Clinica Chimica Acta 495 (2019) 611–624

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/cca

Review

Oncoproteomics: Current status and future opportunities

Yujia He[a,1], Abidali Mohamedali[b,1], Canhua Huang[a], Mark S. Baker[c,*], Edouard C. Nice[a,c,d,**]

[a] West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu 610041, PR China
[b] Department of Molecular Sciences, Faculty of Science and Engineering, Macquarie University, New South Wales 2109, Australia
[c] Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Macquarie University, New South Wales 2109, Australia
[d] Department of Biochemistry and Molecular Biology, Monash University, Clayton, Australia

ARTICLE INFO

Keywords:
Oncoproteomics
Mass spectrometry
The hallmarks of cancer
The omics pipeline
Personalised/precision medicine

ABSTRACT

Oncoproteomics is the systematic study of cancer samples using omics technologies to detect changes implicated in tumorigenesis. Recent progress in oncoproteomics is already opening new avenues for the identification of novel biomarkers for early clinical stage cancer detection, targeted molecular therapies, disease monitoring, and drug development. Such information will lead to new understandings of cancer biology and impact dramatically on the future care of cancer patients.

In this review, we will summarize the advantages and limitations of the key technologies used in (onco) proteogenomics, (the Omics Pipeline), explain how they can assist us in understanding the biology behind the overarching "Hallmarks of Cancer", discuss how they can advance the development of precision/personalised medicine and the future directions in the field.

1. Introduction

Collectively, cancers are the 2nd leading cause of mortality worldwide, but may soon overtake cardiovascular diseases resulting in escalating economic and public health costs to society [1]. Despite the availability of a number of approved diagnostic tests for a range of different organ-of-origin cancer types (e.g. imaging techniques (X-Ray, CT, PET) or molecular markers such as carcinoembryonic antigen (CEA) for colorectal cancer and prostate-specific antigen (PSA) for prostate cancer), many patients are often diagnosed late, well after the onset of clinical symptoms and with accompanying life-threatening implications [2,3]. If cancer has already metastasised to vital organs, the prognosis is generally poor. Therefore, there is a critical need to explore molecular and cellular micro-environmental events that drive pre-cancer and cancer progression, in order to allow earlier detection at a time when the disease is treatable with curative intent, and importantly, to assist in the customization of therapeutic interventions driven by the concept of personalised/precision medicine [4–6].

Abbreviations: CEA, carcinoembryonic antigen; PSA, prostate specific antigen; PTMs, post-translational modifications; MS, mass spectrometry; IHC, immunohistochemistry; AR, antigen retrieval; FFPE, formalin fixed and paraffin embedded; OCT, optimal cutting temperature compound; CIPPE, Chemical Immobilization of Proteins for Peptide Extraction; GI, gastrointestinal; CRC, colorectal cancer; CCAP, CRC-associated fecal proteins; PDX, Patient Derived Xenografts; CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats; MW, molecular weight; NGS, next-generation sequencing; HPLC, high-performance liquid chromatography; MALDI, matrix-assisted laser desorption/ionisation; ESI, electrospray ionisation; OmpT, outer membrane protease T; ICATs, Isotope-coded affinity tag reagents; iTRAQ, isobaric tags for relative and absolute quantification; TMT, tandem mass taga; LDHA, lactate dehydrogenase A; IBT, isobaric tags; IPTL, isobaric peptide termini labelling; SILAC, stable isotope labeling using amino acids in cell culture; OSCC, oral squamous cell carcinoma; LFQ, label-free quantification; DDA, data dependent acquisition; DIA, data independent acquisition; SRM, Selected Reaction Monitoring; MRM, Multiple Reaction Monitoring; PRM, Parallel Reaction Monitoring; SWATH, Sequential Window Acquisition of all Theoretical Mass Spectra; PSM, peptide-spectra matching; SFS, statistical feature selection; DLD, deterministic lateral displacement; TATs, turnaround times; POCT, point-of-care testing; HUPO, Human Proteome Organization; cryo-EM, Cryo-Electron Microscopy; TEM, transmission electron microscopy; CX-MS, cross-linking coupled mass spectrometry; SOM, Self-Organizing Maps; SVM, support vector machines; iPOP, Integrative Personal Omics Profile; SOPs, standard operating procedures; NCI, National Cancer Institute; CPTAC, Clinical Proteomic Tumor Analysis Consortium; APOLLO, Applied Proteogenomics Organizational Learning and Outcomes; ICPC, International Cancer Proteogenome Consortium; ABLE, Accelerated Barocycler Lysis and Extraction; SELDI, surface-enhanced laser desorption/ionisation; CLCA1, calcium-activated chloride channel regulator 1; IPMN, intraductal pancreatic mucinous neoplasms; LCM, laser capture microdissection

* Corresponding author at: Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Macquarie University, New South Wales 2109, Australia.
** Correspondence to: Department of Biochemistry and Molecular Biology, Monash University, Clayton, Australia.
E-mail addresses: mark.baker@mq.edu.au (M.S. Baker), ed.nice@monash.edu (E.C. Nice).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.cca.2019.06.006
Received 25 April 2019; Received in revised form 5 June 2019; Accepted 5 June 2019
Available online 06 June 2019
0009-8981/

It is well established that cancer is a genetic disease that is triggered by the acquisition of somatic DNA lesions with aberrations in the human genome structure and other defects in DNA maintenance and repair. Several strategies and organisations have been developed to detect the aberrant changes in the genome that result in the onset of cancer [7,8]. For instance, The Seven Bridges Cancer Genomics Cloud (CGC; www.cancergenomicscloud.org) was funded as a pilot project by the USA's National Cancer Institute (NCI) to explore novel approaches to facilitate access to massive cancer genomic datasets alongside tools and computational resources to analyse the data [9].

With rapidly evolving next-generation sequencing technologies, the complexities of genomic landscapes of human cancers are progressively being uncovered. This has informed novel approaches for oncology clinical practice, such as assisting in diagnosis, prognosis, and treatment decisions. Large-scale cancer genome sequencing studies have opened the way for an understanding of cancer heterogeneity. However, genotypic information cannot accurately explain how heterogeneous genotypic variants are regulated and drive diverse phenotypic consequences - that is, how the information flows from genomes to functional proteomes in cancer [9–12]. The problem is so acute that although improved therapies are emerging (e.g. antibody therapy), drug resistance and selectivity are significant problems and require further investigation.

Proteins are the functional macromolecules of all cells, interacting with other macromolecules in complex networks controlling signal transduction and biological function. Many proteins are dynamically up- or -down-regulated in cancer biology, making tumour tissue a complex “orchestra” of proteins involved in driving tumour initiation, progression, response/resistance to treatment and recurrence [10]. Whilst the cancer proteome is initially encoded by the relatively static genome, the proteome has considerably higher diversity because of alternative splicing, single amino acid polymorphisms, post-translational modifications and many characteristics of proteins, including their structure, function, expression and disease-related changes, that are difficult to correlate with particular genomic-based alterations [8]. Proteogenomic and other multi-omics approaches are being developed to address this issue. However, sensitive and specific high-throughput methods to probe deeper into the cancer proteome are needed to facilitate an understanding of changes in disease-related signalling pathways, and the identification of associated protein isoforms and post-translational modifications [13].

Oncoproteomics, unravelled by a vast array of various omic technologies (namely the Omics Pipeline, Fig. 1), can be defined as the study of proteins and their interactions during cancer development and progression. It allows protein identification and quantification in complex biological systems and systematically reveals proteomic changes implicated in tumorigenesis, including protein abundance, protein post-translational modifications (PTMs), protein (and other molecular) complexes, biochemical pathway interactions, drug suitability and drug resistance [14]. Oncoproteomics is considered a promising approach for the identification of novel biomarkers for early clinical stage cancer detection, targeted molecular therapies and disease monitoring [15,16]. However, although proteomic approaches have already been widely applied to identify proteins associated with various cancers phenotypes, only a handful of biomarkers are currently FDA-approved (e.g. CA-125 [17], CA19–9 [18], CEA [19]), PSA [20], OVA1 panel [21]). Surprisingly, none of these is routinely used for early cancer stage clinical diagnosis. Several factors contribute to this dismal progress. A significant reason is that human blood, the most commonly used clinical/biological sample, is a complex mixture of proteins with a substantial dynamic concentration range ($\sim 10^{13}$ $\log^{10}$ in concentration). In terms of concentration, human blood plasma is dominated by several liver-derived “housekeeping proteins” that are particularly sensitive to environmental, physiological and disease conditions, making identification of novel low abundance biomarkers a significant challenge.

Additionally, tumour heterogeneity, study design, sample quality, sample collection and selection, the sensitivity and specificity of the chosen proteomic platform, bioinformatic evaluation and reliance on single biomarkers must also be taken into account as confounding factors [22,23]. Consequently, improved cancer treatment remains a long and winding road that needs to be carefully navigated, with a focus on what clinical questions and unmet clinical needs must be addressed. Oncoproteomics, coupled with other multi-omics technologies (Fig. 1) may address these issues.

In this review, we summarize the advantages and limitations of key technologies used in (onco)proteogenomics, explain how they can assist us in understanding the biology behind the overarching “Hallmarks of Cancer” proposed by Hanahan and Weinberg [24,25], illustrate their potential from selected recent studies, and discuss how they assist in the overall development of precision/personalised medicine. Finally, we contemplate perspectives and future directions in the field.

2. Oncoproteomics

New and continually improving technologies are being developed for systematically studying protein abundance, structure, function, post-translational modifications, turnover, and interaction with other proteins/ligands. These include quantitative proteomic methods, high-resolution, high-speed, high-throughput, high-sensitivity mass spectrometry (MS), more robust and specific antibody-based techniques, the use of protein chips, and advanced bioinformatics for data handling and interpretation (Fig. 2). These rapidly evolving technologies hold promise for achieving higher analytical capabilities, greater sensitivity and specificity, more robust and higher throughput sample processing, and greater confidence in the data generated [16,26]. Oncoproteomics typically combines both discovery and translational phases, where suitable biomarkers, biomarkers panels or target proteins identified in the discovery phase can be validated and developed into clinical assays. Typically increasing sample numbers are required as one progresses down the “omics” pipeline [27].

2.1. Samples for oncoproteomics

A wide range of sample types has been used at the discovery phase to provide physicians with actionable information leading to the evidence-based selection of potential biomarkers for early detection, optimal therapy (predictive biomarkers) and improved and more precise prognosis of disease progression (prognostic biomarkers). Samples include animal models, cancer cell lines, tumour tissues, exosomes and different bodily fluids (Table 1). For clinical analysis, specimens obtained using minimally invasive techniques (e.g. blood, urine or feces), or those gathered during an essential clinical investigation (e.g. biopsies) are ideal.

2.1.1. Tissue

Histology, one of the oldest biomedical sciences, has developed significantly since its first use over 200 years ago [28]. To decipher the specific molecular content of a given tissue, immunohistochemistry (IHC) techniques combined with classical microscopy were developed. Following the development of antigen retrieval (AR) procedures to ‘unlock’ the methylene bridges between proteins, both formalin fixed, and paraffin embedded (FFPE) tissues have been extensively applied for IHC studies. Recently, advances in mass spectrometry (MS) and improved separation systems have led to increased uptake of proteomics, enabling the rapid identification of large numbers of proteins (the proteome) with high confidence in a given tissue [29,30].

Tissue is the most relevant biological material for gathering insight into disease mechanisms. It contains a high concentration of potential disease markers compared with other biological samples, as markers concentrations can diminish with increased distance from the site of origin and dilution in the matrix [31,32]. The analysis of tissue proteins

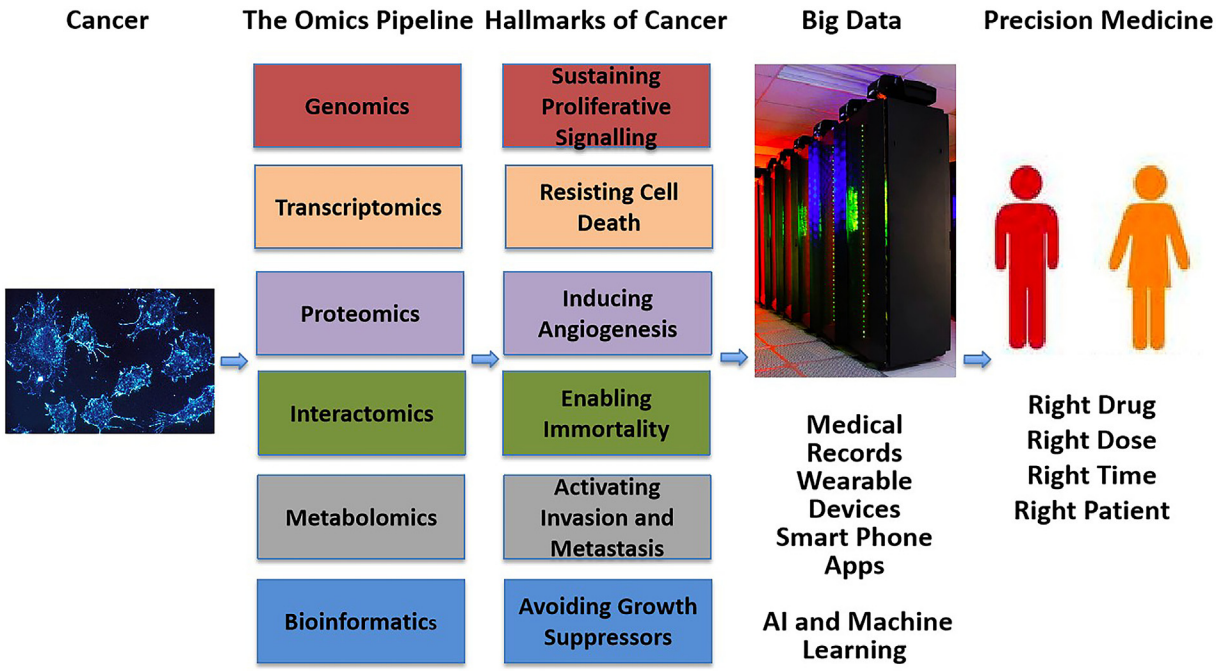

Fig. 1. From Oncoproteomics to Precision Medicine: The Omics Pipeline.
Patient samples will be analysed using genomic, transcriptomic and proteomic approaches (The Omics Pipeline). This in-depth analysis will help us understand the biology underlying the Hallmarks of Cancer [23]. These data will be aligned with the relevant medical records, and the resulting Big Data analysed using techniques such as AI and Machine Learning to support the central dogma of personalised medicine: The Right Drug at the Right Dose at the Right Time for the Right Patient.

is increasingly used for novel biomarker identification, therapeutic drug target discovery, and investigation of molecular mechanisms of disease [33]. For example, using tissue-based proteomics, Meding et al. revealed FXYD3, S100A11, and GSTM3 as novel prognostic markers predicting regional lymph node metastasis in colon cancer patients [34].

