Open access • Journal Article • DOI:10.1002/ELPS.201700330

How many proteins can be identified in a 2DE gel spot within an analysis of a complex human cancer tissue proteome — Source link

Xianquan Zhan, Haiyan Yang, Fang Peng, Jianglin Li ...+10 more authors

Institutions: Central South University, Hunan University, Max Planck Society

Published on: 01 Apr 2018 - Electrophoresis (John Wiley & Sons, Ltd)

Topics: Proteome, Human proteome project, Proteomics, Two-dimensional gel electrophoresis and Gel electrophoresis

Xianquan Zhan[1,2,3,4]
Haiyan Yang[1,2,3]
Fang Peng[1,2,3]
Jianglin Li[5]
Yun Mu[1,2,3]
Ying Long[1,2,3]
Tingting Cheng[1,2,3]
Yuda Huang[1,2,3]
Zhao Li[6]
Miaolong Lu[1,2,3]
Na Li[1,2,3]
Maoyu Li[1,2,3]
Jianping Liu[7]
Peter R. Jungblut[8]

[1]Key Laboratory of Cancer Proteomics of Chinese Ministry of Health, Xiangya Hospital, Central South University, Changsha, Hunan, P. R. China
[2]Hunan Engineering Laboratory for Structural Biology and Drug Design, Xiangya Hospital, Central South University, Changsha, Hunan, P. R. China
[3]State Local Joint Engineering Laboratory for Anticancer Drugs, Xiangya Hospital, Central South University, Changsha, Hunan, P. R. China
[4]The State Key Laboratory of Medical Genetics, Central South University, Changsha, Hunan, P. R. China
[5]Molecular Science and Biomedicine Laboratory, State Key Laboratory for Chemo/Biosensing and Chemometrics, College of Biology, Hunan University, Changsha, Hunan, P. R. China
[6]Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha, Hunan, P. R. China
[7]Bio-Analytical Chemistry Research Laboratory, Modern Analytical Testing Center, Central South University, Changsha, Hunan, P. R. China
[8]Max Planck Institute for Infection Biology, Core Facility Protein Analysis, Berlin, Germany

Received April 30, 2017
Revised November 3, 2017
Accepted November 17, 2017

Research Article

How many proteins can be identified in a 2DE gel spot within an analysis of a complex human cancer tissue proteome?

Two-dimensional gel electrophoresis (2DE) in proteomics is traditionally assumed to contain only one or two proteins in each 2DE spot. However, 2DE resolution is being complemented by the rapid development of high sensitivity mass spectrometers. Here we compared MALDI-MS, LC-Q-TOF MS and LC-Orbitrap Velos MS for the identification of proteins within one spot. With LC-Orbitrap Velos MS each Coomassie Blue-stained 2DE spot contained an average of at least 42 and 63 proteins/spot in an analysis of a human glioblastoma proteome and a human pituitary adenoma proteome, respectively, if a single gel spot was analyzed. If a pool of three matched gel spots was analyzed this number further increased up to an average of 230 and 118 proteins/spot for glioblastoma and pituitary adenoma proteome, respectively. Multiple proteins per spot confirm the necessity of isotopic labeling in large-scale quantification of different protein species in a proteome. Furthermore, a protein abundance analysis revealed that most of the identified proteins in each analyzed 2DE spot were low-abundance proteins. Many proteins were present in several of the analyzed spots showing the ability of 2DE-MS to separate at the protein species level. Therefore, 2DE coupled with high-sensitivity LC-MS has a clearly higher sensitivity as expected until now to detect, identify and quantify low abundance proteins in a complex human proteome with an estimated resolution of about 500 000 protein species. This clearly exceeds the resolution power of bottom-up LC-MS investigations.

Keywords:
Protein species / Proteome / Resolution / Tandem mass spectrometry / Two-dimensional gel electrophoresis
DOI 10.1002/elps.201700330

Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

Two-dimensional gel electrophoresis (2DE) is a traditional and classical technique to separate proteins by two important characteristics that are isoelectric point (pI) and molecular weight (Mr) [1–3]. 2DE is a common and extensively

Correspondence: Dr. Xianquan Zhan, Key Laboratory of Cancer Proteomics of Chinese Ministry of Health, Xiangya Hospital, Central South University, 87 Xiangya Road, Changsha, Hunan 410008 P. R. China
E-mail: yjzhan2011@gmail.com

Abbreviation: CHAPS, 3-(3-cholamidopropyl) dimethylammonio-1-propanesulfonate

used technique in the separation of the components of complex proteomes with more than thousand publications before and after the terms "proteome" and "proteomics" were proposed in 1995 [4]. Most of publications prerequisite that each 2DE gel spot only contains one or two proteins, which is evidenced by randomly selected documents [5–18], and also by the World 2DPAGE database (World-2DPAGE portal: http://world-2dpage.expasy.org/portal) and 2D-PAGE at MPIIB: (http://www.mpiib-berlin.mpg.de/2D-PAGE/). The World 2DPAGE is a dynamic and virtual portal, which integrates over 250 maps for 23 species, totally including nearly 40 000 identified spots, and is the biggest gel-based proteomics dataset available from a single interface [19]. A few publications show that several proteins were contained in some 2DE gel spots [5, 6, 20, 21]. A comprehensive analysis of the Hela cell proteome by SILAC 2DE-LC-LTQ OrbiTrap XL revealed that 676 spots from 816 analyzed ones contained at least two with up to 22 different proteins [21]. Combined with 2DE gel image analysis, conventional 2DE-based comparative proteomics has been extensively used to determine "differentially expressed proteins" under different physiological, pathological, or pathophysiological conditions as reviewed in the selected publications [22–26]. However, the protein composition in a human tissue, cell, or body-fluid proteome is very complex with a very wide abundance range and complicated huge protein speciation caused by splicing, truncation and modification [27]. The protein species concept lets one to in-depth understand a proteome [28–30]. By protein speciation, one protein may be diversified into 1 billion protein species as was estimated for a histone [31]. Therefore, it can be expected that one spot contains not only different protein species but also different proteins. Now 2DE resolution can be complemented by super-high sensitivity mass spectrometers, such as OrbiTrap Velos to elucidate these proteins.

Mass spectrometry is the key technique to identify proteins in a complex human proteome. Its sensitivity significantly affects the number of identified proteins from complex proteomes. In our laboratory, the sensitivity of the old ESI-Q-TOF and MALDI-TOF-TOF is at the level of 10–100 fmol for identification of endogenous proteins from a human proteome. In contrast, the sensitivity of the new OrbiTrap Velos is at the level of 1–10 amol for identification of endogenous proteins from a human proteome. The new OrbiTrap Velos dramatically increases the ability of identification of absolutely low-abundance proteins in a complex human proteome and benefits the identification of low-abundance proteins that are present in a 2DE gel spot in separations of complex human tissue proteomes.

Here we describe that each Coomassie Blue-stained 2DE gel spot contains depending on the identification criteria up to several hundred proteins derived from a pooled sample of three matched gel spots and about 50 proteins derived from a single gel spot. As a major consequence, the quantification of spot intensities is only useful, if it is proofed that the spot contains only one protein or even more consequently only one protein species. With our investigation, this seems rarely to be the case and therefore the only solution at present is to use isotopic labeling for quantification in 2DE/MS based approaches of tissues as was already suggested for complete cells [21]. 2DE-MS with isotopic labeling and the application of high-sensitivity MS enables the quantification of a much larger part of the human proteome as assumed before.

2 Materials and methods

2.1 Tumor tissues and protein extracts

A human glioblastoma tissue (male, 34 years old, IV-grade of astrocytoma), and a human pituitary adenoma tissue [male, 22 years old, invasive pituitary adenoma, ACTH (−), hGH (−), PRL (++), LH (−), and FSH (−)] was obtained from the Department of Neurosurgery of Xiangya Hospital, China, and approved by the Xiangya Hospital Medical Ethics Committee of Central South University, China.

For glioblastoma, proteins were extracted from a portion of glioblastoma tissue (~650 mg) with a volume (3 ml) of protein extraction buffer that contained 7 M urea, 2 M thiourea, 100 mM dithiothreitol (DTT), and 1 mM PMSF. For pituitary adenoma, a portion of pituitary adenoma tissue (~310 mg) was homogenized in 5.64 ml of homogenizing buffer that contained 2 mol/L acetic acid and 1 mL/L β-mercaptoethanol, the homogenate was sonicated (20 s) and lyophilized, and then was stored (−80°C), and then 17 mg lyophilizates were used to extract proteins with a volume (2.2 ml) of protein extraction buffer described above. The extracted protein concentration (glioblastoma: 4.20 μg/μl with a total volume of ~ 3 ml; pituitary adenoma: 4.52 μg/μl with a total volume of ~ 2.2 ml) was determined with a 2D Quant kit (GE Healthcare).

