Chromatography **2014**, *35*, 73-80

Review

Challenges Facing Complete Human Proteome Analysis

Mio IWASAKI[1], Yasushi ISHIHAMA[*2]

[1]*Center for iPS Cell Research and Application, Kyoto University, Shogoin-Kawaracho, Sakyo-ku, Kyoto 606-8507, Japan*

[2]*Graduate School of Pharmaceutical Sciences, Kyoto University, Yoshida-Shimoadachicho, Sakyo-ku, Kyoto 606-8501, Japan*

Abstract

Mass spectrometry (MS) is the most widely used technology for proteome analysis, and MS-based proteomics platforms have identified over 10,000 human proteins in single-state cellular targets. Here, we review recent advances towards complete analysis of the human proteome. The main obstacles are the enormously wide dynamic range and huge complexity of the proteome. New technologies of sample preparation, protein/peptide separation and MS acquisition have been developed to tackle these issues. Pre-fractionation approaches, such as multidimensional separations prior to LC-MS/MS, have been introduced, but the total measurement time increases as the fraction number increases. Alternative approaches using high-resolution LC without pre-fractionation have also been developed, because high sample recovery is expected if the number of sample pretreatment steps is minimized. Although technical and methodological issues remain, human proteome analysis covering 7,000 - 10,000 proteins is currently feasible.

Keywords: Complete human proteome; NanoLC-MS/MS

1. Introduction

The cellular proteome, i.e., the set of proteins expressed in a particular cell under particular condition, is one of the most difficult targets for analysis because of its enormously wide dynamic range and huge complexity. So far, protein expression profiles have often been indirectly measured at the mRNA level, mainly because technologies for mRNA analysis, such as microarrays and high-throughput sequencing, already meet the requirements for complete transcriptome analysis. However, considering that protein abundance is regulated not only by transcription but also by translation [1], direct proteome analysis is essential for detailed investigation of a wide range of cellular functions. In addition, post-translational modifications (PTMs), protein-protein interactions and cellular localizations of proteins cannot be analyzed by transcriptomics. Direct analysis of the proteome thus provides otherwise inaccessible information on the molecular basis of control of cellular functions.

Mass spectrometry (MS) is a powerful tool to identify expressed proteomes, and many biological studies have been performed to examine protein turnover, protein-protein interactions, PTMs and differential proteome profiling of various cells and tissues by means of MS-based proteomics. Since 2009, the depth of proteome analysis has been dramatically increased, and it is currently possible to routinely capture over 5,000 expressed proteins in a single-state proteome.

In this review, we focus on recent advances and strategies for achieving in-depth human proteome analyses by means of MS-based proteomics. First, we describe the general workflow of MS-based shotgun proteomics. We then consider analytical obstacles to the identification of the cellular proteome and discuss recent strategies for overcoming these difficulties. Finally, we discuss the prospects for complete human proteome analysis using MS-based technologies.

2. Workflow for shotgun proteomics

'Shotgun proteomics' was a term coined by the Yates lab to describe MS-based proteomics, based on the analogy to shotgun genome sequencing [2]. This approach has also

[*]Corresponding author: Yasushi ISHIHAMA
Tel: +81-75-753-4555; Fax: +81-75-753-4601
E-mail: yishiham@pharm.kyoto-u.ac.jp

Received: 12 June 2014
Accepted: 2 July 2014
DOI: 10.15583/jpchrom.2014.013

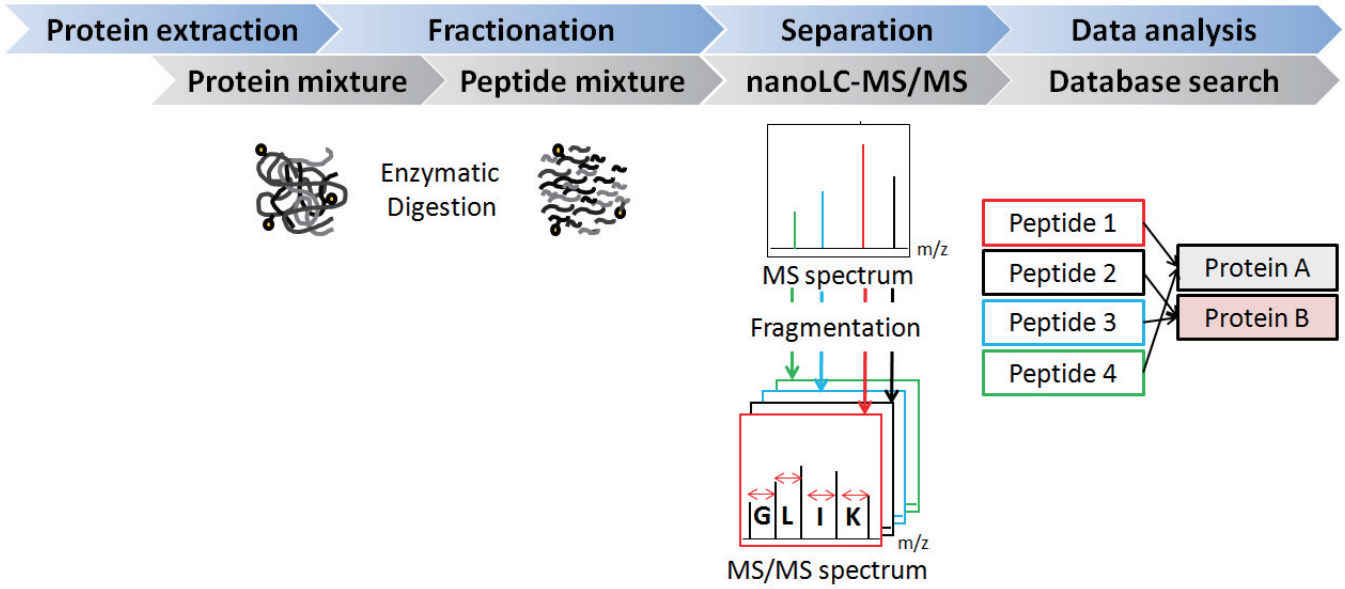

Fig. 1. Workflow of shotgun proteomics. In shotgun proteomics, proteins are extracted from cell pellets or tissues, and digested into peptides by sequence-specific proteases, such as trypsin and Lys-C. Digested peptides are separated by reversed-phase nanoLC and continually electrosprayed into a tandem mass spectrometer. Peptide ions in MS are further isolated as precursor ions. The MS/MS spectra of product ions formed by fragmentation of the precursor ions are recorded to provide information on the amino acid sequences. All peaks in MS and MS/MS spectra are digitized and searched against protein sequence databases by the use of database search engines.