Although tissue proteome aids in the understanding of molecular mechanisms of human disease, some realities restrict their use in clinical research. One is that many tissue proteins have structural or housekeeping functions. These may provide limited physiological or disease-related information and mask the detection of less abundant tissue proteoforms. Several depletion strategies have been applied to

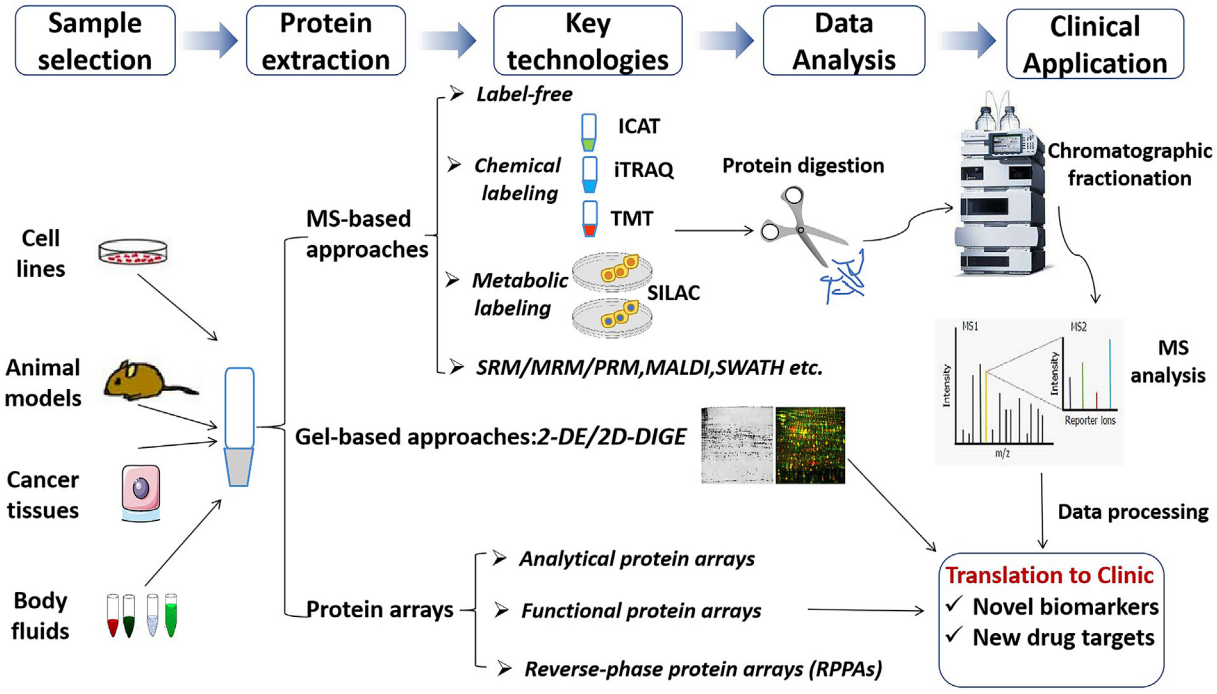

Fig. 2. The oncoproteomic toolbox.
Various sample types can be used for oncoproteomic studies, depending on the assay requirements, protein abundance, and sample heterogeneity. Key technologies used include gel-based approaches, MS-based approaches, and protein arrays. Proteomic profiling can be obtained using mass spectrometry strategies, followed by bioinformatics and statistical analysis of datasets. These data can then be translated into clinical practice to support personalised/precision medicine.

Table 1
Advantages and limitations of commonly used samples in proteomic research.

Sample types	Advantages	Limitations	Reference
Tissue	➢ Highest concentration of potential disease biomarkers ➢ Widely used in the discovery of novel cancer biomarkers.	➢ Interference of disease-unrelated abundant proteins. ➢ Hampered by the extraction of proteins and artificial modifications induced by the fixation process.	[28–34]
Blood	➢ Minimally invasive collection. ➢ Low cost. ➢ Readily accessible. ➢ Extensive application in clinical disease diagnosis.	➢ Interference of disease-unrelated abundant proteins. ➢ Sensitive to environmental changes. ➢ Often accompanied by other physiological process (e.g. inflammation and acute phase reaction).	[2,15,35]
Feces	➢ Non-invasive collection.	➢ Easily influenced by dietary day-to-day variations.	[36–45]
Animal models	➢ Providing in vivo complex environments.	➢ Not naturally endowed with specific features relevant to human diseases. ➢ Long time to construct.	[46–48]
Human xenograft models	➢ Allowing direct evaluation of the biological behavior of human tumour proliferation and metastasis in vivo.	➢ Limited availability of fresh tumour tissues. ➢ Occasionally low survival of the transplanted tissue. ➢ Tumours derived from cell lines do not completely recapitulate normal tumour architecture.	[49]
Cell lines	➢ Easy to work with. ➢ Reduced complexity of cellular proteome.	➢ Lack of tumour microenvironment.	[44,51]
Other biofluids (urine and proximal fluids)	➢ Non-invasive collection. ➢ Unlimited sample volume.	➢ Results need to be normalized for volume.	[23,56–59]

remove large and highly-abundant species (e.g. immunoaffinity depletion, centrifugal ultrafiltration). For example, Alvarez et al. used an integrated capillary LC-ESI-QTOF-MS approach, which showed potential in the study of tissue biology, for the exploration of the low MW and/or lower abundance tissue proteome [32]. Secondly, fresh frozen tissue, regarded as the optimum source for such studies, is not always available. The use of tissues embedded in optimal cutting temperature compound (OCT) or formalin-fixed and paraffin-embedded (FFPE) is hampered by the extraction of the proteins as well as artificial modifications induced by the fixation process [31] and the fact that OCT contains high concentrations of polymers that interfere with MS. Recently, investigators from Johns Hopkins University reported a method named Chemical Immobilization of Proteins for Peptide Extraction (CIPPE) to eliminate interferences from tissues and OCT embedding [30]. This protocol is highly efficient in the removal of polymers and contaminants. CIPPE has been used for extracting peptides for both iTRAQ-labelled and 2D liquid chromatography-tandem mass spectrometry shotgun proteomics studies [30].

2.1.2. Blood

Blood (serum or plasma) samples are widely used to investigate various physiological or pathological states since the collection is minimally invasive and routinely used in the clinic [2]. In cancer, changes in the blood proteome may result from cancer-specific or cancer-related host organ changes in biological pathways, often leading to metabolic or immunological alterations [15]. Indications of cancer status in the blood come from various cellular components, such as circulating tumour cells, exosomes, cell-free DNA/RNA, proteins, peptides, lipids, and metabolites. Among these, proteins detectable in serum or plasma are most commonly used to screen and monitor cancer (e.g. prostate-specific antigen (PSA) for prostate cancer). Furthermore, proteins can also be used to monitor response to therapy and recurrence (e.g. antigens CA125 for ovarian cancer [17], CA19–9 for pancreatic cancer [18] and carcinoembryonic antigen (CEA) for advanced colorectal and ovarian cancer [19]).

However, there are still several challenges using blood (serum or plasma) in proteomics. Firstly, compared with other abundant proteins in the blood, the identification of low abundance biomarkers can be challenging. Blood samples can contain over 10,000 different proteins with concentrations varying over twelve orders of magnitude [35]. Secondly, the process of oncogenesis is often accompanied by inflammation and various acute phase reactions in the liver, which can stimulate cells to produce specific proteins such as hormones, cytokines, or other molecular sensors, masking the measurement of cancer-specific protein biomarkers. Thirdly, the circulatory system is dynamic, rendering the blood proteome sensitive to many types of environmental change (e.g. temperature, starvation, food intake, circadian rhythm, sleep).

2.1.3. Feces

Feces are among the most accessible of biological samples that can be used for biomarker discovery, containing proteins and peptides derived from leakage, secretion or exfoliation of diseased tissues [36]. Feces come into direct contact with both the human gastrointestinal (GI) tract and the gut microbiome, making feces a prominent source to search for biomarkers for all types of GI tract disease [37]. Importantly, biomarkers are present at higher relative concentrations in feces than in blood [33]. Feces collection is non-invasive, does not require the unpleasant bowel preparation needed for a colonoscopy, can be undertaken at home and is not sample limiting. However, there is reticence by a significant proportion of individuals offered stool screening (e.g.FOBT/FIT) to participate, with test uptake as low as 40% in most countries. Education is urgently needed, as diseases like colorectal cancer (CRC) are curable by surgical resection if detected at an early clinical stage.

Feces have been successfully used to monitor gastrointestinal tract pathologies, including CRC [38,39]. For instance, Ang et al. [38] in a pilot study demonstrated the potential of fecal proteomics to identify potential CRC biomarkers using orthogonal multi-dimensional fractionation (1-D SDS-PAGE, RP-HPLC, size exclusion chromatography). They generated a library containing 108 CRC-associated fecal proteins (CCAP) with associated peptide and MS/MS data used for multiplexed SRM analysis. In a recent study [39], Bosch et al., using 315 stool samples from 12 patients with CRC and 10 people without colorectal neoplasia (control samples) and a second series of 81 patients with CRC, 40 with advanced adenomas, and 43 with non-advanced adenomas, as well as 129 control samples, identified novel protein biomarkers in stool that outperform, or complement, hemoglobin in detecting CRC and advanced adenomas.

Another rapidly emerging field is the analysis of the fecal microbiome. It is realised that changes in the microbiome are associated with several diseases, including cancer [40,41]. Indeed, it has recently been reported that different microbiota composition was associated with patients' clinicopathological features in CRC, including metastasis [42,43]. It has been noted that for cancer biomarker analysis, CCAP needs to be discriminated from dietary related proteins and bacterial proteins derived from the intestinal microbiome [44,45], although practically this does not appear to be a problem [36].

2.1.4. Animal models

Experimentation and observation of animal biology can aid understanding of human biology, health, and disease [46]. In vivo environments to understand human disease has relied on the use of animal models, traditionally mice, rats, and rabbits. More recently, animal species with pathology more closely related to the human situation have been used including dogs, pigs, primates [47] and zebrafish [48] that have a similar genetic structure to humans. Recombinant proteins can be used to modulate such systems. Mouse xenografts, in which a patient's tumour tissues are implanted into an immunodeficient or humanised mice (Patient-Derived Xenografts (PDX)) have also been used extensively in oncoproteomics [49]. The use of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) technologies enables animal models of human disease to be manipulated by genome editing [50].

2.1.5. Cell lines

Similar to tumour tissue, tumour cell lines are widely used in cancer research. Cell lines derived from tumours offer several advantages over clinical samples for biomarker identification. They are homogeneous cell populations, which can be quickly grown and can be manipulated by altering gene expression. Furthermore, it is easier to obtain subcellular fractions (e.g. plasma membrane, nucleus, secretome, exosomes) from cell cultures. Additionally, the complexity of the cellular proteome is reduced, simplifying the detection of the low-molecular-weight (MW) species and identification of novel biomarkers [44,51]. For instance, to investigate cellular downstream pathways altered by MUT deficiency, Costanzo et al. used small-interfering RNA to knockdown MUT protein expression in a neuroblastoma cell line (SH-SY5Y). Using label-free quantitative proteomics, they found deregulation in the levels of mitochondrial proteins, such as electron transfer flavoprotein subunit alpha and 2-oxoglutarate carrier, crucially involved in the mitochondrial oxidoreductive homeostasis balance [52]. However, there are downsides to using cell lines. The major problem is that cell lines often show instability after multiple passaging, plus they lack the appropriate tumour microenvironment [51]. A recent multi-omics study looking at HeLa (cervical cancer) cell lines across laboratories (14 stock HeLa samples from 13 international laboratories) showed significant heterogeneity [53]. Such variability could have significant implications on the reproducibility and interpretation of data generated using human cultured cell lines as indicated by a recent study where cell-based variability was found to be a contributing factor to assay discordance in four next-generation sequencing (NGS) assays [54]. More recently, the use of 3D culture techniques has shown promise. For example, 3D endometrial cancer spheroids were obtained by co-culturing endometrial stromal cells with cancerous epithelial cells. The spheroids developed using this approach were phenotypically similar to in vivo endometrial cancer tissue. Proteomic analysis of the co-cultured spheroids revealed 591 common proteins and canonical pathways associated with endometrial biology [55].