2.2 2DE and protein staining

For an 18-cm immobilized pH gradient (IPG) strip (pH 3–10 NL, GE Healthcare), a total of 500 μg (glioblastoma: 119 μL; pituitary adenoma: 110.6 μL) of protein extract was mixed with a volume of 231 μL or 239.4 μL of protein extraction buffer (8 M urea, 2% w/v CHAPS, 100 mM DTT, 0.5% v/v IPG buffer pH 3–10 NL, and a trace of bromophenol blue) for glioblastoma or pituitary adenoma, respectively. For the first dimension – isoelectic focusing (IEF), the precast IPG strips (pH 3–10 NL; 180 × 3 × 0.5 mm), 18 cm IPGstrip holder, and an IPGphor instrument (GE Heathcare) were used, and performed under the conditions of rehydration overnight (~18 h), followed by focusing (20°C) by a multi-step gradient (250 V and 1 h for 125 Vh; 1000 V and 1 h for 500 Vh; 8000 V and 1 h for 4000 Vh; step and hold at 8000 V and 4 h for 32 000 Vh; and a step and hold at 500 V and 0.5 h for 250 Vh) to achieve a final 36 875 Vh (totally ~7.5 h).The IPG strip with the protein sample was equilibrated in a reducing equilibrium buffer (15 mL; 15 min) that contained 375 mmol/L Tris-HCl pH 8.8, 6 mol/L urea, 20 g/L SDS, 20% (v/v) glycerol, 20 g/L DTT, and a trace of bromophenol

blue. The IPG strip was equilibrated in an alkylation equilibrium solution (15 mL; 15 min) that contained 25 g/L iodoacetamide instead of 20 g/L DTT. For the second dimension – SDS-PAGE - an Ettan™ DALT II system (Amersham Pharmacia Biotech), which contains up to 12 gels at a time, 12% PAGE gels (250 mm × 215 mm × 1.0 mm) were used. The gels were run in 10 L of Tris-glycine-SDS electrophoresis buffer that contained 25 mmol/L Tris-base, 192 mmol/L glycine, and 1 g/L SDS with the following conditions: constant 2.5 W/gel for 30 min, and constant 10 W/gel for 340 min. The gel solution for three gels was filled into an Ettan™ DALTsix multi-gel caster (Amersham BioSciences) after mixing 90 mL of 400 g/L acrylamide/bis-acrylamide (29:1 by weight; cross-linking ratio = 3.3%), 75 mL of 1.5 mol/L Tris-HCl pH 8.8, 135 mL of distilled and deionized water, 1.5 mL of 100 g/L ammonia persulfate, and 75 μL of tetramethyl ethylenediamine (TEMED). The 2DE-separated proteins were visualized with Coomassie Brilliant Blue G250 staining [32]. The scanned gel images were analyzed with PDQuest system (Bio-Rad, version 7.1, Hercules, CA) [7, 33, 34].

2.3 Mass spectrometry analysis

The randomly selected protein-gel spots (Figs. 1 and 2) were subjected to in-gel digestion by trypsin (0.32–0.48 μg/spot; Cat No. V5111, Promega) [7]. The tryptic peptide mixture was purified with a ZipTipC18 microcolumn (cat. No. ZTC18S096; Millipore). The purified tryptic peptide mixture was analyzed with LC-ESI-qTOF (micromass Q-Tof micro™, Waters, Micromass UK Limited, Manchester, UK), LTQ-OrbiTrap Velos Pro ETD (Thermo Scientific, Waltham, MA, USA) or MALDI-TOF-TOF MS/MS (Autoflex III, Bruker, USA). MS data are available via ProteomeXchange with identifier PXD004930.

2.3.1 Protein identification for glioblastoma samples

The proteins present in 2DE gel spots of glioblastoma tissue (Fig. 1) were MS-analyzed.

(i) The randomly selected 21 gel spots from the pooled sample of three matched gels (Red number-labeled spots, Fig. 1) were analyzed with an old LC-ESI-quadrupole-time of flight (LC-ESI-qTOF) MS. First the samples were concentrated and desalted by a PepMap C18 pre-column (300 μm i.d. × 5 mm length, LC Packings). Then the peptides were separated in a reversed phase (RP) PepMap C18 column (75 μm i.d. × 15 cm length, LC Packings) with a linear gradient from 100% solvent A (0.1% formic acid) to 50% solvent B (0.1% formic acid in 95% acetonitrile) over 60 min and then 50–90% solvent B for 5 min. Data-dependent mode was applied for MS/MS analysis using the top four largest intensity precursor ions from each survey scan, and managed with MassLynx v 4.0 software. The generated .pkl file from each .raw file and the online MASCOT search engine (version 2.3) were used to search a protein against the Swiss-Prot 2014–7 database (release date July 1, 2014; 546000 sequences; 194259968 residues; Homo sapiens 20274 sequences). Search parameters were a mass tolerance of ± 0.5 Da #13C(0) for precursor ions (MS) and ± 0.5 Da for product ions (MS/MS), allowance for up to one trypsin miscleavage, fixed modification of carbamidomethylation (C), variable modifications consisting of oxidation (M) were used. An individual ion score > 33 indicated significant identity ($p < 0.05$). Also, a blank gel on the margin on a 2DE gel was analyzed in parallel to remove any contaminating protein, including trypsin and keratin.

(ii) The randomly selected 18 gel spots each pooled from three gels (Green number-labeled spots, Fig. 1) and five gel spots (Spots L1-L5, Fig. 1) from a single gel were analyzed with an LC-ESI-LTQ OrbiTrap Velos MS/MS. Each tryptic peptide mixture was analyzed twice with an EASY-nano LC system (Proxeon Biosystems, Odense, Denmark) coupled online with an LTQ-OrbiTrap Velos mass spectrometer (Thermo Scientific, Waltham, MA) under the conditions of a PepMap C18 trap column (300 μm i.d. × 5 mm length; Dionex Corp.), an RP PepMap column (75 μm i.d. × 15 cm length;

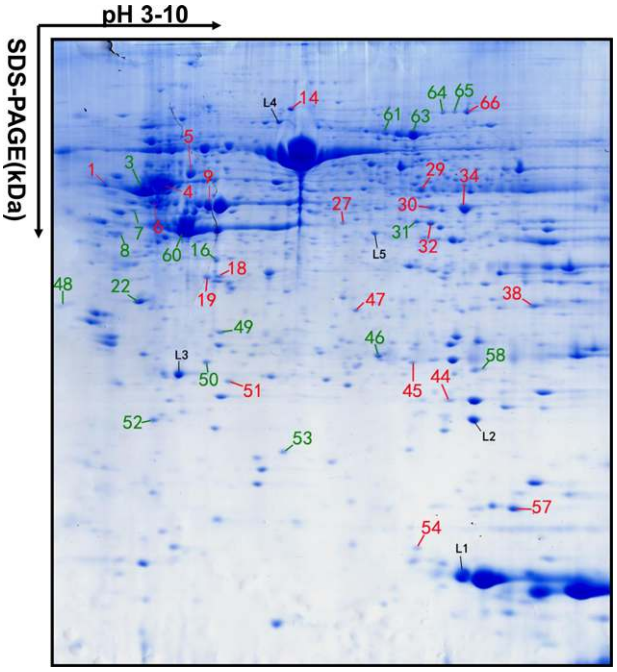

Figure 1. 2DE image of a glioblastoma proteome (500 μg protein per 2DE gel). The separated proteins were stained with Coomassie Blue G250 [32]. Each red- or green-number labeled spot was combined from three matched gel spots for MS/MS analysis. Spots L1-L5 came from one gel for MS/MS analysis. The randomly selected 21 red number-labeled spots were analyzed with LC-ESI-qTOF MS/MS. The randomly selected 18 green number-labeled spots were analyzed with LC-ESI-OrbiTrap Velos MS/MS. The randomly selected spots L1-L5 were analyzed with LC-ESI-OrbiTrap Velos MS/MS and MALDI-TOF-TOF MS/MS.

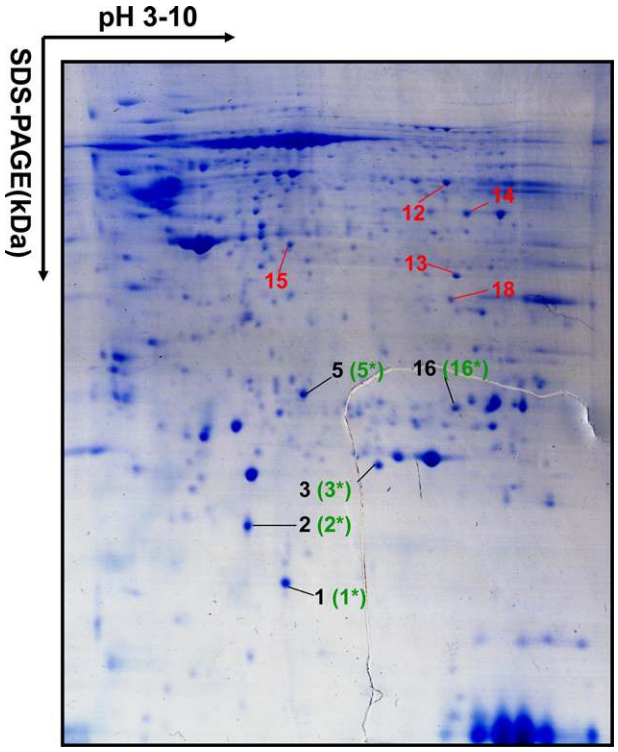

Figure 2. Two-dimensional gel electrophoresis image of a pituitary adenoma proteome (500 μg protein per 2DE gel). The separated proteins were stained with Coomassie Blue G250 [32]. Each black-number labeled spot was from single gel spot for MS/MS analysis. Each green-number labeled spot was combined from two matched gel spots for MS/MS analysis. Each red-number labeled spot was combined from three matched gel spots for MS/MS analysis. Spots 1*, 2*, 3*, 5*, and 16* matched to the corresponding spots 1, 2, 3, 5, and 16, but combined additional two gels. All selected spots were analyzed with LC-ESI-OrbiTrap Velos MS/MS and MALDI-TOF-TOF MS/MS.