been referred to as "bottom-up proteomics", because proteins are not analyzed in intact form (this would be a "top-down approach" [3]), but must be digested into peptides for MS analyses. Both top-down and bottom-up approaches have their pros and cons, but so far top-down approaches have not been fully explored because of inadequate MS sensitivity. Fig. 1 shows a typical workflow of shotgun proteomics. Extracted protein mixtures are first digested into peptides by sequence-specific proteases such as trypsin, Lys-C or Asp-N. Then, the peptide mixtures are separated by reversed-phase nano liquid chromatography and continually electrosprayed [4, 5] into a tandem mass spectrometer (nanoLC-ESI-MS/MS). Peptide ions in MS are further isolated as precursor ions and fragmented into product ions, for which MS/MS spectra are obtained to provide information on the amino acid sequences. All peaks in MS and MS/MS spectra are digitized and searched against protein sequence databases by database search engines such as Mascot [6], SEQUEST [7], Paragon [8] MaxQuant [9] and so on, to identify peptides and proteins. Generation of digested peptides is essential in shotgun proteomics, even though this step enormously increases the complexity as well as the dynamic range of the samples, making complete proteome analysis difficult.

3. Complexity and dynamic range of the proteome

Currently the number of human genes is considered to be in the range of 20,000 – 25,000 [10]. Approximately 20,000 human proteins are registered in the latest version of the UniProtKB/Swiss-Prot (2014-04) protein sequence database. Although it is still difficult to predict how many proteins are indeed translated from the human genome in a single cellular state, a recent paper reported detection of 16,846 transcripts in a single state of human HeLa cells, suggesting that up to 80% of the whole proteome might be expressed in a single cellular state [11]. We can roughly estimate the complexity of shotgun proteomics samples based on the following assumptions.

(1) Abundance distribution: the distribution of proteome abundance is assumed to be Gaussian in a logarithmic copy number scale [12] covering seven orders of magnitude for cell crude extract [1] and ten orders of magnitude for plasma [13].

(2) Total protein number: about 15,000 proteins are assumed to exist in a single-state human cell. For the present purpose, we randomly selected 15,000 human cellular proteins from the UniProtKB/Swiss-Prot database.

(3) MS-detectable tryptic peptides: we assume that trypsin digests the 15,000 proteins without any missed cleavage. Because the *m/z* scan range and the distinguishable peptide charge are limited, we assume that peptides of 600-4,000 Da can be detected by MS, resulting in approximately 640,000 peptides. However, in fact the peptide number would be larger due to the presence of numerous unexpected proteolytic digestion products as well as PTMs (*e.g.* phosphorylation, acetylation, methylation, ubiquitylation, etc.) [14-16]. Based on these considerations, we can conclude that typical shotgun proteomics samples

are likely to contain millions of detectable peptides. In addition, the dynamic range in the original peptide mixture would be further extended owing to differences in ionization efficiency of the peptides, depending upon physicochemical properties such as hydrophobicity and charge [17].

4. Progress in MS instrumentation

In shotgun proteomics, MS data acquisition is conducted by means of an MS survey scan followed by MS/MS scans to measure the *m/z* values of the peptide precursor ion and its fragment ions. Therefore, any improvements in MS performance, such as detection sensitivity, scan speed, mass accuracy, mass resolution and fragmentation efficiency, lead directly to an increase in the number of identified peptides/proteins. Extensive efforts have been made to develop new MS instrumentation, including mass analyzers (linear ion trap, hybrid ion trap-FT ICR, hybrid quadrupole-TOF, hybrid ion trap-orbitrap, hybrid quadrupole-orbitrap), ion sources, fragmentation techniques (CID, ECD, ETD, HCD), and detectors, as well as computing power to speed up the scan cycle. State-of-the-art MS instruments such as Q-Exactive (Thermo) and TripleTOF 5600 (AB Sciex) can carry out 10 - 50 Hz MS/MS scans, corresponding to 36,000 - 180,000 MS/MS scans per hour. Nevertheless, to complete analysis of a shotgun proteomics sample within one hour using nanoLC-MS/MS analysis, at least a 10-fold increase of the scan speed in addition to a much wider dynamic range, without sacrificing sensitivity, would be needed.

The work of Michalski *et al.* illustrates how many peptides can be identified by single-shot LC-MS analysis of a HeLa trypsin digest using an LTQ-Orbitrap Velos (Thermo) [18]. With a 90 min gradient, 101,726 ions with a peptide signature were detected. Less than 20% of these ions were targeted by MS/MS and about half of these led to peptide identification. This result suggests that the scan speed is inadequate to target all detected ions and also that the sensitivity in MS/MS is not enough to identify all targeted peptides. In addition, ionization suppression must be taken into account if the LC separation is not good enough to adequately reduce the co-elution of peptides [19-21]. Furthermore, the purity of the isolated precursor ions is also a critical factor. In most cases, more than one peptide ion was isolated within the selection window for MS/MS, making peptide identification difficult.