2.1.6. Urine

Urine has proved a viable alternative biological sample, especially for biomarker and metabolomics studies. The collection is non-invasive, and the sample volume is not limiting. However, results need to be normalized for volume, typically using creatinine concentration as a standard. Urinary proteome profiling has recently been reported to predict lung cancer from other pathologies using a panel of 5 protein biomarkers [56]. In another study, the urinary glycopeptidome of 238 healthy controls and 969 patients with five different cancer types: bladder, prostate and pancreatic cancer, cholangiocarcinoma, and renal cell carcinoma was analysed. Three O-glycopeptides and five N-glycopeptides showed significant differences in their concentration in the different cancer types, compared to healthy controls [57].

2.1.7. Proximal fluids

Proteomic analysis of proximal fluids (e.g. ductal fluids (pancreatic or biliary cancer), bronchoalveolar lavage (lung cancer), nipple aspirate (breast cancer) seminal prostatic fluid (prostate cancer) or saliva (head and neck cancer)) have been reported [58]. Whereas the concentration of tumour-derived proteins is typically deficient in blood, their abundance could be significantly increased at, or close to, the tissue source. However, obtaining such biosamples may be more invasive than a routine blood collection [23,59].

2.2. Separation techniques

Due to the complex nature of most biological samples, some form of upfront separation is typically required to maximise the depth of proteome coverage. Multiple (multidimensional) orthogonal separation techniques are frequently used in discovery phase experiments [60].

2.2.1. Gel-based approaches

1D-PAGE, 2D-PAGE, and 2D-DIGE (see [61–63] for comprehensive reviews) have been used extensively for differential proteomic analysis of a range of biological samples to investigate the molecular basis of cancer pathogenesis and to identify potential new biomarkers or therapeutic targets. The resolution allows thousands of proteins to be separated simultaneously [64]. In an example of this approach, Kirana and colleagues used the combination of laser microdissection, 2D DIGE with saturation labelling and MALDI TOF/TOF to identify protein biomarkers for predicting CRC spread. They compared stage II CRC patients whose tumours did not metastasise from those whose tumours metastasised to the liver [65]. Based on orthogonal *in cellular* and in vitro functional proteomic approaches, their data showed the importance of the uPAR·αvβ6 interaction in CRC [66]. They went on to show that epithelial and stromal cell urokinase plasminogen activator receptor expression differentially correlated with survival in rectal cancer stage B and C patients using epitope-specific anti-uPAR monoclonal antibodies and IHC [67].

Peptide isoelectric focusing in the first dimension on immobilised pH gradients (peptide IPG-IEF), using SCX or SDS-PAGE to separate peptides or proteins before LC-MS/MS analysis, is advantageous as peptide pIs have been shown to be an accurate primary filtering criterion to reduce false positive or negative identifications [68]. The application of peptide IPG-IEF revealed that greater coverage of the membrane proteome could be achieved using IPG strips with various pH ranges [69].

2.2.2. Micropreparative HPLC

While 2D-PAGE-based approaches have proved useful in proteomics studies, there are limitations with these techniques, including an insufficient resolution for large proteomes, incompatibility with hydrophobic proteins and the time-consuming process required to “pick” and perform in-gel digestion of hundreds of spots for MS identification. Multidimensional micropreparative high-performance liquid chromatography (HPLC) techniques have been developed as an alternative to gel-based approaches and are compatible with the recovery of low-level sample components [60]. High purification factors are obtained by using chromatographic supports with different functional properties (e.g. charge, hydrophobicity, size, affinity).

Sample preparation needs to be optimised for various factors, including sample matrix, sample complexity, and the goal of the research [70]. Techniques such as depletion [71] and ultradepletion [72] as well as enrichment techniques (e.g. antibodies, ligand dyes, metal binding, lectins) have been successfully used to mine deeper into the proteome [60].

2.2.3. High-throughput sample preparation

As the field moves towards the development of clinical assays, there is a requirement for the development of high-throughput sample

processing compatible with downstream MS/MS analysis. Such methods should be simple and ideally capable of automation.

2.3. Mass spectrometry (MS) -based approaches

The potential of oncoproteomics has primarily been realised by the development of matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI) MS for which Tanaka [73] and Fenn [74], respectively, shared the Nobel Prize for Chemistry in 2002. (For comprehensive reviews of these technologies see [75–77]).

Three major strategies have been used for identification purposes. In bottom-up, or shotgun, proteomics, proteins are enzymatically cleaved to generate peptide fragments (typically using trypsin) before MS/MS analysis. In middle-up proteomics, larger peptide fragments (ideally in the 3–15 kDa mass range) are generated by chemical reactions (e.g. CNBr) or proteases (e.g. pepsin) that are specific for less frequent residues, or through restricted enzymatic hydrolysis (e.g. outer membrane protease T (OmpT)). In top-down protocols, proteolysis is omitted, and intact proteins are detected and fragmented in the MS gas phase. Top-down is particularly attractive for the analysis of PTMs [78], or small proteins that do not yield appropriate proteotypic peptides using other approaches.

2.3.1. MALDI imaging

Another technique that is now maturing is that of MALDI imaging. In this technology, thin tissue slices (either frozen or formalin fixed) are mounted on a MALDI target plate that can be moved in an XY orientation. Matrix is applied and a laser rasterised over the tissue surface with MS of molecules present (e.g. proteins, peptides, lipids, small molecules, or glycans) recorded as intensity maps. These can be compared with corresponding H&E or IHC images. As an example, MALDI-MS Imaging, LC-MS/MS and immunohistochemistry were used to screen fresh-frozen and formalin-fixed samples from a patient with head and neck cancer to localise the tumour edges [77]. Another recent study on medullary thyroid carcinomas [79] showed MALDI-MSI could detect the classic immunohistochemical markers used for diagnosis (e.g. calcitonin, thyroglobulin) with good sensitivity and specificity. Additionally, the combination of MALDI-MSI and nLC-ESI-MS/MS analysis allowed further potential markers to be identified and their spatial localisation visualised within tumour regions.

2.4. Quantitative MS

Accurate quantitation is essential for measuring differential disease related to proteomes in oncoproteomics. Several approaches have been developed, as discussed below.

2.4.1. Stable isotope-based methods

The use of stable isotopes allows measurement of peptides or proteins utilising the specific increase in mass in individual samples by incorporating mass tags with stable isotopes, enabling simultaneous analysis and peptide quantification for multiple specimens with direct comparison across samples [80].

Isotope-coded affinity tag reagents (ICATs) paved the way for the development of stable isotope-based quantitative techniques. ICATs are biotinylated iodoacetamide derivatives that react with the sulfhydryl groups of cysteine side chains in denatured peptides or proteins. After tryptic digestion, the labelled peptides are purified by affinity purification using avidin-agarose beads [80,81]. Although widely applied, there are still some weaknesses in the ICAT technology, including complex and time-consuming data analysis, extensive sample fractionation, nonspecific binding to the streptavidin affinity matrix and many subsequent reactions at the same site [8,82].

2.4.2. Chemical labeling

Chemical labelling with isobaric tandem mass tags is an alternative method of quantitative analysis, allowing a comparison of multiple states in a single experiment. Two isobaric tag families are commercially available: Isobaric tags for relative and absolute quantification (iTRAQ) and tandem mass tags (TMT). iTRAQ, first described by Ross et al. [83] in 1999, and widely used in the biomarker discovery field, allows up to eight individual samples to be measured simultaneously. Kuan-lin Huang et al. used iTRAQ for quantitative proteomic profiling of 24 breast cancer patient-derived xenografts (PDX). Integrated proteogenomic analysis indicated a positive correlation between the transcriptomics and proteomics data. The proteogenomic analysis also revealed several receptor tyrosine kinase targets. Drug treatment studies targeting HER2 and components of the PI3K pathway supported the proteogenomic predictions from seven of the xenograft models. Such strategies could enable improved prediction of the efficacy of mechanism-based breast cancer therapeutics [84].

The major distinction between iTRAQ and TMT is that the labels differ in structure and mass [26,85]. In chemical labelling, tagging is achieved by using chemical moieties or tags which are identical in mass, so that all derivatised peptides are isobaric, chromatographically identical and yield a single peak in the primary mass spectrum for both samples (e.g.normal and cancer samples). The relative abundances of the isobaric-tagged peptides are revealed when the moieties fragment during MS/MS to release reporter ions with different masses. Using iodoTMT labelling, Yuan Liu et al. [86] found that lactate dehydrogenase A (LDHA) preferably locates to the nucleus in HPV16-positive cervical tumours, uncovering a non-canonical enzymic activity of nuclear LDHA which epigenetically controls cellular redox balance and cell proliferation, facilitating HPV-induced cervical cancer development. However, the high experimental cost and inconvenient handling for large samples sizes has limited the application of TMT and iTRAQ, especially for the quantitation of modified peptides. To address this, some promising isobaric strategies with high stability and low cost are currently under development, such as 10-plex isobaric tags (IBT) [87] and isobaric peptide termini labelling (IPTL) [88].

Unlike postharvest-methods (e.g. ICAT, iTRAQ, and TMT), stable isotope labelling using amino acids in cell culture (SILAC) incorporates the isotopic label into the protein when the tissue/cells are still metabolically active. The first comprehensive, quantitative proteomic analysis of tissues from whole animals was described by Krüger et al., who extended the SILAC approach in vivo by feeding mice with a $^{13}C_6$-lysine diet. This enabled complete labelling of proteins in all tissues and allowed quantitative protein analysis from whole organs down to subcellular compartments and specific peptides [89]. Yu-Lin Chen and colleagues identified the underlying mechanisms of lymph node metastasis in oral squamous cell carcinoma (OSCC) using a combination of SILAC and LC-MS/MS [90].

A significant shortcoming of metabolic labelling is the requirement for samples to be undergoing active protein synthesis, making it impossible to utilise the original technology for many clinical specimens. To overcome this drawback, super-SILAC has been developed. Here, a mixture of SILAC-labelled cells is used as a spike-in standard to analyse human tumour tissues [91]. Super-SILAC and TMT labelling strategies have recently been used to interrogate proteomic heterogeneity in metastatic lung adenocarcinoma. Using a longitudinal, prospective set of tumour tissues from a metastatic lung adenocarcinoma patient, proteomic and phosphoproteomic data was generated. Targeted MRM assays revealed two novel variant somatic peptides (CDK12-G879 V and FASN-R1439Q) expressed in lung and lymph node metastatic sites, respectively which may influence the overall prognosis and treatment response of the patient. [92]

2.4.3. Label-free quantitative MS

Although isotope-labelling strategies are precise and accurate, the labelling procedures are complicated and expensive. Furthermore, isotope-based quantification is limited by sample number and is not compatible with all experimental designs [93]. By contrast, label-free

quantification (LFQ) approaches are more generally applicable, making them ideally suited for non-targeted discovery proteomics applications [94,95]. Data-dependent acquisition (DDA) or data-independent acquisition (DIA) methods can be used. Although label-free strategies are theoretically scalable to any number of samples, analysis is currently restricted by instrumental capabilities. However, approaches are being developed to improve label-free quantification [91].

2.4.4. SRM/MRM/PRM-based quantitative MS

Targeted approaches, such as Selected Reaction Monitoring (SRM), Multiple Reaction Monitoring (MRM) or Parallel Reaction Monitoring (PRM), are currently gaining popularity, due to their high sensitivity, high specificity, and multiplex potential. They are particularly useful for the validation of biomarker panels, where antibody-based technologies would be too expensive and time-consuming to generate.

During the process of SRM/MRM, several targeted proteotypic peptides and their corresponding transition ions are selected from data generated in the discovery phase. MRM can be applied to multiple peptide fragments, making it capable of detecting multiple isoforms and post-translationally modified species in a single test run [93,96,97].

Sensitivity can be further enhanced by the use of immune-SRM. In this case, enrichment is achieved by the use of antibodies or aptamers. Several companies are now offering these capture methods commercially. It should be noted that the requirements of antibody or assay specificity are less stringent than in ELISA, as specificity is achieved by MS phase of the methodology.