DionexCorp., Sunnyvale, CA, USA) for separation with a gradient at 98% solvent A (0.1% formic acid) and 2% solvent B (0.1% formic acid in 100% acetonitrile) for 5 min, 2 to 40% solvent B for 45 min, 40 to 95% solvent B for 5 min, and 95% solvent B for 10 min (a total of 65 min at 300 nL/min), positive-ion mode with data-dependent automatic survey MS scan and tandem mass spectra acquisition modes, each MS scan in the OrbiTrap analyzer (mass range, m/z 350–1800; resolution = 100 000 at m/z 400) followed by MS/MS of the seven most-intense ions in the LTQ, fragmentation with collision-induced dissociation (CID), managed with Xcalibur software v.2.1 (Thermo Scientific), and data processing with Proteome Discoverer software v.1.3 beta (Thermo Scientific). In addition, an LC washing procedure with a gradient at 5–35% solvent B for 20 min, 35–95% solvent B for 2 min, and 95% solvent B for 8 min (a total of 30 min at 300 nL/min was run after LC-ESI-MS/MS analysis of every two to three samples. For spots L1-L5, the mass data were searched by SEQUEST that is contained in the Thermo Proteome Discoverer 1.3 (version No. 1.3.0.339) against the human protein database Uniprot version 201410.1_HUMAN.fasta with search parameters including the enzyme (trypsin), taxonomy (Homo sapiens), peptide tolerance (±10 ppm), MS/MS tolerance (±0.8 Da), peptide charge (2+, 3+, and 4+), one missed cleavage site allowed, carbamidomethylation of cysteine as a fixed modification, and oxidation of methionine and nitration tyrosine as variable modifications. The numbers of proteins, protein groups, and peptides were filtered for false discovery rates (FDR) <1% and only peptides with rank 1. A protein was identified with the parameter PSM ≥ 1 (PSMs: The total number of identified peptide sequences (PSMs) for the protein, including those redundantly identified). A minimum of two peptides per protein was accepted for identity with the Proteome Discoverer. The protein lists from two technical repetitions were merged, and duplicate proteingroups were removed. For those 18 spots (Green number-labeled spots, Fig. 1), each .raw file was converted to .mgf file for MASCOT (version 2.3.02) searching against human protein database (uniprothuman_20161031.fasta that contains 70940 sequences and 23897047 residues) with fixed modifications (carbamidomethyl (C)), variable modifications (deamidated (NQ) and oxidation (M)), trypsin digestion with maximum two missed cleavages, precursor ion mass tolerance 10 ppm, daughter ion mass tolerance 0.8 Da, monoisotopic peak, and significance threshold of 0.05. Each protein had at least two unique peptides identified. Exponentially modified protein abundance index (emPAI) was calculated for each protein in an analyzed spot to estimate the amount of each protein in a 2DE spot.

(iii) The five gel spots (Spots L1-L5, Fig. 1) from a single gel were also analyzed with MALDI-TOF-TOF MS/MS. Each tryptic peptide mixture (0.5 μL) was spotted on a 384-well MALDI-plate, immediately spotted 0.5 μL of saturated CHCA matrix in 0.1% TFA/50% acetonitrile on the top of the peptide sample, and dried in the air. The air-dried peptides were analyzed with MALDI-TOF-TOF MS/MS. The instrument was managed automatically, 100 laser shots were used to acquire one MS1 spectrum, and six repeated MS1 spectra were accumulated into one synthetic MS1 spectrum. Four most-intense precursor ions from each synthetic MS1 spectrum were selected for MS/MS analysis. For each precursor ion, 100 laser shots were used to acquire one MS/MS spectrum, and four repeated MS/MS spectra were accumulated into one synthetic MS/MS spectrum. The generated .mgf files of the MS and all MS/MS spectra together were used for protein search with online MASCOT MS/MS ion search (version 2.3) against protein database Swiss-Prot 091215 (513877 sequences; 180750753 residues; *Homo sapiens* 513877 sequences; July 20, 2015) with searching parameters (MS/MS Ion Search, enzyme trypsin, fixed modification carbamidomethyl (C), variable modification Oxidation (M), monoisotopic mass value, peptide mass tolerance ± 100 ppm, fragment mass tolerance ± 0.7 Da, and max missed cleavage1). A statistical probability based

on Mowse score was used to determine the identified protein. Ions score was −10*Log(P), where P was the probability that the observed match was a random event, and protein score > 70 was significant ($p < 0.05$).

2.3.2 Protein identification for pituitary adenoma

The proteins that were contained in 2DE gel spots in separation of pituitary adenoma tissue proteomes (Fig. 2) were MS-analyzed by LTQ-OrbiTrap Velos and MALDI-TOF-TOF MS/MS with the same MS parameters as described in the Section 2.3.1, including the randomly selected five gel spots from the pooled sample of three matched spots (red-number labeled spots), five gel spots from the pooled sample of two matched spots (green-number labeled spots), and five gel spots from a single gel (black-number labeled spots) (Green-spots 1*, 2*, 3*, 5*, and 16* matched to black-spots 1, 2, 3, 5, and 16, respectively). The LTQ-OrbiTrap Velos MS/MS data were used to search the human database with Mascot as described in Section 2.3.1 (ii). The MALDI-TOF-TOF MS/MS data were used to search the human database with MASCOT as described in Section 2.3.1 (iii).

3 Results and discussion

Approximately 1100 protein spots were detected in each Coomassie-stained 2DE gel (Figs. 1 and 2). Most of them were distributed within a range of *pI* 4–8 and *Mr* of 15–150 kDa. We randomly picked out 44 spots from glioblastoma and ten spots from pituitary adenoma tissue for MS analysis with different mass spectrometers and evaluated the MS data with the aim to identify as much as possible proteins and to see protein speciation.

3.1 Comparison of different mass spectrometers

For glioblastoma (Table 1), peptides from five spots (Spots L1-L5) from a single gel and 18 gel spots (Spots 3, 7, 8, 16, 22, 31, 46, 48, 49, 50, 52, 53, 58, 60, 61, 63, 64, and 65) from three gels were analyzed with LC-MS/MS on the 2-year-old OrbiTrap Velos mass spectrometer. Peptides from 21 gel spots (Spots 1, 4, 5, 6, 9, 14, 18, 19, 27, 29, 30, 32, 34, 38, 44, 45, 47, 51, 54, 57 and 66) from three gels each were analyzed with a 10-year-old qTOF mass spectrometer. Peptides from five spots (Spots L1-L5) were also analyzed with MALDI-TOF-TOF MS/MS. For the same amount of glioblastoma protein samples (Table 1), much more proteins were identified in OrbiTrap Velos relative to QTOF mass spectrometer or MALDI-TOF-TOF mass spectrometer: an average of 42 proteins/spot vs. 1 protein/spot for a single spot, and depending on the identification criteria an average of up to 334 proteins/spot with at least 1 unique peptide identified, 230 proteins/spot with at least 2 unique peptides identified and 169 proteins/spot with at least three unique peptides identified versus 5 proteins/spot for a pooled sample of three matched spots. The one unique peptide identification had a much larger potential for more proteins to increase the coverage of a proteome, however, it also increases the possibility of wrong identifications and needs further confirmation. Here, all identified non-redundant proteins with at least two unique peptides identified ($n = 2492$) are listed in Supporting Information Table 1.

For pituitary adenomas, peptides from five gel spots (Spots 1, 2, 3, 5, and 16) from a single gel, 5 gel spots (Spots 1*, 2*, 3*, 5*, and 16*) from the pooled sample of two matched spots, and five gel spots (Spots 12, 13, 14, 15, and 18) from the pooled sample of three matched spots were analyzed with LC-ESI-OrbiTrap Velos and MALDI-TOF-TOF MS/MS. For the same amount of pituitary adenoma protein samples (Table 2), a much higher number of proteins was identified in the super-high sensitivity OrbiTrap Velos relative to the conventional MALDI-TOF-TOF mass spectrometer. The number of identified proteins increases in average from 41 proteins/spot to 63 and 96 by allowing at least three, two and one unique peptides identification for single spot identification, respectively. These numbers increase further to 50, 71, and 110, if two spots are pooled, and to 82, 118, and 167 protein spots, if three spots are pooled. All identified non-redundant proteins with at least two peptide identified ($n = 761$) are listed in Supporting Information Table 2.

These results clearly demonstrated that the higher sensitivity mass spectrometer OrbiTrap Velos significantly increased the number of identified proteins under the condition of the same amount of protein samples relative to the old and lower sensitivity mass spectrometers such as qTOF and MALDI-TOF-TOF MS.

3.2 Protein speciation

Multiple factors including splicing and post-translational modification lead to multiple protein speciation corresponding to each gene-coded protein. (i) For glioblastoma, a total of 2492 unique proteins with at least two unique peptides identified was identified from only 44 spots distributed over the whole 2DE pattern (Fig. 1 and Supporting Information Table 1); of them, 802 proteins (802/2492 = 32%) occurred in multiple spots with a range of 2 to 21 spots (Table 3; Supporting Information Table 1), and 23 proteins occurred in more than ten spots, for example, serum albumin occurring in 21 spots, actin cytoplasmic 1 occurring in 17 spots, glial fibrillary acidic protein occurring in 16 spots, 60 kDa mitochondrial heat shock protein, heterogeneous nuclear ribonucleoprotein U and dihydropyrimidinase-related protein 2 occurring in 13 spots, fibrinogen alpha chain, fibronectin, tubulin beta chain, nestin and hemoglobin subunit beta occurring in 12 spots, and MAP7 domain-containing protein 3, alpha-enolase, and moesin occurring in 11 spots. (ii) For pituitary adenoma, a total of 761 unique proteins with at least two unique peptides identified was identified from only ten spots distributed over the whole 2DE pattern (Fig. 2 and Supporting Information Table 2); of them, 157 proteins

Table 1. The number of proteins identified in each 2D gel-spot with mass spectrometry analysis of glioblastoma tissue

LC-ESI-QTOF (Pooled 3 matched spots)		LC-ESI-OrbiTrap vs. MALDI-TOF-TOF (One single gel spot)			LC-ESI-OrbiTrap (Pooled 3 matched spots)			
						Number of proteins identified		
Spot No.	Number of proteins identified	Spot No.	Number of proteins identified (OrbiTrap)	Number of proteins identified (MALDI)	Spot No.	Uni_pep≥1	Uni_pep≥2	Uni_pep≥3
1	8	L1	25	1	3	424	289	231
4	1	L2	23	1	7	356	254	182
5	7	L3	59	0	8	408	280	205
6	6	L4	52	0	16	329	226	175
9	2	L5	51	1	22	352	234	162
14	6				31	410	294	222
18	8				46	296	192	136
19	5				48	328	230	166
27	2				49	392	267	171
29	14				50	365	252	187
30	2				52	273	184	126
32	7				53	199	120	80
34	7				58	368	237	169
38	2				60	225	134	85
44	2				61	372	287	234
45	6				63	251	177	137
47	3				64	326	242	186
51	5				65	347	255	200
54	0							
57	0							
66	1							
Total	94	Total	210	3	Total	6021	4154	3054
Average	5 (94/19)	Average	42 (210/5)	1 (3/3)	Average	334 (6021/18)	230 (4154/18)	169 (3054/18)

Note: Uni_pep = the number of unique peptides. For a pool of three matched spots (right column), proteins were analyzed by OrbiTrap Velos and Mascot, and identified with the criterion of at least one, two and three unique peptides and significance $p < 0.05$. For one single gel spot (middle column), proteins were analyzed by OrbiTrap Velos and Sequest, and identified with the criterion PSM ≥ 1 (PSMs: The total number of identified peptide sequences (PSMs) for the protein, including those redundantly identified).

