizing deregulated enzymatic activities in various cancer models and specimens, as discussed below. Examples of activity-based probes that have been used in cancer studies are provided in TABLE 1.

Serine hydrolases are one of the largest and most diverse enzyme classes in mammalian proteomes and include esterases, thioesterases, lipases, amidases and proteases[46]. Several serine hydrolases have been implicated in tumorigenesis, including fatty acid synthase[10], protein methyl esterase 1 (REF. 47), and urokinase-type (uPA) and tissue-type (tPA) plasminogen activators[48]. Fluorophosphonate probes that target the serine hydrolase superfamily[20,27,28] (TABLE 1) have been used to discover several deregulated enzymes in cancer[49–52]. Using ABPP, we discovered that two serine hydrolases — the uncharacterized enzyme KIAA1363 (REFS 49,52,53) and monoacylglycerol (MAG) lipase (MAGL)[11] — are highly expressed in aggressive human cancer cells and primary tumours. We also used ABPP to develop selective inhibitors of KIAA1363 and MAGL for the functional characterization of these enzymes in cancer (discussed below; FIG. 2). Using ABPP, Shields and colleagues[54] recently determined that the activity of the serine hydrolase retinoblastoma-binding protein 9 (RBBP9) is increased in pancreatic carcinomas, in which it promotes anchorage-independent growth and pancreatic carcinogenesis through overcoming transforming growth factor-β (TGFβ)-mediated anti-proliferative signalling by reducing the phosphorylation levels of SMAD2 and SMAD3.

ABPP has also contributed to our knowledge of serine protease activities in cancer. In the course of characterizing an *in vivo*-derived variant of the human breast cancer line MDA-MB-231, termed 231MFP cells, we determined that these cells possess increased uPA and tPA activity in their secreted proteome[51]. These cells also show increased tumour growth *in vivo*, suggesting that deregulated proteolytic activity could contribute to their increased pathogenicity. Madsen and colleagues[55] compared serine hydrolase activities in high and low intravasating variants of the human fibrosarcoma HT-1080 cell line and found increased uPA activity in the high-intravasating variant, in which the protease controlled tumour cell intravasation. Interestingly, in these examples, alterations in protease activity occurred without significant changes in mRNA[51] or protein[55] expression.

Intravasating
A process in cancer metastasis in which the cancer cells invade through the basement membrane into blood vessels.

Interrogating the activities of other protease families has also provided new insights into deregulated proteolytic processes in cancer. Using epoxide-electrophile probes for cysteine proteases[56] (TABLE 1), Joyce and colleagues[57] found that cathepsin activity is higher in angiogenic vasculature and at the invasive fronts of carcinomas, and that pharmacological ablation of a wide range of cathepsins impaired the angiogenic switch, tumour growth, vascularity and invasiveness. They also found that cysteine cathepsins are increased in human papilloma virus (HPV)-induced cervical carcinomas[57]. Profiling metalloproteinase activities with photoreactive, hydroxamate activity-based probes (TABLE 1) has uncovered neprilysin as a membrane-associated glycoprotein that has increased activity in aggressive human melanoma lines compared with less-aggressive counterparts[31]. Comparison of ubiquitin-specific protease activities using the haemagglutinin-tagged ubiquitin-vinyl methyl ester probe (TABLE 1) revealed that the deubiquitylases ubiquitin-carboxyl esterase-L3 (UCHL3) and UCH37 were upregulated in HPV-positive tumours compared with adjacent normal cervical tissue[58]. Comparison of a heptaoma cell line that expressed a stably replicating hepatitis C virus subgenomic replicon RNA with the parental cell line using ABPP probes composed of an *N*-acetylated amino acid that mimicked the P1 position in the peptide substrates of the protease, identified several differentially regulated enzyme activities, some of which were decreased or increased during HCV replication[59].

ABPP has also been used for imaging enzyme activities. Bogyo and colleagues[60] have introduced quenched near-infrared fluorescent activity-based probes (qNIRF-ABPs) to image cysteine protease activities in tumour xenografts *in vivo* and *ex vivo*. These probes emit a fluorescent signal only after covalently modifying a specific protease target, and they can also be used to monitor small-molecule inhibition of protease targets both biochemically and by direct imaging methods. The same researchers have implemented a similar approach using a highly selective aza-peptidyl asparadinyl epoxide qNIRF-ABP[61] to target a specific lysosomal cysteine protease, legumain, the expression of which is increased in many human cancers. In another study, mice were treated with dexamethasone, which induced apoptosis and caspase activation in the thymus, both of which were visualized *in vivo* using a caspase-directed activity-based probe[62] (TABLE 1). This probe could also detect apoptosis that was induced by the monoclonal antibody Apomab in mice bearing xenografted human colorectal tumours[62]. Furthermore, Blum and colleagues[63] demonstrated that among contrast agents for protease activities (small peptides,

© 2010 Macmillan Publishers Limited. All rights reserved

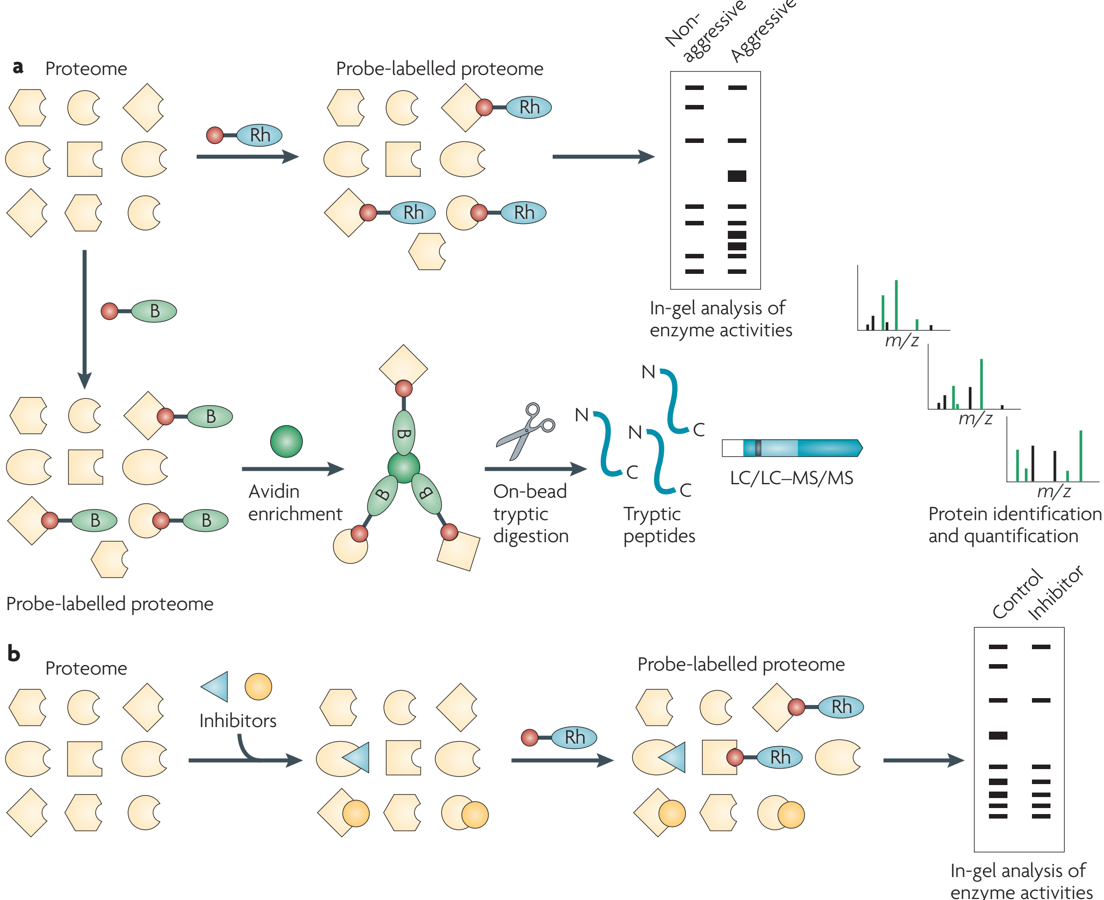


Figure 1 | **Activity-based protein profiling. a** | Activity-based protein profiling (ABPP) uses active site-directed chemical probes to assess the functional state of large numbers of enzymes in native biological systems. Activity-based probes consist of a reactive group (red ball) for targeting a specific set of enzymes and a detection handle (a fluorophore, such as a rhodamine (Rh) or biotin (B)). In a typical ABPP experiment, a proteome is reacted with the activity-based probe and probe-labelled proteins detected by either in-gel fluorescence scanning (for fluorophore-conjugated probes; top) or avidin enrichment, on-bead tryptic digest and liquid chromatograpy–mass spectrometry (LC–MS) analysis (for biotinylated probes; bottom). **b** | ABPP can also be used in a competitive format to evaluate the potency and selectivity of enzyme inhibitors in native biological samples. Inhibitors compete with activity-based probes for enzyme targets, and this competition is read out by loss of fluorescence (for fluorophore-conjugated probes) or MS (for biotinylated probes) signals (not shown). *m/z*, mass to charge ratio.

large polymer-based quenched fluorescent substrates and fluorescently labelled ABPP probes), fluorescent ABPP probes showed more rapid and selective uptake into tumours and overall brighter signal compared with substrate-based probes. These approaches can potentially be used in the clinic to define tumour margins, diagnose tumour grade and assess drug-target occupancy *in vivo*.

ABPP for inhibitor discovery in cancer

Because activity-based probes label the active sites of their enzyme targets, they can form the basis for a competitive screen for enzyme inhibitors[27,64,65]. This competitive ABPP platform has several advantages compared with conventional substrate assays, as inhibitor screens can be conducted directly in complex proteomes and allow concurrent optimization of potency and selectivity against many enzymes in parallel (FIG. 1). Inhibitors can also be developed for uncharacterized enzymes that lack known substrates[53,66]. Beyond its important role in inhibitor discovery, competitive ABPP has been used to identify drug targets and off-targets in preclinical or clinical development to gauge mechanism of action and safety[31,37,67,68].

Competitive ABPP has served as a principal assay for screening directed libraries of inhibitors and for optimizing their selectivity against serine hydrolases that are expressed in cancer cell and tissue proteomes. This effort has led to the identification of two carbamate agents, AS115 (REF. 53) and JZL184 (REF. 69), which are potent and selective inhibitors of KIAA1363 and MAGL, respectively. Competitive ABPP has also been used to explore the full target profile for anti-cancer drugs. Profiling cytochrome P450 enzymes with clickable aryl-alkyne probes (TABLE 1) showed that the aromatase inhibitor anastrazole, which is approved for breast cancer therapy, significantly increases the activity (as determined by probe labelling) of CYP1A2 and decreases the activity of CYP2C19 (REF. 68). These results indicate that anastrazole interacts with multiple P450 enzymes and, in at least one case (CYP1A2), might cooperatively

© 2010 Macmillan Publishers Limited. All rights reserved

Table 1 | **Representative activity-based probes and their application to cancer research**

Structure*	Enzyme class	Applications in cancer
	Serine hydrolases	Identified increased KIAA1363 (REFS 49,52,53) and MAGL[11] activities in aggressive human cancer lines[11,52] and primary tumours[11,49]. Identified increased uPA and tPA serine protease activities in secreted proteomes of aggressive cancer cells[51,52]. Identified increased RBBP9 activity in pancreatic carcinomas[54]
	Metalloproteinases	Identified increased neprilysin activity in aggressive melanoma cell lines[31]
	Cysteine proteases	Identified increased cathepsin cysteine protease activities in pancreatic islet tumours[57]. Used for *in vivo* imaging of tumour cathepsin activity[57,60]
	Kinases	Inhibitor selectivity profiling of kinase inhibitors[35,71]
	Caspases	*In vivo* and *ex vivo* visualization of apoptosis in colon tumour-bearing mice treated with Apomab[62]
	Deubiquitylases	Identified increased carboxy-terminal hydrolase UCHL3 and UCH37 activity in HPV cervical carcinomas[58]
	Cytochrome P450s	Identified the aromatase inhibitor anastrazole as an inducer of CYP1A2 activity[68]

HPV, human papilloma virus; tPA, tissue plasminogen activator; uPA, urinary plasminogen activator. *Blue boxes around structures represent the portion of the activity-based probes that react with the active sites of enzymes.

enhance the binding of other drugs. Characterization of the matrix metalloproteinase (MMP)-directed inhibitor GM6001 (ilomastat) by competitive ABPP revealed that this agent also inhibits several metalloproteinases outside of the MMP family[31], which could explain some of the toxicity issues that broad-spectrum MMP inhibitors have confronted in clinical development[70]. Kinase-directed activity-based probes (TABLE 1) have been used to identify 39 kinase targets of the broad-spectrum inhibitor staurosporine in cancer cell lines[35]. Activity-based probes based on the structure of the PI3K inhibitor wortmannin[37,38], revealed that this natural product also targets members of the polo-like clan of kinases[37,71]. The integration of competitive ABPP platforms into the pre-clinical and clinical development of cancer therapeutics has the potential to clarify the mechanism of action and reduce off-target toxicity for future drug candidates.

Finally, we should note that competitive ABPP experiments have historically been analysed using one-dimensional (1D) SDS–PAGE or MS, which are

© 2010 Macmillan Publishers Limited. All rights reserved

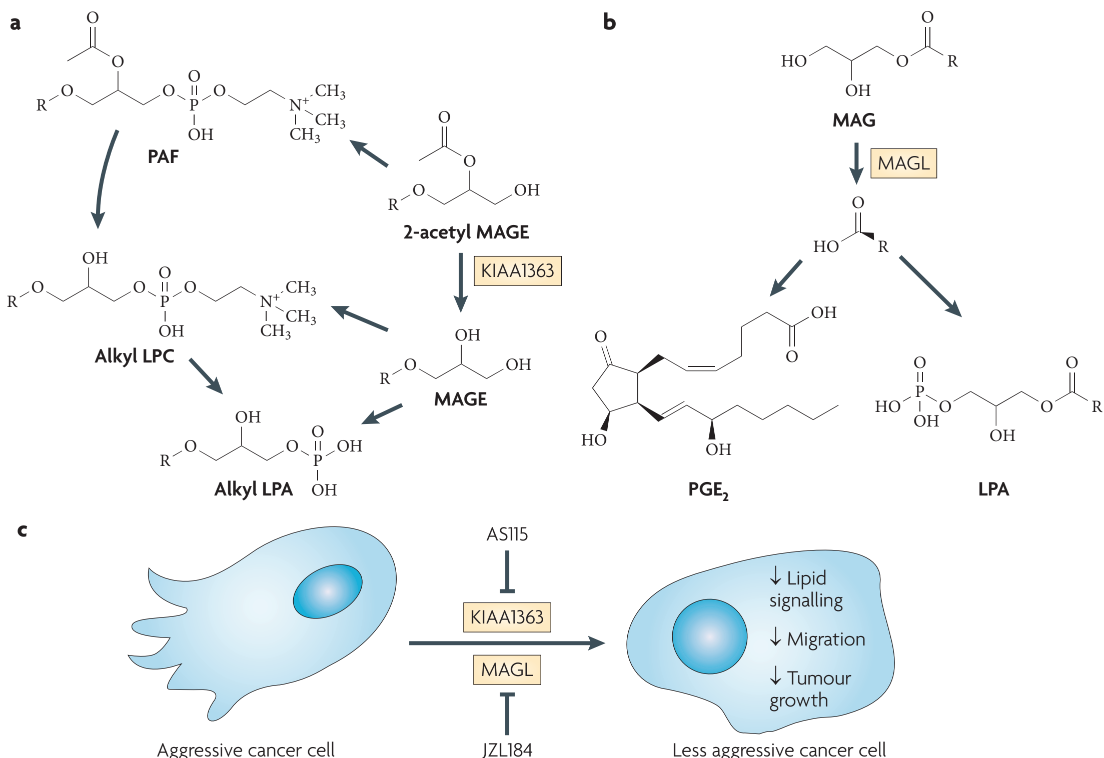


Figure 2 | **Serine hydrolases KIAA1363 and MAGL regulate lipid metabolic pathways that support cancer pathogenesis.** Activity-based protein profiling (ABPP) identified KIAA1363 (part **a**) and monoacylglycerol (MAG) lipase (MAGL) (part **b**) as being increased in aggressive human cancer cells from multiple tumour types. Pharmacological and/or RNA interference ablation of KIAA1363 and MAGL coupled with metabolomic analysis revealed specific roles for KIAA1363 and MAGL in cancer metabolism. Disruption of KIAA1363 by the small-molecule inhibitor AS115 lowered monoalkylglycerol ether (MAGE), alkyl lysophosphatidic acid (alkyl LPA) and alkyl lysophosphatidyl choline (alkyl LPC) levels in cancer cells. Disruption of MAGL by the small-molecule inhibitor JZL184 raised MAG levels and reduced free fatty acid, lysophosphatidic acid (LPA) and prostaglandin E_2 (PGE_2) levels in cancer cells. Disruption of KIAA1363 and MAGL leads to impairments in cancer cell aggressiveness and tumour growth (part **c**). PAF, platelet-activating factor.

limited in throughput to compound libraries of a modest size (200–300 compounds). This limitation has recently been addressed, at least in part, by coupling competitive ABPP with fluorescence polarization (fluopol-ABPP)[72], which provides a homogeneous assay that is compatible with high-throughput screening and which can be adapted to different classes of enzymes and activity-based probes. Fluopol-ABPP has been successfully used to discover selective inhibitors for two cancer-related enzyme targets, the hydrolytic enzyme RBBP9 and the thioltransferase glutathione *S*-transferase omega 1 (GSTO1)[72,73].

Integration of ABPP with other profiling methods

Integration of ABPP with metabolomics. Metabolomics has emerged as a powerful method for broadly assessing the biochemical functions of enzymatic pathways in normal physiology and disease[23,24] (BOX 1). When complemented with selective inhibitors developed through competitive ABPP, metabolomics can be used to not only identify endogenous substrates and products of enzymes, but also metabolites that are upstream or downstream of these immediate substrates and products, allowing the integration of individual enzymatic reactions into the larger metabolic networks of cancer cells. Two examples of how coupling of ABPP with metabolomics has helped in defining contributions made by enzymes in cancer are discussed below.

A role for KIAA1363 in regulating pro-tumorigenic ether lipids. As mentioned above, ABPP studies identified increased KIAA1363 activity in both aggressive human cancer cell lines[53] and primary tumours[49], and have identified AS115 as a potent and selective inhibitor of this enzyme[53]. Untargeted liquid chromatograpy (LC)–MS analysis of lipophilic metabolites from AS115-treated cancer cells revealed that KIA1363 regulates an unusual class of lipids — the monoalkylglycerol ethers (MAGEs)[11]. Previous studies had shown that tumours contained increased levels of MAGE and other ether lipids and identified positive correlations between ether lipid content and tumorigenicity in cancer cells[74,75]. However, the enzymes that regulate ether lipid metabolism in cancer have remained enigmatic. Additional studies showed that KIAA1363 is the principal 2-acetyl MAGE hydrolase in cancer cells, providing one potential pathway by which this enzyme could influence ether lipid content. Importantly, stable knock down of KIAA1363 by RNA interference (RNAi) also led to a reduction in MAGE levels, and this effect was found to further perturb other

Ether lipid
A lipid in which one or more of the oxygens on a glycerol backbone is bonded to an alkyl chain by an ether linkage.

© 2010 Macmillan Publishers Limited. All rights reserved

pro-tumorigenic lipids, including alkyl lysophosphatidic acid (alkyl-LPA)[53]. These metabolic changes in cancer cells correlated with reductions in migration and *in vivo* tumour growth, thus pointing to an important role for the KIAA1363–ether lipid pathway in supporting cancer pathogenesis[53].

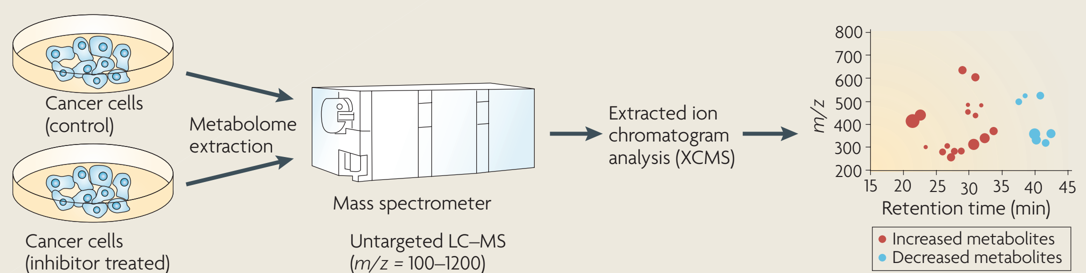
A role for MAGL in regulating fatty acid pathways in cancer cells. Competitive ABPP identified a potent and selective MAGL inhibitor, JZL184 (REF. 69), which increases MAG levels in multiple tissues in mice without concurrent changes in global free fatty acid levels[76]. In contrast to this metabolic profile, MAGL inhibition in aggressive melanoma cells, and ovarian and breast cancer cells, not only raised MAGs but also lowered free fatty acid levels — results that were confirmed by RNAi knock down of MAGL[11]. MAGL reductions also impaired cancer cell migration, invasion and serum-free cell survival *in vitro*, as well as tumour xenograft growth *in vivo*[11]. These effects were rescued by the addition of free fatty acids *in vitro* or treatment with a high-fat diet *in vivo*, which supports a pro-tumorigenic role for MAGL-generated fatty acids. Also consistent with this premise, overexpression of wild-type, but not catalytically dead, MAGL was sufficient to increase free fatty acids and confer malignant properties on non-aggressive cancer cells[11]. Metabolomic profiles revealed that the MAGL–fatty acid pathway feeds into a larger lipid network that includes the pro-tumorigenic signalling molecules LPA and prostaglandin E_2 (PGE_2)[11]. These results suggest that as tumour development progresses, cancer cells with increased MAGL activity produce more cellular fatty acids, which can serve as building blocks for lipid transmitters that further drive cancer malignancy.