2.4.5. Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH)

SWATH is based on the cyclical acquisition of precursor ions with fixed isolation windows that cover the entire *m/z* range. SWATH is an emerging DIA platform capable of label-free quantification with higher accuracy and precision, and capable of comprehensive detection and quantitation of virtually every component in a sample [93,98]. SWATH data consists of highly multiplexed fragment ion maps recorded over a user-defined mass precursor mass range across the complete chromatographic separation. Importantly SWATH has been shown to have excellent inter- and intra-lab variation [99].

SWATH typically relies on spectral reference libraries for peptide identification. Quality and coverage of these libraries are vital for optimum analysis. Recently, three fractionation strategies to increase local library size (protein SDS-PAGE fractionation, peptide high-pH RP-HPLC fractionation, and MS-acquisition gas phase fractionation) were compared to standard library building using replicate DDA injection. Data was evaluated in terms of the number of extracted peptides and proteins, the match quality of peptides and the extraction reproducibility of transitions. The results showed that SWATH quantitative results and interpretations are affected by the fractionation method used [100].

Additionally, SWATH analysis revealed sophisticated metabolic reprogramming in hepatocellular carcinoma tissues during development and progression [101]. In another study [102], SWATH was used to classify breast cancer subtypes with varying degrees of heterogeneity using 96 tissue samples representing five breast cancer subtypes according to conventional classification. Significantly, proteomics analysis revealed varying degrees of heterogeneity within the conventional subtypes, with triple negative tumours being the most heterogeneous. This data demonstrates how large-scale protein-level measurements using next-generation proteomics can assist improved patient stratification for precision medicine.

Bioinformatics plays a crucial role in DIA analysis, especially for peptide-spectra matching (PSM). Programs such as proBAMsuite, a bioinformatics framework for proteomics and proteogenomics research, allow users to browse cancer proteomics datasets. PSMs can be re-annotated using user-specified gene annotation schemes and assembled for both protein and gene identifications [103]. Another example is iSwathX that allows users to combine multiple libraries to interrogate DIA datasets [104]. Bioinformatics will also help confirm the existence of missing proteins (proteins which currently do not have high-stringency evidence (e.g. mass spectrometry) [105] or in dealing with biological and technical heterogeneity in data and statistical feature selection (SFS) [106].

3. Other technologies

Several other technologies supporting oncoproteomics warrant mention.

3.1. Protein arrays

Protein arrays (synonymous with protein microarrays or protein biochips) offer a solution to the high-throughput study of protein interaction networks or immune reactivity [107]. They have also proved to be a suitable tool for antigen and autoantibody profiling in complex samples across many diseases [108]. High-throughput protein microarrays offer a direct approach for simultaneously screening thousands of antigens in an unbiased manner using minimal (μl) amounts of sample.

Three types of protein arrays have been reported: analytical protein microarrays, functional protein microarrays (or target protein arrays) and reverse phase protein microarrays. In analytical microarrays, antibodies are immobilised on the surface, and retained proteins are detected in a sandwich-type assay. In functional arrays, purified proteins, typically recombinant materials, are immobilised on the array surface while in reverse phase arrays complex samples, such as tissue lysates, are immobilised and probed with antibodies against the target protein of interest. These antibodies are typically detected using chemiluminescence, fluorescence or colourimetric readouts. Parallel monitoring of the expression of a particular protein in hundreds to thousands of samples is possible [109].

Aptamer arrays have also been developed. Recently they have been validated for serum profiling to predict clinical outcome of stage IIA colon cancer patients [110]. SomaLogic, a US-based company, is currently developing a range of aptamer chips for cancer diagnosis.

3.2. Microfluidic devices

Typically a microfluidic device comprises a set of micro-scale channels etched into a surface (e.g. glass, silicon or polymers) for manipulating/processing small samples. These can be reactions with required reactants, especially for disease diagnostics or even cell culture with a controllable microenvironment. They allowa precise examination of chemical and biological processes. A range of proteomics-compatible microfluidic devices have been developed (see review by Pedde et al. [111]). A particular focus has been on the isolation of exosomes where immunoaffinity, membrane-based filtration, trapping on nanowires, acoustic nanofiltration, and deterministic lateral displacement (DLD) sorting have been utilised [112]. In one recent application of this technology, self-assembled herringbone nanopatterns were used to isolate exosomes from ovarian cancer patients and identify potential protein biomarkers for early detection and monitoring of the disease [113].

3.3. 3D printing

A typical workflow for the analysis of biological fluids in clinical laboratory frequently requires extraction of the targets from the sample, reduction of the sample volume, and separation on a liquid chromatograph with mass spectrometry detection, which is time-consuming and often incurs significant delays in turnaround times (TATs) [114,115]. Due to the critical need for accurate, precise and rapid diagnostics, improved point-of-care testing (POCT) devices have been developed to enable individual patient self-monitoring, especially using three-

dimensional (3D) printing [116], facilitating telemedicine.

3D printing, a fabrication process allowing the translation of 3D computer designs into physical models through the additive patterning of various material, is regarded as a revolutionary, industry-compatible approach, which helps iterative design-build-test cycles of multiple components, from microfluidic chips to smartphone interfaces, that are geared towards point-of-care applications [117–119]. Recent advancements in 3D printing technologies have enabled the rapid and cheap fabrication of large numbers of highly complex microfluidic devices that can be integrated into various stages of proteomics protocols. In one example, a 3D printed cartridge was designed for targeted MS detection of plasma proteins. It utilised an integrated antibody enrichment column combined with a built-in substrate for protein ionisation and a spray tip for coupling to the MS [120].

3.4. Cryo-electron microscopy (cryo-EM)

In cryogenic electron microscopy (cryo-EM) samples are cooled to cryogenic temperatures and embedded in an environment of vitreous water and analysed using transmission electron microscopy (TEM). This method preserves the biological samples at near-native state. This technology has been described as the “revolution of the century”, and was recognised by the award of the Nobel Prize for Physics to Professors Dubochet, Frank and Henderson in 2017. Structural resolution at better than 3 Å has been reported [121].

It is now realised that the combination of cryo-EM with cross-linking coupled mass spectrometry (CX-MS) allows more in-depth interrogation of protein structures, giving a higher level of detail than either technology in their own right [122]. This results in highly refined 3D structure of protein assemblies at high resolution. Importantly, CX-MS cross-links regions in a protein that are not covalently bound, allowing visualisation of these interactions that are not seen in EM, providing further evidence for MS-based interactomic studies.

4. The big data problem

Mass spectrometry is heavily data-intensive as terabytes of qualitative and quantitative data can be generated from a single sample in a single experiment. This, combined with the availability of more accurate and detailed patient meta-data from samples (e.g. clinical information, treatment, staging) as we move towards precision/personalised medicine, a push to increase sample numbers to match genomic datasets together with the added complexity of multi-omic datasets [123], makes data analysis challenging. Here again, bioinformatic tools play an essential role in analysing, integrating, and understanding such proteomic data [124,125]. However, ultimately, one of our biggest challenges may be how to deal with this flood of information (the Big Data Problem). Three main challenges can be identified. Firstly, there is the statistical challenge of not only extracting useful data [126] but also applying various statistical methodologies to comprehend the data [127]. Secondly, and perhaps more importantly, it is crucial to put the data in a biological context to explain known, expected, or unusual biological phenomena. Finally, it is essential to be able to correlate this data with similar datasets available publicly or with published data often obtained using diverse instrumentation and methodology [128].

Oncoproteomic experiments usually result in the generation of data sets that contain protein identifications and quantifications layered with multiple variables including, but not limited to, patient demographic data, stage of cancer, treatment regimen, etc. After standard univariate or multivariate statistical analysis, differentially expressed proteins are displayed using a variety of conventional visualisation techniques (ROC curves, box plots, heat maps, volcano plots, etc.). The use of bespoke and novel statistical analysis techniques (including machine learning and AI clustering methodologies (e.g. Self-Organizing Maps (SOM), random forest, support vector machines (SVM) [16], Markov clustering) have only recently started to be used as the data sets become more substantial and more reliable [129]. These techniques allow a more in-depth interrogation of proteomics data to generate novel hypotheses. However, these are often blinded studies devoid of biological context.

Additionally, while novel analysis methods such as topological scoring of protein interaction networks [130] and hierarchical clustering [131] have been performed on various datasets, they are not yet routinely used in oncoproteomics. Data usually needs to be validated using some form of orthogonal technique to verify quantitative and qualitative information such as in vitro/in vivo studies, Western blotting, drug or peptide inhibition. Meta-analysis usually combines gene ontology analysis to understand expression data and often other, more sophisticated data analysis software tools such as Ingenuity pathway analysis. Using such approaches, VEGFR2 was shown to promote tumorigenesis in gastric cancer [132] and uncover molecular signatures of breast cancer cell proliferation [133]. Complementary software such as Ingenuity [134], Cytoscape [135] or String [136], KEGG [137] can be used to gain insight into molecular pathways of differentially expressed proteins to give further context. However, this pipeline has received some criticism for lack of stringency, often low sample numbers, lack of concrete and repeatable biological validation, and the scarcity of validated biomarkers that are sensitive and specific enough to discriminate cancer from other disorders or ailments.

5. Hallmarks of cancer: putting oncoproteogenomics data into context

In 2000, Hanahan and Weinberg published their seminal Hallmarks of Cancer review, which essentially devised a framework to understand the basic biologies of cancer [24]. This framework was revised in 2011(almost a decade later) [23] and provided a comprehensive understanding of the complex biological mechanistic landscape of cancer and the multiple effects these can exert. In their latest publication [23], seven biological activities were defined as being acquired during the multistep development of all human tumours. These comprise (1) Selective growth and proliferative advantage, (2) Altered stress response favouring overall survival, (3)Vascularization, (4) Invasion/metastasis, (5) Metabolic rewiring, (6) Generation of a compatible microenvironment, and (7) Immune modulation.

Although based primarily on genomic data [25], this framework has proven invaluable in understanding the mechanisms behind cancer development. To date, only a handful of oncoproteomics studies have utilised the breadth of the Hallmarks of Cancer framework to infer biological correlations. Examples include Gomig et al. [138] who predicted the physiological effects of breast cancer cells using quantitative label-free mass spectrometry on contralateral and adjacent breast tissues to reveal differentially expressed proteins and their predicted impact on pathways and cellular functions in breast cancer whilst others have inferred specific biological pathways that can potentially lead to therapeutic interventions [139].

Recently, several new databases have been developed to help extract meaning from large quantitative datasets like those generated through most proteomics and proteogenomics studies. The Cancer Hallmarks Analytics Tool (CHAT) [140] allows users to mine literature and index the various hallmarks associated with differentially-expressed proteins in their particular dataset. However, these are currently gene-based tools, and it is vital that more specific proteomics-based biological frameworks are developed, ideally at the peptide level, in addition to superimposing data on patient heterogeneity.

6. Drug development

The ‘OMICs’ revolution brings the ability to generate, analyse and decipher large amounts of qualitative and quantitative proteomics data and integrate this with data from other related disciplines. This has brought with it multiple opportunities, ranging from diagnostics and

prognostics to rational drug design and targeting. In the context of precision medicine, these all can significantly improve disease outcomes for cancer patients.

Since over 60% of the drugs currently in clinical use target cell surface proteins [141], membrane proteomics will play a prominent role in designing and implementing novel targeted cancer therapies. Drug development has befitted from oncoproteomics. Firstly, it aids the rational design of the drugs themselves. Secondly, it has enabled the discovery of novel drug targets that may prove to be more efficient than those already approved. Thirdly, it can monitor the clinical effects of drugs including efficacy, off-target effects and the development of drug resistance. Finally, it can enable the selection of a suitable drug or drug combinations for use in a personalised medicine contex. A recent review by Frantzi et al. [142] elaborates the different facets of drug development in cancer and other disorders. They make the significant conclusion that proteomics will play an important role in all elements of the drug discovery to therapy spectrum.

The emergence of humanised antibodies against a range of protein targets has been one of the most significant advances in cancer treatment in recent decades. Drugs such as trastuzumab and imatinib [143] for the treatment of breast cancer and bevacizumab for the treatment of a range of solid tumours [144] have become standard therapies. However, their efficacy can be problematic due to the heterogeneous nature of tumours [145] coupled with factors such as genomic background, modifying genes, drug turnover, and environmental differences, requiring a more personalised, targeted approach [146]. Additionally, frequent drug administration can result in toxicity or drug resistance.

Currently, many targeted studies address the mutational status of tumours [147] and drugs are often designed to target particular fusions or mutations. Oncoproteomics affords the significant potential for developing the next generation of targeted therapies, be they humanised antibodies, proteins, peptides/RNA or small molecules. A deeper understanding of the biological effects resulting from targeted therapies [148] or simply understanding the complex biology and hallmarks of cancer [149] is a mainstay of oncoproteomics. Several potential protein-based biomarkers [96] for diagnosis and prognosis have been reported, as well as identification of novel therapeutic targets [66,150].