(157/761 = 21%) occurred in multiple spots with a range of two to ten spots (Table 3; Supporting Information Table 2). Here, one would find that not as much protein species in pituitary adenoma were found as in glioblastoma. One of possible reasons is that the preparation of pituitary adenoma protein sample used the highly acidic homogenizing buffer (2 mol/L acetic acid), which would cause protein degradation by cleaving between D and P in all proteins. Another main factor might be derived from two different experienced technicians, the glioblastoma experiments were done by a very experienced technician, while pituitary adenoma experiments were done by a novice to possibly lead a loss of sample in the entire experimental process from protein extraction to preparation of tryptic peptide for MS analysis, a much less number of proteins was obtained in pituitary adenoma (average of 118 proteins/spot for the pool of three matched spots, Table 2) than in glioblastoma (average of 230 proteins/spot for the pool of three matched spots, Table 1). Anyway, the present data showed clearly that protein speciation can be recognized already with the low number and randomly picked spots analyzed in this study. To unravel protein speciation, a complete analysis of all spots in a 2DE gel is necessary as was performed earlier [21].

3.3 How many proteins can be found in one spot?

The number of proteins identified with LC-ESI-QTOF, and MALDI-TOF-TOF was much less than that identified with LC-ESI-OrbiTrap Velos (Tables 1 and 2). Therefore, here only the results from OrbiTrap Velos with at least two unique peptides identified were used to discuss the number of proteins identified in one spot. For glioblastoma (Table 1), a total of 210 proteins with an average of 42 proteins per spot were identified from five randomly selected gel spots (Spots L1–L5) from a single gel with the OrbiTrap Velos; and applying the MASCOT criteria for identification with the additional restriction at least two unique peptides identified, a total of

Table 2. The number of proteins identified in each 2DE gel spot with MS analysis of pituitary adenoma tissue

LC-ESI-OrbiTrap versus MALDI-TOF-TOF (One single gel spot)					LC-ESI-OrbiTrap versus MALDI-TOF-TOF (Pooled 2 matched spots)				LC-ESI-OrbiTrap versus MALDI-TOF-TOF (Pooled 3 matched spots)				
	Number of proteins identified (OrbiTrap)					Number of proteins identified (OrbiTrap)				Number of proteins identified (OrbiTrap)			
Spot No.	Uni_pep ≥ 1	Uni_pep ≥ 2	Uni_pep ≥ 3	Number of proteins identified (MALDI)	Spot No.	Uni_pep ≥ 1	Uni_pep ≥ 2	Uni_pep ≥ 3	Number of proteins identified (MALDI)	Spot No.	Uni_pep ≥ 1	Uni_pep ≥ 2	Uni_pep ≥ 3	Number of proteins identified (MALDI)
1	115	76	50	2	1*	82	50	34	1	12	144	102	68	2
2	55	32	24	1	2*	75	48	35	1	13	141	91	73	0
3	92	60	39	0	3*	145	93	67	2	14	184	141	96	1
5	130	86	56	1	5*	124	78	52	1	15	231	156	110	2
16	91	61	38	1	16*	124	84	65	1	18	138	98	65	1
Total	483	315	207	5	Total	550	353	253	6	Total	838	588	412	6
Average	96 (483/5)	63 (315/5)	41 (207/5)	1 (5/4)	Average	110 (550/5)	71 (353/5)	50 (253/5)	1 (6/5)	Average	167 (838/5)	118 (588/5)	82 (412/5)	1 (6/4)

Note: Spots 1*, 2*, 3*, 5*, and 16* matched to the corresponding spots 1, 2, 3, 5, and 16, but combined additional 2 gels. Uni_pep = the number of unique peptides. For each ObriTrap Velos analysis, proteins were searched with Mascot, and identified with the criterion of at least one, two and three unique peptides and significance $p < 0.05$.

Table 3. Protein speciation recognized among 44 randomly picked spots in the glioblastoma 2DE map and ten randomly picked spots in the pituitary adenoma 2DE map in this study

Glioblastoma (total 44 spots analyzed)		Pituitary adenoma (total ten spots analyzed)	
The number of proteins	Speciation (how many spots occurred)	The number of proteins	Speciation (how many spots occurred)
1	21	2	10
1	17	1	9
2	16	2	8
3	13	6	7
5	12	6	6
3	11	9	5
8	10	13	4
10	9	30	3
13	8	88	2
18	7	604	1
24	6		
41	5		
87	4		
150	3		
436	2		
1690	1		
Total: 2492		Total: 761	

4154 proteins with an average of 230 proteins per spot were identified from 18 randomly selected gel spots from a pooled sample of three matched gels with the new OrbiTrap Velos MS/MS. For pituitary adenoma analyzed with LC-ESI-OrbiTrap Velos MS/MS (Table 2), a total of 315 proteins with an average of 63 proteins per spot with at least two unique peptides was identified from five selected spots from a single spot; a total of 353 proteins with an average of 71 proteins per spot with at least two unique peptides was identified from five selected spots from two matched spots; and a total of 588 proteins with an average of 118 proteins per spot with at least two unique peptides was identified from five selected spots from a pool of three matched spots.

These results clearly demonstrated that each 2DE gel spot contained many proteins and not only one or two. Compared to the number of proteins per spot from a single spot (Glioblastoma: average 42 proteins/spot; Pituitary adenoma: average 63 proteins/spot) and from a pooled sample of two matched spots (Pituitary adenoma: average 71 proteins/spot), much more proteins were obtained from a pooled sample of three matched spots (Glioblastoma: average 230 proteins/spot; Pituitary adenoma: average 118 proteins/spot) when they were analyzed on the same OrbiTrap Velos. Even with the unpooled, one single spot analysis it is obvious that many not only one or two proteins were identified in a 2DE spot in the analysis of a complex cancer tissue proteome. Moreover, more low-abundance proteins in a 2D gel spot

were definitely identified with a pool of two or three matched spots relative to a single gel spot. Therefore, the present data clearly demonstrate that a Coomassie Blue-stained 2DE spot contains an average of at least 42 proteins/spot in an analysis of a glioblastoma proteome and an average of at least 63 proteins/spot in an analysis of a pituitary adenoma proteome.

3.4 Estimation of the ratio of each protein in the analyzed spots

Exponentially modified protein abundance index (emPAI) is used to estimate absolute protein amount in proteomics, which is equal to 10^{PAI} minus one [35]. The protein abundance index (PAI) is the number of identified tryptic peptides divided by the number of theoretically observable tryptic peptides [36]. The emPAI value was calculated for each protein identified with at least two unique peptides from 18 glioblastoma spots (Spots 3, 7, 8, 16, 22, 31, 46, 48, 49, 50, 52, 53, 58, 60, 61, 63, 64, and 65 from 3 matched gels) with OrbiTrap Velos (Supporting Information Table 3) and from 15 pituitary adenoma spots (Spots 12, 13, 14, 15, and 18 from three matched gels; Spots 1*, 2*, 3*, 5*, and 16* from two matched gels; and Spots 1, 2, 3, 5, and 16 from a single gel) with OrbiTrap Velos (Supporting Information Table 4). The proteins identified in each spot were categorized by the range of emPAI value (Table 4). The result showed that most of spots contained multiple proteins (12–41 proteins in glioblastoma; 2–16 proteins in pituitary adenoma except for spots 2, 3, and 1* with only one protein) with middle to high abundance (emPAI $\geq$ 1), multiple proteins (53–139 proteins in glioblastoma; 16–73 proteins in pituitary adenoma) with middle abundance (0.1 $\leq$ emPAI $<$ 1), and multiple proteins (47–150 proteins in glioblastoma; 15–67 proteins in pituitary adenoma) with low abundance (0.01 $\leq$ emPAI $<$ 0.1). Moreover, glioblastoma spots 7, 60, and 61 were taken as examples to analyze in detail the ratio of each protein in each analyzed spot (Table 5; Supporting Information Table 3). It clearly showed 27 middle to high-abundance proteins (emPAI $\geq$ 1) with 99.30% of total emPAI value, 105 middle-abundance proteins (0.1 $\leq$ emPAI $<$ 1) with 0.60% of total emPAI value, and 122 low-abundance proteins (0.01 $\leq$ emPAI $<$ 0.1) with 0.10% of total emPAI value in glioblastoma spot 7; 12 middle to high-abundance proteins (emPAI $\geq$ 1) with 99.97% of total emPAI value, 60 middle-abundance proteins (0.1 $\leq$ emPAI $<$ 1) with 0.02% of total emPAI value, and 62 low-abundance proteins (0.01 $\leq$ emPAI $<$ 0.1) with 0.01% of total emPAI value in glioblastoma spot 60; and 35 middle to high-abundance proteins (emPAI $\geq$ 1) with 95.29% of total emPAI value, 116 middle-abundance proteins (0.1 $\leq$ emPAI $<$ 1) with 4.11% of total emPAI value, and 136 low-abundance proteins (0.01 $\leq$ emPAI $<$ 0.1) with 0.60% of total emPAI value in glioblastoma spot 61. These results clearly demonstrate that a large number of low-abundance proteins were identified in each 2DE gel spot with high-sensitivity OrbiTrap Velos mass spectrometer.