Additional metabolomic studies in cancer. Metabolomics has also been a useful technology for characterizing enzymes that are genetically mutated in cancer. An integrated genomic analysis, consisting of sequencing 22,661 protein-coding genes coupled with high-density oligonucleotide array analysis of 22 human glioblastoma (World Health Organization (WHO) grade IV) samples, identified mutations in the enzyme cytosolic isocitrate dehydrogenase 1 (IDH1) as a common feature of a major subset of primary human brain cancers[7]. These mutations produce a single amino acid change in the IDH1 active site, resulting in the loss of the ability of the enzyme to convert isocitrate to α-ketoglutarate. Metabolomics revealed that cancer-associated mutations also result in a new catalytic activity — the NADPH-dependent reduction of α-ketoglutarate to R(2)-2-hydroxyglutarate (2-HG)[7]. Human malignant gliomas with an IDH1 mutation exhibit markedly increased levels of 2HG compared with gliomas without this mutation[7]. These intriguing findings designate 2-HG as a potential onco-metabolite. Consistent with this hypothesis, an excess accumulation of 2-HG has been shown to lead to an increased risk of brain tumour development in patients with inborn errors of 2-HG metabolism[77].

ABPP and proteases. Proteases have long been implicated as drivers of tumorigenesis and are important for early tumour progression, as well as invasion and metastasis[78–81]. Increased protease activity is also useful as a diagnostic marker for many cancers[78–81]. Although proteases catalyse one of the most pervasive post-translational modifications in living systems, most of these enzymes remain incompletely characterized with respect to their endogenous substrates. Whereas technologies such as ABPP are useful for identifying deregulated proteolytic activities, additional methods are required to identify

Box 1 | **Metabolomics**

Metabolomics has emerged as a powerful approach for investigating enzyme function in living systems[23,24]. Metabolomic experiments in the context of enzyme studies typically start with the extraction of metabolites from control and enzyme-disrupted biological systems, followed by metabolite detection and comparative data analysis. For example, lipophilic metabolites can be enriched from cells or tissues by organic extraction. Mass spectrometry (MS) has become a primary analytical method for surveying metabolites in complex biological samples, with upfront separation accomplished by liquid chromatography (LC–MS) or gas chromatography (GC–MS). MS experiments can be carried out using targeted[92] or untargeted[93,94] approaches, depending on whether the objective is to profile and quantitate known metabolites or to broadly scan for metabolites across a large mass range, respectively. As metabolomic experiments generate a large amount of data, powerful software tools are needed for identification and quantitation of ions in LC–MS data sets (see the figure; the mass to charge ratio (*m/z*) is indicated). One such program is XCMS[95], which aligns, quantifies and statistically ranks ions that are altered between two sets of metabolomic data. This program can be used to rapidly identify metabolomic signatures of various disease states or to assess metabolic networks that are regulated by an enzyme using pharmacological or genetic tools that modulate enzyme function. Additional databases assist in metabolite structural characterization, such as HMDB[96,97], METLIN[98,99] and LIPID MAPS[100].



© 2010 Macmillan Publishers Limited. All rights reserved

the substrates of proteases. To address this issue, several proteomic approaches have been developed that globally profile protease–substrate relationships in native biological systems[82–86] (FIG. 3). One group of proteomic methods, N-terminal labelling techniques, relies on chemically tagging neo-N termini that are produced through proteolytic cleavage events. The chemically tagged N-terminal peptides can then be enriched and analysed using MS methods. Many variations of this strategy have been developed and have been instrumental in identifying exact sites of proteolytic cleavage for various proteases, including caspases[84] and MMPs[85]. Wells and colleagues[84] have introduced an engineered enzyme (subtiligase) to tag and enrich nascent N termini in complex protein mixtures. Application of this technique to study apoptosis led to the identification of 292 caspase-cleaved proteins[84]. Overall and colleagues[85] have recently developed the terminal amine isotopic labelling of substrates (TAILS) technique, which uses an amine-reactive polymer to capture internal peptides leaving only labelled neo-N termini and mature blocked N termini for MS analysis. They used this technique to extensively study the substrate profile of MMP2 in mouse fibroblast secretomes and identified 288 potential substrates for this protease[85]. Van Damme *et al.*[82] developed an N-terminal enrichment strategy, which relies on combined fractional diagonal chromatography (COFRADIC), and so identified 93 caspase-mediated cleavage events during CD95-induced apoptosis. Also of note is the N-CLAP method developed by Jaffrey and colleagues[83], which uses a chemical treatment strategy to selectively label N-terminal amines on proteins.

Although N-terminal labelling methods have provided valuable insights into biological systems that are controlled by proteases, they possess some drawbacks. Perhaps most notably, the identification of a cleavage event relies on the detection of a single peptide from the C-terminal portion of a cleaved protein, and therefore does not provide information on the size or stability of the remaining protein fragments. With this consideration in mind, we developed a platform for protease substrate discovery that aims to provide a more complete topological map of proteolytic events occurring in biological systems. This approach, termed the Protein Topography and Migration Analysis Platform (PROTOMAP), consists of the fractionation of active and inactive protease samples by 1D SDS–PAGE, followed by LC–MS analysis of tryptically digested proteins from individual gel bands[86,87] (FIG. 3). Identified proteins are subsequently assembled into two-dimensional peptographs that combine sequence coverage with gel migration information, and when coupled with spectral counting data this yields a semi-quantitative topographical map for all detectable proteins (and their cleavage products) in a sample. PROTOMAP has been used to study the intrinsic apoptotic pathway in Jurkat T cells, resulting in the discovery of more than 250 protein cleavage events in apoptotic cells, most of which had not previously been reported in the literature[86,87]. Interestingly,

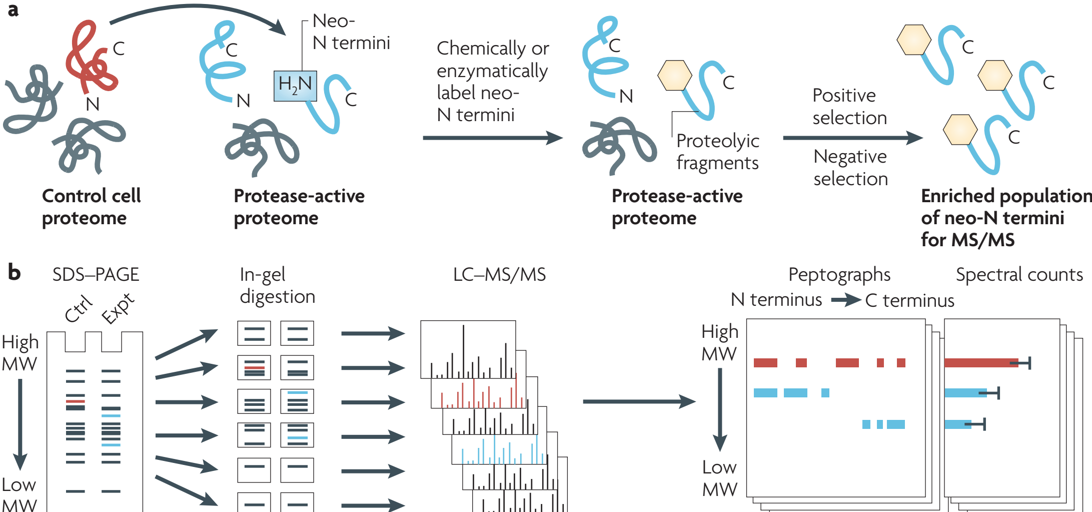


Figure 3 | **Proteomic strategies for mapping protease substrates. a** | Amino terminal labelling techniques use chemical or enzymatic methods to selectively label neo-N termini that are created on protease treatment. The labelled N-terminal peptides can then be enriched through positive selection methods (such as the subtiligase method[84]) or, alternatively, the internal unmodified peptides can be removed through negative selection methods (such as terminal amine isotopic labelling of substrates (TAILS) methods[85]). The remaining pool of enriched labelled neo-N termini can then be analysed using tandem mass spectrometry (MS) and the exact sites of proteolytic cleavage can be assigned. **b** | An alternative proteomic method for protease substrate identification, Protein Topography and Migration Analysis Platform (PROTOMAP [86]) combines one-dimensional (1D) SDS–PAGE fractionation with liquid chromatograpy (LC)–MS analysis. In a typical PROTOMAP experiment, proteomes from control (Ctrl; red) and experimental (protease-active) (Expt; blue) systems are separated by 1D SDS–PAGE. The lanes are cut into bands at fixed intervals, digested with trypsin and analysed by LC–MS/MS to generate data that are integrated into peptographs, which plot sequence coverage for a given protein in the horizontal dimension (N to C terminus; left to right) versus gel migration in the vertical dimension. Spectral count values for each protein in each gel band provide quantitation. Cleaved proteins are identified by shifts in migration from higher to lower molecular weight (MW) in Ctrl versus Expt systems.

© 2010 Macmillan Publishers Limited. All rights reserved

a meta-analysis has revealed substantial congruency in the proteolytic events identified by the subtiligase and PROTOMAP methods in apoptotic cells[73], indicating that these approaches provide complementary ways to characterize proteolytic pathways in biological systems. We envision that coupling protease–substrate discovery platforms with ABPP should offer a versatile and a potentially routine way to map deregulated proteolytic pathways in cancer and other pathophysiological processes. The information acquired by protease–substrate discovery platforms should also guide the development of new substrate-derived ABPP probes for specific proteases.

Conclusions

Chemical labelling methods have become centrepieces for a wide range of proteomic investigations, whether for measuring the expression[88] and the post-translational modification[89–91] of proteins or, as we discuss above, the activity state of proteins. We have attempted to highlight how one such chemoproteomic technology, ABPP, has enabled the discovery of new enzyme activities deregulated in human cancer. Importantly, ABPP also provides a built-in assay that can be used to develop inhibitors to assess the functional role of enzymes in cancer. Activity-based probes have found additional uses as, for example, imaging agents to visualize enzyme activities in tumours. An extension of these studies into a clinical setting could enable rapid imaging, staging and diagnosis of various types of cancers both in biopsy samples and *in vivo*. Looking forwards, we also anticipate that ABPP might have an important role in demystifying the process of target identification for chemical genomic screens in cancer cells. Lead compounds emerging from phenotypic screens could be transformed into ABPP probes by introducing, for example, photoreactive groups and clickable affinity handles to facilitate the identification of protein targets and sites of probe labelling.

With the advent of fluopol-ABPP, inhibitor discovery and optimization can now be carried out in a high-throughput manner, allowing cancer-relevant enzyme targets to be screened against extensive small-molecule libraries, such as those available as part of the US National Institutes of Health-supported Molecular Libraries Screening initiative. By coupling ABPP with other large-scale profiling methods, such as metabolomics and proteomics, important insights can be gained into how certain enzymes are used or hijacked to carry out biochemical tasks that fuel tumorigenesis. Such enzymes could be important targets for the next generation of cancer therapeutics.

1. Deberardinis, R. J., Lum, J. J., Hatzivassiliou, G. & Thompson, C. B. The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell. Metab.* **7**, 11–20 (2008).
2. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033 (2009).
3. Warburg, O. On respiratory impairment in cancer cells. *Science* **124**, 269–270 (1956).
4. Warburg, O. On the origin of cancer cells. *Science* **123**, 309–314 (1956).
A seminal paper describing how cancer cells primarily depend on glycolysis for energy, showing that cancer cells have a fundamentally altered metabolism compared with normal cells.
5. Christofk, H. R. *et al.* The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **452**, 230–233 (2008).
This paper shows that the M2 isoform of pyruvate kinase is responsible for the Warburg effect.
6. Christofk, H. R., Vander Heiden, M. G., Wu, N., Asara, J. M. & Cantley, L. C. Pyruvate kinase M2 is a phosphotyrosine-binding protein. *Nature* **452**, 181–186 (2008).
7. Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744 (2009).
8. Wise, D. R. *et al.* Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. *Proc. Natl Acad. Sci. USA* **105**, 18782–18787 (2008).
9. DeBerardinis, R. J. *et al.* Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc. Natl Acad. Sci. USA* **104**, 19345–19350 (2007).
10. Menendez, J. A. & Lupu, R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nature Rev. Cancer* **7**, 763–777 (2007).
A comprehensive review of the roles of enhanced fatty acid synthase and *de novo* lipogenesis in cancer.
11. Nomura, D. K. *et al.* Monoacylglycerol lipase regulates a fatty acid network that promotes cancer pathogenesis. *Cell* **140**, 49–61 (2010).
This paper used ABPP in combination with metabolomics to show that the lipolytic enzyme MAGL regulates pro-tumorigenic fatty acid products in cancer cells.
12. Mohamed, M. M. & Sloane, B. F. Cysteine cathepsins: multifunctional enzymes in cancer. *Nature Rev. Cancer* **6**, 764–775 (2006).
13. Brown, P. O. & Botstein, D. Exploring the new world of the genome with DNA microarrays. *Nature Genet.* **21**, 33–37 (1999).
14. Golub, T. R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537 (1999).
15. Patterson, S. D. & Aebersold, R. H. Proteomics: the first decade and beyond. *Nature Genet.* **33**, 311–323 (2003).
16. Yates, J. R. Mass spectral analysis in proteomics. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 297–316 (2004).
17. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**, 212–217 (2006).
18. Cravatt, B. F., Simon, G. M. & Yates, J. R., 3rd. The biological impact of mass-spectrometry-based proteomics. *Nature* **450**, 991–1000 (2007).
19. Kobe, B. & Kemp, B. E. Active site-directed protein regulation. *Nature* **402**, 373–376 (1999).
20. Liu, Y., Patricelli, M. P. & Cravatt, B. F. Activity-based protein profiling: the serine hydrolases. *Proc. Natl Acad. Sci. USA* **96**, 14694–14699 (1999).
This paper describes the principles of ABPP and development of probes for the serine hydrolase superfamily.
21. Cravatt, B. F., Wright, A. T. & Kozarich, J. W. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.* **77**, 383–414 (2008).
22. Fonovic, M. & Bogyo, M. Activity based probes for proteases: applications to biomarker discovery, molecular imaging and drug screening. *Curr. Pharm. Des.* **13**, 253–261 (2007).
23. Vinayavekhin, N., Homan, E. A. & Saghatelian, A. Exploring disease through metabolomics. *ACS Chem. Biol.* **5**, 91–103 (2010).
24. Saghatelian, A. *et al.* Assignment of endogenous substrates to enzymes by global metabolite profiling. *Biochemistry* **43**, 14332–14339 (2004).
25. Speers, A. E., Adam, G. C. & Cravatt, B. F. Activity-based protein profiling *in vivo* using a copper(i)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Amer. Chem. Soc.* **125**, 4686–4687 (2003).
26. Speers, A. E. & Cravatt, B. F. Profiling enzyme activities *in vivo* using click chemistry methods. *Chem. Biol.* **11**, 535–546 (2004).
27. Kidd, D., Liu, Y. & Cravatt, B. F. Profiling serine hydrolase activities in complex proteomes. *Biochemistry* **40**, 4005–4015 (2001).
28. Patricelli, M. P., Giang, D. K., Stamp, L. M. & Burbaum, J. J. Direct visualization of serine hydrolase activities in complex proteome using fluorescent active site-directed probes. *Proteomics* **1**, 1067–1071 (2001).
29. Greenbaum, D. *et al.* Chemical approaches for functionally probing the proteome. *Mol. Cell. Proteomics* **1**, 60–68 (2002).
30. Kato, D. *et al.* Activity-based probes that target diverse cysteine protease families. *Nature Chem. Biol.* **1**, 33–38 (2005).
31. Saghatelian, A., Jessani, N., Joseph, A., Humphrey, M. & Cravatt, B. F. Activity-based probes for the proteomic profiling of metalloproteases. *Proc. Natl Acad. Sci. USA* **101**, 10000–10005 (2004).
32. Sieber, S. A., Niessen, S., Hoover, H. S. & Cravatt, B. F. Proteomic profiling of metalloprotease activities with cocktails of active-site probes. *Nature Chem. Biol.* **2**, 274–281 (2006).
33. Chan, E. W., Chattopadhaya, S., Panicker, R. C., Huang, X. & Yao, S. Q. Developing photoactive affinity probes for proteomic profiling: hydroxamate-based probes for metalloproteases. *J. Am. Chem. Soc.* **126**, 14435–14446 (2004).
34. Li, Y. M. *et al.* Photoactivated gamma-secretase inhibitors directed to the active site covalently label presenilin 1. *Nature* **405**, 689–694 (2000).
35. Patricelli, M. P. *et al.* Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry* **46**, 350–358 (2007).
36. Cohen, M. S., Hadjivassiliou, H. & Taunton, J. A clickable inhibitor reveals context-dependent autoactivation of p90 RSK. *Nature Chem. Biol.* **3**, 156–160 (2007).
37. Liu, Y. *et al.* Wortmannin, a widely used phosphoinositide 3-kinase inhibitor, also potently inhibits mammalian polo-like kinase. *Chem. Biol.* **12**, 99–107 (2005).
38. Yee, M. C., Fas, S. C., Stohlmeyer, M. M., Wandless, T. J. & Cimprich, K. A. A cell-permeable activity-based probe for protein and lipid kinases. *J. Biol. Chem.* **280**, 29053–29059 (2005).
39. Kumar, S. *et al.* Activity-based probes for protein tyrosine phosphatases. *Proc. Natl Acad. Sci. USA* **101**, 7943–7948 (2004).
40. Salisbury, C. M. & Cravatt, B. F. Activity-based probes for proteomic profiling of histone deacetylase complexes. *Proc. Natl Acad. Sci. USA* **104**, 1171–1176 (2007).