Fig. 2 shows a typical distribution of protein abundance. As the number of identified proteins increases, the depth of the analysis, *i.e.*, the dynamic range covered by the analysis, is extended and the median of the abundance of identified proteins is decreased. To accomplish deeper proteome analysis with current MS instrumentation, we have to reduce the number of peptide precursor ions per MS survey scan. One conventional strategy is to fractionate proteins and peptides prior to LC-MS/MS analysis [22, 23]. In general, protein pre-fractionation is more effective to reduce the dynamic range of the sample than peptide pre-fractionation [24], but peptide pre-fractionation is easy to automate with higher efficiency. Another strategy is to increase the LC separation efficiency as well as the separation time window in LC-MS/MS. In the field of proteomics, 38-200 cm long columns packed with 1-3 μm C18 particles [25-28] and 200 to 600 cm long C18 monolithic columns [20, 29, 30] have been used for this purpose. These strategies are effective to reduce the number of peptides per MS scan and to increase the number of MS/MS events. Ionization suppression is also reduced due to the reduction of peptide co-elution. As a result, more peptides are selected for MS/MS and more peptides and proteins are identified.

PubMed search with the seach terms “proteomics”, “human” and “comprehensive” yielded approximately 1500 papers. Among them, more than 98% reported less than 5000 human proteins as identified, and only 22 papers successfully identified more than 5,000 proteins. Table 1 lists these papers with details of the conditions used.

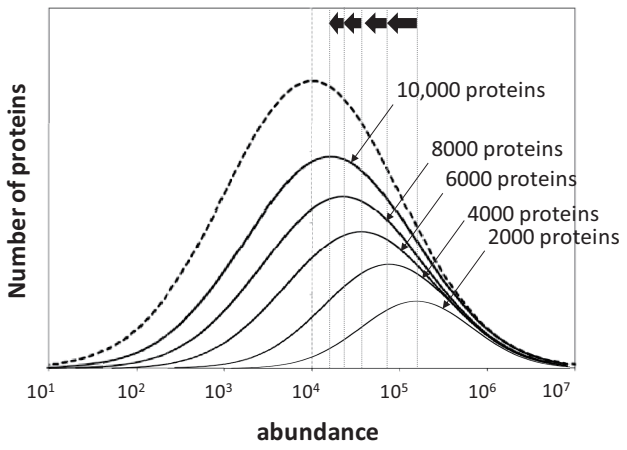

Fig. 2. Distribution of human protein abundance and proteome coverage. The distribution of proteome abundance is assumed to be Gaussian in a logarithmic copy number scale covering seven orders of magnitude, and 15,000 human proteins are assumed to exist in a single state. Proteome coverage is simulated based on random sampling in a data-dependent acquisition mode.

5. Pre-fractionation approach

As mentioned above, fractionation at the protein and/or peptide levels prior to LC-MS analyses is an effective approach to reduce sample complexity. In 2008, the yeast proteome was completely analyzed by de Godoy *et al* [31]. They employed extensive pre-fractionation both at the protein level by SDS-PAGE and at the peptide level by isoelectric focusing (IEF) followed by gas-phase

fractionation in MS. Finally, 523 LC-MS/MS runs (taking about 1 month) were performed to obtain the first complete yeast proteome map. Although the number of human genes is only 3 times larger than the number of yeast genes, the same depth of analysis of the human proteome has not yet been reported.

In 2009, Wiśniewski *et al.* identified more than 5,000 human proteins from single-state cells for the first time. They optimized the sample preparation protocol to increase protein recovery using sodium dodecyl sulfate and performed 24 IEF fractionations to identify 7,093 proteins from 50 μg samples by means of nanoLC-MS/MS measurements over 2 days [32]. Since then, identification of more than 5,000 proteins has been achieved using 6 to 300 fractionations with combinations of multidimensional LC separations (strong cation exchange: SCX, strong anion exchange: SAX), SDS-PAGE, IEF and high-pH reversed-phase (RP) LC [11, 33-49]. In terms of protein identification efficiency, Ding *et al.* achieved the highest efficiency of 680 protein identifications per hour [33] (Fig. 3). They employed short high-pH RPLC to obtain 22 fractions in the first dimension. For the second dimension, sequential 30 min RPLC-MS/MS runs at pH 3 were performed. They identified 8,154 proteins in a total LC-MS measurement time of 12 hours from 80 μg of HeLa cell lysate. Although their pre-fractionation method was very effective, the cumulative number of identified proteins did not exceed 9,000, despite the increased number of fractions. Therefore, they performed the same analyses with three different MS instruments (Q Exactive, TripleTOF 5600 and LTQ-Orbitrap Velos) and finally identified 9,037 proteins (12 h × 3 instruments).

In 2011, the number of protein identifications was taken beyond 10,000 proteins using LC-MS/MS measurements (LTQ-Orbitrap Velos) of 0.1 mg to 10 mg samples over 11 to 30 days [11, 38, 39]. Nagaraj *et al.* reported identification of 10,255 proteins from 72 fractions obtained by gel filtration and SAX from HeLa cell lysate, using different proteases to produce various types of peptides [11]. Beck *et al.* reported identification of 10,006 proteins from U2OS cells using 16 IEF fractions and an exclusion list of already identified peptides [38]. Munoz *et al.* collected 45 to 50 SCX fractions from one ES cell line and two types of iPS cells and fibroblasts to obtain up to 10,628 identified proteins in combination [39]. It is interesting that increasing the fraction number does not always result in a greater number of identified proteins. This is because too many fractions leads to an increase of overlapping proteins/peptides between fractions, suggesting that higher resolution in the first dimension would make it possible to increase the fraction number and hence the proteome coverage. Very recently, Branca *et al.* identified 13,078 human proteins by pre-fractionation using high-resolution IEF with gel strips covering pH intervals of 3.7–4.9, 3.70–4.05, 4.00–4.25, 4.20–4.45 and 4.39–4.99. Each strip was divided into 72 fractions and analyzed by LC-MS/MS on an LTQ Orbitrap Velos [47]. Although they used only 30% of the whole proteome (peptides with pI 3.9-5.0), they successfully identified the largest number of human proteins so far from single-state cells due to their ultrahigh-resolution IEF fractionation (0.0035 pH units per fraction) in the first dimension.