3.5 Protein numbers identified depending on the identification criteria

Each identified protein as described above had at least two unique peptides for OrbiTrap Velos MS/MS data in combination with MASCOT search in analyses of glioblastoma 2DE spots 3, 7, 8, 16, 22, 31, 46, 48, 49, 50, 52, 53, 58, 60, 61, 63, 64, and 65, and of pituitary adenoma 2DE spots 1, 2, 3, 5, 16, 1*, 2*, 3*, 5*, 16*, 12, 13, 14, 15, and 18 (Supporting Information Tables 1–4; Tables 1–5). Here, the protein numbers were further investigated with different criteria concerning minimal number of unique peptides from one to three (Tables 1 and 2; Supporting Information Tables 5 and 6). For glioblastoma (a pooled sample of three matched spots for each OrbiTrap Velos analysis) (Table 1 and Supporting Information Table 5), there were an average of 334 proteins/spot with the criterion of at least one unique peptide identified, of 230 proteins/spot with the criterion of at least two unique peptides identified, and of 169 proteins/spot with the criterion of at least three unique peptides identified, among 18 analyzed spots. For pituitary adenoma (one single gel spot for each OrbiTrap Velos analysis) (Table 2 and Supporting Information Table 6), there was an average of 96 proteins/spot with the criterion of at least 1 unique peptides identified, of 63 proteins/spot with the criterion of at least two unique peptides identified, and of 41 proteins/spot with the criterion of at least three unique peptides identified, among five analyzed spots. For pituitary adenoma (a pooled sample of two matched spots for each OrbiTrap Velos analysis) (Table 2 and Supporting Information Table 6), there was an average of 110 proteins/spot with the criterion of at least one unique peptide identified, of 71 proteins/spot with the criterion of at least two unique peptides identified, and of 50 proteins/spot with the criterion of at least three unique peptides identified, among five analyzed spots. For pituitary adenoma (a pooled sample of three matched spots for each OrbiTrap Velos analysis) (Table 2 and Supporting Information Table 6), there was an average of 167 proteins/spot with the criterion of at least 1 unique peptides identified, of 118 proteins/spot with the criterion of at least two unique peptides identified, and of 82 proteins/spot with the criterion of at least three unique peptides identified, among five analyzed spots. These results show that the criteria concerning minimal number of unique peptides from one to three significantly affected the number of identified proteins. The number of proteins identified in each spot was significantly decreased from the soft criterion "at least one unique peptide" to the strong criterion "at least three unique peptides" (Tables 1 and 2). In order to maximally avoid a wrong identification with the one unique peptide criterion and the possible noise derived from highly sensitive mass spectrometer, the identification criterion "at least two unique peptides" was used in this study.

However, although the number of wrong identifications is surely larger for the one unique peptide identifications, many of the one peptide identifications can be expected to be still correct. One can find that many proteins with one unique peptide identification have very high protein amount,

Table 4. The number of proteins in each emPAI range, which is the estimation of the ratio of each protein with at least 2 unique peptides identified in the analyzed 2DE spot with OrbiTrap Velos MS/MS

Glioblastoma								Pituitary adenoma							
	The number of proteins by emPAI range								The number of proteins by emPAI range						
Spot No.	0.01–0.1 (n)	0.1–1 (n)	1–10 (n)	10–100 (n)	> 100 (n)	Total (n)	The range of p*I*	Spot No.	0.01–0.1 (n)	0.1–1 (n)	1–10 (n)	10–100 (n)	> 100 (n)	Total (n)	The range of p*I*
3	146	105	28	3	7	289	4.31–11.36	1	19	51	5	1	–	76	4.30–11.36
7	122	105	19	6	2	254	4.35–11.52	2	15	16	-	1	–	32	4.35–10.05
8	124	130	22	2	2	280	4.21–11.52	3	29	30	-	1	–	60	4.30–11.36
16	97	89	32	8	–	226	4.35–11.67	5	28	49	9	–	–	86	4.34–11.52
22	81	122	27	3	1	234	4.32–11.27	16	22	37	2	–	–	61	4.35–11.36
31	129	125	36	4	–	294	4.35–11.25	1*	20	29	1	–	–	50	4.64–10.05
46	62	106	19	4	1	192	4.53–11.01	2*	18	27	2	1	–	48	4.78–10.05
48	68	132	29	1	–	230	3.77–11.36	3*	44	42	6	–	1	93	4.30–10.71
49	94	139	29	2	3	267	4.35–11.88	5*	29	42	7	–	–	78	4.30–10.05
50	85	126	40	1	–	252	4.26–11.36	16*	26	43	14	1	–	84	4.64–12.27
52	51	99	33	-	1	184	4.29–11.88	12	43	48	10	1	–	102	4.43–11.53
53	47	61	12	-	–	120	4.48–11.88	13	45	43	2	1	–	91	4.26–10.05
58	71	129	31	4	2	237	4.35–11.64	14	64	72	4	1	–	141	4.35–11.84
60	62	60	6	4	2	134	4.35–10.18	15	67	73	15	1	–	156	4.43–10.05
61	136	116	28	6	1	287	4.51–10.85	18	50	37	10	1	–	98	4.47–10.29
63	97	53	24	2	1	177	4.35–10.29								
64	130	93	18	1	–	242	4.39–11.77								
65	150	85	19	1	–	255	4.47–11.77								

Note: For glioblastoma, each spot was analyzed with combined three matched gel spots. For pituitary adenoma, spots 1, 2, 3, 5, and 16 were analyzed with single gel spot; Spots 1*, 2*, 3*, 5*, and 16* were analyzed with combined two matched gel spots, and matched to the corresponding spots 1, 2, 3, 5, and 16, respectively; and spots 12, 13, 14, 15, and 18 were analyzed with combined three matched gel spots. Each identified protein had at least two unique peptides identified.

Table 5. The estimated ratio of each protein in the selected glioblastoma 2DE spot

	Spot 7			Spot 60			Spot 61		
emPAI range	N	ΣemPAI	Ratio (%)	n	ΣemPAI	Ratio (%)	n	ΣemPAI	Ratio (%)
First main component	1	5579.90	90.51	1	114352.20	98.86	1	656.21	71.59
Second main component	1	294.35	4.77	1	1203.98	1.04			
10–100	6	205.11	3.32	4	62.83	0.05	6	131.29	14.32
1–10	19	43.25	0.70	6	22.66	0.02	28	86.00	9.38
0.1–1	105	37.23	0.60	60	19.91	0.02	116	37.70	4.11
0.01–0.1	122	5.24	0.10	62	2.92	0.01	136	5.42	0.60
Total	254	6165.08	100.00	134	115664.50	100.00	287	916.62	100.00

Note: Each protein was identified with the criterion of at least two unique peptides and significance $p < 0.05$.

and which are even the main component in that analyzed spot (Supporting Information Tables 5 and 6). For example in glioblastoma analysis, 40S ribosomal protein SA fragment (Swiss-Prot: F8WD59) with one unique peptide identification had the highest emPAI value of 2714.12 among 408 proteins identified in spot 8 (Supporting Information Tables 5–3); ubiquitin carboxyl-terminal hydrolase (Swiss-Prot: D6RE83) with one unique peptide identification had the highest emPAI value of 3952.35 among 365 proteins identified in spot 50 (Supporting Information Tables 5–10); nucleoside diphosphate kinase (Swiss-Prot: P15531) with one unique peptide identification had the highest emPAI value of 732.56 among 199 proteins identified in spot 53 (Supporting Information Tables 5–12); the N(G),N(G)-dimethylarginine dimethylaminohydrolase (Swiss-Prot: B4DYP11) with one unique peptide identification had the highest emPAI value of 65.7 among 329 proteins identified in spot 16 (Supporting Information Tables 5–4); and actin cytoplasmic 2 (Swiss-Prot: P63261) with one unique peptide identification had the second highest emPAI value of 106022.6 among 225 proteins identified in spot 60 (Supporting Information Tables 5–14). Therefore, the identification criterion with minimal two unique peptides in this study may actually underestimate the real protein number in the analyzed spots.