7. Precision/personalised/P4 medicine: beyond genomics

Precision or personalised or P4 [151] medicine represents a shift from the current “one size fits all” model of disease treatment to a situation where the right drug at the right dose for the right patient at the right time becomes the new paradigm. There has been much debate over the terms “precision” and “personalised”, where concerns were raised that the word “personalised” could be misinterpreted to suggest treatments and preventions are being developed uniquely for individuals. Although this could be the case (e.g. patient-specific therapeutic cancer vaccines), precision medicine has a clear focus on identifying approaches that are effective for patient groups based on genetic, proteomic, metabolomic, lipidomic, microbiome, environmental and lifestyle factors. Pharmacogenomics, a subset of the precision medicine approach, has emerged in which a comprehensive integrative omics strategy combining proteomics, transcriptomics, genomics, and bioinformatics is used to “fingerprint” patients and help characterise their responses to particular drugs [152].

Developments in oncoproteomics will help fuel the transition towards precision medicine. However, while transition medicine is already in practice and gaining momentum [153], unfortunately, there is currently a widespread misunderstanding (among both the general public, governmental and funding bodies) that precision medicine is solely genetics-based [154,155]. Precision medicine should be a team sport where all areas of human biology (e.g., genomics, transcriptomics, proteomics, metabolomics, lipidomics, microbiomics) must be combined to generate a complete picture and to fight complex non-Mendelian gene diseases, like cancer. Genomics per se informs us about the genetic makeup of a human being, but not whether particular genes are actually expressed (transcriptome), in what amount and whether they are functional (proteome). In particular, genomics does not inform us regarding mechanisms like alternative splicing, single amino acid polymorphisms arising from non-synonymous single-nucleotide polymorphisms or posttranslational modifications such as methylation, phosphorylation, glycosylation, ubiquitination, nitrosylation, acetylation, lipidation and proteolysis. The genome is essentially static while the proteome is dynamic, complex, and far larger, with estimates suggesting it comprises as many as 1 million distinct proteoforms.

Importantly proteomics reveals the total protein (proteome) content of any cell or tissue at a specific instance in time in either a health or disease context. Proteins are the functional molecules of cells/tissues as they control the complex biological pathways (interactomes) required for health, and whose dysregulation frequently leads to disease. While human pathologies are encoded both by the genome and environment, they are all produced by measurable changes in the human proteome, and most drug targets are proteins.

To date, cancer has led the field in the introduction of precision medicine. As described above, recent technological advances have assisted this. Several key publications have shown the potential of proteomics, combined with other omics platforms, in supporting precision medicine. Lee Hood (Institute for Systems Biology, Seattle, USA) launched the Pioneer 100 Persons Wellness Project in 2014. This project included whole-genome sequencing, more than 100 clinical tests (using blood, saliva, urine, and stool), metabolomics, proteomics and analysis of the patient's microbiomes at three-time points with daily activity tracking (sleep patterns, pulse, physical activity) [156]. Associations were divided into data communities, which identified biomarkers in context within biological networks. It was extended to the recruitment of 100,000 patients using the services of a commercial entity (Arivale, Seattle, USA). The use of longitudinal data is a key advantage of the precision medicine approach where the individual becomes his baseline rather than the current situation where a supposedly “normal range” is used. We have noted in our studies that the patient data is often far tighter than that of the so-called healthy controls.

In what will almost certainly prove to be a landmark publication, Mike Snyder (Stanford University), one of the first scientists to sequence his own genome (affectionately known as the Snyderome), used a combination of genomic, transcriptomic, proteomic, metabolomic, and autoantibody profiles to monitor himself over a 14 months period (Integrative Personal Omics Profile, iPOP). This was combined with the use of a number of wearable sensors and detailed medical history. This revealed a susceptibility to several potential health issues including coronary artery disease, basal-cell carcinoma, hypertriglyceridemia and Type 2 Diabetes (which occurred during the study) [157]. This single study generated over 30 terabytes of data, indicating the data-intensive nature of such studies (Big Data). This approach has now been extended to the hPOP (human Personal Omics Profile) study, run in collaboration with the Human Proteome Organization (HUPO). In this Human Proteome Project-endorsed initiative, study volunteers at annual international HUPO Congresses, which alternate regionally, are consented and provide blood, urine, and fecal samples at the Congress using strict standard operating procedures (SOPs). These samples will be used for a comprehensive study incorporating whole genome sequencing, transcriptomics, proteomics (including fecal proteomics to measure the microbiome), and metabolomics. Again this study will provide longitudinal data from delegates who regularly attend these meeting as well as enabling ethnicity studies due to the different rotating Congress locations (Central (Europe), Western (Americas) and Eastern (Asia-Oceania)) [155].

The need for multidisciplinary (e.g. chemists, biochemists, physicists, mathematicians, pathologists, clinicians, computer programmers, information technologists, engineers, instrument manufacturers), often multinational, teams to address P4/PM was proposed by Hood [158].

To assist such studies, several organisations and initiatives have arisen. HUPO was established in 2001 to represent and promote proteomics through international cooperation and collaboration to foster the development of new technologies, techniques, education, and training. Their goal is to generate the map of the protein-based molecular architecture of the human body using currently available and emerging technologies, which will become a resource to help elucidate the biological and molecular function and advance diagnosis and treatment of diseases. This was based on three fundamental pillars: Mass Spectrometry, Antibodies, and Knowledgebase (bioinformatics). Data is typically generated using both Chromosome (C-HPP) and Biology and Disease (BD-HPP) approach. In 2018 Pathology was added as a fourth pillar, recognising the translational nature of many of the projects and the need for active interaction with the pathology and clinical communities.

A number of initiatives have stemmed from the US National Cancer Institute (NCI). The Early Detection Research Network (EDRN brings several institutions working primarily in the biomarker space to help accelerate the translation of biomarker information into clinical applications and to evaluate new ways of identifying early-stage cancer and cancer risk. The Clinical Proteomic Tumour Analysis Consortium (CPTAC), launched in 2011, is a US effort to accelerate the understanding of the molecular basis of cancer through the application of large-scale proteome and genome analysis. Projects have included integrated proteogenomic analysis of colorectal, breast and ovarian cancer which have revealed valuable new insights into these cancer types, including identification of proteomic-centric subtypes, prioritisation of driver mutations by correlative analysis of copy number alterations and protein abundance, and understanding the role of posttranslational modifications in cancer-related pathways [159]. Arising from this, as part of the Obama Cancer Moonshot Task Force, the Applied Proteogenomics Organizational Learning and Outcomes (APOLLO) initiative was established in 2016 to bridge oncology research and care effectively. This was achieved by partnering with the nation's two largest health systems, the Department of Defence and Veterans Administration, to analyse DNA, RNA, and protein expression profiles of 8000 annotated human tissue specimens from a wide variety of organ sites.

The International Cancer Proteogenome Consortium (ICPC) was established as part of the Cancer Moonshot to develop a collaborative forum to study common cancers in different populations to enhance precision oncology and share the data collected with scientists and physicians worldwide. This currently comprises 12 countries and 31 distinct institutions of which 11 were already members of the CPTAC effort. In Australia, as part of this initiative, ProCan was established in Sydney for “industrial scale proteomics” with the goal of generating 10,000 cancer proteomes a year from fresh frozen or formalin fixed paraffin embedded material using Accelerated Barocycler Lysis and Extraction (ABLE) technology for sample processing and multiple SWATH MS instruments for high throughput analysis [160].

8. Towards the clinic

Although the first biomarker panel to be discovered by proteomics (OVA1), using SELDI (surface-enhanced laser desorption/ionisation time-of-flight mass spectrometry) technology [161], was clinically approved in 2009, since then remarkably few potential biomarkers identified by proteomics have made it to the clinic. Despite the large number of putative biomarkers reported, target verification and validation are still a significant hurdle for the clinical translation of oncoprotein data [162]. Several factors contribute to this. These include the time lapse for appropriate robust, high throughput, sensitive and specific, multiplexed MS assays to be developed, inappropriate sample size, poorly designed trials, overfitting of data, lack of funding, regulatory hurdles and reimbursement issues. Additionally, most studies have focussed on blood-based measurements where improved methodologies need to be developed (e.g. ultradepletion [163]), improved fractionation) to overcome the extremely high protein dynamic range in plasma that renders it difficult to measure low-level components.

One hurdle that is well recognised by the pathology community, but perhaps less so by the proteomics fraternity, is that of clinical utility. Key questions include (1) What is the clinical management problem and desired outcome (2) Can the proposed biomarker, or biomarker panel contribute to the solution (3) Is there an existing clinical solution (4) Is the proposed biomarker solution feasible in practice? [164]. These concepts should be visited earlier rather than later in the biomarker development pathway, again highlighting the need for clinicians and pathologists as part of the multidisciplinary teams.

Despite these caveats, several proposed biomarkers are heading towards the clinic. This has been facilitated by the development of MRM/immunoMRM/SWATH technology, coupled with robotic sample handling, enabling high throughput, multiplexed validation and resources. These resources include The Human SRM Atlas [165], the JPT Human Proteome Peptide Catalog (which comprises more than 400,000 light and heavy proteotypic peptides covering 19,840 proteins predicted by the human genome and confirmed by more than 10,000,000 validated MS spectra available from the ProteomicsDB database) (http://www.jpt.com), and The Human Protein Atlas [166].

One multiplexed cancer-related MRM/SRM assay has now entered the clinic. Xpresys Lung (Integrated Diagnostics, Seattle) is a blood-based test in which eleven proteins are assayed to determine the malignant potential of lung nodules. This was based on proteomics study [167] in which a systems biology strategy had been used to identify 371 protein candidates, and MRM assays developed for each. Interestingly, pathway analysis identified several transcriptional regulators (NF2L2, AHR, MYC, and FOS) associated with lung cancer, lung inflammation, and oxidative stress networks.

Several recent proteomics publications have targeted pancreatic cancer, which has the worst survival rate of all the 22 common cancers, and where biomarkers for early detection are urgently required [168,169]. Many of these studies have presented orthogonal validation to support their findings. Thus, Hu et al. [170] followed up on a previous proteomics study that had revealed that calcium-activated chloride channel regulator 1 (CLCA1) was a potential tumour suppressor in pancreatic cancer, with expression inversely correlated with patient survival. CLCA1 expression was then evaluated using tissue microarrays and IHC in a separate cohort of 140 patients with pancreatic ductal adenocarcinoma showing low CLCA1 expression was an independent predictive factor of poor disease-free survival. In another study on intraductal pancreatic mucinous neoplasms (IPMN) [171], MALDI imaging was used in combination with nanoLC-ESI-MS/MS. Differentially expressed proteins were validated in an independent IPMN cohort by IHC on surgical specimens using tissue-microarrays and endoscopic ultrasound fine-needle aspirates.

In a study addressing CRC, another of the top 4 deadliest cancers, a combination of laser capture microdissection (LCM), iTRAQ labelling and 2D LC-MS/MS was used to compare stromal proteomes in different stages of CRC to assess the effect of the tumour microenvironment. Differentially expressed proteins were subjected to cluster and pathway analyses. The differential expression of Tenascin-C and S100A9 were validated using IHC. It was found that expression levels of S100A9 and Tenascin-C correlated with tumour stage and metastatic potential. Additionally, Tenascin-C was found to be highly expressed in metastatic lymph nodes. Taken together, these data clearly show that the “valley of death” between discovery and translation is being overcome, and rapid advances are now being made in progressing proteogenomic data towards the clinic as an adjunct to precision medicine.

9. Future perspectives

Significant technological advancements have been made in proteomics, especially oncoproteomics, over the past decade, single cell

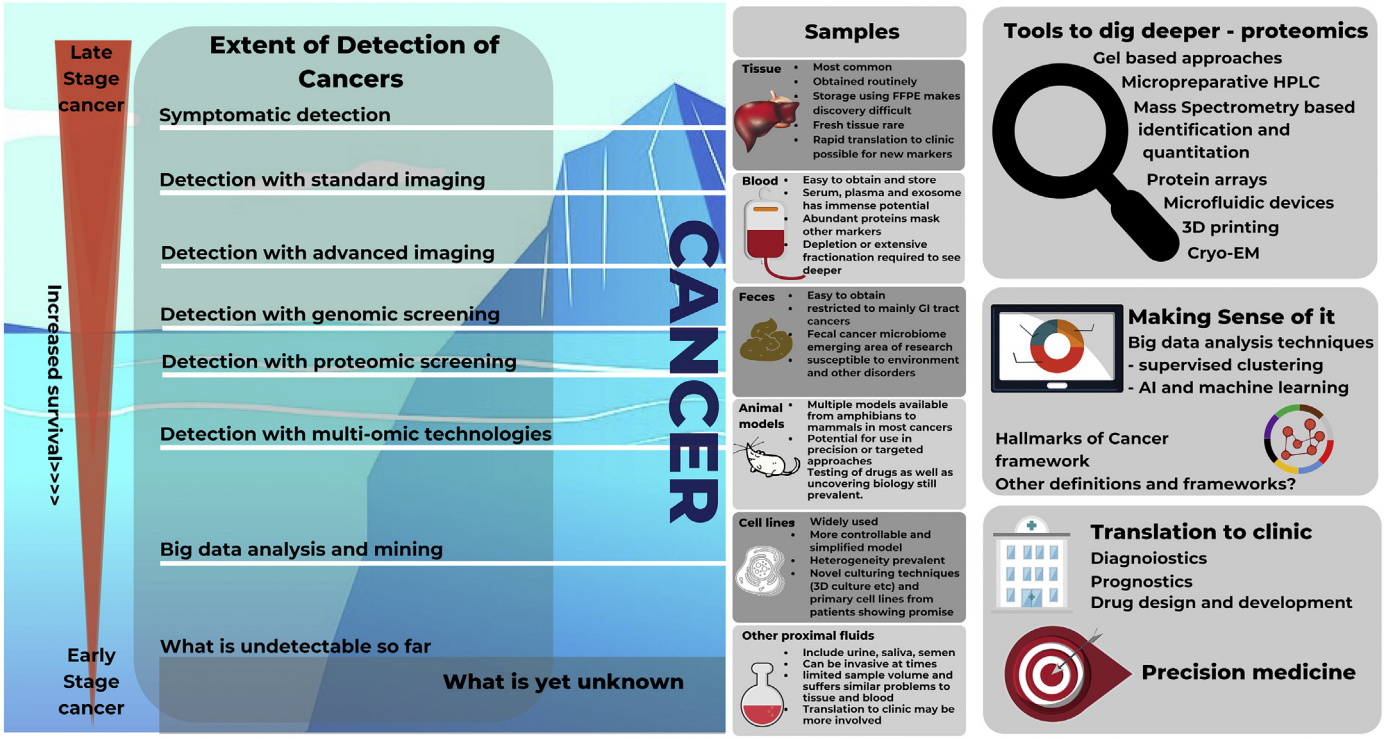

Fig. 3. Oncoproteomics and precision medicine: the tip of the iceberg.
Our current understanding of cancer biology can be represented by the tip of an iceberg. Currently, most cancers are only detected when they are already advanced. Genomic and proteomic screening (using a range of proteomics tools (top right)) on a variety of samples (middle) will enable deeper mining of the proteome (left) leading to an increased understanding of cancer biology and the Hallmarks of Cancer. Significant computing power, coupled with machine learning and AI will be required to address the size and complexity of the data generated (Big Data), leading to improved treatment strategies (Precision medicine). An enlarged version of this figure is available (Supplementary Fig. 1).