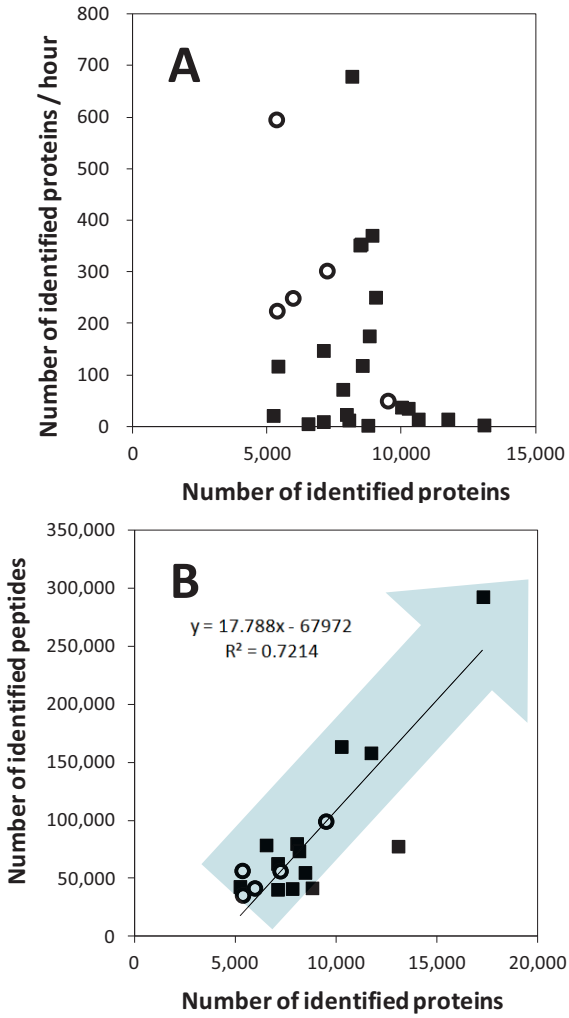

Fig. 3. Identification efficiency of proteins. Meta-analysis of 22 recent publications with >5,000 human protein identification by pre-fractionation approaches (■) or by one-shot approaches (○). A: Identification efficiency is plotted according to the number of identified proteins by number of identified proteins per analysis time (hour). The efficiency decreases greatly at around 10,000-protein identification. The analysis time is based on single MS usage. B: Number of identified peptides is plotted against the number of identified proteins. As the number of proteins to be identified increases, the required number of peptides also increases.

6. Improvement of separation efficiency in nanoLC

Another powerful strategy to increase the proteome coverage is to use high-resolution LC with an expanded separation window in LC-MS/MS without pre-fractionation.

Table 1. Summary of 22 recent publications reporting identification of more than 5,000 human proteins.