3.6 Comparison of protein number identified in glioblastoma 2DE pattern spot 7 from this study with that in human HeLa cell 2DE pattern spot 116 from a previous study

Vimentin is the main component in the list of proteins identified in glioblastoma 2DE pattern spot 7 from this study and in human HeLa cell 2DE pattern spot 116 from documented Thiede's publication [21] (Proteome 2D-PAGE database at MPIIB, human/HeLa cells/Cell proteins/Cell culture; http://web.mpiib-berlin.mpg.de/cgi-bin/pdbs/2d-page/extern/menu_frame.cgi).The LC-ESI MS/MS .raw files acquired from glioblastoma 2DE pattern spot 7 and HeLa cell 2DE pattern spot 116 were converted into .mgf files with the same parameters. The generated .mgf files were used to identify proteins in a search against the same human protein database with Mascot software applying the same criteria, trypsin, missing cleavage, mass tolerance, and statistical significance; however, different modifications between HeLa cells (SILAC labeling and variable propionamide C) and the present investigation (fixed carbamidomethylation) had to be used in addition to oxidation of Met (ox) and pyroglutamination of Gln, respectively. Also, the protein searching results from two repeated runs of HeLa cell spot 116 were combined. A total of 254 proteins were identified with a criterion of at least two unique peptides from glioblastoma 2DE pattern spot 7 (Table 6; Supporting Information Table 7), including vimentin (Swiss-Prot Access No. P08670) that was the main protein with an emPAI value of 5579.90 in this spot, 1 protein with the emPAI value of 294.35, six proteins with the emPAI values in the range from 10 to 100, 19 proteins with the emPAI values in the range from 1 to 10, 27 proteins with the emPAI values in the range from 0.45 to 1, 78 proteins with the emPAI values in the range from 0.10 to 0.45, 54 proteins with the emPAI values in the range

Table 6. Comparison of protein number identified in glioblastoma 2DE spot 7 in this study and human HeLa cell 2DE spot 116 in a previous study [25]

	The number of proteins in glioblastoma spot 7		The number of proteins in HeLa cell spot 116	
emPAI	n	Marks	n	Marks
> 1000	1	Vimentin(P08670): emPAI = 5579.9		
100–1000	1	Glial fibrillary acidic protein(P14136): emPAI = 294.35		
10–100	6			
1–10	19		1	Vimentin (P08670): emPAI = 8.04)
0.45–1.00	27			
0.10–0.45	78		5	
0.05–0.10	54		1	
0.01–0.05	68			
Total	254		7	

Note: Each protein was identified with at least two unique peptides and a significant score ($p < 0.05$).

Table 7. Comparison of analysis parameters between the spot identification of this study and a previous study [25] of a spot in a comparable region of the gel

	Glioblastoma spot 7 in this study	HeLa cell spot 116 in the previous study [25]
First hit	Vimentin	Vimentin
Experiment type	one biol situation	SILAC, two biologic situations
Biologic sample	Tissue	Cell line
2DE type	2DE Immobiline	2DE Ampholyte
Gel size	18 × 25 × 0.1 cm	23 × 30 × 0.15 cm
Protein amount	500 μg × 3 (pooled three matched spots)	400 μg protein (a single gel spot)
LC gradient	0 to 2% B in 2 min, 2 to 40% B in 45 min, 40 to 95% B in 5 min, 95% B 10 min, total of 65 min at 300 nL/min	7 to 40% B in 17 min, 40–50% B in 3 min, total run time 33 min
Mass spectrometer	OrbiTrap Velos	LTQ-OrbiTrap XL
Size of mgf file	PF02.mgf: 164.824 MB	116.mgf: 0.806 MB

from 0.05 to 0.10, and 68 proteins with the emPAI value of 0.01-0.05. A total of seven proteins were identified with a criterion of at least two unique peptides from human HeLa cell 2DE pattern spot 116 (Table 6; Supporting Information Table 8), including vimentin (Swiss-Prot Access No. P08670) that was the main protein with an emPAI value of 8.04 in this spot, five proteins with the emPAI values in the range from 0.10 to 0.45, and one protein with the emPAI values in the range from 0.05 to 0.10. These data show clearly that much more significantly identified proteins were found in the glioblastoma 2DE spot 7 that contained vimentin as the main protein in this study compared to the previously documented human HeLa cell 2DE spot 116 that also contained vimentin as the main protein [21].

This huge difference in protein numbers between glioblastoma spot 7 in this study and HeLa cell spot 116 in the previous study [21] was obviously derived from different experiment parameters (Table 7), including different experimental type, biological sample, 2DE type, gel size, protein amount, LC gradient, and mass spectrometer. Compared to the previously documented HeLa cell experimental study [21], the smaller gel size, more protein amount, longer LC gradient strategy, and more sensitive peak detection criteria for OrbiTrap Velos in this study resulted in a much larger .mgf file (Glioblastoma .mgf file: 164.824 MB versus HeLa cell .mgf: 0.806 MB) leading to much more identified proteins. It implies that the above mentioned parameters are responsible for the increase of the number of proteins identified. Of course, the difference in efficacy of LC gradient strategy and mass spectrometer sensitivity between these two studies can only be confirmed by a direct comparison of the same amount of peptides from the same spot from the same gel.

3.7 Rationality and in-depth consideration of the many proteins identified within a spot

This present study clearly demonstrated that many proteins can be identified in a Coomassie-stained 2DE gel spot within an analysis of a complex human cancer tissue proteome. There are several factors influencing the number of detected proteins per spot:

(i) Protein speciation. It is well-known that the human genome contains about 20 000–25 000 genes [37]. In contrast, because of protein speciation the human proteome contains a much larger number of protein species as compared with the number of genes. For example, one billion potential protein species were already estimated for one histone [31]) due to post-translational modifications (estimated with the prerequisite of 32 binding sites with different modifications each) [28, 38, 39]. In addition, the copy number of different proteins is much different to result in a huge range of protein abundance, with an estimation of the dynamic range of protein amount for prokaryotes of 10exp6 and for eukaryotes of 10exp12 [30, 40, 41]. For this study, the dynamic range of identified protein amount in each spot was within 10exp2 to 10exp6. Thus, it can be expected that there are many proteins with very similar pI and Mr that co-migrate into a 2DE gel spot. High amounts of direct protein extractions or prefractionated, enriched protein extracts have to be analyzed to access the low abundant protein species.

(ii) Protein-loading amount per gel and pooling of matched gel spots. In order to be compatible with Coomassie staining, a high amount (500 μg) of protein extract was used for each 2DE analysis compared with earlier studies where only 70 μg protein/gel [7] or 150 μg protein/gel [12] were used, and spots from up to three gels were combined for one LC-MS/MS analysis. This enables the detection of lower abundance proteins. If the identification limit for a certain protein is 1 fmol it can be principally identified if present with 1 fmol in 500 μg protein. If only 250 μg protein is investigated it will not be identified because with 0.5 fmol it is below the identification limit.

(iii) Noise. Each analytical procedure is accompanied by noise. In proteomics sample preparation, protein separation and protein identification all produce noise. Sample preparation methods are prone to keratin contaminations. In the case of LC-MS for identification of the proteins, for example carryover effects cannot be ignored, especially if high-sensitive MS is used. This effect is reduced by washing the LC columns thoroughly between the runs. Here we want to focus on the contribution of 2DE to noise. As in every separation technique in 2DE the noise increases by increasing the protein amount to be separated. A spot represents the top of a two-dimensional Gaussian curve. These Gaussian curves may have toes that can extend very far from the top. Thus at high protein-loading amount, under a given spot there may be many proteins coming from the Gaussians of

other spots. Proteins of neighboring spots in a 2DE pattern contribute more to this cross-Gaussians spreading than far-away spots, depending on their amount. Further noise is provoked by pooling of spots, as cutting several spots increases both the amount of proteins, and thus of contaminating ones, and also increases the likelihood that different parts of the spots might be taken on the three picks. An earlier study showed that each protein species is distributed as a Gaussian curve and it cannot be excluded that parts of a certain protein species occur in a nearby protein spot resulting in a neighbor spot contamination [42].

For example, the glioma samples spots 3, 7, 8, 16, 22, 31, and 46 were automatically run sequentially with a 30-min LC washing described in Section 2.3.1.ii after every two to three samples. Venn diagrams (Fig. 3) show the overlap of protein identified among sample spots 3, 7, 8, and 16, and among sample spots 16, 22, 31, and 46. A total of 71 overlapped proteins was found between spots 3 (71/289 = 24%) and 7(71/254 = 28%), 71 overlapped proteins between spot 7 (71/254 = 28%) and 8 (71/280 = 25%), and 44 overlapped proteins between spot 8 (44/280 = 15%) and 16 (44/226 = 19%) (Fig. 3A). A total of 46 overlapped proteins was found between spots 16 (46/226 = 20%) and 22 (46/234 = 19%), 34 overlapped proteins between spot 22 (34/234 = 15%) and 31 (34/294 = 12%), and 33 overlapped proteins between spot 31 (33/294 = 11%) and spot 46 (33/192 = 17%) (Fig. 3B). These findings clearly show overlapping proteins in the range between11 and 28% between two sequentially run samples. An interesting phenomenon is that the distance among spots 3, 7, 8, and 16 is shorter than the distance among spots 16, 22, 31, and 46 (Fig. 1), and the overlapping rate (range of 19–28%) among spots 3, 7, 8, and 16 is greater than the overlapping rate (11–20%) among spots 16, 22, 31, and 46 (Fig. 3), which shows the influence of 2DE. However, one must realize that those overlapped proteins between two runs were derived by multiple factors including carryover effects or LC memory effects between the two runs, 2DE caused factors such as neighbor-spot protein contamination, and biochemical reasons such as protein speciation. But it can be concluded that most of the proteins (72–89%) were only present in a single one spot.

(iv) Sensitivity of the readout, the mass spectrometer. The new, high sensitivity OrbiTrap mass spectrometer significantly improved the detection of lower abundance proteins than our 10-year-old QTOF and MALDI-TOF-TOF mass spectrometers although the highly sensitive mass spectrometers would result in the problem of noise that proteomics is facing now with. In order to maximally overcome the noise problem, a reasonable signal to noise (S/N) ratio, and at least two statistically significant unique peptides were used for identification of a protein. If noise is the reason for the high number of proteins per spot, the identified proteins should be mostly high abundant proteins. Therefore, we tried to determine the amount of low abundant proteins within the identified proteins. For that a human protein abundance data base that contains 7309 proteins identified with MS/MS which has been established in human osteosarcoma tissue culture cell line U2OS in 2011 [43] was used. The proteins identified with the criterion of at least 1, 2, and 3 unique peptides from the pool of three matched spots in selected glioblastoma gel spots 3, 7, 8, 16, 22, 31, and 46 were evaluated with the U2OS protein abundance data [43] (Table 8). An average of 36% (135 proteins per spot) of glioblastoma proteins with at least one unique peptide were found in U2OS protein abundance data, 45% (115 proteins per spot) of glioblastoma proteins with at least two unique peptides were found in U2OS protein abundance data, 49% (93 proteins per spot) of glioblastoma proteins with at least three unique peptides were found in U2OS protein abundance data. Those verified proteins are mainly due to multiple factors: (a) Many different protein species between human glioblastoma tissue and osteosarcoma tissue culture cell line U2OS can be expected. (b) U2OS protein abundance data were established in 2011, while glioblastoma proteins were analyzed in 2016 with higher sensitivity MS. Anyway, around 45% (115 proteins per spot) of glioblastoma proteins per spot with at least two unique peptides were found in human U2OS protein abundance data. And, important for the noise question, a detailed analysis of predicted protein copies found most of the proteins were low to moderate abundance proteins (Table 9). Therefore, those proteins identified in glioblastoma gel spots do not represent noise but mainly real proteins.