© 2010 Macmillan Publishers Limited. All rights reserved

41. Salisbury, C. M. & Cravatt, B. F. Optimization of activity-based probes for proteomic profiling of histone deacetylase complexes. *J. Am. Chem. Soc.* **130**, 2184–2194 (2008).
42. Hekmat, O., Kim, Y. W., Williams, S. J., He, S. & Withers, S. G. Active-site peptide "fingerprinting" of glycosidases in complex mixtures by mass spectrometry. Discovery of a novel retaining beta-1,4-glycanase in *Cellulomonas fimi*. *J. Biol. Chem.* **280**, 35126–35135 (2005).
43. Vocadlo, D. J. & Bertozzi, C. R. A strategy for functional proteomic analysis of glycosidase activity from cell lysates. *Angew. Chem. Int. Ed Engl.* **43**, 5338–5342 (2004).
44. Adam, G. C., Cravatt, B. F. & Sorensen, E. J. Profiling the specific reactivity of the proteome with non-directed activity-based probes. *Chem. Biol.* **8**, 81–95 (2001).
45. Weerapana, E., Simon, G. M. & Cravatt, B. F. Disparate proteome reactivity profiles of carbon electrophiles. *Nature Chem. Biol.* **4**, 405–407 (2008).
46. Simon, G. M. & Cravatt, B. F. Activity-based proteomics of enzyme superfamilies: serine hydrolases as a case study. *J. Biol. Chem.* **285**, 11051–11055 (2010).
47. Puustinen, P. *et al.* PME-1 protects extracellular signal-regulated kinase pathway activity from protein phosphatase 2A-mediated inactivation in human malignant glioma. *Cancer Res.* **69**, 2870–7 (2009).
48. Andreasen, P. A., Egelund, R. & Petersen, H. H. The plasminogen activation system in tumor growth, invasion, and metastasis. *Cell. Mol. Life Sci.* **57**, 25–40 (2000).
49. Jessani, N. *et al.* A streamlined platform for high-content functional proteomics of primary human specimens. *Nature Methods* **2**, 691–697 (2005).
50. Jessani, N., Niessen, S., Mueller, B. M. & Cravatt, B. F. Breast cancer cell lines grown *in vivo*: what goes in isn't always the same as what comes out. *Cell Cycle* **4**, 253–255 (2005).
51. Jessani, N. *et al.* Carcinoma and stromal enzyme activity profiles associated with breast tumor growth *in vivo*. *Proc. Natl Acad. Sci. USA* **101**, 13756–13761 (2004).
52. Jessani, N., Liu, Y., Humphrey, M. & Cravatt, B. F. Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness. *Proc. Natl Acad. Sci. USA* **99**, 10335–10340 (2002).
53. Chiang, K. P., Niessen, S., Saghatelian, A. & Cravatt, B. F. An enzyme that regulates ether lipid signaling pathways in cancer annotated by multidimensional profiling. *Chem. Biol.* **13**, 1041–1050 (2006).
This paper uses ABPP coupled with metabolomics to determine that the uncharacterized enzyme KIAA1363 regulates ether lipid metabolism in aggressive cancer cells.
54. Shields, D. J. *et al.* RBBP9: a tumor-associated serine hydrolase activity required for pancreatic neoplasia. *Proc. Natl Acad. Sci. USA* **107**, 2189–2194 (2010).
55. Madsen, M. A., Deryugina, E. I., Niessen, S., Cravatt, B. F. & Quigley, J. P. Activity-based protein profiling implicates urokinase activation as a key step in human fibrosarcoma intravasation. *J. Biol. Chem.* **281**, 15997–16005 (2006).
56. Greenbaum, D., Medzihradszky, K. F., Burlingame, A. & Bogyo, M. Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem. Biol.* **7**, 569–581 (2000).
57. Joyce, J. A. *et al.* Cathepsin cysteine proteases are effectors of invasive growth and angiogenesis during multistage tumorigenesis. *Cancer Cell* **5**, 443–453 (2004).
This paper uses ABPP to identify the cathepsin class of cysteine proteases as important contributors to cancer pathogenesis.
58. Rolen, U. *et al.* Activity profiling of deubiquitinating enzymes in cervical carcinoma biopsies and cell lines. *Mol. Carcinog.* **45**, 260–269 (2006).
59. Blais, D. R. *et al.* Activity-based proteome profiling of hepatoma cells during hepatitis C virus replication using protease substrate probes. *J. Proteome Res.* **9**, 912–923 (2009).
60. Blum, G., von Degenfeld, G., Merchant, M. J., Blau, H. M. & Bogyo, M. Noninvasive optical imaging of cysteine protease activity using fluorescently quenched activity-based probes. *Nature Chem. Biol.* **3**, 668–677 (2007).
This paper describes the use of ABPP probes for near-infrared imaging of protease activities in tumours in living animals.
61. Lee, J. & Bogyo, M. Development of near-infrared fluorophore (NIRF)-labeled activity-based probes for *in vivo* imaging of legumain. *ACS Chem. Biol.* **5**, 233–243 (2010).
62. Edgington, L. E. *et al.* Noninvasive optical imaging of apoptosis by caspase-targeted activity-based probes. *Nature Med.* **15**, 967–973 (2009).
63. Blum, G., Weimer, R. M., Edgington, L. E., Adams, W. & Bogyo, M. Comparative assessment of substrates and activity based probes as tools for non-invasive optical imaging of cysteine protease activity. *PLoS ONE* **4**, e6374 (2009).
64. Greenbaum, D. C. *et al.* Small molecule affinity fingerprinting. A tool for enzyme family subclassification, target identification, and inhibitor design. *Chem. Biol.* **9**, 1085–1094 (2002).
65. Leung, D., Hardouin, C., Boger, D. L. & Cravatt, B. F. Discovering potent and selective reversible inhibitors of enzymes in complex proteomes. *Nature Biotechnol.* **21**, 687–691 (2003).
66. Li, W., Blankman, J. L. & Cravatt, B. F. A functional proteomic strategy to discover inhibitors for uncharacterized hydrolases. *J. Am. Chem. Soc.* **129**, 9594–9595 (2007).
67. Wright, A. T. & Cravatt, B. F. Chemical proteomic probes for profiling cytochrome p450 activities and drug interactions *in vivo*. *Chem. Biol.* **14**, 1043–1051 (2007).
68. Wright, A. T., Song, J. D. & Cravatt, B. F. A suite of activity-based probes for human cytochrome P450 enzymes. *J. Am. Chem. Soc.* **131**, 10692–10700 (2009).
69. Long, J. Z. et *al.* Selective blockade of 2-arachidonoylglycerol hydrolysis produces cannabinoid behavioral effects. *Nature Chem. Biol.* **5**, 37–44 (2009).
70. Coussens, L. M., Fingleton, B. & Matrisian, L. M. Matrix metalloproteinase inhibitors and cancer: trials and tribulations. *Science* **295**, 2387–2392 (2002).
71. Liu, Y., Jiang, N., Wu, J., Dai, W. & Rosenblum, J. S. Polo-like kinases inhibited by wortmannin. Labeling site and downstream effects. *J. Biol. Chem.* **282**, 2505–2511 (2007).
72. Bachovchin, D. A., Brown, S. J., Rosen, H. & Cravatt, B. F. Identification of selective inhibitors of uncharacterized enzymes by high-throughput screening with fluorescent activity-based probes. *Nature Biotechnol.* **27**, 387–394 (2009).
73. Bachovchin, D. A. *et al.* Oxime esters as selective, covalent inhibitors of the serine hydrolase retinoblastoma-binding protein 9 (RBBP9). *Bioorg Med. Chem. Lett.* **20**, 2254–2258 (2010).
74. Roos, D. S. & Choppin, P. W. Tumorigenicity of cell lines with altered lipid composition. *Proc. Natl Acad. Sci. USA* **81**, 7622–7626 (1984).
75. Snyder, F. & Wood, R. Alkyl and alk-1-enyl ethers of glycerol in lipids from normal and neoplastic human tissues. *Cancer Res.* **29**, 251–257 (1969).
76. Long, J. Z., Nomura, D. K. & Cravatt, B. F. Characterization of monoacylglycerol lipase inhibition reveals differences in central and peripheral endocannabinoid metabolism. *Chem. Biol.* **16**, 744–753 (2009).
77. Van Schaftingen, E., Rzem, R. & Veiga-da-Cunha, M. L. -2-hydroxyglutaric aciduria, a disorder of metabolite repair. *J. Inherit Metab. Dis.* **32**, 135–142 (2009).
78. Koblinski, J. E., Ahram, M. & Sloane, B. F. Unraveling the role of proteases in cancer. *Clin. Chim. Acta* **291**, 113–135 (2000).
79. Janicke, F. *et al.* Urokinase (uPA) and its inhibitor PAI-1 are strong and independent prognostic factors in node-negative breast cancer. *Breast Cancer Res. Treat* **24**, 195–208 (1993).
80. Kessenbrock, K., Plaks, V. & Werb, Z. Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* **141**, 52–67 (2010).
81. Tandon, A. K., Clark, G. M., Chamness, G. C., Chirgwin, J. M. & McGuire, W. L. Cathepsin D and prognosis in breast cancer. *N. Engl. J. Med.* **322**, 297–302 (1990).
82. Van Damme, P. *et al.* Caspase-specific and nonspecific *in vivo* protein processing during Fas-induced apoptosis. *Nature Methods* **2**, 771–777 (2005).
83. Xu, G., Shin, S. B. & Jaffrey, S. R. Global profiling of protease cleavage sites by chemoselective labeling of protein N-termini. *Proc. Natl Acad. Sci. USA* **106**, 19310–19315 (2009).
84. Mahrus, S. *et al.* Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell* **134**, 866–876 (2008).
This paper describes an advanced N-terminal labelling method to map the precise sites of proteolytic cleavage in complex cellular systems.
85. Kleifeld, O. *et al.* Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nature Biotechnol.* **28**, 281–288.
86. Dix, M. M., Simon, G. M. & Cravatt, B. F. Global mapping of the topography and magnitude of proteolytic events in apoptosis. *Cell* **134**, 679–691 (2008).
This paper uses an integrated gel and MS platform to map proteolytic cleavage events and their effect on protein levels and structure.
87. Simon, G. M., Dix, M. M. & Cravatt, B. F. Comparative assessment of large-scale proteomic studies of apoptotic proteolysis. *ACS Chem. Biol.* **4**, 401–408 (2009).
88. Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol.* **17**, 994–999 (1999).
89. Khidekel, N., Ficarro, S. B., Peters, E. C. & Hsieh-Wilson, L. C. Exploring the O-GlcNAc proteome: direct identification of O-GlcNAc-modified proteins from the brain. *Proc. Natl Acad. Sci. USA* **101**, 13132–13137 (2004).
90. Martin, B. R. & Cravatt, B. F. Large-scale profiling of protein palmitoylation in mammalian cells. *Nature Methods* **6**, 135–138 (2009).
91. Hang, H. C. *et al.* Chemical probes for the rapid detection of Fatty-acylated proteins in Mammalian cells. *J. Am. Chem. Soc.* **129**, 2744–2745 (2007).
92. Ward, P. S. *et al.* The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting α-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **17**, 225–234.
93. Saghatelian, A. & Cravatt, B. F. Global strategies to integrate the proteome and metabolome. *Curr. Opin. Chem. Biol.* **9**, 62–68 (2005).
94. Vinayavekhin, N. & Saghatelian, A. Untargeted metabolomics. *Curr. Protoc Mol. Biol.* **30**, 1–24 (2010).
95. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* **78**, 779–787 (2006).
96. Wishart, D. S. *et al.* HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **37**, D603–D610 (2009).
97. Wishart, D. S. *et al.* HMDB: the Human Metabolome Database. *Nucleic Acids Res.* **35**, D521–D526 (2007).
98. Sana, T. R., Roark, J. C., Li, X., Waddell, K. & Fischer, S. M. Molecular formula and METLIN Personal Metabolite Database matching applied to the identification of compounds generated by LC/TOF-MS. *J. Biomol. Tech.* **19**, 258–266 (2008).
99. Smith, C. A. *et al.* METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **27**, 747–751 (2005).
100. Schmelzer, K., Fahy, E., Subramaniam, *S. &* Dennis, E. A. The lipid maps initiative in lipidomics. *Methods Enzymol.* **432**, 171–183 (2007).

Acknowledgements
This work was supported by the US National Institutes of Health (CA087660 and CA132630), the American Cancer Society (D.K.N.), the California Breast Cancer Research Foundation (M.M.D.), the ARCS Foundation (M.M.D.) and the Skaggs Institute for Chemical Biology.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Benjamin F. Cravatt's homepage:
http://www.scripps.edu/chemphys/cravatt/

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2010 Macmillan Publishers Limited. All rights reserved

INNOVATION

Molecular imaging by mass spectrometry — looking beyond classical histology

Kristina Schwamborn and Richard M. Caprioli

Abstract | Imaging mass spectrometry (IMS) using matrix-assisted laser desorption ionization (MALDI) is a new and effective tool for molecular studies of complex biological samples such as tissue sections. As histological features remain intact throughout the analysis of a section, distribution maps of multiple analytes can be correlated with histological and clinical features. Spatial molecular arrangements can be assessed without the need for target-specific reagents, allowing the discovery of diagnostic and prognostic markers of different cancer types and enabling the determination of effective therapies.

Mass spectrometry (MS) has been the focus of technology development and application for imaging for several decades, with early work in secondary ion mass spectrometry (SIMS) reported in the 1960s[1,2]. SIMS uses an atomic or molecular cluster beam to irradiate a sample with concomitant desorption of ions, which are subsequently analysed to measure their mass to charge ratio (m/z)[3]. This was later followed by laser desorption methods that used a focused laser to achieve the desorption process. One of the limitations of these techniques is that molecules with a molecular mass of more than approximately 2,000 Da cannot be desorbed intact, and so applications typically address the analysis of elements, atomic clusters, drugs and other low molecular mass compounds within tissue sections[4–6]. Matrix-assisted laser desorption ionization (MALDI)-based imaging mass spectrometry (IMS) was introduced in 1997 and provided the capability of measuring both low and high molecular mass compounds, such as proteins with masses of greater than 50,000 Da[7]. In general, the practical image resolution for the desorption technologies for the measurement of intact molecules is around 1 μm, and this is mostly limited by sensitivity.

Much of the early work in MALDI IMS was focused on proteomics in which proteins of molecular masses from 2 to 80 kDa or higher can be ablated and directly analysed intact from tissue sections. Their relative abundance and spatial distribution throughout the section can be measured with a lateral resolution of up to 10 μm using current technology[8,9]. Formalin-fixed and paraffin-embedded (FFPE) samples, constituting most of the tissue collected and stored by pathologists worldwide, can also be analysed by MALDI IMS by combining commonly used antigen retrieval techniques, such as heat-induced antigen retrieval coupled with tryptic digestion of proteins *in situ*[10,11]. In contrast to more commonly used imaging modalities, such as immunohistochemistry (IHC), MALDI IMS allows for the unbiased analysis of tissue sections for the discovery and identification of proteins that are linked to specific tissue types or disease phenotypes as it does not require target-specific reagents[12]. The spatial distribution of hundreds of proteins can be evaluated in parallel and visualized from the same section[13]. By leaving molecular distributions intact, MALDI IMS enables the discovery of spatial molecular arrangements in disease to help assess, for example, the aggressiveness of the disease in discrete areas in the tissue and ultimately to substantially enable the prediction of patient outcome.

MALDI IMS is becoming widely accepted, as shown by the rapidly increasing number of publications in recent years. This interest is driven by the need to gain a better and more fundamental understanding of molecular events involving proteins as well as other types of biological compounds in the development and progression of disease, especially cancer. As the proteome is more complex and dynamic than the genome, it is an effective sensor for ongoing molecular alterations[14,15]. MALDI IMS technology has the capability to overcome some of the limitations of other approaches in the identification of new biomarker molecules that are involved in the development and progression of cancer. Its major strength lies in its ability to measure molecules with direct correlation to anatomical features in tissues at the cellular level, bypassing cumbersome and time-consuming preparation techniques such as laser capture microdissection[16]. This is of particular interest for the analysis of small histological features that would require many sections to be microdissected to acquire the necessary number of cells for proteomic analysis. Another key feature of MALDI IMS is its ability to discover molecular signatures of disease; these signatures typically comprise 5–20 or more different proteins that together result in robust diagnostic patterns[17]. Several studies have shown the usefulness and advantages of this technology in the field of cancer research as an aid to diagnosis or for ascertaining prognosis of different cancer types, as well as in the determination of the effectiveness of a therapeutic regime. This Perspective article highlights some recent publications using MALDI IMS in the field of cancer research, but it is not intended as a complete review of the field.

Technology

MS is recognized as a powerful tool for analysing various different analytes, including small molecules, lipids, DNA segments, peptides and proteins[18–24]. MALDI IMS is amenable to the analysis of high molecular mass biologicals of 100 kDa or more, as well as small molecules with a molecular mass of

© 2010 Macmillan Publishers Limited. All rights reserved

less than 1 kDa. For analysis by MALDI, a sample is mixed or coated with an energy-absorbing matrix and subsequently irradiated with a laser beam. In positive ionization mode, singly protonated molecular ions ($[M+H]^+$) are generated from analytes in the sample, whereas in negative ionization mode singly deprotonated ions ($[M-H]^-$) are generated. These are subsequently detected by one of several different types of mass analysers. Time-of-flight (TOF) analysers are commonly used in MALDI IMS to measure the ionized analytes. Following acceleration at a fixed potential[25], ions are separated and recorded based on their mass (molecular mass) to charge ratio (*m/z*) (FIG. 1). As the charge for MALDI is typically one, the analyser read-out measures the protonated or deprotonated molecular mass of the molecule.

The introduction of MALDI IMS first enabled the analysis of intact biological tissue sections in either an ordered array of spots to obtain an image or at selected positions of interest (known as histology directed profiling)[7,26]. The major advantages of these approaches are that they enable the direct correlation of mass spectra with anatomical or pathological features as tissue sections remain intact throughout the analysis and can be stained after MS analysis if necessary[27,28]. Combining well-established antigen retrieval techniques and on-tissue tryptic digestion allows FFPE tissue samples to be analysed by MALDI IMS analysis through the measurement of the resulting peptides[11]. FFPE samples constitute the vast majority of samples that are available in worldwide tissue banks in which millions of samples exist. This type of fixation has also become the standard practice in pathology — samples are processed and embedded in a standardized manner and can be easily stored at room temperature. Moreover, samples from rare diseases are available in addition to the corresponding clinical information[29].

Sample handling and preparation are crucial for achieving high quality and reproducibility in MALDI IMS studies[30,31]. Preanalytical steps should be standardized and carried out with great care to achieve a high degree of reproducibility. Thin tissue sections can be collected on various sample targets (such as a gold plate or glass slide with conductive surface) and subsequently coated with the matrix. Matrix application can be achieved in several different ways such as dry coating, sublimation, spray coating, and manual and robotic spotting[32–35]. The preferred method of matrix application depends on the analyte. Although dry coating and sublimation are the favoured matrix application techniques for lipid analysis, they are not optimal for protein and peptide analyses; these samples require spray coating or spotting techniques. For these analytes, high quality mass spectra are obtained when the matrix solution crystallizes on the tissue and thereby 'traps' these analytes in the crystals themselves. In FFPE tissue, matrix application is preceded by on-tissue tryptic digestion to release peptide domains that are not cross-linked[36]. A typical workflow for the analysis of fresh frozen tissue is shown in FIG. 2. The acquired data can be visualized by depicting single proteins or peptides (producing protein or peptide images) or by using statistical methods to combine multiple proteins in a model to reveal a signature of key molecular changes within the tissue (producing classification images).

The use of different classification algorithms facilitates data analysis to generate images of tissue sections through two different general approaches: unsupervised (if no knowledge of the tissue composition is available) and supervised (if spatial details of the section are known)[37,38]. Using virtual z-stacks and three-dimensional volume rendering allows the reconstruction of MALDI IMS data to generate three-dimensional volumes, and these can be co-registered with additional data such as block face optical images and magnetic resonance imaging data[39,40].

In summary, MALDI IMS is a versatile technology that can be used for the direct analysis of biological compounds from tissue sections. Although MALDI has been successfully implemented for certain aspects of DNA analysis (such as the analysis of DNA methylation, small oligonucleotides and single nucleotide polymorphisms[22]), analysis of intact DNA from tissue sections has not yet been achieved. MALDI IMS studies have mostly focused on proteins and small molecules in a wide variety of applications, ranging from fundamental biology to clinical studies. Proteomic data obtained from intact tissue sections can be directly correlated with the histology.

Despite impressive advances in MALDI IMS technology, challenges remain to further increase the sensitivity in order to sample deeper into the proteome. The proteins detected are typically at mid- to high-abundance levels. In addition, higher spatial resolution needs to be achieved, ultimately

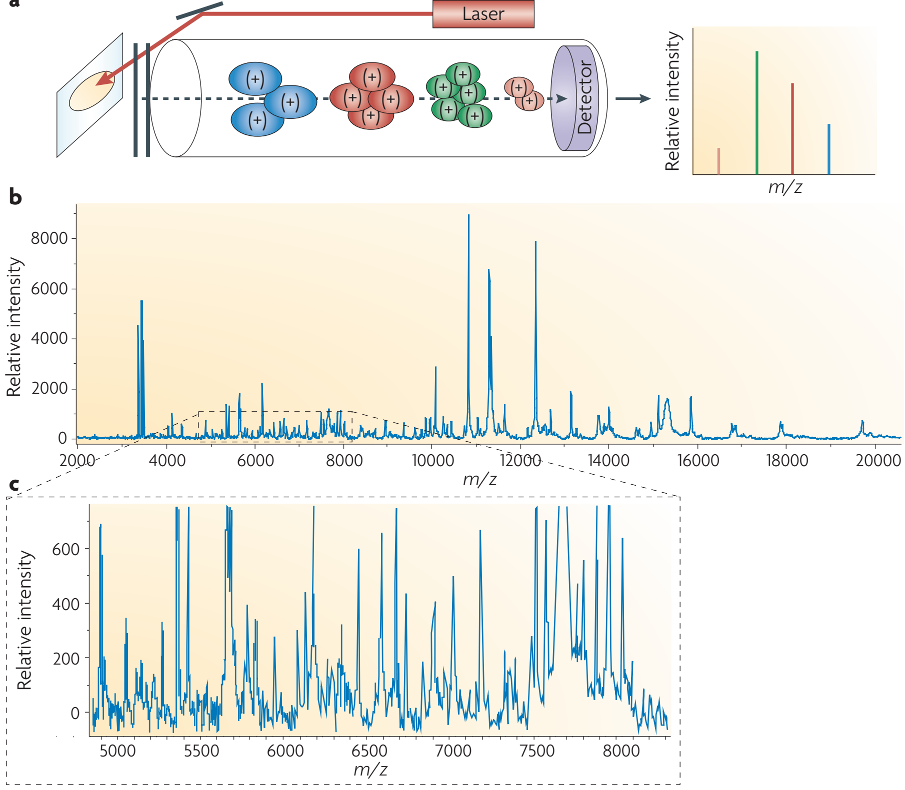


Figure 1 | **Principles of MALDI-TOF IMS. a** | Schematic outline of a typical separation of analytes in a linear matrix-assisted laser desorption ionization (MALDI)-time-of-flight (TOF) imaging mass spectrometer (IMS) based on their mass to charge ratio (*m/z*). **b** | Typical mass spectrum in the mass range between 2 and 20 kDa. **c** | A magnification of the mass range between 5 and 8 kDa.

© 2010 Macmillan Publishers Limited. All rights reserved

enabling subcellular analyses. Current limitations in matrix application (such as the size of matrix drops applied to the tissue), as well as laser beam sizes, limit the achievable lateral image resolution to 10–20 μm. Although IHC can attain higher spatial resolution, MALDI IMS has the capability of analysing hundreds of analytes in parallel. A particular strength of MALDI IMS is its independence from the requirement of antibodies. As analytes can be visualized without the need for upstream labelling approaches, discovery becomes an integral advantage of MALDI IMS.

Oncology applications

Applications of MALDI IMS in the field of cancer are diverse and are mainly focused on the analysis of peptides, proteins and small molecules. Some recent studies are discussed below that emphasize specific aspects of MALDI IMS technology. For protein analysis, three different types of studies can be distinguished: diagnostic studies comparing different tissue types (such as tumour versus normal) to aid in pathological diagnosis; prognostic studies to categorize patients with long- or short-term survival; and drug response studies to predict a patient's response to a certain treatment.

MALDI IMS-based studies have been used to elucidate molecular signatures from samples with different tumour types and grades, including brain[41], oral[42], lung[36,43], breast[44,45], gastric[46], pancreatic[47], renal[48], ovarian[49,50] and prostate cancer[28,51]. These studies have been conducted in a retrospective manner and are mainly focused on the identification of molecular signatures of a disease or disease status. As these molecular signatures are comprised of peptides and proteins, identification of the proteins is important, and this can be validated by IHC or western blot. As the technology provides an unbiased method for the analysis of samples as well as multiplexed spatially resolved molecular information, it allows the study of differences in molecular expression levels between adjacent anatomical and pathological structures without the need for labelling approaches, as is the case for IHC. As no a priori knowledge of the proteins is necessary, the discovery of new, unexpected proteins and pathways can be accomplished.