Reference no.	Reference	# identified proteins	# identified peptides	Cell	Methods used	Column	Gradient time	MS used	sample amount	Total LC-MS runs (days)
32	Wisniewski JR., Nat. Methods, 2009	7,093	40,582	HeLa	12 IEF fractions	75 µm i.d., 15 cm long (ReproSil-Pur C18-AQ, 3 µm)	240 min	LTQ-Orbitrap XL	50 µg	48 h (2 days)
34	van Hoof D., Cell Stem Cell, 2009	5,222	43,017	ESC (HuESC-7)	24 SCX fractions, Phosphopeptides enrichment (TiO_2 2D-LC)	50 µm i.d., 20 cm long (ReproSil-Pur C18-AQ, 3 µm)	100 min	LTQ-Orbitrap XL	1000 µg	144 runs (10 days)
35	Lundberg E., Mol. Syst. Biol., 2010	5,399	--	U2 OS, A-431, U-251 MG	SDS-PAGE (15 slices), 7 SAX fractions	75 µm i.d., unknown length (ReproSil-Pur C18-AQ, 3 µm)	100 min (gel) 180 min (SAX)	LTQ-Orbitrap XL	unknown	46 h (2 days)
36	Rigbolt KTG., Sci. Signal., 2011	6,521	78,819	ESC (HUES9, Odense-3)	SDS-PAGE (20 slices), 15 SCX fractions, phosphopeptides enrichment (TiO_2)	75 µm i.d., 20 cm long (ReproSil-Pur C18-AQ, 3 µm)	90 min	LTQ-FT Ultra, LTQ-Orbitrap, LTQ-Orbitrap XL	> 100 µg	Replicate of 378 runs (47 days)
50	Thakur SS., Mol. Cell. Proteomics, 2011	5,376	35,155	HEK293	no fractionation	75 µm i.d., 50 cm long (ReproSil-Pur C18-AQ, 1.8 µm)	480 min	LTQ-Orbitrap Velos	6 µg	3 runs (1 day)
37	Phanstiel DH., Nat. Methods, 2011	7,952	--	ESC lines (H1, H9), iPSC (DF19.7), Fibroblast (NFF)	SCX fractions, phosphopeptides enrichment (magnetic beads)	50 µm i.d., 15 cm long (Alltech C18)	120 min	dcQLT-Orbitrap	unknown	< 14 days
46	Ting L., Nat. Methods, 2011	8,803	41,804	HeLa	20 SCX fractions	100 µm i.d., 20 cm long (Maccel C18 AQ, 3 µm)	150 min	LTQ-Orbitrap Velos	> 150 µg	20 runs (2 days)
11	Nagaraj N., Mol. Syst. Biol., 2011	10,255	163,784	HeLa	gel fractionation, GluC digestion, 72 SAX fractions	75 µm i.d., 40 cm long (ReproSil-Pur C18-AQ, 1.8 µm)	240 min	LTQ-Orbitrap Velos	10 mg	72 runs (12 days)
38	Beck M., Mol. Syst. Biol., 2011	10,006	--	U2OS	16 IEF fractions, inclusion list mode	75 µm i.d., 15 cm long (Magic C18-AQ, 3 µm)	120 min	hybrid LTQ-FT-ICR, LTQ-Orbitrap Velos	127 µg	127 runs (11 days)
39	Munoz J., Mol. Syst. Biol., 2011	10,628	--	ESC (HES-3). iPSC (IMR90, 4Skin), Fibroblast (IMR90, 4Skin)	45 or 50 SCX fractions	50 µm i.d., 40 cm long (ReproSil-Pur C18-AQ, 3 µm)	90 or 180 min	LTQ-Orbitrap XL ETD LTQ-Orbitrap Velos	unknown	348 runs (30 days)
29	Iwasaki M., J. Chromatogr. A, 2012	5,970	41,319	HeLa	no fractionation	100 µm i.d., 400 cm long (C18 monolithic column)	480 min	TT5600	12 µg	3 runs (1 day)
40	Geiger T., Mol. Cell. Proteomics, 2012	8,543	--	HEK293	6 SAX fractions	75 µm i.d., 20 cm long (ReproSil-Pur C18-AQ, 1.8 µm)	240 min	LTQ-Orbitrap Velos	40 µg	Triplicate of 6 runs (3 days)
		11,731	158,294	A549, GAMG, HEK293, HeLa, HepG2, K562, MCF7, RKO, U2OS, LnCap and Jurkat					440 µg	198 runs (33 days)
41	Geiger T., Cancer Res., 2012	8,750	--	HMEC, MCF10a, HMT-3522-S1, HCC1937, HCC1143, HCC1599, HCC202, HCC2218, MFM223, MDA-MB-453, MCF7	12 IEF fractions, 6 SAX fractions	75 µm i.d., 15 cm long (ReproSil-Pur C18-AQ, 3 µm)	90 min	LTQ-OrbiTrap XL	1,500 µg	Triplicate of 720 runs (135 days)
42	Arabi A., Nat. Commun., 2012	7,816	41,238	HCT116	2 subcellular fractions, 72 IEF fractions	100 µm i.d., 15 cm long (C18, 5 µm)	45 min	LTQ-Orbitrap Velos	60 µg	144 runs (4.5 days)
43	Boisvert FM., Mol. Cell. Proteomics, 2012	8,041	80,098	HeLa	subcellular fractions (Cytoplasm, nuclei, and nucleoli), SDS-PAGE (16 bands for in-gel digestion)	75 µm i.d., 15 cm long (PepMap C18 nano column)	65 min	LTQ-OrbiTrap XL LTQ-Orbitrap Velos	1,152 µg	576 runs (26 days)
44	Muraoka S., J. Proteome Res., 2013	7,092	62,865	Breast Cancer Tissue	36 SCX fractions	100 µm i.d., 20 cm long (L-column2 C18, 3 µm)	135 min	LTQ-Orbitrap Velos	180 µg / patient	(30 - 60 days)
54	Yamana R., J. Proteome Res., 2013	7,241	56,121	iPSC (32R1)	no fractionation	100 µm i.d., 200 cm long (C18 monolithic column)	480 min	TT5600	4 µg	3 runs (1 day)
		9,510	98,977	iPSC (32R1, 201B7, 414C2, 585A1, 606A1), Fibroblast (aHDF1388, aHDF1419, Tig120slc)					96 µg	24 runs (8 days)
45	Wisniewski JR., Proteomics Clin. Appl., 2013	8,500	--	HeLa	6 SAX fractions	75 µm i.d., 20 cm long (ReproSil-Pur C18-AQ, 1.8 µm)	240 min	Q-Exactive	6 µg	6 runs (1 day)
		8,906	--	HCT116					6 µg	6 runs (1 day)
		8,453	55,144	microdissected colonic adenoma tissue					6 µg	6 runs (1 day)
51	Pirmoradian M., Mol. Cell. Proteomics, 2013	5,354	56,390	A375	no fractionation	75 µm i.d., 50 cm long (PepMap RSLC C18)	180 min	Q-Exactive	9 µg	3 runs (9 hours)
33	Ding C., Mol. Cell. Proteomics, 2013	8,154	73,680	HeLa	22 short pH 10 RP fractions, pH 3 RP-LC-MS/MS runs	75 µm i.d., 10 cm long (C18 particles, 3 µm)	30 min	Q-Exactive	80 µg	22 runs (0.5 day)
		9,037	--					TT5600 Q-Exactive LTQ-Orbitrap Velos	240 µg	66 runs (1.5 day)
47	Branca RMM., Nat. Methods, 2014	13,078	77,785	A431	subcellular fractions (whole, nuclear, organellar, soluble), 72 high resolution IEF (pH3.7-4.9) fractions	100 µm i.d., 15 cm long (C18, 5 µm)	45 min	LTQ-Orbitrap Velos	unknown	2,160 runs (67.5 day)
48	Kim MS, Nature, 2014	17,294	293,000	17 human adult tissues, 7 Fetal tissues, 6 hematopoietic cell types	SDS-PAGE, 22 high pH RP fractions	75 µm i.d., 10 cm long on Velos 75 µm i.d., 20 cm long on Elite	60 min	LTQ-Orbitrap Velos LTQ-Orbitrap Elite	unknown (less than 450 µg / sample)	> 2,000 runs

This approach is called 'one-shot proteomics' or 'single-shot proteomics'. Generally, the usage of smaller particles and longer columns increases the separation efficiency. Compared to a pre-fractionation-based approach, the one-shot approach is expected to have higher recovery/sensitivity because it does not require any additional step prior to LC-MS. However, the one-shot approach with small particle-packed long columns has the drawback that both smaller particles and longer columns lead to higher back-pressure which must be kept below the pressure limit of the LC system employed. Considering that standard 15 cm columns packed with 3 μm particles (75-100 μm inner diameter) generate 12-15 MPa at 1 mm/s linear velocity, the maximum length that can be used in general HPLC systems is around 35-50 cm at room temperature.

So far, four papers have reported identification of more than 5,000 human proteins by one-shot LC-MS/MS analyses without pre-fractionation. In 2011, Thakur *et al.* used a 50 cm column (75 μm i.d.) packed with 1.8 μm particles to identify 5,376 proteins from 6 μg of HEK293 cells in triplicate analyses within 24 hours [50]. Similar results were obtained by Pirmoradian *et al.* in 2013 with a much higher throughput (9 hour gradient time) from 9 μg of A375 cells (triplicate analyses of 3-h gradient LC-MS/MS run). They used almost the same analytical settings, but each experimental procedure was optimized [51].