Many proteins in a gel spot are a real situation, which is also evidenced by a previous study [21]. However, even if the gels look nice to the eye, there may be invisible artifacts. Due to a part of the proteins not focused or sticking to the gel matrix, degradation because of the high concentrations, possibly incomplete reduction/alkylation due to a depletion effect, incomplete reduction and alkylation, and still interaction of proteins the found protein molecules may not represent the in vivo situation. However, these problems exist in all 2DE procedures and several of them also in LC analyses, and are very difficult to avoid. The goal of this study is to elucidate how many proteins on earth are identified in an excised 2DE gel spot. It clearly demonstrates the reality that many proteins exist in a 2DE gel spot.

Furthermore, these data lead to an in-depth consideration of 2DE-based proteomics: (i) The common method, which selects the first ranked MS-identified protein in a 2DE gel spot, actually loses much protein information. (ii) The accuracy of 2DE-based comparative proteomics to identify differentially abundant proteins by optical density measurements was challenged because it cannot be decided which protein is present in different amount when many proteins are contained in the differential 2DE gel spot. Thus, other validation methods such as Western blot are necessary. (iii) When 2DE is looked as preseparation technique (multiple but not one protein in a 2DE

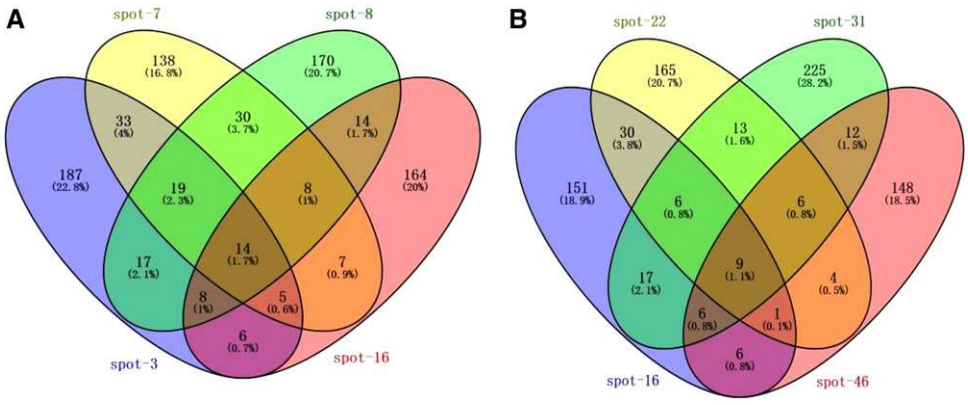

Figure 3. Overlap analysis of proteins identified between sequentially LC runs of samples. A. Venn diagrams of sequentially runs of sample spots 3, 7, 8 and 16. B. Venn diagrams of sequentially runs of sample spots 16, 22, 31, 46.

Table 8. Prediction of proteins identified from the pool of three matched spots in selected glioblastoma tissue gel spots, with human osteosarcoma cell line U2OS protein abundance data [43]

Spot No.	MS-identified proteins (Uni_pep≥1)	Predicted proteins with U2OS protein abundance data*	Ratio of predicted proteins to identified proteins	MS-identified proteins (Uni_pep≥2)	Predicted proteins with U2OS protein abundance data*	Ratio of predicted proteins to identified proteins	MS-idnetified proteins (Uni_pep≥3)	Predicted proteins with U2OS protein abundance data*	Ratio of predicted proteins to identified proteins
3	424	149	35%	289	130	45%	231	110	48%
7	356	134	37%	254	112	44%	182	88	48%
8	408	131	32%	280	113	40%	205	87	42%
16	329	120	36%	226	105	46%	175	86	49%
22	352	136	38%	234	109	46%	162	85	52%
31	410	175	42%	294	155	53%	222	132	59%
46	296	104	35%	192	81	43%	136	69	50%
Average	367	135	36%	252	115	45%	187	93	49%

Note: Uni_pep = the number of unique peptides. Proteins were analyzed by OrbiTrap Velos and MASCOT, and identified with the criterion of at least one, two and three unique peptides and significance $p < 0.05$. *The MS-identified proteins were predicted with 7309 protein abundance data in human osteosarcoma tissue culture cell line U2OS [43].

spot) to construct a 2DE reference map of the proteome, 2DE coupled with super-high sensitivity mass spectrometer such as OrbiTrap Velos can significantly increase the coverage of that proteome. For the present study, protein abundance analysis revealed most identified proteins in each analyzed spot were of low or of extremely low abundance. (iv) 2DE is a very good method to visualize the variants/isoforms [8, 27, 44, 45] or protein species [21] of a given protein in a complex human proteome; for example, 24 growth hormone protein species were present in a 2DE map of pituitary tissues [8, 44, 45], and 29 vimentin protein species in a 2DE map of HeLa cells [21]. The present study revealed 802 proteins in glioblastoma occurred in two or more spots with a total of randomly analyzed 44 spots, and 157 proteins in pituitary adenoma tissue occurred in two or more spots with a total of randomly analyzed ten spots. Furthermore, according to our result with an average of over 50 proteins per spot for glioblastoma and pituitary adenoma analyses, one can speculate that about 500 000 different protein species could be quantified with SILAC-2DE-LC/MS within one gel [21] with a resolution power of 10 000 spots per gel with 30 × 40 cm large gels [3]. In contrast, bottom-up 2DLC-MS/MS combination can only identify maximally 20 000 proteins, which is limited by the human genome and the impossibility to reach the protein species level. Thus, isotopic labeling combined with 2DE-LC/MS shows its strong power to detect and quantify protein species in a human proteome.

Moreover, the present finding changed the traditional concept that the number of accessible proteins is an order of magnitude lower in 2DE-MS as compared with bottom-up LC-MS [46]. The reason of that traditional concept is mainly addressed to inaccessible low abundance proteins, low Mr proteins, high Mr proteins, and very basic and very acidic proteins in 2DE-MS relative to bottom-up ICAT-LC-MS [47] and only one to two proteins contained in a 2DE spot [5–19]. This argument is found in many publications and is surely a wrong statement because at least low Mr proteins are better accessible by 2DE-MS [48]. Additionally, in the present investigation we clearly reveal that with a higher sensitivity MS also proteins with low abundance ($0.01 \leq$ emPAI < 0.1) and extreme pI values (range from pI 3.77 to 12.27) can be accessed (Table 4, Supporting Information Tables 3 and 4). Therefore, 2DE coupled with highly sensitive LC-MS/MS is a very strong approach to maximize the number of accessible proteins and even protein species in a human proteome.

Table 9. The number of proteins in each predicted protein abundance range, which is predicted with human osteosarcoma cell line U2OS protein abundance data [43], for the selected glioblastoma tissue gel spots with at least one, two and three unique peptides identified with OrbiTrap Velos MS/MS

	The number of predicted protein abundance range identified with at least one unique peptide					The number of predicted protein abundance range identified with at least two unique peptides					The number of predicted protein abundance range identified with at least three unique peptides				
Spot No.	Very low (n)	Low (n)	Moderate (n)	High (n)	Total (n)	Very low (n)	Low (n)	Moderate (n)	High (n)	Total (n)	Very low (n)	Low (n)	Moderate (n)	High (n)	Total (n)
3	30	22	41	56	149	28	19	34	49	130	23	14	30	43	110
7	17	17	52	48	134	16	14	41	41	112	11	9	34	34	88
8	14	18	52	47	131	14	17	45	37	113	13	16	30	28	87
16	15	13	42	50	120	13	11	35	46	105	12	9	27	38	86
22	9	23	48	56	136	6	18	39	46	109	4	14	29	38	85
31	23	25	70	57	175	20	21	61	53	155	17	17	50	48	132
46	8	7	50	39	104	7	7	33	34	81	5	4	27	33	69

Note: Proteins were analyzed by OrbiTrap Velos and MASCOT, and identified with the criterion of at least one, two and three unique peptides and significance $p < 0.05$. * The protein abundance was predicted with 7309 protein abundance data in human osteosarcoma tissue culture cell line U2OS [43]. Very low-abundance means < 500 copies; Low abundance means 500–5000 copies; Moderate abundance means 5000–100 000 copies; and High abundance means > 100 000 copies [43].

4 Concluding remarks

Multiple proteins (Glioblastoma: an average of at least 42 proteins/spot; Pituitary adenoma: an average of at least 63 proteins/spot) were present within Coomassie-stained 2DE gel spots in analyses of glioblastoma and pituitary adenoma proteomes. For conventional 2DE-based comparative proteomics of complex human tissues with optical density quantification, our results further support the finding that the accuracy of a differentially abundant protein is misleading because one cannot accurately determine which protein is present with different amount. This study identified lots of low- or extremely low-abundance proteins in each analyzed 2DE spot, and discovered that 802 glioblastoma proteins occurred in two or more spots with randomly analyzed 44 spots and 157 pituitary adenoma proteins occurring in two or more spots with randomly analyzed ten spots. Moreover, isotope labeling such as SILAC coupled with 2DE-LC/MS as discussed above shows a super high power in detection and quantification of different proteins in a complex proteome, and 2DE coupled with immunoblot and mass spectrometry is a good choice for analysis of protein variants/isoforms or protein species of a given protein. Therefore, 2DE coupled with isotope labeling and a high sensitivity mass spectrometer has a much stronger power in identification and quantification of protein variants, protein species, and low abundance proteins of the human tissue proteome as estimated before our investigation.