Peptide and protein analysis — diagnostic studies. Studies using MALDI IMS have revealed differences in protein expression between tumour tissue and the immediately adjacent normal tissue[52]. Most interestingly, these differences could still be observed in the surrounding histologically normal tissue,

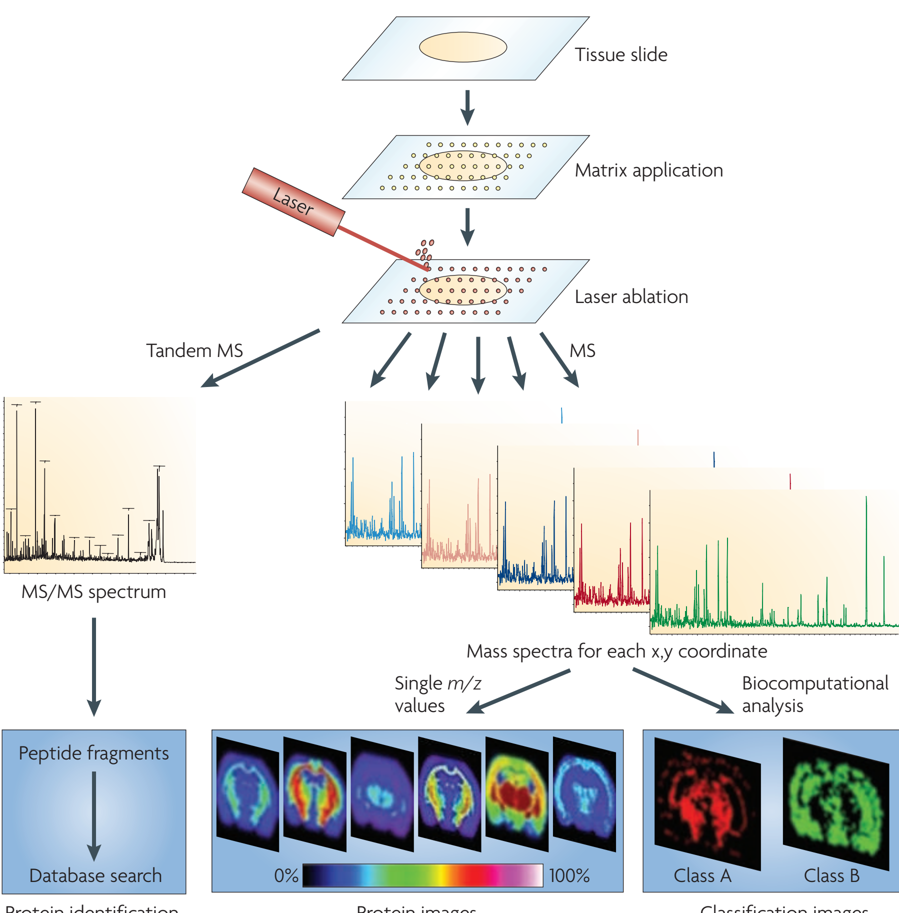


Figure 2 | **Workflow for MALDI IMS analysis.** Schematic outline of a typical workflow for fresh frozen tissue samples. Sample pretreatment steps include cutting and mounting the tissue section on a conductive target. Matrix is applied in an ordered array across the tissue section and mass spectra are generated at each x,y coordinate for protein analysis or tandem MS (MS/MS) spectra for protein identification. Further analytical steps include the visualization of the distribution of a single protein within the tissue (protein image) or statistical analysis to visualize classification images as well as database searching to identify the protein. The scale represents the relative intensity of the protein. IMS, imaging mass spectrometry; MALDI, matrix-assisted laser desorption ionization; MS, mass spectrometry; m/z, mass to charge ratio.

decreasing with distance to 1.5 cm beyond the tumour[53]. For example, the molecular assessment of clear cell renal cell carcinoma (ccRCC) samples ($n = 34$), comprising tumour and adjacent normal tissue in the same section, revealed molecular changes in the normal tissue that was adjacent to the tumour that were similar to changes observed in the tumour itself[48]. One group of proteins, which has been shown to be involved in the mitochondrial electron transport system (such as cytochrome *c*; $m/z = 12{,}272$), was consistently underexpressed in tumour and adjacent histologically normal tissue. These findings directly address the issue of the molecular assessment of surgical tumour margins and the desire to reduce local tumour recurrence.

The application of MALDI IMS for diagnostic purposes to differentiate between tumour and normal tissue or different tumour subtypes has become an emerging field as investigators seek to find better markers to aid diagnosis. In a study on prostate cancer, investigators compared a total of 31 fresh frozen prostate cancer and 41 normal prostate biopsy samples to identify differentially expressed peptides[51]. Combining 3 peptides in a genetic algorithm-based model resulted in the correct classification of cancerous areas in 85% of the discovery set (cancer = 11; normal = 10) and 81% of the validation set (cancer = 23; normal = 31). One of the proteins identified was MEKK2 ($m/z = 4{,}355$). These findings were further validated by western blot analysis and IHC, both of which verified that MEKK2 was overexpressed in prostate cancer tissue and cell lines.

In a similar study of ovarian cancer, fresh frozen ovarian cancer tissue samples ($n = 25$) were analysed in comparison to benign

© 2010 Macmillan Publishers Limited. All rights reserved

ovarian tumours ($n = 23$)[49]. A fragment of the 11S proteasome activator complex REGα (also known as proteasome activator subunit 1 (PSME1); $m/z = 9{,}744$) was found to be overexpressed in the cancer samples. Results were confirmed by western blot analysis as well as IHC. IHC revealed a distinct and diagnosis-dependent localization in cellular compartments, with cytoplasmic localization of REGα in carcinomas as well as nuclear staining but no cytoplasmic staining in 76.9% of benign tumours. This is the first report describing high expression levels of REGα in ovarian cancer. The 11S regulator complex (PA28; of which REGα is a subunit) binds to the 20S proteasome, and downregulation of PA28 in tumour cells has been shown to result in an impaired presentation of tumour antigens[54].

Two separate studies of non-small-cell lung cancer (NSCLC) were able to identify proteomic patterns to accurately classify and predict histopathological groups (such as adenocarcinoma and squamous cell carcinoma)[36,43]. In the first study, investigators analysed fresh frozen samples from 79 lung tumours and 14 normal lung tissues[43]. Using a training set of 42 tumour samples and 8 normal samples, investigators created a model that could correctly classify all samples in the training set as well as all samples in an independent validation set (37 cancer and 6 normal samples). Moreover, expression profiles of two proteins also allowed the classification of tumours with and without lymph node metastasis with 85% and 75% accuracy in the training and validation sets, respectively. Combining 15 peaks in a proteomic pattern enabled the distinction between patients with NSCLC with poor ($n = 25$) and good ($n = 41$) prognosis ($p<0.0001$). Two proteins ($m/z = 10{,}519$ and $m/z = 4{,}964$) that were highly expressed in primary NSCLC were identified as SUMO2 and thymosin-β4, X-linked (TMSβ4X), respectively.

Another study of NSCLC demonstrated the use of FFPE tissue microarray (TMA) samples for high-throughput analysis and classification by means of on-tissue tryptic digestions and MALDI IMS[36]. Using a support vector machine algorithm-based model enabled the correct classification of all squamous cell carcinoma ($n = 22$) and all adenocarcinoma ($n = 18$) samples. MALDI-tandem mass spectrometry (MS/MS) analysis directly from the TMA section allowed the identification of approximately 50 proteins. For example, heat shock protein β1 (HSPB1) identified by three tryptic peptides, was found to be almost exclusively expressed in squamous cell carcinomas. All three tryptic peptides showed a similar ion density distribution across the TMA, thereby corroborating the derivation of these peptides from the same protein. Two other peptides ($m/z = 1{,}407.7$ and $m/z = 1{,}410.7$) were identified as a fragment of cytokeratin 6 (CK6) and CK5 (also known as keratin 5), respectively. These two antigens are usually expressed in squamous cell carcinoma, whereas adenocarcinomas are mainly CK5- and CK6-negative. Commercially available antibodies stain both antigens. However, a wide variation in the expression of both cytokeratins in patients was revealed when comparing MALDI IMS-based peptide images of CK5 and CK6 with IHC staining using an antibody against CK5 and CK6 (REF. 75) (FIG. 3).

In breast cancer, MALDI IMS has been used to classify ERBB2 (also known as HER2) receptor status on stored fresh frozen tissue samples. The assessment of ERBB2 expression in breast cancer is crucially important for treatment decisions in newly diagnosed patients with primary breast cancer to predict which patients are most likely to respond to trastuzumab therapy[44]. Fresh frozen samples from 48 patients were analysed. The expression of ERBB2 was evaluated by IHC and fluorescence *in situ* hybridization in the training set (ERBB2-positive $n = 15$ and ERBB2-negative $n = 15$) and by IHC only in the validation set (ERBB2-positive $n = 6$ and ERBB2-negative $n = 12$). Classification of the ERBB2 receptor status in the training set using an artificial neuronal network-based model resulted in 87% sensitivity and 93% specificity. Applying the same model to predict the ERBB2 receptor status in the validation set resulted in the correct classification of 16 of the 18 cases (83% sensitivity and 92% specificity). In addition, a protein overexpressed in ERBB2-positive samples was identified as cysteine-rich intestinal protein 1 (CRIP1; $m/z = 8{,}403$). This result confirms previous findings of the co-expression of *CRIP1* and *ERBB2* mRNAs[55].

Peptide and protein analysis — prognostic studies. In a study on patient outcome, samples from 108 patients with glioma and 19 non-cancer patients were subjected to MALDI IMS analysis to obtain protein signatures that correlated with tumour histology and patient survival[41]. In addition to being able to distinguish between non-cancer subjects and patients with glioma with an average of >92% accuracy, investigators could also identify a proteomic signature (24 peaks) to differentiate between patients from two prognostic groups: a short-term survival group (52 patients; mean survival of <15 months) and a long-term survival group (56 patients; mean survival of >90 months). Multivariate analysis showed a strong correlation between the identified protein expression signature and patient survival, demonstrating that the protein expression signature is an independent indicator of patient survival. Additionally, for a subgroup of the patients with glioma — those with glioblastoma ($n = 57$) — two different peaks

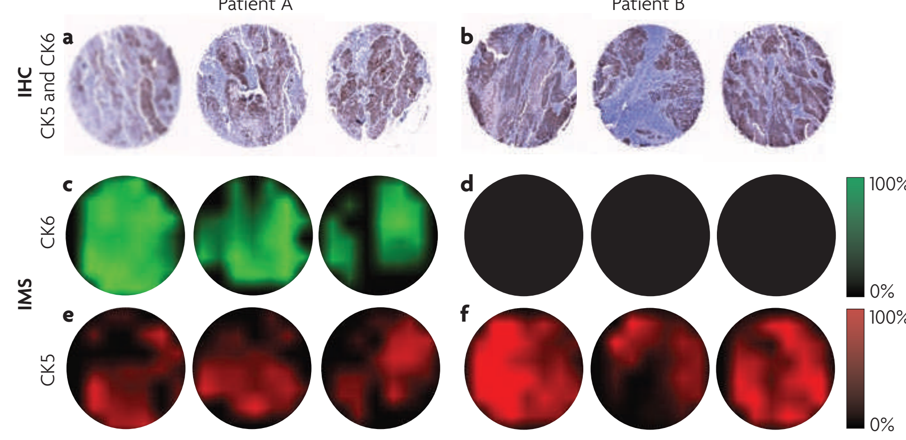


Figure 3 | **Comparison of IHC and MALDI IMS results.** Immunohistochemical (IHC) staining with an antibody against cytokeratin 5 (CK5) and CK6 of triplicate lung cancer biopsy samples from two different patients (patient A or patient B) showing positive staining in all biopsy samples (parts **a** and **b**). The corresponding peptide image of a 1407.7 Da fragment of CK6 (green; parts **c** and **d**) reveals this peptide to be present only in one patient, whereas the 1410.7 Da fragment of CK5 (red; parts **e** and **f**) is present in all biopsy samples[75]. The scale represents the relative intensity of the peptide. IMS, imaging mass spectrometry; MALDI, matrix-assisted laser desorption ionization. Image courtesy of M.R. Groseclose, Vanderbilt University, USA.

© 2010 Macmillan Publishers Limited. All rights reserved

could separate patients with short-term survival (28 patients; average survival of 10.9 months) from patients with long-term survival (29 patients; average survival of 16.8 months). This proteomic signature could also be verified as an independent predictor of patient survival by a multivariate Cox proportional hazards model. In total, six proteins could be identified by high-performance liquid chromatography (HPLC) and MALDI-MS/MS in this study: calcyclin (also known as S100A6; *m/z* = 10,092), dynein light chain LC8-type 2 (DLC8B; also known as DYMLL2; *m/z* = 10,262), calpactin I light chain (also known as S100A10; *m/z* = 11,073), phosphoprotein-enriched in astrocytes (PEA15; *m/z* = 15,035), fatty acid-binding protein 5 (FABP5; *m/z* = 15,076) and tubulin-folding cofactor A (TBCA; *m/z* = 17,268). Two of these proteins, S100A6 and DLC8B, were among the classifiers for overall patient prognosis of all patients with glioma, with S100A6 overexpressed in patients with short-term survival and DLC8B predominant in patients with long-term survival. S100A6, a protein thought to be involved in tumorigenesis and cell proliferation, has previously been described as overexpressed in other tumours, such as colon cancer, and its expression level in epithelial cells was found to be directly proportional to the increase in malignancy[56,57]. DLC8B has been reported to interact with BIM, a pro-apoptotic protein, to negatively regulate its apoptotic function[58].

Peptide and protein analysis — drug response studies. Another goal of MALDI IMS-based proteomic studies is the identification of markers that correlate with patient response to a therapeutic regime. MALDI IMS, together with gene expression profiling, has been used to identify markers of taxane sensitivity in patients with breast cancer receiving a neoadjuvant therapeutic treatment regime of the taxane paclitaxel with radiation[45]. Fresh frozen pretreatment tissue samples from 19 patients were subjected to histology-directed profiling. Of these 19 patients, 6 achieved a pathological complete response and 13 showed residual disease or no response after treatment. Comparing spectra from tumour regions between the two groups revealed three highly overexpressed (>30-fold) features (*m/z* = 3,371, *m/z* = 3,442 and *m/z* = 3,485) and four features of lower expression (*m/z* = 5,667, *m/z* = 6,955, *m/z* = 7,007 and *m/z* = 15,348) in the responder group. The three significantly overexpressed proteins had previously been identified as α-defensins (DEFα1, *m/z* = 3,371; DEFα2, *m/z* = 3,442; and DEFα3, *m/z* = 3,485) in proteomic analyses of different types of cancer, such as head and neck cancer[59]. To verify the tumour cells as the source of the DEFα expression and to eliminate the possibility of an artefact caused by blood contamination of the tissue or the presence of neutrophils, IHC was carried out. Samples from patients with pathological complete responses exhibited a positive staining pattern of DEFα in the tumour cells, whereas samples from non-responders showed little or no staining of DEFα.

In another study it was shown that early proteomic changes assessed by MALDI IMS could predict the treatment response of mammary tumours in transgenic mice[60]. Fresh frozen tumour sections from mouse mammary tumour virus (MMTV);*Erbb2*-transgenic mice treated with different doses of a small molecule tyrosine kinase inhibitor of epidermal growth factor receptor (EGFR), erlotinib, and/or an ERBB2-blocking antibody (trastuzumab), were analysed by MALDI IMS at various time points and compared with tumour sections from untreated mice. Tumours from mice treated with erlotinib showed a dose- and time-dependent decrease in TMSβ4X (*m/z* = 4,965), as well as ubiquitin (*m/z* = 8,565) and an increase of a fragment of E-cadherin-binding protein E7 (also known as c-CBL-like protein 1 (CBLL1); *m/z* = 4,794) when compared with tumours from untreated mice. Those drug-induced changes precede the inhibition of proliferation or cell death, as measured by IHC (using the detection of proliferating cell nuclear antigen (PCNA)) and terminal deoxynucleotidyl transferase (Tdt)-mediated nick end labelling (TUNEL). Additional proteins found to be more highly expressed in tumours treated with both drugs include a fragment of calmodulin, acyl-CoA binding protein (also known as DBI), calgizzarin (also known as S100A11), histone H3 and histone H4.

Lipid analysis. Interest in assessing lipid changes in disease has increased considerably owing to their important functions in signal transduction and energy storage[61]. Altered levels of lipids are found in many human diseases, including cancer[62]. MALDI IMS for spatially resolved analysis of lipids has also grown rapidly as the technology has developed. Owing to their low molecular mass (<1,000 Da), *in situ* analysis can be hampered by interference from other components in the tissue and through the normal tissue preservation protocols used in histology[61]. Initial MALDI IMS studies using fresh frozen tissue sections suggest that regional differences in concentrations for specific molecular species of lipids can be detected in tissues[63,64]. Investigators were able to analyse the distribution of mitochondrion-specific lipids, the cardiolipins, in different rat organ sections[65]. Work from our laboratory in studies of biopsy samples from patients with ccRCC showed marked differences in lipid distributions between tumour and normal regions (unpublished observations, S. Puolitaival and R.M.C.). In total, 78 matched pairs were analysed in positive and negative ionization mode. Most linoleic acid-containing phospholipids were found to be more abundant in tumour regions, whereas sphingomyelins were more abundant in normal regions. Although these findings are interesting in terms of understanding the total molecular changes that occur in tumorigenesis, their specific role in this process remains unclear. In a recently published non-cancer-related study, MALDI IMS was used to successfully characterize the spatial and temporal distribution of phospholipid species during mouse embryo implantation[66].

Applications of MALDI IMS to drug discovery. The assessment of the distribution and metabolism of drug candidates in targeted tissues and throughout the body is of central interest to drug discovery and in the understanding of drug responses[67,68]. Drug analysis uses MS/MS that allows both the molecular species and the structure-specific fragments to be simultaneously monitored to increase sensitivity and better validate the identification of signals with high confidence. Single reaction monitoring (SRM) and multiple reaction monitoring (MRM) techniques further enable this analysis to be accomplished in a high-throughput manner[69]. These techniques monitor the structure-related composition of the drug (precursor or parent ion) in the mass spectrometer to form one (SRM) or multiple (MRM) specific fragment ions.

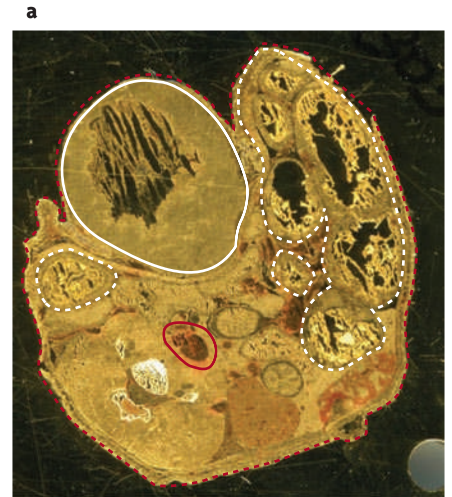
Minimizing the tissue handling and washing steps that are typically a part of sample pretreatment protocols for the analysis of proteins is crucial for the detection of small molecules, as such treatment can degrade or compromise the level of the drug of interest. Initial studies by MALDI IMS have shown the possibility of detecting not only the distribution of the drug itself but also the simultaneous distribution of its individual metabolites in whole-body tissue sections[70]. FIGURE 4 shows an example of small

© 2010 Macmillan Publishers Limited. All rights reserved

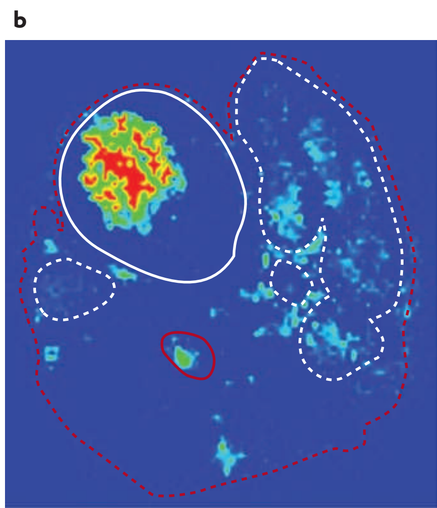
molecule imaging in whole-animal sections in which the orally administered anti-cancer drug is present in the digestive system as well as the tumour (personal communication, M. L. Reyzer, R. A. Smith and H. C. Manning).