To further increase separation efficiency, monolithic silica columns currently seem to be the best option in terms of column length and feasible LC pressure. Monolithic columns have small skeletons and relatively large through-pores, which result in much lower back-pressure compared to particle-packed columns, while they retain a separation efficiency comparable to that of 3 μm particle-packed columns [52]. In 2012, we reported the identification of 5,970 proteins using a 400 cm long monolithic column in triplicate analyses of 12 μg of HeLa cell digest within 24 h [29]. After optimization of the sample preparation methods [53], as many as 7,241 proteins were identified in one day (triplicate analyses with 8-hour gradient time), and 9,510 proteins from 5 types of iPS cells and 3 types of fibroblasts in 8 days [54].

7. Perspectives: Toward complete human proteome analysis

Very recently, two groups independently reported the largest human proteome draft maps to date [48, 49]. Pandey's group profiled 30 human samples including 17 adult tissues, 7 fetal tissues and 6 purified primary haematopoietic cells, resulting in identification of 17,294 gene products. Kuster's group used both publicly available LC-MS data and their own data to profile 18,097 gene products. These groups identified proteins from >2,000 and 16,857 LC-MS runs, respectively. As more and more data are accumulated from the proteomics community, these draft maps should rapidly become comprehensive. Then, the next step would be to establish a routine method to profile single-state human cellular proteomes of interest with reasonable throughput and acceptable sample size. Currently, single-state human proteome analysis covering 7,000 - 10,000 proteins is feasible, although some technical and methodological issues remain. However, it is still a significant challenge to identify more than 10,000 proteins, because the identification efficiency of current technology is greatly decreased at this level of identification, as shown in Fig. 3A. To accomplish identification of 15,000 proteins, approximately 200,000 non-redundant peptides need to be identified, as predicted from Fig. 3B. This should be possible if 'one-shot'-type long-gradient LCMS analysis is combined with high-resolution pre-fractionation, though the total measurement time might be long. Breakthroughs in MS, LC and other technologies will be needed to achieve complete single-state human cellular proteome analysis.

Recently, it has been reported that ESI sensitivity is increased by employing lower LC flow rates and narrower analytical columns [55, 56]. In addition, organic solvents also increase the sensitivity when they are added to the LC mobile phase [57]. We reported that the use of hydrophilic interaction chromatography (HILIC)-MS with meter-long ureidopropyl-modified silica monolithic columns dramatically increased the MS intensities of peptides, compared to RPLC-MS, due to the higher content of acetonitrile [58].

It is still the case that only a limited number of laboratories can carry out proteome analyses reproducibly and comprehensively [59]. Moreover, the cost of operating a high-end MS is in the range of several hundred to one thousand dollars per day [60]. However, trans-omics comparisons, such as in-depth proteomics combined with other omics data, can provide unique opportunities for biologists to uncover biological mechanisms. Given the demand for simplified and high-throughput analytical platforms, we anticipate rapid progress in this field. Even with current technologies, it will be possible to routinely identify at least 10,000 proteins within 10 hours in the near future [12, 51].

Acknowledgements

This work is supported by a Japan Society for the Promotion of Science (JSPS) Fellowship for Young Scientists to M.I., and a JSPS Grant-in-Aid for Scientific Research to Y.I.

References

[1] Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M.