analysis is becoming a reality, and our understanding of the complex biology underlying the hallmarks of cancer and disease heterogeneity is rapidly expanding. However, we are still only seeing "the tip of the iceberg" in terms of clinical application and precision medicine (Fig. 3). Although several potential biomarkers and biomarker panels have been and will continue to be, identified, the approval of these more widely in the clinical setting has been slow. The use of techniques such as SRM and SWATH now enables rapid validation in a multiplexed manner with appropriate sample numbers. Improved validation methodology will result in a drastic increase in the number of approved assays entering the clinic. Interactomics will continue to play an important role, especially in further understanding cancer biology and identifying novel biomarkers and drug targets. MS and bioinformatics, coupled with techniques like CyroEM, will populate this vital area. A meta-analysis of large cancer data sets and data-driven in silico studies [172] will further increase our understanding of cancer biology, while predictive modelling of drug treatment [173] will facilitate individualised prediction of patient responses to treatment. This will facilitate drug development, which will in turn support personalised medicine, addressing both health and disease. Undoubtedly a significant hurdle will be that of the Big Data, arising from not only the large body of oncoproteogenomic data, that will be generated, but also large heterogeneous datasets from other resources such as electronic health records and data obtained from smartphone applications and personal monitoring devices (Fig. 1), the so called "Health Avatar" [174]. The use of increasingly powerful supercomputers coupled with artificial intelligence and machine learning methods will be fundamental to overcoming this problem [175] and will identify trends and associations that are not obvious in smaller data sets (Fig. 3). New methods specifically for optimising data collection, storage, cleaning, processing, and interpretation will continue to be developed [176]. The need for multidisciplinary teams has been recognised. This will lead to a heightened need for collaborative networks sharing both their data and expertise, with the uptake of standardised SOPs and methodologies. Bodies such as CPTAC and HUPO will be essential to oversee this and ensure international agreement.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cca.2019.06.006.

References

[1] R. Yuan, X. Zhu, G. Wang, et al., Cancer as a robust intrinsic state shaped by evolution: a key issues review, Rep. Prog. Phys. 80 (2017).
[2] C.A.K. Borrebaeck, Precision diagnostics: moving towards protein biomarker signatures of clinical utility in cancer, Nat. Rev. Cancer 17 (2017) 199–204.
[3] Y. Zhou, G.A. Abel, W. Hamilton, et al., Diagnosis of cancer as an emergency: a critical review of current evidence, Nat. Rev. Clin. Oncol. 14 (2017) 45–56.
[4] J.M. Lee, E.C. Kohn, Proteomics as a guiding tool for more effective personalized therapy, Ann. Oncol. 21 (2010) 205–210.
[5] P. Jin, J. Lan, K. Wang, et al., Pathology, proteomics and the pathway to personalised medicine, Expert Rev. Proteomics. 15 (2018) 231–243.
[6] W. Hamilton, F.M. Walter, G. Rubin, et al., Improving early diagnosis of symptomatic cancer, Nat. Rev. Clin. Oncol. 13 (2016) 740–749.
[7] H.D. Shukla, J. Mahmood, Z. Vujaskovic, Integrated proteo-genomic approach for early diagnosis and prognosis of cancer, Cancer Lett. 369 (2015) 28–36.
[8] L. Zhou, Q. Li, J. Wang, et al., Oncoproteomics: trials and tribulations, Proteomics Clin. Appl. 10 (2016) 516–531.
[9] J.W. Lau, E. Lehnert, A. Sethi, et al., The cancer genomics cloud: collaborative, reproducible, and democratized - a new paradigm in large-scale computational research, Cancer Res. 77 (21) (2017) e3–e6.
[10] S. Yi, S. Lin, Y. Li, et al., Functional variomics and network perturbation: connecting genotype to phenotype in cancer, Nat. Rev. Genet. 18 (2017) 395–410.
[11] J.A. Alfaro, A. Sinha, T. Kislinger, et al., Onco-proteogenomics: Cancer proteomics joins forces with genomics, Nat. Methods 11 (2014) 1107–1113.
[12] S. Roychowdhury, A.M. Chinnaiyan, Translating cancer genomes and transcriptomes for precision oncology, CA Cancer J. Clin. 66 (2015) 75–88.
[13] M.A. Gillette, K.R. Clauser, P. Wang, et al., Proteogenomics connects somatic mutations to signalling in breast cancer, Nature. (2016) 1–19.
[14] S. Pan, T.A. Brentnall, K. Kelly, et al., Tissue proteomics in pancreatic cancer study: discovery, emerging technologies, and challenges, Proteomics. 13 (3–4) (2013) 710–721.