A recently published study used MALDI IMS to detect the localization of the alkylating agent oxaliplatin and its derivates in heated intra-operative chemotherapy-like treated rat kidneys[71]. The lowest concentration of oxaliplatin detectable by MALDI IMS was determined as 0.23 ± 0.05 mg of total oxaliplatin per gram of tissue using a 15 μm- thick tissue section. Imaging results showed that oxaliplatin and its derivates could be detected within the kidney cortex only. Penetration of the drug into the medulla was not observed. Another study imaged the distribution of vinblastine within rat whole-body sections in order to visualize the distribution of the drug in different organs[72]. Ion density maps of vinblastine (precursor ion; $m/z = 811.4$) and its fragments (for example, $m/z = 793$) showed a higher distribution in the liver, kidney and tissue surrounding the gastrointestinal tract. However, especially in the kidney, differences between the distribution of the precursor ion together with its fragment at $m/z = 751$ were observed when compared with the other seven fragments, with the first showing a relatively high signal intensity in the renal pelvis. The distribution of orally administered erlotinib and its metabolites was investigated in tissue sections from rat liver, spleen and muscle by MALDI IMS[67]. The highest drug concentrations were found in liver sections together with the detection of one of the major drug metabolites. In addition, direct quantitative analysis of the drug on tissue (liver, spleen and muscle) by MALDI IMS was carried out and compared with standard LC–MS/MS on tissue homogenates. The ratio of total ion intensities from liver and spleen, as estimated by MALDI IMS analysis, was in good agreement with the ratio calculated by LC–MS/MS analysis. One study of tumour response to drug treatment also analysed the spatial localization of elortinib directly in tissue sections by SRM of the transition $m/z = 394.2$ to $m/z = 278.1$ (REF. 60). The drug was predominantly found to be distributed in the more vascularized peripheral areas of the tumour section 16 hours after drug administration.

a



b



c

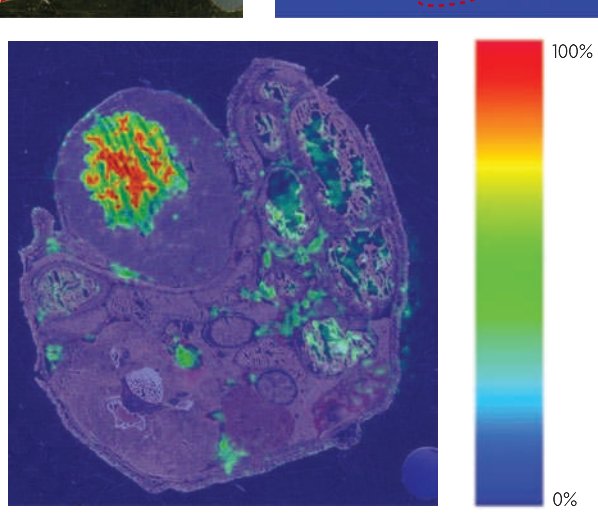


Figure 4 | **Small molecule imaging.** Unstained transverse whole mouse section (part **a**), image of the drug distribution (part **b**) and overlay (part **c**). The orally administered anti-cancer drug is present in high concentrations in the tumour and can also be found in lower concentrations in the digestive system and blood vessels. The solid white line represents the tumour; dashed white line the gastrointestinal tract; solid red line the major blood vessels and the dashed red line the outline of the whole section. The scale represents the relative intensity of the drug. Personal communication, M. L. Reyzer, R. A. Smith and H. C. Manning, Vanderbilt University, USA.

Conclusions and perspectives

MALDI IMS is a new technology with respect to defined clinical applications. Its outstanding molecular recognition capabilities should substantially benefit molecular pathology as the technology develops further. Such development would include the implementation of biocomputational tools as well as the building and validation of extensive databases that correlate molecular signatures with histological and clinical data such as patient outcome. Many studies have shown the strength of this emerging technology and its ability to be used to aid in the diagnosis and prognosis of several different cancer types. In addition to discovering potential new drug targets, MALDI IMS also offers the possibility of analysing the distribution and effect of drug candidates directly in tissues and whole animals. The most important characteristics of MALDI IMS in a discovery setting are its multiplexicity and its independence from target-specific reagents such as antibodies. Further improvements in the technology will bring an increased ability to measure proteins of lower expression levels in tissue sections. Also, as image resolution increases, smaller laser spot sizes (that are subcellular, for example) will require even greater advances in sensitivity as the amount of material analysed per spot will decrease. Progress in many areas has already been shown; for example, in the development of special sample pretreatment protocols to allow the detection of hydrophobic and membrane-bound proteins[73,74].

As this technology has proved its feasibility and versatility, the next phase in its clinical use will be its translation to effective applications in the clinic. More comprehensive studies with many hundreds of samples are needed to validate the robustness and potential use in a clinical setting in support of personalized medicine. The technology holds

© 2010 Macmillan Publishers Limited. All rights reserved

great promise for establishing protein or lipid signatures as robust aids in disease diagnosis and most importantly for prediction of a patient's response to therapy and overall prognosis. MALDI IMS, perhaps in conjunction with other approaches, can clearly bring the art and practice of molecular pathology to a new and more effective level.

Kristina Schwamborn and Richard M. Caprioli are at the Mass Spectrometry Research Center, Department of Biochemistry, Vanderbilt University, Nashville, Tennessee 37232-2195, USA.
Correspondence to R.M.C.
e-mail: richard.m.caprioli@vanderbilt.edu

doi:10.1038/nrc2917
Published online 19 August 2010

1. Liebl, H. Ion microprobe mass analyzer. *J. Appl. Phys.* **38**, 5277–5283 (1967).
2. Castaing, R. & Slodzian, G. Microanalysis by secondary ionic emission. *J. Microsc.* **1**, 395–410 (1962).
3. Esquenazi, E., Yang, Y. L., Watrous, J., Gerwick, W. H. & Dorrestein, P. C. Imaging mass spectrometry of natural products. *Nat. Prod. Rep.* **26**, 1521–1534 (2009).
4. Fragu, P., Klijanienko, J., Gandia, D., Halpern, S. & Armand, J. P. Quantitative mapping of 4′-iododeoxyrubicin in metastatic squamous cellcarcinoma by secondary ion mass spectrometry (SIMS) microscopy. *Cancer Res.* **52**, 974–977 (1992).
5. Pacholski, M. L. & Winograd, N. Imaging with mass spectrometry. *Chem. Rev.* **99**, 2977–3006 (1999).
6. Chandra, S. & Lorey, D. R. SIMS ion microscopy in cancer research: single cell isotopic imaging for chemical composition, cytotoxicity and cell cycle recognition. *Cell. Mol. Biol. (Noisy-le-grand)* **47**, 503–518 (2001).
7. Caprioli, R. M., Farmer, T. B. & Gile, J. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Anal. Chem.* **69**, 4751–4760 (1997).
8. Seeley, E. H. & Caprioli, R. M. Molecular imaging of proteins in tissues by mass spectrometry. *Proc. Natl Acad. Sci. USA* **105**, 18126–18131 (2008).
9. Chaurand, P., Schriver, K. E. & Caprioli, R. M. Instrument design and characterization for high resolution MALDI-MS imaging of tissue sections. *J. Mass Spectrom.* **42**, 476–489 (2007).
10. Lemaire, R. *et al.* Direct analysis and MALDI imaging of formalin-fixed, paraffin-embedded tissue sections. *J. Proteome Res.* **6**, 1295–1305 (2007).
11. Ronci, M. *et al.* Protein unlocking procedures of formalin-fixed paraffin-embedded tissues: application to MALDI-TOF imaging MS investigations. *Proteomics* **8**, 3702–3714 (2008).
12. Zimmerman, T. A., Monroe, E. B., Tucker, K. R., Rubakhin, S. S. & Sweedler, J. V. Chapter 13: imaging of cells and tissues with mass spectrometry: adding chemical information to imaging. *Methods Cell Biol.* **89**, 361–390 (2008).
13. Chaurand, P., Sanders, M. E., Jensen, R. A. & Caprioli, R. M. Proteomics in diagnostic pathology: profiling and imaging proteins directly in tissue sections. *Am. J. Pathol.* **165**, 1057–1068 (2004).
14. Celis, J. E. & Gromov, P. Proteomics in translational cancer research: toward an integrated approach. *Cancer Cell* **3**, 9–15 (2003).
15. Hanash, S. Disease proteomics. *Nature* **422**, 226–232 (2003).
16. Emmert-Buck, M. R. *et al.* Laser capture microdissection. *Science* **274**, 998–1001 (1996).
17. Caprioli, R. M. Deciphering protein molecular signatures in cancer tissues to aid in diagnosis, prognosis, and therapy. *Cancer Res.* **65**, 10642–10645 (2005).
18. Siuzdak, G. The emergence of mass spectrometry in biochemical research. *Proc. Natl Acad. Sci. USA* **91**, 11290–11297 (1994).
19. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
20. Schiller, J. *et al.* Matrix-assisted laser desorption and ionization time-of-flight (MALDI-TOF) mass spectrometry in lipid and phospholipid research. *Prog. Lipid Res.* **43**, 449–488 (2004).
21. Harvey, D. J. *et al.* Comparison of fragmentation modes for the structural determination of complex oligosaccharides ionized by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **9**, 1556–1561 (1995).
22. Tost, J. & Gut, I. G. DNA analysis by mass spectrometry-past, present and future. *J. Mass Spectrom.* **41**, 981–995 (2006).
23. Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **60**, 2299–2301 (1988).
24. Tanaka, K. *et al.* Protein and polymer analyses up to *m/z*100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 151–153 (1988).
25. Mann, M., Hendrickson, R. C. & Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **70**, 437–473 (2001).
26. Cornett, D. S. *et al.* A novel histology-directed strategy for MALDI-MS tissue profiling that improves throughput and cellular specificity in human breast cancer. *Mol. Cell. Proteomics* **5**, 1975–1983 (2006).
27. Crecelius, A. C. *et al.* Three-dimensional visualization of protein expression in mouse brain structures using imaging mass spectrometry. *J. Am. Soc. Mass Spectrom.* **16**, 1093–1099 (2005).
28. Schwamborn, K. *et al.* Identifying prostate carcinoma by MALDI-imaging. *Int. J. Mol. Med.* **20**, 155–159 (2007).
29. Becker, K. F. *et al.* Quantitative protein analysis from formalin-fixed tissues: implications for translational clinical research and nanoscale molecular diagnosis. *J. Pathol.* **211**, 370–378 (2007).
30. Cornett, D. S., Reyzer, M. L., Chaurand, P. & Caprioli, R. M. MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. *Nature Methods* **4**, 828–833 (2007).
31. Schwartz, S. A., Reyzer, M. L. & Caprioli, R. M. Direct tissue analysis using matrix-assisted laser desorption/ionization mass spectrometry: practical aspects of sample preparation. *J. Mass Spectrom.* **38**, 699–708 (2003).
32. Puolitaival, S. M., Burnum, K. E., Cornett, D. S. & Caprioli, R. M. Solvent-free matrix dry-coating for MALDI imaging of phospholipids. *J. Am. Soc. Mass Spectrom.* **19**, 882–886 (2008).
33. Chaurand, P. *et al.* Integrating histology and imaging mass spectrometry. *Anal. Chem.* **76**, 1145–1155 (2004).
34. Aerni, H. R., Cornett, D. S. & Caprioli, R. M. Automated acoustic matrix deposition for MALDI sample preparation. *Anal. Chem.* **78**, 827–834 (2006).
35. Hankin, J. A., Barkley, R. M. & Murphy, R. C. Sublimation as a method of matrix application for mass spectrometric imaging. *J. Am. Soc. Mass Spectrom.* **18**, 1646–1652 (2007).
36. Groseclose, M. R., Massion, P. P., Chaurand, P. & Caprioli, R. M. High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry. *Proteomics* **8**, 3715–3724 (2008).
37. Deininger, S. O., Ebert, M. P., Futterer, A., Gerhard, M. & Rocken, C. MALDI imaging combined with hierarchical clustering as a new tool for the interpretation of complex human cancers. *J. Proteome Res.* **7**, 5230–5236 (2008).
38. Hanselmann, M. *et al.* Toward digital staining using imaging mass spectrometry and random forests. *J. Proteome Res.* **8**, 3558–3567 (2009).
39. Andersson, M., Groseclose, M. R., Deutch, A. Y. & Caprioli, R. M. Imaging mass spectrometry of proteins and peptides: 3D volume reconstruction. *Nature Methods* **5**, 101–108 (2008).
40. Sinha, T. K. *et al.* Integrating spatially resolved three-dimensional MALDI IMS with *in vivo* magnetic resonance imaging. *Nature Methods* **5**, 57–59 (2008).
41. Schwartz, S. A. *et al.* Proteomic-based prognosis of brain tumor patients using direct-tissue matrix-assisted laser desorption ionization mass spectrometry. *Cancer Res.* **65**, 7674–7681 (2005).
42. Patel, S. A. *et al.* Imaging mass spectrometry using chemical inkjet printing reveals differential protein expression in human oral squamous cell carcinoma. *Analyst* **134**, 301–307 (2009).
43. Yanagisawa, K. *et al.* Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* **362**, 433–439 (2003).
44. Rauser, S. *et al.* Classification of HER2 receptor status in breast cancer tissues by MALDI imaging mass spectrometry. *J. Proteome Res.* **9**, 1854–1863 (2010).
45. Bauer, J. A. *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.* **16**, 681–690 (2010).
46. Morita, Y. *et al.* Imaging mass spectrometry of gastric carcinoma in formalin-fixed paraffin-embedded tissue microarray. *Cancer Sci.* **101**, 267–273 (2010).
47. Djidja, M. C. *et al.* MALDI-ion mobility separation-mass spectrometry imaging of glucose-regulated protein 78 kDa (Grp78) in human formalin-fixed, paraffin-embedded pancreatic adenocarcinoma tissue sections. *J. Proteome Res.* **8**, 4876–4884 (2009).
48. Oppenheimer, S. R., Mi, D., Sanders, M. & Caprioli, R. M. A molecular analysis of tumor margins by MALDI mass spectrometry in renal carcinoma. *J. Proteome Res.* **9**, 2182–2190 (2010).
49. Lemaire, R. *et al.* Specific MALDI imaging and profiling for biomarker hunting and validation: fragment of the 11S proteasome activator complex, reg α fragment, is a new potential ovary cancer biomarker. *J. Proteome Res.* **6**, 4127–4134 (2007).
50. Kang, S. *et al.* Molecular proteomics imaging of tumor interfaces by mass spectrometry. *J. Proteome Res.* **9**, 1157–1164 (2010).
51. Cazares, L. H. *et al.* Imaging mass spectrometry of a specific fragment of mitogen-activated protein kinase/extracellular signal-regulated kinase kinase kinase 2 discriminates cancer from uninvolved prostate tissue. *Clin. Cancer Res.* **15**, 5541–5551 (2009).
52. Herring, K. D., Oppenheimer, S. R. & Caprioli, R. M. Direct tissue analysis by matrix-assisted laser desorption ionization mass spectrometry: application to kidney biology. *Semin. Nephrol.* **27**, 597–608 (2007).
53. Caldwell, R. L., Gonzalez, A., Oppenheimer, S. R., Schwartz, H. S. & Caprioli, R. M. Molecular assessment of the tumor protein microenvironment using imaging mass spectrometry. *Cancer Genom. Proteom.* **3**, 279–288 (2006).
54. Sijts, A. *et al.* The role of the proteasome activator *PA28* in MHC class I antigen processing. *Mol. Immunol.* **39**, 165–169 (2002).
55. Wilson, K. S., Roberts, H., Leek, R., Harris, A. L. & Geradts, J. Differential gene expression patterns in *HER2*/neu-positive and -negative breast cancer cell lines and tissues. *Am. J. Pathol.* **161**, 1171–1185 (2002).
56. Bronckart, Y. *et al.* Development and progression of malignancy in human colon tissues are correlated with expression of specific Ca^{2+}-binding S100 proteins. *Histol. Histopathol.* **16**, 707–712 (2001).
57. Lesniak, W., Slomnicki, L. P. & Filipek, A. S100A6 - new facts and features. *Biochem. Biophys. Res. Commun.* **390**, 1087–1092 (2009).
58. Puthalakath, H., Huang, D. C., O'Reilly, L. A., King, S. M. & Strasser, A. The proapoptotic activity of the *Bcl-2* family member Bim is regulated by interaction with the dynein motor complex. *Mol. Cell* **3**, 287–296 (1999).
59. Roesch-Ely, M. *et al.* Proteomic analysis reveals successive aberrations in protein expression from healthy mucosa to invasive head and neck cancer. *Oncogene* **26**, 54–64 (2007).
60. Reyzer, M. L. *et al.* Early changes in protein expression detected by mass spectrometry predict tumor response to molecular therapeutics. *Cancer Res.* **64**, 9093–9100 (2004).
61. Jackson, S. N. & Woods, A. S. Direct profiling of tissue lipids by MALDI-TOFMS. *J. Chromatogr B Analyt. Technol. Biomed. Life Sci.* **877**, 2822–2829 (2009).
62. Wenk, M. R. The emerging field of lipidomics. *Nature Rev. Drug Discov.* **4**, 594–610 (2005).
63. Murphy, R. C., Hankin, J. A. & Barkley, R. M. Imaging of lipid species by MALDI mass spectrometry. *J. Lipid Res.* **50**, S317–S322 (2009).
64. Meriaux, C., Franck, J., Wisztorski, M., Salzet, M. & Fournier, I. Liquid ionic matrixes for MALDI mass spectrometry imaging of lipids. *J. Proteomics* **73**, 1204–1218 (2010).
65. Wang, H. Y., Jackson, S. N. & Woods, A. S. Direct MALDI-MS analysis of cardiolipin from rat organs sections. *J. Am. Soc. Mass Spectrom.* **18**, 567–577 (2007).
66. Burnum, K. E. *et al.* Spatial and temporal alterations of phospholipids determined by mass spectrometry during mouse embryo implantation. *J. Lipid Res.* **50**, 2290–2298 (2009).

© 2010 Macmillan Publishers Limited. All rights reserved

67. Signor, L. *et al.* Analysis of erlotinib and its metabolites in rat tissue sections by MALDI quadrupole time-of-flight mass spectrometry. *J. Mass Spectrom.* **42**, 900–909 (2007).
68. Cornett, D. S., Frappier, S. L. & Caprioli, R. M. MALDI-FTICR imaging mass spectrometry of drugs and metabolites in tissue. *Anal. Chem.* **80**, 5648–5653 (2008).
69. Kitteringham, N. R., Jenkins, R. E., Lane, C. S., Elliott, V. L. & Park, B. K. Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **877**, 1229–1239 (2009).
70. Khatib-Shahidi, S., Andersson, M., Herman, J. L., Gillespie, T. A. & Caprioli, R. M. Direct molecular analysis of whole-body animal tissue sections by imaging MALDI mass spectrometry. *Anal. Chem.* **78**, 6448–6456 (2006).
71. Bouslimani, A., Bec, N., Glueckmann, M., Hirtz, C. & Larroque, C. Matrix-assisted laser desorption/ionization imaging mass spectrometry of oxaliplatin derivatives in heated intraoperative chemotherapy (HIPEC)-like treated rat kidney. *Rapid Commun. Mass Spectrom.* **24**, 415–421 (2010).
72. Trim, P. J. *et al.* Matrix-assisted laser desorption/ionization-ion mobility separation-mass spectrometry imaging of vinblastine in whole body tissue sections. *Anal. Chem.* **80**, 8628–8634 (2008).
73. Grey, A. C., Chaurand, P., Caprioli, R. M. & Schey, K. L. MALDI imaging mass spectrometry of integral membrane proteins from ocular lens and retinal tissue. *J. Proteome Res.* **8**, 3278–3283 (2009).
74. Norris, J. L., Porter, N. A. & Caprioli, R. M. Combination detergent/MALDI matrix: functional cleavable detergents for mass spectrometry. *Anal. Chem.* **77**, 5036–5040 (2005).
75. Groseclose, M.R. *High-throughput analysis of tissue microarrays of disease: combining in situ proteomics with MALDI imaging mass spectrometry.* Thesis, Vanderbilt Univ. (2009).

Acknowledgements
The authors would like to thank E. H. Seeley for brain images, M. R. Groseclose for lung images, and M. L. Reyzer, H. C. Manning and R. A. Smith for whole mouse images.

Competing interests statement
The authors declare no competing financial interests.

DATABASES
National Cancer Institute Drug Dictionary: http://www.cancer.gov/drugdictionary/
erlotinib | oxaliplatin | paclitaxel | trastuzumab | vinblastine
Pathway Interaction Database: http://pid.nci.nih.gov/ERBB2

FURTHER INFORMATION
Richard M. Caprioli's homepage: http://www.mc.vanderbilt.edu/root/vumc.php?site=msrc

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

the ERBB2 (also known as HER2)-specific monoclonal antibody trastuzumab (Herceptin; Genentech) in ERBB2-positive breast cancers[3] are such examples.

The potential benefits of this personalized approach to the treatment of disease are considerable. They include the identification of improved biological targets using validated biomarker studies, the capacity to increase the likely success of clinical trials by preselecting the patient population and the fact that this will in turn reduce the time, cost and the likelihood of failure of clinical trials[4]. Information from validated biomarker studies also allows the re-introduction of drugs that have failed in a clinical trial setting or that have been withdrawn from the market to be re-applied in a more targeted way. Similarly, biomarker studies might also offer the potential to avoid adverse side effects, and this would, in turn, lead to higher compliance with various treatment regimes.

The need for large integrated biobanks

One of the biggest limiting factors to the successful translation of basic scientific cellular and molecular studies into improved patient outcome has been the lack of access to large, appropriate and well-annotated cohorts of human tissue[5,6]. Focused disease-specific institutional biobanks have had some success in translational and personalized medicine (as described above). However, owing to the complex and heterogeneous nature of cancer, it is now clear that much larger biobanks are required[7,8]. For genetic main-effect studies 2,000–5,000 samples are needed, for lifestyle main-effect studies 2,000–20,000 samples are required and for gene–lifestyle interaction studies 20,000–50,000 samples are required[8]. Only when these larger resources are available can we truly understand the interactions between gene, environment, lifestyle and disease and translate this knowledge into the clinic through innovative diagnostics, therapeutics and preventive strategies for cancer. These larger resources can only be achieved by the integration of existing biobanks that already have a wealth of information and samples. However, there are many obstacles and challenges associated with integration, including technical, logistical, ethical and legal ones.

Groups across both North America and Europe have started to address these obstacles and challenges to move this process forwards. Initially, national programmes were established that linked previously collected biobank samples. These included the Canadian Tumour Repository Network (CTRNet; see the CTRNet website; Further

SCIENCE AND SOCIETY

Integrating biobanks: addressing the practical and ethical issues to deliver a valuable tool for cancer research

R. William G. Watson, Elaine W. Kay and David Smith

Abstract | Cancer is caused by complex interactions between genes, environment and lifestyles. Biobanks of well-annotated human tissues are an important resource for studying the underlying mechanisms of cancer. Although such biobanks exist, their integration to form larger biobanks is now required to provide the diversity of samples that are needed to study the complexity and heterogeneity of cancer. Clear guidelines and policies are also required to address the challenges of integrating individual institutional or national biobanks and build public trust. This Science and Society article highlights some of the main practical and ethical issues that are undergoing discussion in the integration of tissue biobanks for cancer.