The MS data have been deposited to the ProteomeXchange Consortium via the PRIDE [49] partner repository with the dataset identifier PXD004930. The spot number in the present article corresponding to the spot number in PRIDE Archive is listed in Supporting Information Table 9.

The authors acknowledge financial support from the National Natural Science Foundation of China (Grant No. 81572278 and 81272798 to X.Z.), the Hunan Provincial Natural Science Foundation of China (Grant No. 14JJ7008 to X. Z.), and the Xiangya Hospital Funds for Talent Introduction (to X.Z.), and China "863" Plan Project (Grant No. 2014AA020610-1 to X. Z.).

The authors have declared no conflict of interest.

X.Z. conceived the concept, designed experiments and the entire manuscript, instructed experiments, analyzed data, supervised results, wrote and revised manuscript, and was responsible for its financial supports and the corresponding works. H.Y., F.P., Y.M, Y.L., T.C., Y.H., M.L., and N.L. participated in partial experiments and data analysis. J.L., M.L., and J.L. participated in mass spectrometry experiments and database searching. Z.L. collected tumor tissue samples. P.R.J contributed to the concept, the evaluation of the data, the design and the revision of the manuscript. All authors approved the final manuscript.

5 References

[1] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[2] Görg, A., Obermaier, C., Boguth, C., Harder, A., Scheibe, B., Wildgruber, R., Weiss, W., *Electrophoresis* 2000, *21*,1037–1053.

[3] Klose, J., Kobalz, U., *Electrophoresis* 1995, *16*, 1034–1059.

[4] Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. L., Humphery-Smith, I., *Electrophoresis* 1995, *16*, 1090–1094.

[5] Rabilloud, T., *Proteomics* 2013, *13*, 2065–2068.

[6] Campostrini, N., Areces, L. B., Rappsilber, J., Pietrogrande, M. C., Dondi, F., Pastorino, F., Ponzoni, M., Righetti, P. G., *Proteomics* 2005, *5*, 2385–2395.

[7] Zhan, X., Desiderio, D. M., *Proteomics* 2003, *3*, 699–713.

[8] Moreno, C. S., Evans, C. O., Zhan, X., Okor, M., Desiderio, D. M., Oyesiku, N. M., *Cancer Res.* 2005, *65*, 10214–10222.

[9] Zhan, X., Desiderio, D. M., Wang, X., Zhan, X., Guo, T., Li, M., Peng, F., Chen, X., Yang, H., Zhang, P., Li, X., Chen, X., *Electrophoresis* 2014, *35*, 2184–2194.

[10] Zhan, X., Desiderio, D. M., *Clin. Chem.* 2003, *49*, 1740–1751.

[11] Liu, J., Zhan, X., Li, M., Li, G., Zhang, P., Xiao, Z., Shao, M., Peng, F., Hu, R., Chen, Z., *BMC Med. Genomics* 2012, *5*, 62.

[12] Wang, X., Guo, T., Peng, F., Long, Y., Mu, Y., Yang, H., Ye, N., Li, X., Zhan, X., *Electrophoresis* 2015, *36*, 1289–1304.

[13] Karsani, S. A., Saihen, N. A., Zain, R. B., Cheong, S. C., Rahman, M. A., *Proteome Sci.* 2014, *12*, 3.

[14] Yan, J. X., Devenish, A. T., Wait, R., Stone, T., Lewis, S., Fowler, S., *Proteomics* 2002, *2*, 1682–1698.

[15] Sanchez, J. C., Chiappe, D., Converset, V., Hoogland, C., Binz, P. A., Paesano, S., Appel, R. D., Wang, S., Sennitt, M., Nolan, A., Cawthorne, M. A., Hochstrasser, D. F., *Proteomics* 2001, *1*, 136–163.

[16] Jung, E., Hoogland, C., Chiappe, D., Sanchez, J. C., Hochstrasser, D. F., *Electrophoresis* 2000, *21*, 3483–3487.

[17] Reymond, M. A., Sanchez, J. C., Hughes, G. J., Riese, J., Tortola, S., Peinado, M. A., Kirchner, T., Hohenberger, W., Hochstrasser, D. F., Kockerling, F., *Electrophoresis* 1997, *18*, 2842–2848.

[18] Sarto, C., Marocchi, A., Sanchez, J. C., Giannone, D., Frutiger, S., Golaz, O., Wilkins, M. R., Doro, G., Cappellano, F., Hughes, G. J., Hochstrasser, D. F., Mocarelli, P., *Electrophoresis* 1997, *18*, 599–604.

[19] Hoogland, C., Mostaguir, K., Appel, R. D., Lisacek, F., *J. Proteomics* 2008, *71*, 245–248.

[20] Lim, H, Eng, J, Yates, J. R. 3rd, Tollaksen, S. L., Giometti, C. S., Holden, J. F., Adams, M. W., Reich, C. I., Olsen, G. J., Hays, L. G., *J. Am. Soc. Mass Spectrom.* 2003, *14*, 957–970.

[21] Thiede, B., Koehler, C. J., Strozynski, M., Treumann, A., Stein, R., Zimny-Arndt, U., Schmid, M., Jungblut, P. R., *Mol. Cell. Proteomics* 2013, *12*, 529–538.

[22] Jungblut, P. R., Zimny-Arndt, U., Zeindl-Eberhart, E., Stulik, J., Koupilova, K., Pleissner, K. P., Otto, A., Müller, E. C., Sokolowska-Köhler, W., Grabher, G., Stöffler, G., *Electrophoresis* 1999, *20*, 2100–2110.

[23] Murphy, S., Dowling, P., Ohlendieck, K., *Proteomes.* 2016, *4*, pii: E27.

[24] Agresta, A. M., De Palma, A., Bardoni, A., Salvini, R., Iadarola, P., Mauri, P. L., *Proteomics Clin. Appl.* 2016, *0*, 457–469.

[25] Zhan, X., Wang, X., Cheng, T., *Front Endocrinol. (Lausanne)* 2016, *7*, 54.

[26] Butterfield, D. A., Gu, L., Di Domenico, F., Robinson, R. A., *Mass Spectrom. Rev.* 2014, *33*, 277–301.

[27] Rogowska-Wrzesinska, A., Le Bihan, M. C., Thaysen-Andersen, M., Roepstorff, P., *J. Proteomics* 2013, *88*,4–13.

[28] Jungblut, P. R., Thiede, B., Schlüter, H., *J. Proteomics* 2016, *134*, 1–4.

[29] Schlüter, H., Apweiler, R., Holzhütter, H. G., Jungblut, P. R., *Chem. Cent. J.* 2009, *3*, 11.

[30] Jungblut, P. R., Holzhütter, H. G., Apweiler, R., Schlüter, H., *Chem. Cent. J.* 2008, *2*, 16.

[31] Jungblut, P. R., *J. Proteomics* 2014, *107*, 98–102.

[32] Peng, F., Li, J., Guo, T., Yang, H., Li, M., Sang, S., Li, X., Desiderio, D. M., Zhan, X., *J. Am. Soc. Mass Spectrom.* 2015, *26*, 2062–2076.

[33] Zhan, X., Desiderio, D. M., *Electrophoresis* 2003, *24*, 1834–1846.

[34] Zhan, X., Desiderio, D. M., *Electrophoresis* 2003, *24*, 1818–1833.

[35] Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., Mann, M., *Mol. Cell. Proteomics* 2005, *4*, 1265–1272.

[36] Rappsilber, J., Ryder, U., Lamond, A. I., Mann, M., *Genome Res.* 2002, *12*: 1231–1245.

[37] Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S.

R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J., Szustakowki, J., International Human Genome Sequencing Consortium, *Nature* 2001, *409*, 860–921.

[38] 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.

[39] Kelleher, N. L., *J. Am. Soc. Mass Spectrom.* 2012, *23*, 1617–1624.

[40] Swindell, W. R., Remmer, H. A., Sarkar, M. K., Xing, X., Barnes, D. H., Wolterink, L., Voorhees, J. J., Nair, R. P., Johnston, A., Elder, J. T., Gudjonsson, J. E, *Genome Med.* 2015, *7*, 86.

[41] Mehdi, A. M., Patrick, R., Bailey, T. L., Bodén, M., *Mol. Cell. Proteomics* 2014, *13*, 1330–1340.

[42] Schmidt, F., Schmid, M., Jungblut, P. R., *J. Am. Soc. Mass Spectrom.* 2003, *14*, 943–956.

[43] 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.

[44] Zhan, X., Giorgianni, F., Desiderio, D. M., *Proteomics* 2005, *5*, 1228–1241.

[45] Kohler, M., Thomas, A., Püschel, K., Schänzer, W., Thevis, M., *J. Proteome Res.* 2009, *8*, 1071–1076.

[46] Han, D. K, Eng, J., Zhou, H., Aebersold, R., *Nat. Biotechnol.* 2001, *19*, 946–951.

[47] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., Aebersold, R., *Nat. Biotechnol.* 1999, *17*, 994–999.

[48] Schmidt, F., Donahoe, S., Hagens, K., Mattow, J., Schaible, U. E., Kaufmann, S. H., Aebersold, R., Jungblut, P. R., *Mol. Cell Proteomics* 2004, *3*, 24–42.

[49] Vizcaíno, J. A., Csordas, A., del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q. W., Wang, R., Hermjakob, H., *Nucleic Acids Res.* 2016, *44*, D447–D456.