[15] K.K. Jain, Innovations, challenges and future prospects of oncoproteomics, Mol. Oncol. 2 (2008) 153–160.
[16] W.C.S. Cho, Contribution of oncoproteomics to cancer biomarker discovery, Mol. Cancer 6 (2007) 1–13.
[17] N. Zwakman, R. van de Laar, T. van Gorp, et al., Perioperative changes in serum CA125 levels: a prognostic factor for disease-specific survival in patients with ovarian cancer, J. Gynecol. Oncol. 28 (2017) 1–12.
[18] D. Dong, L. Jia, L. Zhang, et al., Periostin and CA242 as potential diagnostic serum biomarkers complementing CA19.9 in detecting pancreatic cancer, Cancer Sci. 109 (2018) 2841–2851.
[19] W. Lin, D. Cao, K. Shen, Prognostic significance of preoperative serum Cea in primary mucinous ovarian carcinoma: a retrospective cohort study, Cancer Manag. Res. 10 (2018) 6913–6920.
[20] G. De Angelis, H.G. Rittenhouse, S.D. Mikolajczyk, et al., Twenty years of PSA: from prostate antigen to tumor marker, Rev. Urol. 9 (3) (2007) 113–123.
[21] A.K. Füzéry, J. Levin, M.M. Chan, et al., Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges, Clin.Proteomics. 10 (1) (2013) 13.
[22] C. Panis, L. Pizzatti, G.F. Souza, et al., Clinical proteomics in cancer: where we are, Cancer Lett. 382 (2016) 231–239.
[23] K.E. Hung, K.H. Yu, Proteomic approaches to Cancer biomarkers, Gastroenterology. 138 (2010) 46–51.e1.
[24] D. Hanahan, R.A. Weinberg, The hallmarks of cancer, Cell. 100 (1) (2000) 57–70.
[25] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, Cell. 144 (2011) 646–674.
[26] R.J. Lipscombe, K.-C. Tan, R.A. Syme, et al., Analysis of reproducibility of proteome coverage and quantitation using isobaric mass tags (iTRAQ and TMT), J. Proteome Res. 16 (2016) 384–392.
[27] P.E. Geyer, L.M. Holdt, D. Teupser, et al., Revisiting biomarker discovery by plasma proteomics, Mol. Syst. Biol. 13 (9) (2017) 942.
[28] L. Peng, D.I. Cantor, C. Huang, et al., Tissue and plasma proteomics for early stage cancer detection, Mol. Omi. 14 (2018) 405–423.
[29] D. Baiwir, G. Mazzucchelli, M. Fléron, et al., Tissue proteomics for the next decade? Towards a molecular dimension in histology, Omi. A J. Integr. Biol. 18 (2014) 539–552.
[30] P. Shah, B. Zhang, C. Choi, et al., Tissue proteomics using chemical immobilization and mass spectrometry, Anal. Biochem. 469 (2015) 27–33.
[31] V. Broeckx, L. Peeters, E. Maes, et al., Formalin-fixed paraffin-embedded tissue: the holy grail of clinical proteomics, Proteomics Clin. Appl. 8 (2014) 735–736.
[32] M.T.B. Alvarez, D.J. Shah, C.D. Thulin, et al., Tissue proteomics of the low-molecular weight proteome using an integrated cLC-ESI-QTOFMS approach, Proteomics. 13 (2013) 1400–1411.
[33] Z.R. Gregorich, Y.-H. Chang, L. Hwang, et al., New mass-spectrometry-compatible degradable surfactant for tissue proteomics, J. Proteome Res. 14 (2015) 1587–1599.
[34] S. Meding, B. Balluff, M. Elsner, et al., Tissue-based proteomics reveals FXYD3, S100A11 and GSTM3 as novel markers for regional lymph node metastasis in colon cancer, J. Pathol. 228 (2012) 459–470.
[35] Y. Zhang, B.R. Fonslow, B. Shan, et al., Protein analysis by shotgun/bottom-up proteomics, Chem. Rev. 113 (4) (2013) 2343–2394.
[36] K. Wang, C. Huang, E.C. Nice, Proteomics, genomics and transcriptomics: their emerging roles in the discovery and validation of colorectal cancer biomarkers, Expert Rev. Proteomics. 11 (2014) 179–205.
[37] T. Bennike, S. Birkelund, A. Stensballe, et al., Biomarkers in inflammatory bowel diseases: current status and proteomics identification strategies, World J. Gastroenterol. 20 (2014) 3231–3244.
[38] C.S. Ang, J. Rothacker, H. Patsiouras, et al., Use of multiple reaction monitoring for multiplex analysis of colorectal cancer-associated proteins in human feces, Electrophoresis. 32 (2011) 1926–1938.
[39] L.J.W. Bosch, M. De Wit, T.V. Pham, et al., Novel stool-based protein biomarkers for improved colorectal cancer screening, Ann. Intern. Med. 167 (12) (2017) 855–866.
[40] Z.Y. Kho, S.K. Lal, The human gut microbiome - a potential controller of wellness and disease, Front. Microbiol. 9 (2018) 1835.
[41] H. Lin, Q.Y. He, L. Shi, et al., Proteomics and the microbiome: pitfalls and potential, Expert Rev. Proteomics. 16 (6) (2019) 501–511.
[42] R. Li, R. Zhou, H. Wang, et al., Gut microbiota-stimulated cathepsin K secretion mediates TLR4-dependent M2 macrophage polarization and promotes tumor metastasis in colorectal cancer, Cell Death Differ. (2019) 1.
[43] Y. Wu, L. Shi, Q. Li, et al., Microbiota diversity in human colorectal Cancer tissues is associated with Clinicopathological features, Nutr. Cancer 0 (2019) 1–9.
[44] P. Álvarez-Chaver, O. Otero-Estévez, M.P. de la Cadena, et al., Proteomics for discovery of candidate colorectal cancer biomarkers, World J. Gastroenterol. 20 (2014) 3804–3824.
[45] J.S. Lichtman, A. Marcobal, J.L. Sonnenburg, et al., Host-centric proteomics of stool: a novel strategy focused on intestinal responses to the gut microbiota, Mol. Cell. Proteomics 12 (2013) 3310–3318.
[46] E. Bendixen, Animal models for translational proteomics, PROTEOMICS - Clin. Appl. 8 (2014) 637–639.
[47] L.B. Schook, T.M. Fan, K.M. Schachtschneider, et al., Of mice, dogs, pigs, and men: choosing the appropriate model for Immuno-oncology research, ILAR J. (2018 Nov 22), https://doi.org/10.1093/ilar/ily014 (Epub ahead of print).
[48] P. Letrado, I. De Miguel, I. Lamberto, et al., Zebrafish: speeding up the cancer drug discovery process, Cancer Res. 78 (2018) 6048–6058.
[49] M.G. Terp, H.J. Ditzel, Application of proteomics in the study of rodent models of cancer, Proteomics Clin. Appl. 8 (2014) 640–652.
[50] A. Zarei, V. Razban, S.E. Hosseini, et al., Creating cell and animal models of human disease by genome editing using CRISPR/Cas9, J. Gene Med. 21 (4) (2019) e3082.
[51] O. Méndez, J. Villanueva, Challenges and opportunities for cell line secretomes in cancer proteomics, Proteomics Clin. Appl. 9 (2015) 348–357.
[52] M. Costanzo, A. Cevenini, E. Marchese, et al., Label-free quantitative proteomics in a Methylmalonyl-CoA mutase-silenced neuroblastoma cell line, Int. J. Mol. Sci. 19 (2018).
[53] Y. Liu, Y. Mi, T. Mueller, et al., Multi-omic measurements of heterogeneity in HeLa cells across laboratories, Nat. Biotechnol. 37 (3) (2019) 314–322.
[54] D. Stetson, A. Ahmed, X. Xu, et al., Orthogonal comparison of four plasma NGS tests with tumor suggests technical factors are a major source of assay discordance, JCO Precis. Oncol. (2019), https://doi.org/10.1200/PO.18.00191.
[55] A.A.A. Al-Juboori, A. Ghosh, M.F.B. Jamaluddin, et al., Proteomic analysis of stromal and epithelial cell communications in human endometrial cancer using a unique 3D co-culture model, Proteomics. (2019) e1800448, , https://doi.org/10.1002/pmic.201800448 [Epub ahead of print].
[56] C. Zhang, W. Leng, C. Sun, et al., Urine proteome profiling predicts lung Cancer from control cases and other tumors, EBioMedicine. 30 (2018) 120–128.
[57] I. Belczacka, M. Pejchinovski, M. Krochmal, et al., Urinary Glycopeptide analysis for the investigation of novel biomarkers, Proteomics Clin. Appl. (2019) 1800111, , https://doi.org/10.1002/prca.201800111 (Epub 2018 Oct 18).
[58] P.N. Teng, N.W. Bateman, B.L. Hood, et al., Advances in proximal fluid proteomics for disease biomarker discovery, J. Proteome Res. 9 (2010) 6091–6100.
[59] S.M. Hanash, S.J. Pitteri, V.M. Faca, Mining the plasma proteome for cancer biomarkers, Nature. 452 (2008) 571–579.
[60] E.C. Nice, J. Rothacker, J. Weinstock, et al., Use of multidimensional separation protocols for the purification of trace components in complex biological samples for proteomics analysis, J. Chromatogr. A 1168 (2007) 190–210.
[61] R. Westermeier, 2D gel-based proteomics: there's life in the old dog yet, Arch. Physiol. Biochem. 122 (5) (2016) 236–237.
[62] F. Ning, X. Wu, W. Wang, Exploiting the potential of 2DE in proteomics analyses, Expert Rev. Proteomics. 13 (10) (2016) 901–903.
[63] T. Kondo, Cancer biomarker development and two-dimensional difference gel electrophoresis (2D-DIGE), Biochim. Biophys. Acta, Proteins Proteomics 1867 (2019) 2–8.
[64] F.G.G. Vercauteren, L. Arckens, R. Quirion, Applications and current challenges of proteomic approaches, focusing on two-dimensional electrophoresis, Amino Acids 33 (2007) 405–414.
[65] L. Peng, J.P. Keating, R. Miller, et al., Combination of laser microdissection, 2D-DIGE and MALDI-TOF MS to identify protein biomarkers to predict colorectal cancer spread, Clin. Proteomics 16 (2019) 1–13.
[66] D. Birch, R. Frank, M.S. Baker, et al., Characterization of the interaction between heterodimeric αvβ6 integrin and Urokinase plasminogen activator receptor (uPAR) using functional proteomics, J. Proteome Res. 13 (2014) 5956–5964.
[67] S.B. Ahn, C. Chan, O.F. Dent, et al., Epithelial and stromal cell urokinase plasminogen activator receptor expression differentially correlates with survival in rectal cancer stages B and C patients, PLoS One 10 (2015) 1–19.
[68] J.M. Chick, P.A. Haynes, M.P. Molloy, et al., Characterization of the rat liver membrane proteome using peptide immobilized pH gradient isoelectric focusing, J. Proteome Res. 7 (3) (2008) 1036–1045.
[69] L.R. McQuade, U. Schmidt, D. Pascovici, et al., Improved membrane proteomics coverage of human embryonic stem cells by peptide IPG-IEF, J. Proteome Res. 8 (12) (2009) 5642–5649.
[70] W. Zhou, L.A. Liotta, E.F. Petricoin, Cancer metabolism and mass spectrometry-based proteomics, Cancer Lett. 356 (2015) 176–183.
[71] C. Tu, P.A. Rudnick, M.Y. Martinez, et al., Depletion of abundant plasma proteins and limitations of plasma proteomics, J. Proteome Res. 9 (10) (2010) 4982–4991.
[72] S.H. Tan, A. Mohamedali, A. Kapur, et al., Ultradepletion of human plasma using chicken antibodies: A proof of concept study, J. Proteome Res. 12 (6) (2013) 2399–2413.
[73] K. Tanaka, H. Waki, Y. Ido, 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 (1988) 151–153.
[74] J.B. Fenn, M. Mann, C.K. Meng, et al., Electrospray ionization for mass spectrometry of large biomolecules, Science. 246 (1989) 64–71.
[75] J.R. Yates, C.I. Ruse, A. Nakorchevsky, Proteomics by Mass Spectrometry: Approaches, Advances, and Applications, Annu. Rev. Biomed. Eng. 11 (2009) 49–79.
[76] P. Lössl, M. van de Waterbeemd, A.J. Heck, The diverse and expanding role of mass spectrometry in structural and molecular biology, EMBO J. 35 (24) (2016) 2634–2657.
[77] R. Matthiesen, J. Bunkenborg, Introduction to mass spectrometry-based proteomics, Methods Mol. Biol. 1007 (2013) 1–45.
[78] Z.R. Gregorich, Y. Ge, Top-down proteomics in health and disease: challenges and opportunities, Proteomics. 14 (10) (2014) 1195–1210.
[79] A. Smith, M. Galli, I. Piga, et al., Molecular signatures of medullary thyroid carcinoma by matrix-assisted laser desorption/ionisation mass spectrometry imaging, J. Proteome 191 (2019) 114–123.
[80] O. Chahrour, D. Cobice, J. Malone, Stable isotope labelling methods in mass spectrometry-based quantitative proteomics, J. Pharm. Biomed. Anal. 113 (2015) 2–20.
[81] B. Wang, G. Hom, S. Zhou, et al., The oxidized thiol proteome in aging and cataractous mouse and human lens revealed by ICAT labeling, Aging Cell 16 (2017) 244–261.
[82] J. Ayté, H. Molina, A. Domènech, et al., Monitoring in vivo reversible cysteine oxidation in proteins using ICAT and mass spectrometry, Nat. Protoc. 9 (2014)