It is widely accepted that although basic scientific studies carried out using cell lines and animal models can be informative about the cellular and molecular aspects of cancer there is a clear requirement to confirm this in human samples. The concept of patient-specific and disease-specific ('targeted') therapy has expanded rapidly in recent years. Many researchers believe that this concept of personalized medicine will provide the solution to the considerable challenges posed to the clinical treatment of cancer. The move from the traditional 'one size fits all' approach for the treatment of cancer to targeted approaches seems to offer genuine hope for improved patient outcomes. There are a few examples for which the concept of a highly effective drug treatment targeted towards a specific limited patient population has become reality. The use of imatinib (Gleevec; Norvartis) in chronic myeloid leukaemia[1], the use of monoclonal antibodies that target the epidermal growth factor receptor (EGFR) in patients with EGFR-expressing metastatic colon cancer[2] and the use of

© 2010 Macmillan Publishers Limited. All rights reserved

67. Signor, L. *et al.* Analysis of erlotinib and its metabolites in rat tissue sections by MALDI quadrupole time-of-flight mass spectrometry. *J. Mass Spectrom.* **42**, 900–909 (2007).
68. Cornett, D. S., Frappier, S. L. & Caprioli, R. M. MALDI-FTICR imaging mass spectrometry of drugs and metabolites in tissue. *Anal. Chem.* **80**, 5648–5653 (2008).
69. Kitteringham, N. R., Jenkins, R. E., Lane, C. S., Elliott, V. L. & Park, B. K. Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **877**, 1229–1239 (2009).
70. Khatib-Shahidi, S., Andersson, M., Herman, J. L., Gillespie, T. A. & Caprioli, R. M. Direct molecular analysis of whole-body animal tissue sections by imaging MALDI mass spectrometry. *Anal. Chem.* **78**, 6448–6456 (2006).
71. Bouslimani, A., Bec, N., Glueckmann, M., Hirtz, C. & Larroque, C. Matrix-assisted laser desorption/ionization imaging mass spectrometry of oxaliplatin derivatives in heated intraoperative chemotherapy (HIPEC)-like treated rat kidney. *Rapid Commun. Mass Spectrom.* **24**, 415–421 (2010).
72. Trim, P. J. *et al.* Matrix-assisted laser desorption/ionization-ion mobility separation-mass spectrometry imaging of vinblastine in whole body tissue sections. *Anal. Chem.* **80**, 8628–8634 (2008).
73. Grey, A. C., Chaurand, P., Caprioli, R. M. & Schey, K. L. MALDI imaging mass spectrometry of integral membrane proteins from ocular lens and retinal tissue. *J. Proteome Res.* **8**, 3278–3283 (2009).
74. Norris, J. L., Porter, N. A. & Caprioli, R. M. Combination detergent/MALDI matrix: functional cleavable detergents for mass spectrometry. *Anal. Chem.* **77**, 5036–5040 (2005).
75. Groseclose, M.R. *High-throughput analysis of tissue microarrays of disease: combining in situ proteomics with MALDI imaging mass spectrometry.* Thesis, Vanderbilt Univ. (2009).

Acknowledgements
The authors would like to thank E. H. Seeley for brain images, M. R. Groseclose for lung images, and M. L. Reyzer, H. C. Manning and R. A. Smith for whole mouse images.

Competing interests statement
The authors declare no competing financial interests.

DATABASES
National Cancer Institute Drug Dictionary: http://www.cancer.gov/drugdictionary/
erlotinib | oxaliplatin | paclitaxel | trastuzumab | vinblastine
Pathway Interaction Database: http://pid.nci.nih.gov/ERBB2

FURTHER INFORMATION
Richard M. Caprioli's homepage: http://www.mc.vanderbilt.edu/root/vumc.php?site=msrc

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

the ERBB2 (also known as HER2)-specific monoclonal antibody trastuzumab (Herceptin; Genentech) in ERBB2-positive breast cancers[3] are such examples.

The potential benefits of this personalized approach to the treatment of disease are considerable. They include the identification of improved biological targets using validated biomarker studies, the capacity to increase the likely success of clinical trials by preselecting the patient population and the fact that this will in turn reduce the time, cost and the likelihood of failure of clinical trials[4]. Information from validated biomarker studies also allows the re-introduction of drugs that have failed in a clinical trial setting or that have been withdrawn from the market to be re-applied in a more targeted way. Similarly, biomarker studies might also offer the potential to avoid adverse side effects, and this would, in turn, lead to higher compliance with various treatment regimes.

The need for large integrated biobanks

One of the biggest limiting factors to the successful translation of basic scientific cellular and molecular studies into improved patient outcome has been the lack of access to large, appropriate and well-annotated cohorts of human tissue[5,6]. Focused disease-specific institutional biobanks have had some success in translational and personalized medicine (as described above). However, owing to the complex and heterogeneous nature of cancer, it is now clear that much larger biobanks are required[7,8]. For genetic main-effect studies 2,000–5,000 samples are needed, for lifestyle main-effect studies 2,000–20,000 samples are required and for gene–lifestyle interaction studies 20,000–50,000 samples are required[8]. Only when these larger resources are available can we truly understand the interactions between gene, environment, lifestyle and disease and translate this knowledge into the clinic through innovative diagnostics, therapeutics and preventive strategies for cancer. These larger resources can only be achieved by the integration of existing biobanks that already have a wealth of information and samples. However, there are many obstacles and challenges associated with such integration, including technical, logistical, ethical and legal ones.

Groups across both North America and Europe have started to address these obstacles and challenges to move this process forwards. Initially, national programmes were established that linked previously collected biobank samples. These included the Canadian Tumour Repository Network (CTRNet; see the CTRNet website; Further

SCIENCE AND SOCIETY

Integrating biobanks: addressing the practical and ethical issues to deliver a valuable tool for cancer research

R. William G. Watson, Elaine W. Kay and David Smith

Abstract | Cancer is caused by complex interactions between genes, environment and lifestyles. Biobanks of well-annotated human tissues are an important resource for studying the underlying mechanisms of cancer. Although such biobanks exist, their integration to form larger biobanks is now required to provide the diversity of samples that are needed to study the complexity and heterogeneity of cancer. Clear guidelines and policies are also required to address the challenges of integrating individual institutional or national biobanks and build public trust. This Science and Society article highlights some of the main practical and ethical issues that are undergoing discussion in the integration of tissue biobanks for cancer.

It is widely accepted that although basic scientific studies carried out using cell lines and animal models can be informative about the cellular and molecular aspects of cancer there is a clear requirement to confirm this in human samples. The concept of patient-specific and disease-specific ('targeted') therapy has expanded rapidly in recent years. Many researchers believe that this concept of personalized medicine will provide the solution to the considerable challenges posed to the clinical treatment of cancer. The move from the traditional 'one size fits all' approach for the treatment of cancer to targeted approaches seems to offer genuine hope for improved patient outcomes. There are a few examples for which the concept of a highly effective drug treatment targeted towards a specific limited patient population has become reality. The use of imatinib (Gleevec; Norvartis) in chronic myeloid leukaemia[1], the use of monoclonal antibodies that target the epidermal growth factor receptor (EGFR) in patients with EGFR-expressing metastatic colon cancer[2] and the use of

© 2010 Macmillan Publishers Limited. All rights reserved

information), which was established to link cancer researchers with provincial tumour banks, thus creating new opportunities for translational cancer research to improve cancer outcomes in Canada and beyond. This network gave researchers access to tissue and clinical data. The Organisation of European Cancer Institutes (OECI) TuBaFrost database (see the TuBaFrost website; Further information) was established in 2003 to network European frozen tissue pathology banks for cancer research. In this initiative, the tissue collection process was standardized, a code of conduct for the exchange of residual human material for research was developed based on European legislation and a web-based sample request process was developed. Other examples include the Spanish tumour bank, EuroBoNet (see the EuroBoNet website; Further information) and the Office of Biorepositories and Biospecimen Research (OBBR; see the OBBR website; Further information) in the USA. These initial biobanks led to further expansions and networking to create resources across Europe and the United States. The establishment of the European Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) programme (see the BBMRI website; Further information) illustrates moves to coordinate existing biobanking activities across Europe[9,10] (BOX 1). The development of the Cancer Bioinformatics Grid (caBIG®) infrastructure (see the Cancer Bioinformatics Grid website; Further information) is connecting research organizations across the United States[11] (BOX 2). The National Comprehensive Cancer Network (NCCN; see the NCCN website; Further information) has collected patient data outcomes for breast cancer, non-small-cell lung cancer, colorectal cancer, non-Hodgkin's lymphoma and ovarian cancer. The data, along with records of patient treatments and patient outcomes, have allowed retrospective comparative studies to be conducted. The analysis of subpopulations in these databases has already led to changes in clinical practice[12,13].

There are many reports outlining the complexity of biobanking that provide strong recommendations and the identification of best practice for all aspects of the process[14] (see the National Cancer Institute Best Practices for Biospecimen Resources website; Further information). FIGURE 1 outlines the steps involved in biobanks. As the requirements for complex multi-institutional and international collections to study cancer processes have been established, this article focuses on some of the important practical and ethical issues related to the integration of biobanks.

Box 1 | The Biobanking and Biomolecular Resources Research Infrastructure

The European project Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) was established in 2008 to network European biobanks with the aim of improving resources for biomedical research and therefore contributing to the improved prevention, diagnosis and treatment of disease. The resource includes 261 biobanks across 23 countries with a total of more than 16 million samples. It is only possible to achieve this using a federated network of centres in European countries, which is best described as a distributed hub structure of existing biobanks. This will provide the flexibility to facilitate expansion and multiple uses (see the BBMRI website; Further information).

The mission of the BBMRI is:

- To benefit European health care, medical research and, ultimately, the health of the citizens of the European Union
- To have a sustainable legal and financial conceptual framework for a pan-European Biobank infrastructure
- To increase scientific excellence and efficacy of European research in the life sciences, especially in biomedical research
- To expand and secure the competitiveness of European research and industry in a global context, especially in the fields of medicine and biology

Practical implementations to integration

Definition and use of biobanks. There are many types (BOX 3) and definitions of biobanks, which are informed by the goals, objectives or statements of purpose, and so differ from institution to institution. Defined in the most general terms, a biobank is a collection of biological specimens and corresponding participant data. The goals of the biobank will therefore determine the types of material collected, the scope of the research and the consent used to collect the samples. A clear definition of the scope of a biobank is essential to define informed consent for its participants, determine standard operating procedures for sample collection and storage, define the data management system and determine the use of the samples. All of these characteristics are not only required for a high quality resource but also for building public trust, avoiding the misuse of the resource and setting clear expectations for the contributors to and users of the biobank. However, providing a clear definition of the scope of a biobank is a problem when integrating biobanks that have been collected across different institutions or countries for different reasons. There is a clear need for such flexibility, owing to the ever-expanding development of technology and understanding of the complexity of cancer, but this flexibility has effects on patient consent, standard operating procedures, information technology management systems, and the use and distribution of samples.

Standard operating procedures. The careful and well-documented processing and annotation of samples is essential to provide a useful resource for scientific interrogation. Valid data are defined as those that are reproducible, but variation is an enemy to such reproducibility. The importance of standard operating procedures has been well established to prevent variation in the patient consent, collection, processing, storage and distribution of samples. Several important documents have been produced that clearly define the best practices for biobanks[14,15] (see the National Cancer Institute Best Practices for Biospecimen Resources website; Further information). However, balancing the standardization and flexibility of biobanks is of considerable importance. The main challenge related to biobank integration is the consistency of standard operating procedures across multiple sites, so that differences observed across patient populations are not affected by where and how the samples were processed. There are ongoing international collaborations addressing these issues; for example, the Data Schema and Harmonisation Platform for Epidemiological Research (DataSHaPER; see the DataSHaPER website; Further information) has been developed through the Public Population Project in Genomics (P3G) and Promoting Harmonisation of Epidemiological Biobanks in Europe (PHOEBE) and involves 25 international biobanks. These international collaborations have defined several core high priorities for improved study integration, such as sample collection, preliminary processing and storage, as well as the importance of documenting key aspects of the standard operating procedure, such as the time of collection and processing so that this variable is recorded and linked with the sample. These networks have all identified the need for quality assurance systems that operate

© 2010 Macmillan Publishers Limited. All rights reserved

across the sites and that can address compliance with, and the implementation of, standard operating procedures, but their implementation has not yet been agreed.

Information technology management systems. A central component of any biobank is the bioinformatics and data management system. It is clear that to harness the full potential of any biological sample, information needs to be collected not only about the standard operating procedures and pathological status, but also the demographic, diagnostic, medical and family history and clinical outcome data of the participant. For such data, the patient must be identifiable to a member of the biobank so that the information can be updated over time. Issues associated with such information extend beyond the storage of data and linking it to a specific participant and also include data reporting, data searching and mining, data accessibility and network security, and require personnel to enter, manage and maintain the data in the future. There are a large number of commercially and institutionally available data management systems but there is no common consensus on which system delivers an overall solution. Biobank information management systems (BIMS) are emerging as a proposed solution that allow the storing, tracking and most importantly the searching of several sources of data at the same time. The Karolinska Institute has developed such a system, and it has been proposed that this BIMS model could be used to manage the data from the planned LifeGene project, which will include 500,000 Swedish participants and will follow them for decades using both questionnaires and testing of individuals. Another such project is the UK Biobank, which will gather samples from 500,000 participants with extensive medical and family histories and follow them for 30 years. This project is expected to collect 15 million aliquots of blood and DNA. However, it is clear that a strong and close interaction between the information technology development team, data managers, biobank managers, clinicians and researchers is required to build a system that is fit for purpose and that also contains the appropriate security features to protect participant confidentiality and so foster public trust. Individual biobanks need to agree to use standard identification schemata, data formats, data quality assurance and control processes, database architectures and common security processes when building their systems.

The integration of databases is one of the most challenging issues for establishing biobank networks. Ideally, data should be exchangeable across multiple national and international sites using a web-based system, and facilitated by agreed consent, ontologies, naming conventions, data formats and definitions that are underpinned by common standard operating procedures for data collection, uploading, management and audit. Unfortunately, there is no such agreement; however, the World Wide Biobank Summit has indicated that biobanks should agree on minimum data sets that are interchangeable between biobanks and identify the complete ontology and multi-lingual definitions of the data set. In addition, access policies and levels of security to protect the identity and confidentiality of patients must be agreed. This can be achieved using network security and access control that limit different users to different aspects of the data. For example, data managers would be able to see patient-identifiable information that is applicable to their institution only, so that they can update the appropriate information, and researchers or external collaborators would see the de-identified codes of participants as well as the relevant clinical and scientific data that are applicable only to their level of access and requirements. This approach would help to address issues of confidentiality and reassure participants of privacy, which is central to building public trust. However, it is clear that an important requirement for these information technology systems is that they are built in a flexible way to accommodate the ever-evolving requirements of participants and users of biobanks.

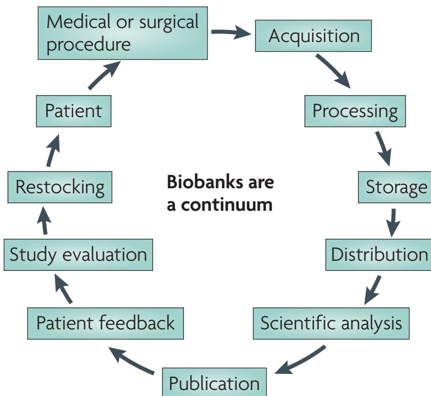


Figure 1 | **The steps involved in establishing a biobank.** The steps involved in biobanks are a continuum, starting with participant consent, leading to the collection, processing, storage and distribution of samples. The process finishes with the publishing of data and informing participants of how their samples have been used, which then supports the continuation of recruitment to biobanks.

Ethical issues associated with integration

Patient consent. Informed consent is a fundamental and key mechanism put in place to protect the interests, welfare and rights of research participants and it is generally accepted as an absolute imperative for the collection of human samples[16,17]. Central to the consent process is information that educates the participant about the research and how samples will be collected and used, along with details of potential benefits and risks. This information allows participants to make an informed decision regarding their involvement in a study. A clearly defined biobank makes it easier to inform participants about the process and the use of their samples; however, with increasing pressure for flexibility in the use and sharing of samples, the consent process has become more complicated. In addition, biobanks are typically established for long-term use, therefore the details of future research projects cannot be known. Obtaining new informed consent for each new project is in most cases difficult, if not practically impossible. However, it has

Box 2 | **The Cancer Bioinformatics Grid (caBIG®)**

The Cancer Bioinformatics Grid (caBIG®) was launched in 2004 by the US National Cancer Institute and is already enabling many research organizations to move their clinical and basic research programmes forwards. The caBIG® programme aims to create a virtual web of interconnected data, individuals and organizations, with the ultimate goal of improving the interaction of those involved in treatment-focused research, leading to improved patient outcome. It is essentially an infrastructure for connecting research organizations that allows for a systems approach to biomedical research. It relies on the widespread use of sophisticated and powerful information technology, which offers improved data management and data exchange and sharing. This will facilitate a move from the traditional self-contained or isolated model of research organizations to the concept of large networks that allow faster, larger and standardized collaborative research. The CaTissue suite developed by caBIG® is a biobanking managing system designed to collect, manage, process, annotate, request and distribute biospecimens and associated information (see the Cancer Bioinformatics Grid website; Further information).

© 2010 Macmillan Publishers Limited. All rights reserved

been argued that general consent to future research projects is not sufficient. To what extent can such consent about the unknown future be accurately described as informed consent?[18–20] One proposed solution to this problem is to accept general consent for future research if the participants have the option to withdraw their data and if future research projects are approved by an Institution Review Board (IRB) or Research Ethics Committee (REC). It can be argued that adopting a somewhat lower standard than that of classic informed consent could be justified because the risks to individuals resulting from research on their human biological material do not involve physical harm, as is the case with clinical trials[21] (BOX 4). In addition, it has been proposed that owing to the benefits that are associated with the knowledge generated through these potential future studies, and the fact that this information could accrue to all individuals as well as future generations, it may be possible to justify an expansive use of participants' samples for future studies[22].

In recent research conducted in the United States on stored biological materials, it was found that biobanks used consent forms that offered individuals the option of designating how their biological materials could be used with varying degrees, including whether the materials could be stored; whether participants could be contacted again in the future; whether the materials could be used for genetic testing; and whether the materials could be shared with researchers who were not part of the original research team. In addition, a few biobanks offered options for participants to designate how their biological materials could be used in future research — for example, by specifying whether identifiers could be retained, or specifying the research topics for which the materials could be used[23].

In support of Elger and Caplan's[17] suggestion that general consent should be the international standard, the European Council has taken the view that broad consent to future research use is acceptable[24]. This implies that individual donors renounce any ongoing rights to exercise control over the uses of their donated materials and the source itself[25]. This point has been emphasized by Shickle[26] who states that "providing that there is proper disclosure and so on, then the choice for the individual is to participate on the terms offered or not. There is a 'negative right' not to be included in the research without consent. There is no 'positive right' for a biobank to be run in such a way just because an individual would like it to be so."

Box 3 | **Types of cancer biobanks**

A population-based biobank is defined as a large repository of donated human DNA and/or its information, collected from volunteers with and without cancer, which is used to identify the genes that contribute to human disease. This is an essential resource if we are to understand the genetic risk factors that are associated with cancer development and the genetic profile of the patient that is associated with the development of cancer. These resources also need to include high-quality lifestyle data.

Disease-based biobanks are defined as a collection of biological material from patients with cancer, and are essential to understand the molecular and cellular development of cancer at a specific stage.

These types of biobanks can be collected for clinical purposes; for example, they could be established for medical purposes, such as a blood bank. Specimens can also be collected as a by-product of diagnostic or curative procedures; for example, from hospital pathology departments. Biobanks can also be collected for specific research investigations; for example, clinical trials.

Investigator access. The scientific potential of biobank samples and data can often be fully exploited only if their use is not confined to individual research projects specified in advance. Donors should be able to give generalized consent to the use of their samples and data for the purposes of medical, including genetic, research. For this reason, donors should also be able to consent to the transfer of samples and data from biobanks to third parties for the purposes of medical research. Most biobanks provide samples and data to various researchers who are often not directly affiliated with the biobank holders. The transfer of samples and sharing of data raises several ethical questions. If samples and data are physically transferred outside the biobank, the recipient might use the material for purposes other than those to which the sample donors have consented or the samples might be transferred to third parties, thus putting the privacy of donors at risk. If recipients use samples for their own purposes, the investments and efforts of the original biobank holder could be threatened. Biobank material is a valuable resource and the ethical issues of allocating such material emerge if the samples and data are misused. However, if the transfer of samples is too restrictive because of transfer agreements that include complex and costly control mechanisms, the resulting inaccessibility of biobank material does not allow a large number of researchers to use this resource in the interests of science and for the good of the public[21].

However, except in circumstances prescribed by law, it has been strongly argued that the transfer of samples or information must take place in a simple coded or double-coded way, with the recipient of double-coded samples having no access to the code. Should the recipients' research require an association with personalized data, this may be provided only by an official of the biobank to which the donors originally entrusted their samples and data. The rationale for this is that an alternative anonymization strategy, which permanently destroys the information that links a sample and donor identity, risks the loss of information of scientific value, as there is no way to prevent multiple inclusions of the same participant, and this approach does not allow retroactive validation and demonstration of reproducibility. To facilitate clear autonomy, transfers of samples and data to third parties must be fully documented for future reference according to standard operating procedures[27].

Subject to the consent of donors, the transfer of samples should be permissible provided that the recipient is subject to standards of donor protection and quality assurance equivalent to those that are applicable to the original institution in charge of the biobank. Transfers of existing biobanks to third parties with the inclusion of personalized donor data should be possible only with the approval of a REC or IRB[28].