Nature **2011**, *473*, 337-342.
[2] Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R. *Nat. Biotechnol*. **1999**, *17*, 676-682.
[3] Durbin, K. R.; Tran, J. C.; Zamdborg, L.; Sweet, S. M.; Catherman, A. D.; Lee, J. E.; Li, M.; Kellie, J. F.; Kelleher, N. L. *Proteomics* **2010**, *10*, 3589-3597.
[4] Yamashita, M.; Fenn, J. B. *J. Phys. Chem.* **1984**, *88*, 4451-4459.
[5] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.
[6] Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551-3567.
[7] Yates, J. R.; Eng, J. K.; McCormack, A. L.; Schieltz, D. *Anal. Chem.* **1995**, *67*, 1426-1436.
[8] Shilov, I. V.; Seymour, S. L.; Patel, A. A.; Loboda, A.; Tang, W. H.; Keating, S. P.; Hunter, C. L.; Nuwaysir, L. M.; Schaeffer, D. A. *Mol. Cell. Proteomics* **2007**, *6*, 1638-1655.
[9] Cox, J.; Mann, M. *Nat. Biotechnol.* **2008**, *26*, 1367-1372.
[10] Clamp, M.; Fry, B.; Kamal, M.; Xie, X.; Cuff, J.; Lin, M. F.; Kellis, M.; Lindblad-Toh, K.; Lander, E. S. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 19428-19433.
[11] Nagaraj, N.; Wisniewski, J. R.; Geiger, T.; Cox, J.; Kircher, M.; Kelso, J.; Pääbo, S.; Mann, M. *Mol. Syst. Biol.* **2011**, *7*, 548.
[12] Zubarev, R. A. *Proteomics* **2013**, *13*, 723-726.
[13] Anderson, N. L.; Anderson, N. G. *Mol. Cell. Proteomics* **2002**, *1*, 845-867.
[14] Chalkley, R. J.; Baker, P. R.; Hansen, K. C.; Medzihradszky, K. F.; Allen, N. P.; Rexach, M.; Burlingame, A. L. *Mol. Cell. Proteomics* **2005**, *4*, 1189-1193.
[15] Picotti, P.; Aebersold, R.; Domon, B. *Mol. Cell. Proteomics* **2007**, *6*, 1589-1598.
[16] Nielsen, M. L.; Savitski, M. M.; Zubarev, R. A. *Mol. Cell. Proteomics* **2006**, *5*, 2384-2391.
[17] Nishikaze, T.; Takayama, M. *Int. J. Mass Spectrom*. **2007**, *268*, 47-59.
[18] Michalski, A.; Cox, J.; Mann, M. *J. Proteome Res.* **2011**, *10*, 1785-1793.
[19] Tolstikov, V. V.; Lommen, A.; Nakanishi, K.; Tanaka, N.; Fiehn, O. *Anal. Chem*. **2003**, *75*, 6737-6740.
[20] Iwasaki, M.; Miwa, S.; Ikegami, T.; Tomita, M.; Tanaka, N.; Ishihama, Y. *Anal. Chem*. **2010**, *82*, 2616-2620.
[21] Shi, T.; Fillmore, T. L.; Gao, Y.; Zhao, R.; He, J.; Schepmoes, A. A.; Nicora, C. D.; Wu, C.; Chambers, J. L.; Moore, R. J.; Kagan, J.; Srivastava, S.; Liu, A. Y.; Rodland, K. D.; Liu, T.; Camp, D. G.,2nd; Smith, R. D.; Qian, W. J. *Anal. Chem*. **2013**, *85*, 9196-9203.
[22] Cargile, B. J.; Sevinsky, J. R.; Essader, A. S.; Stephenson, J. L.; Bundy, J. L. *J. Biomol. Tech.* **2005**, *16*, 181-189.
[23] Motoyama, A.; Yates, J. R. *Anal. Chem*. **2008**, *80*, 7187-7193.
[24] Ishihama, Y. *J. Chromatogr. A* **2005**, *1067*, 73-83.
[25] Shen, Y.; Zhao, R.; Berger, S. J.; Anderson, G. A.; Rodriguez, N.; Smith, R. D. *Anal. Chem.* **2002**, *74*, 4235-4249.
[26] Shen, Y.; Zhang, R.; Moore, R. J.; Kim, J.; Metz, T. O.; Hixson, K. K.; Zhao, R.; Livesay, E. A.; Udseth, H. R.; Smith, R. D. *Anal. Chem*. **2005**, *77*, 3090-3100.
[27] Jerkovich, A. D.; Mellors, J. S.; Jorgenson, J. W. *LC-GC Europe* **2003**, *16*, 20-23.
[28] Köcher, T.; Swart, R.; Mechtler, K. *Anal. Chem*. **2011**, *83*, 2699-2704.
[29] Iwasaki, M.; Sugiyama, N.; Tanaka, N.; Ishihama, Y. *J. Chromatogr. A* **2012**, *1228*, 292-297.
[30] Yamana, R.; Iwasaki, M.; Wakabayashi, M.; Nakagawa, M.; Yamanaka, S.; Ishihama, Y. *J. Proteome Res.* **2013**, *12*, 214-221.
[31] de Godoy, L. M.; Olsen, J. V.; Cox, J.; Nielsen, M. L.; Hubner, N. C.; Fröhlich, F.; Walther, T. C.; Mann, M. *Nature* **2008**, *455*, 1251-1254.
[32] Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. *Nat. Methods* **2009**, *6*, 359-362.
[33] Ding, C.; Jiang, J.; Wei, J.; Liu, W.; Zhang, W.; Liu, M.; Fu, T.; Lu, T.; Song, L.; Ying, W.; Chang, C.; Zhang, Y.; Ma, J.; Wei, L.; Malovannaya, A.; Jia, L.; Zhen, B.; Wang, Y.; He, F.; Qian, X.; Qin, J. *Mol. Cell. Proteomics* **2013**, *12*, 2370-2380.
[34] Van Hoof, D.; Muñoz, J.; Braam, S. R.; Pinkse, M. W.; Linding, R.; Heck, A. J.; Mummery, C. L.; Krijgsveld, J. *Cell Stem Cell* **2009**, *5*, 214-226.
[35] Lundberg, E.; Fagerberg, L.; Klevebring, D.; Matic, I.; Geiger, T.; Cox, J.; Algenas, C.; Lundeberg, J.; Mann, M.; Uhlen, M. *Mol. Syst. Biol.* **2010**, *6*, 450.
[36] Rigbolt, K. T.; Prokhorova, T. A.; Akimov, V.; Henningsen, J.; Johansen, P. T.; Kratchmarova, I.; Kassem, M.; Mann, M.; Olsen, J. V.; Blagoev, B. *Sci. Signal.* **2011**, *4*, rs3.
[37] Phanstiel, D. H.; Brumbaugh, J.; Wenger, C. D.; Tian, S.; Probasco, M. D.; Bailey, D. J.; Swaney, D. L.; Tervo, M. A.; Bolin, J. M.; Ruotti, V.; Stewart, R.; Thomson, J. A.; Coon, J. J. *Nat. Methods* **2011**, *8*, 821-827.
[38] Beck, M.; Schmidt, A.; Malmstroem, J.; Claassen, M.; Ori, A.; Szymborska, A.; Herzog, F.; Rinner, O.; Ellenberg, J.; Aebersold, R. *Mol. Syst. Biol*. **2011**, *7*, 549.
[39] Munoz, J.; Low, T. Y.; Kok, Y. J.; Chin, A.; Frese, C. K.; Ding, V.; Choo, A.; Heck, A. J. *Mol. Syst. Biol.* **2011**, *7*, 550.
[40] Geiger, T.; Wehner, A.; Schaab, C.; Cox, J.; Mann, M.