1131–1145.
[83] P.L. Ross, Y.N. Huang, J.N. Marchese, et al., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents, Mol. Cell. Proteomics 3 (12) (2004) 1154–1169.
[84] S.Q. Sun, K. Ye, S. Li, et al., Proteogenomic integration reveals therapeutic targets in breast cancer xenografts, Nat. Commun. 8 (2017) 14864.
[85] R. Moulder, S.D. Bhosale, D.R. Goodlett, et al., Analysis of the plasma proteome using iTRAQ and TMT-based isobaric labelling, Mass Spectrom. Rev. 37 (2018) 583–606.
[86] W. Gao, H. Wang, W. Hu, et al., Nuclear lactate dehydrogenase a senses ROS to produce α-hydroxybutyrate for HPV-induced cervical tumor growth, Nat. Commun. 9 (1) (2018) 4429.
[87] Y. Ren, Y. He, Z. Lin, et al., Reagents for isobaric labeling peptides in quantitative proteomics, Anal. Chem. 90 (2018) 12366–12371.
[88] J. Waldbauer, L. Zhang, A. Rizzo, et al., diDO-IPTL: a peptide-labeling strategy for precision quantitative proteomics, Anal. Chem. 89 (2017) 11498–11504.
[89] M. Krüger, M. Moser, S. Ussar, et al., SILAC mouse for quantitative proteomics uncovers Kindlin-3 as an essential factor for red blood cell function, Cell. 134 (2008) 353–364.
[90] Y.-L. Chen, W.-L. Wu, C.-W. Jang, et al., Interferon-stimulated gene 15 modulates cell migration by interacting with Rac1 and contributes to lymph node metastasis of oral squamous cell carcinoma cells, Oncogene. 38 (23) (2019) 4480–4495.
[91] P. Rudnick, D. Fenyö, S. Awasthi, et al., Quantitative mass spectrometry to interrogate proteomic heterogeneity in metastatic lung adenocarcinoma and validate a novel somatic mutation CDK12–G879V, Mol. Cell. Proteomics. 18 (4) (2019) 622–641.
[92] P. Rudnick, D. Fenyö, S. Awasthi, et al., Quantitative mass spectrometry to interrogate proteomic heterogeneity in metastatic lung adenocarcinoma and validate a novel somatic mutation CDK12-G879V, Mol. Cell. Proteomics (2019).
[93] J. Ye, L. Guo, Y. Zhang, et al., SWATH enables precise label-free quantification on proteome scale, Proteomics. 15 (2015) 1215–1223.
[94] J. Doblmann, F. Dusberger, R. Imre, et al., ApQuant: accurate label-free quantification by quality filtering, J. Proteome Res. 18 (2019) 535–541.
[95] L.J.E. Goeminne, K. Gevaert, L. Clement, Experimental design and data-analysis in label-free quantitative LC/MS proteomics: a tutorial with MSqRob, J. Proteome 171 (2018) 23–36.
[96] V.A. Hristova, D.W. Chan, Cancer biomarker discovery and translation: proteomics and beyond, Expert Rev. Proteomics. 16 (2019) 93–103.
[97] X. Domingo-Almenara, A. Nordström, A. Thomas, et al., XCMS-MRM and METLIN-MRM: a cloud library and public resource for targeted analysis of small molecules, Nat. Methods 15 (2018) 681–684.
[98] L. Krasny, P. Bland, N. Kogata, et al., SWATH mass spectrometry as a tool for quantitative profiling of the matrisome, J. Proteome 189 (2018) 11–22.
[99] B.C. Collins, C.L. Hunter, Y. Liu, et al., Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry, Nat. Commun. 8 (1) (2017).
[100] E. Govaert, K. Van Steendam, S. Willems, et al., Comparison of fractionation proteomics for local SWATH library building, Proteomics. 17 (15–16) (2017) 291, https://doi.org/10.1038/s41467-017-00249-5.
[101] Y. Gao, X. Wang, Z. Sang, et al., Quantitative proteomics by SWATH-MS reveals sophisticated metabolic reprogramming in hepatocellular carcinoma tissues, Sci. Rep. 7 (2017) 45913.
[102] A.P. Bouchal, O.T. Schubert, J. Faktor, et al., Breast cancer classification based on proteotypes obtained by SWATH mass spectrometry, bioRxiv, 2019, https://doi.org/10.1101/583443.
[103] M.C. Chambers, D.C. Liebler, D.L. Tabb, et al., proBAMsuite, a bioinformatics framework for genome-based representation and analysis of proteomics data, Mol. Cell. Proteomics 15 (2015) 1164–1175.
[104] Z. Noor, J.X. Wu, D. Pascovici, et al., ISwathX: an interactive web-based application for extension of DIA peptide reference libraries, Bioinformatics. 35 (3) (2019) 538–539.
[105] P.D. Verhaert, M. Connor, S. Ranganathan, et al., Accelerating the search for the missing proteins in the human proteome, Nat. Commun. 8 (2017) 14271.
[106] W.W. Bin Goh, L. Wong, Advanced bioinformatics methods for practical applications in proteomics, Brief. Bioinform. 20 (2019) 347–355.
[107] J. Lee, M. Magee, J. Labaer, et al., Emerging Protein Array Technologies for Proteomics, vol. 10, (2016), pp. 65–75.
[108] D. Garrido-Martín, C. Droste, P. Nilsson, et al., Screening and validation of novel biomarkers in Osteoarticular pathologies by comprehensive combination of protein Array technologies, J. Proteome Res. 16 (2017) 1890–1899.
[109] M. Masuda, T. Yamada, Signaling pathway profiling using reverse-phase protein array and its clinical applications, Expert Rev. Proteomics. 14 (2017) 607–615.
[110] J.W. Huh, H.C. Kim, I. Sohn, et al., Serum protein profiling using an aptamer array predicts clinical outcomes of stage IIA colon cancer: a leave-one-out cross-validation, Oncotarget. 7 (13) (2016) 16338–16348.
[111] R.D. Pedde, H. Li, C.H. Borchers, et al., Microfluidic-mass spectrometry interfaces for translational proteomics, Trends Biotechnol. 35 (10) (2017) 954–970.
[112] J.C. Contreras-Naranjo, H.J. Wu, V.M. Ugaz, Microfluidics for exosome isolation and analysis: enabling liquid biopsy for personalized medicine, Lab Chip 17 (21) (2017) 3558–3577.
[113] P. Zhang, X. Zhou, M. He, et al., Ultrasensitive detection of circulating exosomes with a 3D-nanopatterned microfluidic chip, Nat. Biomed. Eng. (2019), https://doi.org/10.1038/s41551-019-0356-9 [Epub ahead of print].
[114] C. Florkowski, A. Don-Wauchope, N. Gimenez, et al., Point-of-care testing (POCT) and evidence-based laboratory medicine (EBLM)-does it leverage any advantage in clinical decision making? Crit. Rev. Clin. Lab. Sci. 54 (7–8) (2017) 471–494.
[115] F. Li, N.P. MacDonald, R.M. Guijt, et al., Multimaterial 3D printed fluidic device for measuring Pharmaceuticals in Biological Fluids, Anal. Chem. 91 (2019) 1758–1763.
[116] A.D. Quinn, D. Dixon, B.J. Meenan, Barriers to hospital-based clinical adoption of point-of-care testing (POCT): a systematic narrative review, Crit. Rev. Clin. Lab. Sci. 53 (2016) 1–12.
[117] G.J. Labonia, K.R. Ludwig, C.B. Mousseau, et al., ITRAQ quantitative proteomic profiling and MALDI-MSI of Colon Cancer spheroids treated with combination chemotherapies in a 3D printed fluidic device, Anal. Chem. 90 (2) (2018) 1423–1430.
[118] H.N. Chan, M.J.A. Tan, H. Wu, Point-of-care testing: applications of 3D printing, Lab Chip 17 (2017) 2713–2739.
[119] R. Amin, S. Knowlton, A. Hart, et al., 3D-printed microfluidic devices, Biofabrication. 8 (2016) 1–16.
[120] C. Zhang, T. Glaros, N.E. Manicke, Targeted protein detection using an all-in-one mass spectrometry cartridge, J. Am. Chem. Soc. 139 (32) (2017) 10996–10999.
[121] M.A. Herzik, M. Wu, G.C. Lander, Achieving better-than-3-Å resolution by single-particle cryo-EM at 200 keV, Nat. Methods 14 (11) (2017) 1075–1078.
[122] S. Pal, K. Ganesan, S. Eswaran, Chemical crosslinking-mass spectrometry (CXL-MS) for proteomics, antibody-drug conjugates (ADCs) and Cryo-Electron microscopy (cryo-EM), IUBMB Life 70 (10) (2018) 947–960.
[123] I.V. Hinkson, T.M. Davidsen, J.D. Klemm, et al., Corrigendum: A Comprehensive Infrastructure for Big Data in Cancer Research: Accelerating Cancer Research and Precision Medicine, 5 (2017), p. 3389.
[124] B. Chapman, A. Livk, A. Hunter, et al., Classification of fish samples via an integrated proteomics and bioinformatics approach, Proteomics. 13 (2013) 3124–3130.
[125] S. Bencharit, M.B. Border, Where are we in the world of proteomics and bioinformatics? Expert Rev. Proteomics. 9 (2012) 489–491.
[126] M.P. van Ooijen, V.L. Jong, M.J.C. Eijkemans, et al., Identification of differentially expressed peptides in high-throughput proteomics data, Brief. Bioinform. 19 (5) (2018) 971–981.
[127] A. Cambiaghi, M. Ferrario, M. Masseroli, Analysis of metabolomic data: tools, current strategies and future challenges for omics data integration, Brief. Bioinform. 18 (3) (2017) 498–510.
[128] L. Martens, J.A. Vizcaíno, Opinion a Golden age for working with public proteomics data, Trends Biochem. Sci. 42 (5) (2017) 333–341.
[129] D. Grapov, J. Fahrmann, K. Wanichthanarak, et al., Rise of deep learning for genomic, proteomic, and metabolomic data integration in precision medicine, Omi. A J. Integr. Biol. 22 (10) (2018) 630–636.
[130] M.E. Sardiu, J.M. Gilmore, B.D. Groppe, et al., Topological scoring of protein interaction networks, Nat. Commun. 10 (1) (2019) 1118.
[131] I. Angelidis, L.M. Simon, I.E. Fernandez, et al., An atlas of the aging lung mapped by single cell transcriptomics and deep tissue proteomics, Nat. Commun. 10 (1) (2019) 963.
[132] L. Lian, X. Li, M. Xu, et al., VEGFR2 promotes tumorigenesis and metastasis in a pro-angiogenic- independent way in gastric cancer, BMC Cancer 19 (1) (2019) 183.
[133] F.J. Velloso, A.R. Campos, M.C. Sogayar, et al., Proteome profiling of triple negative breast cancer cells overexpressing NOD1 and NOD2 receptors unveils molecular signatures of malignant cell proliferation, BMC Genomics 20 (1) (2019) 152.
[134] A. Krämer, J. Green, J. Pollard, et al., Causal analysis approaches in ingenuity pathway analysis, Bioinformatics. 30 (4) (2014) 523–530.
[135] P. Shannon, A. Markiel, O. Ozier, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (11) (2003) 2498–2504.
[136] D. Szklarczyk, A. Franceschini, S. Wyder, et al., STRING v10: protein-protein interaction networks, integrated over the tree of life, Nucleic Acids Res. 43 (2015) D447–D452.
[137] M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (1) (2000) 27–30.
[138] T. Helen, B. Gomig, I. João, et al., Quantitative label-free mass spectrometry using contralateral and adjacent breast tissues reveal differentially expressed proteins and their predicted impacts on pathways and cellular functions in breast cancer, J. Proteome 199 (2019) 1–14.
[139] M.J. Watt, A.K. Clark, L.A. Selth, et al., Suppressing fatty acid uptake has therapeutic effects in preclinical models of prostate cancer, Sci. Transl. Med. 11 (478) (2019).
[140] S. Baker, I. Ali, I. Silins, et al., Data and text mining Cancer hallmarks analytics tool (CHAT): a text mining approach to organize and evaluate scientific literature on cancer, Bioinformatics. 33 (24) (2017) 3973–3981.
[141] C. Finan, A. Gaulton, F.A. Kruger, et al., The Druggable Genome and Support for Target Identification and Validation in Drug Development, vol. 1166, (2017), pp. 1–16.
[142] M. Frantzi, A. Latosinska, H. Mischak, Proteomics in Drug Development: The Dawn of a New Era?, Clinical Applications. (2019) e1800087.
[143] I. Kümler, M.K. Tuxen, D.L. Nielsen, A systematic review of dual targeting in HER2-positive breast cancer, Cancer Treat. Rev. 40 (2) (2014) 259–270.
[144] G. Roviello, T. Bachelot, C.A. Hudis, et al., ScienceDirect the role of bevacizumab in solid tumours: a literature based meta-analysis of randomised trials, Eur. J. Cancer 75 (2017) 245–258.
[145] A. Afghahi, G.W. Sledge, Targeted therapy for Cancer in the genomic era, Cancer J. 21 (2015) 294–298.
[146] P. Lewis, T.A. Yap, Optimizing precision oncology and immunotherapy strategies: moving into the next stage of Cancer Medicine, J. Immunother. Precis. Oncol.

(2019) 4–6.
[147] S.D. Richman, B. Jasani, Predictive biomarkers and targeted therapies in colorectal Cancer, Predict. Biomarkers Oncol. (2019) 423–430.
[148] C.-Y. Liu, W.-L. Wang, H.-P. Yang, et al., Varlitinib downregulates HER/ERK signaling and induces apoptosis in triple negative breast Cancer cells, Cancers (Basel). 11 (2019) 105.
[149] H.R. De Boer, M. Pool, E. Joosten, et al., Quantitative proteomics analysis identi fi es MUC1 as an effect sensor of EGFR inhibition, Oncogene. 38 (9) (2019) 1477–1488.
[150] R.G. Saldanha, M.P. Molloy, K. Bdeir, et al., Proteomic identification of lynchpin Urokinase plasminogen activator receptor protein interactions associated with epithelial Cancer malignancy research articles, J. Proteome Res. (2007) 1016–1028.
[151] G.S. Ginsburg, J.J. Mccarthy, Personalized medicine: revolutionizing drug discovery and patient care, Trends Biotechnol. 19 (2001) 491–496.
[152] D. Fenyö, G. Menschaert, PROTEOGENOMICS FROM A BIOINFORMATICS ANGLE: A GROWING FIELD, Mass Spectrom. Rev. 36 (5) (2017) 584–599.
[153] G.S. Ginsburg, P. Medicine, K.A. Phillips, et al., Precision Medicine: from science to value, Health Aff. 37 (2018) 694–701.
[154] R.W. Peck, Precision Medicine is not just genomics: the right dose for every patient, Annu. Rev. Pharmacol. Toxicol. 58 (2018) 105–122.
[155] E.C. Nice, Challenges for omics technologies in the implementation of personalized medicine, Expert Rev. Precis. Med. Drug Dev. 3 (4) (2018) 229–231.
[156] N.D. Price, A.T. Magis, J.C. Earls, et al., A wellness study of 108 individuals using personal, dense, dynamic data clouds, Nat. Biotechnol. 35 (8) (2017) 747–756.
[157] R. Chen, G.I. Mias, J. Li-Pook-Than, et al., Personal omics profiling reveals dynamic molecular and medical phenotypes, Cell. 148 (6) (2012) 1293–1307.
[158] L. Hood, Systems biology and P4 Medicine: past, present, and future, Rambam Maimonides Med. J. 4 (2) (2013) e0012.
[159] P.A. Rudnick, S.P. Markey, J. Roth, et al., A description of the clinical proteomic tumor analysis consortium (CPTAC) common data analysis pipeline, J. Proteome Res. 15 (2016) 1023–1032.
[160] N. Lucas, A.B. Robinson, M.M. Espersen, et al., Accelerated Barocycler lysis and extraction (ABLE) sample preparation for clinical proteomics by mass spectrometry, J. Proteome Res. 18 (1) (2019) 399–405.
[161] Z. Zhang, R.C. Bast, Y. Yu, et al., Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer, Cancer Res. 64 (16) (2004) 5882–5890.
[162] M. De Wit, R.J.A. Fijneman, H.M.W. Verheul, et al., Proteomics in colorectal cancer translational research: biomarker discovery for clinical applications, Clin. Biochem. 46 (2013) 466–479.
[163] S. Tan, A. Mohamedali, A. Kapur, et al., Ultradepletion of human plasma using chicken antibodies: a proof of concept study, J. Proteome Res. 12 (6) (2013) 2399–2413.
[164] P.J. Monaghan, S. Robinson, D. Rajdl, et al., Practical guide for identifying unmet clinical needs for biomarkers, EJIFCC. 29 (2) (2018) 129–137.
[165] U. Kusebauch, D.S. Campbell, E.W. Deutsch, et al., Human SRMAtlas: a resource of targeted assays to quantify the complete human proteome, Cell. 166 (3) (2016) 766–778.
[166] F. Pontén, K. Jirström, M. Uhlen, The human protein atlas - a tool for pathology, J. Pathol. 216 (4) (2008) 387–393.
[167] X. Li, C. Hayward, P. Fong, et al., A blood-based proteomic classifier for the molecular characterization of pulmonary nodules, Sci. Transl. Med. 5 (207) (2013) 207ra142.
[168] A. Root, P. Allen, P. Tempst, Protein biomarkers for early detection of pancreatic ductal adenocarcinoma: Progress and challenges, Cancers. 10 (3) (2018).
[169] D. Ansari, W. Torén, Q. Zhou, Proteomic and genomic profiling of pancreatic cancer, Cell Biol. Toxicol. (2019), https://doi.org/10.1007/s10565-019-09465-9 [Epub ahead of print].
[170] D. Hu, D. Ansari, Q. Zhou, et al., Calcium-Activated Chloride Channel Regulator 1 as a Prognostic Biomarker in Pancreatic Ductal Adenocarcinoma, (2018), pp. 1–6.
[171] V. Rebours, J. Le Faouder, S. Laouiremb, et al., In situ proteomic analysis by MALDI imaging identifies ubiquitin and thymosin-b4 as markers of malignant intraductal pancreatic mucinous neoplasms, Pancreatology. 14 (2) (2014) 117–124.
[172] N. Malod-Dognin, J. Petschnigg, S.F.L. Windels, et al., Towards a data-integrated cell, Nat. Commun. 10 (1) (2019) 805.
[173] L.A. Ogilvie, C. Wierling, T. Kessler, et al., Predictive modeling of drug treatment in the area of personalized medicine, Cancer Informat. 14 (Suppl. 4) (2015) 95–103.
[174] M. Prosperi, J.S. Min, J. Bian, et al., Big data hurdles in precision medicine and precision public health, BMC Med. Inform. Decis. Mak. 18 (2018) 1–15.
[175] B. Mesko, The role of artificial intelligence in precision medicine, Expert Rev. Precis. Med. Drug Dev. 2 (2017) 239–241.
[176] T. Hulsen, S.S. Jamuar, A.R. Moody, et al., From big data to precision Medicine, Front. Med. 6 (2019) 1–14.