Patient access. An individual's right to access their genetic data is linked to control over their own identity. This varies across biobanks, as do the ethical justifications for the differences. In Estonia, personal genetic data is available to participating individuals on request or the participants can decide that the information should be withheld from them. Likewise, in Iceland, patients are entitled to information about their health (for example, condition, prognosis, risks and benefits of treatment) on request. The UK Biobank takes a different and less open approach. Although the Biobank Protocol acknowledges that individuals will have the legal right to access their personal data if required, it also makes it clear that they

© 2010 Macmillan Publishers Limited. All rights reserved

will not routinely receive any individual information that relates to their blood samples (including biochemistry and genetic findings). The assumption that participants should only be provided with health information about themselves in a clinical situation in which a trained professional can provide appropriate interpretation and guidance lies behind this prohibition. The justification for this position is that it would not be constructive and might even be harmful to provide health information but no interpretation, counselling and support. The UK Biobank cannot provide such counselling and support because it is purely a research initiative[29]. The German Ethics Council recommends that if individual communication of research results to the donor is agreed, then they must also be told as part of the information to be given that they must divulge these details in certain circumstances — for example, when concluding new employment or insurance contracts in the future. In addition, when such individual communication to the donor has been agreed, the findings must be imparted by a person with the appropriate counselling skills, especially when communicating the results of genetic diagnosis[28]. However, most biobanks take a different position regarding the right of donors' access to data. Participants are kept informed about research that is currently in progress or completed. These findings are usually made available through regular information sheets, newsletters or a dedicated website. It is clear that continuing public engagement and the feedback of findings is necessary for an acceptable public profile of biobanks[30]. It is recommended that such an approach should be taken following the integration of biobanks, as the ability to identify individual samples shared across resources would not be possible owing to anonyomization. However, if participants are informed at the time of donating samples of their ability to access research data, then new mechanisms need to be identified to relay information from future studies to participants from the collaborating users of the biobank.

Withdrawing samples or data from biobanks. The right of participants to withdraw their personal data reflects the basic principles regulating medical research according to the Nuremberg Code and the Declaration of Helsinki. However, exercising this right conflicts with the interest of science and industry to maintain the statistical integrity of population-based databases, particularly given the aims of using these databases for longitudinal studies. There are two important problems to withdrawing: namely, what can be withdrawn and when it can be withdrawn[31]. Although all studies discuss the right to withdraw from research generally, almost one-third failed to discuss this right specifically with respect to research involving biological materials. Therefore, specific discussion of the right to withdraw biological materials may be necessary to preserve this basic right regarding participation in research[23].

The UK Biobank offers a set of graded options for withdrawal (complete withdrawal, discontinued participation and no further contact requested), which attempts to balance the interests of the participant for data removal with the interests of the scientific community. In Estonia, there seems little room for negotiation, as donors have the right to have their data deleted from the database on request, and any violation of this right, including coercion to participate, is punishable as a criminal offence[29]. The German Ethics Council argues that donors must have the right to withdraw their consent to the use of their samples and data at any time. It should not be possible to waive this right. However, there should be a provision for donors to allow samples and data to continue to be used in the case of withdrawal if they are anonymized — that is, if the link to the participant's identity has been eliminated[28]. This is particularly important if the biobank has been integrated into a larger network and samples are shared across that network.

Participants need to be made aware of the fact that it will not be possible for samples and data that are included in completed research to be extricated from the research results and destroyed. The right to withdraw from a biobank does not include the right to withdraw research results that have already been accumulated. Rather, it means a prohibition of obtaining new data and analysis from the samples. Even if biobanks decide to retain existing data in an impersonalized form, questions arise as to whether complete and irreversible anonymization can be considered as a solution to the problem. When can withdrawal take place? One obvious point in time when withdrawal will be impossible is when samples and data have been analysed and are part of a publicised research result. Also, at this point, the participant's data will have been merged with the other data of that particular study and integrated into the result to create new data sets. However, participants could potentially withdraw at any point before this occurs[31].

The comparative analysis of practices in Europe and the United States has shown that, although the exact details may be different, all biobanks examined deem the provision of a withdrawal option to be important. The fact that informed consent is requested and that information is provided about

Box 4 | **Donor information required for consent**

In light of ethical controversies, informing participants about the details of storage and research involving their samples is crucial. It is widely agreed that sample donors must be informed about:

- The voluntary nature of participation
- The type of consent used (informed consent for the primary clinical or research purpose, or a general consent for future research)
- The circumstances of sampling, including the risks and benefits of the procedure
- The aims of storage (clinical purpose or research purpose) and the nature, extent and duration of the proposed use, including the possibility of genetic analysis
- The extent of and conditions for the possible future transfer of samples and data
- The measures taken to protect confidentiality (who will have access to the samples and information and the risks for individual donors and groups)
- The form of data storage and combination
- The anonymization or pseudonymization of samples and data and other ancillary donor protection measures and any provision for state access to samples and data
- The right to withdraw consent
- The limits of withdrawal and the fate of samples and data in this instance and if the biobank closes down
- Whether individual or aggregated research results will be disclosed to sample donors
- The possible consequences of the communication of results of genetic analyses for the donor and their relatives, including possible obligations to divulge (for example, to insurance institutions)
- The use of biobank samples for commercial purposes (patients planned, benefit sharing planned, and if so, in which form)
- Issues of payment of expenses, remuneration or benefit sharing

© 2010 Macmillan Publishers Limited. All rights reserved

the mechanisms of withdrawal also plays a decisive part in ensuring that participant autonomy is maintained[23]. However, the withdrawal of samples that have been shared with network partners has not yet been resolved and so is a challenge to the integration of biobanks.

The roles of IRBs and RECs. As can be observed, the role of IRBs and RECs seems to be crucial for the management of biobanks. Therefore, before conducting a research project involving the use of biobank samples and data, the consent of an IRB or REC should be necessary for studies in which bodily substances are to be collected from a donor's body for research purposes; the project calls for identified samples (that are linked to the individual in a way that makes them immediately identifiable) to be transferred to external researchers, and entire existing biobanks are to be transferred to third parties with identified samples included.

The involvement of an IRB or REC, and the need for its favourable opinion, is intended to ensure that a narrowly worded consent is not exceeded, that a consent in broad terms is not inappropriately given an even wider interpretation and that exceptional situations in which consent may be waived are not illegitimately invoked[15]. To ensure the integration of biobanks, IRBs and RECs need to be proactive in cooperating with one another. To some extent, this process is already underway. The European Forum for Good Clinical Practice (EFGCP) (see the European Forum for Good Clinical Practice website; Further information) is an example of the development of common policies and standardized procedures for European RECs. Such initiatives need to be encouraged and supported for biobanking integration so that IRBs and RECs can work from a common set of standards, procedures and documentation that will streamline ethical applications without compromising the underlying ethics. IRBs and RECs are becoming the guardians of biobanks as they move towards more integration.

Perspectives

Biobanks have the potential to be resources that will facilitate an understanding of the interactions between genes, environment and lifestyle in cancer. It is now clear that the integration of existing biobanks into large resources is required to translate this knowledge into clinical use through the development of innovative preventive strategies, diagnostics and therapeutics. There are many obstacles and challenges to such integration, but the BBMRI and caBIG® are addressing these issues to implement a federated network of biobanks. Many of the technical and logistical issues have been addressed; however, IRBs, RECs and scientific management groups that are responsible for these biobanks are still grappling with the social and ethical issues. These issues include consent, withdrawal of samples and data, patient access to information, and research access to information and samples. The roles of IRBs and RECs are becoming central to the management of biobanks, and their integration and the establishment of harmonized international policies are still required to facilitate the smooth sharing and integration of biobanks. If such issues are not dealt with, the public will withdraw its trust and individuals will start to revoke their consent, which would fundamentally affect the quality and completeness of such resources, preventing the ability to draw scientifically valid conclusions from them.

R. William G. Watson is at the UCD School of Medicine and Medical Science, University College Dublin, Ireland.

Elaine W. Kay is at the Department of Pathology, Beaumont Hospital, Ireland, and the Royal College of Surgeons in Ireland.

David Smith is at the Department of General Practice, Royal College of Surgeons in Ireland.

Correspondence to R.W.G.W
e-mail: william.watson@ucd.ie

doi:10.1038/nrc2913
Published online12 August 2010

1. Cohen, M. H. *et al.* Approval summary for imatinib mesylate capsules in the treatment of chronic myelogenous leukaemia. *Clin. Cancer Res.* **8**, 935–942 (2002).
2. Siena, S., Sartore-Bianchi. A., di Nicolantonio, F., Balfour, J. & Bardelli, A. Biomarkers predicting clinical outcome of epidermal growth factor receptor-targeted therapy in metastatic colorectal cancer. *J. Natl Cancer Inst.* **101**, 1308–1324 (2009).
3. Lemieux, J., Provencher, L., Latreilly, J., Rayson, D. & Wang, B. The role of neoadjuvant Her2-targeted therapies in HER2-overexpressing breast cancers. *Curr. Oncol.* **16**, 48–57 (2009).
4. Tan, D. S. W. *et al.* Biomarker-driven early clinical trials in oncology. *Cancer J.* **15**, 406–420 (2009).
5. Bonavida, B. Preface. Antibody therapies for cancer. *Oncogene* **26**, 3592–3593 (2007).
6. Taube, S. E. *et al.* A perspective on challenges and issues in biomarker development and drug and biomarker co development. *J. Natl Cancer Inst.* **101**, 1453–1463 (2009).
7. Caporaso, N. E. Why have we failed to find the low penetrance genetic constituents of common cancers? *Cancer Epidemiol. Biomarkers Prev.* **11**, 1544–1549 (2002).
8. Burton, P. R. *et al.* Size matters: just how big is BIG? Quantifying realistic sample size requirements for human geneome epidemiology. *Int. J. Epidemiol.* **38**, 263–273 (2009).
9. Bevilacqua, G. *et al.* The role of the pathologist in tissue banking: European consensus expert group report. *Virchows Arch.* **456**, 449–542010.
10. Riegman, P. H. J., de Jong, B. W. D. & Llombart-Bosch, A. The organization of European cancer institute pathobiology working group and its support for European biobanking infrastructure for translational cancer research. *Cancer Epidemiol. Biomarkers Prev.* **19**, 923–926 (2010).
11. Buetow, K. H. An infrastructure for interconnecting research institutions. *Drug Disc. Today* **14**, 605–610 (2009).
12. Goosner, M. National Comprehensive Cancer Networks (NCCN) — big data basis outcomes research begins to yield results. *J. Natl. Cancer Inst.* **102**, 215–217 (2010).
13. Cambon-Thomsen, A. Accessing the impact of biobanks. *Nature Genet.* **34**, 25–26 (2003).
14. International Society of Biological and Environmental Repositories. Best practices for repositories: collection, storage, retrieval and distribution of biological materials for research. *Cell Preserv. Technol.* **6**, 3–56 (2008).
15. The German Ethics Council Biobanks for Research, Opinion. Deutscher Ethikrat [online] (2004).
16. Sheikh, A. A. Genetic research and human biological samples: some legal and ethical considerations. *Med. Law* **23**, 897–912 (2004).
17. Elger, B. S. & Caplan, A. L. Consent and anonymization in research involving biobanks. *EMBO J.* **7**, 661–666 (2006).
18. Annas, G. Rules for research on human genetic variation – lessons from Iceland. *N. Engl. J. Med.* **342**, 1830–1833 (2000).
19. Chadwick, R. & Berg, K. Solidarity and equity: new ethical frameworks for genetic databases. *Nature Rev. Genet.* **2**, 318–321 (2001).
20. O'Neill, O. Informed consent and genetic information. *Stud. Hist. Philos. Biol. Biomed. Sci.* **32**, 689–704 (2001).
21. Ravitsky, V., Fiester, A. & Caplan, A. (eds) *The Penn Center Guide to Bioethics* (Springer Publishing, New York, USA, 2009).
22. Hakimian, R. & Korn, D. Ownership and use of tissue specimens for reseach. *JAMA* **292**, 2500–2505 (2004).
23. Wolf, L. E., Bouley, T. A. & McCulloch, C. E. Genetic research and stored biological materials: ethics and practice. *IRB* **32**, 7–18 (2010).
24. Council of Europe Draft explanatory memorandum to the draft recommendation on research on biological materials of human origin, Strasbourg. Council of Europe Steeting committee on Bioethics [online] (2006).
25. Campbell, A. V. The ethical challenges for genetic databases: safeguarding altruism and trust. *Kings Law J.* **18**, 373–382 (2007).
26. Shickle, D. The consent problem with DNA biobanks. *Stud. Hist. Philos. Biol. Biomed. Sci.* **37**, 503–519 (2006).
27. Helgesson, G., Dillner, J., Carlson, J., Bartram, C. R. & Hansson, M. G. Ethical framework for previously collected biobank samples. *Nature Biotechnol.* **25**, 973–976 (2007).
28. Slater, B. & Jones, M. Biobanks and bioethics: politics of legitimation. *J. Eur. Pub. Pol.* **12**, 710–732 (2005).
29. Putnina, A. Exploring the articulation of agency: Population genome project in Latvia. International Summer Academy on Technology Studies [online] (2002).
30. Gertz, R. Withdrawing from participating in a biobank — a comparative study. *Eur. J. Health Law* **15**, 381–389 (2008).
31. Hansson, M. G. *et al.* Should donors be allowed to give broad consent to future biobank research? *Lancet Oncol.* **3**, 266–269 (2006).

Competing interests statement
The authors declare no competing financial interests.

DATABASES
National Cancer Institute Drug Dictionary:
http://www.cancer.gov/drugdictionary
imatinib | trastuzumab

FURTHER INFORMATION
Canadian Tumour Repository Network (CTRNet): http://www.ctrnet.ca
Cancer Bioinformatics Grid: http://cabig.cancer.gov
DataSHaPER: http://www.datashaper.org
EuroBoNet: http://www.eurobonet.eu
European Biobanking and Biomolecular Resources Research Infrastructure (BBMRI): http://www.bbmri.eu
European Forum for Good Clinical Practice: http://www.efgcp.be
National Cancer Institute Best Practices for Biospecimen Resources: http://biospecimens.cancer.gov/bestpractices
National Comprehensive Cancer Network (NCCN): http://www.nccn.org
Office of Biorepositories and Biospecimen Research (OBBR): http://biospecimens.cancer.gov
TuBaFrost: http://www.tubafrost.org

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2010 Macmillan Publishers Limited. All rights reserved

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.

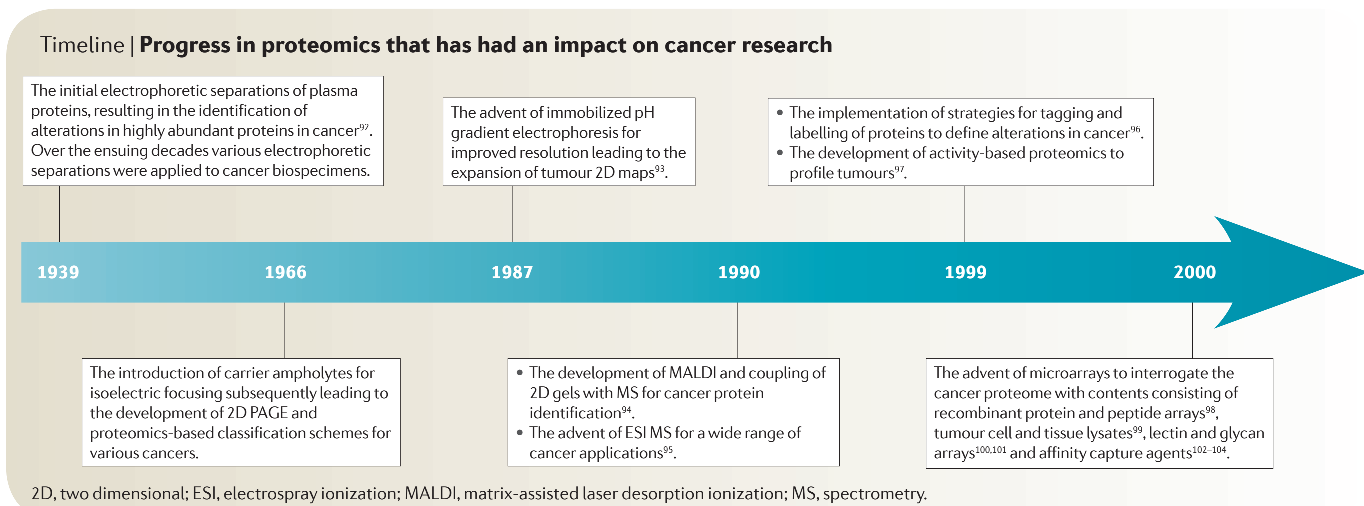
Illustrative of protein profiling to detect alterations in cancer half a century ago is a publication in *Nature* that defined patterns of globulin expression in serum among subjects with myeloma and macroglobulinemia[1] (TIMELINE). The study relied on what was then a new technique of starch gel electrophoresis, 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.

© 2010 Macmillan Publishers Limited. All rights reserved

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.

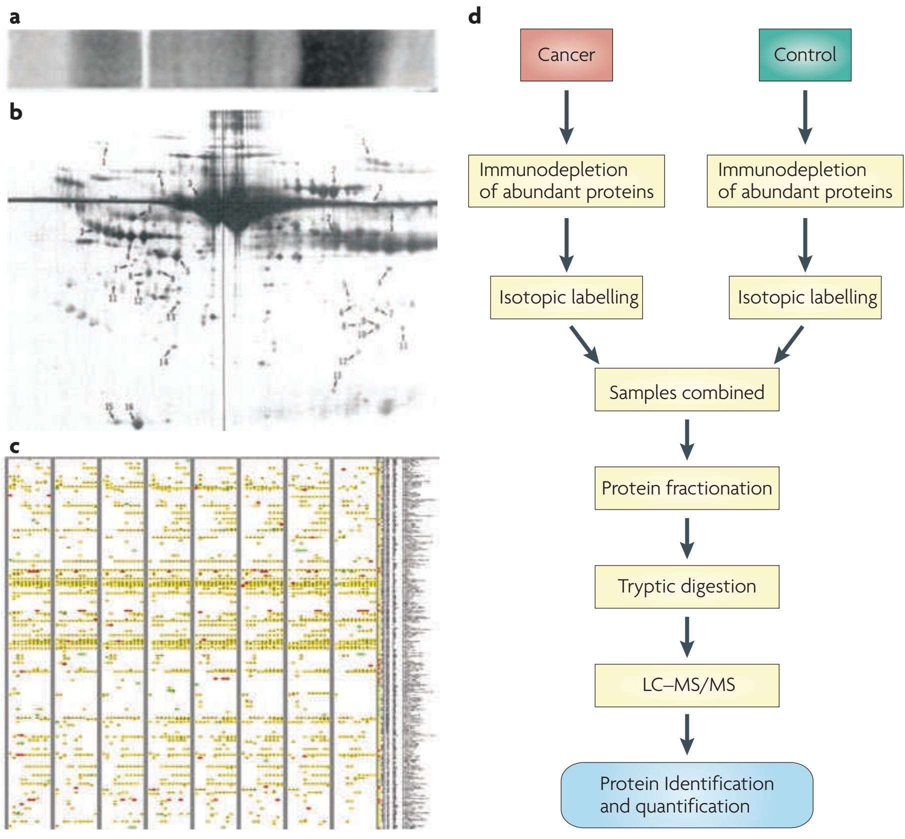


Figure 1 | **Increased resolving power of proteomics technologies as applied to serum and plasma. a** | Detection of a few major bands using one-dimensional (1D) electrophoresis as shown in a study of myeloma and macroglobulinemia[1]. **b** | Detection of a few hundred protein spots by 2D PAGE and the 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

© 2010 Macmillan Publishers Limited. All rights reserved

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.

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 phosphorylated 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

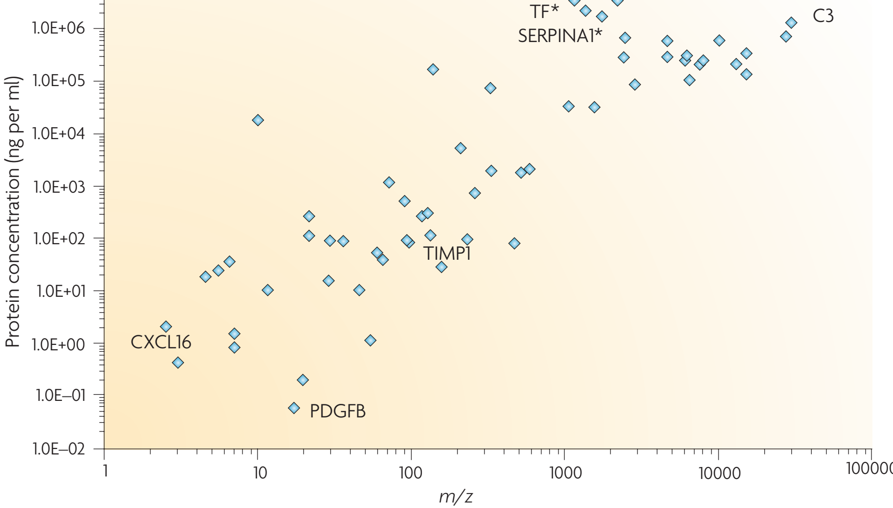


Figure 2 | **Depth of analysis of the plasma proteome.** Identification of 1,442 proteins in plasma that span more than 6 logs of protein abundance as determined by mass spectral counts and independent assays of a subset of proteins identified in a study of plasma from a pancreatic cancer mouse model[91].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

© 2010 Macmillan Publishers Limited. All rights reserved

(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 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.