Mol. Cell. Proteomics **2012**, *11*, M111.014050.
[41] Geiger, T.; Madden, S. F.; Gallagher, W. M.; Cox, J.; Mann, M. *Cancer Res*. **2012**, *72*, 2428-2439.
[42] Arabi, A.; Ullah, K.; Branca, R. M.; Johansson, J.; Bandarra, D.; Haneklaus, M.; Fu, J.; Aries, I.; Nilsson, P.; Den Boer, M. L.; Pokrovskaja, K.; Grander, D.; Xiao, G.; Rocha, S.; Lehtio, J.; Sangfelt, O. *Nat. Commun*. **2012**, *3*, 976.
[43] Boisvert, F. M.; Ahmad, Y.; Gierliński, M.; Charrière, F.; Lamont, D.; Scott, M.; Barton, G.; Lamond, A. I. *Mol. Cell. Proteomics* **2012**, *11*, M111.011429.
[44] Muraoka, S.; Kume, H.; Adachi, J.; Shiromizu, T.; Watanabe, S.; Masuda, T.; Ishihama, Y.; Tomonaga, T. *J. Proteome Res.* **2013**, *12*, 208-213.
[45] Wiśniewski, J. R.; Duś, K.; Mann, M. *Proteomics Clin. Appl.* **2013**, *7*, 225-233.
[46] Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. *Nat. Methods* **2011**, *8*, 937-940.
[47] Branca, R. M.; Orre, L. M.; Johansson, H. J.; Granholm, V.; Huss, M.; Perez-Bercoff, A.; Forshed, J.; Kall, L.; Lehtio, J. *Nat. Methods* **2014**, *11*, 59-62.
[48] Kim, M. S.; Pinto, S. M.; Getnet, D.; Nirujogi, R. S.; Manda, S. S.; Chaerkady, R.; Madugundu, A. K.; Kelkar, D. S.; Isserlin, R.; Jain, S.; Thomas, J. K.; Muthusamy, B.; Leal-Rojas, P.; Kumar, P.; Sahasrabuddhe, N. A.; Balakrishnan, L.; Advani, J.; George, B.; Renuse, S.; Selvan, L. D.; Patil, A. H.; Nanjappa, V.; Radhakrishnan, A.; Prasad, S.; Subbannayya, T.; Raju, R.; Kumar, M.; Sreenivasamurthy, S. K.; Marimuthu, A.; Sathe, G. J.; Chavan, S.; Datta, K. K.; Subbannayya, Y.; Sahu, A.; Yelamanchi, S. D.; Jayaram, S.; Rajagopalan, P.; Sharma, J.; Murthy, K. R.; Syed, N.; Goel, R.; Khan, A. A.; Ahmad, S.; Dey, G.; Mudgal, K.; Chatterjee, A.; Huang, T. C.; Zhong, J.; Wu, X.; Shaw, P. G.; Freed, D.; Zahari, M. S.; Mukherjee, K. K.; Shankar, S.; Mahadevan, A.; Lam, H.; Mitchell, C. J.; Shankar, S. K.; Satishchandra, P.; Schroeder, J. T.; Sirdeshmukh, R.; Maitra, A.; Leach, S. D.; Drake, C. G.; Halushka, M. K.; Prasad, T. S.; Hruban, R. H.; Kerr, C. L.; Bader, G. D.; Iacobuzio-Donahue, C. A.; Gowda, H.; Pandey, A. *Nature* **2014**, *509*, 575-581.
[49] Wilhelm, M.; Schlegl, J.; Hahne, H.; Moghaddas Gholami, A.; Lieberenz, M.; Savitski, M. M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; Mathieson, T.; Lemeer, S.; Schnatbaum, K.; Reimer, U.; Wenschuh, H.; Mollenhauer, M.; Slotta-Huspenina, J.; Boese, J. H.; Bantscheff, M.; Gerstmair, A.; Faerber, F.; Kuster, B. *Nature* **2014**, *509*, 582-587.
[50] Thakur, S. S.; Geiger, T.; Chatterjee, B.; Bandilla, P.; Fröhlich, F.; Cox, J.; Mann, M. *Mol. Cell. Proteomics* **2011**, *10*, M110.003699.
[51] Pirmoradian, M.; Budamgunta, H.; Chingin, K.; Zhang, B.; Astorga-Wells, J.; Zubarev, R. A. *Mol. Cell. Proteomics* **2013**, 12, 3330-3338.
[52] Miyamoto, K.; Hara, T.; Kobayashi, H.; Morisaka, H.; Tokuda, D.; Horie, K.; Koduki, K.; Makino, S.; Núñez, O.; Yang, C.; Kawabe, T.; Ikegami, T.; Takubo, H.; Ishihama, Y.; Tanaka, N. *Anal. Chem*. **2008**, *80*, 8741-8750.
[53] Masuda, T.; Tomita, M.; Ishihama, Y. *J. Proteome Res.* **2008**, *7*, 731-740.
[54] Yamana, R.; Iwasaki, M.; Wakabayashi, M.; Nakagawa, M.; Yamanaka, S.; Ishihama, Y. *J. Proteome Res*. **2013**, *12*, 214-221.
[55] Ficarro, S. B.; Zhang, Y.; Lu, Y.; Moghimi, A. R.; Askenazi, M.; Hyatt, E.; Smith, E. D.; Boyer, L.; Schlaeger, T. M.; Luckey, C. J.; Marto, J. A. *Anal. Chem*. **2009**, *81*, 3440-3447.
[56] Heemskerk, A. A.; Busnel, J. M.; Schoenmaker, B.; Derks, R. J.; Klychnikov, O.; Hensbergen, P. J.; Deelder, A. M.; Mayboroda, O. A. *Anal. Chem*. **2012**, *84*, 4552-4559.
[57] Hahne, H.; Pachl, F.; Ruprecht, B.; Maier, S. K.; Klaeger, S.; Helm, D.; Medard, G.; Wilm, M.; Lemeer, S.; Kuster, B. *Nat. Methods* **2013**, *10*, 989-991.
[58] Horie, K.; Kamakura, T.; Ikegami, T.; Wakabayashi, M.; Kato, T.; Tanaka, N.; Ishihama, Y. *Anal. Chem*. **2014**, *86*, 3817-3824.
[59] Bell, A. W.; Deutsch, E. W.; Au, C. E.; Kearney, R. E.; Beavis, R.; Sechi, S.; Nilsson, T.; Bergeron, J. J.; Group, H. T. S. W. *Nat. Methods* **2009**, *6*, 423-430.
[60] Ahrens, C. H.; Brunner, E.; Qeli, E.; Basler, K.; Aebersold, R. *Nat. Rev. Mol. Cell. Biol.* **2010**, *11*, 789-801.