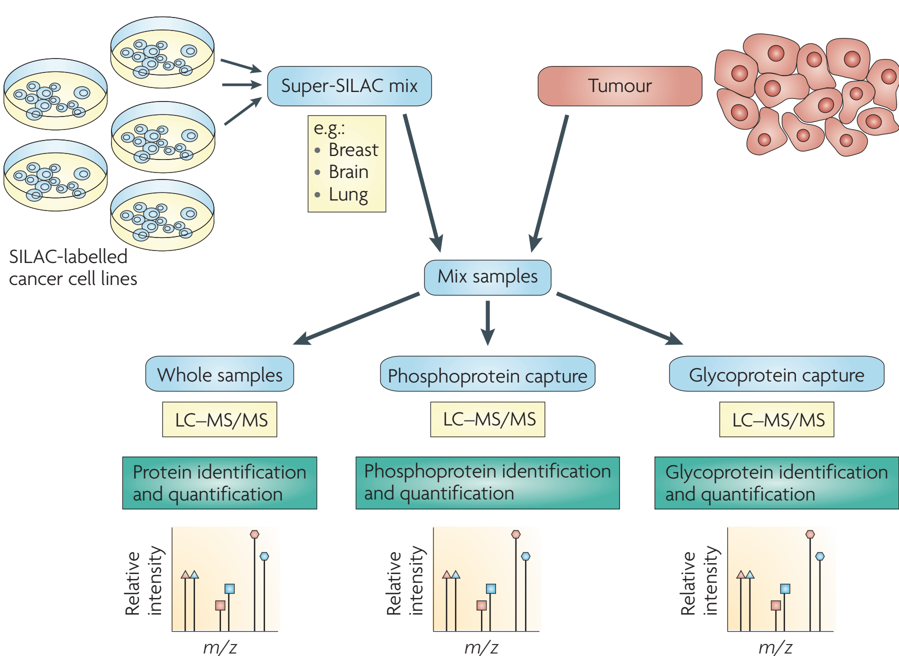


Figure 3 | **Quantitative profiling of proteins in tumour tissue using a modified SILAC strategy[32].** Stable isotope labelling by amino acids in cell culture (SILAC)-labelled proteins are extracted from cancer cell lines, mixed with tumour tissue lysates that are subjected to mass spectrometry (MS)-based 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.

© 2010 Macmillan Publishers Limited. All rights reserved

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.

© 2010 Macmillan Publishers Limited. All rights reserved

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

© 2010 Macmillan Publishers Limited. All rights reserved

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.

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

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.

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).

© 2010 Macmillan Publishers Limited. All rights reserved

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).

© 2010 Macmillan Publishers Limited. All rights reserved

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

© 2010 Macmillan Publishers Limited. All rights reserved

Nature Reviews Cancer | AOP, published online 12 August 2010; doi:10.1038/nrc2919

LEUKAEMIA

Comfortably MSI2–NUMB

Little is known about the mechanisms governing the progression of chronic myelogenous leukaemia (CML) from the slow-growing chronic phase to the aggressive blast crisis. Blast crisis is the stage at which CML is most refractory to treatment, so a greater understanding of its molecular hallmarks is crucial for the development of more effective therapies.

A team led by Tannishtha Reya focused on a signature trait of blast crisis — a block in cellular differentiation leading to an accumulation of immature cells. Reya and colleagues reasoned that, in these cells, the transition to blast crisis might rely on the attenuation of normal differentiation cues. To test this theory, her group modelled the chronic and blast crisis phases of CML by transplanting irradiated recipient animals with haematopoietic stem cells that were transduced with *BCR–ABL* alone (chronic phase) or *BCR–ABL* together with *NUP98–HOXA9* (blast crisis). Armed with these tools, they showed that the RNA binding protein Musashi 2 (MSI2) was dramatically upregulated during blast crisis compared with chronic phase disease. This was a compelling result given the known role of MSI2 as a regulator of asymmetric division in stem and progenitor cells, but was it actually required for progression to blast crisis? Reya and colleagues turned to mice harbouring a targeted disruption in *Msi2* and noted impaired leukaemic growth *in vivo*. Next, short hairpin RNA against *Msi2* was introduced into established blast crisis cells, resulting in decreased disease propagation in recipient mice. These results suggest a fundamental requirement for MSI2 in the establishment and maintenance of blast crisis.

During development, MSI2 represses NUMB, a key regulator of cell commitment and differentiation, so the authors asked whether this was also the case in CML. Accordingly, NUMB levels were significantly downregulated at blast crisis compared with chronic phase, and increased when *Msi2* expression was inhibited. Crucially, forced expression of NUMB in *BCR–ABL*; *NUP98–HOXA9*-transduced cells attenuated disease progression *in vivo*, and leukaemias that did develop were of a more differentiated phenotype. So, perturbations of the MSI2–NUMB axis in CML disrupts the equilibrium between the differentiated and undifferentiated states, and this propels the disease towards blast crisis.

What about human CML? Reya's team analysed 120 samples of CML and found that high levels of MSI2, together with reduced NUMB expression, correlated with advanced, aggressive disease. Further analysis of patients with blast crisis showed that high MSI2 strongly associated with the risk of relapse and risk of death, underscoring the potential of MSI2 as a prognostic tool for CML. Whether MSI2 proves to be relevant in other cancer types remains to be seen, but this study certainly introduces MSI2 as a new heavyweight in myeloid leukaemia and suggests that the modulation of MSI2 activity might be an effective therapeutic strategy.

Safia Ali Danovi

ORIGINAL RESEARCH PAPER Ito, T. *et al.* Regulation of myeloid leukaemia by the cell-fate determinant Musashi. *Nature*, 18 Jul 2010 (doi:10.1038/nature09171)

CORBIS

© 2010 Macmillan Publishers Limited. All rights reserved

Nature Reviews Cancer | AOP, published online 12 August 2010; doi:10.1038/nrc2920

GENOMICS

A broad brush

Gene signatures have indicated the existence of several subtypes of breast cancer, and a few signatures have proved to have prognostic value, but overall translation into the clinic has been limited. Anne-Lise Borresen-Dale and colleagues looked at the bigger picture in terms of changes occurring to whole chromosomes as well as chromosomal segments and identified rearrangement patterns that might have prognostic value, independently of gene signature and histology.

Part of the problem with moving gene signatures into the clinic has been the use of different platforms to generate these data — data produced on different platforms are not comparable. To overcome this limitation, these authors used comparative genomic hybridization data from four different breast cancer data sets to analyse gains and losses of whole chromosome arms (whole arm aberration index; WAAI) and changes within chromosome arms (complex arm aberration index; CAAI) using two new algorithms. All data sets were clinically annotated, which enabled chromosomal changes to be mapped to histology for all four data sets and outcome for two of the data sets.

Different subtypes of breast cancer (based on gene signatures) are often associated with different levels of chromosomal changes, so the authors split the tumours on the basis of those with whole arm gain of 1q or loss of 16q (group A), regional loss on 5q and regional gain on 10p (group B), those with all of these changes (group AB) and those with none of these changes (group C). WAAI and CAAI scores were then assessed for each of these groups and they were further subdivided on the basis of high or low CAAI scores.

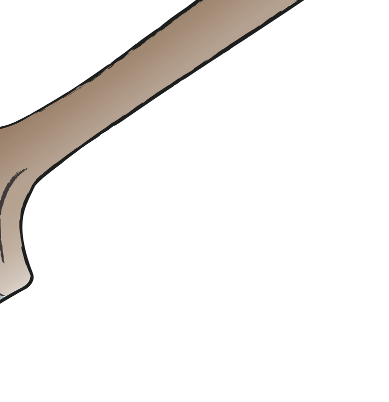
Breast cancer subtypes, as defined by gene expression analyses, correlate with the degree and type of genomic changes. Oestrogen receptor (ER)-positive tumours of low or intermediate grade were enriched in group A, and these also correlated with luminal A and luminal B breast cancer subtypes. Group B contained tumours that were more frequently high grade, aneuploid, had mutations in *TP53* and were dominated by basal-like breast cancers. Group AB were intermediate or high-grade tumours with the highest level of whole arm alterations, and they negatively correlated with the luminal A subtype. Group C had the fewest whole arm alterations, with approximately half the number of the grade 1 tumours, and had tumours of a special histological type, such as lobular or mucinous tumours. However, group C also contained ERBB2-expressing tumours and basal-like breast cancers.

The WAAI and CAAI also provided additional prognostic information; for example, patients with a group B tumour had double the risk of dying from their disease compared with patients with a group A tumour, and this was independent of lymph node status, tumour grade and size, and treatment. Patients in groups B and AB with high CAAI scores who did not receive adjuvant therapy had a poorer prognosis; such patients might benefit from more aggressive treatment. These and other data indicate that a high CAAI score might prove to be an independent prognostic marker for breast cancer.

The authors hope that future analyses of genomic changes in breast cancer might help us to understand the relationship between breast cancer heterogeneity and the cell of origin.

Nicola McCarthy

ORIGINAL RESEARCH PAPER Russens, H. G. *et al.* Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Sci. Transl Med.* **2**, 38ra47 (2010)



© 2010 Macmillan Publishers Limited. All rights reserved

Nature Reviews Cancer | AOP, published online 12 August 2010; doi:10.1038/nrc2921

HYPOXIA

The HIF2α puzzle

The transcriptional regulators hypoxia-inducible factor 1α (HIF1α) and HIF2α are overexpressed in many cancers, and it has been suggested that the HIFs could be potential anti-cancer therapeutic targets. Two studies from Celeste Simon and colleagues now describe differing roles of HIF2α in tumour development, indicating that therapeutically targeting this protein may not be straightforward and that context-dependent effects should be taken into account.

Tumour-associated macrophages (TAMs) migrate towards and accumulate in hypoxic regions of tumours and have been implicated in tumour progression, although their role is controversial as previous reports have described both beneficial and adverse effects of TAMs on tumour growth. As overexpression of HIF2α in TAMs has previously been correlated with high tumour grade and poor prognosis, Imtiyaz *et al.* examined the effects of myeloid-specific deletion of *Hif2a* in mouse models of inflammation-associated hepatocellular carcinoma or colitis-associated cancer. In both models, the authors observed that deletion of *Hif2a* in macrophages led to the decreased infiltration of TAMs into tumours, reduced tumour cell mitosis and delayed tumour growth, suggesting that HIF2α is needed for efficient TAM recruitment to tumours and that these inflammatory cells have a role in driving cancer progression.

The results of this first study are consistent with other studies that have indicated a tumour-promoting role for HIF2α; however, a study from Mazumdar *et al.* reports the surprising result that HIF2α can also have a tumour suppressive role. The authors deleted *Hif2a* in a *Kras*^{G12D} mouse model of lung cancer and found that loss of HIF2α promoted tumour progression. What transcriptional targets of HIF2α might be responsible for its tumour suppressive effect? The authors identified several downregulated HIF2α target genes, including the putative tumour suppressor *Scgb3a1*. They extended their studies to human lung adenocarcinoma A549 cells that contain an activating *KRAS* mutation, and observed downregulation of *SCGB3A1* after short hairpin RNA-mediated *HIF2A* knockdown and increased growth of these cells as tumour xenografts in immunocompromised mice. Restoration of *HIF2A* expression or ectopic expression of *SCGB3A1* in the xenografted A549 cells led to the suppression of tumour growth. HIF2α directly binds to sites in the *SCGB3A1* promoter, and levels of bound HIF2α are increased in hypoxic conditions. Moreover, the levels of *HIF2A* and *SCGB3A1* transcripts correlate in human non-small-cell lung cancer samples, supporting the idea that HIF2α directly regulates *SCGB3A1* transcription. SCGB3A1 has previously been reported to inhibit AKT signalling, and consistent with this, the authors observed increased levels of activated AKT after *HIF2A* knockdown, suggesting that AKT signalling is important for the increased proliferation of *HIF2A*-knockdown cells.

These authors suggest that although HIF2α overexpression can contribute to tumour progression, reduction of HIF2α below a crucial threshold level might have paradoxical tumour-promoting effects by reducing the expression of tumour suppressor genes, such as *SCGB3A1*, and this might have important consequences for HIF-targeted therapeutic strategies.

Meera Swami

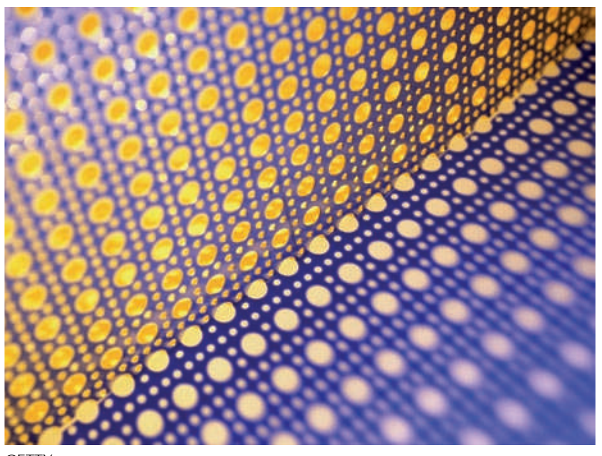
ORIGINAL RESEARCH PAPERS Imtiyaz, H. Z. *et al.* Hypoxia-inducible factor 2α regulates macrophage function in mouse models of acute and tumor inflammation. *J. Clin. Invest.* 19 Jul 2010 (doi:10.1172/JCI39506) | Mazumdar, J. *et al.* HIF-2a deletion promotes Kras-driven lung tumor development. *Proc. Natl Acad. Sci. USA* 21 Jul 2010 (doi:10.1073/pnas.1001296107)

© 2010 Macmillan Publishers Limited. All rights reserved

Nature Reviews Cancer | AOP, published online 19 August 2010; doi:10.1038/nrc2922

PROTEOMICS

A discovery strategy for novel cancer biomarkers



GETTY

Patient responses to targeted therapy can be predicted by examining the modulation of oncogenic signalling pathways in response to specific drugs, which can also identify new pathway-specific biomarkers. The validity of this approach is illustrated in a new study that uses a proteomics-based strategy to identify novel biomarkers that indicate a response of cancer cells to PI3K pathway inhibitors.

Andersen and colleagues used immunoaffinity precipitation followed by mass spectrometry to identify differential serine–threonine phosphorylation events in stable isotope labelling by amino acids in cell culture (SILAC)-labelled, PTEN-null PC-3 prostatic cancer cells that were treated with inhibitors of PDK1, AKT or a dual inhibitor of PI3K and mTOR. This approach allowed the quantification of 375 phosphopeptides that were relevant to PI3K pathway signalling, of which 71 were drug inhibited and 11 were reduced by all 3 PI3K pathway inhibitors.

The authors identified 'PI3K pathway nodes', which were the phosphorylation targets that were the most reduced by drug treatment. For example, phosphorylation of the ribosomal protein RPS6 was the most strongly inhibited by all three drugs, and PRAS40 was the most reduced phosphopeptide by AKT and PI3K–mTOR inhibition. To define common signalling modules, the authors analysed the phosphoproteins that were affected by all three inhibitors, and found that the modules most commonly inhibited were related to cytoskeletal functions, such as cell polarity and cytoskeletal organization. Additional modules commonly downregulated by the PDK1 and AKT inhibitors were related to vesicle transport and protein translation. These findings were consistent with the known inhibitory effects of these drugs on cancer cell motility, migration and invasion.

The authors identified phosphorylation sites that were uniquely downregulated by each PI3K inhibitor, several of which were novel. They suggested that these sites could be used as biomarkers for assessing responses to PI3K inhibitors and, as proof of principle, they evaluated the newly identified site phospho-Thr246-PRAS40 — which was reduced by the AKT inhibitor — as a biomarker. Andersen *et al.* generated an antibody against this epitope, and found that high levels of PRAS-Thr246 phosphorylation predicted sensitivity to an AKT inhibitor in lung and breast cancer cell lines. Moreover, the phospho-Thr246-PRAS40 epitope is more stable than the phospho-Ser473-AKT epitope that is commonly used in clinical diagnosis of tumours with AKT pathway activation, suggesting that this new biomarker might be more suitable for clinical evaluation.

Therefore, the signalling pathway-based strategy described in this study allows the discovery of drug-specific biomarkers and has important clinical implications. The authors suggest that these biomarkers can be used to guide treatment decisions, and that globally analysing signalling pathways using a differential phosphoprofiling approach could be a way to assess the off-target effects of existing cancer drugs.

Meera Swami

ORIGINAL RESEARCH PAPER Andersen, J. N. *et al.* Pathway-based identification of biomarkers for targeted therapeutics: personalized oncology with PI3K pathway inhibitors. *Sci. Trans. Med.* **2**, 43a55 (2010)
FURTHER READING Kolch, W. & Pitt, A. Functional proteomics to dissect tyrosine kinase signalling pathways in cancer. *Nature Rev. Cancer* **10**, 618–629 (2010).

© 2010 Macmillan Publishers Limited. All rights reserved

Nature Reviews Cancer | AOP, published online 19 August 2010; doi:10.1038/nrc2923

TUMORIGENESIS

Ground zero

Human prostate cancer is characterized by an expansion of epithelial luminal cells and an absence of epithelial basal cells. This has fed the common belief that this type of cancer has a luminal origin, especially because the lack of *in vivo* models that could properly recapitulate what happens in the human malignancy had meant that it could not be proved otherwise. Owen Witte and colleagues have now developed a 'human in mouse' model of prostate cancer that turns the luminal origin theory on its head.

Two main epithelial cell populations are found in human prostate tissue: basal cells, which can be identified by high levels of keratin 5 (K5) and K14 expression, and luminal cells, which are characterized by high levels of K18 and K8 expression. Starting with benign tissue from biopsy samples from patients with prostate cancer, the authors used the basal and luminal keratins as cell markers to sort the two different cell populations. They then transduced both cell populations with a lentivirus that encoded activated AKT and the transcriptional regulator ERG, as well as another lentivirus that encoded the androgen receptor (AR), and these proteins are known to be commonly altered in prostate cancer. The transduced cells were subcutaneously injected into immunodeficient mice. After 16 weeks, only the xenografts from basal cells resulted in the development of adenocarcinoma, which recapitulated the histological and molecular features of the human malignancy. The prostate adenocarcinomas showed an absence of basal cells and an expansion of luminal cells that expressed prostate-specific antigen (PSA), α-methylacyl-coA racemase (AMACR) and AR, all of which are markers of prostatic intraepithelial neoplasia and prostate cancer.

The results of this study establish a new *in vivo* model for investigating prostate cancer and, by identifying basal cells from the human prostate as a possible origin for this type of cancer, open up new possibilities for the identification of new markers for early detection.

Teresa Villanueva

ORIGINAL RESEARCH PAPER Goldstein, A. S. *et al.* Identification of a cell of origin for human prostate cancer. *Science* **329**, 568–571 (2010)

“Owen Witte and colleagues have now developed a 'human in mouse' model of prostate cancer that turns the luminal origin theory on its head.”

SUPERSTOCK

© 2010 Macmillan Publishers Limited. All rights reserved

Nature Reviews Cancer | AOP, published online 19 August 2010; doi:10.1038/nrc2926

THERAPY

Lethal cycling

Finding the Achilles heel of specific cancer types is a major goal in the development of targeted therapies. Now, recent evidence indicates that a new synthetic lethal strategy could be useful for oncogenic mutant KRAS-driven lung cancer.

IMAGESTATE

Following their genetic identification in yeast, cyclin-dependent kinases (CDKs) were thought to be essential, ubiquitous drivers of cell cycle progression. Although this remains true for the mitotic CDK (CDK1), knockout mouse studies showed that the interphase CDKs (CDK2, CDK4 and CDK6) are dispensable for viability and are only required for specialized, tissue-specific roles. This has raised questions about the validity of CDK-targeting agents as anti-proliferative therapies.

A team led by Mariano Barbacid examined the dependence on interphase CDKs of lung cancer driven by KRAS-G12V, a frequent lesion in human cancers of the lung, pancreas and colon. *In vitro* experiments using CDK-knockout mouse embryo fibroblasts (MEFs) revealed that deletion of *Cdk2*, *Cdk4* or *Cdk6* prevents the growth of KRAS-G12V-expressing MEFs in low-serum conditions and that knock down of any of these CDKs in human lung cancer cell lines is selectively lethal in KRAS-G12 mutant versus KRAS wild-type cells. However, follow-up *in vivo* studies in germline *Cdk2*- or *Cdk6*-knockout mice failed to prevent KRAS-G12V-induced lung tumorigenesis. By contrast, deletion of *Cdk4* resulted in marked senescence of KRAS-G12V-expressing preneoplastic lung cells, blocking their proliferation and preventing the development of high-grade lung tumours *in vivo*. Moreover, the loss of CDK4 function by either genetic excision or pharmacological inhibition in mice with established KRAS-G12V lung tumours significantly slowed tumour progression, providing further evidence of the therapeutic potential of this approach and implicating CDK4 in both the initiation and progression of KRAS-G12V-induced lung tumours. Interestingly, although KRAS-G12V expression was also induced in other tissues of the mouse, such as the pancreas and colon, these tissues do not undergo hyperplasia or develop tumours, and senescence was not induced in the absence of CDK4. This suggests that the induction of senescence after CDK4 loss could be specific to aberrantly proliferating tissues and so might be selective for tumour cells.

This work shows that certain oncogenic mutations can increase the cellular dependence on the interphase CDKs, thus renewing interest in CDK-targeted therapies. However, the success of these drugs in the clinic will depend on the context of the tumour tissue type and the specific driving oncogenes. Finally, it will be interesting to see whether this response to CDK4 inhibition will be sufficient to observe clinical responses in patients with oncogenic KRAS-mutant tumours.

Darren J. Burgess

ORIGINAL RESEARCH PAPER Puyol, M. *et al.* A synthetic lethal interaction between K-Ras oncogenes and Cdk4 unveils a therapeutic strategy for non-small cell lung carcinoma. *Cancer Cell* **18**, 63–73 (2010)

© 2010 Macmillan Publishers Limited. All rights reserved